TBS joint optimization to serve mmWave high altitude UAVs: a counterfactual MAB approach

Pravallika Katragunta, Michel Barbeau, Joaquin Garcia-alfaro, Evangelos Kranakis

TBS Joint Optimization to Serve mmWave High Altitude UAVs: A Counterfactual MAB Approach

Pravallika Katragunta*, Michel Barbeau*, Joaquin Garcia-Alfaro†, and Evangelos Kranakis*

* School of Computer Science, Carleton University, K1S 5B6, Ottawa, Ontario, Canada

† Télécom SudParis, Institut Polytechnique de Paris, 91120, Palaiseau, France

***Abstract*—High-altitude uncrewed aerial vehicles (UAVs) with millimeter wave communication are well-suited for fifth-generation (5G) and beyond applications. UAVs may cause significant interference to ground user equipment (GUE). We consider using a moving tethered aerial base station (TBS) as an alternative to a terrestrial base station. We consider, the UAV and GUE locations as contexts to perform joint TBS location, UAV and GUE power allocation optimization in a three-dimensional environment. We propose a contextual multi-armed bandit framework using a novel counterfactual Thompson sampling (CTS) algorithm. We compare its performance against a joint optimization using vanilla Thompson sampling (TS) and single optimization TS (SOTS) approaches. Our results show that the CTS approach converges faster. We conclude that the CTS-based approach achieves better interference mitigation for both aerial and ground users.**

***Index Terms*—tethered aerial base station, unmanned aerial vehicle, multi-armed bandit, 5G and beyond, counterfactual causal inference.**

I. Introduction

Integration of uncrewed aerial vehicles (UAVs) into cellular networks is of significant interest to fifth generation (5G) and the upcoming sixth generation (6G) wireless communications [1]. In recent years, drones are used in applications such as remote drug and face mask deliveries, to improve safety in retirement villages during pandemics like COVID-19 [2]. The UAVs move rapidly in a semi-autonomous manner. They can cover large areas from high altitudes at low cost providing new opportunities.

Traditional cellular networks are primarily designed to effectively provide ground coverage. As a result, cellular-connected UAVs flying at high altitudes frequently encounter connectivity issues due to large distances from a base station (BS) and ground-tilted antennas [2], [3]. Also, high power allocation to such UAVs is not a viable option as they can interfere with neighbouring ground cells and surrounding UAVs [3]. Hence, addressing such connectivity challenges is critical to promote practical usage of the cellular-connected UAVs in future wireless networks. Recent publications [4], [5], mention uncrewed aerial vehicle (UAV)-BS or relay node as an alternative to a terrestrial BS to provide ground user equipment (GUE) services. But, such a moving BS role is impractical to high altitude UAVs due to their limited connectivity and battery standards [6]. In [7], [8], authors consider using tethered aerial base station (TBS) to serve GUEs. They have limited three dimensional (3D) mobility, compared to a UAV-BS, but they have no battery limitations. In this work, we focus on using TBS to support aerial communications for high altitude UAVs. A TBS is connected to the ground station through a tether which provides data and energy. It can be placed at high points such as building rooftops. They can favour free space path loss (FSPL) conditions. However, simultaneously serving high altitude UAVs and GUEs with limited interference is a challenging problem. This can be addressed by jointly determining the optimal power allocation of UAVs, GUEs and their link distances from TBS resulting in lower interference on the neighbouring cells. Thus, a joint optimization framework for TBS is necessary to serve high altitude UAVs and GUEs effectively in a 3D environment.

In this work, we propose a contextual multi-armed bandit (CMAB) strategy for the TBS joint optimization framework to serve high altitude UAVs and GUEs, simultaneously in downlink communications. The main contribution of this paper is to model the learning framework at TBS using counterfactual-CMAB (C-CMAB) and CMAB approaches. We consider the problem under 5G millimeter-Wave (mmWave) communications with appropriate signal-to-interference-plus noise ratio (SINR) threshold requirements. Note that we design the C-CMAB scheme by combining the traditional CMAB approach with unobserved confounders using counterfactual causal reasoning as shown in Fig. 1. Here, unobserved confounders are the variables that influence both independent and dependent variables like contextual information and joint optimization parameters, respectively. Our simulation results show that counterfactual causal inference can be beneficial for quick convergence of the TBS optimization problem.

The rest of this paper is organized as follows. Section II reviews the related work. Section III presents the system and communication model followed by the problem formulation. Section V discusses implementation details. Section VI provides simulation results. Section VII concludes the paper.

II. Related Work

Different studies highlight the importance of dedicated BS deployments, interference and handover management for high altitude UAVs operating as user equipment (UE) [9]–[11]. The field trails in [9] demonstrate a strong uplink (UL) interference over ground users with the increase in UAVs altitude levels.

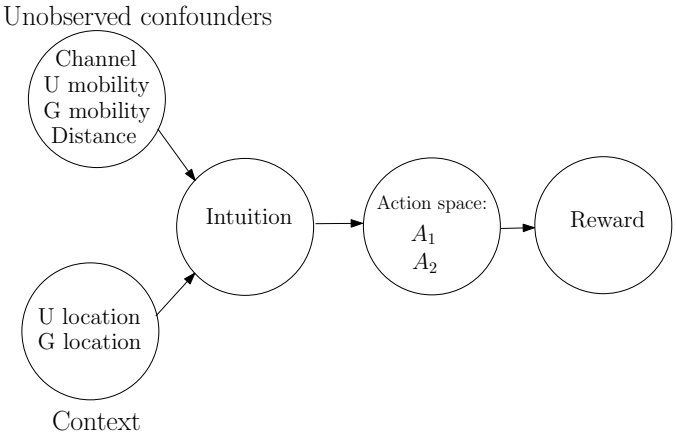


Fig. 1: Causal graph for counterfactual decision making.

In [10], authors investigate the handover issues for high speed UAVs flying at 300 m to 400 m using fourth-generation networks. These experiments show that reliable connectivity is required for high-speed flying UAVs. The authors in [11] consider up-tilted antennas on rooftop-mounted BSs and achieve high data rates for UAVs flying at a maximum altitude of 120 m. This shows that reduced link distances and dedicated BS deployments enhance connectivity for high altitude UAVs.

The authors in [4], [5] consider UAV-BS as a dedicated BSs or relay node for serving GUE. In [4], for serving moving vehicles with different velocities, the authors study the joint optimization of power and movement for UAV-BS and reduce their power consumption by 26 percent. This work highlights the joint optimization importance using a UAV-BS at a fixed altitude of 100 m. The authors in [5] use a dedicated UAV relay node in vehicular communications using MAB learning and maximize data rates by identifying optimal drone locations. This shows the learning-based optimization significance in improving drone connectivity for moving vehicles. However, the work limits only to a single vehicular source node location and does not consider interference.

Recent studies also consider TBS as a dedicated BS for GUE [7], [8]. The authors in [7] analyze the potentiality of using TBS in 6G cellular networks. The results show that TBS with 120 m tether length outperforms the maximum coverage probability of UAV-BS by nearly 30 percent. In [8], authors propose a heuristic algorithm for optimal 3D placement and transmit power of TBS using green antennas. However, a heuristic joint optimization-based interference mitigation may be less practical for large coverage environments. Instead, a learning-based joint optimization for TBS can minimize interference effects quickly and reliably.

Recent works incorporate counterfactual causal inference into classic MAB approaches [12], [13]. The authors in [12] show the benefit in CMAB performance with causal fairness over the classic CMAB algorithm. In [13], authors develop a decision-supportive system to diagnose Apple diseases by modelling a C-CMAB approach with human intuition to identify the correct diagnosis. Their results show that the unobserved con-founders affect the choice of outcome and outperform the traditional CMAB approach.

In our work, we consider TBS as the dedicated BS to serve high altitude moving UAVs and GUE. We consider TBS as the reinforcement learning (RL) agent and incorporate C-CMAB to jointly optimize the TBS location and power allocation at UAV and GUE. Under this approach, we investigate the con-founders as shown in Fig. 1 and propose a counterfactual Thompson sampling (CTS) learning algorithm by incorporating human intuition to find the optimal strategy.

III. System and Communication Model

In this section, we discuss the system model and communication model between TBS, GUE and high altitude UAVs. The objective of this problem is to perform effective interference management by jointly optimizing TBS location and UAVs, GUE power allocations.

A. System Model

We consider a single TBS, single UAV and a single GUE as shown in Fig. 2 and consider downlink (DL) communication between TBS to UAV and GUE. We observe the random

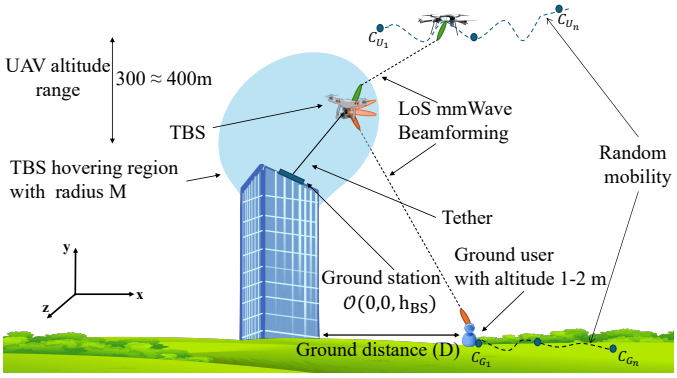


Fig. 2: System model.

mobility of UAV and GUE in a 3D Cartesian grid coverage within their regions $\mathbb{U}_{\text{UAV}}$ and $\mathbb{U}_{\text{GUE}}$ while the TBS are hovering in its limited 3D semi-spherical grid coverage area A_1.The UAVs are flying at altitudes ranging between 300 m to 400 m and GUE are moving randomly on the ground at 1 m and 2 m heights. Each UAV and GUE acts as a receiver (Rx) while the TBS acts as a transmitter (Tx) and is connected using a tether to a ground station (GS) at a position $\mathcal{O}(0,0,h_{\text{BS}})$ in a grid environment with elevation angle ξ, azimuthal angle ϕ and hovering radius δ with respect to GS. The ground distance between GS and GUE is D. The TBS locations in the grid area in Cartesian form are represented as

$$l_{\text{tbs}} = (\delta\cos(\xi_l)\cos(\phi_l), \delta\cos(\xi_l)\sin(\phi_l), \delta\sin(\xi_l) + h_{\text{BS}}) \tag{1}$$

where $l_{\text{tbs}} \in A_1$, $\xi_l \in [0, \pi/2]$, and $\phi_l \in [-\pi, \pi]$. Every time a UAV and GUE move to a new location, the SINR for the TBS-UAV link and TBS - GUE links are measured. This work considers only interference between UAVs and GUEs for

simplicity. Using these apriori SINR measurements, the TBS learns the optimal location to support reliable connectivity to each UAV and GUE grid location at the same time. We consider a simple random mobility model to evaluate the performance of the UAVs as we select locations of randomly moving UAVs as contexts.

B. Communication Model

We consider a separate narrow band line of sight (LoS) and multipath links between TBS-aerial UE and TBS-ground UE, respectively, as shown in Fig. 2. Here, TBS acts as Tx while the UAV and GUE act as the Rx. Each of the Rx and Tx are equipped with single radio frequency (RF) chain N_r and N_t uniform linear array (ULA) antennas, respectively. At each time instant, the TBS can serve a UAV and GUE with a single RF chain of the antenna arrangement. Let $\theta_{\rm tx}$ be the angle of departure (AoD) at Tx side, θ_{r1} and θ_{r2} are the angle of arrival (AoA) at UAV-Rx and GUE-Rx, respectively. Baseband equivalent of the received signal at UAV and GUE are individually given by

$$y_1[k] = \underbrace{\sqrt{P_1}\ \boldsymbol{\omega}_{\rm R1}^H \mathbf{H}_1 \boldsymbol{\omega}_{\rm T} x[k]}_{r_1[k]} + \nu_1[k],$$
$$y_2[k] = \underbrace{\sum_{m=0}^{M} \sqrt{P_2}\ \boldsymbol{\omega}_{\rm R2}^H \mathbf{H}_{2,m} \boldsymbol{\omega}_{\rm T} x[k]}_{r_2[k]} + \nu_2[k], \tag{2}$$

where P_1 and P_2 are the transmission powers considered for UAV and GUE, M represent the number of multi-paths or reflections in the TBS-GUE environment. For UAV and GUE, $\mathbf{H}_1 \triangleq \eta_1 \mathbf{a}_R(\theta_{\rm r1})\mathbf{a}_T^H(\theta_{\rm tx})$ and $\mathbf{H}_{2,m} \triangleq \eta_{2,m}\mathbf{a}_R(\theta_{\rm r2,m})\mathbf{a}_T^H(\theta_{\rm tx,m})$ are the channel matrices, $\mathbf{a}_R(\theta_{\rm r1}) \in \mathbb{C}^{N_r^{\rm UAV}}$, $\mathbf{a}_T(\theta_{\rm tx,m}) \in \mathbb{C}^{N_t}$ and $\mathbf{a}_R(\theta_{\rm r2,m}) \in \mathbb{C}^{N_r^{\rm GUE}}$, $\mathbf{a}_T(\theta_{{\rm tx},m}) \in \mathbb{C}^{N_t}$ are the array response vector pairs corresponding to $m^{\rm th}$ communication links, respectively. $\theta = \theta_{\rm r1}, N = N_r^{\rm GUE}$ for $\mathbf{a}_R(\theta_{\rm r2})$, and $\theta = \theta_{\rm tx}, N = N_t$, for $\mathbf{a}_T(\theta_{\rm tx})$, respectively. $\boldsymbol{\omega}_{\rm R1} \in \mathbb{C}^{N_r^{\rm UAV}}$ and $\boldsymbol{\omega}_{\rm R2} \in \mathbb{C}^{N_r^{\rm GUE}}$, $\boldsymbol{\omega}_{\rm T} \in \mathbb{C}^{N_t}$ are the transmitting and receiving unit-norm beamforming vectors, $\Delta = \lambda/2$ is the antenna element spacing, $\lambda = c/f$ is the wavelength, c is the velocity of light, f is the mmWave carrier frequency. $\nu_1[k], \nu_2[k] \sim \mathcal{CN}(0, WN_0)$ are the additive white Gaussian noise (AWGN) noises with zero mean and two-sided power spectral density $\frac{N_0}{2}$, $x[k]$ is the $k^{\rm th}$ sample of the time-domain transmitted signal from TBS to both UAV and GUE respectively, with equal bandwidth W and $\frac{1}{K}\sum_{k=0}^{\rm K}|x[k]|^2 = 1$. In this work, we assume η_1 and η_2 to follow free space propagation (FSP) LoS and multipath conditions, respectively between TBS and UAV, GUE radio units [14]. We define $r_1[k] = \sqrt{P_1}\,\boldsymbol{\omega}_{\rm R1}^H \mathbf{H}_1 \boldsymbol{\omega}_{\rm T} x[k]$ and $r_2[k] = \sum_{m=0}^{M}\sqrt{P_2}\boldsymbol{\omega}_{\rm R2}^H \mathbf{H}_{2,m}\boldsymbol{\omega}_{\rm T} x_2[k]$ and consider succesive interference cancellation (SIC) at GUE by assuming UAV as the strong UE [15]. Then, the SINR for UAV and GUE (assuming $P_1 < P_2$) are given as

$$\text{SINR}_{\rm U} = \frac{P_1|r_1[k]|^2}{P_2|r_2[k]|^2 + N_0}, \quad \text{SINR}_{\rm G} = \frac{P_2|r_2[k]|^2}{N_0}. \tag{3}$$

IV. Problem Formulation

We consider a 5G new radio (NR) mmWave communication DL from a TBS to a UAV. The UAV and GUE are denoted as U, with location u, and G, with location g. The sets $\mathbb{U}_{\rm UAV}$ and $\mathbb{U}_{\rm GUE}$ denote all U and G locations. The TBS acts as a learning agent. It is initially placed randomly. Starting from the initial context (u, g), the TBS repeatedly applies actions, until it reaches a location satisfying a SINR threshold for both U and G. Let A and rew denote the action set and reward parameters. At any instant, the action a in A of the TBS agent corresponds to the selection of its location, from A_1, and a power allocation, from A_2, for the pair (U, G). Thus, a equals to the pair (a_1, a_2), with a_1 in A_1 and a_2 in A_2. Here, a_2 is equal to the pair (p_1, p_2). The set A_2 is equal to $\{(p_1, p_2) : p_1 \in P_1, p_2 \in P_2)\}$, where P_1 and P_2 are the power allocation sets for U and G. We assume that the TBS provides the power allocation estimates for U and G. We model the joint optimization problem at the TBS using the CMAB framework, to learn sequential decision-making while U and G are mobile in a discrete-time setting. At any instant, the location pairs of U and G are treated as a context c equal to the pair (u, g), where $C = \{(u, g) : u \in \mathbb{U}_{\rm UAV}, g \in \mathbb{U}_{\rm GUE}\}$. Let a denote the action applied by the TBS agent. For a given context and action pair (c, a), let $U_{\rm SINR}(c, a)$ and $G_{\rm SINR}(c, a)$ denote the SINRs measured for the TBS-UAV and TBS-GUE links. Let rew represent the reward associated with the SINRs for the locations u and g and the action a. For the given context c, the optimal action is chosen as the one that receives the maximum reward. The objective is to find the optimal action a_c^* for each context c with maximum reward $rew(c, a)$ and regret Reg calculated as follows:

$$a_c^* = \arg\max_a \mathbb{E}[rew(c, a)], \text{s.t.} \tag{4}$$
$$rew(c, a) = \begin{cases} 1 & \text{if } U_{\rm SINR}(c, a) \geq \gamma_1 \text{ and } G_{\rm SINR}(c, a) \geq \gamma_2, \\ 0 & \text{otherwise} \end{cases}$$
$$Reg = \sum_{r=1}^{R}(p_r\ -\ q_r) \tag{5}$$

where γ_1 and γ_2 are the SINR threshold requirements for U and G. Regarding the choice of the value for γ_1, it is determined taking the average SINR for all locations in the set $\mathbb{U}_{\rm UAV}$. For γ_2, it is done in a similar way using the set $\mathbb{U}_{\rm GUE}$. p_r and q_r represent the agent's possible maximum reward and the received reward at each round r. Here, R denotes the total number of rounds performed.

We define the average regret-reduction ratio (ARR) metric as $\text{ARR} = 100 \cdot \left(\frac{1}{|C|}\sum_c \left(1 - \frac{Reg_{cmab}(c)}{Reg_{ref}(c)}\right)\right)$. $Reg_{cmab}(c)$ denotes the regret performance of the CMAB framework. $Reg_{ref}(c)$ represents the regret performance of a referred scheme with which CMAB is compared against, for instance, single optimization TS (SOTS), traditional TS-based joint optimization, and the conventional terrestrial BS approaches. We use this metric in section VI.

A. Counterfactual Causal Reasoning

Different factors may prevent an agent from obtaining the maximum reward. Addressing the cause-and-effect of such factors using counterfactual reasoning results in faster convergence with maximum reward [16]. The TBS agent learns from observed factors, such as U and G locations. However, unobserved factors, such as channel information, have impact on the outcome. The distance between U and G locations and their mobility can be considered by counterfactual reasoning.

We introduce counterfactual reasoning observing that a TBS location at any instant favors U or G, or both. Incorporating such inference into the TBS agent, as depicted in Fig. 1, leverages human intuition to reduce the exploration of the action space, leading to faster convergence. To capture intuition, we divide the TBS grid space into two layers, limited by an altitude. The location space is separated into two layers, namely, U_{fav} and G_{fav}, referred to as U's and G's favorable location sets. The flowchart of Fig. 3 captures the intuition. This approach allows the TBS to choose the optimal arm from the U_{fav} or G_{fav} sets by maximizing rewards through CTS-based learning.

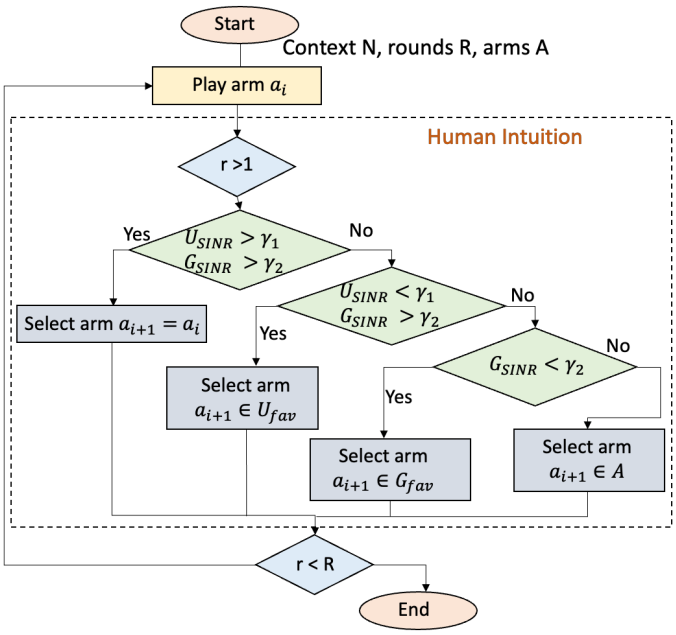


Fig. 3: Intuition flowchart. When arm a_i is played in round $r > 1$, the next arm a_{i+1} is selected following these conditions.

V. Implementation

Every TBS location is an arm of the bandit machine. At each time instant, the location pair (U, G) acts as the context. The TBS agent chooses an optimal arm from $\mathcal{A}$ and computes the reward. The objective of this approach is to identify and choose an optimal arm establishing reliable connectivity for (U, G) contextual location with the largest expected reward.

We jointly optimize the power allocation of U and G besides the TBS location. We observed that Thompson sampling (TS) performs better than upper confidence bound (UCB). Hence, we employ the TS algorithm [17], without any causality, with Eq. (4) formulation as our TS-based CMAB framework. Our counterfactual CMAB approach uses the CTS algorithm by incorporating the TS algorithm with human intuition.

A. Counterfactual Thompson Sampling

At any instant, the TBS plays an arm $a \in \mathcal{A}$ and computes the reward. If the reward is 0, then TBS chooses the next arm from the U_{fav} or G_{fav} location sets based on the conditions given in Fig. 3. The complete steps followed by CTS-based learning are shown in Algorithm 1.

Algorithm 1 *CTS-based CMAB learning*

Input: Context set C, Arm set $\mathcal{A}$
Initialization:
$SumR \leftarrow 0$
$AverageR \leftarrow 0$
$Count \leftarrow 0$
Beta Distribution$(\alpha, \beta) \leftarrow (1, 1)$
for $c \leftarrow 1$ to $NumberOfContexts$ **do**
  for $r \leftarrow 1$ to R **do**
    Sampling from the distribution of each action
    $A' \leftarrow A$
    Select a with maximum CTS(A', α, β)
    if $r > 1$ & $U_{\text{SINR}}(c, a) \geq \gamma_1$ & $G_{\text{SINR}}(c, a) \geq \gamma_2$ **then**
      $reward \leftarrow 1$
      $A' \leftarrow \{a\}$
    else if $r > 1$ & $U_{\text{SINR}}(c, a) \geq \gamma_1$ & $G_{\text{SINR}}(c, a) \leq \gamma_2$ **then**
      $reward \leftarrow 0$
      $A' \leftarrow U_{fav}$ arms
    else if $r > 1$ & $U_{\text{SINR}}(c, a) \leq \gamma_1$ & $G_{\text{SINR}}(c, a) \geq \gamma_2$ **then**
      $reward \leftarrow 0$
      Select $A' \leftarrow G_{fav}$ arms
    else
      $reward \leftarrow 0$
      Select $A' \leftarrow A$
    $SumR(c, a) \leftarrow SumR(c, a) + reward(c, a)$
    $AverageR(c, a) \leftarrow \frac{SumR(c,a)}{Count(c,a)}$
    if reward = 1 **then**
      $\alpha(c, a) \leftarrow \alpha(c, a) + 1$
    else
      $\beta(c, a) \leftarrow \beta(c, a) + 1$

VI. Simulation Results

We first implement the SOTS schemes for UE power allocation with a terrestrial BS and a static TBS. This helps understanding the TBS dynamics for the joint optimization problem. Secondly, we compare the performance of the TS-based and CTS-based approaches in terms of their regret and SINR to UEs. We assume that U and G move randomly in their area $\mathbb{U}_{\text{UAV}}$ and $\mathbb{U}_{\text{GUE}}$. For simplicity, we consider a single U and a single G with five randomly selected grid positions for each UE resulting in $|C|$ equal to 25 contextual location pairs. This approach can be extended to a large number of aerial, ground users and contexts. For the human intuition, we use 130 m as the separation altitude to divide the TBS locations into U_{fav} and G_{fav}. The simulation parameters are listed in Table I.

TABLE I: Simulation parameters.

Parameter	Value
mmWave carrier frequency f	28 GHz
antenna element spacing Δ	0.5λ
Transmit power P_1(dBm)	$\{0, 2, 4, 6, 8, 10\}$
Transmit power P_2(dBm)	$\{10, 12, 14, 16, 18, 20\}$
Bandwidth W (MHz)	150
Transmit antenna elements N_{tx}	8
Receiver antenna elements N_{rx}	UAV = 2,GUE = 4
Noise power density $\left(N_0^{\text{UAV}}, N_0^{\text{G-UE}}\right)$	$(-100, -160)$ dBm/Hz
γ_1, γ_2 (in dB)	10, 5
TBS Max.hovering radius	110 m
TBS latitude angle (in °)	$[0, 180]$
TBS longitude angle (in °)	$[-180, 180]$
Coverage xloc, yloc UAV,GUE (in m)	$[-500 : 10 : 500]$
Coverage zloc UAV (in m)	$[300 : 10 : 400]$
Coverage zloc GUE (in m)	$\{1, 2\}$
$\mathbb{U}_{\text{UAV}}$, $\mathbb{U}_{\text{GUE}}$ (in m)	{xloc, yloc, zloc}
Number of contexts $\lvert C\rvert$	25
tbs grid locations $\lvert A_1\rvert$	107
No. of (P_1, P_2) pairs, $\lvert A_2\rvert$	6
No. of Arms, $\lvert A\rvert = \lvert A_1\rvert \times \lvert A_2\rvert$	642
Rounds R	10000
UAV channel	FSP-LoS [14]
GUE channel	FSP-LoS, FSP-nLoS [14]

A. SOTS Performance using Terrestrial BS and static TBS

Fig. 4 depicts the cumulative SINR across all the contextual distances between U and G for both terrestrial BS-based and static TBS-based SOTS, respectively. From the terrestrial BS plots, we observe that G achieves higher SINR compared to U due to their closer proximity to the terrestrial BS. Similarly, the static TBS favors the SINR of U compared to G. Moreover, the performance of the static TBS approach is dependent on the selected static TBS location, hindering reliable connectivity requirements. Thus, relying on a terrestrial BS and a static TBS-based SOTS do not favour simultaneous interference mitigation at both U and G UEs.

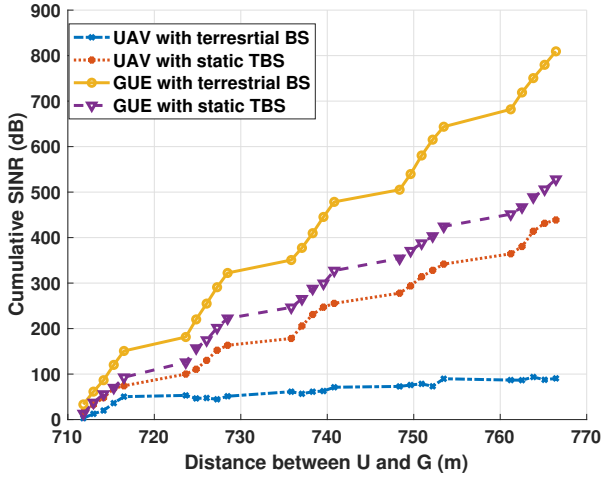


Fig. 4: Cumulative SINR performance of U and G using static TBS-based and terrestrial BS-based SOTS schemes.

B. Regret and UE Power allocation using TBS Joint Optimization Schemes

Fig. 5 shows the average regret over $\lvert C\rvert$ contexts and 10k rounds for random sampling-based, TS-based, CTS-based joint optimization approaches, and static TBS-based SOTS, respectively. We observe that all the joint optimization approaches have faster convergence with lower average regret performance compared to the static TBS-based SOTS. Furthermore, we observe that the CTS and traditional TS-based schemes converge significantly faster than the random sampling strategy with an ARR of 62.85% and 48.40%, respectively. We also observed that the runtime for CTS is 177.19 seconds, whereas for TS, it is taking 540.10 seconds to run the program to completion. Thus, learning the TBS optimal locations alongside the power allocations for each (U, G) pair is critical in satisfying the given SINR threshold requirements. Similarly, we observe that

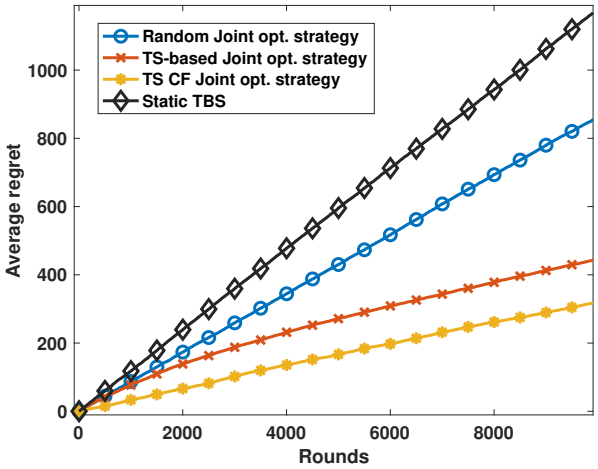


Fig. 5: Average regret over entire contexts.

the proposed CTS joint optimization strategy converges faster over the traditional-TS joint optimization scheme with an ARR of 27.99%.

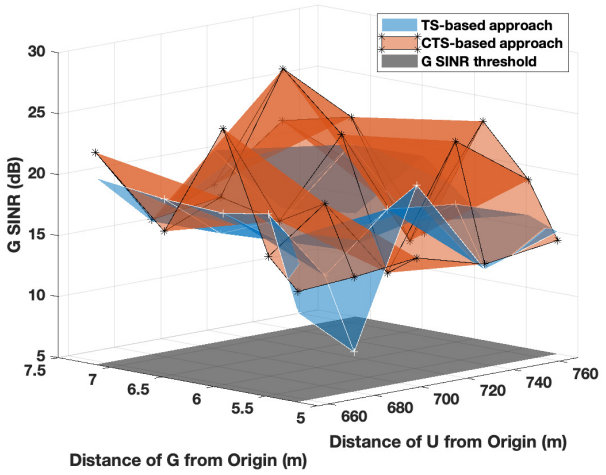
C. SINR performance CTS-based approach over Unobserved Confounders

Fig. 6 displays the U and G SINRs with TBS optimal arms for the TS-based and CTS-based approaches. Firstly, we observe that both these approaches achieve reliable SINRs for both U and G by satisfying their respective threshold requirements. Furthermore, the CTS-approach achieves relatively higher SINRs compared to the TS-based scheme for both U and G. We observe that CTS algorithm has a relatively improved exploration over large action spaces $\mathcal{A}$ and chooses arms with higher power allocation values. Thus, embedding counterfactual reasoning into the TS algorithm through intuition results in achieving reliable SINRs with better interference mitigation for both aerial and ground users.

We note that the TBS and origin are located on the same building, as shown in Fig. 2. From the plot in Fig. 6a, we observe that for a given U located from the origin, the SINR values increase by ≈ 10 dB as the G user moves ≈ 2 m away

from the TBS grid locations. Similarly, for a given G located at the origin, the SINR values increase by ≈ 15 dB as U moves away from the TBS grid locations. Similar observations can be realized in the Fig. 6b CTS plots. Thus, we observe that the mobility and separation distance of U and G influence one another (as shown in Fig. 1) in achieving the reliable SINR threshold requirements.

(a) SINR of U with TBS optimal arms satisfying γ_1 threshold requirements.



(b) SINR of G with TBS optimal arms satisfying γ_2 threshold requirements.

Fig. 6: SINR plots with optimal TBS arms in TS-based and CTS-based approaches across all (U, G) distance pairs from $O(0, 0, 0)$.

VII. Conclusion

In this paper, we proposed a C-CMAB framework to perform joint TBS location and UE power optimization for effectively serving high altitude UAVs. Under this framework, we designed the CTS-based approach by incorporating counterfactual reasoning into a TS algorithm using human intuition. We compared the performances of TS-based and CTS-based joint optimization against terrestrial BS-based and static TBS-based SOTS approaches. Also, we observed that the CTS algorithm converges faster with an ARR of 27.99% compared to the TS-based approach. We conclude that the CTS-based approach achieves reliable SINRs with better interference mitigation for both aerial and ground users. In our future works, we intend to extend these results by considering multiple ground users, UAV swarms, and multiple TBS.

Acknowledgments — This research was supported by Ericsson, Mitacs, and natural sciences and engineering research Council (NSERC) of Canada.

References

[1] G. Geraci, A. Garcia-Rodriguez, M. M. Azari, A. Lozano, M. Mezzavilla, S. Chatzinotas, Y. Chen, S. Rangan, and M. Di Renzo, "What will the future of UAV cellular communications be? A flight from 5G to 6G," IEEE communications surveys and tutorials, vol. 24, no. 3, pp. 1304–1335, 2022.

[2] R. Merkert and J. Bushell, "Managing the drone revolution: A systematic literature review into the current use of airborne drones and future strategic directions for their effective control," Journal of Air Transport Management, vol. 89, p. 101929, 2020.

[3] D. Mishra and E. Natalizio, "A survey on cellular-connected UAVs: Design challenges, enabling 5G/B5G innovations, and experimental advancements," Computer Networks, vol. 182, p. 107451, 2020.

[4] M. Nikooroo and Z. Becvar, "Optimizing transmission and propulsion powers for flying base stations," in 2020 IEEE Wireless Communications and Networking Conference (WCNC), 2020, pp. 1–8.

[5] P. Pourbaba, S. Ali, K. S. Manosha, and N. Rajatheva, "Multi-armed bandit learning for full-duplex UAV relay positioning for vehicular communications," in 2019 16th International Symposium on Wireless Communication Systems (ISWCS). IEEE, 2019, pp. 188–192.

[6] A. Fotouhi, M. Ding, and M. Hassan, "Dronecells: Improving spectral efficiency using drone-mounted flying base stations," Journal of Network and Computer Applications, vol. 174, p. 102895, 2021.

[7] M. Kishk, A. Bader, and M.-S. Alouini, "Aerial Base Station Deployment in 6G Cellular Networks Using Tethered Drones: The Mobility and Endurance Tradeoff," in IEEE Vehicular Technology Magazine, vol. 15, no. 4, 2020, pp. 103–111.

[8] Z. Lou, A. Elzanaty, and M.-S. Alouini, "Green tethered uavs for emf-aware cellular networks," in IEEE Transactions on Green Communications and Networking, vol. 5, no. 4, 2021, pp. 1697–1711.

[9] S. Homayouni, M. Paier, C. Benischek, G. Pernjak, M. Reichelt, and C. Fuchsjäger, "Field trials and design insights of cellular-connected drones," in 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall). IEEE, 2021, pp. 1–6.

[10] L. G. Militaru, D. Popescu, and L. Ichim, "4G/LTE Issues of Low Altitude UAV Flying Systems," in 2020 24th International Conference on System Theory, Control and Computing (ICSTCC). IEEE, 2020, pp. 874–879.

[11] S. Kang, M. Mezzavilla, A. Lozano, G. Geraci, S. Rangan, V. Semkin, W. Xia, and G. Loianno, "Coexistence of UAVs and Terrestrial Users in Millimeter-Wave Urban Networks," in 2022 IEEE Globecom Workshops (GC Wkshps). IEEE, 2022, pp. 1158–1163.

[12] W. Huang, L. Zhang, and X. Wu, "Achieving Counterfactual Fairness for Causal Bandit," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 6, 2022, pp. 6952–6959.

[13] G. Sottocornola, F. Stella, and M. Zanker, "Counterfactual Contextual Multi-Armed Bandit: a Real-World Application to Diagnose Apple Diseases," arXiv preprint arXiv:2102.04214, 2021.

[14] A. A. Khuwaja, Y. Chen, N. Zhao, M.-S. Alouini, and P. Dobbins, "A Survey of Channel Modeling for UAV Communications," IEEE Communications Surveys and Tutorials, vol. 20, no. 4, pp. 2804–2821, 2018.

[15] D. Tse and P. Viswanath, Fundamentals of wireless communication. Cambridge University Press, 2005.

[16] B. Shneiderman, Human-centered AI. Oxford University Press, 2022.

[17] D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, Z. Wen et al., "A tutorial on thompson sampling," Foundations and Trends® in Machine Learning, vol. 11, no. 1, pp. 1–96, 2018.