Traffic Offloading under Outage QoS Constraint in Heterogeneous Cellular Networks

Thant Zin Oo, Nguyen H. Tran, Tuan LeAnh, S. M. Ahsan Kazmi, Tai Manh Ho, and Choong Seon Hong,
Department of Computer Engineering, Kyung Hee University, Korea,
Email: {tzoo, nguyenth, latuan, ahsankazmi, hmtai, cshong}@khu.ac.kr.

***Abstract*—Heterogeneous cellular networks offload the mobile data traffic to small cell base stations to reduce the workload on the macro base stations. Our objective is to maximize the sum rate of the down-links for the whole network under outage QoS constraint. To achieve the objective, we have to jointly solve the user association problem and resource allocation problem. We formulate the two problems into a joint optimization problem and convert it into an equivalent game theoretic formulation. We employ payoff based log linear learning and propose an algorithm that converges to one of the existing Nash equilibrium. We then provide extensive simulation results to verify the performance of our proposed algorithm.**

***Index Terms*—Heterogeneous Cellular Networks, HetNets, Interference Mitigation, User Association, Resource Allocation.**

I. Introduction

In recent years, the demand for wireless data services has grown exponentially with the proliferation of mobile devices. The mobile operators must increase their network capacities to handle the increasing demand. The vision for fifth generation (5G) wireless network is to increase capacity by 1000-times. One component of this vision is the deployment of multi-tier dense small-cell base stations (SBSs) overlaid on the existing macro-cell base stations (MBSs) as a heterogeneous cellular network (HetNet). According to Shannon-Hartley theorem, the nearer the transmitter is to the receiver, the less path loss the data transmission will suffer, and thus, will have better SINR. Smaller cell size translates into shorter transmission range where transmitter requires less power. Therefore, the interference in the network is reduced and more cells can reuse the same frequencies (sub-channels). However, HetNet introduces the disparity in cell sizes into an already complex network environment. A multi-tier HetNet may include macro base stations (MBSs), pico base stations (PBSs) and femto base stations (FBSs) where PBSs and FBSs are classified as SBSs.

As in the case with other wireless networks, the major challenges for HetNet small cells are user association (UA), resource allocation (RA) and interference mitigation (IM). The unique problem in HetNet small cells is that the number of choices or configurations increases exponentially with the number of deployed SBSs. The existing centralized management currently employed in cellular networks can no longer cope with the massive overhead in computation and signaling required by the HetNet small cells.

We propose a distributed network management scheme employing self-organization mechanisms described in [1], [2], [3]. Self-organization reduces the amount of manual intervention involved in network planning. Using self-organization, small cells can learn from their environment and autonomously adjust their configuration strategies towards achieving optimal performance. More importantly, self-organization mechanisms can be implemented distributedly without complete information and thus are scalable with network size [1]. Game theoretic approaches [1], [3], [2] and [4] are usually employed to analyze these self-organized mechanisms due to its unique ability to model the strategic interactions between competing interests of UEs and BSs.

Our contributions in this paper are as follows: First, we develop a UE traffic demand model. Second, we formulate the traffic offloading under outage QoS constraint into an optimization problem. Third, we convert the optimization problem into an equivalent game model. Fourth, we employ payoff based log linear learning and present a joint UE association and resource allocation algorithm. And finally, we perform simulations to verify the convergence of our proposed algorithm to a NE solution. The rest of the paper is organized as follows: In Section II, we present our system model. We formulate the problem in Section III. In Section IV, we propose our joint UA and RA algorithm. We present our simulation results in Section V. And finally, we conclude this paper in Section VI.

II. System Model

1) Traffic Model: We propose construct traffic model as follows: As depicted in Fig. 1, a region, $\mathcal{A} \subset \mathbb{R}^2$, is partitioned into several sub-regions (or partitions) which represent buildings and are mutually exclusive and collectively exhaustive, i.e. $\mathcal{A} = \bigcup_{r=0}^{N_A} \mathcal{A}_r$, where $\mathcal{A}_0$ represents the outdoors. For each sub-region $\mathcal{A}_r$, we assume that the user request arrivals follow a homogeneous Poisson point process with parameter κ_r per unit area. Let $\mathcal{U}_r$ denotes the set of UEs (requests) in sub-region $\mathcal{A}_r$, then the number of UEs, $|\mathcal{U}_r| = \kappa_r a_r$, where a_r denotes the area of the sub-region $\mathcal{A}_r$. And, the total number of UEs for the entire region $\mathcal{A}$, $|\mathcal{U}| = \sum_{r=0}^{N_A} |\mathcal{U}_r|$. Furthermore, let $\psi_i \in \Psi$ denotes the requested data rate (bits per second), i.e. QoS, of each UE i where Ψ is the discrete set of QoS levels.

This research was supported by Basic Science Research Program through National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2014R1A2A2A01005900).

Dr. C. S. Hong is the corresponding author.

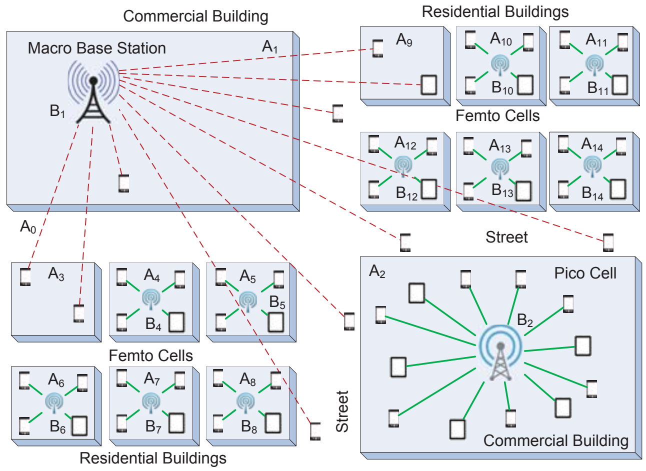


Fig. 1: Network Model

2) Base Stations: As described in Fig. 1, the region $\mathcal{A}$ is served by a single MBS and a number of SBSs consisting of PBSs and FBSs. Depending on the power level, the set of BS can be classified into: $\mathcal{B} = \mathcal{B}_\text{m} \cup \mathcal{B}_\text{s} = \mathcal{B}_\text{m} \cup \mathcal{B}_\text{p} \cup \mathcal{B}_\text{f}$. SBSs are usually positioned at the buildings (or sub-regions) with hot spots. Some buildings have no SBS and the UEs located within the buildings will be served by the MBS. The outdoors region, $\mathcal{A}_0$, is also served by MBS. We assume that each BS $j \in \mathcal{B}$ transmits with a constant power P_j watts and all BSs are connected to a high speed back haul with negligible delay (either wired or wireless). Let $\mathcal{S}$ denotes the set of available sub-channels[1] and each BS $j \in \mathcal{B}$ can access, at most $|\mathcal{S}|$ sub-channels, that will be allocated to its associated UEs.

3) Channel Model: The dynamic nature of the wireless transmission medium is captured by channel power gain. The positive channel power gain between UE i and BS j can be modeled as: $H_{ij} = G_{ij}F_{ij}, \ \forall (i,j) \in \mathcal{U} \times \mathcal{B}$, where G_{ij} embodies the large scale fading components such as path loss, log normal shadowing and antenna gains, whereas F_{ij} represents small scale multi-path Rayleigh fading. For each association epoch, G_{ij} can be regarded as a constant and F_{ij} can be modeled as statistically independent random variable which is exponentially distributed with unit variance. We assumed that each UE i is capable of measuring H_{ij} for all BSs $j \in \mathcal{B}$.

4) Spectral Efficiency and Data Rate: Let $\mathcal{B}_\mathcal{I} \subseteq \mathcal{B}$ denotes the set of interfering BSs to the link between UE i and BS j. Then, the instantaneous SINR received at UE i from BS j on sub-channel k is given as:

$$\Gamma_{ijk} = \frac{P_j H_{ij}}{\sum_{m \in \mathcal{B}_\mathcal{I}} P_m H_{im} + W N_0} \tag{1}$$

where the constant W denotes the bandwidth of the sub-channel and N_0 denotes the thermal noise spectral power[2].

Accordingly, the spectral efficiency between UE i and BS j on sub-channel k can be calculated as: $C_{ijk} = \log_2(1+\Gamma_{ijk})$. Thus, the achievable instantaneous data rate per sub-channel between UE i and BS j can be calculated as (WC_{ijk}). Then, the individual instantaneous data rate achieved between UE i and BS j can be calculated as:

$$R_{ij} = W \sum_{j \in \mathcal{B}} \sum_{k \in \mathcal{S}} x_{ij} y_{ik} C_{ijk}, \quad \forall j \in \mathcal{B}, \ \forall k \in \mathcal{S}, \tag{2}$$

where $x_{ij} \in \{0,1\}$ and $y_{ik} \in \{0,1\}$ are the binary decision variables for UA and RA, respectively.

5) Interference Mitigation and Frequency Reuse: Due to the disparity in transmission power[3], MBS causes significant interference to links between UEs and SBSs. We propose a dynamic spectrum partitioning scheme based on spectrum conflict graph and spectrum reuse graph. Let $\mathcal{B}_j^{(\mathcal{C})}$ and $\mathcal{B}_j^{(\mathcal{R})}$ denote the sets of conflicting BSs and reusing BSs for BS $j \in \mathcal{B}$. If d_{jm} denotes the distance between BS j and m and r_j and r_m denote the cell radii of BS j and m, respectively,

$$m \in \begin{cases} \mathcal{B}_j^{(\mathcal{C})}, & \text{if } d_{jm} \le r_j + r_m, \\ \mathcal{B}_j^{(\mathcal{R})}, & \text{otherwise}, \end{cases} \tag{3}$$

for $j \neq m, \ \forall j, m \in \mathcal{B}$. And the spectrum partitioning constraint for BS $j \in \mathcal{B}$ is given as:

$$\mathcal{S}_j \cap \mathcal{S}_m = \emptyset, \qquad \forall j \in \mathcal{B}, \ \forall m \in \mathcal{B}_j^{(\mathcal{C})}, \tag{4}$$

$$\left| \bigcup\nolimits_j \mathcal{S}_j \right| \le |\mathcal{S}|, \qquad \forall j \in \mathcal{B}, \tag{5}$$

where $\mathcal{S}_j$ and $\mathcal{S}_m$ denote the sub-channels assigned to BSs j and m, respectively. Thus, for BS j, $\mathcal{B}_\mathcal{I} \equiv \mathcal{B}_j^{(\mathcal{R})}$ because highest interfering BSs have already been isolated by (4). Furthermore, (5) ensures that the number of sub-channels allocated to UEs by BSs do not exceed the budget, i.e. total number of sub-channels available to the network.

6) Outage Probability and QoS Constraint: For UE i served by BS j, the individual outage event can be defined as the event where the instantaneous data rate drops below a certain threshold ψ_i. Then, for the small scale Rayleigh fading environment, the outage probability, denoted by $Q_{ij} = \Pr\{R_{ij} < \psi_i\}$, can be calculated as [5]

$$Q_{ij} = 1 - \exp\left(-\frac{WN_0}{P_j H_{ij}} \Psi_{ij}\right) \prod_{m \in \mathcal{B}_\mathcal{I}} \left(1 + \frac{P_m H_{im}}{P_j H_{ij}} \Psi_{ij}\right)^{-1}, \tag{6}$$

where $\Psi_{ij} = 2^{\frac{\psi_i}{b_{ij}}} - 1$ and $b_{ij} = \sum_{k \in \mathcal{S}} y_{ik}$.

Due to the stochastic behavior of the wireless channels, it is not possible to guarantee a constant instantaneous rate to a UE. Nevertheless, we can guarantee a certain data rate, ψ_i, with a certain probability, θ as follows:

$$\Pr\{R_{ij} \le \psi_i\} \le \theta_i. \tag{7}$$

[1]We will use sub-channel and sub-carrier interchangeably. In LTE systems, the notion of resource blocks (RBs) is used where 1 RB consists of 12×15kHz sub-carriers. A 20MHz channel contains 100 RBs (1200 sub-carriers) with only 18 MHz of usable spectrum for UEs.

[2]Thermal noise floor for 1 Hz bandwidth at room temperature ($20\,^\circ$ C) is -174 dBm and the bandwidth, $W = 15$ kHz.

[3]In practice, the typical transmit power of MBS is around 43 dBm, and that of SBSs is $20 \sim 30$ dBm lower.

III. Problem Formulation

We define the traffic offloading problem under outage QoS provisioning as sum rate maximization (**SRM**) problem:

$$
\begin{aligned}
\underset{\mathbf{x},\mathbf{y}}{\text{maximize:}} \quad & W\sum_{k\in\mathcal{S}}\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{U}} x_{ijk}C_{ijk} & \\
\text{subject to:} \quad & \sum_{j\in\mathcal{B}} x_{ij} \le 1,\ \forall i\in\mathcal{U}, & \text{(AC)} \\
& Q_{ij} \le \theta_i,\ \forall i\in\mathcal{U}, & \text{(QC)} \\
& \mathcal{S}_j \cap \mathcal{S}_m = \emptyset,\ \forall j\in\mathcal{B},\ \forall m\in\mathcal{B}_j^{(\mathcal{C})}, & \text{(IC)} \\
& \left|\bigcup\nolimits_j \mathcal{S}_j\right| \le |\mathcal{S}|,\ \forall j\in\mathcal{B}, & \text{(RC)} \\
& x_{ij}\in\{0,1\},\ y_{ik}\in\{0,1\}. & \text{(Var)}
\end{aligned}
\tag{8}
$$

The first constraint is referred to as the association constraint (AC) and it states that a UE can only be associated with at most one BS. The second constraint is the outage QoS constraint (QC) which is given in (7). The third constraint is interference constraint mentioned in (4) with dynamic spectrum partitioning scheme. The fourth constraint, referred to as resource constraint (RC), ensures that the sub-channels allocated to UEs by BSs do not exceed total number of sub-channels available to the network. Then finally, $\mathbf{x}$ (size $|\mathcal{U}|\times|\mathcal{B}|$) and $\mathbf{y}$ (size $|\mathcal{U}|\times|\mathcal{S}|$) are the variables or control knobs of the optimization problem which jointly performs UE association, resource allocation and spectrum partitioning. The elements x_{ij} and y_{ik} can only have binary values ('0' or '1'). Note that the first and second constraints are UE specific and BSs are responsible for the third and fourth constraints. This optimization problem is combinatorial and NP-hard.

IV. Game Theoretic Perspective

A. Game Theoretic Model

(**SRM**) presented in Sec. III can be modeled by the following game in normal-form as:

$$\mathcal{G} = \left(\{\mathcal{U},\mathcal{B}\}, \{\mathbf{x}_i,\mathbf{y}_i\}, \{R_{ij}\}\right),\ (i\in\mathcal{U},\ j\in\mathcal{B}), \tag{9}$$

where the payoff R_{ij} is defined in (2). Moreover,

- the players are the set of UEs and BSs, $\{\mathcal{U},\mathcal{B}\}$.
- the actions of the players, UEs and BSs,
  - UE i sends request to BS j for UA and RA, $\{\mathbf{x}_i\}_{i\in\mathcal{U}}$.
  - BS j either accepts or rejects the request, $\{\mathbf{y}_j\}_{j\in\mathcal{B}}$.
- the payoffs, R_{ij} are data rates achieved by UEs.

$\mathcal{G}$ is a non-cooperative and static game. Potential games [1], [2], [3] are a special class of non-cooperative and static games. The game is played as follows.

Firstly, every UE i measures the received power from each BS j within its transmission range from its pilot signal. From the measured received signal strength, UE i calculates the minimum required sub-channels for its traffic demand that satisfies the outage QoS constraint in (7).

Secondly, UE i sends its request to the BS j that it has chosen to associate with. The action of each UE i, $\mathbf{x}_i$ is a vector (size $|\mathcal{B}|\times 1$). When UE i sends request to BS j, its corresponding element, $x_{ij}=1$. The request contains the required data rate, the number of sub-channels needed to satisfy its QoS (7) and the time duration of the request. Note that UE i can only send one request at any one time (8)-(AC).

Thirdly, BS j will either accept or reject the traffic request from the UE i depending on its available resources (4). The acceptance of a request is indicated by the allocation of sub-channels in the form of a reply with sub-channel frequencies to UE i, i.e. $\mathbf{y}_i$ (size $|\mathcal{S}|\times 1$. The BS j will simply not reply in case of a rejection, i.e. $\mathbf{y}_i$ be a vector of zeros.

Lastly, the BS j will update the binary matrix Υ which is a two-dimensional matrix of size $|\mathcal{B}|\times|\mathcal{S}|$. Υ is used to monitor the availability of resources and to ensure the constraint (4) is not violated. The updated Υ matrix is then broadcast to other BS via the back haul.

B. Log Linear Learning

Inspired by the payoff based log-linear learning [3], we propose a joint UE association and resource allocation (UA-RA) algorithm which is presented in Alg. 1. $\omega_i(t)$ is the exploration rate of each UE i and the exploitation rate $\nu_i(t)$ is given as:

$$\nu_i(t) = \frac{\exp(\beta\ R_{ij'})}{\exp(\beta\ R_{ij'}) + \exp(\beta\ R_{ij})}. \tag{10}$$

V. Performance Evaluation

We perform simulations in Matlab to evaluate the performance of Alg. 1. We placed 20 BSs in $200\ \mathrm{m}\times 200\ \mathrm{m}$ area. The BSs constitute 1 MBS, 2 PBSs and 17 FBSs. Using PPP, we randomly created 300 UEs whose traffic demands follow a binomial distribution such that $\psi_i \in \{.05, 0.1, \cdots, 0.95, 1\}$ Mbps. For evaluation, we compare Alg. 1 with the following schemes: (i) MBS only scheme without any HetNet deployment (ii) maximum SINR scheme and (iii) cell range extension (CRE) scheme with $3\,\mathrm{dB}$ SINR bias.

We performed an experiment to test the convergence of Alg. 1. For this experiment, we randomly created $|\mathcal{U}|=300$ UEs and fixed the decay in exploration rate $\omega_{\text{step}}=0.02$. The simulation results are depicted in Fig. 2a. We subdivided the results of Alg. 1 into respective BS tiers as shown in Fig. 2a.

At the initial state, all UEs are served by the MBS, and thus, MBS data rate is higher than that of SBSs. As the time progress, UEs move to low power SBSs as displayed in Fig. 2a by the decrease in MBS data rate and the increase in PBSs and FBSs data rates. During $0 < t \le 100$, we can observe the rapid offloading of traffic from MBS to SBSs by the sharp decline in MBS data rates as displayed in Fig. 2a. As more UEs are offloaded to SBSs, more sub-carriers are reused. Note that the network topology specifies how much sub-carriers can be reused where the network must satisfy the interference constraint given in (4). During $100 < t \le 400$, we can observe high fluctuations corresponding to some UEs switching back and forth among BS tiers. During $400 < t \le 600$, the fluctuations decrease as more UEs stop their exploration. And finally, $t > 600$, Alg. 1 converges to one of the feasible configurations.

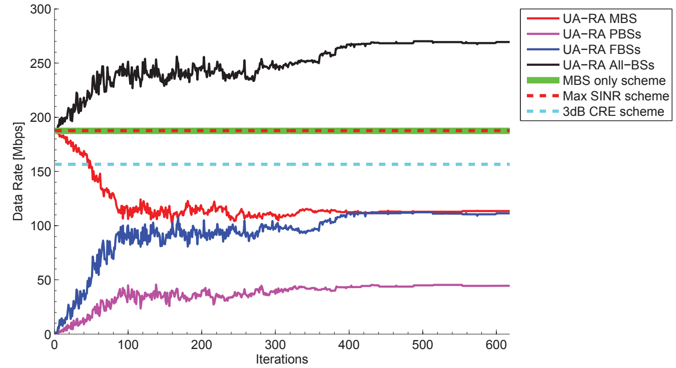
Algorithm 1: UA-RA algorithm

Let $\mathcal{U}_p \subseteq \mathcal{U}$ be set of participating UEs.
Initialization: *$\mathcal{U}_p := \mathcal{U}$ and each UE $i \in \mathcal{U}$*
  Measures pilot signals from $\forall j \in \mathcal{B}$.
  Calculates required sub-channels using (7).
  Sends request to MBS which serves requests FIFO.
  Calculates its utility R_{ij} using (2).
for *$t \in \{1, 2, \cdots, T\}$, and $\forall i \in \mathcal{U}_p$* **do**
  Exploration: *If UE i did not explore in (t),*
    With probability $\omega_i(t)$,
      UE i chooses a feasible BS $j' \neq j,\ j', j \in \mathcal{B}$.
      UE i sends request to BS j'.
      if *enough resources are available at BS j'* **then**
        BS j' find sub-channels using (4).
        BS j' will reply with sub-channels, $\mathbf{y}_i^{(j')}$.
      else
        BS j' will not reply, i.e. $\mathbf{y}_i^{(j')} = [0, 0, \cdots, 0]$.
      $\omega_i(t) = \max\{0,\ \omega_i(t) - \omega_{\text{step}}\}$
    With probability $1 - \omega_i(t)$,
      UE i stays with current BS j.
    UE i calculates its utility $R_{ij'}$ using (2).
  Exploitation: *If UE i explored in (t),*
    UE i calculates $\nu_i(t)$ using (10).
    UE i will choose either BS j' or j as follows:
      Stay with BS j' with prob. $\nu_i(t)$.
      Revert back to BS j, with prob. $1 - \nu_i(t)$.
  BS j updates resource usage bitmap, Υ.
  BS j broadcasts Υ to all BSs.
  if $\omega_i(t) = 0$ **then**
    $\mathcal{U}_p := \mathcal{U}_p \setminus \{i\}$, remove UE i from $\mathcal{U}_p$.

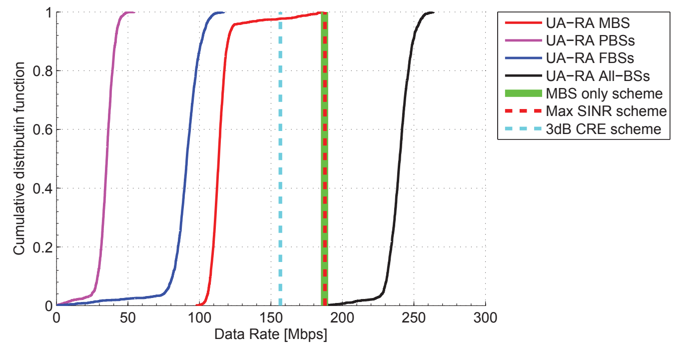
Fig. 2a also displays the three comparison schemes and Alg. 1 outperforms all three comparison schemes. Note that MBS only scheme without any HetNet deployment and max SINR scheme have similar performances. Furthermore, since all three comparison schemes are deterministic, they have no fluctuations as evidence by the unit step CDFs depicted in Fig. 2b which depicts the empirical CDF of Alg. 1 and three comparison schemes.

VI. Conclusions

In this paper, we studied about self-organization for heterogeneous cellular networks to offload user traffic from MBS. We propose a network model based on the variability of the traffic. We model the problem into sum rate maximization problem and convert it into an equivalent game theoretic model. Inspired by log linear learning, we propose a joint user association and resource allocation algorithm. Using the proposed algorithm, the network self-organizes itself to a Nash equilibrium. We perform simulations to verify the convergence of our proposal. Simulation results prove that our proposed joint UA and RA algorithm requires little overhead and will converge to an NE.



(a) Data rate versus iterations, $|\mathcal{U}| = 300$, $\omega_{\text{step}} = 0.02$.



(b) Empirical CDF of data rate, $|\mathcal{U}| = 300$

Fig. 2: Comparison of UA-RA with other schemes.

References

[1] M. Bennis, S. Perlaza, P. Blasco, Z. Han, and H. Poor, "Self-organization in small cell networks: A reinforcement learning approach," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 7, pp. 3202–3212, July 2013.

[2] J. Marden, G. Arslan, and J. Shamma, "Joint strategy fictitious play with inertia for potential games," *Automatic Control, IEEE Transactions on*, vol. 54, no. 2, pp. 208–220, Feb 2009.

[3] J. Marden and J. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, Sept 2010, pp. 1171–1172.

[4] D. Niyato and E. Hossain, "Dynamics of network selection in heterogeneous wireless networks: An evolutionary game approach," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 4, pp. 2008–2017, May 2009.

[5] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *Wireless Communications, IEEE Transactions on*, vol. 1, no. 1, pp. 46–55, Jan 2002.