This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3116223

Sleeping Contextual/Non-Contextual Thompson Sampling MAB for mmWave D2D Two-Hop Relay Probing

Ehab Mahmoud Mohamed, *Member, IEEE,* Sherief Hashima, *Senior Member, IEEE,* Kohei Hatano, Mostafa M. Fouda, *Senior Member, IEEE*, and Zubair Md Fadlullah, *Senior Member, IEEE*

***Abstract*—Millimeter wave (mmWave) band, i.e., 30 to 300 GHz, is characterized by short range transmissions and vulnerability to path blockage necessitating the use of relaying. Probing more relays finds out the best relay having the highest spectral efficiency but at the expense of increasing the probing overhead due to excessive beamforming training (BT) causing a decrease in the overall throughput. In this paper, mmWave two-hop relaying will be formulated as a single player multi-armed bandit (MAB) problem enabling one relay probing while maximizing the achievable spectral efficiency. Moreover, the relays could not establish the mmWave link due to blockage for instance will be identified as sleeping relays and eliminated from the rest of the MAB game. Thus, sleeping non contextual MAB (S-MAB) algorithm, namely sleeping Thompson sampling (S-TS) will be proposed to handle the problem. Furthermore, by utilizing the multiband capability of standardized WiGig devices containing both 2.4/5 GHz WiFi and 60 GHz mmWave bands, WiFi information will be used as contexts of the MAB game. Therefore, sleeping contextual MAB (S-CMAB) algorithm, namely S-CTS, will be proposed as well. Numerical and regret analysis ensure the superior performance of the S-CMAB algorithm over the S-MAB counterpart and the existing mmWave relay probing solutions accompanied with high convergence rates.**

***Index Terms*—mmWave relay probing, Contextual MAB, TS, CTS.**

I. Introduction

MILLIMETER wave (mmWave) band, i.e., 30 to 300 GHz, is one of the main enabling technologies of the current fifth generation (5G) mobile networks and the upcoming beyond 5G (B5G) and six generation (6G) [1]–[3]. However, mmWave transmissions suffer from harsh channel impairments due its highly operating frequencies, which reduces its transmission range and gets it vulnerable to path blockage [4], [5]. Although antenna beamforming training (BT) is advocated as an efficient strategy to compact the fragile mmWave channel, it consumes a considerable overhead to find out the best transmit (TX) / receive (RX) beam pair [6]–[9]. From standardization point of view, IEEE 802.11ad [10] and IEEE 802.11ay [11] become the ratified standards of wireless gigabit (WiGig) operating at 60 GHz for wireless local area network (WLAN) applications. Besides, device to device (D2D) relaying [12]–[14] turns to be a reasonable and promising approach to extend the mmWave D2D coverage and rout around blockages. However, investigating the best relay maximizing the spectral efficiency of the mmWave link from source to destination using relay probing consumes a considerable BT overhead reducing the achievable throughput of the mmWave relaying. Thus, a tradeoff exists between exploring more relays and increasing the attainable throughput. In [7], the authors firstly investigated the mmWave D2D two-hop relaying problem, and they stated that this problem is a pure threshold policy, where a fixed-point equation was formulated to find out the predefined spectral efficiency. Despite the pioneer work presented in [7], it has the following shortcomings: 1) The distribution of the spectral efficiency as well as the blockage probability should be known beforehand which is impractical in real scenarios. 2) The relay probing process should be stop once reaching the predefined spectral efficiency although better spectral efficiency can be obtained by one of the other non-probed relays. To further enhance the relay probing process, the authors in [8] proposed to utilize the out-of-band information provided by wireless fidelity (WiFi) to assist mmWave relay selection thanks to the standardized multi-band WiGig devices comprising both 2.4/5 GHz WiFi [14] and 60 GHz mmWave interfaces. However, the scheme presented in [8] still needs a high number of real-time probed relays. Moreover, the average throughputs of both schemes are still far from the optimal one [8], where the maximum spectral efficiency can be achieved by only probing the best relay.

Multi-armed bandit (MAB) is an efficient online learning tool, where a player intends to maximize its average reward, i.e., profit, through playing over multiple arms [15], [16]. Typically, MAB games contain a fundamental tradeoff be-

E. M. Mohamed is with the Electrical Engineering Department, College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Addwasir 11991, Saudi Arabia and the Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt (email: ehab_mahmoud@aswu.edu.eg).

S. Hashima is with the Computational Learning Theory Team, RIKEN-Advanced Intelligent Project, Fukuoka 819-0395, Japan, and the Department of Engineering and Scientific Equipment, Egyptian Atomic Energy Authority, Cairo, Inshas 13759, Egypt (e-mail: sherief.hashima@riken.jp).

K. Hatano is with Faculty of Arts and Science, Kyushu University, Fukuoka 819-0395, Japan and with the Computational Learning Theory Team, RIKEN-Advanced Intelligent Project, Fukuoka 819-0395, Japan (email: hatano@inf.kyushu-u.ac.jp).

M. M. Fouda is with the Department of Electrical and Computer Engineering, College of Science and Engineering, Idaho State University, Pocatello, ID 83209, USA (email: mfouda@ieee.org).

Z. M. Fadlullah is with the Department of Computer Science, Lakehead University; and Thunder Bay Regional Health Research Institute (TBRHRI), Thunder Bay, Ontario, Canada (email: Zubair.Fadlullah@lakeheadu.ca).

This work was supported by **JSPS KAKENHI** Grant Numbers **JP19H04174** and **JP21K14162**, respectively.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3116223

tween always exploiting the arms that gave high average profits so far or exploring new ones [16]. Many algorithms exist in literature to efficiently realize the MAB game such as upper confidence bound (UCB) [17] and Thompson sampling (TS) [18]. Contextual MAB (CMAB) is a powerful type of MAB games, where the player enhances its arm selection policy via utilizing arms' features vector, named as contexts [19]. Linear UCB (LinUCB) [19] and contextual TS (CTS) [20] algorithms are proposed in literature to implement the CMAB hypothesis. It is well known in literature that TS and CTS algorithms have performance guarantee better than UCB and LinUCB. This is the reason why we will only focus on TS based algorithms to handle the mmWave relay probing in this paper [18], [20].

In this paper, motivated by the exploitation-exploration dilemma of the MAB games which meets the aforementioned tradeoff of the mmWave relay probing, the problem of mmWave D2D two-hop relay probing is considered as a single player online MAB game. In this formulation, the source node will act as the player aiming to maximize its long-term spectral efficiency from source (S) to destination (D), which emulates the reward of the game. This is done via playing over the available relays (R), which are the arms of the bandit. Thus, the main contributions of this paper can be listed as follows:

- The optimization problem of the mmWave D2D two-hop relay probing will be formulated as a MAB game to find out the best candidate relay through online learning. This will boost the relay probing process over time while keeping the BT overhead at the minimum level as only one relay node will be probed at a time.
- Sleeping non-contextual MAB (S-MAB) algorithm, namely S-TS, will be proposed to address the formulated problem. In this scheme, relay nodes unable to establish the mmWave S-R-D link, e.g., blocked relays, will be considered as sleeping arms and removed from the rest of the MAB game. This will give the proposed scheme the capacity of learning idle relays such as those experiencing harsh blockage and eliminating them, which cannot be achieved using the conventional approaches explained above.
- The direct relationship between WiFi and mmWave link statistics proved in [8], [21] motivates us to exploit WiFi statistics as contexts of the mmWave relays. Thus, S-CMAB algorithm, namely S-CTS, will be proposed to further enhance the performance of the constructed MAB game at no additional cost thanks to the multi-band WiGig devices. Moreover, it investigates the performance improvements over the non-contextual counterpart, i.e., the S-MAB policy. In the proposed S-CMAB strategy, the instantaneous value of the WiFi received power in addition to its average value and variance up to time t are used as contexts of the mmWave relays.
- Regret analysis is conducted to bound the performance of the proposed S-TS and S-CTS mathematically. Also, extensive numerical simulations are conducted to compare the performances of the proposed S-MAB and S-CMAB approaches and prove their effectiveness over the existing techniques given in [7] and [8]. Moreover, the convergence rates of the proposed MAB schemes towards the optimal performance will be investigated as well. In this paper, S-MAB/ S-TS and S-CMAB /S-CTS will be used interchangeably.

The rest of this paper is organized as follows, Section II gives the related works, and Section III gives the proposed system model including the optimization problem formulation. Section IV gives the proposed S-MAB and S-CMAB approaches including the proposed S-TS, and S-CTS algorithms. Section V gives the regret analysis of the proposed S-TS and S-CTS algorithms. Section VI gives the conducted numerical simulations followed by the concluded remarks in Section VII.

II. Related Works

To overcome the mmWave short range transmissions as well as its susceptibility to path blockage, mmWave relaying was investigated in literature [22]–[28]. Stochastic geometry was used by the authors in [22] and [23] to study the improvements in mmWave D2D relay networks and to bound the ranging performance for adjusting the placement of the mmWave relay nodes. In [24], the authors proposed a buffer aided relaying to improve the delay performance of mmWave machine-to machine (M2M) communications. In [25], the problem of power allocations in conjunction with full duplex relaying was investigated by the means of multi-objective combinatorial optimization. In [26], to minimize the total transmission time as well as overcoming mmWave blockage, multi-hop relaying was proposed. In [27], the authors studied the error rate, capacity, and coverage of decode and forward mmWave relaying for both line-of-sight (LoS) and non-LoS (NLoS) scenarios. Full duplex relaying in conjunction with mmWave transmission was investigated by the authors in [28], while self-interference cancellation is performed by the means of orthogonal matching precoder. Despite the deep investigations of mmWave relaying provided by the previous research works [22]–[28], the problem of relay probing was highly relaxed as their main concern was to analyze the performance of mmWave relay networks and its related radio resource management. Thus, all channel information is assumed to be known beforehand ignoring the incredible BT overhead required to obtain such information. At the best of our knowledge, only the schemes given in [7] and [8] considered the problem of mmWave relay probing with the aforementioned drawbacks. Recently, different MAB schemes have been widely applied in numerous wireless communication challenges, especially in D2D communications as surveyed in [29], [30]. Multi-player MABs have been employed to control the transmitted power of the direct communications between two D2D users for interference mitigation and performance improvement [31], [32]. In [33], the authors formulated the mmWave D2D neighborhood discovery problem as a budget constrained stochastic MAB framework. Multi-player MAB also has been applied in UAV selections in disaster area scenario [34]. Furthermore, in [35], a multi-user MAB framework has been applied for relay selection in underwater acoustic sensor networks without any former knowledge about channel settings. An energy efficient relay selection technique

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3116223

based on multi-player MAB, which solves the permutation problem, for hierarchical wireless sensor networks (WSNs) is discussed in [36]. In [37], the problem of relay selection in acoustic underwater communications is formulated as a CMAB problem using the contexts of the underwater environment. At the best of our knowledge, the consideration of mmWave two-hop relaying as a sleeping MAB problem while utilizing the WiFi information as contexts is firstly introduced in this paper.

III. System Model

In this section, we will give the network architecture of the mmWave D2D two-hop relaying, the used WiFi/mmWave link models, and the optimization problem formulation.

A. MmWave D2D Two-Hop Relaying Network Architecture

Fig. 1 shows the network architecture of the mmWave D2D two-hop relaying, where WiFi/mmWave nodes are distributed in the base station (BS) area, e.g., Macro-cell, Pico-cell, etc. The BS is responsible for managing/controlling the operation of the D2D communications such as scheduling the D2D requests using control signaling provided by the BS2D control link. However, the initiation, termination and relay probing of the D2D relay links are done locally by the distributed nodes. Although power line communication (PLC) is a promising technology for future communications systems [38], it cannot be replaced by mmWave links in the system model under consideration. This is because it will not be convenient for the mmWave movable relay nodes, e.g., smart phones. In Fig. 1, if node S intends to establish a mmWave D2D link with node D while node D is far away or experiences blockage, node S will relay its information through one of the distributed relays, e.g., R_1, R_2, and R_3, using either mmWave LoS or NLoS paths. Accordingly, relay probing using BT should be conducted between node S and the distributed relays and between the relays and node D to find out the best TX/RX beam pairs maximizing the achievable spectral efficiency of the $S-R-D$ link. Then, node S selects the best relay node R_i^* having the maximum spectral efficiency. Apparently, probing more relays results in finding out the best one having the maximum spectral efficiency but at the expense of increasing the BT overhead and decreasing the achievable throughput eventually. Ideally, the best relay should be obtained by just one relay probing. The works given in [7] and [8] tackled this problem but with the aforementioned drawbacks, which we will overcome using online learning as will be explained throughout this paper.

B. WiFi and MmWave Link Models

In this subsection, the used WiFi and mmWave link models including the mmWave blockage model will be explained.

1) WiFi Link Model: Without loss of generality, the 5.25 GHz WiFi link model given in [39] will be used throughout this paper, where the received power, P_r^w in dBm at RX separated by d meters from TX is expressed as:

$$P_r^w\,[\text{dBm}] = P_t^w\,[\text{dBm}] - \beta_w - 10n_w \log_{10}(d) - \delta_w, \quad (1)$$

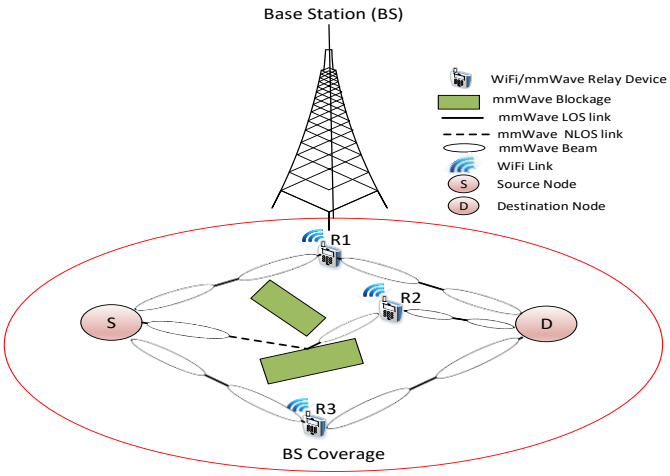

Fig. 1. Network architecture of mmWave D2D Two-Hop relaying.

where $P_t^w\,[\text{dBm}]$, β_w, n_w, δ_w are the WiFi TX power in dBm, path loss at the reference distance d_0, path loss exponent, and the log-normal shadowing term, respectively. Based on the measurements given in [39], $\beta_w = -47.4$ dB, $d_0 = 5$m, $n_w = 2.33$ and $\delta_w \sim \mathcal{N}(0, \sigma_w)$ with zero mean and standard deviation σ_w of 6 dB.

2) MmWave Link Model and Blockage Probability: For mmWave link model, we utilize that given in [40] and [41], where the mmWave RX power, P_r^m, at a distance d from mmWave TX considering both TX and RX beamforming gains in addition to both LoS and NLoS paths, can be expressed as:

$$\begin{aligned} P_r^m =& P_t^m A_{TX}(\theta,\theta_{TX})\, A_{RX}(\varphi,\varphi_{RX}) \\ & \Big(\varepsilon\left(\mathbb{P}_{\text{LoS}}(d)\right) 10^{-0.1\beta_m^{\text{LoS}}} d^{-n_m^{\text{LoS}}} \mathcal{Z}_{m,\mathcal{LN}}^{\text{LoS}} \\ & + \nu\left(\mathbb{P}_{\text{NLoS}}(d)\right) 10^{-0.1\beta_m^{\text{NLoS}}} d^{-n_m^{\text{NLoS}}} \mathcal{Z}_{m,\mathcal{LN}}^{\text{NLoS}}\Big), \end{aligned} \quad (2)$$

where P_t^m is the mmWave TX power, $A_{TX}(\theta,\theta_{TX})$ indicates the beamforming gain at the TX as a function of the angle of departure (AoD) θ, and the TX beam boresight angle θ_{TX}. $A_{RX}(\varphi,\varphi_{RX})$ indicates the RX beamforming gain as a function of the angle of arrival (AoA) φ and the RX beam boresight angle φ_{RX}. In this paper, the 2D steerable antenna model with Gaussian main loop profile given in [42] is used for expressing A_{TX} and A_{RX}, where A_{TX} is expressed as:

$$\begin{aligned} A_{TX}(\theta,\theta_{TX}) &= A_0 \exp\left(-4\ln(2)\left(\frac{\theta-\theta_{TX}}{\theta_{-3\text{dB}}}\right)^2\right), \\ A_0 &= \left(\frac{1.6162}{\sin\left(\frac{\theta_{-3\text{dB}}}{2}\right)}\right)^2, \end{aligned} \quad (3)$$

where A_0 is the maximum gain and $\theta_{-3\text{dB}}$ is -3dB beamwidth. By interchanging θ by φ, θ_{TX} by φ_{RX}, and $\theta_{-3\text{dB}}$ by $\varphi_{-3\text{dB}}$, same expression is used for calculating $A_{RX}(\varphi,\varphi_{RX})$. $\Lambda_{\text{LoS}} = \varepsilon\left(\mathbb{P}_{\text{LoS}}(d)\right) 10^{-0.1\beta_m^{\text{LoS}}} d^{-n_m^{\text{LoS}}} \mathcal{Z}_{m,\mathcal{LN}}^{\text{LoS}}$ and $\Lambda_{\text{NLoS}} = \nu\left(\mathbb{P}_{\text{NLoS}}(d)\right) 10^{-0.1\beta_m^{\text{NLoS}}} d^{-n_m^{\text{NLoS}}} \mathcal{Z}_{m,\mathcal{LN}}^{\text{NLoS}}$ indicate both the LoS component and the NLoS component of the received mmWave

signal, respectively. β_m^{LoS}, β_m^{NLoS} are the LoS, NLoS path losses at reference distance $d_0 = 5$m, and n_m^{LoS}, n_m^{NLoS} are the path loss exponents of the LoS, NLoS paths, respectively. $\mathcal{Z}_{m,\mathcal{LN}}^{\text{LoS}} \sim \mathcal{LN}\left(0, (\delta_{m,\mathcal{LN}}^{\text{LoS}})^2\right)$ and $\mathcal{Z}_{m,\mathcal{LN}}^{\text{NLoS}} \sim \mathcal{LN}\left(0, (\delta_{m,\mathcal{LN}}^{\text{NLoS}})^2\right)$ are log-normal random variables (R.Vs) with zero mean and variances of $\left(\delta_{m,\mathcal{LN}}^{\text{LoS}}\right)^2$ and $\left(\delta_{m,\mathcal{LN}}^{\text{NLoS}}\right)^2$ representing the shadowing terms. $\delta_{m,\mathcal{LN}}^{\text{LoS}} = 0.1\sigma_m^{\text{LoS}} \ln 10$ and $\delta_{m,\mathcal{LN}}^{\text{NLoS}} = 0.1\sigma_m^{\text{NLoS}} \ln 10$ as deduced in [43], where σ_m^{LoS} and σ_m^{NLoS} are the standard deviations of the normal distributions of the LoS and NLoS shadowing terms.

In (2), $\varepsilon\left(\mathbb{P}_{\text{LoS}}(d)\right)$ and $\nu\left(\mathbb{P}_{\text{NLoS}}(d)\right)$ are two Bernoulli R.Vs implementing the blockage of the LoS and NLoS paths, where $\mathbb{P}_{\text{LoS}}(d)$ and $\mathbb{P}_{\text{NLoS}}(d)$ are the probabilities of LoS and NLoS as functions of the separation distance d, noted that $\mathbb{P}_{\text{NLoS}}(d) = 1 - \mathbb{P}_{\text{LoS}}(d)$. For modeling $\mathbb{P}_{\text{LoS}}(d)$, we utilize the model given in [22], which is applicable for both indoors and outdoors. In this model, blockages of cylindrical shapes are distributed in the environment according to 2D Poisson point process (PPP). Thus, the probability that there are no blockages intersecting the LoS path between mmWave TX and RX separated by d m, can be expressed as:

$$\mathbb{P}_{\text{LoS}}(d) = \mu e^{-\eta d}, \tag{4}$$

where $\mu = e^{-\pi\lambda\xi\mathbb{E}(\Omega)^2}$ and $\eta = 2\lambda\xi\,\mathbb{E}(\Omega)$. ξ is the thinning factor, which is equal to 1 in case of indoor scenarios, and λ is the obstacles density. Ω is an R.V expressing the radius of the obstacles, and $\mathbb{E}(.)$ indicates the average value or expectation operation. The detailed derivation of this equation can be found in [22].

3) MmWave D2D Two-Hop Relay Probing Optimization Problem: The aim of the mmWave D2D relay probing is to find out the best relay, R_i^*, maximizing the throughput from S node to D node, TH_{S-R_i-D}, via probing N_P relays out of N total relays. This can be expressed mathematically as follows:

$$R_i^* = \max_{1\le R_i\le N_P}\left(\text{TH}_{S-R_i-D}\right) = \max_{1\le R_i\le N_P}\left(o_{N_P}\omega\psi_{S-R_i-D}\right)$$
$$\text{s.t. } R_i \in \phi_R$$

where

$$o_{N_P} = \left(\frac{T_d}{T_{N_P}+T_d}\right), \tag{5}$$

ϕ_R is the space of all available relays with cardinality of N, o_{N_P} is the overhead resulting from probing N_P relays, T_d is the time required for data transmission, and T_{N_P} is the BT overhead of the probing process. ω is the available bandwidth and ψ_{S-R_i-D} is the spectral efficiency in bit/sec/Hz corresponding to relay R_i. By assuming decode and forward half duplex relaying ψ_{S-R_i-D} can be given as:

$$\psi_{S-R_i-D} = \min\left(0.5\log_2(1+\gamma_{S-R_i}), 0.5\log_2(1+\gamma_{R_i-D})\right), \tag{6}$$

where $\gamma_{S-R_i} = \frac{P_{r,S-R_i}^m}{\sigma_0^2}$ and $\gamma_{R_i-D} = \frac{P_{r,R_i-D}^m}{\sigma_0^2}$ are signal to noise power ratios (SNRs) from node S to relay node R_i, and from R_i to node D. $P_{r,S-R_i}^m$ and P_{r,R_i-D}^m are the mmWave received powers at R_i from S and at D from R_i, respectively, and σ_0^2 indicates the noise power.

For the relay probing process, we follow that given in [7] and [8]. In this strategy, S node starts probing relay R_i using BT with a time duration of T_{BT}. If $P_{r,S-R_i}^m \ge P_{r,th}^m$, where $P_{r,th}^m$ is the threshold power sufficient for constructing the mmWave $S-R_i$ link, the R_i-D link is probed by node R_i; otherwise, S node will probe another candidate relay. Based on this strategy, T_{N_P} given in (5) can be calculated as:

$$T_{N_P} = \sum_{i=1}^{N_P} T_{BT}\left(1 + \mathbf{1}_{P_{r,th}^m}(P_{r,S-R_i}^m)\right), \tag{7}$$

where $\mathbf{1}_{P_{r,th}^m}\left(P_{r,S-R_i}^m\right)$ is an indicator function, which equals to 1 if $P_{r,S-R_i}^m \ge P_{r,th}^m$ and 0 otherwise.

From (5), increasing N_P will definitely increase the opportunity of finding out the best relay having the maximum ψ_{S-R_i-D}, but it will increase T_{N_P} as given in (7) and reduce the achievable throughput accordingly. In this paper, the high analogy between the mmWave relay probing tradeoff and the exploitation-exploration dilemma of the MAB problem motivates us to reformulate it as a MAB game. Then, both S-MAB and S-CMAB algorithms are proposed to efficiently address it, as given in the following section.

IV. Proposed S-MAB and S-CMAB Approaches

In this section, the optimization formulation of the mmWave relaying as a sleeping bandit MAB game will be presented followed by the proposed S-TS and S-CTS algorithms.

A. Sleeping MAB Formulation of MmWave Relay Probing

The tradeoff exists in the mmWave relay probing motivates us to reformulate it as a MAB problem, where the S node acts as the player aiming to maximize its long-term spectral efficiency, i.e., the rewards, through playing over the distributed relays, i.e., the arms of the bandit. The relays could not establish the relay link will be considered as sleeping arms and removed from the current MAB game. Mathematically speaking, this maximization problem can be formulated as:

$$\max_{\mathbb{I}(1),\ldots,\mathbb{I}(T_H)} \frac{\omega}{T_H}\sum_t\sum_i \mathbb{I}_{R_{i,t}}(o_{1_P,t}\psi_{S-R_{i,t}-D}),\ o_{1_P,t} = \frac{T_d}{T_{1_P,t}+T_d}, \tag{8}$$

s.t.

1) $T_H \in (0, Z^+)$
2) $R_{i,t} \in \phi_R$
3) $R_{i,t} \notin \phi_S$
4) $\sum_i \mathbb{I}_{R_{i,t}} = 1, 1 \le i \le N$,

where $T_H > 0$ represents the total time horizon, and Z^+ is the set of positive integer numbers. $\mathbb{I}_{R_{i,t}}$ is a linkage indicator which is equal to 1 if $R_{i,t}$ is chosen for establishing the mmWave relay link at time t and zero otherwise. The second and third constraints in (8), i.e., $R_{i,t} \in \phi_R$ and $R_{i,t} \notin \phi_S$, indicate that the selected relay should be within the set of available relays ϕ_R while it is not within the

set of sleeping relays ϕ_S. The last constraint in (8), i.e., $\sum_i \mathbb{I}_{R_{i,t}} = 1, 1 \leq i \leq N$, refers that only one relay could be selected at a time t. $o_{1_P,t}$ is the overhead resulting from probing one relay as only one relay node is probed at a time in MAB formulation, and $T_{1_P,t} = T_{BT}(1 + \mathbf{1}_{P^m_{r,th}}(P^m_{r,S-R_{i,t}}))$ is the consumed time for probing $R_{i,t}$ at time t as given in (6) where $N_P = 1$. In this paper, after probing relay $R_{i,t}$, $\Pi_{R_{i,t}} = \min\left(P^m_{r,S-R_{i,t}}, P^m_{r,R_{i,t}-D}\right)$ is evaluated, and $R_{i,t}$ is identified as a sleeping relay if the following inequality holds:

$$\Pi_{R_{i,t}} < P^m_{r,th}, \tag{9}$$

B. Proposed S-TS Algorithm

Herein, we will introduce the proposed S-TS algorithm that can address the MAB based relay probing optimization problem given in (8). TS algorithm is a pure Bayesian policy, where prior/posterior distributions based on a predefined probabilistic model are constructed for the rewards of the played arms. TS has performance guarantee better than UCB especially when the assumed probabilistic model matches that of the actual reward distribution. At the beginning of the TS algorithm, parameters are initialized for the stated model. Then, random samples are drawn from the obtained arms' distributions, and the arm corresponding to the maximum sample value will be played. After obtaining its corresponding reward, the probabilistic model parameters of the chosen arm are updated for the next round of the MAB game. Algorithm 1 gives the proposed S-TS algorithm for mmWave relay probing. Due to the log-normal shadowing terms of mmWave link models given in (2), the reward, i.e., spectral efficiency, of each relay is assumed to be taken from Gaussian distribution, i.e., $\mathcal{N}\left(\mathbb{E}\left(\psi_{S-R_{i,t}-D}\right), \sigma^2_{R_{i,t}}\right)$, where $\mathbb{E}\left(\psi_{S-R_{i,t}-D}\right)$ and $\sigma^2_{R_{i,t}}$ are its mean and variance. At the beginning of the S-TS algorithm, i.e., $t = 0$, the number of selections $x_{R_{i,t}}, \mathbb{E}\left(\psi_{s-R_{i,t}-D}\right)$ and $\sigma^2_{R_{i,t}}$ are initialized by 0, 0, and 1, respectively, for each candidate relay. Also, the space of non-sleeping relays is set to equal to the space of all available relays, i.e., $\phi_{NS,t} = \phi_R$. During the MAB game, $1 \leq t \leq T_H$, random samples $\delta_{R_{i,t-1}}$ are drawn from the Gaussian distributions of $\phi_{NS,t-1}$ relays. Then, the relay $R^*_{i,t}$ having the maximum sample value will be selected by the algorithm:

$$R^*_{i,t} = \underset{\phi_{NS,t-1}}{\arg\max}\left(\delta_{R_{i,t-1}}\right). \tag{10}$$

Subsequently and based on (9), $R^*_{i,t}$ is probed and tested if it is a sleeping relay or not. If it is considered as a sleeping relay, it will be removed from the current $\phi_{NS,t}$ set, i.e., $\phi_{NS,t} = \phi_{NS,t-1} - \{R^*_{i,t}\}$. Otherwise, its corresponding $\psi_{S-R^*_{i,t}-D}$ is obtained and its number of selections $x_{R^*_{i,t}}$ and model parameters, $\mathbb{E}(\psi_{S-R^*_{i,t}-D})$ and $\sigma^2_{R^*_{i,t}}$, are updated as given in Algorithm 1. For updating $\mathbb{E}(\psi_{s-R^*_{i,j}-D})$ and $\sigma^2_{R^*_{i,t}}$, we followed the Gaussian distribution approach given in [42] where $\mathbb{E}(\psi_{S-R^*_{i,j}-D}) = \frac{1}{x_{R^*_{i,t}}}\sum_{j=1}^{x_{R^*_{i,t}}}\psi_{S-R^*_{i,j}-D}$ and $\sigma^2_{R^*_{i,t}} = \frac{1}{x_{R^*_{i,t}+1}}$.

Algorithm 1: S-TS for mmWave two-hop relay probing

Input: ϕ_R
Initialization: $t = 0, \mathbb{E}(\psi_{S-R_{i,t}-D}) = 0, x_{R_{i,t}} = 0, \sigma^2_{R_{i,t}} = 1, \phi_{NS,t} = \phi_R$

```
for t = 1 : T_H do
    1. Sample δ_{R_{i,t-1}}, ∀R_{i,t-1} ∈ φ_{NS,t-1}, from normal distributions
       N(E(ψ_{S-R_{i,t-1}-D}), σ²_{R_{i,t-1}})
    2. Select a relay node R*_{i,t} = arg max_{φ_{NS,t-1}} (δ_{R_{i,t-1}})
    3. Eliminate the sleeping relay and update φ_{NS} or obtain its corresponding
       reward and update its model parameters
    if Π_{R*_{i,t}} < P^m_{r,th} then
        R*_{i,t} is a sleeping relay and
        φ_{NS,t} = φ_{NS,t-1} − {R*_{i,t}}
    else
        Obtain ψ_{S-R*_{i,t}-D}
        x_{R*_{i,t}} = x_{R*_{i,t-1}} + 1
        E(ψ_{S-R*_{i,j}-D}) = (1/x_{R*_{i,t}}) Σ_{j=1}^{x_{R*_{i,t}+1}} ψ_{S-R*_{i,j}-D}
        σ²_{R*_{i,t}} = 1/x_{R*_{i,t}+1}
    end
end
```

C. Proposed S-CTS Algorithm

Thanks to the multiband capable standardized WiGig devices, compromising both WiFi and mmWave bands, along with the direct relationship between WiFi and WiGig link statistics proved in [8], [21], we will utilize WiFi signal information as contexts of the mmWave relays. This will improve the mmWave relay probing process over the non-contextual counterpart towards enhancing the relaying throughput. In this framework, we will utilize the instantaneous WiFi received power, and its average value and variance up to time t as the context vector $\boldsymbol{b}_{R_{i,t}}$ of the mmWave relay $R_{i,t}$. This can be expressed as:

$$\begin{aligned}
\boldsymbol{b}_{R_{i,t}} &= \left[b_{1R_{i,t}}, b_{2R_{i,t}}, b_{3R_{i,t}}\right]^T, \\
b_{1R_{i,t}} &= \min\left(P^w_{r,S-R_{i,t}}, P^w_{r,R_{i,t}-D}\right), \\
b_{2R_{i,t}} &= \min\left(\mathbb{E}\left(P^w_{r,S-R_{i,t}}\right), \mathbb{E}\left(P^w_{r,R_{i,t}-D}\right)\right), \\
b_{3R_{i,t}} &= \min\left(\text{var}\left(P^w_{r,S-R_{i,t}}\right), \text{var}\left(P^w_{r,R_{i,t}-D}\right)\right),
\end{aligned} \tag{11}$$

where $(.)^T$ means transpose and $P^w_{r,S-R_{i,t}}$ is the instantaneous WiFi power received at $R_{i,t}$ from S as given in (1). $\mathbb{E}\left(P^w_{r,S-R_{i,t}}\right)$ and $\text{var}\left(P^w_{r,S-R_{i,t}}\right)$ indicate its average and variance values up to time t. Same definitions are applicable for $P^w_{r,R_{i,t}-D}$, $\mathbb{E}\left(P^w_{r,R_{i,t}-D}\right)$ and $\text{var}\left(P^w_{r,R_{i,t}-D}\right)$. In (11), we used the minimum value of the WiFi signal statistics of the $S-R_{i,t}$ and the $R_{i,t}-D$ links because the mmWave spectral efficiency is based on the minimum SNR of both links as given in (6). The CMAB hypothesis relies on estimating the reward of an arm given its context vector, where the reward is assumed to be a linear function of the context vector. In the case of mmWave

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3116223

relaying, this is expressed as:

$$\mathbb{E}\left[\psi_{S-R_{i,t}-D} \backslash \boldsymbol{b}_{R_{i,t}}\right] = \boldsymbol{b}_{R_{i,t}}^{T} \mathbf{Q}_{R_i}^{*}, \tag{12}$$

where $\mathbf{Q}_{R_i}^{*}$ is the optimal value of the linearity parameter. This assumption is consistent with the results given in [8] and [21], where the spectral efficiency of the mmWave link is found to be linearly related with its received WiFi RSS. The aim of the CMAB algorithms is to estimate the value of $\mathbf{Q}_{R_i}^{*}$ through online learning, where LinUCB [19] and CTS [20] are known as efficient CMAB algorithms using two different procedures for approximating $\mathbf{Q}_{R_i}^{*}$. Because CTS outperforms LinUCB [20], we will modify it by proposing S-CTS to address the formulated CMAB mmWave relaying problem while considering the sleeping relays. Like the TS algorithm, S-CTS is a pure Bayesian strategy, where $\widehat{\mathbf{Q}}_{R_i}$ is assumed to have a prior/posterior distribution given the context vector $\boldsymbol{b}_{R_{i,t}}$. In this paper, as mmWave link model follows normal distribution, multi-variate Gaussian distributions are assumed for $\widehat{\mathbf{Q}}_{R_i}$ with mean $\mathbb{E}\left(\widehat{\mathbf{Q}}_{R_i}\right)$ and variance $\alpha^2 \mathbf{B}_{R_i}^{-1}$, where α is a design parameter. In this context, $\mathbb{E}\left(\widehat{\mathbf{Q}}_{R_i}\right)$ is given as [20]:

$$\mathbb{E}(\widehat{\mathbf{Q}}_{R_i}) = \boldsymbol{B}_{R_i}^{-1} \boldsymbol{c}_{R_i}, \tag{13}$$

$$\boldsymbol{c}_{R_i} = \psi_{S-R_i-D} \boldsymbol{b}_{R_i}, \tag{14}$$

where $\mathbf{B}_{R_i}$ is the matrix of context vectors of size $k \times l$ containing the past k context vectors of length l, and $\mathbf{I}_d$ is the identity matrix of size $l \times l$, where $l = 3$ as defined in (11). Algorithm 2 summarizes the proposed S-CTS algorithm, where the inputs to the algorithm are the parameter α and the set of all available relays ϕ_R. At every time t, after parameters initialization, i.e., setting $\mathbf{B}_{R_i}, \mathbb{E}\left(\widehat{\mathbf{Q}}_{R_i}\right), \boldsymbol{c}_{R_i}$ and ϕ_{NS}, random l dimensional samples $\widehat{\mathbf{Q}}_{R_{i,t}}$ for every relay in $\phi_{NS,t-1}$ are drawn from the multi-variate Gaussian distributions $\mathcal{N}\left(\mathbb{E}\left(\widehat{\mathbf{Q}}_{R_{i,t-1}}\right), \alpha^2 \mathbf{B}_{R_{i,t-1}}^{-1}\right)$. Then, the relay node maximizing the following equation is selected:

$$R_{i,t}^{*} = \arg \max_{\phi_{NS,t-1}} \left(\boldsymbol{b}_{R_{i,t}}^{T} \widehat{\mathbf{Q}}_{R_{i,t}}\right). \tag{15}$$

If the selected relay node is identified as a sleeping relay based on (9), it will be eliminated from $\phi_{NS,t-1}$. Otherwise, its corresponding $\psi_{S-R_{i,t}^{*}-D}$ is obtained and its related parameters for posterior distribution, i.e., $\mathbf{B}_{R_{i,t}^{*}}, \boldsymbol{c}_{R_{i,t}^{*}}$, and $\mathbb{E}\left(\widehat{\mathbf{Q}}_{R_{i,t}^{*}}\right)$, are updated eventually for the next round of relay selection as given in Algorithm 2.

Algorithm 2: S-CTS for mmWave relay probing

Input: $\alpha \in \mathbb{R}^{+}, \phi_R$

Initialization: $\mathbf{B}_{R_i} \leftarrow \mathbf{I}_d, \mathbb{E}(\widehat{\mathbf{Q}}_{R_i}) \leftarrow \mathbf{0}_{l\times 1}$, $\boldsymbol{c}_{R_i} \leftarrow \mathbf{0}_{l\times 1}$ for $\forall R_i \in \phi_R, \phi_{NS,t} = \phi_R$

for $t = 1, 2, 3, \ldots, T_H$ **do**

- 1. Sample $\widehat{\mathbf{Q}}_{R_{i,t}}$ from distribution $\mathcal{N}\left(\mathbb{E}\left(\widehat{\mathbf{Q}}_{R_{i,t-1}}\right), \alpha^2 \mathbf{B}_{R_{i,t-1}}^{-1}\right)$ for $\forall R_i \in \phi_{NS,t-1}$, and observe $\boldsymbol{b}_{R_{i,t}}$
- 2. Select a relay node $R_{i,t}^{*} = \arg\max_{\phi_{NS,t-1}} \left(\boldsymbol{b}_{R_{i,t}}^{T} \widehat{\mathbf{Q}}_{R_{i,t}}\right)$
- 3. Eliminate the sleeping relay and update ϕ_{NS} or obtain its corresponding reward and update its model parameters
- **if** $\Pi_{R_{i,t}^{*}} < P_{r,\text{th}}^{m}$ **then**
  - $R_{i,t}^{*}$ is a sleeping relay and $\phi_{NS,t} = \phi_{NS,t-1} - \left\{R_{i,t}^{*}\right\}$
- **else**
  - Obtain $\psi_{S-R_{i,t}^{*}-D}$
  - $\mathbf{B}_{R_{i,t}^{*}} \leftarrow \mathbf{B}_{R_{i,t-1}^{*}} + \boldsymbol{b}_{R_{i,t}^{*}} \boldsymbol{b}_{R_{i,t}^{*}}^{T}$
  - $\boldsymbol{c}_{R_{i,t}^{*}} \leftarrow \boldsymbol{c}_{R_{i,t-1}^{*}} + \psi_{S-R_{i,t}^{*}-D} \boldsymbol{b}_{R_{i,t}^{*}}$
  - $\mathbb{E}\left(\widehat{\mathbf{Q}}_{R_{i,t}^{*}}\right) = \mathbf{B}_{R_{i,t}^{*}}^{-1} \boldsymbol{c}_{R_{i,t}^{*}}$
- **end**

end

V. Regret Analysis

In this section, we provide the expected regret of Algorithms 1 and 2. To formalize the regret analysis, we use the availability set $\phi_{NS,t}$. At time t ($\in T_H$), $R_{i,t}$ denotes the arm selected (by either algorithm) such that $R_{i,t}^{*} \in \arg\max_{R_{i,t}\in\phi_{NS,t}} \mu_{R_{i,t}}$ refers to the most plausible (i.e., optimal) arm. Therefore, the expected regret, $\mathcal{R}_{\phi_{NS}}$, for the sequence of the non-sleeping sets (i.e., $\{\phi_{NS,t}, \forall\ t \in T_H\}$), may be expressed as follows,

$$\mathcal{R}_{\phi_{NS}}(T_H) = \mathbb{E}\left[\sum_{t=1}^{T_H} \mu_{R_{i,t}^{*}} - \mu_{R_{i,t}}\right], \tag{16}$$

where $\mu_{R_{i,t}^{*}}$ and $\mu_{R_{i,t}}$ denote the mean rewards of the optimal arm (i.e., $R_{i,t}^{*}$ with hindsight) and the chosen arm (i.e., actually selected $R_{i,t}$).The expectation $\mathbb{E}$ is derived from the random choices of $R_{i,t}$ made by the algorithms. Then, the expected regret for the worst-case non-sleeping set can be represented as,

$$\mathcal{R}_{\text{S-TS}}(T_H) = \max_{\phi_{NS,t}} \mathcal{R}_{\phi_{NS}}(T_H). \tag{17}$$

First, we analyze the regret bound of Algorithm 1 based on the remark from [44], [45] that permits adaptation of the Bernoulli Thompson sampling algorithm to the general stochastic bandits case, i.e., when the rewards for arm i can be generated from an arbitrary unknown distribution with support [0, 1] to provide regret analysis of our considered sleeping MAB case. Assume that the difference of means of rewards belongs to arms i and j is denoted by $\Delta_{i,j}$. In other words, $\Delta_{i,j} = (\mu_{R_i} - \mu_{R_j})$. For generalization, note that t is dropped from the suffix of R_i and R_j. Then, the non-sleeping relay set-based regret can be re-expressed as follows.

$$\mathcal{R}_{\phi_{NS}}(T_H) \leq \sum_{i=1}^{N_p-1} \sum_{j=i+1}^{N_p} \frac{32 \ln\left(a_{j,T_H}\right)}{\Delta_{i,j}^{2}} \cdot \Delta_{i,i+1} + O(1), \tag{18}$$

where a_{j,T_H} indicates the availability count of the j^{th} arm (i.e., R_j) until time T_H. Interested readers may refer to the work in [45] for a detailed proof of (18).

For simplicity and without any loss of generalization, we may substitute a_{j,T_H} by T_H to derive $\mathcal{R}(T_H)$. Next, based on [46], $\sum_{i<j} \frac{\Delta_{i,i+1}}{\Delta_{i,j}^2}$ can be substituted by $2\sum_{i=1}^{N_P-1} \frac{1}{\Delta_{i,i+1}}$. Thus, (18) can be rewritten as:

$$\mathcal{R}_{\text{S-TS}}(T_H) \leq 64\ln(T_H) \cdot \sum_{i=1}^{N_P-1} \frac{1}{\Delta_{i,i+1}} + O(1). \quad (19)$$

Next, we evaluate the regret bound for Algorithm 2 which consists of both the contextual and sleeping MAB. For the contextual MAB part, the algorithm chooses $R_{i,t}^* = \arg\max_{\phi_{NS,t-1}} \left(\boldsymbol{b}_{R_{i,t}}^T \widehat{\mathbf{Q}}_{R_{i,t}}\right)$. Let $\Delta_i'(t)$ denotes the difference between the mean rewards of the optimal arm and arm i at time t as below,

$$\Delta_i'(t) = \boldsymbol{b}_{R_{i,t}^*}^T \widehat{\mathbf{Q}}_{R_{i,t}^*} - \boldsymbol{b}_{R_{i,t}}^T \widehat{\mathbf{Q}}_{R_{i,t}}. \quad (20)$$

Now, assuming that $\eta_{i,t} = r_i(t) - \boldsymbol{b}_{R_{i,t}}^T \widehat{\mathbf{Q}}_{R_{i,t}}$, where the first term $r_i(t)$ indicates the obtained reward by playing arm i at time t while the second term indicates its expected reward. This parameter is assumed to be conditionally α-sub-Gaussian where $\alpha \geq 0$ such that the following holds,

$$\forall\gamma \in \mathbb{R}, \mathbb{E}\left[e^{\gamma\eta_{i,t}} \mid \left\{\boldsymbol{b}_{R_{i,t}}, \mathbf{B}_{R_{i,t-1}}\right\}_{i=1}^{N_P}\right] \leq \exp\left(\frac{\gamma^2\alpha^2}{2}\right), \quad (21)$$

whenever $r_i(t) \in \left[\boldsymbol{b}_{R_{i,t}}^T \widehat{\mathbf{Q}}_{R_{i,t}} - \alpha, \boldsymbol{b}_{R_{i,t}}^T \widehat{\mathbf{Q}}_{R_{i,t}} + \alpha\right]$ [47]. Furthermore, we assume that the L2 norm is ≤ 1 for each of the vectors, $\boldsymbol{b}_{R_{i,t}}^T$, $\widehat{\mathbf{Q}}_{R_{i,t}}$, and Δ_i', so that scale-free regret bounds may be derived as follows:

$$\mathcal{R}_{\text{CTS}}(T_H) \leq l\sqrt{T_H \log(N_P)}\left(\ln(T_H) + \sqrt{\ln(T_H)\ln\left(\frac{1}{\delta}\right)}\right), \quad (22)$$

where l and $(1-\delta)$ denote the dimension of the context vector and the arm selection probability, respectively. Also, note that $0 < \delta < 1$. Interested readers may refer to [20] for the detailed proof for $\mathcal{R}_{\text{CTS}}$. Next, we further incorporate the effect of the sleeping MAB from (19) to the context-based one in (22) to derive the overall regret bound for Algorithm 2 as follows,

$$\mathcal{R}_{\text{S-CTS}}(T_H) \leq l\sqrt{T_H \log(N_P)}\left(\ln(T_H) + \sqrt{\ln(T_H)\ln\left(\frac{1}{\delta}\right)}\right) \left(64\ln(T_H) \cdot \sum_{i=1}^{N_P-1} \frac{1}{\Delta_{i,i+1}}\right). \quad (23)$$

It is worth noting that a much tighter bound than that derived in (23) could be possibly found, and this is currently an active area of research. As the regret bound of both S-TS and S-CTS algorithms is of order $o(T_H)$, the average regret converges to zero with respect to the optimal comparator in hindsight.

VI. Numerical Analysis

In this section, we will conduct numerical simulations to prove the effectiveness of the proposed S-MAB and S-CMAB approaches. In the simulation settings, 50 relays are uniformly distributed in a square area of 30×30 m^2, where S and D nodes are located in the opposite comers of the simulation area. All nodes are multi-band capable containing both the 5.25 GHz and 60 GHz bands. Without loss of generality and for fair comparisons, only mmWave LoS paths are assumed as it is the only scenario considered by the benchmark scheme given in [7]. Other simulation parameters are given in Table I.

TABLE I
Simulation Parameters.

Parameter	Value
ω	2.16 GHz [8]
T_d, T_{BT}, T_H	50 msec [8], 1 Sec [7], 500
σ_m^{LoS}, β_m^{LoS}, n_m^{LoS}	10.3 dB, 54.9 dB, 2.22 [8]
P_t^w, P_t^m and $P_{r,thr}^m$	20 dBm [8], 10 dBm [8], −78 dBm [8]
ξ and Ω	1 [22] and uniform [0.3 – 0.6] m [22]
$\theta_{-3\text{dB}}$, $\varphi_{-3\text{dB}}$	20°, 20°
α	0.1

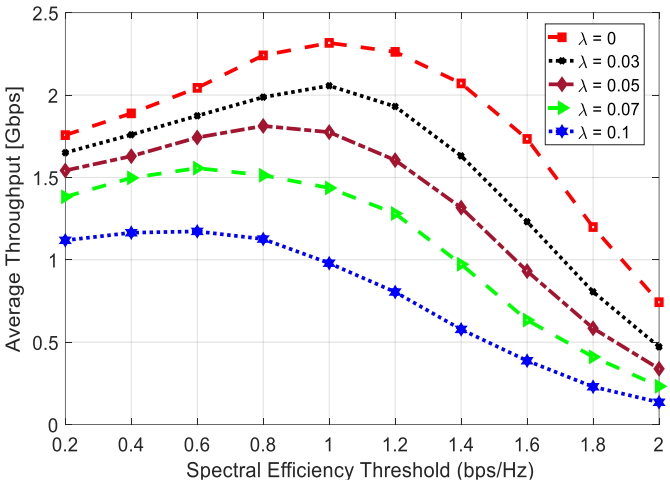

Fig. 2. Average throughput against spectral efficiency threshold at different values of λ.

Fig. 2 shows the average throughput in Gbps against the target spectral efficiency threshold in bps/Hz at different values of λ. In this figure, $\lambda = 0$ indicates no LoS blockage at all while $\lambda = 0.1$ indicates LoS blockage probability of 70%. Obviously, a tradeoff exists between increasing the target spectral efficiency threshold and the obtained average throughput. For low target values of spectral efficiency, low values of average throughput are obtained. As we increase the target spectral efficiency, the average throughput is increased till reaching a peak point after which the average throughput decreases again. This due to the increase in BT overhead coming from probing many relays. Moreover, as the value of λ is increased, the average throughput is decreased at all target values of spectral efficiency due to the effect of LoS blockage.

Fig. 3 tunes the value of α in the S-CTS algorithm at different values of λ. Generally, as λ is increased, the average throughput is decreased at all values of α due to the increase in the LoS blockage probability. At low values of α, low variance of the multivariate Gaussian distribution occurs, which decreases the exploration capability of the S-CTS algorithm and

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3116223

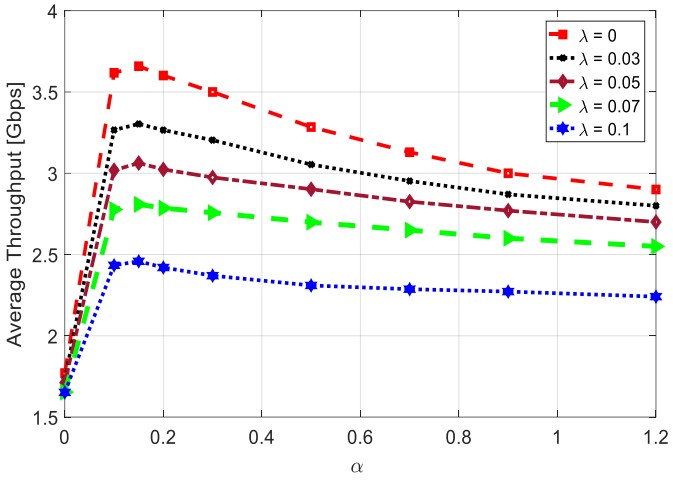

Fig. 3. Average throughput against the value of α at different values of λ.

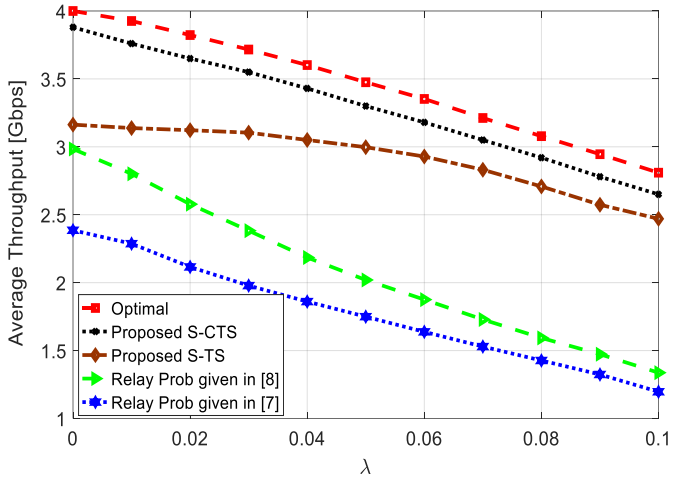

Fig. 4. Average throughput comparisons against the value of λ.

gets the exploitation term the most dominant. In consequence, low average throughput is obtained at low values of α. As α is increased, the variance of the multivariate Gaussian distribution is increased enhancing the exploration capability of the S-CTS algorithm and improves the obtained average throughput accordingly. As the α value is further increased, the exploration term becomes the most dominant causing the average throughput to reduce again. From Fig. 3, $\alpha = 0.1$ is selected as the optimal value for the proposed S-CTS algorithm as given in Table I.

Fig. 4 shows the average throughput of the relay probing schemes involved in the comparisons at different values of λ. Also, the optimal performance, where the maximum spectral efficiency is obtained by just probing the best relay, is given. As λ is increased, the average throughput of the schemes are decreased influenced by the high blockage probability. As shown in this figure, the proposed S-CTS nearly matches the optimal performance while that proposed in [7] shows the worst performance. The average throughput values of the scheme given in [7] are to the peaks of the curve given in Fig. 2. Moreover, the proposed S-CTS outperforms the

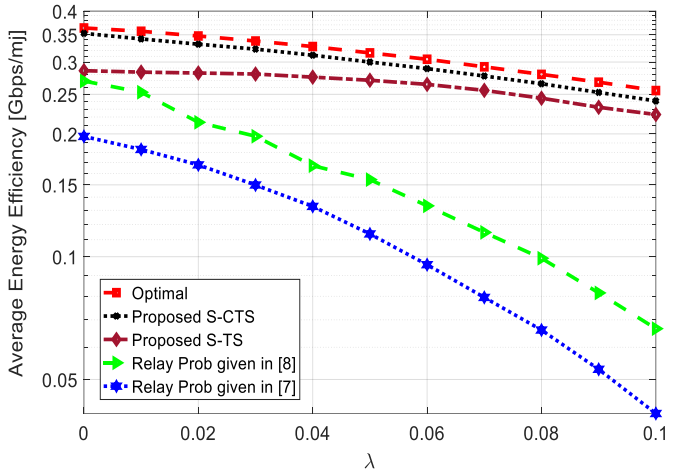

Fig. 5. Average energy efficiency comparisons against the value of λ.

proposed S-TS due to the use of additional WiFi context information. At $\lambda = 0$, the proposed S-CTS and S-TS obtain 97% and 79% of the optimal performance, respectively. However, the benchmark schemes given in [7] and [8] obtain 59.5% and 74.5% of the optimal performance, respectively. At $\lambda = 0.1$, these values become 95%, 88%, 47.8%, and 70%, for S-CTS, S-TS, and schemes given in [7] and [8], respectively.

Fig. 5 shows the average energy efficiency (EE) in Gbps/mj of the schemes involved in the comparisons against the value of λ. By neglecting the tiny overhead of the WiFi signaling, the average EE is calculated as:

$$\text{EE} = \frac{\text{TH}}{P_t^{\text{m}} \left(T_{N_P} + T_d\right)}, \tag{24}$$

where TH is the average throughput given in Fig. 4 and T_{N_P} is given in (7). Again, as the value of λ is increased, EE of all compared schemes are decreased due to the effect of harsh blockage. In the case of the optimal and the proposed S-TS and S-CTS schemes, the number of propped relays N_P is always equal to 1 irrespective the value of λ. However, Fig. 6 shows the average number of probed relays against the value of λ of the schemes given in [7] and [8]. Obviously, as λ is increased, both schemes require a high number of probed relays to obtain their corresponding average throughput given in Fig. 4. Moreover, the scheme given in [7] requires a higher number of probed relays than that given in [8]. This is the reason why the EE of both schemes decreases quickly as λ is increased as shown in Fig. 5. However, the proposed S-CTS almost matches the optimal performance and better than S-TS due to the utilization of WiFi contexts. At $\lambda = 0$, the proposed S-CTS reaches 97% of the optimal performance while the proposed S-TS reaches 79%. However, the schemes given in [7] and [8] reach 54% and 74% of the optimal performance, respectively. At $\lambda = 0.1$, these values become 95%, 88%, 16.08%, and 26%, for S-CTS, S-TS, the schemes given in [7] and [8], respectively.

Figs. 7 and 8 show the spectral efficiency convergence rate of the proposed S-CTS and S-TS schemes against the optimal value at $\lambda = 0$ and $\lambda = 0.1$, respectively. For the sake of comparisons, the spectral efficiencies of the schemes

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3116223

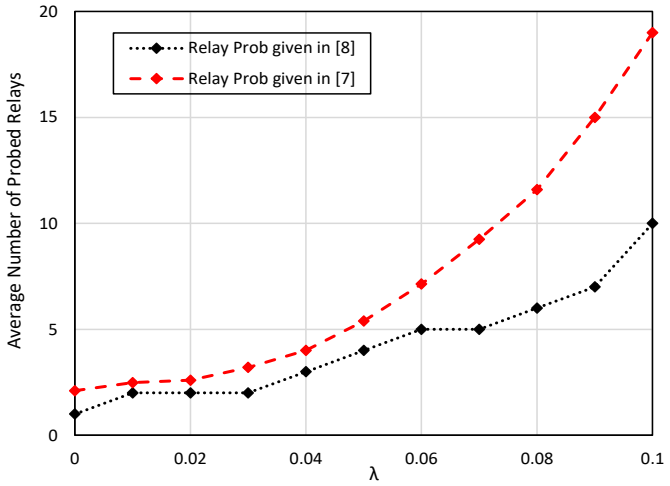

Fig. 6. Average number of probed relays of the benchmark schemes given in [7], [8].

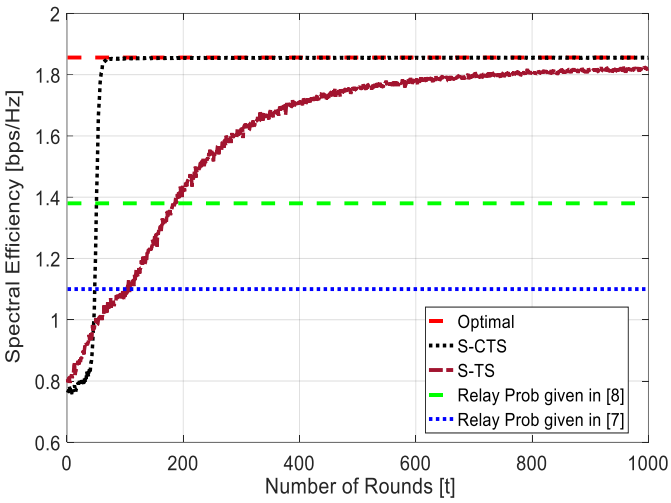

Fig. 7. Spectral efficiency convergence rate against the number of rounds t at $\lambda = 0$.

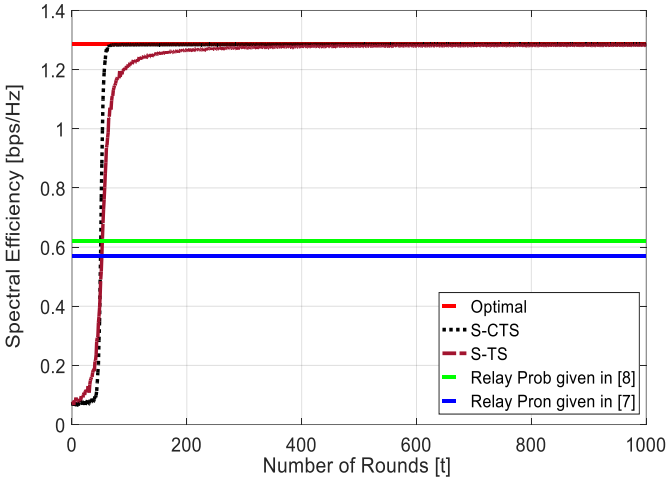

Fig. 8. Spectral efficiency convergence rate against the number of rounds t at $\lambda = 0.1$.

given in [7] and [8] are also presented. From these figures, at low number of rounds, the S-TS shows slightly better convergence rate than S-CTS because the S-CTS starts to learn the relation between the WiFi contexts and the achievable spectral efficiencies of the played relays. After the algorithm learns this relationship, the S-CTS starts to converge faster than the S-TS algorithm as shown in Figs. 7 and 8. Yet, the proposed S-CTS matches the optimal performance at both λ values after 100 rounds. However, the proposed S-TS never reaches the optimal performance even after 1000 rounds at $\lambda = 0$, while it reaches the optimal performance after 400 rounds at $\lambda = 0.1$. The schemes given in [7] and [8] are far from the optimal performance at both values of λ

For complexity analysis, the time consumed by the relay probing schemes come from two main sources. The first one comes from the relay probing time, and the second one comes from computational complexity of the algorithms. The first source is considered as the major source of time complexity as one relay probing using BT consumes about 50 msec as given in [8] and stated in Table I. At each round, the proposed MAB algorithms, i.e., S-TS and S-CTS, probe one relay only at a time for all λ values, which is the same as the optimal relay probing strategy. However, the benchmark schemes proposed in [7], [8] require an average number of probed relays of 2, 1 and 19, 10, at $\lambda = 0$ and $\lambda = 0.1$ respectively, as given in Fig. 6. Thus, at $\lambda = 0$, the total relay probing time, i.e., BT from S-R and from R-D, of the compared schemes will be 100 msec, 100 msec, 200 msec, and 100 msec for S-CTS, S-TS, the schemes given in [7] and [8], respectively. These values become 100 msec, 100 msec, 1.9 sec, and 1 sec at $\lambda = 0.1$, respectively. This comes with a near-optimal performance of the proposed S-CTS as given in Figs. 7 and 8.

For the second source of time complexity, the main sources of computational complexity for S-TS come from sampling 1-dimensional Gaussian random variable and updating its related parameters with complexity of $O(|\emptyset_{NS}| + 1)$ where $|\emptyset_{NS}|$ is the cardinality of the non-sleeping relays space. By analogy, for the S-CTS algorithm, the two main sources of computational complexity come for sampling the multivariate Gaussian distribution and updating its related parameters including inverse matrix calculation. Thus, the total complexity of the S-CTS is of order $O\left(l^2 |\emptyset_{NS}| + l^3\right)$, where l is the length of the context vector. The term $l^2 |\emptyset_{NS}|$ corresponds to sampling the multivariate Gaussian distribution while the term l^3 corresponds to updating the multivariate Gaussian parameters. In this paper, the dimensions of l and $|\emptyset_{NS}|$ are both small enough. Thus, we can assure that, the proposed S-TS and S-CTS have a low execution complexity. Compared to the other benchmark schemes, the scheme given in [7] requires solving a fixed-point equation using iterative method assuming both the spectral efficiency distribution and blockage probability are known beforehand, which are impractical in real scenarios. The other scheme given in [8] requires two phases: namely, offline phase and online relay probing phase. This consumes a considerable execution complexity through probability calculations in the offline phase, and large number of probed relays in the online phase as given in Fig. 6. Thus, the proposed MAB schemes outperform that proposed in [7]

and [8] in time complexity as well as relaying performance.

MmWave unmanned aerial vehicle (UAV) relaying network is considered as one of the promising use cases of the proposed MAB based relay probing. In this scenario, access UAVs are distributed to fully cover a post-disaster area where the cellular network is completely damaged or malfunctioned. After collecting essential information from its dedicated coverage zone in the post-disaster area, access UAVs should relay their collected information through one of the distributed gateway UAVs towards the nearest survival cellular base station. Herein, mmWave relay probing problem come to the scene and the proposed MAB based algorithms provide efficient solution to this problem over the existing ones in [7] and [8]. As the proposed MAB based schemes prob one relay at a time, the battery energy of the UAVs, which is a critical factor in UAV communication, will be highly preserved compared to the use of the existing techniques. Furthermore, the capability of excluding sleeping arms, i.e., the gateway UAVs with non-sufficient remaining battery capacity, will highly preserve the energy of the UAV network and extends its overall lifetime.

VII. Conclusion

In this paper, the problem of mmWave two hop relaying was investigated. Due to the analogy between the tradeoff exists in mmWave relaying and the exploitation-exploration dilemma inherent in MAB hypothesis, the problem was formulated as a MAB game. Moreover, idle relays were identified as sleeping bandits and eliminated during the game. S-TS algorithm was proposed to implement the formulated S-MAB based mmWave relaying. Thanks to the multi-band capable mmWave devices and the direct relationship between WiFi and mmWave link statistics, WiFi signal information was utilized as contexts of the mmWave relays. Therefore, S-CTS algorithm was proposed to handle the mmWave relaying problem as well. The proposed S-MAB and S-CMAB based approaches using the proposed S-TS and S-CTS algorithms outperformed the existing approaches in relaying performance as well as time complexity. Moreover, the proposed S-CTS nearly matched the optimal performance at all tested blockage environments. The results presented in this study opens the door for applying online learning especially MAB games to address several of mmWave communication challenges.

References

[1] K. Sakaguchi, E. M. Mohamed, H. Kusano, and et al., "Millimeter-wave wireless LAN and its extension toward 5G heterogeneous networks," *IEICE Transactions on Communications*, vol. E98.B, no. 10, pp. 1932–1948, 2015, doi: 10.1587/transcom.E98.B.1932.

[2] E. M. Mohamed, K. Sakaguchi, and S. Sampei, "Wi-fi coordinated WiGig concurrent transmissions in random access scenarios," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 357–10 371, 2017, doi: 10.1109/TVT.2017.2738198.

[3] S. Zhang, J. Liu, H. Guo, M. Qi, and N. Kato, "Envisioning device-to-device communications in 6g," *IEEE Network*, vol. 34, no. 3, pp. 86–91, 2020.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013, doi: 10.1109/ACCESS.2013.2260813.

[5] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, 2013, doi: 10.1109/TAP.2012.2235056.

[6] A. Abdelreheem, E. M. Mohamed, and H. Esmaiel, "Location-based millimeter wave multi-level beamforming using compressive sensing," *IEEE Communications Letters*, vol. 22, no. 1, pp. 185–188, 2018, doi: 10.1109/LCOMM.2017.2766629.

[7] N. Wei, X. Lin, and Z. Zhang, "Optimal relay probing in millimeter-wave cellular systems with device-to-device relaying," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10 218–10 222, 2016, doi: 10.1109/TVT.2016.2552239.

[8] E. M. Mohamed, B. M. Elhalawany, H. S. Khallaf, M. Zareei, A. Zeb, and M. A. Abdelghany, "Relay probing for millimeter wave multi-hop D2D networks," *IEEE Access*, vol. 8, pp. 30 560–30 574, 2020, doi: 10.1109/ACCESS.2020.2972614.

[9] E. M. Mohamed, "Millimeter wave beamforming training: A reinforcement learning approach," *International Journal of Electronics and Telecommunications*, vol. 67, no. 1, pp. 95–102, 2021.

[10] "IEEE standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements-part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 GHz band," *IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012)*, pp. 1–628, 2012, doi: 10.1109/IEEESTD.2012.6392842.

[11] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017, doi: 10.1109/MCOM.2017.1700393.

[12] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in lte-advanced networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 1923–1940, 2015.

[13] J. Liu, H. Nishiyama, N. Kato, and J. Guo, "On the outage probability of device-to-device-communication-enabled multichannel cellular networks: An rss-threshold-based perspective," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 163–175, 2016.

[14] F. Tang, Z. M. Fadlullah, N. Kato, F. Ono, and R. Miura, "AC-POCA: anticoordination game based partially overlapping channels assignment in combined uav and d2d-based networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1672–1683, 2018.

[15] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002, doi: 10.1023/A:1013689704352.

[16] J.-Y. Audibert, R. Munos, and C. Szepesvári, "Exploration–exploitation tradeoff using variance estimates in multi-armed bandits," *Theoretical Computer Science*, vol. 410, no. 19, pp. 1876–1902, 2009, doi: 10.1016/j.tcs.2009.01.016.

[17] I. Francisco-Valencia, J. R. Marcial-Romero, and R. M. Valdovinos-Rosas, "A comparison between UCB and UCB-tuned as selection policies in GGP," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 5073–5079, 2019, doi: 10.3233/JIFS-179052.

[18] S. Agrawal and N. Goyal, "Further optimal regret bounds for thompson sampling," in *Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2013.

[19] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th International Conference on World Wide Web*, New York, NY, USA, 2010, pp. 661–670, doi: 10.1145/1772690.1772758.

[20] S. Agrawal and N. Goyal, "Thompson sampling for contextual bandits with linear payoffs," 2014.

[21] E. M. Mohamed, M. A. Abdelghany, and M. Zareei, "An efficient paradigm for multiband wigig d2d networks," *IEEE Access*, vol. 7, pp. 70 032–70 045, 2019, doi: 10.1109/ACCESS.2019.2918583.

[22] S. Wu, R. Atat, N. Mastronarde, and L. Liu, "Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 2251–2265, 2018, doi: 10.1109/TCOMM.2017.2787990.

[23] G. Ghatak, A. De Domenico, and M. Coupechoux, "Relay placement for reliable ranging in cooperative mm-wave systems," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1324–1327, 2019, doi: 10.1109/LWC.2019.2915824.

[24] Z. Chen and D. Smith, "Mmwave m2m networks: Improving delay performance of relaying," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 577–589, 2021, doi: 10.1109/TWC.2020.3026710.

[25] B. Ma, H. Shah-Mansouri, and V. W. S. Wong, "Full-duplex relaying for d2d communication in millimeter wave-based 5g networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4417–4431, 2018, doi: 10.1109/TWC.2018.2825318.

[26] Y. Wang, Y. Niu, H. Wu, B. Ai, Z. Zhong, D. O. Wu, and T. Juhana, "Relay assisted concurrent scheduling to overcome blockage in full-duplex millimeter wave small cells," *IEEE Access*, vol. 7, pp. 105 755–105 767, 2019, doi: 10.1109/ACCESS.2019.2931876.

[27] K. Belbase, C. Tellambura, and H. Jiang, "Coverage, capacity, and error rate analysis of multi-hop millimeter-wave decode and forward relaying," *IEEE Access*, vol. 7, pp. 69 638–69 656, 2019, doi: 10.1109/ACCESS.2019.2919099.

[28] Y. Zhang, M. Xiao, S. Han, M. Skoglund, and W. Meng, "On precoding and energy efficiency of full-duplex millimeter-wave relays," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1943–1956, 2019, doi: 10.1109/TWC.2019.2900038.

[29] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5g small cells," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 64–73, 2016, doi: 10.1109/MWC.2016.7498076.

[30] S. Hashima, B. M. ElHalawany, K. Hatano, K. Wu, and E. M. Mohamed, "Leveraging machine-learning for d2d communications in 5g/beyond 5g networks," *Electronics*, vol. 10, no. 2, 2021, doi: 10.3390/electronics10020169.

[31] F.-C. Kuo, C. Schindelhauer, H.-C. Wang, W.-J. Lin, and C.-C. Tseng, "D2d resource allocation with power control based on multi-player multi-armed bandit," *Wireless Personal Communications*, vol. 113, no. 3, pp. 1455–1470, 2020, doi: 10.1007/s11277-020-07313-2.

[32] A. Ortiz, A. Asadi, M. Engelhardt, A. Klein, and M. Hollick, "Cbmos: Combinatorial bandit learning for mode selection and resource allocation in d2d systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2225–2238, 2019, doi: 10.1109/JSAC.2019.2933764.

[33] S. Hashima, K. Hatano, E. Takimoto, and E. Mahmoud Mohamed, "Neighbor discovery and selection in millimeter wave D2D networks using stochastic mab," *IEEE Communications Letters*, vol. 24, no. 8, pp. 1840–1844, 2020, doi: 10.1109/LCOMM.2020.2991535.

[34] E. M. Mohamed, S. Hashima, A. Aldosary, K. Hatano, and M. A. Abdelghany, "Gateway selection in millimeter wave uav wireless networks using multi-player multi-armed bandit," *Sensors*, vol. 20, no. 14, 2020, doi: 10.3390/s20143947.

[35] X. Li, J. Liu, L. Yan, S. Han, and X. Guan, "Relay selection for underwater acoustic sensor networks: A multi-user multi-armed bandit formulation," *IEEE Access*, vol. 6, pp. 7839–7853, 2018, doi: 10.1109/ACCESS.2018.2801350.

[36] J. Zhang, J. Tang, and F. Wang, "Cooperative relay selection for load balancing with mobility in hierarchical wsns: A multi-armed bandit approach," *IEEE Access*, vol. 8, pp. 18 110–18 122, 2020, doi: 10.1109/ACCESS.2020.2968562.

[37] X. Li, J. Liu, L. Yan, S. Han, and X. Guan, "Relay selection in underwater acoustic cooperative networks: A contextual bandit approach," *IEEE Communications Letters*, vol. 21, no. 2, pp. 382–385, 2017, doi: 10.1109/LCOMM.2016.2625300.

[38] L. Yang, X. Yan, S. Li, D. B. da Costa, and M.-S. Alouini, "Performance analysis of dual-hop mixed plc/rf communication systems," *IEEE Systems Journal*, pp. 1–12, 2021.

[39] X. Gao, J. Zhang, G. Liu, D. Xu, P. Zhang, Y. Lu, and W. Dong, "Large-scale characteristics of 5.25 ghz based on wideband mimo channel measurements," *IEEE Antennas and Wireless Propagation Letters*, vol. 6, pp. 263–266, 2007, doi: 10.1109/LAWP.2007.897513.

[40] A. M. et al., "Channel models for 60 ghz wlan systems," *IEEE document 802.11-09-0334r6*, 2010.

[41] K. Belbase, C. Tellambura, and H. Jiang, "Two-way relay selection for millimeter wave networks," *IEEE Communications Letters*, vol. 22, no. 1, pp. 201–204, 2018, doi: 10.1109/LCOMM.2017.2759106.

[42] F. Wilhelmi, C. Cano, G. Neu, B. Bellalta, A. Jonsson, and S. Barrachina-Muñoz, "Collaborative spatial reuse in wireless networks via selfish multi-armed bandits," *Ad Hoc Networks*, vol. 88, pp. 129–141, 2019, doi: 10.1016/j.adhoc.2019.01.006.

[43] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2196–2211, 2015, doi: 10.1109/JSAC.2015.2435357.

[44] S. Agrawal and N. Goyal, "Analysis of thompson sampling for the multi-armed bandit problem," in *Proceedings of the 25th Annual Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, S. Mannor, N. Srebro, and R. C. Williamson, Eds., vol. 23. Edinburgh, Scotland: JMLR Workshop and Conference Proceedings, 25–27 Jun 2012, pp. 39.1–39.26.

[45] A. Chatterjee, G. Ghalme, S. Jain, R. Vaish, and Y. Narahari, "Analysis of thompson sampling for stochastic sleeping bandits," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, Aug. 2017.

[46] R. Kleinberg, A. Niculescu-Mizil, and Y. Sharma, "Regret bounds for sleeping experts and bandits," *Machine Learning*, vol. 80, no. 2–3, pp. 245–272, 2010.

[47] S. Filippi, O. Cappé, A. Garivier, and C. Szepesvári, "Parametric bandits: The generalized linear case," in *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'10, 2010, p. 586–594.

Ehab Mahmoud Mohamed (Member, IEEE) received the B.E. and M.E. degrees in electrical engineering from South Valley University, Egypt, in 2001 and 2006, respectively, and the Ph.D. degree in information science and electrical engineering from Kyushu University, Japan, in 2012. From 2013 to 2016, he has joined Osaka University, Japan, as a Specially Appointed Researcher. Since 2017, he has been an Associate Professor with Aswan University, Egypt. He has also been an Associate Professor with Prince Sattam Bin Abdulaziz University, Saudi Arabia, since 2019. His current research interests include 5G, B5G and 6G networks, cognitive radio networks, millimeter wave transmissions, Li-Fi technology, MIMO systems, and underwater communication. He is a technical committee member of many international conferences and a reviewer of many international conferences, journals, and transactions. He is the General Chair of the IEEE ITEMS'16 and IEEE ISWC'18.

Sherief Hashima received his B.Sc. and M.Sc. degrees in Electronics and Communication Engineering (ECE), with class of honors, in 2004, 2010 from Tanta and Menoufiya University, Egypt, respectively. He obtained his Ph.D degree from Egypt-Japan University of Science & Technology (EJUST), Alexandria, EGYPT at 2014. He is a post-doctoral researcher, computational learning theory team, RIKEN AIP, Japan since July 2019. He is working as assistant professor at the Engineering and scientific equipment Department, Nuclear Research Center (NRC), Egyptian Atomic Energy Authority (EAEA), Egypt since 2014. From Jan-June 2018, he was a visiting researcher at Center for Japan-Egypt Cooperation in Science and Technology, Kyushu University. He is a technical committee member in many international conferences and a reviewer in many international conferences, journals and transactions. His research interests include wireless communications, machine learning, online learning, 5G, B5G, and 6G systems, image processing, millimeter waves, nuclear instrumentation, and internet of things. He is an IEEE senior member and AAAI member.

Kohei Hatano received Ph.D. from Tokyo Institute of Technology in 2005. Currently, he is an associate professor at Research and Development Division in Kyushu University Library. He is also the leader of the Computational Learning Theory team at RIKEN AIP. His research interests include machine learning, computational learning theory, online learning and their applications.

Mostafa M. Fouda (Senior Member, IEEE) is currently an Assistant Professor with the Department of Electrical and Computer Engineering at Idaho State University, ID, USA. He also holds the position of Associate Professor at Benha University, Egypt. He received his Ph.D. degree in Information Sciences from Tohoku University, Japan in 2011. His research interests include cyber security, machine learning, blockchain, IoT, 6G networks, and smart grid communications. He has served on the technical committees of several IEEE conferences. He served as the Track Co-Chair of IEEE VTC2021-Fall. He is also a Reviewer in several IEEE Transactions and Magazines. He is an Editor of IEEE Transactions on Vehicular Technology (TVT) and an Associate Editor of IEEE Access.

Zubair Md Fadlullah (Senior Member, IEEE) is currently an Associate Professor with the Computer Science Department, Lakehead University, and a Research Chair of the Thunder Bay Regional Health Research Institute (TBRHRI), Thunder Bay, Ontario, Canada. He was an Associate Professor at the Graduate School of Information Sciences (GSIS), Tohoku University, Japan, from 2017 to 2019. His main research interests are in the areas of emerging communication systems, such as 5G New Radio and beyond, deep learning applications on solving computer science and communication system problems, UAV based systems, smart health technology, cyber security, game theory, smart grid, and emerging communication systems. He received several best paper awards at conferences including IEEE/ACM IWCMC, IEEE GLOBECOM, and IEEE IC-NIDC. He is currently editor of IEEE Transactions on Vehicular Technology (TVT), IEEE Network Magazine, IEEE Access, IEEE Open Journal of the Communications Society, and Ad Hoc & Sensor Wireless Networks (AHSWN) journal. Dr. Fadlullah is a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE), and IEEE Communications Society (ComSoc).