Received 8 February 2024; accepted 27 February 2024. Date of publication 5 March 2024; date of current version 21 March 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3372551

Joint UAV Trajectory Planning and LEO-Sat Selection in SAGIN

EHAB MAHMOUD MOHAMED [1] (Member, IEEE), MOHAMMAD AHMED ALNAKHLI [1] (Member, IEEE), AND MOSTAFA M. FOUDA [2] (Senior Member, IEEE)

[1]Department of Electrical Engineering, College of Engineering in Wadi Addawasir, Prince Sattam Bin Abdulaziz University, Wadi ad-Dawasir 11991, Saudi Arabia

[2]Department of Electrical and Computer Engineering, College of Science and Engineering, Idaho State University, Pocatello, ID 83209, USA

Corresponding author: E. M. MOHAMED (e-mail: ehab_mahmoud@aswu.edu.eg)

This work was supported by the Prince Sattam Bin Abdulaziz University under Project PSAU/2024/01/78915.

ABSTRACT In this paper, the problem of joint unmanned aerial vehicle (UAV) trajectory planning and low-orbit satellites (LEO-Sats) selection in space-air-ground integrated networks (SAGIN) will be investigated. This problem is of utmost importance when SAGIN is exploited for post-disaster relief services, where ground base stations (GBSs) within the post-disaster area are completely damaged or malfunctioned. In this scenario, UAV will provide wireless connectivity for the victims, while LEO-Sats will relay the UAV data to the nearest survival GBS. UAV trajectory should be optimized to maximize the collected data from the victims subject to its limited battery capacity, while it should jointly select the best LEO-Sat in each visited location within its optimized trajectory. The selected LEO-Sat should maximize UAV's achievable data rate while maintaining a long remaining visible time to minimize frequent LEO-Sats' handovers. In this paper, an online learning approach using multi-armed bandit (MAB) models will be proposed to address this highly dynamic problem. As LEO-Sat selection should be performed after UAV arrives at a dedicated location in its optimized trajectory, the problem is divided into two MAB stages. In the first stage, the battery constraint UAV trajectory optimization is modeled as budget constraint MAB (BC-MAB) game using BC-upper confidence bound (BC-UCB) algorithm. In the second stage, LEO-Sat selection in each visited location is modeled as contextual MAB with variable arms (CMAB-VA) game using LinUCB-VA algorithm. Numerical analysis confirms the superior performance of the proposed approach over candidate benchmarks.

INDEX TERMS Unmanned aerial vehicles (UAV), low Earth orbit (LEO) satellite, contextual bandit, space air ground integrated (SAGIN).

I. INTRODUCTION

THE TRANSITION from fifth generation (5G) to sixth generation (6G) wireless networks will mark a monumental shift, as delineated in [1], [2]. Notably, 6G systems are anticipated to mandate minimum data rate of 1Tbps, latency below 1msec, and spectrum efficiency of 1Gbps/m^2. Comprehensive details on these specifications, among others, are elucidated in [1]. Cutting-edge technologies, including reconfigurable intelligent surfaces (RIS) [3], harnessing of high-frequency domains such as millimeter wave (mmWave) and tera hertz (THz) bands [4], networks empowered by artificial intelligence (AI) [5], the convergence of cloud and fog networking [6], cell-free network setups [7], and the melding of space, air, and ground integrated networks (SAGIN) [8], collectively contribute to the realization of 6G networks. Within the scope of this paper, our attention predominantly converges on SAGIN, with a spotlight on the cooperation between low earth orbit satellites (LEO-Sats) and unmanned aerial vehicles (UAVs) while using AI to facilitate its operation.

LEO-Sats stationed at altitudes between 500 to 2,000 kilometers above the Earth, bring forth notable merits in terms of consistent connectivity such as swift Internet connections, voice and video conferencing, instant messaging, and data

relay, even in the most isolated regions [9]. Several endeavors, like Iridium, Globalstar, and the contemporary Starlink, have been initiated to harness LEO-Sats for delivering voice and Internet services [10]. Similarly, UAVs come equipped with a plethora of distinctive attributes, notably their aerial agility and navigation ability [11]. These attributes have garnered significant interest, paving the way for UAVs in diverse domains such as emergency response, calamity mitigation, aerial imaging, vehicular traffic regulation, postal deliveries, and as airborne communication hubs [12], [13]. For wireless communications, UAVs are adopted as flying base stations (BSs) or aerial relays for providing/ relaying wireless connectivity to far away or hard-to-reach regions. Recently, the integration between LEO-Sats and UAVs attracted a lot of researchers to take advantage of both platforms towards a complete SAGIN [14], [15], [16], [17], [18], [19].

In this paper, a SAGIN is proposed to cover a post disaster region, e.g., after earthquakes or floods. In this scenario, the terrestrial infrastructure is completely/partially destroyed including the cellular system setup, which causes destruction/malfunction to the ground base stations (GBSs) located within this catastrophic area. Thus, direct communication links between ground devices (GDs) of the survivors and the terrestrial networks cannot be maintained. Instead, a SAGIN is proposed to facilitate the rescue services, where a UAV scans the catastrophic area by taking photos, recording life videos, providing wireless connections to survivors' GDs, etc. However, both UAV and survivors' GDs cannot directly communicate with the nearby survival GBSs. This is due to the limited coverage range/cell radius of the nearby survival GBSs as typically cellular networks are planned with minimal mutual coverage (interference) between adjacent cells. In addition, a post-disaster area after earthquakes or floods is typically about several km^2 containing many destroyed/malfunctioned GBSs, which are far from nearby survival GBSs. Furthermore, it is time and energy consuming to let the UAV collect data from the area then flies toward the nearest survival GBs to offload it and then flies back to resume its rescue mission, especially in such kind of time-sensitive rescue application. Instead, LEO-Sats in SAGIN can provide an adequate solution for directly relaying UAV's communication to the nearest survival GBS. However, UAV trajectory planning (UTP) should be optimized to maximize the GDs' achievable data rates within the UAV's limited battery capacity. Concurrently, whenever the UAV reaches at a new location in its trajectory, it should opt for a LEO-Sat from its visible LEO-Sat constellation (LEO-SatCon) that maximizes the data rate of the LEO-UAV connection. This choice should consider the traffic needs of GDs within this location and ensure extended LEO-Sat visibility to minimize frequent LEO-UAV handovers. Thus, the main contributions of this paper can be summarized as follows:

- A SAGIN is proposed for facilitating the rescue operations in post-disaster scenarios, where joint UTP and LEO-Sat selection will be investigated. This optimization problem presents major challenges from two primary aspects. First, the UAV lacks knowledge about the traffic demands of GDs at each location, especially in the post-disaster scenarios. This uncertainty adds complexity to the UAV's tasks, given its limited battery life. Second, the UAV cannot determine the LEO-Sats covering a specific location until it arrives there, due to the rapid movement of both. Compounding this, the swift dynamics of LEO-Sats mean that some become visible to the UAV while others vanish from its sight within the same location. In this shifting environment, the UAV needs to choose a LEO-Sat that not only offers a data rate compatible with its current traffic needs but also guarantees extended visibility duration.
- To efficiently address this highly dynamic and time dependent optimization problem, online learning will be approached in this paper by modeling the problem as a multi-armed bandit (MAB) game [20]. Since LEO-Sat selection should be performed after UAV arrives at its intended location, we split the problem into two MAB stages. In the first MAB stage, UTP problem will be modeled as a budget constraint MAB (BC-MAB) game [21] to optimize the next location in UAV's trajectory. In the second MAB stage, LEO-Sat selection at UAV's current location will be considered as contextual MAB game [22], [23] with variable arms (CMAB-VA). To implement the proposed two-stage MAB model, budget constraint upper confidence bound (BC-UCB) will be proposed to implement the energy efficient UTP in the first MAB stage. For LEO-Sat selection in the second MAB stage, LinUCB algorithm with variable arms (LinUCB-VA) will be proposed to implement it.
- By means of numerical analysis, the proposed "BC-UCB" algorithm outperforms benchmark techniques for UTP optimization. Also, the proposed "LinUCB-VA" outperforms the state-of-the-art LEO-Sat selection techniques, which are only based on maximizing one of LEO-Sat features, such as maximum received power (MRP), maximum elevation angle (MEA), maximum available bandwidth (MBW), and maximum remaining visible time (MRVT) based LEO-Sat selection as given in [24], [25].

The remainder of this paper is organized as follows; Section II explores the literature review. Section III gives the proposed system model including link models and optimization problem formulation. Section IV gives the proposed two-stage MAB approach including both BC-UCB and LinUCB-VA algorithms. Section V gives the conducted numerical analysis followed by the concluding remarks in Section VI.

II. RELATED WORKS

Lately, SAGIN has been spotlighted as a key facilitator for forthcoming 6G networks [8]. The intricacies of SAGIN arise

from the merger of multiple network tiers. This encompasses space networks such as GEO-Sats, medium earth orbit -Sats (MEO-Sats), and LEO-Sats; aerial networks like high-altitude platform systems (HAPS), UAVs, and airships; and conventional land-based networks, namely Macro, Micro, and Pico BSs [8]. Each of these networks possesses unique specifications, making their seamless integration a tough endeavor. In [26], a model was introduced where content service providers managed content requests for SAGIN. The authors of [27] suggested a SAGIN that employed free space optics (FSO) for space-to-air connections and radio frequency (RF) for air-to-ground links. The authors of [28] presented an outage probability study for a SAGIN that includes GEO-Sat, HAPs, and terrestrial users. The authors of [29] offered a SAGIN design to amplify maritime communication by directing space signals to sea vessels via commercial passenger planes, with deep learning techniques optimizing the routing within this SAGIN.

Within the scope of SAGIN, there is a particular interest by researchers in integrating LEO-Sats with UAVs. In [14], the focus was on resource allocation and 3D trajectory fine-tuning, using UAVs as airborne BSs and delegating computing assignments to LEO-Sats. The authors of [15] introduced a cloud-edge framework tailored for the Internet of Vehicle Things (IoV) in SAGIN, where factors like service delays, energy consumption, resource usage, and security functions were holistically optimized. In [16], a SAGIN model utilizing UAVs as intermediaries to relay satellite transmissions to terrestrial receivers was proposed. UAVs featured phased-array antennas for satellite reception and deployed non-orthogonal multiple access (NOMA) for data relay to ground users. The authors of [17] described a communication strategy where FSO bridged the LEO-Sat to UAV link, and mmWave connected UAVs to ground users. In [18], the proposal was to harness both UAVs and LEO-Sats to collect data from distant sensors in the Internet of Remote Things (IoRT) within the framework of 6G networks. In [19], the idea was to use LEO-Sats and UAV caching to amplify ground network connectivity and content delivery capabilities. Recently in [30], the authors of this paper proposed LEO assisted UAV distributions in post-disaster scenario.

For UAV resource allocation and power control in SAGIN, the authors in [31] proposed hovering altitude adjustment and power control for UAVs to optimize resource allocation in UAVs network while considering interferences coming from space and ground tiers. However, this work considered neither UTP nor LEO-Sat selection. Also, the authors in [32] proposed a SAG-IoRT network, where several UAVs are used to collect IoRT information and relay it to the ground network via LEO-Sat links. The authors optimized sub-channels selection, power control and UAVs relay deployment for maximizing system energy efficiency. However, only one LEO-Sat was adopted, without considering the problem of LEO-Sat selection or UTP optimization. To the best of our knowledge, no existing studies in SAGIN, especially those focusing on LEO-UAV integration, have addressed the issue of joint UTP and LEO-Sat selection like the current work. This issue is pivotal given the rapid dynamics of LEO-Sats and UAVs especially in post-disaster scenarios.

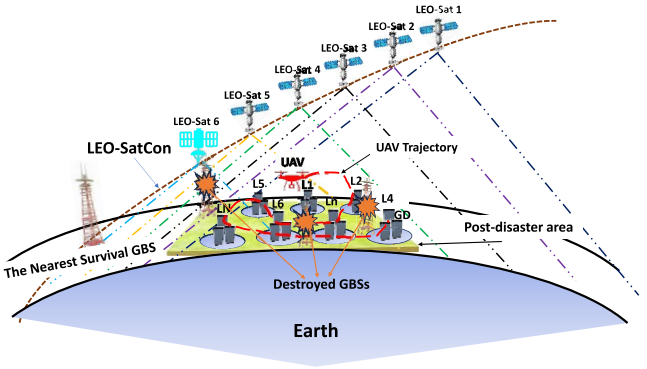

FIGURE 1. Proposed SAGIN system model for post-disaster rescue operation including joint UTP and LEO-Sat selection.

III. PROPOSED SYSTEM MODEL

Figure 1 shows the proposed SAGIN system model for post-disaster rescue operation including UTP and LEO-Sat selection. Several GBSs are destroyed inside the post disaster area making the direct connections between UAV or survivors' GDs to the terrestrial network almost impossible. Thus, UAV V is employed to cover a region containing several locations of post-disaster survivors, e.g., L_1 .., L_n, .., L_N . UAV trajectory should be optimized to cover N locations for maximizing its achievable data rate subject to its limited battery budget without any prior knowledge about the traffic needs of these locations. For covering a specific location, the UAV will hover at the center of this location at a fixed altitude to provide maximum coverage to the randomly distributed survivors within it. Simultaneously, at every visited location, UAV should select one of its M visible LEO-Sats in LEO-SatCon. Again, the UAV has no-prior knowledge about the set of LEO-Sats covering a certain location until it reaches. Moreover, this set of LEO-Sats is dynamically changing with time due to LEO-Sats movement. For example, when covering L_1 at a time instant t, the UAV falls within the coverage of LEO-Sats 1 to 5 but not covered by LEO-Sat 6. UAV should weigh various LEO-Sat characteristics for selecting the best one. These include the elevation angle of the LEO-Sat, its remaining visibility duration, its peak received power, and the remaining bandwidth it offers. Making a choice based solely on one of these features might lead to a suboptimal LEO-UAV communication link as given in [24], [25]. Given these changing conditions and varying LEO-Sats/UAV attributes, an effective method is essential to help the UAV for optimizing its UTP while picking the best LEO-Sat at each visited location n in time t. This should maximize the data rate of the whole LEO-UAV-GDs link, while considering the

remaining visible time of LEO-Sats and the limited battery budget of the UAV.

In the proposed system model, a three-tire paradigm using UAV as an intermediate node between LEO-Sat and survivors' GDs is utilized instead of two-tier paradigm via directly connecting GDs with LEO-Sats. This is because UAV typically uses communication technologies and frequency bands available for mobile communications and being used by survivors' GDs. However, in the two-tier case, GDs should have the capability of communicating with LEO-Sats within their assigned frequency bands, a functionality which is not available in most commercial cellular devices. Also, as UAV hovers at low altitude of tens of meters above GDs, low power communications can be maintained between UAV and survivors' GDs, which highly reduces their energy consumptions and prolongs the lifetimes of their batteries in consequence. In the opposite side, as the separation distances between LEO-Sats and GDs are of hundreds of kilometers, GDs should transmit at very high powers, which highly drains their batteries. In post-disaster scenarios, GDs' battery lives, and their energy consumptions are of most significance as the electricity network will be destroyed/malfunctioned with low chances for re-charging GDS' batteries. In the following, the channel models of UAV-GD and LEO- UAV links will be provided. Then, the optimization problem of joint UTP and LEO-Sat selection will be formulated.

A. UAV-GD LINK MODEL

For UAV-GD communication link, we utilized the simple link model given in [33], and without loss of generality we assume uplink transmission. In this model, the total path loss, $\beta_{Vk_n}(r_{Vk_n})$, in dB between UAV V and GD k in location n as a function of their separation distance r_{Vk_n}, can be expressed as [33]:

$$\beta_{Vk_n}(r_{Vk_n}) = \sum_{l\in\{LoS,NLos\}} \mathbb{P}^l \beta^l_{Vk_n}(r_{Vk_n}), \tag{1}$$

where

$$\beta^l_{Vk_n}(r_{Vk_n}) = 20\log\left(\frac{4\pi f_G r_{Vk_n}}{c}\right) + \alpha^l, \tag{2}$$

$l \in \{LoS, NLos\}$ indicates the line of sight (LoS) and non-LoS (NLoS) path components, while $\beta^l_{Vk_n}(r_{Vk_n})$ stands the $l-th$ path loss component of $\beta_{Vk_n}(r_{Vk_n})$. Also, f_G, c and α^l are the operating frequency of the UAV-GD link, the speed of light and the system loss due to path component l. $\mathbb{P}^l$ indicates the probability of the $l-th$ path component, where $\mathbb{P}^{LOS} = 1 - \mathbb{P}^{NLOS}$, and it can be defined as follows [33]:

$$\mathbb{P}^{LoS} = \left[1 + a\exp\left(-b\left(\varphi_{Vk_n} - a\right)\right)\right]^{-1}. \tag{3}$$

The environmental features are represented by parameters a and b. The elevation angle between UAV V and GD k in location n, denoted as φ_{Vk_n} in (3), is determined as $\varphi_{Vk_n} = \tan^{-1}(\frac{h_V}{r_{HVk_n}})$. Herein, h_V stands for the UAV's altitude, and r_{HVk_n} signifies the horizontal separation between UAV V and GD k in location n. Thus, the uplink UAV-GD data rate Ψ_{Vk_n} can be expressed as follows:

$$\Psi_{Vk_n} = \frac{W_V}{K_n}\log_2\left(1 + \frac{P_{tG}10^{\beta_{Vk_n}(r_{Vk_n})/10}}{\sigma^2}\right), \tag{4}$$

where P_{tG} is the GD transmit (TX) power, and σ^2 is the additive wight gaussian noise (AWGN) power. W_V is the total UAV's available bandwidth, and K_n is the total number of GDs in location n. In this context, we assumed no interference among GDs in location n as an interference mitigation technique, e.g., frequency division multiple access (FDMA), is used to coordinate their transmission to UAV as given in (4).

For UAV energy consumption, it mainly comes from three sources, namely flying, hovering and communication, where we neglected the other physical factors for simplicity and due to their minor effects [34]. Actually, there are eight sources of UAVs' energy consumptions as given in details in [34]. However, flying, hovering, and communication energy consumptions are the most dominant ones as shown in [34], with flying consumes more energy than other sources [34]. Thus, UAV's energy consumption $E_{Vn-1,n}$ for covering location n can be expressed as follows [3]:

$$E_{Vn-1,n} = P_f T_{f_{n-1,n}} + (P_h + P_c)T_{h_n}, \tag{5}$$

$$T_{f_{n-1,n}} = \frac{r_{n-1,n}}{S_V}, T_{h_n} = \frac{Td_n}{\Psi_{Vn}}, \quad \Psi_{V_n} = \sum_{k_n=1}^{K_n}\Psi_{Vk_n}, \tag{6}$$

where P_f, P_h, P_c are the UAV's flying, hovering, and communication powers, respectively. In this context, both P_f and P_h given in (5) are related to the mass of the UAV, the gravitational force, the radius of the propeller, and the air density. In addition, P_f depends on the deviation angle between the UAV vertical axis and the Z axis as given in [34]. For more details about various sources of UAV power consumption and their mathematical details, interested readers are advised to check [34]. $T_{f_{n-1,n}}$ is the UAV's flying duration between its current location $n-1$ and the target location n as a function of their separation distance $r_{n-1,n}$ divided by UAV's flying speed S_V in m/s as given in (6). Also, T_{h_n} is the hovering time duration of UAV over location n, which is equal to the total traffic demand of location n, Td_n, divided by its total data rate as given in (6).

B. LEO-UAV LINK MODEL

In the LEO-UAV link model, the fading channel framework outlined in [35] is adopted. Due to the relative motion of LEO-Sat and UAV, the channel model is highly time dependent [35]. This model recognizes the interactions of signals with common terrestrial obstructions like trees, buildings, and hills adjacent to the UAV. The fading pattern changes from Rician fading with higher received power at steeper elevation angles to Rayleigh fading with lower received

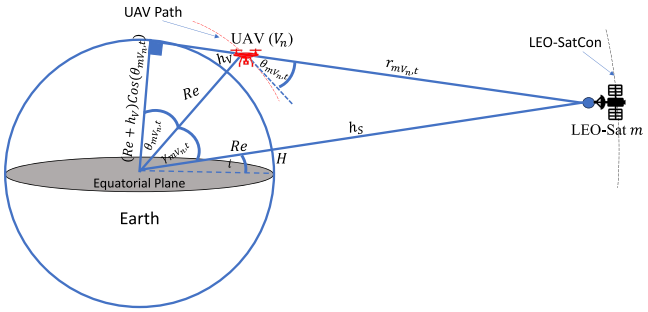

FIGURE 2. Geometrical representation of LEO-UAV communication link.

signal power at shallower elevation angles as described in [35], where the received (RX) power is formulated as [35]:

$$P_{rmV_n,t} = P_{tV} G \left(\frac{c}{4\pi f_V r_{mV_n,t}} \right)^2 \tag{7}$$

where $P_{rmV_n,t}$ is the time dependent RX power by LEO-Sat $m \in \mathcal{M}_{V_n,t}$ from UAV V when covering location n at time t assuming uplink transmissions. $\mathcal{M}_{V_n,t}$ is the set of LEO-Sats from LEO-SatCon covering UAV V at location n in time t.

P_{tV} and f_V represent UAV TX power and operating frequency, respectively. Also, G includes the product of both TX and RX antenna gains. $r_{mV_n,t}$ represents the time dependent separation distance between LEO-Sat m and UAV V hovering at location n, which is expressed as [36]:

$$\begin{aligned} r_{mV_n,t} = &-(R_e + h_V)\sin\left(\theta_{mV_n,t}\right) \\ &+\sqrt{(R_e + h_S)^2 - (R_e + h_V)^2 \cos^2\left(\theta_{mV_n,t}\right)}, \end{aligned} \tag{8}$$

where, R_e, h_V, h_S, and $\theta_{mV_n,t}$ stand the earth's radius, the UAV's altitude, the LEO-Sat's altitude, and the LEO-Sat's m elevation angle towards UAV V when covering location n in time t. These values are geometrically represented by Fig. 2. In this figure, i indicates the inclination angle of the LEO-SatCon from the equatorial plane, H represents the subsatellite point on the Earth's surface, and the time dependent trace angle is represented by $\gamma_{mV_n,t}$. In this context, $\theta_{mV_n,t}$ is expressed as [36]:

$$\theta_{mV_n,t} = \tan^{-1}\left(\frac{\cos\left((\omega_E \cos(i) - \omega_S)t + \psi_{mV_n,0}\right)\cos\left(\gamma_{V_n}\right) - \left(\frac{R_e + h_U}{R_e + h_S}\right)}{\sin\left(\cos^{-1}\left(\cos\left((\omega_E \cos(i) - \omega_S)t + \psi_{mV_n,0}\right)\cos\left(\gamma_{V_n}\right)\right)\right)} \right), \tag{9}$$

where, ω_E represents the Earth's rotational angular velocity. Also, the angular velocity of the LEO-Sat, ω_S, is connected to the LEO-SatCon orbit radius, and is written as [36]:

$$\omega_S = \sqrt{\frac{\rho}{(R_e + h_S)^3}}, \tag{10}$$

where, ρ stands for Kepler's constant, with a value of 3.986e+5 $\mathrm{km}^3/\mathrm{sec}^2$. The angle $\psi_{mV_n,t}$, illustrated in Fig. 3, refers to the angle on the Earth's surface that links H to the Q point. At $t = 0$, $\psi_{mV_n,t}$ becomes $\psi_{mV_n,0}$ which is included in (9). Moreover, γ_{V_n} is the trace angle, defining the smallest angle for the angular distance between the ground

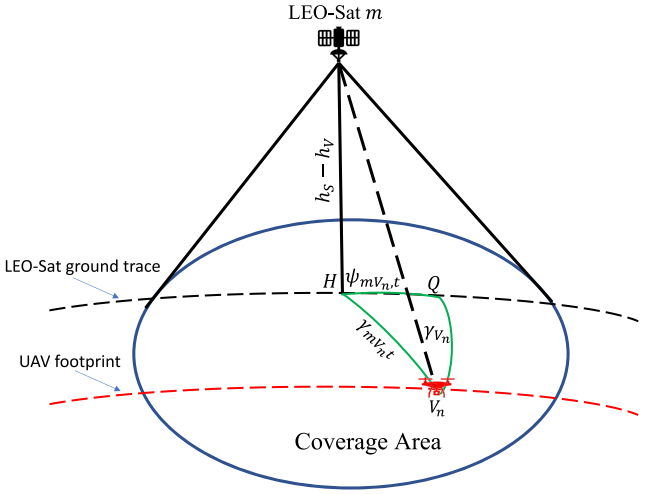

FIGURE 3. Trace angle and coverage area of LEO-Sat.

trace of the LEO-Sat and the UAV's footprint located at n, V_n, as depicted in Fig. 3. This angle is determined by the UAV's geographical position relative to the LEO-Sat. Appendix proves (9) and the value for $\psi_{mV_n,0}$. The primary source for this derivation comes from [36], but our focus is on the LEO-UAV situation, as opposed to the LEO-GD context assumed in [36]. Considering multi-path effect, the expression for the probability density function (PDF) of P_{rmV_n} is given as [35]:

$$f\left(P_{rmV_n}\right) = Ze^{-Z\left(P_{rmV_n}+1\right)} I_0\left(2Z\sqrt{P_{rmV_n}}\right). \tag{11}$$

In which Z signifies the proportion of the LoS to the multipath signals received. The term I_0 is the zero-order Bessel function. When the LEO-Sat and UAV are in relative motion at a speed of S_{mV}, the received signal characteristics change due to the Doppler effect. Under this circumstance, the Doppler frequency f_D has a relationship with the wavelength of the UAV TX signal λ_V and $\theta_{mV_n,t}$ as follows [35]:

$$f_D = \frac{|S_{mV}|}{\lambda_V} \cos\left(\theta_{mV_n,t}\right). \tag{12}$$

To acquire the impact of the Doppler effect based on $\theta_{mV_n,t}$, we can represent the Doppler power spectral density as follows [35]:

$$\delta(f_D) = \begin{cases} \frac{1}{\pi f_{D\max}\sqrt{1-\left(\frac{f}{f_{D\max}}\right)^2}}, & |f| < f_{D\max} \\ 0, & \text{elsewhere}, \end{cases} \tag{13}$$

where $f_{D\max}$ is the maximum Doppler frequency. Thus, the data rate of LEO-UAV link, when covering location n at time t, reflecting the detailed statistical link model given in (7) to (13) can be expressed as follows:

$$X_{mV_n,t} = W_{mV_n,t} \log_2\left(1 + \frac{P_{rmV_n,t}}{\sigma^2}\right), \tag{14}$$

where $W_{mV_n,t}$ indicates the available bandwidth of LEO-Sat m covering V_n at time t, and σ^2 is the AWGN power.

For simplicity, we assumed that no source of interference affects the LEO-UAV link using high interference mitigation techniques.

Apparently, the aforementioned UAV-GD and LEO-UAV channel models are different from the terrestrial network ones. They are mainly based on channel geometry like elevation angles, UAV altitude, environmental parameters, LEO-Sat altitude, Earth raduis, angular velocity of both Earth and LEO-Sat, etc. All these parameters are not affecting the channel models of the terrestrial networks as they are mainly based on the separation distance between TX and RX.

C. OPTIMIZATION PROBLEM FORMULATION

In this subsection, the problem of joint UTP and UAV centric LEO-Sat selection will be formulated. The main objective of the optimization problem is to jointly find the next location on the UAV's trajectory n_t^* and select the most appropriate LEO-Sat m_t^* covering it. The aim is to maximize the data rate of the whole LEO-UAV-GDs link while maintaining long remining LEO-Sat visible time subject to UAV's limited battery capacity and LEO-Sats' features. Mathematically speaking, this optimization problem can be formulated as follows:

$$\left\{n_t^*, m_t^*\right\} = \arg \max_{\mathcal{N}, \mathcal{M}_{V_n,t}} \left(\Sigma_{mV_n,t} \mathcal{T}_{mV_n,t}\right), \tag{15}$$

where

$$\Sigma_{mV_n,t} = \min\left(\Psi_{V_n,t}, X_{mV_n,t}\right)$$

s.t.

$$\mathcal{M}_{V_n,t} \subset \mathcal{M}_{SatCon} \tag{15.a}$$
$$E_{Vn-1,n,t} < E_{V\text{Min}} \tag{15.b}$$
$$\mathcal{T}_{mV_n,t} \in \{0, \mathcal{T}_{V_n\max}\} \tag{15.c}$$
$$\theta_{mV_n,t} \in \{\theta_0, \theta_{V_n\max}\} \tag{15.d}$$
$$W_{mV_n,t} \in \{0, W_{L\max}\} \tag{15.e}$$
$$f_D \in \{0, f_{D\max}\} \tag{15.f}$$
$$S_V \leq S_{V\max} \tag{15.g}$$

where $\mathcal{T}_{mV_n,t}$ represents the time dependent remaining visible time of LEO-Sat m when communicating with UAV V in location n. The constraint (15.a) means that $\mathcal{M}_{V_n,t}$ should be a subset of the total set of LEO-Sats $\mathcal{M}_{SatCon}$ in the LEO-SatCon. Typically, $\mathcal{M}_{V_n,t}$ is dynamically changing with time as some LEO-Sats will enter UAV visibility while others will leave it. The constraint (15.b) means that the UAV energy required to move from its current location $n-1$ to its next location n at time t in its trajectory should be less than its minimum battery energy $E_{V\text{Min}}$. $E_{V\text{Min}}$ preserves the minimum energy enabling the UAV to fly to its start location for battery re-charging. The constraint (15.c) indicates that $\mathcal{T}_{mV_n,t}$ has a range between 0 and $\mathcal{T}_{V_n\max}$, where 0 signifies that LEO-Sat m is no longer visible to the UAV at location n, while $\mathcal{T}_{V_n\max}$ represents the longest duration LEO-Sat m remains visible. $\mathcal{T}_{V_n\max}$ is equivalent to double the time taken to achieve the peak elevation angle, $\theta_{V_n\max}$. As per [36], $\mathcal{T}_{V_n\max}$ is calculated as:

$$\mathcal{T}_{V_n\max} = \frac{-2\psi_{mV_n,0}}{((\omega_E \cos(i) - \omega_S))}. \tag{16}$$

The constraint (15.d) specifies that the elevation angle, $\theta_{mV_n,t}$, ranges between its lowest value, θ_0, and its highest value, $\theta_{V_n\max}$. The calculation for $\theta_{V_n\max}$ is provided in [36] as:

$$\theta_{V_n\max} = \tan^{-1}\left(\frac{\cos(\gamma_{V_n}) - \left(\frac{R_e+h_V}{R_e+h_S}\right)}{\sin(\gamma_{V_n})}\right). \tag{17}$$

A comprehensive derivation of $\mathcal{T}_{V_n\max}$ and $\theta_{V_n\max}$ is available in [36]. The constraint (15.e) implies that the accessible bandwidth of LEO-Sat m, $W_{mV_n,t}$, must not exceed its maximum available value of $W_{L\max}$. Similarly, the constraint (15.f) indicates that the Doppler frequency falls between 0 and its uppermost value $f_{D\max}$, where (12) is used to compute $f_{D\max}$ when $\theta_{mV_n,t} = \theta_0$. The last constraint given in (15.g) indicates that S_V is bounded by its maximum value $S_{V\max}$.

The optimization problem presented in (15) is a profoundly nonlinear dynamic challenge. This kind of problems cannot be solved using traditional optimization techniques or even exhaustively searching all solution spaces at every time t. Simply because the set of visible LEO-Sats covering n_t^* is not exactly known unless UAV moves to that location. Nevertheless, this set of visible LEO-Sats is dynamically changing with time. Even conventional methods used in satellite-to-mobile user selection/handover cannot efficiently address the problem of LEO-Sat selection alone [24], [25]. This is because those traditional techniques typically focus on optimizing a single feature of the LEO-Sat, like MRP, MEA, MBW, or MRVT [24], [25]. However, for an effective solution to LEO-Sat selection, it is essential to take all these features into account at once, as dictated by the constraints.

IV. PROPOSED TWO-STAGE MAB APPROACH

As LEO-Sat selection should be performed after UAV arrives at its next location in its trajectory, the problem can be split into two stages. At the first stage, UTP is optimized subject to the limited UAV energy by finding the next location in its trajectory, while the most appropriate LEO-Sat is selected for that location subject to LEO-Sats' features in the second stage. Also, as the problem presented in (15) is a sequential time-based optimization challenge influenced by the characteristics and dynamicity of the LEO-Sats bounded by UAV's battery capacity, two stage online learning approach using two MAB models will be proposed. In the first stage, a BC-MAB model implemented by BC-UCB algorithm is proposed to find the next appropriate location in the UAV trajectory, while CMAB-VA implemented by LinUCB-VA algorithm is proposed to select the most appropriate LEO-Sat at that location in the second stage.

A. MAB MODELS

MAB is an effective online learning tool that emulates the choices for a gambler playing with a multi-armed slot machine. In this scenario, the gambler observes the rewards obtained by playing different arms of the machine [20]. The objective for the player is to consistently choose the arm that has yielded the highest reward thus far, while also trying out new arms. This is often referred to as exploitation and exploration balance in MAB games [37]. Strategies like epsilon-greedy (ϵ-greedy), upper confidence bound (UCB), and Thompson sampling (TS) [38] are recognized as effective approaches for implementing MAB models. The significance of MAB models has grown recently due to their applicability in areas like advertising, medicine, and telecommunications, among others. There have also been various adaptations of the basic MAB model to suit different real-world scenarios. In some cases, choosing an arm incurs a cost constrained by the player's available budget. This model is referred to as the BC-MAB, as presented in [21]. Another adaptation, the contextual MAB model, leverages features of the arms to hasten the decision-making process for the optimal arm, using algorithms like LinUCB [22]. Additionally, in some MAB scenarios, certain arms might become inaccessible, which is called MAB game with sleeping arms [23]. These MAB models' variations and more offer solutions and optimization strategies for complex applications, like the one discussed in this paper.

While deep reinforcement learning (DRL) is a powerful approach for sequential decision-making, it often requires extensive computational resources, substantial training datasets, and may exhibit challenges in terms of interpretability. SAGIN has a dynamic nature especially in post-disaster scenario, where real-time decision is a critical matter plus no prior information is typically available about the environment including the number of survivors and their traffic demands. Thus, MAB models were considered more practical and computationally efficient for this specific application over DRL due to the following reasons. 1) MAB models do not need any prior intensive offline training or even any pre-information about the environment like DRL. 2) No frequent updates to the constructed offline databases based on environment changes are needed as befalls in DRL. 3) Given the constraints of the UAV's onboard processing, UAV's limited battery budget, and the real-time demands of post-disaster scenarios, MAB models were deemed more suitable than DRL for efficient real-time decision-making.

B. PROPOSED TWO-STAGE MAB APPROACH

The proposed two-stage MAB approach presented in Algorithm 1 consists of BC-MAB implemented using BC-UCB algorithm in the first stage while CMAB-VA implemented using LinUCB-VA in the second stage. The inputs to the algorithm are the design parameter Ω, $\mathcal{N}$, $\mathcal{M}_{SatCon}$, the full UAV battery capacity $C_{V\text{Max}}$, and the minimum UAV energy $E_{V\text{Min}}$. The outputs of the algorithm are the selected location n_t^* and the selected LEO-Sat m_t^*

Algorithm 1 Two-Stage MAB Approach for Joint UTP and LEO-Sat Selection

Output: n_t^*, m_t^*
Input: $\Omega \in \mathbb{R}^+$, $C_{V\text{Max}}$, $E_{V\text{Min}}$, $\mathcal{N}$, $\mathcal{M}_{SatCon}$
Initialization:

- UAV visits each location in $\mathcal{N}$ at once from its start location, and observes its achievable sum rate $\Psi_{V_n,t}$ and energy consumption $E_{V_{1,n},t}$.
- UAV backs to its start location and set $\{n-1\} = 1$, $t = N$, $\overline{\Psi}_{V_n,t} = \Psi_{V_n,t}$, $\overline{E}_{V_{1,n},t} = E_{V_{1,n},t}$, $j_{n,t} = 1\ \forall n \in \mathcal{N}$, $E_{VR,t} = C_{V\text{Max}}$

While $E_{VR,t} > E_{V\text{Min}}$
$t = t + 1$

- **First Stage:** UTP using BC-UCB Algorithm
  1. $\nu_t = \min_{\mathcal{N}} \left(\overline{E}_{V_{n-1,n},t-1}\right)$
  2. $n_t^* = \arg\max_{\mathcal{N}} \left(\frac{\overline{\Psi}_{V_n,t-1}}{\overline{E}_{V_{n-1,n},t-1}} + \frac{1}{\nu_t} \left(1 + \frac{1}{\nu_t - \sqrt{\frac{\ln(t-1)}{j_{n,t-1}}}}\right) \sqrt{\frac{\ln(t-1)}{j_{n,t-1}}} \right)$
  3. Observe $\Psi_{V_{n^*},t}$ and $E_{V_{n-1,n^*},t}$
  4. $j_{n^*,t} = j_{n^*,t-1} + 1$
  5. $\overline{\Psi}_{V_{n^*},t} = \frac{1}{j_{n^*,t}} \sum_{i=1}^{j_{n^*,t}} \Psi_{V_{n^*},i}$
  6. $\overline{E}_{V_{n-1,n^*},t} = \frac{1}{j_{n^*,t}} \sum_{i=1}^{j_{n^*,t}} E_{V_{n-1,n^*},i}$
  7. $E_{VR,t} = E_{VR,t-1} - E_{V_{n-1,n^*},t}$
  8. $\{n-1\} = \{n_t^*\}$
- **Second Stage:** LEO-Sat selection using LinUCB-VA Algorithm for the selected n_t^* location.
  1. Enumerate $\mathcal{M}_{V_{n^*},t} \subset \mathcal{M}_{SatCon}$, where $\theta_{mV_{n^*},t} \geq \theta_0$ and then observe $\mathbf{y}_{mV_{n^*},t}\ \forall m \in \mathcal{M}_{V_{n^*},t}$
     for $\forall m \in \mathcal{M}_{V_{n^*},t}$
       If m is a new satellite in UAV visibility
         $\mathbf{A}_{mV_{n^*}} \leftarrow \mathbf{I}_b$
         $\mathbf{\Gamma}_{mV_{n^*}} \leftarrow 0_{l\times 1}$
       end If
       If $\theta_{mV_{n^*},t} < \theta_0$,
         LEO-Sat m is considered as a sleeping satellite and removed from $\mathcal{M}_{V_{n^*},t}$
       end If
       $\widehat{\mathbf{q}}_{mV_{n^*}} = \mathbf{A}_{mV_{n^*}}^{-1} \mathbf{\Gamma}_{mV_{n^*}}$
     End for
  2. $m_t^* = \arg\max_{\mathcal{M}_{V_{n^*},t}} \left(\mathbf{y}_{mV_{n^*},t}^T \widehat{\mathbf{q}}_{mV_{n^*}} + \Omega \sqrt{\mathbf{y}_{mV_{n^*},t}^T \mathbf{A}_{mV_{n^*}}^{-1} \mathbf{y}_{mV_{n^*},t}} \right)$
  3. **Observe** $X_{m^*V_n,t}$
  4. $\mathbf{A}_{m^*V_{n^*},t} \leftarrow \mathbf{A}_{m^*V_{n^*},t} + \mathbf{y}_{m^*V_{n^*},t} \mathbf{y}_{m^*V_{n^*},t}^T$
  5. $\mathbf{\Gamma}_{m^*V_{n^*},t} \leftarrow \mathbf{\Gamma}_{m^*V_{n^*},t} + X_{m^*V_n,t} \mathbf{y}_{m^*V_{n^*},t}$

End While

at time t. For initialization, the UAV explores all locations, $\forall n \in \mathcal{N}$, from its start position, $n = 1$, and observes their achievable sum rates $\Psi_{V_n,t}$ and energy consumptions $E_{V_{1,n},t}$. Then, UAV backs to its start location, and the following parameters are initialized: $t = N$, $\{n-1\} = 1$, $\overline{\Psi}_{V_n,t} = \Psi_{V_n,t}$, $\overline{E}_{V_{n-1,n},t} = E_{V_{n-,n},t}$, $j_{n,t} = 1\ \forall n \in \mathcal{N}$, $E_{VR,t} = C_{V\text{Max}}$. $j_{n,t}$ indicates how many times location n was selected up to time t, and $E_{VR,t}$ indicates the remining UAV energy at time t. Also, $\overline{\Psi}_{V_n,t}$ and $\overline{E}_{V_{n-1,n},t}$ are the average data rate of location n and energy consumptions of UAV when moving from location $n-1$ to location n at time t, respectively. These

values are initially set to $\Psi_{V_{n},t}$ and $E_{V_{n-,n},t}$ calculated using (5) and (6), respectively. After initialization, the two-stage MAB approach is conducted as long as $E_{VR,t} > E_{V\text{Min}}$ as follows.

1) FIRST STAGE: UTP USING BC-UCB ALGORITHM

In this BC-MAB stage, the UAV will act as the player, the arms of the MAB game will be the set of distributed locations $\mathcal{N}$, and the rewards will be the UAV achievable data rates $\Psi_{V_{n},t}$. Also, the game is bounded by UAV's energy budget $E_{V\text{Min}}$. In this context, BC-UCB algorithm is proposed to find the next UAV location n_t^* in its trajectory. UCB is the most effective MAB algorithm that can well address the balance between exploration and exploitation [20]. In UCB, exploitation refers to the average rewards garnered from the arms that have been played, while exploration denotes the number of times each arm has been played. Essentially, the algorithm aims to maximize the confidence in the selected arm by minimizing its uncertainty. In the BC-UCB variant, the arms budget is added to the naive UCB equation, where not only the average reward but also the average budget of the arms is included in the exploitation term. Also, a parameter indicating the minimum average budget over all arms is included in the exploration term. The proposed BC-UCB algorithm is influenced by that presented in [39]. At the beginning of the algorithm after increasing t by 1, the parameter ν_t is evaluated. Then, the location n_t^* maximizing the following formula is selected to be the next location in the UAV's trajectory:

$$n_t^* = \arg\max_{\mathcal{N}} \left(\frac{\overline{\Psi}_{V_{n},t-1}}{\overline{E}_{V_{n-1,n},t-1}} + \frac{1}{\nu_t} \left(1 + \frac{1}{\nu_t - \sqrt{\frac{\ln(t-1)}{j_{n,t-1}}}} \right) \sqrt{\frac{\ln(t-1)}{j_{n,t-1}}} \right), \tag{18}$$

Then, the values of $\Psi_{V_{n^*},t}$ and $E_{V_{n-1,n^*},t}$ corresponding to n_t^* are observed by the UAV. Then, its related parameters $j_{n^*,t}$, $\overline{\Psi}_{V_{n^*},t}$ and $\overline{E}_{V_{n-1,n^*},t}$ are updated as given in Algorithm 1. Also, $E_{VR,t}$ is updated, and the index of location $n-1$ is set to equal n_t^* for the next round of location selection as given in Algorithm 1.

2) SECOND STAGE: LEO-SAT SELECTION USING LINUCB-VA ALGORITHM

After choosing n_t^*, UAV flies towards it to collect its data and then selects the most appropriate LEO-Sat m_t^* from $\mathcal{M}_{V_{n^*},t}$ to cover it. In this stage, the player will also be the UAV, $\mathcal{M}_{V_{n^*},t}$ will act as the arms of the MAB game, and the rewards will be the achievable data rates of the LEO-UAV links. In should be noted that these arms are variable with time due to appearance of new LEO-Sats and disappearance of old ones. Also, $\theta_{mV_{n^*},t}$, $\mathcal{T}_{mV_{n^*},t}$ and $W_{mV_{n^*},t}$ will be the features of the arms, while the effect of f_D is included in $\theta_{mV_{n^*},t}$ as given in (12). In this paper, to implement this CMAB-VA game, LinUCB-VA algorithm is proposed as given in Algorithm 1. The initial version of LinUCB was introduced in [21] for CMAB models, where it linearly projects the time dependent context vector onto the reward space. The proposed LinUCB-VA is an advanced version of LinUCB that incorporates the dynamic nature of LEO-Sat arms. Thus, the b dimension features vector of LEO-Sat m when covering location n_t^* can be represented as:

$$\mathbf{y}_{mV_{n^*},t} = \left[\theta_{mV_{n^*},t}, W_{mV_{n^*},t}, \mathcal{T}_{mV_{n^*},t}\right]^T. \tag{19}$$

For each time instant t, the GBS in the SAGIN communicates this feature vector to the UAV. Within this framework, $\theta_{mV_{n^*},t}$ can be deduced using (9). $\mathcal{T}_{mV_{n^*},t}$ starts from its maximum value at $t=0$ when LEO-Sat m becomes visible to the UAV at location n_t^* and reduces at each subsequent time instant until it becomes zero when LEO-Sat m disappears from UAV visibility. The value of $W_{mV_{n^*},t}$ for each LEO-Sat m can be relayed to the UAV by the GBS via its connected LEO-Sat at time t.

CMAB was first presented in [21] to address personalized article suggestions by factoring in historical user choices associated with article content. In this model, the payoff $X_{mV_{n^*},t}$ is said to be correlated with the features vector $\mathbf{y}_{mV_{n^*},t}$. In the problem under consideration, this relation holds true since $X_{mV_{n^*},t}$ connects to $\theta_{mV_{n^*},t}$, $W_{mV_{n^*},t}$, and $\mathcal{T}_{mV_{n^*},t}$. Both $W_{mV_{n^*},t}$ and $\theta_{mV_{n^*},t}$ directly affect $X_{mV_{n^*},t}$; as they rise, so does $X_{mV_{n^*},t}$, and vice versa. Also, as the LEO-Sat moves, both $\theta_{mV_{n^*},t}$ and $\mathcal{T}_{mV_{n^*},t}$ fluctuate, consequently impacting $X_{mV_{n^*},t}$. As per [21], the relationship between $X_{mV_{n^*},t}$ and $\mathbf{y}_{mV_{n^*},t}$ can be written as follows:

$$\mathbb{E}\left[X_{mV_{n^*},t} \backslash \mathbf{y}_{mV_{n^*},t}\right] = \mathbf{y}_{mV_{n^*},t}^T \mathbf{q}_{mV_{n^*}}^*, \tag{20}$$

where $\mathbb{E}$ signifies expectation, and $\mathbf{q}_{mV_{n^*}}^*$ is an indeterminate coefficient. The CMAB's goal is to predict $\mathbf{q}_{mV_{n^*}}^*$ values via continuous online learning. This involves using linear regression, where $\widehat{\mathbf{q}}_{mV_{n^*}}$ is computed as [21]:

$$\widehat{\mathbf{q}}_{mV_{n^*}} = \mathbf{A}_{mV_{n^*}}^{-1} \mathbf{\Gamma}_{mV_{n^*}}, \tag{21}$$

$$\mathbf{A}_{mV_{n^*}} = \mathbf{D}_{mV_{n^*}}^T \mathbf{D}_{mV_{n^*}} + \mathbf{I}_b, \tag{22}$$

$$\mathbf{\Gamma}_{mV_{n^*}} = \Psi_{mV_{n^*},t} \mathbf{y}_{mV_{n^*}}, \tag{23}$$

$\mathbf{D}_{mV_{n^*}}$ is an $l \times b$ contextual vectors matrix, where each l row contains b contextual vector. $\mathbf{I}_b$ represents the $b \times b$ identity matrix. As proven in [21]:

$$\left| \mathbf{y}_{mV_{n^*},t}^T \widehat{\mathbf{q}}_{mV_{n^*}} - \mathbb{E}\left[X_{mV_{n^*},t} \backslash \mathbf{y}_{mV_{n^*},t}\right] \right| \leq \Omega \sqrt{\mathbf{y}_{mV_{n^*},t}^T \mathbf{A}_{mV_{n^*}}^{-1} \mathbf{y}_{mV_{n^*},t}}. \tag{24}$$

The design parameter Ω balances between the exploitation represented by $(\mathbf{y}_{mV_{n^*},t}^T \widehat{\mathbf{q}}_{mV_{n^*}})$ and the exploration denoted by $(\Omega \sqrt{\mathbf{y}_{mV_{n^*},t}^T \mathbf{A}_{mV_{n^*}}^{-1} \mathbf{y}_{mV_{n^*},t}})$. Merging these two terms yields the selected LEO-Sat index m_t^* at time t [21]:

$$m_t^* = \arg\max_{\mathcal{M}_{V_{n^*},t}} \left(\mathbf{y}_{mV_{n^*},t}^T \widehat{\mathbf{q}}_{mV_{n^*}} + \Omega \sqrt{\mathbf{y}_{mV_{n^*},t}^T \mathbf{A}_{mV_{n^*}}^{-1} \mathbf{y}_{mV_{n^*},t}} \right). \tag{25}$$

The details of the proposed LinUCB-VA algorithm implementing the second stage of the proposed MAB game,

namely CMAB-VA, are given in Algorithm 1. It takes Ω and $\mathcal{M}_{SatCon}$ as inputs, delivering the chosen LEO-Sat m_t^*. As the first step, the set of visible LEO-Sats $\mathcal{M}_{V_{n^*},t}$ satisfying the condition $\theta_{mV_{n^*},t} \geq \theta_0$ is enumerated by UAV at location n. This involves two phases: initially, the UAV verifies the LEO-Sat index; if new, the related parameters $\mathbf{A}_{mV_{n^*}}$ and $\mathbf{\Gamma}_{mV_{n^*}}$ are set as detailed in Algorithm 1. If $\theta_{mV_{n^*},t}$ for any existing LEO-Sats falls below θ_0, they are flagged as sleeping and excluded from $\mathcal{M}_{V_{n^*},t}$, showcasing LinUCB-VA's variable arm feature. Subsequently, $\widehat{\mathbf{q}}_{mV_{n^*}}$ values are calculated for all visible LEO-Sats as in (21). Upon determining $\widehat{\mathbf{q}}_{mV_{n^*}}$, m_t^* is calculated using (25). Then, its associated reward $X_{m^*V_n,t}$ is observed, and its related parameters are updated as per Algorithm 1.

C. REGERT ANALYSIS OF THE PROPOSED TWO-STAGE MAB APPROACH

To bound the performance of MAB algorithms, regret analysis provides an effective metric. It is mathematically represented as follows [22]:

$$\mathcal{R}(T_H) = \mathbb{E}\left[\sum_{t=1}^{T_H} \pi_{n^*,t} - \pi_{n,t}\right], \tag{26}$$

where T_H indicates the time horizon over which the MAB algorithm was conducted. $\pi_{n^*,t}$ is the reward of the optimal arm selected by an oracle algorithm, and $\pi_{n,t}$ is the arms' reward selected by the MAB algorithm. For the proposed two-stage MAB algorithm, T_H is not an absolute value, but it depends on the value of $E_{V\text{Min}}$ where the MAB algorithm should be terminated, i.e., $T_H(E_{V\text{Min}})$. For worst case scenario, its total regret bound can be written as follows:

$$\mathcal{R}_{Two-Stage\ MAB} = \max\left(\mathcal{R}_{\text{BC}-\text{UCB}}, \mathcal{R}_{LinUCB-VA}\right), \tag{27}$$

where $\mathcal{R}_{\text{BC}-\text{UCB}}$ is the regret bound of its first stage while $\mathcal{R}_{LinUCB-VA}$ is the regret bound of the second stage.

For the proposed BC-UCB stage, its regret bound is like that belongs to the UCB-BV2 algorithm given in [40]. Based on the analysis given there, this regret bound is expressed as follows:

$$\mathcal{R}_{BC-UCB} \leq \left(\frac{\eta^r_{n^*}}{\eta^c_{n^*}}(E_{V\text{Min}}+1) - \eta^c_{n^*}\mathbb{E}(T_H(E_{V\text{Min}}))\right) + \left(\eta^r_{n^*}\mathbb{E}(T_H(E_{V\text{Min}})) - \mathbb{E}\left[\sum_{t=1}^{T_H(E_{V\text{Min}})} \Psi_{V_n,t}\right]\right) \tag{28}$$

where $\eta^r_{n^*}$ is the mean reward, i.e., UAV data rate, corresponding to the optimal location, and $\eta^c_{n^*}$ is its corresponding UAV energy cost. After deducing the value of $\mathbb{E}(T_H(E_{V\text{Min}}))$ detailed in [40], the regret bound is found to be as follows:

$$\mathcal{R}_{BC-UCB} \leq \mathcal{O}\left(\sqrt{\ln(E_{V\text{Min}})}\right), \tag{29}$$

where $\mathcal{O}$ notation represents the order.

For LinUCB-VA, the regret bound of basic LinUCB algorithm is deduced in [41] as follows:

$$\mathcal{R}_{LinUCB} \leq \mathcal{O}\left(\sqrt{T_H b\ ln^3(\Delta T_H \ln(T_H)/\Omega)}\right), \tag{30}$$

where the total number of available arms is represented by Δ. For the proposed LinUCB-VA, the same regret bound is applied except that Δ is replaced by $|\mathcal{M}_{SatCon}|$, which is the cardinality of all available LEO-Sats in the LEO-SatCon as a worst-case scenario. Also, T_H should be replaced by $T_H(E_{V\text{Min}})$. Thus, $\mathcal{R}_{LinUCB-VA}$ can be written as follows:

$$\mathcal{R}_{LinUCB-VA} \leq \mathcal{O}\left(\sqrt{T_H(E_{V\text{Min}}) b\ ln^3\left(\frac{|\mathcal{M}_{SatCon}| T_H(E_{V\text{Min}}) \ln(T_H(E_{V\text{Min}}))}{\Omega}\right)}\right). \tag{31}$$

As both $\mathcal{R}_{BC-UCB}$ and $\mathcal{R}_{LinUCB-VA}$ are of $\mathcal{O}(\sqrt{\ln(E_{V\text{Min}})})$, we can conclude that the regret bound of the proposed two-stage MAB approach is of $\mathcal{O}(\sqrt{\ln(E_{V\text{Min}})})$.

V. NUMERICAL ANALYSIS

To prove the effectiveness of the proposed two-stage MAB approach, extensive numerical simulations were performed. The locations of GDs are randomly distributed with arbitrary γ_{V_n} values in the range $[0, \gamma_0]$. The number of GDs per location is randomly allocated in the range [1], [25]. The LEO-SatCon contains several numbers of LEO-Sats at altitude of 1000 Km with orbital inclination angle of 60° and θ_0 of 10°. The LEO-Sat speed is set to 8000 m/sec while UAV speed is set to 20 m/sec at an altitude of 100 m. Other important simulation parameters are listed in Table 1.

A. PERFORMANCE OF UTP

In this section of numerical analysis, we evaluate the effectiveness of the proposed BC-MAB in optimizing the UAV trajectory against the distributed locations while setting the number of LEO-Sats equals 4. As a benchmark, we provide the upper bound of the UAV data rate performance, which corresponds to scenario where the UAV consistently chooses the location with the highest possible rate. We also simulate the performance of the naive UCB without BC.

Lastly, we examine the approach where the UAV chooses its next location on its trajectory at random, i.e., "Rand" scheme.

Figure 4 shows the average UAV data rate in [Mpbs] against the number of distributed locations ranging from 4 to 49 locations. As the number of locations increases, the average UAV data rate of the upper bound, BC-UCB and UCB schemes are slightly increased due to the high chance of finding locations with higher data rates. On the other hand, Rand scheme shows constant UAV data rate despite the tested number of locations due to the randomly selected location. As shown by this figure, UCB has almost the same data rate performance as upper bound performance because it aims to select the arm, i.e., location, maximizing

TABLE 1. Simulation parameters.

Parameter	Value
γ_{V_n} and γ_0	Randomly generated in the range of $[0, \gamma_0]$ and 0.3780 [18]
K_n	Uniformly random in the range [1, 25]
P_{tG} and P_{tV}	1 and 10 Watt
R_e, h_s and h_V	6371 Km, 1000 Km and 100m [18]
θ_0 and i	10° and 60° [18]
ω_E and ω_S	7.3e-5 and 9.977e-04 [36]
Z	10 dB [36]
G	15 dB [36]
σ_0	-174 + 10log10(W) + 7 [4]
W_V	60 MHz
$W_{mV_n,t}$	Randomly generated in the range of $[0, W_{Lmax}]$, W_{Lmax} = 20 MHz
P_f, P_h, and P_c	4, 2, and 1 Watt [18]
S_V and $\lvert S_{mV}\rvert$	20 m/sec and 780 m/sec [35]
$f_{D\max}$	6.1452 KHz [35]
Ω	0.2
f_G and f_V	2.4 GHz and 2 GHz
α^{LoS} and α^{NLoS}	0.1 and 21 dB [34]
a and b	4.88 and 0.429 [34]
$E_{V\text{Min}}$ and $C_{V\text{Max}}$	100 and 10,000
Td_n	Randomly generated in the range of [0,10 Gbps]

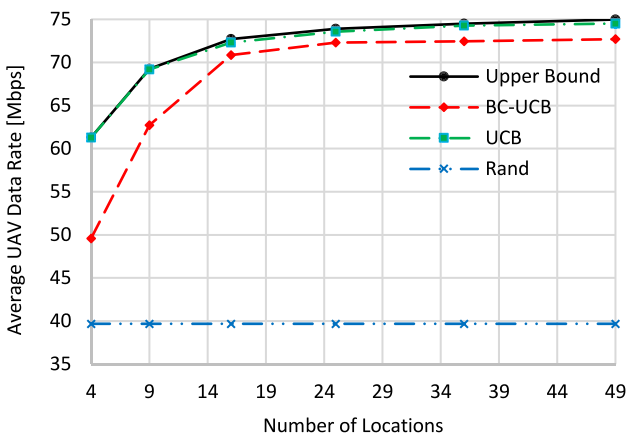

FIGURE 4. Average UAV data rate against the number of locations.

the achievable UAV data rate through the successive online training. However, as the proposed BC-UCB compromises between UAV achievable data rate and its energy cost when selecting the next location, it has lower average data rate performance than UCB and upper bound. As Rand scheme selects the locations at random, it has the lowest data rate performance, and it is constant despite the tested number

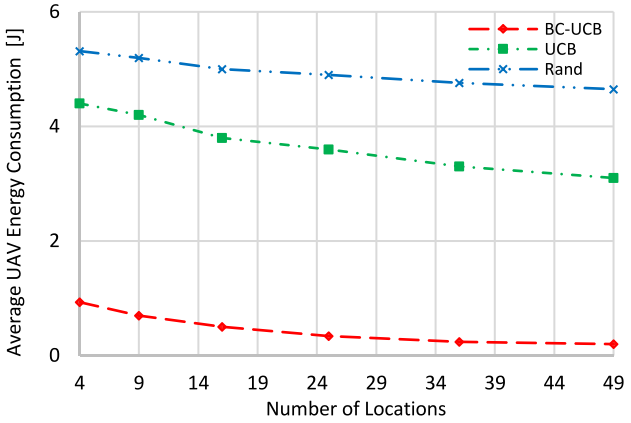

FIGURE 5. Average UAV energy consumption against number of locations.

of locations. From Fig. 4 and at $n = 4$, UAV average data rate of upper bound, UCB, BC-UCB and Rand selection are 61.3 Mbps, 61.2 Mbps, 49.59 Mbps, and 39.69 Mbps, respectively. These values become 75 Mbps, 74.51 Mbps, 72.7 Mbps, and 39.69 Mbps, when $n = 49$, respectively.

Figure 5 shows the UAV energy consumption in joule against the number of locations. As the number of locations is increasing, the energy efficiency performances of all schemes are decreasing. This is because more locations are distributed with nearer inter distances which reduces the UAV flying energy consumptions. Moreover, for both BC-UCB and UCB algorithms, as the number of locations increases, the UAV data rate increases as given in Fig. 4, which also contributes to lowering the hovering time and the hovering energy consumption as given in (6). The proposed BC-UCB has the lowest energy consumption performance while Rand selection has the highest one. This comes from the budget constraint functionality of the proposed algorithm which enables it to highly reducing UAV's energy consumption over the other schemes involved in the comparison. From Fig. 5 and at $n = 4$, the UAV energy consumption of BC-UCB, UCB and Rand schemes are 0.093 J, 4.4 J, and 5.31 J, respectively. These values become 0.2 J, 3.1 J, and 4.65 J, when $n = 49$, respectively.

Figure 6 illustrates the average energy efficiency of UAVs in Mbps/J, which is evaluated by dividing the UAV data rate by its energy consumption. As the number of locations increases, the UAV energy efficiency of all the compared schemes shows an upward trend, primarily driven by the boost or constant in case of using Rand selection in UAV data rate and a reduction in energy consumption. Among the schemes compared, the proposed BC-UCB stands out with the highest energy efficiency performance, while the Rand scheme lags behind with the lowest energy efficiency. At $n = 4$ locations, the use of the BC-UCB scheme results in approximately 3.82 times and 7.13 times higher UAV energy efficiency compared to the use of the UCB and Rand schemes, respectively. These values grow to 15.1 and 42.6 times, respectively, as the number of locations expands.

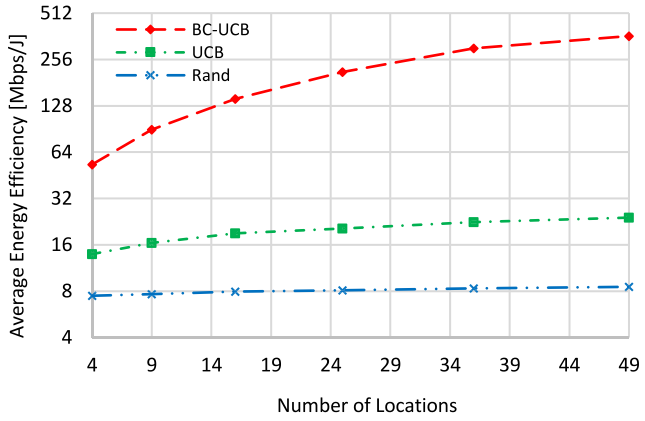

FIGURE 6. Average UAV energy efficiency against the number of locations.

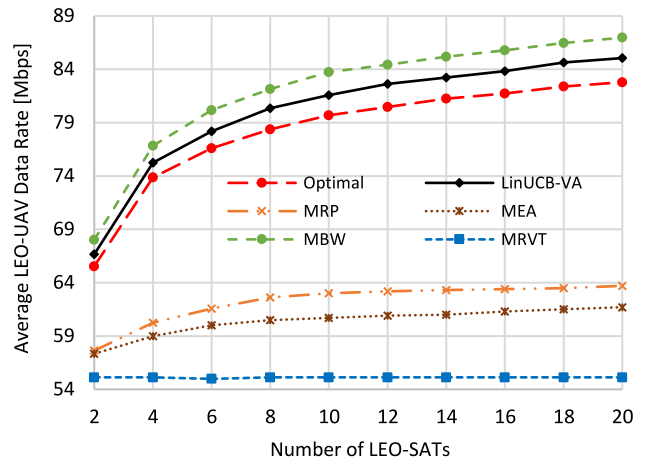

FIGURE 7. Average LEO-Sat data rate against the number of LEO-Sats.

B. PERFORMANCE OF LEO-SAT SELECTION

In this part of numerical analysis, we study the performance of LEO-Sat selection against the number of LEO-Sats using 25 locations. As there is no approach proposed in literature for UAV-centric LEO-Sat handover/selection to the best of our knowledge, the performance of the proposed LinUCB-VA is compared with the state-of-the-art LEO-Sat handover/selection schemes that maximize a specific LEO-Sat feature like MEA, MRP, MBW and MRVT approaches stated in [24], [25]. Optimal performance is also given, where it exhaustively searches all available LEO-Sats $\mathcal{M}_{V_{n^*},t}$ and selects the one maximizes (15). Indeed, it gives the best performance of (15) but at the expense of a bunch of overhead. This overhead comes from the need to communicate with all LEO-Sats in $\mathcal{M}_{V_{n^*},t}$ individually before selecting the best one at every time t while considering $\mathcal{M}_{V_{n^*},t}$ is time updatable.

Figure 7 shows the average data rate of the LEO-UAV link against the available number of LEO-Sats ranging from 2 to 20. The average LEO-UAV data rate of all compared schemes, except MRVT, increases when increasing the number of LEO-Sats because of the increasing chance of finding out a LEO-Sat with higher LEO-UAV link speed.

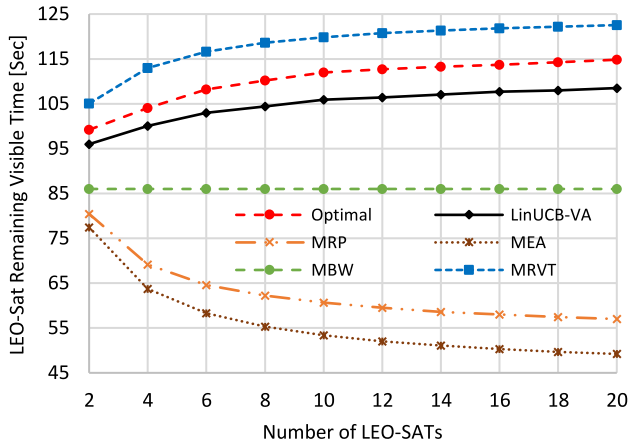

FIGURE 8. Remaining LEO-Sat visible time against the number of LEO-Sats.

From this figure, MBW has the best performance as LEO-Sat bandwidth has the dominant effect in deciding the data rate. In contrast, MRVT delivers the lowest performance, and this performance remains almost constant. This is attributed to the fact that a LEO-Sat with the maximum remaining visible time is typically the one that has just entered UAV visibility. This LEO-Sat often has the smallest elevation angle, resulting in the lowest achievable data rate, as depicted in Fig. 7. Moreover, as the remining visible time has no effect on the LEO-UAV data rate, it remains constant despite the number of used LEO-Sats. The optimal performance does not guarantee the highest average data rate performance due to its dual objective of maximizing the product of data rate and remaining visible time. Meanwhile, the proposed LinUCB-VA scheme demonstrates comparable average data rate performance to that achieved by MBW and it is even better than the optimal scheme. This comes from considering all LEO-Sat features when selecting the most appropriate one. Furthermore, it is interesting to highlight that MRP demonstrates superior average data rate performance compared to MEA. This comes from considering various crucial parameters related to LEO-UAV communication links, including antenna gain, operating frequency, path loss, and other factors in addition to the elevation angle as outlined in (7). For the case of two LEO-Sats, the average data rates for the optimal, proposed LinUCB-VA, MRP, MEA, MBW, and MRVT schemes are 65.54 Mbps, 66.7 Mbps, 57.6 Mbps, 57.3 Mbps, 68 Mbps, and 55.12 Mbps, respectively. These values increase to 82.8 Mbps, 85 Mbps, 63.7 Mbps, 61.7 Mbps, 87 Mbps, and 55.13 Mbps when using 20 LEO-Sats.

Figure 8 demonstrates the remaining visible time in sec of the selected LEO-Sat against the number of used LEO-Sats. As expected MRVT demonstrates the highest performance as it always selects the LEO-Sat with the highest remaining visible time. Also, its performance is increasing when increasing the number of LEO-Sats due to the high chance of finding out a LEO-Sat with a higher remining visible time. As the policies of both optimal and

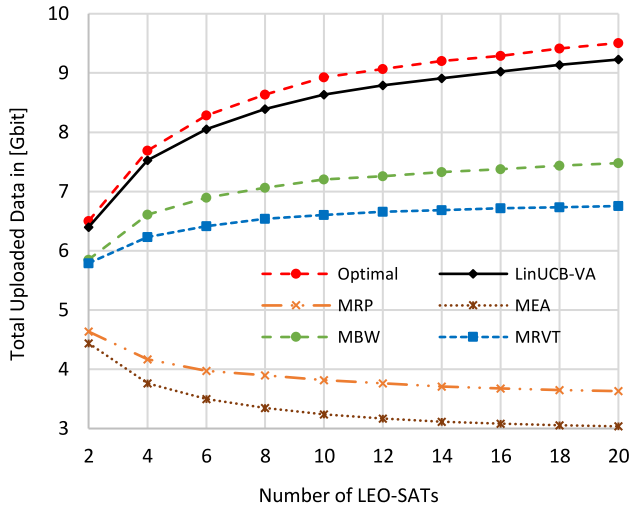

FIGURE 9. Total uploaded data against the number of LEO-Sats.

the proposed LinUCB-VA schemes are based on maximizing the remaining visible time when selecting a LEO-Sat, they demonstrate a comparable performance to MRVT. As MBW is mainly based on maximizing the LEO-Sat available bandwidth and has no policy regarding its remaining visible time, it has a constant performance. It is important to notice that the performances of both MRP and MEA are declining when raising the number of LEO-Sats. This comes from the fact that LEO-Sats with higher elevation angles have lower remaining visible time and vice versa. Thus, as we increase the number of LEO-Sats, a LEO-Sat with a higher elevation angle will be selected results in a lower remaining visible time. As MEA is only based on the elevation angle in its LEO-Sat selection, it shows the lowest remaining visible time performance. It is even lower than MRP, which is based on many other factors besides the elevation angle. From Fig. 8 and for two LEO-Sats, the LEO-Sat remaining visible time for the optimal, proposed LinUCB-VA, MRP, MEA, MBW, and MRVT schemes are 99.2 sec, 96 sec, 80.45 sec, 77.45 sec, 86 sec, and 105.04 sec, respectively. These values turn to be 114.82 sec, 108.5 sec, 57 sec, 49.21 sec, 86 sec, and 122.54 sec when using 20 LEO-Sats.

Figure 9 gives the performance of total uploaded data in Gbit, which is equal to the multiplication of average data rate and the remaining visible time of the selected LEO-Sat, as the main objective of (15). The optimal scheme demonstrates the best performance due to its brute search policy, but at the expense of intensive overhead as previously explained. The proposed LinUCB-VA shows a comparable performance of the optimal performance by only observing the reward of the previously selected LEO-Sats in addition to the current features of the candidate LEO-Sats as given in Algorithm 1 without any need for exhaustively searching all available LEO-Sats. Moreover, it is noticed that MBW and MRVT have better performances than MEA and MRP, because MBW and MRVT have predominant effects on maximizing the achievable data rate and the remaining visible time, respectively. As the data rate has a higher effect than the remining visible time on maximizing (15), MBW has better performance than MRVT. Finally, as MRP includes the effect of elevation angle in addition to other crucial factors affecting the LEO-UAV link, it has better performance than MEA which maximizes the elevation angle alone. Also, both are decreasing, affected by their remaining visible time performances given in Fig. 8. From Fig. 9 and for the case of two LEO-Sats, the total uploaded data for the optimal, proposed LinUCB-VA, MRP, MEA, MBW, and MRVT schemes are 6.5 Gbit, 6.4 Gbit, 4.64 Gbit, 4.44 Gbit, 5.85 Gbit, and 5.79 Gbit, respectively. These values turn to be 9.51 Gbit, 9.23 Gbit, 3.63 Gbit, 3.04 Gbit, 7.48 Gbit, and 6.76 Gbit when using 20 LEO-Sats.

C. COMPLEXITY ANALYSIS

The complexity of the proposed two-stage MAB approach comes from the complexity of the proposed BC-UCB and the LinUCB-VA algorithms. For BC-UCB, its computational complexity comes from selecting location maximizes its policy and updating its associated parameters with computational complexity of $\mathcal{O}(N+1)$, which is equal to that belonging to the naive UCB as stated in [3]. For the proposed LinUCB-VA, its computational complexity is like the computational complexity of LinUCB algorithm presented in [21], [41], which is of order $\mathcal{O}(b^2|\mathcal{M}_{V_{n^*},t}|)$, where b is a small number. For communication overhead, the proposed LinUCB-VA algorithm only needs to communicate with the selected LEO-Sat at each time t with complexity $\mathcal{O}(1)$. Compared to the other benchmark schemes, i.e., MRP, MBW, MEA, and MRVT and the optimal policy, their computational complexity is of order $\mathcal{O}(|\mathcal{M}_{V_{n^*},t}|)$ coming from selecting the LEO-Sat maximizes their own criterion. However, they suffer from high communication overhead as they need to communicate with all visible LEO-Sats $\mathcal{M}_{V_{n^*},t}$ before selecting the best one at each time t based on their policies. In this context, the communication overhead is the most dominant as it is related to communication protocol delays. Thus, we can conclude that the proposed LinUCB-VA algorithm reduces the communication overhead by almost $|\mathcal{M}_{V_{n^*},t}|$ while obtaining a comparable performance to the optimal policy and highly advanced over other benchmarks.

For the point of implementation cost in actual deployment, both algorithms BC-UCB and LinUCB-VA are implemented in the UAV as it is the player of both MAB models. However, as given in the above-mentioned complexity analysis, both algorithms have linear computational complexities with N and $|\mathcal{M}_{V_{n^*},t}|$. Thus, in actual deployment, both algorithms will be implemented using a low-cost micro-controller located in the UAV. Based on the proposed algorithms, this micro-controller decides the UAV next location and the selected LEO-Sat at each time step. This autonomous implementation will greatly reduce the communication overhead and latency, which is more appropriate for post-disaster rescue applications, where time is an important factor.

D. LIMITATIONS OF THE PROPOSED APPROACH

In the scenes of integrating LEO-Sats with UAVs in the proposed SAGIN, the proposed approach has the following limitations. 1) LEO-Sats handovers were not considered in the current paper, where we only assumed LEO-Sat selection based on maximum LEO-UAV achievable rate and long LEO-Sat visible time to reduce frequent LEO-UAV handovers. However, the impact of LEO-UAV handovers on the overall system delay performance should be investigated. 2) It is assumed that the LEO-Sats' features vector $\mathbf{y}_{mV_{n^*},t}$ should be carefully estimated at each location n and time step t. The error in this estimated vector will affect the performance of the proposed scheme. 3) The integration of LEO-Sats into the overall SAGIN framework may pose technical limitations. That is different LEO-Sat management and communication protocols, frequencies, or standards may need to be considered, and interoperability issues could arise. All these current limitations and others will be the subject of our future investigations.

VI. CONCLUSION

In this paper, we have investigated the problem of joint UTP and LEO-Sat selection in SAGIN. The optimization problem of this problem was formulated under its constraints of limited UAV battery budget and LEO-Sat features. To address this highly dynamic problem, an online learning approach in the form of a two-stage MAB model was proposed. In the first MAB stage, the energy aware UTP was modeled as a BC-MAB game, and the BC-UCB algorithm was proposed to implement it. In the second MAB stage, LEO-Sat selection was modeled as CMAB-VA game, and LinUCB-VA was proposed to realize it. By the means of numerical simulations. We proved the effectiveness of the proposed BC-UCB and LinUCB-VA over other benchmarks. In this context, the proposed BC-UCB showed better energy efficiency performance than naive UCB with the same computational complexity. Also, the proposed LinUCB-VA had a comparable performance to the optimal policy with much lower computational complexity and outperformed all other LEO-Sat selection schemes.

APPENDIX

In this derivation, we adopt the methodology presented in [36], but we focus on the LEO-UAV scenario rather than the LEO-GD scenario described in [36]. As the following derivation is general for any LEO-Sat m and UAV location n, we omit the subscript mV_n in the derivation for notation simplicity. Based on Fig. 2, we can formulate the subsequent equation:

$$\cos(\theta_t + \gamma_t) = \left(\frac{R_e + h_V}{R_e + h_S}\right)\cos(\theta_t). \tag{32}$$

The expansion of the left-hand side will lead to the following equation:

$$\tan(\theta_t) = \frac{\cos(\gamma_t) - \left(\frac{R_e+h_V}{R_e+h_S}\right)}{\sin(\gamma_t)}. \tag{33}$$

To determine γ_t, let us consider the trace angle and the coverage area of the LEO-Sat as illustrated in Fig. 3 and by referencing the cosine law, it can be inferred that [36]:

$$\cos(\gamma_t) = \cos(\psi_t)\cos\left(\gamma_{V_n}\right). \tag{34}$$

Then,

$$\gamma_t = \cos^{-1}\left(\cos(\psi_t)\cos\left(\gamma_{V_n}\right)\right). \tag{35}$$

In this regard, $\frac{d}{dt}|\psi_t|$ represents the angular velocity of the LEO-Sat in the Earth centered fixed frame, i.e., ω. This can be formulated as:

$$|\psi_t| = \int \omega dt, \tag{36}$$

where

$$\omega \approx \omega_S - \omega_E \cos(i). \tag{37}$$

Consequently, ψ_t can be expressed as follows [36]:

$$\psi_t = -(\omega_S - \omega_E \cos(i))t + \psi_0. \tag{38}$$

By substituting (38) in (35),

$$\gamma_t = \cos^{-1}\left(\cos((\omega_E \cos(i) - \omega_S)t + \psi_0)\cos\left(\gamma_{V_n}\right)\right). \tag{39}$$

By inserting (39) in (33) allows us to infer (9).

Additionally, from (34), and by setting $t = 0$, ψ_0 can be written as follows:

$$\psi_0 = \cos^{-1}\left(\frac{\cos(\gamma_0)}{\cos\left(\gamma_{V_n}\right)}\right), \tag{40}$$

where γ_0 is the angle related to the onset of LEO-Sat visibility. Recall (32), and setting $t = 0$, then $\theta_t = \theta_0$ and γ_0 can be expressed as follows:

$$\gamma_0 = \cos^{-1}\left(\left(\frac{R_e + h_V}{R_e + h_S}\right)\cos(\theta_0)\right) - \theta_0. \tag{41}$$

From (41), it is noted that γ_0 is fixed for a certain value of θ_0 as it is only based on geometrical values R_e, h_V and h_S.

REFERENCES

[1] S. Hashima et al., "On softwarization of intelligence in 6G networks for ultra-fast optimal policy selection: Challenges and opportunities," *IEEE Netw.*, vol. 37, no. 2, pp. 190–197, Mar./Apr. 2023, doi: 10.1109/MNET.103.2100587.

[2] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021, doi: 10.1109/OJCOMS.2021.3057679.

[3] E. M. Mohamed, S. Hashima, and K. Hatano, "Energy aware multiarmed bandit for millimeter wave-based UAV mounted RIS networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 6, pp. 1293–1297, Jun. 2022, doi: 10.1109/LWC.2022.3164939.

[4] E. M. Mohamed, K. Sakaguchi, and S. Sampei, "Wi-Fi coordinated WiGig concurrent transmissions in random access scenarios," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10357–10371, Nov. 2017, doi: 10.1109/TVT.2017.2738198.

[5] B. M. ElHalawany, S. Hashima, K. Hatano, K. Wu, and E. M. Mohamed, "Leveraging machine learning for millimeter wave beamforming in beyond 5G networks," *IEEE Syst. J.*, vol. 16, no. 2, pp. 1739–1750, Jun. 2022, doi: 10.1109/JSYST.2021.3089536.

[6] S.-C. Hung, H. Hsu, S.-Y. Lien, and K.-C. Chen, "Architecture harmonization between cloud radio access networks and fog networks," *IEEE Access*, vol. 3, pp. 3019–3034, 2015, doi: 10.1109/ACCESS.2015.2509638.

[7] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive MIMO networks: A survey of opportunities, challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 611–652, 1st Quart., 2022, doi: 10.1109/COMST.2021.3135119.

[8] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 4th Quart., 2018, doi: 10.1109/COMST.2018.2841996.

[9] I. Leyva-Mayorga et al., "LEO small-satellite constellations for 5G and beyond-5G communications," *IEEE Access*, vol. 8, pp. 184955–184964, 2020.

[10] I. d. Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica*, vol. 159, pp. 123–135, Jun. 2019, doi: 10.1016/j.actaastro.2019.03.040.

[11] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.

[12] 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, p. 3947, Jul. 2020, doi: 10.3390/s20143947.

[13] A. Amrallah, E. M. Mohamed, G. K. Tran, and K. Sakaguchi, "UAV trajectory optimization in a post-disaster area using dual energy-aware bandits," *Sensors*, vol. 23, p. 1402, Jan. 2023, doi: 10.3390/s23031402.

[14] Z. Hu et al., "Joint resources allocation and 3D trajectory optimization for UAV-enabled space-air-ground integrated networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 11, pp. 14214–14229, Nov. 2023, doi: 10.1109/TVT.2023.3280482.

[15] B. Cao et al., "Edge–cloud resource scheduling in space–air–ground-integrated networks for Internet of Vehicles," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 5765–5772, Apr. 2022, doi: 10.1109/JIOT.2021.3065583.

[16] N. Wang, F. Li, D. Chen, L. Liu, and Z. Bao, "NOMA-based energy-efficiency optimization for UAV enabled space-air-ground integrated relay networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4129–4141, Apr. 2022, doi: 10.1109/TVT.2022.3151369.

[17] P. K. Singya and M.-S. Alouini, "Performance of UAV-assisted multiuser terrestrial-satellite communication system over mixed FSO/RF channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 2, pp. 781–796, Apr. 2022, doi: 10.1109/TAES.2021.3111787.

[18] Z. Jia, M. Sheng, J. Li, D. Niyato, and Z. Han, "LEO-satellite-assisted UAV: Joint trajectory and data collection for Internet of Remote Things in 6G aerial access networks," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9814–9826, Jun. 2021, doi: 10.1109/JIOT.2020.3021255.

[19] D.-H. Tran, S. Chatzinotas, and B. Ottersten, "Satellite- and cache-assisted UAV: A joint cache placement, resource allocation, and trajectory optimization for 6G aerial networks," *IEEE Open J. Veh. Technol.*, vol. 3, pp. 40–54, Jan. 2022, doi: 10.1109/OJVT.2022.3142170.

[20] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.* vol. 47, pp. 235–256, May 2002, doi: 10.1023/A:1013689704352.

[21] S. Hashima, K. Hatano, E. Takimoto, and E. M. Mohamed, "Neighbor discovery and selection in millimeter wave D2D networks using stochastic MAB," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1840–1844, Aug. 2020, doi: 10.1109/LCOMM.2020.2991535.

[22] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 661–670.

[23] E. M. Mohamed, S. Hashima, K. Hatano, M. M. Fouda, and Z. M. Fadlullah, "Sleeping contextual/non-contextual Thompson sampling MAB for mmWave D2D two-hop relay probing," *IEEE Trans. Veh. Technol.*, vol. 70, no. 11, pp. 12101–12112, Nov. 2021, doi: 10.1109/TVT.2021.3116223.

[24] Z. Wu, F. Jin, J. Luo, Y. Fu, J. Shan, and G. Hu, "A graph-based satellite handover framework for LEO satellite communication networks," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1547–1550, Aug. 2016, doi: 10.1109/LCOMM.2016.2569099.

[25] P. K. Chowdhury, M. Atiquzzaman, and W. Ivancic, "Handover schemes in satellite networks: State-of-the-art and future research directions," *IEEE Commun. Surveys Tuts.*, vol. 8, no. 4, pp. 2–14, 4th Quart., 2006, doi: 10.1109/COMST.2006.283818.

[26] P. Qin, M. Wang, X. Zhao, and S. Geng, "Content service oriented resource allocation for space–air–ground integrated 6G networks: A three-sided cyclic matching approach," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 828–839, Jan. 2023, doi: 10.1109/JIOT.2022.3203793.

[27] R. Samy, H.-C. Yang, T. Rakia, and M.-S. Alouini, "Space-air-ground FSO networks for high-throughput satellite communications," *IEEE Commun. Mag.*, vol. 61, no. 3, pp. 82–87, Mar. 2023, doi: 10.1109/MCOM.002.2200018E.

[28] J. Ye, S. Dang, B. Shihada, and M.-S. Alouini, "Space-air-ground integrated networks: Outage performance analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7897–7912, Dec. 2020, doi: 10.1109/TWC.2020.3017170.

[29] D. Liu, J. Zhang, J. Cui, S.-X. Ng, R. G. Maunder, and L. Hanzo, "Deep learning aided routing for space-air-ground integrated networks relying on real satellite, flight, and shipping data," *IEEE Wireless Commun.*, vol. 29, no. 2, pp. 177–184, Apr. 2022, doi: 10.1109/MWC.003.2100393.

[30] E. M. Mohamed, "LEO satellite assisted UAV distribution using combinatorial bandit with fairness and budget constraints," *PLoS One*, vol. 18, Aug. 2023, Art. no. e0290432, doi: 10.1371/journal.pone.0290432.

[31] J. Wang, C. Jiang, Z. Wei, C. Pan, H. Zhang, and Y. Ren, "Joint UAV hovering altitude and power control for space-air-ground IoT networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1741–1753, Apr. 2019, doi: 10.1109/JIOT.2018.2875493.

[32] Z. Li et al., "Energy efficient resource allocation for UAV-assisted space-air-ground Internet of Remote Things networks," *IEEE Access*, vol. 7, pp. 145348–145362, 2019, doi: 10.1109/ACCESS.2019.2945478.

[33] J. Sabzehali, V. K. Shah, Q. Fan, B. Choudhury, L. Liu, and J. H. Reed, "Optimizing number, placement, and backhaul connectivity of multi-UAV networks," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21548–21560, Nov. 2022, doi: 10.1109/JIOT.2022.3184323.

[34] H. V. Abeywickrama, B. A. Jayawickrama, Y. He, and E. Dutkiewicz, "Comprehensive energy consumption model for unmanned aerial vehicles, based on empirical studies of battery performance," *IEEE Access*, vol. 6, pp. 58383–58394, 2018, doi: 10.1109/ACCESS.2018.2875040.

[35] D. D. Mrema and S. Shimamoto, "Performance of quadrifilar helix antenna on EAD channel model for UAV to LEO satellite link," in *Proc. Int. Conf. Collab. Technol. Syst. (CTS)*, 2012, pp. 170–175, doi: 10.1109/CTS.2012.6261045.

[36] Y. Seyedi and F. Rahimi, "A trace-time framework for prediction of elevation angle over land mobile LEO satellites networks," *Wireless Pers. Commun.* vol. 62, pp. 793–804, Feb. 2012, doi: 10.1007/s11277-010-0094-5.

[37] J-Y. Audibert, R. Munos, and C. Szepesvári, "Exploration–exploitation tradeoff using variance estimates in multi-armed bandits," *Theor. Comput. Sci.*, vol. 410, no. 19, pp. 1876–1902, 2009.

[38] E. M. Mohamed, M. Alnakhli, S. Hashima, and M. Abdel-Nasser, "Distribution of multi mmWave UAV mounted RIS using budget constraint multi-player MAB," *Electronics* vol. 12, no. 1, p. 12, 2023, doi: 10.3390/electronics12010012.

[39] W. Ding, T. Qin, X.-D. Zhang, and T.-Y. Liu, "Multi-armed bandit with budget constraint and variable costs," in *Proc. 27th AAAI Conf. Artif. Intell.*, 2013, pp. 232–238.

[40] W. Chu, L. Li, L. Reyzin, and R. Schapire, "Contextual bandits with linear payoff functions," in *Proc. Mach. Learn. Res.*, 2011, pp. 208–214. [Online]. Available: https://proceedings.mlr.press/v15/chu11a.html

[41] E. M. Mohamed, S. Hashima, K. Hatano, S. A. Aldossari, M. Zareei, and M. Rihan, "Two-hop relay probing in WiGig device-to-device networks using sleeping contextual bandits," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1581–1585, Jul. 2021.

EHAB MAHMOUD MOHAMED (Member, IEEE) received the Ph.D. degree in information science and electrical engineering from Kyushu University, Japan, in 2012. He is currently a Full Professor with the Department of Electrical Engineering, College of Engineering in Wadi Addwasir, Prince Sattam bin Abdulaziz University, Saudi Arabia. He is also a Full Professor with the Department of Electrical Engineering, Aswan University Egypt. He has been a Specially Appointed Researcher with Osaka University Japan from 2013 to 2016. He has 170+ journals and conference papers. His current research interests include 5G/ 6G networks, cognitive radio networks, millimeter wave transmissions, Li-Fi technology, MIMO systems, and reconfigurable intelligent surfaces. He was the General Chair of the IEEE ITEMS'16 and IEEE ISWC'18 and a guest editor in many highly ranked transactions/Journals. He is a technical committee member of many international conferences and a reviewer of many international conferences, Journals, and transactions, especially highly ranked IEEE transactions/Journals.

MOHAMMAD AHMED ALNAKHLI (Member, IEEE) received the B.Sc. degree in electrical engineering from Umm Al-Qura University, Makkah, Saudi Arabia, in 2008, the first M.Sc. degree in electronic communication and computer engineering from the Electrical and Computer Engineering Department, University of Nottingham, Nottingham, U.K., in 2011, and the second M.Sc. degree in network and security and the Ph.D. degree in electrical and computer engineering from the Stevens Institute of Technology, Hoboken, NJ, USA, in 2017 and 2019, respectively. He is an Assistant Professor with the Department of Electrical Engineering, Prince Sattam bin Abdulaziz University, Saudi Arabia. His current research interests include spectrum management and energy consumption in next-generation wireless networks, radio resource allocation in UAVs and LEO satellites, Internet of Things, energy and spectrum efficient wireless networks, and 5G, 5GB and 6G systems and cognitive radio networks.

MOSTAFA M. FOUDA (Senior Member, IEEE) received the B.S. (as the valedictorian) and M.S. degrees in Electrical Engineering from Benha University, Egypt, in 2002 and 2007, respectively, and the Ph.D. degree in information sciences from Tohoku University, Japan, in 2011. He is currently an Associate Professor with the Department of Electrical and Computer Engineering with Idaho State University, ID, USA. He also holds the position of Full Professor with Benha University. He was an Assistant Professor with Tohoku University and a Postdoctoral Research Associate with Tennessee Technological University, TN, USA. He has coauthored more than 230 technical publications. He has received several research grants, including NSF Japan-U.S. Network Opportunity 3. His current research focuses on cybersecurity, communication networks, signal processing, wireless mobile communications, smart healthcare, smart grids, AI, and IoT. He has guest-edited a number of special issues covering various emerging topics in communications, networking, and health analytics. He currently serves on the editorial board of the IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE ACCESS.