IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 3, 2019, accepted December 27, 2019, date of publication January 14, 2020, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966657

Thompson Sampling-Based Channel Selection Through Density Estimation Aided by Stochastic Geometry

WANGDONG DENG, (Student Member, IEEE), SHOTARO KAMIYA, (Student Member, IEEE), KOJI YAMAMOTO, (Member, IEEE), TAKAYUKI NISHIO, (Member, IEEE), AND MASAHIRO MORIKURA, (Member, IEEE)

Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

Corresponding author: Koji Yamamoto (kyamamot@i.kyoto-u.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP18H01442, and in part by the KDDI Foundation.

ABSTRACT We propose a sophisticated channel selection scheme based on multi-armed bandits and stochastic geometry analysis. In the proposed scheme, a typical user attempts to estimate the density of active interferers for every channel via the repeated observations of signal-to-interference power ratio (SIR), which demonstrates the randomness induced by randomized interference sources and fading effects. The purpose of this study involves enabling a typical user to identify the channel with the lowest density of active interferers while considering the communication quality during exploration. To resolve the trade-off between obtaining more observations on uncertain channels and using a channel that appears better, we employ a bandit algorithm called Thompson sampling (TS), which is known for its empirical effectiveness. We consider two ideas to enhance TS. First, noticing that the SIR distribution derived through stochastic geometry is useful for updating the posterior distribution of the density, we propose incorporating the SIR distribution into TS to estimate the density of active interferers. Second, TS requires sampling from the posterior distribution of the density for each channel, while it is significantly more complicated for the posterior distribution of the density to generate samples than well-known distribution. The results indicate that this type of sampling process is achieved via the Markov chain Monte Carlo method (MCMC). The simulation results indicate that the proposed method enables a typical user to determine the channel with the lowest density more efficiently than the TS without density estimation aided by stochastic geometry, and ϵ-greedy strategies.

INDEX TERMS Channel selection, multi-armed bandit, Thompson sampling, stochastic geometry, Markov chain Monte Carlo method.

I. INTRODUCTION

Given explosive growth in wireless communications, the existing wireless networks are insufficient to satisfy significant demand for broad-bandwidth access driven by modern mobile traffic, such as multimedia transmissions and cloud computing tasks [1]. To cope with the exponential growth of mobile broadband data traffic, an important issue involves enhancing the spectrum utilization is a serious issue (e.g., device to device communication, heterogeneous networks) [2]. Conversely, these extensive uses of communications can lead to new challenges related to resource sharing and interference between communication nodes [3]. Problem solutions include efficient channel selection that plays an important role in the adoption of interference mitigation and performance improvement.

To efficiently identify the optimal available channel, a user should be able to monitor and sense the surroundings, and learn information about the unknown environment (e.g., the information about randomized interference source). It should be noted that seeking better channels requires more observations on uncertain channels, while excessive observations lead to less transmission opportunities on a channel that appears better. Hence, there is a dilemma between exploration

The associate editor coordinating the review of this manuscript and approving it for publication was Usama Mir.

and exploitation. The aforementioned perspective scenario attracts significant research effort to develop learning techniques that can optimize the trade-off between the exploration and the exploitation of environment and resources [4].

A relevant class of problem formulation corresponds to the multi-armed bandits (MAB), which is a powerful tool for online learning theory [5], [6]. The set of solutions is termed as bandit algorithms and play an important role in balancing the trade-off between exploration and exploitation. In most MAB frameworks, given a set of arms (actions), a player pulls an arm at each time to receive some reward. The rewards are not available for the player in advance. However, upon pulling any arm, the instantaneous reward of that arm is revealed [7]. There are a variety of applications of bandit algorithms to resource allocation in wireless networks [8]–[11].

However, to the best of the authors' knowledge, most of the applications of bandit algorithms require some necessary conditions or empirical parameter tunings. For example, the rewards in the upper confidence bound (UCB) [12] algorithms are required to satisfy the moment conditions, such as uniform distribution; and the performance of ϵ-greedy [13] depends on the configuration of ϵ, which controls the degree of exploration. When rewards for the arms are generated from arbitrary distributions (e.g., interference power and signal-to-interference power ratio (SIR)), a straightforward application without parameter dependence of these arbitrary distributions is difficult because the environment information of the channels vary over time, thereby resulting in a more complicated reward distributions.

Conversely, Thompson sampling (TS) [14] is an old heuristic based on Bayesian inference that selects an arm based on posterior samples of each arm; it has attracted significant attention for its empirically excellent performance [15]. Thus, TS exhibits more generality due to its flexibility to incorporate the complicated reward distributions of the arms. The reward distribution can be set by preferences based on different application scenarios.

In [11], Zhao et.al. proposed a TS-based antenna state selection scheme. They used the normalized signal-to-noise ratio (SNR) as the reward. Thus, the algorithm proposed in [16] can be applied in their system. This idea is also helpful for our study. However, as previously mentioned, the environment information (i.e., SIR) in the channels vary over time and the dynamic range is significant due to the spatial randomness of the interferers. Hence, it is inefficient to identify the channel with smallest interference only by the normalized SIR or normalized interference powers. With respect to capturing the randomness due to the topology of interferences, stochastic geometry is an important mathematical tool that provides the SIR distribution under the consideration of a spatial randomness of interference sources, and spatial averages calculated over a large number of nodes at different locations or over many network realizations [17]. It is useful to capture the relationship between the observation data and the system parameters (e.g., SIR and density of active interferers).

In the study, we propose a channel selection scheme based on TS and stochastic geometry. We provide a new perspective based on density estimation. In our system model, we assume that the locations of active transmitters in each channel follow a homogenous Poisson point process (HPPP) with a respective density. Our objective is to enable a typical user to identify the channel with the lowest density of other active transmitters (i.e., interferers) while considering the communication quality during exploration. It is noted that the values of densities are not given in advance. Hence, the user have to update the posterior distribution of density based on the measured performances (i.e., the measured SIRs associated with transmissions) to estimate its value.

In order to elaborate the TS-based scheme, we introduce two techniques that are related to reward distributions and a sampling method. First, to capture the relationship between measured SIR and density of active interferers more structurally, we employ the SIR distribution derived by stochastic geometry, and provide a method to incorporate this type of a statistical model into Thomson sampling. However, the resulting posterior distribution of each density parameter that appears in the process of TS, is not a well-known distribution (e.g., Gauss distribution and beta distribution), and thus we cannot draw samples in a simple way. The second technique is a sampling method termed as Markov chain Monte Carlo (MCMC) method [18], via which we overcome the difficulty of sampling from complicated posterior distributions. The MCMC-based sampling allows us to draw samples from the posterior distributions obtained through stochastic geometry analysis and Bayesian inference, and thus the TS-based algorithm performs well.

The contributions of the study are summarized as follows:

- Given the significant dynamic range of SIR in the channels, we provide a new perspective based on density estimation to the channel selection problem. In detail, we propose a framework to utilize the SIR distribution derived by stochastic geometry as a reward distribution in TS. In our system, a typical user can update the posterior distribution of densities based on the measured SIR and efficiently identify the channel with the lowest density.
- To overcome the difficulty of drawing samples from the complicated posterior distribution of density, we propose to employ the MCMC method which is a general and powerful tool for sampling from complicated distributions with high dimensionality of the sample space.
- We demonstrate that the proposed scheme resolves the exploration-exploitation trade-off more efficiently than the ϵ-greedy and the TS without density estimation aided by stochastic geometry through simulations. It is noted that although the performance of ϵ-greedy scheme sensitively depends on the parameters that are to be tuned, ϵ, the proposed scheme does not require such a learning parameter.

The rest of the study is organized as follows. In Section II, we introduce the system model and problem formulation. In

Section III, we present the proposed algorithm, and describe the stochastic geometry and sampling method. Section IV shows the algorithms selected for comparison, and Section V provides the simulation results. Finally, we conclude the study in Section VI.

II. SYSTEM MODEL

We consider K available channels on the Euclidean plane $\mathbb{R}^2$ and denote the index set of these channels by $\mathcal{C} = \{1, 2, \ldots, K\}$. We assume that these channels are orthogonal, and thus each channel is independent from each other and the locations of the active transmitters in kth channel form an independent HPPP Φ_k of density λ_k. For simplicity purposes, we assume $\lambda_1 < \lambda_2 < \cdots < \lambda_K$.

We specifically focus on a user in $\mathbb{R}^2$, who attempts to identify the channel with the lowest density, i.e., Φ_1. We term this typical user as the *learning user*. We assume that the learning user does not possess information on the density parameters $\{\lambda_k\}$ in advance. Without loss of generality, we can assume that the learning user is placed at the origin o and it attempts to communicate with the transmitter at a distance of r. The desired and interference signals experience path loss with an exponent of α and Rayleigh fading, i.e., the channel gain is constant during a time slot and is exponentially distributed with a mean of 1.

Our objective is to enable the learning user to efficiently identify the channel with the fewest interferers, i.e., channel $1 \in \mathcal{C}$. However, in general, the learning user is generally unaware of the density of interferers directly and can only sense the surrounding environment. In the system, the learning user updates the posterior distributions of densities to estimate the densities of interferes in each channel through the observation of SIR associated with the communication after channel selection. Additionally, we assume that the locations of the interferers continuously change as time progresses. In other words, the set of the active interferers varies over time while maintaining a constant proportion of all potential interferers. When the aggregate interference at the learning user in channel k is expressed by

$$I_k = \mathbb{E}\left(\sum_{x\in\Phi_k} h_x \|x\|^{-\alpha}\right), \tag{1}$$

where $\|\cdot\|$ denotes the Euclidean norm, h_x denotes the fading coefficient between the interferes at $x \in \Phi_k$ and the learning user [19].

III. PROPOSED SCHEME

We formulate the aforementioned channel selection problem as an MAB problem and solve it via the TS algorithm, which selects the optimal arm by optimizing a random sample from the posterior distributions. It is noted that in the context of an MAB problem, *arm* is used to denote an action to be selected. Hereinafter, we use *channel* and *arm* interchangeably.

In this section, we first briefly describe the TS algorithm, and propose a TS-based channel selection scheme in Section III-A. Sections III-B and III-C focus on the explanation of the mathematical preliminaries, i.e., stochastic geometry and sampling method MCMC, respectively.

A. THOMPSON SAMPLING

The basic idea of TS involves assuming a simple prior distribution on the underlying parameters of the reward distribution of every arm, and selecting an arm based on its posterior probability of being the optimal arm at every time step. The general TS involves the following elements [20]:

- A set of interested parameters θ; In the study, the parameter corresponds to density λ.
- An assumed prior distribution $P(\theta)$ on these parameters. This term can be removed, and this is indicated in Section III-C.
- Past observation data $\mathcal{D}$ for the arms played in the past time steps; In the study, the observation data corresponds to the measured SIR. Hereinafter, we use *observed* and *measured* interchangeably.
- An assumed likelihood function $P(\mathcal{D} \mid \theta)$, which gives the probability of the observation data, given a parameter θ;
- A posterior distribution $P(\theta \mid \mathcal{D}) \propto P(\mathcal{D} \mid \theta)\, P(\theta)$.

By sampling actions (playing arms) based on the optimal posterior probability, the algorithm continues to sample all actions that could plausibly be optimal, while shifting sampling away from those that are unlikely to be optimal [21]. Thus, the algorithm gradually discards the arms that are considered to underperform and finally converges to the optimal arm.

The proposed scheme is summarized in Algorithm 1. First, the arbitrary initial samples are set at the beginning of the algorithm. After selecting the channel with the lowest value of $\lambda_i, \forall i \in \mathcal{C}$, the learning user measures the SIR of the selected channel, and the posterior distribution of density in the selected channel is updated. It is noted that, as previously

Algorithm 1 Thompson Sampling With Density Estimation Aided by Stochastic Geometry ($\mathcal{C} = \{1, 2, \ldots, K\}$, $\mathcal{P} = \{P_1, P_2, \ldots, P_K\}$)

Initialization: Determine the initial samples $\lambda_i[0], \forall i \in \mathcal{C}$, where $\lambda_i[0]$ is a non-zero positive number.

for $t = 0, 1, \ldots, T$ **do**
  Select channel $k = arg\,min_{i\in\mathcal{C}}\, \lambda_i[t]$ to be observed.
  The learning user attempts to connect to channel k, and observes the corresponding *SIR*.
  Update the posterior distribution P_k of λ_k according to (7).
  for every channel $i \in \mathcal{C}$ **do**
    Draw the next samples $\lambda_i[t+1]$ independently according to the updated posterior distribution P_i.
  end for
end for

mentioned, in each iteration, the locations of the other transmitters in every channel are changed, based on HPPPs with respective densities, $\lambda_1, \lambda_2, \ldots, \lambda_K$ at Step 3. Subsequently, the new sample is obtained by employing MCMC. The algorithm then goes back to Step 2 and the learning user once again chooses a channel with the lowest value of $\lambda_i, \forall i \in \mathcal{C}$. It should be noted that MCMC is executed several times and only takes the last sample at the Step 6. This is because maintaining sufficient sampling intervals can mitigate the interdependence between samples [18].

B. STOCHASTIC GEOMETRY

As previously mentioned, the information of the posterior distributions is essential for TS algorithm to solve the problem. In this subsection, we derive the posterior distribution of the densities λ using stochastic geometry and Bayesian inference.

Stochastic geometry is extensively applied to evaluate the system performance of the wireless networks. Specifically, stochastic geometry models random topologies based on a point process (e.g., HPPP) to derive a direct and tractable mathematical expressions of the performance metrics (e.g., SIR distribution, transmission success probability, etc.) without loss of accuracy. In this section, we treat the cases without and with fading separately.

1) INTERFERENCE DISTRIBUTION WITHOUT FADING

In this case, the signal power is constant, and thus the SIR is only determined by the interference power. Hence, it is only necessary to consider the interference distribution. Based on [19], [22], the probability density function (pdf) of interference in Poisson networks without fading where the value of α is 4 is given as follows:

$$f_I(x) = \frac{\pi\lambda}{2x^{2/3}} \exp\left(-\frac{\pi^3\lambda^2}{4x}\right), \tag{2}$$

As previously mentioned, the environment information (i.e., interference power) of the selected channel is observed in each iteration in Algorithm 1. Let ith measured interference power be denoted by x_i. Hence, the likelihood function of λ can be expressed as follows:

$$\begin{aligned} P(I \mid \lambda) &= \prod_{i=1}^{N} f_I(x_i) \\ &= \left(\frac{\pi\lambda}{2}\right)^N \prod_{i=1}^{N} x_i^{-2/3} \exp\left(-\frac{\pi^3\lambda^2}{4}\sum_{i=1}^{N}\frac{1}{x_i}\right), \end{aligned} \tag{3}$$

where N denotes the number of observations.

Based on Bayes's theorem, the posterior distribution of λ is expressed as follows:

$$P(\lambda \mid I) = \frac{P(I \mid \lambda)\, P(\lambda)}{\int_\lambda P(I \mid \lambda)\, P(\lambda)\, \mathrm{d}\lambda}, \tag{4}$$

where $P(\lambda)$ denotes the prior distribution.

2) SIR DISTRIBUTION WITH RAYLEIGH FADING

In this case, the effects of Rayleigh fading are considered. Based on Theorem 5.7 in [19], the pdf of SIR is expressed as follows:

$$f_{SIR}(x) = \frac{2c\lambda}{\alpha} x^{2/\alpha - 1} \exp(-c\lambda x^{2/\alpha}), \tag{5}$$

where $SIR = S/I$, and S denotes the signal power, which is exponentially distributed with mean $r^{-\alpha}$. $c = \pi r^2\, \Gamma(1 + 2/\alpha)\, \Gamma(1 - 2/\alpha)$.

It is noted that, $f_{SIR}(x)$ denotes the function of x and λ. As previously mentioned, the SIR of the selected channel is observed at Step 3 in Algorithm 1. Let ith measured SIR be denoted by x_i. Hence, the likelihood function of λ is expressed as:

$$\begin{aligned} P(SIR \mid \lambda) &= \prod_{i=1}^{N} f_{SIR}(x_i) \\ &= \left(\frac{2c\lambda}{\alpha}\right)^N \prod_{i=1}^{N} x_i^{2/\alpha - 1} \exp\left(-c\lambda \sum_{i=1}^{N} x_i^{2/\alpha}\right), \end{aligned} \tag{6}$$

where N also denotes the number of the observations.

Similarly, based on Bayes' theorem, the posterior distribution of λ is expressed as follows:

$$P(\lambda \mid SIR) = \frac{P(SIR \mid \lambda)\, P(\lambda)}{\int_\lambda P(SIR \mid \lambda)\, P(\lambda)\, \mathrm{d}\lambda}, \tag{7}$$

where $P(\lambda)$ denotes the prior distribution.

C. MARKOV CHAIN MONTE CARLO METHODS

In order to obtain samples from the posterior distributions in the TS algorithm, we employ a sampling method termed as MCMC [18].

To draw samples, a simpler distribution $q(z)$, that is sometimes termed as a *proposal distribution* is required. In the MCMC, the proposal distribution is symmetric, i.e., $q(z_A \mid z_B) = q(z_B \mid z_A)$ for all values of z_A and z_B. At each step t, a candidate sample z' is drawn from the proposal distribution, and $z[t+1]$ is updated by z' with following the probability:

$$\min\left\{1, \frac{p(z')}{p(z[t])}\right\}, \tag{8}$$

where $p(z)$ denotes the target distribution. If the candidate sample z' is discarded, $z[t+1]$ is set to $z[t]$ and another candidate sample will be drawn. This rule is also termed as the Metropolis-Hastings criterion.

In the study, we utilize *random walk* to obtain samples from the posterior distribution $P(\lambda \mid SIR)$ (i.e., target distribution). Hence, at each step t, the candidate sample λ' is given as follows:

$$\lambda' = \lambda[t] + e, \tag{9}$$

where e denotes a sample from the normal distribution $\mathcal{N}(0, 1)$.

Algorithm 2 ϵ-Greedy With Maximum Likelihood Estimation ($\mathcal{A} = \{\{\hat{\lambda_1}[t]\}, \{\hat{\lambda_2}[t]\}, \ldots, \{\hat{\lambda_K}[t]\}\}$)

Input: the value of ϵ
for $t = 0, 1, \ldots, T$ **do**
  if $t = 0$ **then**
    Randomly select a channel from the available channels.
  end if
  Select a channel from the available channels with probability ϵ to be observed. Otherwise, select the channel with the minimum mean $k = arg\,min_{i\in\mathcal{N}} \mathbb{E}[\{\hat{\lambda_i}[t]\}]$.
  Observe *SIR* of the selected channel k.
  Calculate the maximum likelihood value $\hat{\lambda_k}[t]$ of the selected channel according to the (13).
end for

Algorithm 3 Thompson Sampling Without Density Estimation Aided by Stochastic Geometry

For each channel $i = 1, \ldots, N$ set $S_i(1) = 0, F_i(1) = 0$.
for $t = 0, 1, \ldots, T$ **do**
  for each channel $i = 1, \ldots, K$ **do**
    Sample $\theta_i(t)$ from the Beta $(S_i + 1, F_i + 1)$ distribution.
    Select channel $i(t) := \arg\max_i \theta_i(t)$ and observe *SIR*.
    $\tilde{r}_t \leftarrow$ normalized *SIR*
    Perform a Bernoulli trial with success probability $\tilde{r}_t$ and observe output $r_t \in \{0, 1\}$.
  end for
end for
if $r_t = 1$ **then**
  $S_i(t) = S_i(t) + 1$.
else
  $F_i(t) = F_i(t) + 1$.
end if

From (6), (7), (8), the acceptance probability in MCMC for the proposed algorithm is expressed as follows:

$$p = \min\left\{1, \frac{P(\lambda' \mid SIR)}{P(\lambda \mid SIR)}\right\} = \min\left\{1, \frac{P(SIR \mid \lambda')}{P(SIR \mid \lambda)}\right\} = \min\left\{1, \left(\frac{\lambda'}{\lambda}\right)^N \exp\left(-ce\sum_{n=1}^{N}(SIR)^{2/\alpha}\right)\right\}. \quad (10)$$

It is noted that the denominator and $P(\lambda)$ in (7) are removed. Hence, it is feasible to calculate this acceptance probability.

Similarly, in the case of without fading effects, the acceptance probability in MCMC is expressed as follows:

$$p = \min\left\{1, \frac{P(\lambda' \mid I)}{P(\lambda \mid I)}\right\} = \min\left\{1, \frac{P(I \mid \lambda')}{P(I \mid \lambda)}\right\} = \min\left\{1, \left(\frac{\lambda'}{\lambda}\right)^N \exp\left(-\frac{\pi^3}{4}\sum_{i=1}^{N}\frac{1}{x_i}(\lambda'^2 - \lambda^2)\right)\right\}, \quad (11)$$

where candidate sample λ' is also given by $\lambda' = \lambda + e$.

TABLE 1. Algorithms for comparison.

With density estimation (w/ DE)	Without density estimation (w/o DE)
TS (w/ DE)	TS (w/o DE)
ϵ-greedy (w/ DE)	ϵ-greedy (w/o DE)

TABLE 2. Simulation parameters.

System area	$1000\,\text{m} \times 1000\,\text{m}$
Number of channels K	3
Densities of interferers λ	$\{10^{-4}, 1.5\cdot10^{-4}, 2\cdot10^{-4}\}$ / $\{10^{-4}, 1.1\cdot10^{-4}, 1.2\cdot10^{-4}\}$
Path loss exponent α	4
Number of simulation steps	2000
Number of sampling intervals	10
Communication distance r	10 m
Identical transmission power P	23 dBm
Exploration parameter ϵ	0.01,0.1,0.5

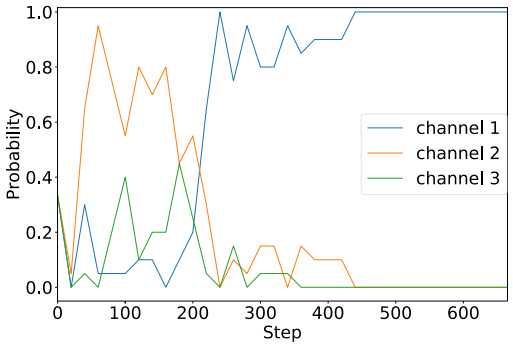


FIGURE 1. Probability of selected channels (TS (w/ DE)) in the case where the set of channel densities is given by $\{10^{-4}, 1.5\cdot10^{-4}, 2\cdot10^{-4}\}$.

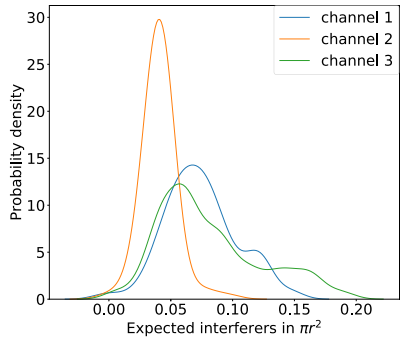
IV. ALGORITHMS FOR COMPARISON

In order to prove the effectiveness of the proposed scheme, we compare its performance with a conventional algorithm termed ϵ-greedy [13] and the TS without density estimation aided by stochastic geometry. Specifically, ϵ-greedy is a method that selects an available action with probability ϵ and selects a greedy action which is designed by users, otherwise. A naive approach is that the greedy action involves selecting the channel with the largest mean of the measured SIR. In the study, for the purpose of fairness, we provide another ϵ-greedy algorithm with maximum likelihood estimation (MLE) of λ for comparison. We assume that the greedy action selects a channel with the minimum mean of the maximum likelihood values of λ.

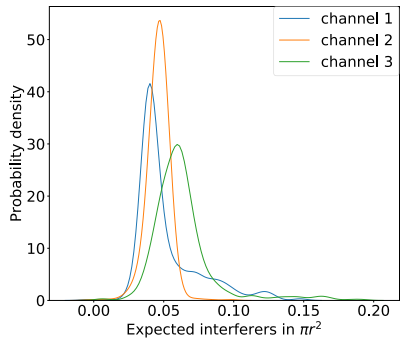
A. ϵ-GREEDY WITH MLE FOR CHANNEL SELECTION

Based on MLE, the log-likelihood function of λ is given as follows:

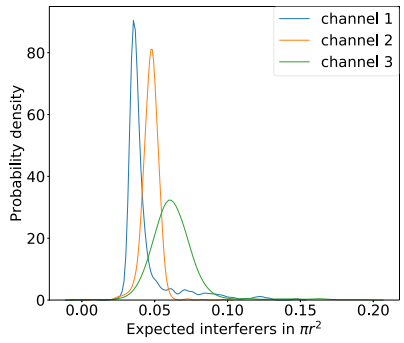
$$\ln[P(SIR \mid \lambda)] = N\ln\lambda + N\ln\left(\frac{2c}{\alpha}\right) + \sum_{i=1}^{N}\ln x_i^{2/\alpha-1} - c\lambda\sum_{i=1}^{N}x_i^{2/\alpha}, \quad (12)$$



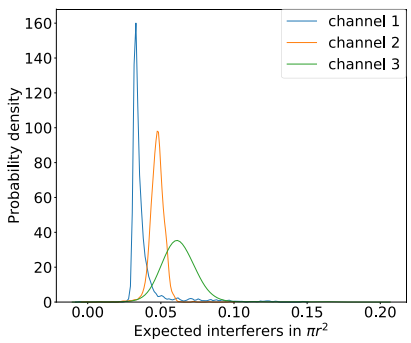
(a) Step 1–100.



(b) Step 1–500.



(c) Step 1–1000.



(d) Step 1–2000.

FIGURE 2. Kernel density estimation of expected number of interferers in πr^2, $\lambda \pi r^2$.

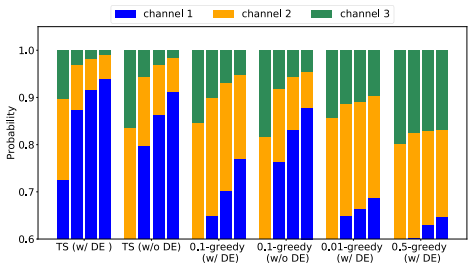


FIGURE 3. Comparison of TS algorithms and ϵ-greedy algorithms in the case where the set of channel densities is given by $\{10^{-4}, 1.5 \cdot 10^{-4}, 2 \cdot 10^{-4}\}$.

By taking the derivative with respect to λ, and setting the derivative to 0, we obtain the following expression:

$$\hat{\lambda} = \frac{N}{c \sum_{i=1}^{N} x_i^{2/\alpha}}, \tag{13}$$

which denotes the maximum likelihood value.

We summarized the ϵ-greedy algorithm with MLE in Algorithm 2. It should be noted that the maximum likelihood value of the selected channel is updated in each iteration.

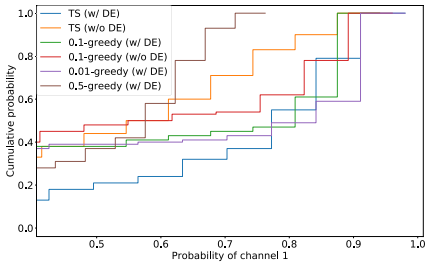
B. TS WITHOUT DENSITY ESTIMATION THROUGH STOCHASTIC GEOMETRY

When rewards for arm i are generated from an arbitrary unknown distribution with support $[0, 1]$ and mean μ_i, TS can be modified to adapt to the general stochastic bandits case [16]. The main idea is that the algorithm performs a Bernoulli trial with success probability $\tilde{r}_t$ after observing the reward $\tilde{r} \in [0, 1]$. Furthermore, the SIR of each channel in our system is bounded, and thus we can use normalized SIR at each time as the reward.

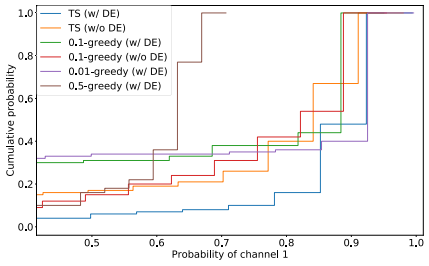
The algorithm in this case is summarized in Algorithm 3 as reference [16]. Let $S_i(t)$ and $F_i(t)$ denote the number of successes and failures in the Bernoulli trials, respectively, until time t. The remaining algorithm is identical to that for Bernoulli bandits.

V. SIMULATION RESULTS

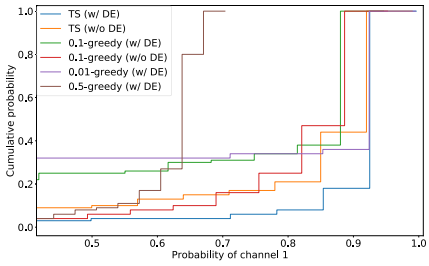
We evaluate and compare the performance of Algorithm 1, the aforementioned naive approach, Algorithm 2, and 3 via simulations. All algorithms are listed in Table 1, and the example parameters are listed in Table 2. In the simulations, three available channels are set, and the desirable channel corresponds to channel 1. The locations of the transmitters in these channels are change based on HPPPs with respective densities, λ_1, λ_2, and λ_3 in each iteration. All transmitters are assumed to exhibit the same configurable transmission power levels and to experience Rayleigh fading with a mean of 1 in the case with fading effects.



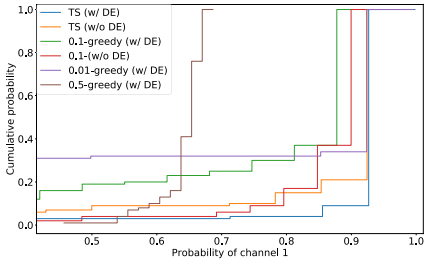
(a) Step 1–100.



(b) Step 1–500.



(c) Step 1–1000.



(d) Step 1–2000.

FIGURE 4. Empirical cumulative distribution function in the case where the set of channel densities is given by $\{10^{-4}, 1.5 \cdot 10^{-4}, 2 \cdot 10^{-4}\}$.

We demonstrate the simulation results where the fading effects (i.e., Rayleigh fading) is considered in Section V-A and Section V-B. The case without fading effects is described in Section V-C.

A. PERFORMANCE OF TS WITH DENSITY ESTIMATION

Fig. 1 shows the probability of the selected channels at intervals of twenty steps. Evidently, the probability of channel 1

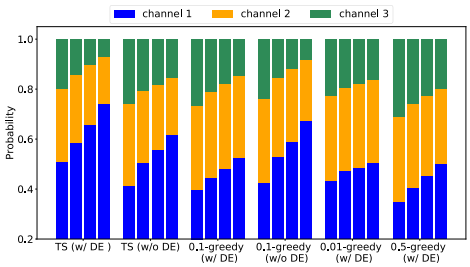


FIGURE 5. Comparison of TS algorithms and ϵ-greedy algorithms in the case where the set of channel densities is given by $\{10^{-4}, 1.1 \cdot 10^{-4}, 1.2 \cdot 10^{-4}\}$.

(which denotes the desired channel) converges to one in a short time. Although the action concentrates on channel 2 at the beginning of the simulation, the proposed algorithm can escape from that and eventually identifies the optimal channel as time progresses.

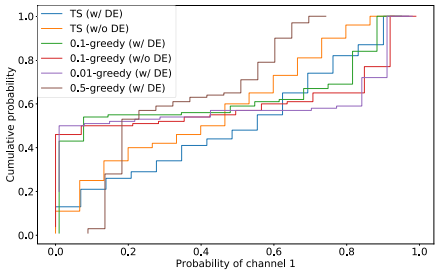
Figs. 2(a)-2(d) illustrate the kernel density estimation of the samples of expected number interferers in πr^2 for four periods. As shown in Fig. 2(a), it is intuitive that channel 2 is the optimal channel. However, when time progresses, the shapes of the distributions change because the posterior distribution of selected channel is updated. Finally, the desired channel with the lowest density is identified along with the updates of posterior distribution. Additionally, it should be noted that the means of $\lambda\pi r^2$ in the channels are close to the true values.

B. COMPARISON OF TS ALGORITHMS AND ϵ-GREEDY ALGORITHMS

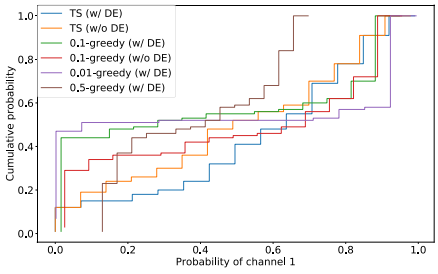
Fig. 3 shows a comparison of TS algorithms and ϵ-greedy algorithms ($\epsilon = 0.01, 0.1, 0.5$). The result is based on 100 independent simulations of each algorithm. It is noted that, 0.1-greedy (w/o DE) denotes the aforementioned naive approach without MLE. For every algorithm, the proportions of the selected channels in the four periods (i.e., step 100, 500, 1000, and 2000) are shown by four stacked histograms, and these results indicate the mean proportions of 100 independent simulations. We focus on the blue bars that reflect the proportion of channel 1. Evidently, the proposed algorithm (i.e., TS (w/ DE)) exhibits the optimal performance. The proportion of channel 1 in the proposed algorithm exceeds 0.7 at step 100.

Figs. 4(a)-4(d) show the empirical cumulative distribution function of the probability of channel 1 in 100 independent simulations in Fig. 3. It is observed that TS (w/ DE) keeps the optimal performance in the four periods. As shown in Fig. 4(d), over 90% of the simulations in the proposed algorithm, the probability of channel 1 exceeds 0.9.

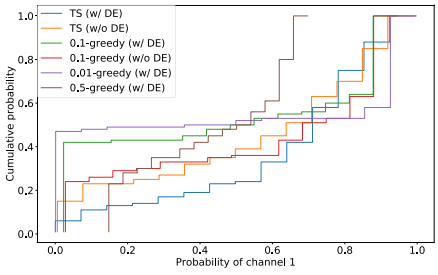
Fig. 5 shows a comparison of TS algorithms and ϵ-greedy algorithms ($\epsilon = 0.01, 0.1, 0.5$), when the values of densities in channels are $\{10^{-4}, 1.1 \cdot 10^{-4}, 1.2 \cdot 10^{-4}\}$. In this case, the environment information of the available channels are similar, and thus, it is difficult to identify the optimal channel.



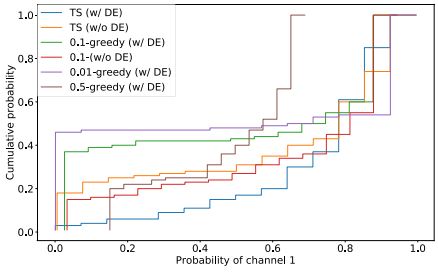
(a) Step 1–100.



(b) Step 1–500.



(c) Step 1–1000.



(d) Step 1–2000.

FIGURE 6. Empirical cumulative distribution function in the case where the set of channel densities is given by $\{10^{-4}, 1.1 \cdot 10^{-4}, 1.2 \cdot 10^{-4}\}$.

It is observed that the proportion of channel 1 for every algorithm is reduced. However, the proposed algorithm still exhibits the optimal performance in this case.

Figs. 6(a)-6(d) show the empirical cumulative distribution function of the probability of channel 1 in 100 independent simulations in Fig. 5. It is observed that the performance of ϵ-greedy algorithms are biased. For example, the probability

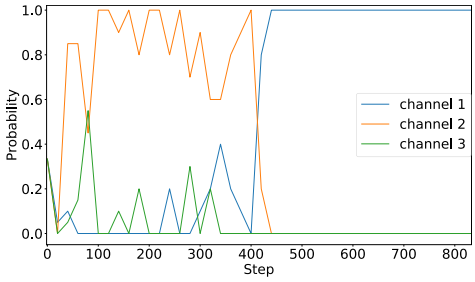


FIGURE 7. Probability of selected channels (TS (w/ DE)) when the interference distribution without fading is considered, and the set of channel densities is given by $\{10^{-4}, 1.5 \cdot 10^{-4}, 2 \cdot 10^{-4}\}$.

of channel 1 exceeds 0.9 in 50% of the simulations, especially, in 0.01-greedy algorithm. However, the mean proportion of channel 1 in Fig. 5 is lower than 0.5. Conversely, over 80% of the simulations in the proposed algorithm, the probability of channel 1 exceeds 0.6, and the mean proportion of channel 1 exceeds 0.7 in Fig. 5. Thus, the results indicate that the proposed algorithm exhibits the overall optimal performance.

Additionally, Figs. 3 and 5 show that the performance of ϵ-greedy algorithms rely on the value of ϵ. Conversely, it is not necessary for the proposed algorithm to configure the values of the parameters in advance. Therefore, the proposed algorithm exhibits a higher generality than the ϵ-greedy algorithms. With respect to the density estimation, the results indicate that the performance of TS algorithm can be improved by considering the density estimation. Conversely, it does not perform well in ϵ-greedy algorithms.

C. INTERFERENCE DISTRIBUTION WITHOUT FADING

The key parameters in this case are identical to that in Table 2 except the part of Rayleigh fading. In the simulation, three available channels are set, and the desirable channel also corresponds to channel 1.

Fig. 7 shows the probability of the selected channel at intervals of twenty steps where the set of channel densities is given by $\{10^{-4}, 1.5 \cdot 10^{-4}, 2 \cdot 10^{-4}\}$. Evidently, the result is similar to Fig. 1. The action focuses on exploration at the beginning of the simulation, and eventually converges to the best channel as time progresses. Therefore, the results indicate that the proposed scheme exhibits high generality when the distribution given by stochastic geometry analysis is considered.

VI. CONCLUSION

In the study, we formulate the channel selection problem of as an MAB problem and solve it with a TS-based algorithm by employing density estimation based on stochastic geometry and MCMC. The user can utilize the SIR distribution derived by stochastic geometry to update the posterior distribution of densities to estimate the densities of the channels and efficiently identify the optimal channel. We compare the performance of the proposed algorithm, the ϵ-greedy algorithms,

and TS without density estimation through stochastic geometry. The results indicate that the proposed algorithm can identify the optimal channel more accurately and efficiently, and exhibits a steady performance. Additionally, the proposed algorithm is not restricted to the posterior distribution of λ and can be widely applied to other models with different posterior distributions (e.g., the distribution of interference with other fading as Nakagami fading). Future works include verifying the performance of the proposed algorithm in a non-stationary system.

ACKNOWLEDGMENT

This article was presented in part at the 2020 IEEE Consumer Communications and Networking Conference.

REFERENCES

[1] W. Bao and B. Liang, "Stochastic geometric analysis of user mobility in heterogeneous wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2212–2225, Oct. 2015.
[2] P. Phunchongharn, E. Hossain, and D. I. Kim, "Resource allocation for device-to-device communications underlaying LTE-advanced networks," *IEEE Wireless Commun.*, vol. 20, no. 4, pp. 91–100, Aug. 2013.
[3] S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Operational region of D2D communications for enhancing cellular network performance," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5984–5997, Nov. 2015.
[4] K. Avrachenkov, L. Cottatellucci, and L. Maggi, "Slow fading channel selection: A restless multi-armed bandit formulation," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, Oct. 2012, pp. 1083–1087.
[5] J. Zhu, Y. Song, D. Jiang, and H. Song, "Multi-armed bandit channel access scheme with cognitive radio technology in wireless sensor networks for the Internet of Things," *IEEE Access*, vol. 4, pp. 4609–4617, 2016.
[6] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao, and J. Wang, "Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1689–1713, Apr. 2013.
[7] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5G small cells," *IEEE Wireless Commun.*, vol. 23, no. 3, pp. 64–73, Jun. 2016.
[8] V. Toldov, L. Clavier, V. Loscri, and N. Mitton, "A Thompson sampling approach to channel exploration-exploitation problem in multihop cognitive radio networks," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Valencia, Spain, Dec. 2016, pp. 1–6.
[9] S. Boldrini, L. De Nardis, G. Caso, M. Le, J. Fiorina, and Sep. , "MuMAB: A Multi-Armed Bandit Model for Wireless Network Selection," *Algorithms*, vol. 11, no. 2, p. 13, Jan. 2018.
[10] A. Feki and V. Capdevielle, "Autonomous resource allocation for dense LTE networks: A Multi Armed Bandit formulation," in *Proc. IEEE 22nd Int. Symp. Pers., Indoor Mobile Radio Commun.*, Toronto, ON, Canada, Jan. 2011, pp. 66–70.
[11] T. Zhao, M. Li, and M. Poloczek, "Fast reconfigurable antenna state selection with hierarchical thompson sampling," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[12] E. Kaufmann, O. Cappé, and A. Garivier, "On Bayesian upper confidence bounds for bandit problems," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, La Palma, Spain, Apr. 2012, pp. 592–600.
[13] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[14] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, nos. 3–4, pp. 285–294, Dec. 1933.
[15] J. Komiyama, J. Honda, and H. Nakagawa, "Optimal regret analysis of Thompson sampling in stochastic multi-armed bandit problem with multiple plays," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Lille, France, vol. 37, Jul. 2015, pp. 1152–1161.
[16] S. Agrawal and N. Goyal, "Analysis of Thompson sampling for the multi-armed bandit problem," in *Proc. Annu. Conf. Learn. Theory (PMLR)*, vol. 23, S. Mannor, N. Srebro, and R. C. Williamson, Eds. Edinburgh, Scotland, Jun. 2012, pp. 39.1–39.26.
[17] H. Elsawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996–1019, Jun. 2013.
[18] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2011.
[19] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[20] S. Agrawal and N. Goyal, "Near-optimal regret bounds for thompson sampling," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, Scottsdale, AZ, USA, Apr. 2013, pp. 99–107.
[21] D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen, "A tutorial on thompson sampling," *FNT Mach. Learn.*, vol. 11, no. 1, pp. 1–96, Jul. 2018.
[22] E. Sousa and J. Silvester, "Optimum transmission ranges in a direct-sequence spread-spectrum multihop packet radio network," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 5, pp. 762–771, Jun. 1990.

WANGDONG DENG (Student Member, IEEE) received the B.E. degree in physics and telecommunications engineering from South China Normal University, in 2015. He is currently pursuing the M.I. degree with the Graduate School of Informatics, Kyoto University. He received the VTS Japan Young Researcher's Encouragement Award, in 2019.

SHOTARO KAMIYA (Student Member, IEEE) received the B.E. degree in electrical and electronic engineering from Kyoto University, in 2015, and the M.E. degree from the Graduate School of Informatics, Kyoto University, in 2017, where he is currently pursuing the Ph.D. degree. He is a Student Member of the IEICE.

KOJI YAMAMOTO (Member, IEEE) received the B.E. degree in electrical and electronic engineering, and the master's and Ph.D. degrees in informatics from Kyoto University, in 2002, 2004, and 2005, respectively. From 2004 to 2005, he was a Research Fellow with the Japan Society for the Promotion of Science (JSPS). From 2008 to 2009, he was a Visiting Researcher at Wireless@KTH, Royal Institute of Technology (KTH), Sweden. Since 2005, he has been with the Graduate School of Informatics, Kyoto University, where he is currently an Associate Professor. His research interests include radio resource management, game theory, and machine learning. He is a member of the Operations Research Society of Japan and a Senior Member of the IEICE. He received the PIMRC 2004 Best Student Paper Award, in 2004, and the Ericsson Young Scientist Award, in 2006, and the Young Researcher's Award, the Paper Award, and the SUEMATSU-Yasuharu Award from the IEICE of Japan, in 2008, 2011, and 2016, respectively, and the IEEE Kansai Section GOLD Award, in 2012. He was a Tutorial Lecturer in ICC 2019. He was also the Track Co-Chair of APCC 2017, CCNC 2018, APCC 2018, and CCNC 2019, and the Vice Co-Chair of the IEEE ComSoc APB CCC. He serves as an Editor of the IEEE Wireless Communications Letters and the *Journal of Communications and Information Networks*.

TAKAYUKI NISHIO (Member, IEEE) received the B.E. degree in electrical and electronic engineering, and the master's and Ph.D. degrees in informatics from Kyoto University, in 2010, 2012, and 2013, respectively. He is currently an Assistant Professor in communications and computer engineering with the Graduate School of Informatics, Kyoto University. From 2016 to 2017, he was a Visiting Researcher with the Wireless Information Network Laboratory (WINLAB), Rutgers University, USA. His current research interests include machine learning-based network control, machine learning in wireless networks, and heterogeneous resource management.

MASAHIRO MORIKURA (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in electronics engineering from Kyoto University, Kyoto, Japan, in 1979, 1981, and 1991, respectively. He joined NTT, in 1981, where he was engaged in the research and development of TDMA equipment for satellite communications. From 1988 to 1989, he was with the Communications Research Centre, Canada, as a Guest Scientist. From 1997 to 2002, he was also active in the standardization of the IEEE 802.11a based wireless LAN. He is currently a Professor with the Graduate School of Informatics, Kyoto University. His current research interests include WLANs and M2M wireless systems. He received the Paper Award and the Achievement Award from IEICE, in 2000 and 2006, respectively, the Education, Culture, Sports, Science and Technology Minister Award, in 2007, the Maejima Award, in 2008, and the Medal of Honor with Purple Ribbon from Japan's Cabinet Office, in 2015.