Faster Exploration and Exploitation for Communication Environment Awareness in Starlink

Quanwei Zhang*, Zhiming Huang†, Jinwei Zhao*, Ali Ahangarpour*, Jianping Pan*
*Department of Computer Science, University of Victoria, BC, Canada
†Paul G. Allen School of Computer Science & Engineering, University of Washington, WA, USA
Email: quanweizhang@uvic.ca, zhimingh@cs.washington.edu, {clarkzjw, aliahangarpour, pan}@uvic.ca

***Abstract*—Environmental awareness is critical for maintaining robust connectivity in *low Earth orbit (LEO)* satellite networks. Starlink terminals construct a 123×123 obstruction map by passively recording signal quality from connected satellites. However, this process is slow: even under continuous operation, meaningful map coverage typically requires 6 to 12 hours, impractically slow for dynamic environments or short-term deployments. Building on Starlink's existing capability, we investigate how intelligent satellite selection can further accelerate obstruction map construction, without requiring architectural changes. We formulate the problem as a multi-armed bandit with sleeping arms and introduce SkyBandit, a lightweight online learning algorithm that balances link quality with exploration of uncharted sky regions. SkyBandit achieves a regret bound of $O(\sqrt{NT \ln T})$, where N is the number of observed orbits over T rounds. Experiments using a refined version of xeoverse LEO emulator along with real Starlink measurement data demonstrate that SkyBandit reduces obstruction map convergence time to just around 3 hours, while preserving high-quality connections. To our knowledge, this is the first work to formalize and optimize obstruction map construction in LEO networks via principled sequential learning.**

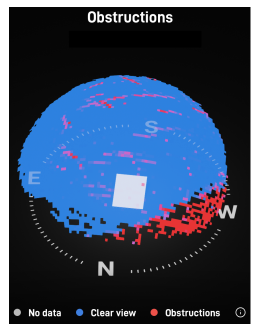


Fig. 1. Obstruction map obtained from the Starlink mobile app. The red pixels indicate obstructions.

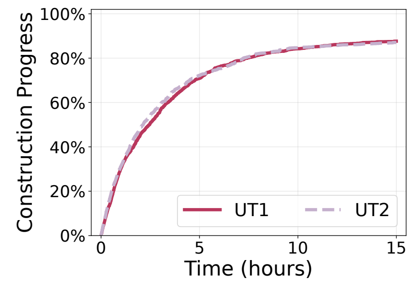


Fig. 2. Construction progress of the obstruction map over time for two directly accessible Starlink user terminals (UT1: July 15, 2025 20:44–July 16, 17:45 UTC; UT2: July 15, 2025 23:00–July 17, 01:01 UTC). Both UTs will be used for evaluation; the construction process uses the converged map as the benchmark.

I. Introduction

Fundamentally, effective decision-making hinges on the ability to perceive obstacles in the environment and respond to them appropriately. In this light, *awareness of obstructions* emerges not merely as a technical detail but as a fundamental capability and design principle: mapping environmental constraints to enable truly informed and adaptive behavior.

This principle extends naturally to satellite-to-ground communication. In Starlink's *low Earth orbit (LEO)* network, each *user terminal (UT)* maintains an *obstruction map* (see Fig. 1), which is a polar projection of the sky discretized into a 123×123 pixel grid. When the terminal connects to a satellite, it evaluates link quality using the *received signal strength indicator (RSSI)*; if the RSSI falls below a predefined threshold, the corresponding pixel is marked as obstructed. Over time, this process accumulates information about the occluded regions in the terminal's *field of view (FoV)*.

While Starlink's obstruction map was originally intended to guide installation by helping users avoid obstructions such as trees, buildings, and other obstacles, recent research has revealed its broader value for system observability and reverse engineering. Because the obstruction map encodes historical satellite connectivity, it implicitly captures trajectory traces of previously assigned satellites [1]–[3]. If satellite identifiers were accessible rather than currently obfuscated, these traces could be matched much more easily with publicly available ephemeris data such as *two-line element (TLE)* sets to reconstruct precise satellite paths. This would enable a wide range of advanced capabilities, including localization, predictive handover modeling, anomaly detection, and environment-aware replay for protocol emulation (see Sec. II).

Yet currently, the obstruction map's slow convergence remains a central bottleneck, delaying the terminal's ability to perceive and adapt to its surrounding obstructions. As shown in Fig. 2, our measurements from July 2025 reveal that even under continuous operation, it typically takes 6 to 12 hours to reach just 80% coverage of the terminal's FoV. According to Starlink's official documentation, full convergence may take up to a week [4]. This sluggish construction severely limits the system's responsiveness in dynamic or short-lived deployments, where obstructions may change over time or where rapid setup is required. As a result, the terminal may spend hours operating suboptimally, repeatedly connecting to blocked satellites, missing clearer alternatives, and suffering avoidable interruptions or degraded performance.

A faster obstruction map construction would unlock significant practical benefits for both existing mechanisms and emerging capabilities. For example, according to Starlink's

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), in part by the Canada Foundation for Innovation (CFI), and in part by the British Columbia Knowledge Development Fund (BCKDF) of Canada.

recent update [5], UTs now support beam switching, enabling handovers to secondary satellites when the primary link becomes obstructed. With a more rapidly constructed obstruction map, terminals could proactively identify and connect to satellites with clearer visibility, rather than relying on reactive switching triggered by link degradation. This would reduce the likelihood of switching from one obstructed link to another and enhance overall link stability, especially in partially obstructed or dynamically changing environments.

This motivates a natural extension: given that Starlink UTs already possess beam-switching capabilities and have access to tens of concurrently visible satellites, it becomes both feasible and valuable to incorporate an exploration–exploitation decision logic, one that balances maximizing immediate link quality with actively exploring less-known sky regions. Such a mechanism could significantly accelerate obstruction map construction while maintaining robust connectivity, all within the constraints of Starlink's existing architecture.

To realize this potential, we formulate the satellite selection problem faced by a UT as a sequential decision-making problem, where at each step the terminal must choose between exploiting a high-quality link to ensure communication quality or exploring a new, potentially suboptimal direction to expand its awareness of obstructions and improve future connectivity. To address this sequential decision-making challenge, we propose a principled framework that accelerates obstruction map construction while preserving service quality. Our key contributions are as follows:

- In Sec. III, we introduce a novel formulation based on *multi-armed bandits with sleeping arms (sleeping bandits)* for the UT's sequential satellite selection problem, where each orbit is modeled as an arm and availability is governed by the UT's FoV, satellite motion, and Earth's rotation. Our formulation introduces two key modeling innovations tailored to the obstruction map construction task: (i) a policy-driven filtering mechanism that allows designers to restrict the set of available orbits to prioritize unexplored directions, and (ii) a reward abstraction framework that decouples performance objectives from algorithm design, enabling flexible integration of metrics such as *round-trip time (RTT)*. This setup creates a general and configurable decision framework for balancing exploration and link quality without changing Starlink's existing system architecture.
- In Sec. IV, we propose **SkyBandit**, an algorithm which combines ideas from both *follow-the-perturbed-leader (FTPL)* and *upper confidence bound (UCB)* frameworks to operate near the frontier of the exploration–exploitation trade-off. Furthermore, SkyBandit is designed to handle the sleeping arms settings where an arm's availability can be either stochastic or adversarial.
- In Sec. V, we present a novel worst-case analysis based on a refined anti-concentration technique, establishing that SkyBandit achieves a regret bound of $O(\sqrt{NT \ln T})$, where N denotes the number of distinct orbits observed over T rounds. This improves upon classical bounds that depend on the total number of arms, **by depending only on the subset actually observed**.
- In Sec. VI, we integrate real Starlink obstruction map measurements with the refined version of xeoverse emulator [6] to evaluate our approach under both clear-sky and heavy-obstruction conditions. Our results demonstrate that the proposed method cuts obstruction map convergence time to just around 3 hours.

II. Related Work

To the best of our knowledge, there is currently no existing work that explicitly focuses on accelerating the construction of the obstruction map in LEO satellite systems. However, several prior studies have demonstrated the importance of obstruction maps in both theoretical and empirical analyses. In this section, we first review work that highlights the critical role of obstruction maps in Starlink, and then briefly discuss how bandit-based methods have been applied in satellite communication, particularly for connectivity optimization.

Applications of Obstruction Map: Recent research has leveraged Starlink's obstruction map as a valuable data source for understanding real-world satellite behavior. Prior work uses obstruction maps to characterize satellite visibility constraints and environmental blockages, and to analyze their impact on satellite assignment and network performance [2], [7]. Other studies further exploit obstruction maps to infer serving satellite identities based on geometric or trajectory alignment [3]. More recently, Liu et al. [8] use obstruction maps primarily to identify serving satellite IDs and examine how these identities correlate with measured throughput performance.

Beyond measurement studies, Zhang et al. [9] formulate the design of satellite–ground link topologies in LEO mega-constellations as a low-latency interconnection problem, highlighting that real-time knowledge of link visibility, essentially an obstruction or sky-visibility map, is indispensable for minimizing delay and jitter.

While these studies underscore the analytical value of obstruction maps, they do not address the overhead associated with constructing them. In particular, the long convergence time of the obstruction map, often spanning 6 to 12 hours, remains a critical bottleneck, especially in dynamic or mobile environments.

Bandits in Satellite: Among the algorithmic tools developed for decision-making under uncertainty, multi-armed bandits have shown promise in modeling dynamic satellite and link selection in space-based networks. Mohamed et al. [10] tackle LEO satellite selection for UAVs using a contextual bandit framework that explicitly considers time-varying visibility and link quality, while Wu et al. [11] apply sleeping bandits to dynamic path selection in space–ground integrated networks with evolving node availability. Related efforts have also explored online learning formulations for satellite handovers [12], orbital-plane resource allocation [13], and adaptive access-point selection in highly dynamic, partially observable environments [14], [15]. However, these works primarily focus on connectivity optimization rather than accelerating obstruction map construction through actively

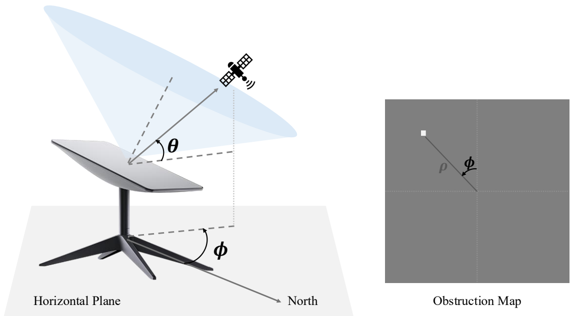


Fig. 3. The UT's FoV is shown in blue. θ and ϕ denote the elevation and azimuth angles to the satellite. In this work, the obstruction map is Earth-referenced, with the top-center pixel pointing toward geographic north.

guided exploration of unobserved sky regions, which is a direction our work explicitly and directly addresses.

III. System Model and Problem Formulation

A. System Model

In Starlink, a UT refers to the user-side phased-array antenna system that establishes and maintains connectivity with passing satellites. The UT's FoV defines the region of the sky within which satellites can be observed and served, as illustrated by the blue shaded area in Fig. 3. When the UT successfully connects to a satellite within its FoV, characterized by elevation angle θ and azimuth angle ϕ, the obstruction map marks the corresponding sky direction as clear. Specifically, the updated pixel is located at a radial distance ρ from the map center, offset by ϕ degrees from geographic north, where ρ is an empirically defined projection radius given by $\rho = \frac{90^\circ - \theta}{\beta}$. The constant β depends on the UT's FoV size [3], and we have $\beta = 1.29$ for our UTs. It is worth noting that Starlink obstruction maps can be represented in either an Earth-referenced or a terminal-referenced frame; in this work, we focus on the Earth-referenced frame, which is typically used by stationary UTs under non-roaming subscription plans, where the top-center pixel corresponds to geographic north.

Each UT maintains a 123×123 obstruction map, in which each pixel represents a fixed sky direction. The UT continuously records the RSSI of the current satellite link and classifies the associated direction as either clear or obstructed [16], [17]. Specifically, if the RSSI falls below a predefined threshold, suggesting signal degradation potentially caused by blockage, the corresponding pixel is marked as obstructed (typically shown in red); otherwise, it is marked as clear (illustrated as white in this work). Directions that have not yet been observed remain unexplored (e.g., black).

In the current Starlink architecture, satellite selection is centrally controlled, and the UT has no control over explored directions. To accelerate obstruction map construction, we consider a minimal extension that allows the UT to apply an exploration–exploitation logic when selecting from currently visible satellites. Importantly, the obstruction labeling mechanism remains unchanged: the UT still relies on RSSI measurements to classify obstruction map pixels as obstructed.

B. Problem Formulation

To model this active selection process, we formulate a sleeping bandit problem over orbital tracks. Let $\mathcal{O}$ denote the set of all orbits (the arms of the bandit), where each orbit corresponds to a fixed orbital plane in the Walker-Delta constellation, uniquely determined by its inclination and right ascension of the ascending node (RAAN) [18]. Given the short experimental timescale, RAAN drift due to orbital perturbations is negligible, and the orbital plane definition remains effectively constant. A sleeping bandit model captures the sequential decision-making process in which the arm (i.e., action) set evolves over time. Moreover, our modeling introduces two key innovations tailored to the obstruction-map construction task:

- **Controllable Filtering of Available Orbits:** The set of visible orbits at any given time is determined by satellite motion, the UT's FoV, and Earth's rotation. On top of these physical constraints, our approach allows the UT to apply policy-driven filtering to the visible orbits. For example, the UT could exclude or deprioritize orbits that correspond to sky directions already explored. This provides a mechanism to inject exploration bias into the action set, accelerating obstruction map coverage without requiring changes to the underlying system architecture.
- **Reward-based Orbit Selection for Link Quality Optimization:** Among the filtered available orbits, the UT aims to select the one that offers the best tradeoff between exploration and communication performance. To make this framework general and adaptable to different objectives, we assume that each orbit $o \in \mathcal{O}$ is associated with an unknown reward distribution $\mathcal{D}_o$ with a support in $[0, 1]$ and a mean $r(o)$. This reward can encode any desired performance metric, e.g., low *round-trip time (RTT)*, high throughput, low packet loss, or a combination thereof.

This setup yields a flexible decision-making framework, allowing designers to (i) define custom filtering rules over the available orbits $\mathcal{O}_t$ to guide exploration, and (ii) specify reward functions aligned with different performance objectives.

With the above two key innovations, the decision-making process proceeds as follows. At each discrete time step $t = 1, \ldots, T$ (each step representing an interval between handovers), the UT has a subset of orbits $\mathcal{O}_t \subseteq \mathcal{O}$ that are currently *available* (i.e., satisfy the physical constraints and filtering policies). Once the UT selects an orbit o_t at time t, it must still choose a specific satellite on that orbit for the actual connection. At the end of the time step t, the UT observes a random reward $r^t_{o_t} \sim \mathcal{D}_{o_t}$.

By following these principles, our approach supports a spectrum of policies, ranging from purely exploration-driven strategies that prioritize rapid obstruction map completion to performance-aware ones that balance map coverage with service quality. The framework further allows adaptive interpolation between these extremes, for example by prioritizing exploration during off-peak hours (e.g., overnight) when service demand is low.

To evaluate the UT's decision-making performance, we use a regret metric defined with respect to a dynamic optimum. Let $o_t^* := \arg\max_{o\in\mathcal{O}_t} r_o$ be the optimal orbit at time t (i.e. the orbit with the highest expected reward in the available set at that time, which may change in different time steps). Our objective is to minimize the cumulative regret after T rounds:

$$R(T) := \mathbf{E}\left[\sum_{t=1}^{T} r(o_t^\star) - r(o_t)\right], \tag{1}$$

where the expectation is taken with respect to the algorithm and the reward of the arms.

IV. The SkyBandit Algorithm

To accelerate obstruction map construction while preserving link quality, we propose a learning algorithm, i.e., SkyBandit, that operates under the problem formulation described in Sec. III. SkyBandit incorporates three key mechanisms: (i) a targeted orbit filtering strategy aligned with the goal of efficient environment discovery, (ii) a perturbed confidence-based selection rule that balances exploration and exploitation in the sleeping bandit setting, and (iii) a spatial-correlation-based satellite-selection strategy.

A. Available Orbit Set Determination

Recall that in each round t, the UT observes a subset of visible orbits, denoted as $\mathcal{O}_t^{\text{raw}}$, derived from the satellite ephemerides and the UT's FoV. However, not all of these orbits are equally valuable for obstruction map construction.

To prioritize informative regions of the sky, we define the filtered availability set $\mathcal{O}_t \subseteq \mathcal{O}_t^{\text{raw}}$ as follows: For each orbit $o \in \mathcal{O}_t^{\text{raw}}$, we predict the future sky trajectory of its satellites over the next handover interval (e.g., 15 seconds) using publicly available TLE data. Then, we compare these trajectories with the current obstruction map maintained by the UT. If at least one satellite on orbit o is expected to traverse a partially unexplored region (i.e., at least one uncolored pixel in its predicted path), then orbit o is included in $\mathcal{O}_t$. If no such orbits are found, we fall back to setting $\mathcal{O}_t := \mathcal{O}_t^{\text{raw}}$ to ensure coverage continuity and service availability.

This mechanism enables the UT to inject directional exploration into its action space while remaining compliant with the physical constraints of visibility and satellite availability.

B. Orbit Selection

Given the filtered orbit set $\mathcal{O}_t$, SkyBandit chooses an orbit $o_t \in \mathcal{O}_t$ by combining ideas from both the FTPL and UCB algorithms. The core idea is simple yet effective: SkyBandit samples a shared standard Gaussian random variable $w \sim \mathcal{N}(0,1)$ at the very beginning of the algorithm. For each available orbit $o \in \mathcal{O}_t$, the UT computes a reward estimate based on *perturbed* confidence bound $\bar{r}_o$:

$$\bar{r}_t(o) \leftarrow \hat{r}_{n_o(t)}(o) + w \cdot \sqrt{\frac{\alpha \ln t}{n_o(t)+1}}, \tag{2}$$

where $\hat{r}_{n_o(t)}(o)$ is the empirical mean reward of orbit o and $n_o(t)$ is the number of times that orbit o has been selected at

Algorithm 1: SkyBandit

Input: The TLE file
Output: Online decision sequence for satellite selection; updated obstruction map

Set $t \leftarrow 1$;
Initialize pull counts $n_o(1) \leftarrow 0$ and $\hat{r}_0(o) \leftarrow 1$ for all $o \in \mathcal{O}$; Initialize obstruction map $\mathcal{M}$ to be entirely undiscovered;
Draw a Gaussian random variable $w \sim \mathcal{N}(0,1)$;
while *terminal is online and handover trigger occurs* **do**
    Observe all available orbits within FoV $\mathcal{O}^{\text{raw}}$
    Construct orbit selection candidate set $\mathcal{O}_t \subseteq \mathcal{O}^{\text{raw}}$ according to the TLE file and current obstruction map
    for *each* $o \in \mathcal{O}_t$ **do**
        Compute $\hat{r}_{n_o(t)}(o) \leftarrow \frac{1}{t-1}\sum_{s=1}^{t-1} 1[o_s = o] r_o^s$;
        Compute perturbed score:
        $$\bar{r}_t(o) \leftarrow \hat{r}_{n_o(t)}(o) + w \cdot \sqrt{\frac{\alpha \ln t}{n_o(t)+1}};$$
    Select $o_t \leftarrow \arg\max_{o\in\mathcal{O}_t} \bar{r}_t(o)$;
    Conduct satellite selection procedure;
    Update $n_o(t+1) = \begin{cases} n_o(t)+1, & o = o_t, \\ n_o(t), & o \neq o_t; \end{cases}$
    Update obstruction map using observation;
    $t \leftarrow t+1$;

the beginning of time step t. The parameter $\alpha > 0$ controls the exploration rate.

As shown in (2), SkyBandit employs a randomized perturbation mechanism to control the width of the confidence interval via a tunable exploration parameter α. Unlike the classic UCB algorithm, which typically requires $\alpha \geq 1.5$ to satisfy Hoeffding-type concentration bounds, our approach allows for any value $\alpha > 0$, e.g., $\alpha = 0.8$ in our implementation. This enables tighter confidence intervals and supports a more adaptive trade-off between exploration and exploitation. The *slightly* reduced α value leads to improved empirical performance while preserving sufficient anti-concentration guarantees for theoretical analysis.

This randomized perturbation mechanism achieves a better adaptive behavior, particularly in sparse or dynamically constrained environments like our problem. As shown in Sec. V-A, it also yields a regret bound that scales with the number of observed orbits rather than the total number of possible arms, offering stronger guarantees in realistic settings where only a subset of orbits is available at any time.

C. Satellite Selection Strategy

Each orbit typically contains several satellites that are approximately uniformly spaced. After an orbit o_t is selected, the

UT must choose a specific satellite to connect to. Our satellite-selection strategy is guided by the observation that obstructions tend to exhibit spatial correlation rather than being uniformly distributed (see Sec. V-B).

To exploit this structure, the UT prioritizes satellites whose future sky trajectories: (i) pass through regions adjacent to already known unobstructed pixels, thereby expanding coverage in promising directions; or (ii) avoid areas near previously identified obstructions, reducing the likelihood of failed connections. This spatially informed strategy allows the UT to make more efficient use of each handover opportunity, either by consolidating regions of known clear sky or by cautiously probing new directions with a higher probability of being clear.

V. Analytical Results

A. Regret Analysis

Recall that $\mathcal{O}_t$ is changing over time, and thus the optimal orbit changes in each round. We want to characterize the gap between the orbit output by the learning algorithm and the optimal orbit. To this end, we analyze the (cumulative) regret of our learning algorithm with regard to the optimal action. Let $N := |\mathcal{O}_1 \cup \mathcal{O}_2 \cup \cdots \cup \mathcal{O}_T|$ denote the total number of available orbits over T rounds. Then, the regret defined in (1) is guaranteed by the following theorem.

Theorem 1. *For any sequence $\mathcal{O}_1, \ldots, \mathcal{O}_T$, the regret between the selection of SkyBandit and the optimal selection in each round is upper bounded by $O(\sqrt{NT \ln T})$ for any $T \geq N$.*

Proof. The proof follows three steps.

1) **Decompose Regret:** First, we define $\mathcal{E} := \{\forall t \in [T] : \hat{r}_{n_{o_t}(t)}(o_t) - r(o_t) \leq \sqrt{\frac{3 \ln T}{n_{o_t}(t)+1}}\}$ the event that the empirical mean reward is close to the true mean reward of the selected arm in each round, and denote by $\overline{\mathcal{E}}$ the complementary event of $\mathcal{E}$. Then, we can decompose the regret as follows.

$$\begin{aligned}
R(T) &:= \mathbf{E}\left[\sum_{t=1}^{T} (r(o_t^\star) - r(o_t))\,\mathbf{1}[\mathcal{E}]\right] + \mathbf{E}\left[\sum_{t=1}^{T} (r(o_t^\star) - r(o_t))\,\mathbf{1}[\overline{\mathcal{E}}]\right] \\
&\leq \mathbf{E}\left[\sum_{t=1}^{T} (r(o_t^\star) - r(o_t))\,\mathbf{1}[\mathcal{E}]\right] + T \cdot \Pr(\overline{\mathcal{E}}) \\
&\leq \underbrace{\mathbf{E}\left[\sum_{t=1}^{T} (r(o_t^\star) - \tilde{r}_t(o_t))\,\mathbf{1}[\mathcal{E}]\right]}_{=:(a)} + \underbrace{\mathbf{E}\left[\sum_{t=1}^{T} (\tilde{r}_t(o_t) - r(o_t))\,\mathbf{1}[\mathcal{E}]\right]}_{=:(b)} + 1,
\end{aligned}$$

where in the last inequality we use the fact that $\Pr(\overline{\mathcal{E}}) \leq \frac{1}{T}$ by the union bound and Hoeffding's inequality (see Lemma 2 in the Appendix).

2) **Bounding (a):** Term (a) represents the core analytical challenge and constitutes the main contribution of this proof. As shown in Lemma 3 in the Appendix, it is bounded by $O(\sqrt{NT \ln T})$ through a refined anti-concentration analysis of the shared Gaussian perturbation w used in SkyBandit. This analysis enables better control over the exploration bonus compared to classical confidence-based methods.

3) **Bounding (b):** Term (b) captures the gap between the empirical and true mean rewards of the selected arms. Using standard concentration inequalities, we bound this term by $O(\sqrt{NT \ln T})$, as formally established in Lemma 4 in the Appendix.

□

For sleeping bandits, Kleinberg et al. [19] establish a problem-dependent lower bound that can be tuned to yield a worst-case lower bound of $\Omega(\sqrt{NT \ln T})$ using our notations. Thus, our regret bound is minimax-optimal. In practice, the number of distinct orbits observed is often significantly smaller than the total, making our bound particularly well-suited for this realistic scenario.

B. Spatial Correlation

Claim 1. *Obstructions in the sky are not randomly distributed but exhibit strong spatial autocorrelation, manifesting as localized clusters rather than isolated or uniformly scattered pixels.*

To validate this, we use two standard measures of spatial autocorrelation: Moran's I [20], which quantifies global spatial similarity, and Geary's C [21], which captures local variability. Both metrics are computed over explored pixels (i.e., those labeled as obstructed or clear). While our main analysis focuses on UT2 (as denoted in Fig. 2), we further validated our findings by examining two additional UTs from the public dataset [22], located in Dallas and Stanford. All three UTs consistently report high Moran's I (0.908, 0.948, and 0.727 for Dallas, Stanford, and UT2, respectively) and low Geary's C (0.104, 0.064, and 0.285), confirming strong spatial autocorrelation and pronounced clustering of obstructions. Notably, the Dallas and Stanford terminals show even stronger spatial correlation than UT2, which further reinforces our claim. These results confirm that neighboring pixels tend to share similar visibility states, which is consistent with the intuition that obstructions arise from localized features such as buildings, trees, or terrain. It is worth noting that this analysis was not conducted for UT1, as it was installed in an unobstructed-sky environment without obstructions.

VI. Emulation with Measurement Dataset

A. Experimental Setup

We build upon xeoverse [6], [23], [24], an open-source real-time emulation platform for large-scale LEO satellite constellations, and extend it to support our proposed SkyBandit framework. xeoverse leverages Mininet [25] to instantiate satellites, ground stations, and UTs as virtual nodes running full Linux network stacks. Its modular design includes a Constellation Topology Manager for real-time satellite position and link availability tracking, an Adaptive Routing Engine, and a Physical Layer Modeling module that captures RF propagation effects. The platform supports dynamic topology evolution driven by satellite motion and time-varying visibility. In our implementation, we focus on *ground-to-satellite links (GSLs)* and retain xeoverse's obstruction map module for accurate terminal FoV modeling. We further integrate two new components: an Obstruction Learning module for environment-aware visibility inference and an Intelligent Exploration and Selection module that implements the SkyBandit algorithm within the real-time emulation loop.

xeoverse is originally designed for UT-to-UT experiments and makes several simplifying assumptions that diverge from real Starlink operations. In practice, Starlink user traffic is mainly routed through terrestrial gateways and the ground backbone, while ISL forwarding is not the primary data path. In contrast, xeoverse builds the constellation using a single shell and assumes that UT-to-UT traffic is carried entirely via inter-satellite links, which reduces realism when UTs connect to different shells or when cross-shell ISLs are unavailable. To address these issues, we decouple UT interactions in our experiment design and expand the constellation to include all available Starlink shells. This revised setup preserves the fidelity of obstruction-aware GSL modeling while enabling a more realistic evaluation of SkyBandit under heterogeneous visibility and connectivity conditions.

B. Baseline Methods

We evaluate three baseline satellite selection strategies without orbit selection phase, including

- **Highest Elevation**, which is the baseline that selects the visible satellite with the highest elevation within the UT's FoV. As observed in [3], Starlink typically prefers higher-elevation satellites during handovers. We therefore adopt this method as a representative extreme case, since higher elevation generally implies lower atmospheric attenuation, improved link stability, and longer intervals of high-quality connectivity.
- **Weighted Random**, which introduces probabilistic selection among eligible satellites, where each satellite's selection probability is inversely proportional to its distance from the terminal. The weight calculation assigns higher probabilities to closer satellites while maintaining non-zero selection chances for all eligible candidates: $w_{s_i} = (l_{\max} - l_i + 1)$, where $l_{\max}$ is the maximum distance among all eligible satellites and d_i is the distance to satellite s. The probability of s_i being selected is calculated as $\frac{w_{s_i}}{\sum_{j=1}^{n}(l_{\max}-l_j+1)}$.
- **SB-EXP3** [26], which is a state-of-the-art sleeping bandits algorithm designed for settings with time-varying arm availability. Each orbit o maintains an exponential weight q_o, initialized to 1. At each decision epoch, orbits are selected probabilistically from the available set $\mathcal{O}_t$ with probability $p_o = \frac{q_o}{\sum_{j\in\mathcal{O}_t} q_j}$. After observing the reward, weights are updated via $q_o \leftarrow q_o \cdot \exp(-\eta \cdot \ell_o/p_o)$, where the loss is defined as $\ell_o = 1 - r_o$, where we omit the superscript t for simplicity. Then the satellite with maximum elevation within the orbit is selected. Although this probabilistic selection implicitly balances exploration and exploitation without explicit bonus terms, its exploration efficiency is limited by treating orbits as independent arms without considering spatial correlation among sky regions.

TABLE I
EXPLORATION EFFICIENCY BETWEEN STARLINK AND SKYBANDIT ON COMMUNICATION-USABLE SKY REGIONS FOR UT1 AND UT2.

Method	UT	Hours to $\mathcal{R}_{\text{exp}} \geq 0.8 \downarrow$	$\mathcal{R}_{\text{exp}}$ at 6h $\uparrow$
Starlink	UT1	7.46	0.761
	UT2	5.68	0.808
SB-EXP3	UT1	$5.49 \pm 7.7\times10^{-2}$	$0.824 \pm 3.9\times10^{-3}$
	UT2	$4.42 \pm 1.1\times10^{-1}$	$0.888 \pm 5.8\times10^{-3}$
SkyBandit	UT1	$3.01 \pm 2.2\times10^{-2}$	$0.984 \pm 8.2\times10^{-4}$
	UT2	$3.19 \pm 3.4\times10^{-2}$	$0.971 \pm 1.2\times10^{-3}$

$\downarrow$ ($\uparrow$) indicates that lower (higher) values are better.
$\mathcal{R}_{\text{exp.}}$ is the exploration ratio defined in (3).
All values are reported as mean $\pm$ 95% CI over 20 independent trials.

C. Implementation Details

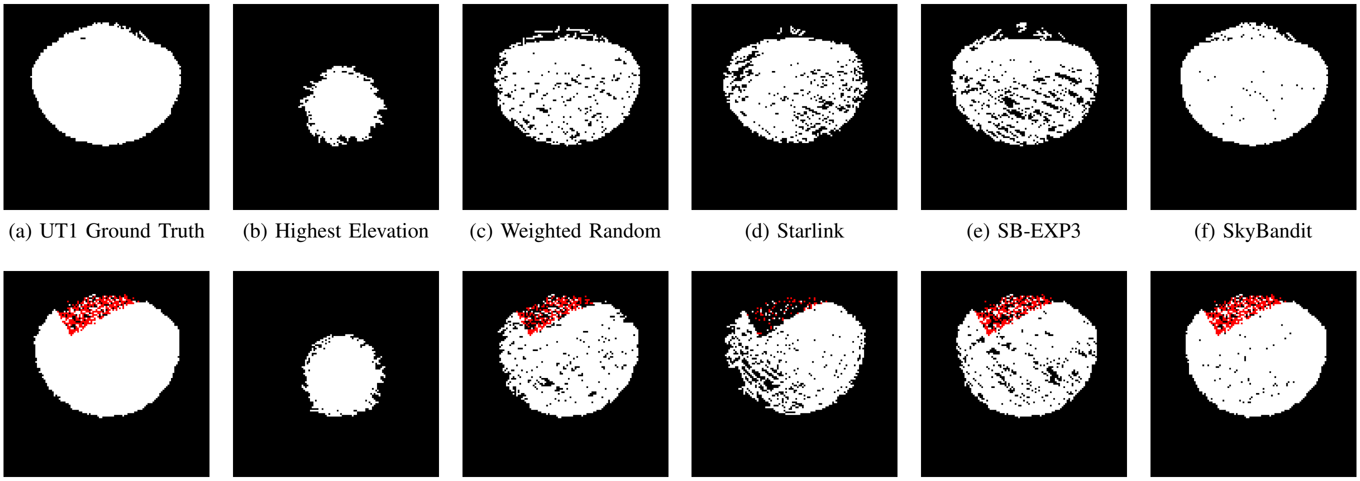
To faithfully capture the actual constraints of each UT, whether software- or hardware-induced, we restrict the simulated UT's FoV to the exact sky regions observed in the ground truth obstruction maps for both UT1 and UT2. During simulation, connection attempts are validated as follows: unobstructed areas (white pixels) allow successful connections, obstructed areas (red pixels) result in failures, and unknown regions (black pixels) are treated as invisible and considered outside the FoV. This approach ensures that hardware and practical limitations affecting the real UT's FoV are properly accounted for, making our simulation results directly comparable to real-world Starlink user terminal behavior.

To achieve continuous and more realistic coverage of satellite trajectories, we interpolate satellite positions at 0.1-second intervals between observation timestamps, following Starlink's continuous map update paradigm. This fine-grained sampling ensures that even rapid satellite movements are fully captured, preventing gaps in obstruction learning. Each interpolated point is mapped to its corresponding grid cell and classified according to the constrained ground truth, guaranteeing comprehensive coverage of each satellite's path within the user terminal's FoV. For performance evaluation, we measure network latency by issuing ping commands within the Mininet topology and recording the RTT. The RTT metric is processed using a sliding window of the most recent 10 samples, which allows our learning algorithm to reflect current network conditions, smooth out transient fluctuations, and maintain constant memory usage throughout the simulation. This approach ensures both practical responsiveness and computational efficiency in our evaluation framework.

D. Results

We evaluate the performance of our proposed SkyBandit algorithm and baseline methods across multiple metrics using data from the aforementioned two user terminals UT1 and UT2. All results are reported after 6 hours of continuous operation for each algorithm.

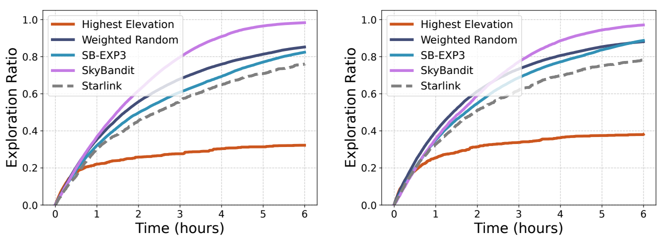
To provide an intuitive comparison of algorithm performance, we directly visualize the learned obstruction maps for each algorithm and scenario in Fig. 4, alongside the ground



(a) UT1 Ground Truth (b) Highest Elevation (c) Weighted Random (d) Starlink (e) SB-EXP3 (f) SkyBandit

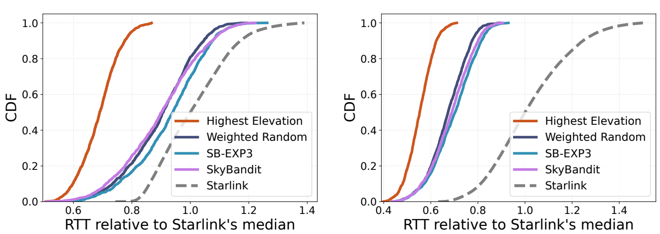
(g) UT2 Ground Truth (h) Highest Elevation (i) Weighted Random (j) Starlink (k) SB-EXP3 (l) SkyBandit

Fig. 4. Learned obstruction maps for UT1 and UT2 scenarios, with each algorithm running for 6 hours. **Pixel colors: Red**: Obstruction, **Black**: No data, **White**: Clear (unobstructed). All obstruction maps are represented in the Earth-referenced coordinate frame.



(a) UT1 (b) UT2

Fig. 5. Exploration ratio over time for UT1 and UT2, averaged over 20 trials.



(a) UT1 (b) UT2

Fig. 6. CDFs of RTT normalized by the median RTT of the Starlink baseline.

truth maps obtained via gRPC API calls. The visualization illustrates how different satellite selection strategies influence the spatial coverage of the user terminal's sky. As shown in the figure, the Highest Elevation strategy exhibits very limited sky coverage because its decisions are inherently restricted to regions where satellites reach the highest elevation, thus narrowing the scope of exploration. The Weighted Random strategy achieves broader spatial coverage through proximity-weighted sampling, but produces scattered gaps due to its lack of feedback-driven adaptation. In contrast, the bandit-based policies, SB-EXP3 and SkyBandit, leverage reward feedback to guide exploration over time: SB-EXP3 improves coverage but remains patchy, while SkyBandit produces the most complete and consistent obstruction maps across both UT scenarios. Consequently, SkyBandit noticeably improves exploration efficiency by learning where to explore rather than merely exploring broadly.

For each algorithm, we show its exploration ratio over time in Fig. 5. Specifically, the exploration ratio at time τ is calculated as

$$\mathcal{R}_{\text{exp.}}(\tau) = \frac{|\{(i,j) : \mathcal{M}_{i,j}(\tau) = 0 \wedge (i,j) \in \mathcal{V}\}|}{|\mathcal{V}|}, \quad (3)$$

where $\mathcal{M}_{i,j}(\tau)$ is the pixel in the i-th row and j-th column of the obstruction map at time τ, taking the value 0 for unobstructed, 1 for obstructed, and -1 for unobserved. The set $\mathcal{V}$ comprises all pixels that are usable for communication (i.e., those marked as white in the ground truth) within the UT's field of view.

Table I compares the exploration efficiency of Starlink's native strategy, SB-EXP3, and SkyBandit for UT1 and UT2. SB-EXP3 improves over Starlink by probabilistically selecting available orbits, reducing the time to reach $\mathcal{R}$exp. $\geq$ 0.8 from 7.46 to 5.49 hours for UT1 and from 5.68 to 4.42 hours for UT2, and achieving higher exploration ratios after 6 hours. However, SkyBandit further significantly accelerates exploration, reaching $\mathcal{R}$exp. $\geq$ 0.8 in only 3.01 hours for UT1 and 3.19 hours for UT2, corresponding to improvements of 59.7% and 43.8% over Starlink. After 6 hours, SkyBandit attains the highest exploration ratios (0.984 for UT1 and 0.971 for UT2). The similar exploration times across UTs indicate that SkyBandit effectively focuses on unobstructed sky regions, enabling efficient and consistent exploration across both clear-sky and obstructed environments. Overall, these results demonstrate that while SB-EXP3 benefits from adaptive probabilistic selection, SkyBandit's spatially informed design enables significantly faster and more consistent exploration across both clear-sky and obstructed environments.

These results demonstrate that SkyBandit not only accelerates the discovery of clear sky regions but also achieves more comprehensive coverage within a fixed period, thereby improving environment awareness over the baseline approach. Fig. 6 compares RTT distributions normalized by the median RTT of the real Starlink system, i.e., RTTnorm $= \frac{\text{RTT}}{\text{median}(\text{RTT}_{\text{Starlink}})}$, as absolute RTTs between emulation and real networks are not directly comparable. This normalization is used to assess whether the proposed algorithms introduce systematic RTT degradation relative to the operational baseline. We observe that the normalized RTT distributions of the proposed algorithms closely track the Starlink baseline without a systematic right shift, indicating that their adoption preserves latency behavior within the range observed in the operational system.

E. Deployment Considerations

Current Starlink systems rely on centralized network-side mechanisms to manage user terminal handovers and serving satellite assignments, rather than allowing terminals to autonomously control satellite selection. As a result, SkyBandit is not intended to replace existing centralized scheduling in operational networks. Nevertheless, our approach remains practically deployable: SkyBandit operates entirely at the user terminal and requires no modification to Starlink's infrastructure. A terminal can run SkyBandit when newly installed at a location or during off-peak hours to learn obstruction-aware satellite preferences, which can then be seamlessly used alongside the existing centralized handover mechanism.

VII. Conclusion

We present SkyBandit, an online learning framework for accelerating obstruction map construction in Starlink's LEO network. Building on the system's existing capability for proactive satellite selection, we reformulate map construction as a sleeping bandit problem over dynamically visible orbits.

SkyBandit integrates ideas from FTPL and UCB through a single shared Gaussian perturbation, enabling better exploration–exploitation trade-offs. Our analysis establishes a regret bound of $O(\sqrt{NT\ln T})$, where N is the number of observed orbits, offering tighter guaranties than classical bounds that depend on the full action space.

Evaluated using real Starlink traces and the refined version of xeoverse emulator, SkyBandit reduces obstruction map convergence time to just around 3 hours, while preserving link quality. Importantly, the approach is fully compatible with the current Starlink architecture, demonstrating the potential of learning-based decision-making for enhancing environment awareness and connectivity in large-scale LEO constellations. The authors have provided public access to their code at https://github.com/quanwei-zhang/OpenSkyBandit.

Appendix

Let $\mathcal{O}_1^T := \bigcup_{t=1}^T \mathcal{O}_t$ be the union of all the available arms. Then, by the definition of N, we have that $N = |\mathcal{O}_1^T|$, and $\sum_{o\in\mathcal{O}_1^T} n_o(T) = T$.

Lemma 1. *For any sequence $o_1, \ldots, o_t$, we have that*

$$2\sqrt{T} - 2N \leq \sum_{t=1}^{T} \sqrt{\frac{1}{n_{o_t}(t)+1}} \leq 2\sqrt{NT} - N.$$

Proof. Let $\tau_o(s)$ denote the round where o is pulled for the s time, i.e., $n_o(\tau_o(s)) = s-1$. Then, we have

$$\begin{aligned}\sum_{t=1}^{T} \sqrt{\frac{1}{n_{o_t}(t)+1}} &= \sum_{o\in\mathcal{O}_1^T} \sum_{s=1}^{n_o(T)} \sum_{t=\tau_o(s)}^{\tau_o(s+1)-1} \mathbf{1}[o_t = o] \sqrt{\frac{1}{n_o(t)+1}} \\ &= \sum_{o\in\mathcal{O}_1^T} \sum_{s=1}^{n_o(T)} \sqrt{\frac{1}{s}}\end{aligned}$$

Since

$$\int_1^{n_o(T)} \sqrt{\frac{1}{s}} ds \leq \sum_{s=1}^{n_o(T)} \sqrt{\frac{1}{s}} \leq 1 + \int_1^{n_o(T)} \sqrt{\frac{1}{s}} ds,$$

we have that

$$2\left(\sqrt{n_o(T)} - 1\right) \leq \sum_{s=1}^{n_o(T)} \sqrt{\frac{1}{s}} \leq 1 + 2\left(\sqrt{n_o(T)} - 1\right).$$

Sum over $o \in \mathcal{O}_t^T$ and we have the lower bound as follows

$$\sum_{o\in\mathcal{O}_1^T} 2\left(\sqrt{n_o(T)} - 1\right) \geq 2\sqrt{\sum_{o\in\mathcal{O}_1^T} n_o(T)} - 2N = 2\sqrt{T} - 2N,$$

where the last inequality we use the fact that $\sum_{o\in\mathcal{O}_1^T} n_o(T) = T$. On the other hand, the upper bound can be obtained as follows by using the Cauchy–Schwarz inequality:

$$\sum_{o\in\mathcal{O}_1^T} \left(1 + 2\left(\sqrt{n_o(T)} - 1\right)\right) \leq 2\sqrt{N \sum_{o\in\mathcal{O}_1^T} n_o(T)} - N \leq 2\sqrt{NT} - N$$

□

Lemma 2. *The event $\overline{\mathcal{E}}$ occurs with probability at most $\frac{1}{T}$.*

Proof. By using the union bound and Hoeffding's inequality, we have that

$$\begin{aligned}\Pr(\overline{\mathcal{E}}) &\leq \sum_{t=1}^{T} \Pr\left(\hat{r}_{n_{o_t}(t)}(o_t) - r(o_t) \geq \sqrt{\frac{3\ln T}{n_{o_t}(t)+1}}\right) \\ &\leq \sum_{t=1}^{T} \Pr\left(\hat{r}_{n_{o_t}(t)}(o_t) - r(o_t) \geq \sqrt{\frac{3\ln T}{2n_{o_t}(t)}}\right) \\ &\leq \sum_{t=1}^{T} \sum_{s=1}^{t} \sum_{o\in\mathcal{O}_1^T} \mathbf{1}[o_t = o] \Pr\left(\hat{r}_s(o) - r(o) \geq \sqrt{\frac{3\ln T}{2s}}\right) \\ &\leq \sum_{t=1}^{T} \sum_{s=1}^{t} \sum_{o\in\mathcal{O}_1^T} \mathbf{1}[o_t = o] \frac{1}{T^3} \leq \sum_{t=1}^{T} \frac{1}{T^2} \leq \frac{1}{T}.\end{aligned}$$

□

Lemma 3. *The optimism part (a) is bounded as follows.*

$$\mathbf{E}\left[\sum_{t=1}^{T} \left(r(o_t^\star) - \bar{r}_t(o_t)\right) \mathbf{1}[\mathcal{E}]\right] \leq 500\sqrt{NT\ln T}.$$

Proof. Now, let's bound term (a) as follows. Let $\mathbf{E}_w[\cdot] = \mathbf{E}[\cdot \mid w]$ be the expectation conditioned on w, where the only randomness is the reward.

$$(a) = \mathbf{E}\left[\sum_{t=1}^{T}\left(r(o_t^\star) - \bar{r}_t(o_t)\right)\mathbf{1}[\mathcal{E}]\right]$$
$$= \mathbf{E}\left[\underbrace{\mathbf{E}_w\left[\sum_{t=1}^{T}\left(r(o_t^\star) - \bar{r}_t(o_t)\right)\mathbf{1}[\mathcal{E}]\right]}_{=:\alpha}\right].$$

If $\alpha \leq 0$, the above item is trivially bounded by 0. Thus, we consider the case where $\alpha > 0$. By Markov's inequality, we have that for any general random variable X:

$$\Pr(X \geq \alpha) = \Pr(X^+ \geq \alpha) \leq \frac{\mathbf{E}[X^+]}{\alpha},$$

where $X^+ := \max\{0, X\}$.

Now let

$$X := \left(\sum_{t=1}^{T}\bar{r}_t(o_t) - \mathbf{E}_w\left[\sum_{t=1}^{T}\bar{r}_t(o_t)\right]\right)\mathbf{1}[\mathcal{E}].$$

Then, we have that

$$\alpha \leq \frac{\mathbf{E}[X^+]}{\Pr(X\geq\alpha)} \leq \frac{\mathbf{E}\left[\left(\sum_{t=1}^{T}\bar{r}_t(o_t) - \mathbf{E}_w\left[\sum_{t=1}^{T}\bar{r}_t(o_t)\right]\right)^+\mathbf{1}[\mathcal{E}]\right]}{\Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t) \geq \sum_{t=1}^{T}r(o_t^\star)\right)},$$

where the denominator is obtained as follows

$$\begin{aligned}
&\Pr(X \geq \alpha)\\
&= \Pr\left(\left(\sum_{t=1}^{T}\bar{r}_t(o_t) - \mathbf{E}_w\left[\sum_{t=1}^{T}\bar{r}_t(o_t)\right]\right)\mathbf{1}[\mathcal{E}] \geq \mathbf{E}_w\left[\sum_{t=1}^{T}\left(r(o_t^\star) - \bar{r}_t(o_t)\right)\mathbf{1}[\mathcal{E}]\right]\right)\\
&\geq \Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t) - \mathbf{E}_w\left[\sum_{t=1}^{T}\bar{r}_t(o_t)\right] \geq \mathbf{E}_w\left[\sum_{t=1}^{T}\left(r(o_t^\star) - \bar{r}_t(o_t)\right)\right]\right)\\
&\geq \Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t) \geq \mathbf{E}_w\left[\sum_{t=1}^{T}r(o_t^\star)\right]\right) = \Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t) \geq \sum_{t=1}^{T}r(o_t^\star)\right).
\end{aligned}$$

Bounding the numerator: Next, by the definition of $\bar{r}_t(o_t)$, we have.

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{t=1}^{T}\bar{r}_t(o_t) - \mathbf{E}_w\left[\sum_{t=1}^{T}\bar{r}_t(o_t)\right]\right)^+\mathbf{1}[\mathcal{E}]\right]\\
&\leq \mathbf{E}\left[\left(2\left|\sum_{t=1}^{T}\hat{r}_{n_{o_t}(t)}(o_t) - r(o_t)\right| + |w|\sqrt{0.8\ln T}\left|\sum_{t=1}^{T}\sqrt{\frac{1}{n_{o_t}(t)+1}} - \mathbf{E}_w\sum_{t=1}^{T}\sqrt{\frac{1}{n_{o_t}(t)+1}}\right|\right)\mathbf{1}[\mathcal{E}]\right]\\
&\leq \mathbf{E}\left[2\left|\sum_{t=1}^{T}\sqrt{\frac{3\ln T}{n_{o_t}+1}}\right| + |w|\sqrt{0.8\ln T}\left|\sum_{t=1}^{T}\sqrt{\frac{1}{n_{o_t}(t)+1}} - \mathbf{E}_w\sum_{t=1}^{T}\sqrt{\frac{1}{n_{o_t}(t)+1}}\right|\right].
\end{aligned}$$

Since Lemma 1 gives that

$$\sum_{t=1}^{T}\sqrt{\frac{1}{n_{o_t}(t)+1}} \leq 2\sqrt{NT} - N.$$

Thus, we have that

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{t=1}^{T}\bar{r}_t(o_t) - \mathbf{E}_w\left[\sum_{t=1}^{T}\bar{r}_t(o_t)\right]\right)^+\mathbf{1}[\mathcal{E}]\right]\\
&\leq \mathbf{E}\left[2\left|2\sqrt{NT} - N\right| + |w|\sqrt{0.8\ln T}\left|\left(2\sqrt{NT} - 2\sqrt{T} + N\right)\right|\right]\\
&\leq 6\sqrt{NT\ln T} + N\sqrt{\ln T} \leq 5\sqrt{NT\ln T},
\end{aligned}$$

where we use the fact that $|w|$ is a half-normal distribution whose expectation is $\sqrt{\frac{2}{\pi}}$, and the assumption that $T \geq N$.

Bound the denominator: Since $\bar{r}_t(o_t) \geq \bar{r}_t(o_t^\star)$, we have that

$$\Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t) \geq \sum_{t=1}^{T}r(o_t^\star)\right) \geq \Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t^\star) \geq \sum_{t=1}^{T}r(o_t^\star)\right).$$

Now, we define event $\mathcal{H} := \{\forall t \in [T], r(o_t^\star) - \hat{r}_{n_{o_t^\star}(t)}(o_t^\star) \leq \sqrt{\frac{2\ln t}{n_{o_t^\star}(t)+1}}\}$. The probability of $\mathcal{H}$ happens is at least:

$$\begin{aligned}
&\Pr(\mathcal{H}) = 1 - \Pr(\overline{\mathcal{H}})\\
&\geq 1 - \sum_{t=1}^{T}\Pr\left(\hat{r}_{n_{o_t^\star}(t)}(o_t^\star) - r(o_t^\star) \leq -\sqrt{\frac{2\ln t}{n_{o_t^\star}(t)+1}}\right)\\
&= 1 - \left(0 + \sum_{t=2}^{T}\Pr\left(\hat{r}_{n_{o_t^\star}(t)}(o_t^\star) - r(o_t^\star) \leq -\sqrt{\frac{2\ln t}{2n_{o_t^\star}(t)}}\right)\right)\\
&\geq 1 - \sum_{t=2}^{\infty}\exp\left(-2\ln t\right)\sum_{O_t\subseteq\mathcal{O}}\Pr(\mathcal{O}_t) = 2 - \frac{\pi^2}{6} \geq 0.35.
\end{aligned}$$

Then, we have that

$$\begin{aligned}
&\Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t^\star) \geq \sum_{t=1}^{T}r(o_t^\star)\right) \geq \Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t^\star) \geq \sum_{t=1}^{T}r_t(o_t^\star), \mathcal{H}\right)\\
&= \Pr\left(\sum_{t=1}^{T}\bar{r}_t(o_t^\star) \geq \sum_{t=1}^{T}r(o_t^\star) \mid \mathcal{H}\right)\Pr(\mathcal{H})\\
&\geq \Pr\left(w\sum_{t=1}^{T}\sqrt{\frac{0.8\ln t}{n_{o_t^\star}(t)+1}} \geq \sum_{t=1}^{T}\sqrt{\frac{2\ln t}{n_{o_t^\star}(t)+1}}\right)\cdot 0.35\\
&= \left(1 - \Phi\left(\sqrt{\frac{5}{2}}\right)\right)\cdot 0.35 \geq 0.01.
\end{aligned}$$

Therefore, (a) is bounded by $\frac{5\sqrt{NT\ln T}}{0.01} = 500\sqrt{NT\ln T}$. □

Lemma 4. *The deviation part (b) is bounded as follows.*

$$\mathbf{E}\left[\sum_{t=1}^{T}\left(\bar{r}_t(o_t) - r(o_t)\right)\mathbf{1}[\mathcal{E}]\right] \leq 2\sqrt{3NT\ln T}.$$

Proof. By the definition of $\bar{r}_t(o_t)$, we have that

$$\begin{aligned}
\text{(b)} &= \mathbf{E}\left[\sum_{t=1}^{T}\left(\hat{r}_{n_{o_t}(t)}(o_t) + w\sqrt{\frac{0.8\ln t}{n_{o_t}(t)+1}} - r(o_t)\right)\mathbf{1}[\mathcal{E}]\right]\\
&\leq \mathbf{E}\left[\sum_{t=1}^{T}\left(\hat{r}_{n_{o_t}(t)}(o_t) - r(o_t)\right)\mathbf{1}[\mathcal{E}]\right] + \mathbf{E}\left[w\sum_{t=1}^{T}\sqrt{\frac{0.8\ln t}{n_{o_t}(t)+1}}\right]\\
&\leq \mathbf{E}\left[\sum_{t=1}^{T}\sqrt{\frac{3\ln T}{n_{o_t}(t)+1}}\right] + \mathbf{E}\left[w\sum_{t=1}^{T}\sqrt{\frac{0.8\ln t}{n_{o_t}(t)+1}}\right]\\
&\leq 2\sqrt{3NT\ln T} - N\sqrt{3\ln T} + \mathbf{E}[w]\left(2\sqrt{0.8NT\ln T} - N\sqrt{0.8\ln T}\right)\\
&= 2\sqrt{3NT\ln T} - N\sqrt{3\ln T} \leq 2\sqrt{3NT\ln T},
\end{aligned}$$

where the second inequality is due to the definition of event $\mathcal{E}$, and the last equality is due to the definition of w such that $\mathbf{E}[w] = 0$. □

References

[1] Z. Lai, Z. Li, Q. Wu, H. Li, W. Liu, Y. Liu, X. Xie, Y. Li, and J. Liu, "Mind the Misleading Effects of LEO Mobility on End-to-End Congestion Control," in *Proceedings of the 23rd ACM Workshop on Hot Topics in Networks*, pp. 34–42, 2024.

[2] H. B. Tanveer, M. Puchol, R. Singh, A. Bianchi, and R. Nithyanand, "Making Sense of Constellations: Methodologies for Understanding Starlink's Scheduling Algorithms," in *Companion of the 19th International Conference on Emerging Networking EXperiments and Technologies*, pp. 37–43, 2023.

[3] A. Ahangarpour, J. Zhao, and J. Pan, "Trajectory-based Serving Satellite Identification with User Terminal's Field-of-View," in *Proceedings of the 2nd International Workshop on LEO Networking and Communication*, pp. 55–60, 2024.

[4] Starlink, "Understanding Obstructions and Network Performance." Starlink Support, 2025. [Online]. Available: https://www.starlink.com/gb/support/article/71707228-cea9-52d5-6134-f3de8cc7437f.

[5] Starlink, "Starlink Beam Switching," 2025. [Online]. Available: https://starlink.com/updates/starlink-beam-switching.

[6] M. M. Kassem and N. Sastry, "$x$eoverse: A Real-Time Simulation Platform for Large LEO Satellite Mega-Constellations," in *Proceedings of 2024 IFIP Networking Conference*, pp. 1–9, IEEE, 2024.

[7] L. Izhikevich, M. Tran, K. Izhikevich, G. Akiwate, and Z. Durumeric, "Democratizing LEO Satellite Network Measurement," in *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, pp. 1–26, ACM New York, NY, USA, 2024.

[8] Z. Liu, F.-X. G. Reidys, S. Tanveer, and D. Vasisht, "Vivisecting Starlink Throughput: Measurement and Prediction," *Proceedings of the ACM on Networking*, vol. 3, no. CoNEXT4, pp. 1–23, 2025.

[9] Y. Zhang, Q. Wu, Z. Lai, and H. Li, "Enabling Low-Latency-Capable Satellite-Ground Topology for Emerging LEO Satellite Networks," in *Proceedings of 2022-IEEE Conference On Computer Communications*, pp. 1329–1338, IEEE, 2022.

[10] E. M. Mohamed and M. Rihan, "Bandit Approach for Unmanned Aerial Vehicle-centric Low Earth Orbit Satellite Selection," *Elsevier Digital Signal Processing*, vol. 151, p. 104546, 2024.

[11] H. Wu, Y. Xu, S. Cao, J. Liu, H. Takakura, and S. Norio, "Sleeping Multi-Armed Bandit-Based Path Selection in Space-Ground Semantic Communication Networks," in *2025 IEEE Wireless Communications and Networking Conference*, pp. 1–6, IEEE, 2025.

[12] E. M. Mohamed, S. Hashima, K. Hatano, and H. S. Khallaf, "LEO-Assisted Aerial Deployment in Post-Disaster Scenarios Using a Combinatorial Bandit and Genetic Algorithmic Approach," *Electronics*, vol. 12, no. 24, p. 4964, 2023.

[13] L.-H. Shen, Y. Ho, K.-T. Feng, L.-L. Yang, S.-H. Wu, and J.-M. Wu, "Hierarchical Multi-agent Multi-armed Bandit for Resource Allocation in Multi-LEO Satellite Constellation Networks," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, pp. 1–5, IEEE, 2023.

[14] V. Saxena, J. Jaldén, J. E. Gonzalez, M. Bengtsson, H. Tullberg, and I. Stoica, "Contextual Multi-armed Bandits for Link Adaptation in Cellular Networks," in *Proceedings of the 2019 Workshop on Network Meets AI & ML*, pp. 44–49, 2019.

[15] 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 Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 12101–12112, 2021.

[16] M. S. McCormick, "Determination of Electronic Beam Steering Angles," 2024.

[17] T. L. e. a. Robinson, "Apparatuses and Methods for Facilitating Satellite Visibility for Earth-based Antenna Systems," 2024.

[18] B. Matthiesen, N. Razmi, I. Leyva-Mayorga, A. Dekorsy, and P. Popovski, "Federated Learning in Satellite Constellations," *IEEE Network*, vol. 38, no. 2, pp. 232–239, 2023.

[19] R. Kleinberg, A. Niculescu-Mizil, and Y. Sharma, "Regret Bounds for Sleeping Experts and Bandits," *Machine Learning*, vol. 80, no. 2-3, pp. 245–272, 2010.

[20] P. A. Moran, "Notes on Continuous Stochastic Phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950.

[21] R. C. Geary, "The Contiguity Ratio and Statistical Mapping," *The incorporated statistician*, vol. 5, no. 3, pp. 115–146, 1954.

[22] J. Zhao and J. Pan, "LENS: A LEO Satellite Network Measurement Dataset," in *Proceedings of the 15th ACM Multimedia Systems Conference*, MMSys '24, (New York, NY, USA), p. 278–284, Association for Computing Machinery, 2024.

[23] D. Ron, F. A. Yusufzai, S. Kwakye, S. Roy, N. Sastry, and V. K. Shah, "Time-Dependent Network Topology Optimization for LEO Satellite Constellations," in *Proceedings of 2025 IEEE Conference on Computer Communications*, pp. 1–10, IEEE, 2025.

[24] University of Surrey Networks and Systems Research Group, "$x$eoverse: Real-time Simulation Platform for Large LEO Satellite Mega-Constellations." https://netsys.surrey.ac.uk/softwares/xeoverse/, 2025. Accessed: Dec. 2025.

[25] N. Handigol, B. Heller, V. Jeyakumar, B. Lantz, and N. McKeown, "Reproducible Network Experiments Using Container-based Emulation," in *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pp. 253–264, 2012.

[26] Q. M. Nguyen and N. Mehta, "Near-Optimal Per-Action Regret Bounds for Sleeping Bandits," in *International Conference on Artificial Intelligence and Statistics*, pp. 2827–2835, PMLR, 2024.