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.

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) with the already widely deployed WiFi APs is very attractive to operators and a key aspect of their strategy. In fact, WiFi access points (APs) along with femtocells are projected to carry over 60% of all the global data traffic by 2015. 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.

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 $\alpha$. 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. 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], 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.

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

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

**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. Much progress has been made in modeling association areas of Poisson Voronoi (PV), 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.

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 \(k\)th tier of the \(m\)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., \(\mathcal{V}_{m} = \{k : \lambda_{mk} + \lambda_{nk} \neq 0\}\). 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 \(\mathcal{V}_{m}^{o} \triangleq \bigcup_{m=1}^{M} \bigcup_{k \in \mathcal{V}_{m}^{o}} (m, k)\). 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 \(P_{mk}\) over bandwidth \(W_{mk}\). 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 \(x\) is \(P_{mk} h x^{-\alpha_{k}}\) where \(h\) is the channel power gain. The random
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

$$ (i, j) = \arg \max_{(m, k) \in \mathcal{V}^o} T_{mk} Z_{mk}^{r_m}, $$

(1)

where $T_{mk}$ 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 $T_{1k} \gg T_{2k}$, then more traffic is routed through RAT-1 as compared to RAT-2. Special cases for the choice of association weights, $T_{mk}$, include:

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

Note that “≡” 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 [X]. For notational brevity the following normalized parameters are defined

$$ T_{mk} \triangleq \frac{T_{mk}}{T_{ij}}, \quad \hat{P}_{mk} \triangleq \frac{P_{mk}}{P_{ij}}, \quad \hat{B}_{mk} \triangleq \frac{B_{mk}}{B_{ij}}, \quad \hat{\alpha}_k \triangleq \frac{\alpha_k}{\alpha_j}. $$

The association model described above leads to the formation of association areas 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

$$ C_{x_{ij}} = \left\{ y \in \mathbb{R}^2 : \|y - x\| \leq \left( \frac{T_{ij}}{T_{mk}} \right)^{1/\alpha_j} \|y - Z_{mk}\|^{\alpha_j/\alpha_i} \right\} \quad \forall (m, k) \in \mathcal{V}^o. $$

(2)

The readers familiar with the field of spatial tessellations would recognize that the random tessellation formed by the collection $\{C_{x_{ij}}\}$ 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 $T_{23} = 20$ dB and $T_{13} = 10$ 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

$$ R_{ij} = \frac{W_{ij}}{N_{ij}} \log (1 + \text{SINR}_{ij}), $$

(3)

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

<table>
<thead>
<tr>
<th>Table I: Notation Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Notation</strong></td>
</tr>
<tr>
<td>$M$</td>
</tr>
<tr>
<td>$K$</td>
</tr>
<tr>
<td>$(m, k)$</td>
</tr>
<tr>
<td>$\mathcal{V}^o$</td>
</tr>
<tr>
<td>$\Phi_{mk} ; \Phi_{mk}^o ; \Phi_u$</td>
</tr>
<tr>
<td>$\lambda_{mk} ; \lambda_{mk}^o ; \lambda_u$</td>
</tr>
<tr>
<td>$T_{mk} ; T_{mk}^o$</td>
</tr>
<tr>
<td>$P_{mk} ; P_{mk}^o$</td>
</tr>
<tr>
<td>$B_{mk} ; B_{mk}^o$</td>
</tr>
<tr>
<td>$\alpha_j ; \alpha_k$</td>
</tr>
<tr>
<td>$\sigma_m^2$</td>
</tr>
<tr>
<td>$W_{mk}$</td>
</tr>
<tr>
<td>$\tau_{mk}$</td>
</tr>
<tr>
<td>$\rho_{mk}$</td>
</tr>
<tr>
<td>$N_{mk}$</td>
</tr>
<tr>
<td>$C_{mk} ; C$</td>
</tr>
<tr>
<td>$S_{mk} ; S$</td>
</tr>
<tr>
<td>$R_{mk} ; R$</td>
</tr>
</tbody>
</table>
A. Load Characterization

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

\[ G_{N_{ij}}(z) = \mathbb{E} \left[ \exp \left( \lambda_u C_{ij} (z - 1) \right) \right], \]

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 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 \(C_{0ij}\) 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

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

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

\[ A_{ij} = 2\pi \lambda_i \int_0^\infty z \exp \left( -\pi \sum_{(m,k) \in \mathcal{V}_o} G_{ij}(m,k) z^{2/\alpha_k} \right) dz, \]

where

\[ G_{ij}(m,k) = \lambda_{mk} \hat{T}_{mk}^{2/\alpha_k}. \]

If \(\alpha_k \equiv \alpha\), then the association probability is simplified to

\[ A_{ij} = \frac{\lambda_{ij}}{\sum_{(m,k) \in \mathcal{V}_o} G_{ij}(m,k)}. \]

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 \(i^{th}\) RAT is given by \(A_i = \sum_{j \in \mathcal{V}_o} A_{ij}\). This probability is also the average fraction of the traffic offloaded, referred henceforth as traffic offload fraction, to the \(i^{th}\) RAT.

**Remark 2.** \(A_{ij}\) 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 \(\frac{1}{A_{ij}}\). 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 \(j^{th}\) tier of the \(i^{th}\) RAT can be approximated as

\[ C_{ij} = C \left( \frac{\lambda_i}{A_{ij}} \right), \]

where \(\alpha_k\) is the average fraction of the traffic offloaded, referred henceforth as traffic offload fraction, to the \(i^{th}\) RAT.
\[ S = \sum_{(i,j) \in V^o} 2\pi \lambda_{ij} \int_0^\infty y \exp \left( -\frac{\tau_{ij}}{\text{SNR}_{ij}(y)} - \pi \left( \sum_{k \in V_i} D_{ij}(k, \tau_{ij}) y^{2/\alpha_k} + \sum_{(m,k) \in V^o} G_{ij}(m,k) y^{2/\alpha_k} \right) \right) dy, \]  

(19)

where \( C(y) \) is the area of a typical cell of a Poisson Voronoi of density \( y \) (a scale parameter).

**Remark 3.** The approximation is trivially exact for a single tier, single RAT scenario, i.e. for \( |V| = 1 \).

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

\[ C \left( \frac{\lambda_{ij}}{A_{ij}} \right) = C \left( \frac{\sum_{(m,k) \in V^o} \lambda_{mk}}{A_{ij}} \right). \]

(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 \( \sum_{(m,k) \in V^o} \lambda_{mk} \).

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

\[ f_{C_{ij}}(c) = \frac{3.5 \lambda_{ij}}{\Gamma(3.5) A_{ij}} \lambda_{ij}^2 \exp \left( -3.5 \lambda_i c \right), \]

(11)

where \( \Gamma(x) = \int_0^\infty \exp(-t) t^{x-1} dt \) 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 other – apart from the typical – users (\( N_{o,ij} \)) associated with the tagged AP need to be characterized.

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

\[ G_{N_{o,ij}}(z) = 3.5^{4.5} \left( 3.5 + \frac{\lambda_{u\ij}}{\lambda_{ij}} (1 - z) \right)^{-4.5}. \]

(12)

Furthermore, the moments of \( N_{o,ij} \) are given by

\[ E \left[ N_{o,ij}^k \right] = \sum_{k=1}^{\infty} \left( \frac{\lambda_{u\ij}}{\lambda_{ij}} \right)^k S(n,k) E \left[ C^{k+1}(1) \right], \]

(13)

where \( S(n,k) \) are Stirling numbers of the second kind\(^5\).

**Proof:** See Appendix E.

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

\(^5\)The notation of Stirling numbers given by \( S(n,k) \) should not be confused with that of SINR coverage, \( S \).
The rate coverage expression for the most general setting requires a single numerical integral and use of lookup tables for Z and Γ. Since both the terms \(P(N_{ij} = n)\) and \(S_{ij}(t(n))\) 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, \(N_{\text{max}}\). We found \(N_{\text{max}} = 4\lambda_u/k\) 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 (22) (at the top of page), where

\[ N_{ij} = \mathbb{E}[N_{ij}] = 1 + \frac{1.28\lambda_u A_{ij}}{\lambda_{ij}}. \]

**Proof:** Using (3), the first moment of load as \(\mathbb{E}[N_{ij}]/\mathbb{E}[C^2(1)]\) where \(\mathbb{E}[C^2(1)] = 1.28\) (20). Using an approximation for (22) with \(\mathbb{E}[S_{ij}(t|\rho_{ij} N_{ij})]\), 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 \((\alpha_k \equiv 1)\), the rate coverage is

\[ \mathcal{R} = \sum_{(i,j) \in V^o} \sum_{k \in V_i} I_{ij}(k, t|\rho_{ij} N_{ij}) + \sum_{(m,k) \in V^o} G_{ij}(m,k). \]
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 $\lambda$. 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 $\rho$. 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 $\lambda$, 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 $T_{mk} = P_{mk}B_{mk}$ corresponding to the biased received power based association $\{2\}$, where $B_{mk}$ is the association bias for $(m, k)$. The effective resources at an AP are assumed to be uniformly $W_{mk} = 10$ 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 $B_{11} = 0$ 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, $V = \{(1, 1); (2, 3); (2, 3')\}$, $\lambda_{11} = 1$ 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 rate distribution in a two-RAT three-tier setting with $V = \{(1, 1); (1, 2); (2, 2); (2, 3)\}$, $\lambda_{11} = 1$ 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, $B_{12} = B_{22} = 5$ 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 $\{25\}$. 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 $(2, 3)$, $V = \{(1, 1); (2, 3)\}$ 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 $T_{mk} = P_{mk}B_{mk}$. For general settings, the optimum association biases $\{B_{mk}\}$ 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 $q^{th}$ tier of RAT-1 overlaid with $r^{th}$ tier of RAT-2, i.e. $V = \{(1, q); (2, r)\}$. 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.

A. Offloading for Optimal SIR Coverage

Proposition 1. Ignoring thermal noise (interference limited scenario, \(\sigma^2 \rightarrow 0\)), assuming equal path loss coefficients \((\hat{\alpha}_k \equiv 1)\), the value of association bias \(\frac{B_{1q}}{B_{1q}^*}\) maximizing SIR coverage is

\[
b_{\text{opt}} = \frac{P_{1q}}{P_{2r}} \left( \frac{Z(\tau_{1q}, \alpha, 1)}{aZ(\tau_{2r}, \alpha, 1)} \right)^{\alpha/2}, \tag{28}
\]

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

\[
A_2 = \frac{Z(\tau_{1q}, \alpha, 1)}{Z(\tau_{2r}, \alpha, 1) + Z(\tau_{1q}, \alpha, 1)}. \tag{29}
\]

The corresponding SIR coverage is

\[
Z(\tau_{2r}, \alpha, 1) + Z(\tau_{1q}, \alpha, 1) + Z(\tau_{2r}, \alpha, 1)Z(\tau_{1q}, \alpha, 1). \tag{30}
\]

Proof: See Appendix

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 \(Z\) 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 2) is used.

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

\[
b_{\text{opt}} = \arg \max_b \left\{ \left( Z(t_q(\hat{\rho}_{1q}, \tilde{N}_{1q}), \alpha, 1) + 1 + a(\hat{P}_{2r}, b)^{2/\alpha} \right)^{-1} \right\}, \tag{31}
\]

where \(a = \lambda_{2r}/\lambda_{1q}\) and \(b = B_{2r}/B_{1q}\).

Proof: The optimum association bias can be found by maximizing the expression obtained from Corollary 2 using \(V = \{(1, q); (2, r)\}\), \(\lambda_{2r} = a\lambda_{1q}\), and \(B_{2r} = bB_{1q}\).

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. \(V = \{(1, 1); (2, 3)\}\). 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 $\tau_{11} = 2$ dB and $\tau_{23} = 6$ dB. It is clear that for any fixed value of association bias, $S$ 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, $S$ is sub-optimal for all density values except for the optimum density, $\lambda_{\text{opt}} = P_{11} - 2 / P_{23} - 2_{23}^{2/\alpha}(Z_{(a_{11}, a_{11})}) / Z_{(a_{23}, a_{23})}$

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 $\lambda_n = 200$ users/km², the rate threshold $\rho_{mk} = 256$ Kbps, the effective bandwidth $W_{mk} = 10$ 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 5th percentile rate $\rho_{50}$, with $\mathcal{R}_{|\rho_{50}} = 0.95$ (i.e. 95% of the user population receives a rate greater than $\rho_{50}$) 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 5th percentile rate for a setting with $\mathcal{V} = (\{(1,1);(1,2);(2,2);(2,3)\}$, $\lambda_{11} = 1 \text{ BS/km}^2, \lambda_{12} = \lambda_{12} = 5 \text{ BS/km}^2, B_{12} = B_{23} = 5 \text{ dB}$ as a function of association bias of $(2,3)$. It can be seen that the choice of association weights 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: $W_{11} = 15$ MHz, $W_{23} = 5$ MHz, $\rho_{11} = 256$ Kbps and $\rho_{23} = 512$ Kbps.
- Case 2: $W_{11} = 5$ MHz, $W_{23} = 15$ 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 ($W_{ij}$), 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.

VI. Conclusion

In this paper, we presented a tractable model to analyze the effect of offload 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
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 5th percentile rate with \( V = \{(1, 1); (2, 3)\} \).

Fig. 10: Effect of association bias for third tier of RAT-2 APs on 5th percentile rate with \( \lambda_{12} = \lambda_{22} = 5 \lambda_{11} \), \( B_{12} = B_{22} = 5 \) dB.

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

The optimum offloading bias and traffic offload fraction can 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 $A_{ij}$ is the association probability of a typical user with RA T-tier pair of $(i, j)$, then

$$A_{ij} = \mathbb{P}\left( \bigcap_{(m,k) \in \mathcal{V}^n \setminus \{(i,j)\}} \left( T_{ij} Z_{ij}^{-\alpha i} > T_{mk} Z_{mk}^{-\alpha k} \right) \right), \quad (32)$$

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

$$A_{ij} = \prod_{(m,k) \in \mathcal{V}^n \setminus \{(i,j)\}} \mathbb{P}\left( T_{ij} Z_{ij}^{-\alpha i} > T_{mk} Z_{mk}^{-\alpha k} \right), \quad (33)$$

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

$$P(Z_{mk} > z) = \mathbb{P}(\Phi_{mk} \cap B(0, z) = \emptyset) = e^{-\pi \lambda \text{area}}, \quad (35)$$

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

$$f_{Z_{mk}}(z) = \frac{d}{dz}\left(1 - \mathbb{P}(Z_{mk} > z)\right) = 2\pi \lambda \text{area} \exp(-\pi \lambda \text{area} z^2), \quad \forall z \geq 0. \quad (36)$$

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

$$A_{ij} = 2\pi \lambda \text{area} \int z \exp(-\pi \lambda \text{area} z^2) \, dz, \quad (37)$$

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

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

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 $\Phi_a$ which is the same as the original distribution since $\Phi_a$ is a PPP [19, Sec. 4.4]. Thus, using Lemma 1 and (39), the PGF of the other users in the tagged AP is

$$G_{N_{ij}}(c) = \mathbb{E}\left[ \exp\left( \lambda u C'_{ij}(c-1) \right) \right] = \int_{c>0} \exp(\lambda u(c-1)) \frac{3.5^{3.5} \lambda_{ij} \left( \lambda_{ij}/\lambda_{ij} \right)^{3.5} \exp\left(-3.5 \lambda_{ij}/\lambda_{ij} \right)}{\mathbb{E}[C_{ij}]} \, dc \quad (40)$$

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

$$P(N_{ij} = n + 1) = P(N_{ij} = n) = \frac{G_{N_{ij}}^{(n)}(0)}{n!} = \frac{3.5^{3.5} \Gamma(n + 4.5) \left( \lambda u A_{ij}/\lambda_{ij} \right)^n \exp\left(3.5 + \lambda u A_{ij}/\lambda_{ij}\right)}{\Gamma(4.5)} \quad (42)$$

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

$$\mathbb{E}\left[N_{ij}^n\right] = \mathbb{E}\left[\mathbb{E}\left[N_{ij}^n | C'_{ij}\right]\right] = \mathbb{E}\left[\sum_{k=0}^{n} (\lambda u C'_{ij})^k S(n, k)\right] = \sum_{k=1}^{n} \lambda^k \mathbb{E}\left[C'_{ij}^k\right]. \quad (43)$$

Using (39) and the area approximation (9)

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

and thus

$$\mathbb{E}\left[N_{ij}^n\right] = \sum_{k=1}^{n} \left(\lambda u A_{ij}/\lambda_{ij}\right)^k S(n, k) \mathbb{E}\left[C_{ij}^{k+1}\right]. \quad (46)$$

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

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

$$= P(Z_{ij} > y, \text{user is associated with } (i, j)) / P(\text{user is associated with } (i, j)). \quad (47)$$
Now using Lemma 2

\[ \mathbb{P}(Z_{ij} > y, \text{ user is associated with } (i, j)) \]

\[ = 2\pi \lambda_{ij} \int_{y}^{\infty} z \exp \left( -\pi \sum_{(m, k) \in V^o} G_{ij}(m, k) z^{2/\alpha_k} \right) \, dz. \]  

(48)

Using (47) and (48) we get

\[ \mathbb{P}(Y_{ij} > y) \]

\[ = 2\pi \lambda_{ij} \int_{y}^{\infty} z \exp \left( -\pi \sum_{(m, k) \in V^o} G_{ij}(m, k) z^{2/\alpha_k} \right) \, dz, \]  

(49)

which leads to the PDF of \( Y_{ij} \)

\[ f_{Y_{ij}}(y) = \frac{2\pi \lambda_{ij}}{\mathcal{A}_{ij}} y \exp \left( -\pi \sum_{(m, k) \in V^o} G_{ij}(m, k) y^{2/\alpha_k} \right). \]  

(50)

### APPENDIX D

**Proof of Lemma 3** The SINR coverage of a user associated with an AP of \((i, j)\) is

\[ S_{ij}(\tau_{ij}) = \int_{y>0} \mathbb{P}(\text{SINR}_{ij}(y) > \tau_{ij}) f_{Y_{ij}}(y) \, dy. \]  

(51)

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

\[ \mathbb{P} \left( \frac{P_{ij}h_{ij}y^{-\alpha_j}}{\sum_{k \in V_i} I_{ik} + \sigma_i^2} > \tau_{ij} \right) \]

\[ = \mathbb{P} \left( h_y > y^{\alpha_j} P_{ij}^{-1} \tau_{ij} \left( \sum_{k \in V_i} I_{ik} + \sigma_i^2 \right) \right) \]

\[ = \mathbb{E} \exp \left( -y^{\alpha_j} \tau_{ij}^{-1} \sum_{k \in V_i} I_{ik} + \sigma_i^2 \right) \]

\[ \overset{(a)}{=} \exp \left( -\frac{\tau_{ij}}{\text{SNR}_{ij}(y)} \right) \prod_{k \in V_i} \mathbb{E}_k \left[ \exp \left( -y^{\alpha_j} \tau_{ij} P_{ij}^{-1} I_{ik} \right) \right], \]  

(52)

\[ = \exp \left( -\frac{\tau_{ij}}{\text{SNR}_{ij}(y)} \right) \prod_{k \in V_i} M_{I_{ik}} \left( y^{\alpha_j} \tau_{ij} P_{ij}^{-1} \right), \]  

(53)

where \( \text{SNR}_{ij}(y) = \frac{P_{ij}h_{ij}y^{-\alpha_j}}{\sigma_i^2} \) and (a) follows from the independence of \( I_{ik} \) and \( M_{I_{ik}}(s) \) is the moment-generating function (MGF) of the interference. Expanding the interference term, the MGF of interference is given by

\[ M_{I_{ik}}(s) = \mathbb{E}_{\Phi_{ik}, \Phi_{ik}', h_x, h_x'} \left[ \exp \left( -s P_{ik} \left\{ \sum_{x \in \Phi_{ik} \cap o} h_x x^{-\alpha_k} + \sum_{x' \in \Phi_{ik}'} h_{x'} x'^{-\alpha_k} \right\} \right) \right], \]

(54)

\[ \overset{(a)}{=} \mathbb{E}_{\Phi_{ik}} \left[ \prod_{x \in \Phi_{ik} \cap o} M_{h_x} \left( s P_{ik} x^{-\alpha_k} \right) \right] \times \mathbb{E}_{\Phi_{ik}'} \left[ \prod_{x' \in \Phi_{ik}'} M_{h_{x'}} \left( s P_{ik} x'^{-\alpha_k} \right) \right], \]

(55)

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 unity 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

\[ T_{ij} y^{-\alpha_j} = T_{ik} z_{ik}^{-\alpha_k} \quad \text{or} \quad z_{ik} = \left( T_{ik} \right)^{1/\alpha_k} y^{1/\alpha_k}. \]  

(56)

Using change of variables with \( t = (s P_{ik})^{-2/\alpha_k} x^2 \), the integral can be simplified as

\[ \int_{z_{ik}}^{\infty} \frac{2x}{1 + (s P_{ik})^{-2/\alpha_k} x^{\alpha_k}} \, dx = (s P_{ik})^{2/\alpha_k} \int_{0}^{\infty} \frac{dt}{(t^{1/\alpha_k} + 1)^{\alpha_k/2}} \]

\[ = Z(s P_{ik}, \alpha_k), \]  

(57)

where

\[ Z(a, b, c) = a^{2/3} b \int_{0}^{\infty} \frac{du}{(u^{1/3} + 1)^{2/3}}. \]

This gives the MGF of interference

\[ M_{I_{ik}}(s) = \exp \left( -\pi (s P_{ik})^{2/\alpha_k} \right) \times \left\{ \lambda_{ik} Z \left( 1, \alpha_k, \frac{z_{ik}^{\alpha_k}}{s P_{ik}} \right) + \lambda_{ik}' Z \left( 1, \alpha_k, 0 \right) \right\}. \]  

(58)

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

\[ \mathbb{P}(\text{SINR}_{ij}(y) > \tau_{ij}) \]

\[ = \exp \left( -\frac{\tau_{ij}}{\text{SNR}_{ij}(y)} - \pi \sum_{k \in V_i} y^{2/\alpha_k} D_{ij}(k, \tau_{ij}) \right). \]  

(59)
where

\[ D_{ij}(k, \tau_{ij}) = b^{2/(\alpha k)} \left\{ \lambda_{ki} Z(\tau_{ij}, \alpha_k, k^{1+1/\alpha_k}) + \lambda_{kj} Z(\tau_{ij}, \alpha_k, 0) \right\}. \]

Using (51) along with Lemma 4 gives

\[ \mathbf{S}_{ij}(\tau_{ij}) = \frac{2\pi \lambda_{ij}}{A_{ij}} \int_{y>0} y \exp \left( -\frac{\tau_{ij}}{\text{SR}_{ij}(y)} \right) \right. 
- \left. \pi \left\{ \sum_{k \in V_i} D_{ij}(k, \tau_{ij}) y^{2/(\alpha k)} + \sum_{(m,k) \in V} G_{ij}(m,k) y^{2/(\alpha k)} \right\} dy. \]

Using law of total probability we get

\[ \mathbf{S} = \sum_{(i,j) \in V} \mathbf{S}_{ij}(\tau_{ij}) A_{ij}, \quad (67) \]

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

APPENDIX

Proof of Proposition 4

In the described setting SIR coverage can be written as

\[ \mathbf{S} = \sum_{(i,j) \in V} \frac{\lambda_{ij}}{\sum_{k \in V_i} D_{ij}(k, \tau_{ij}) + \sum_{(m,k) \in V} G_{ij}(m,k)}, \quad (68) \]

and with \( V = \{(1,q), (2,r)\} \), \( \lambda_{2r} = a\lambda_{1q} \), and \( B_{2r} = bB_{1q} \)

\[ \mathbf{S} = \frac{\lambda_{1q} Z(\tau_{1q}, \alpha, 1) + \lambda_{1q} + \lambda_{2r}(P_{2r} B_{2r})^{2/\alpha}}{\lambda_{2r} Z(\tau_{2r}, \alpha, 1) + \lambda_{2r} + \lambda_{1q}(P_{1q} B_{1q})^{2/\alpha}} \]

\[ = \frac{Z(\tau_{1q}, \alpha, 1) + 1 + a(P_{2r} b)^{2/\alpha}}{Z(\tau_{2r}, \alpha, 1) + 1 + \alpha/(a(P_{2r} b)^{2/\alpha})}. \]

The gradient of \( S \) with respect to association bias \( \nabla_b S \) is zero at

\[ b_{\text{opt}} = \arg \max_b \left\{ \left( Z(\tau_{1q}, \alpha, 1) + 1 + a(P_{2r} b)^{2/\alpha} \right)^{-1} \right. 
+ \left. \left( Z(\tau_{2r}, \alpha, 1) + 1 + \alpha/(a(P_{2r} b)^{2/\alpha}) \right)^{-1} \right\} \]

\[ = \frac{P_{1q}}{P_{2r}} \left( \frac{Z(\tau_{1q}, \alpha, 1)}{a Z(\tau_{2r}, \alpha, 1)} \right)^{\alpha/2}. \]

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

\[ A_2 = \frac{\lambda_{2r}}{G_{2r}(r)} = a \left( \frac{P_{1q}}{P_{2r} b_{\text{opt}}} \right)^{2/\alpha} = \frac{Z(\tau_{1q}, \alpha, 1)}{Z(\tau_{2r}, \alpha, 1) + Z(\tau_{1q}, \alpha, 1)}. \quad (69) \]

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


