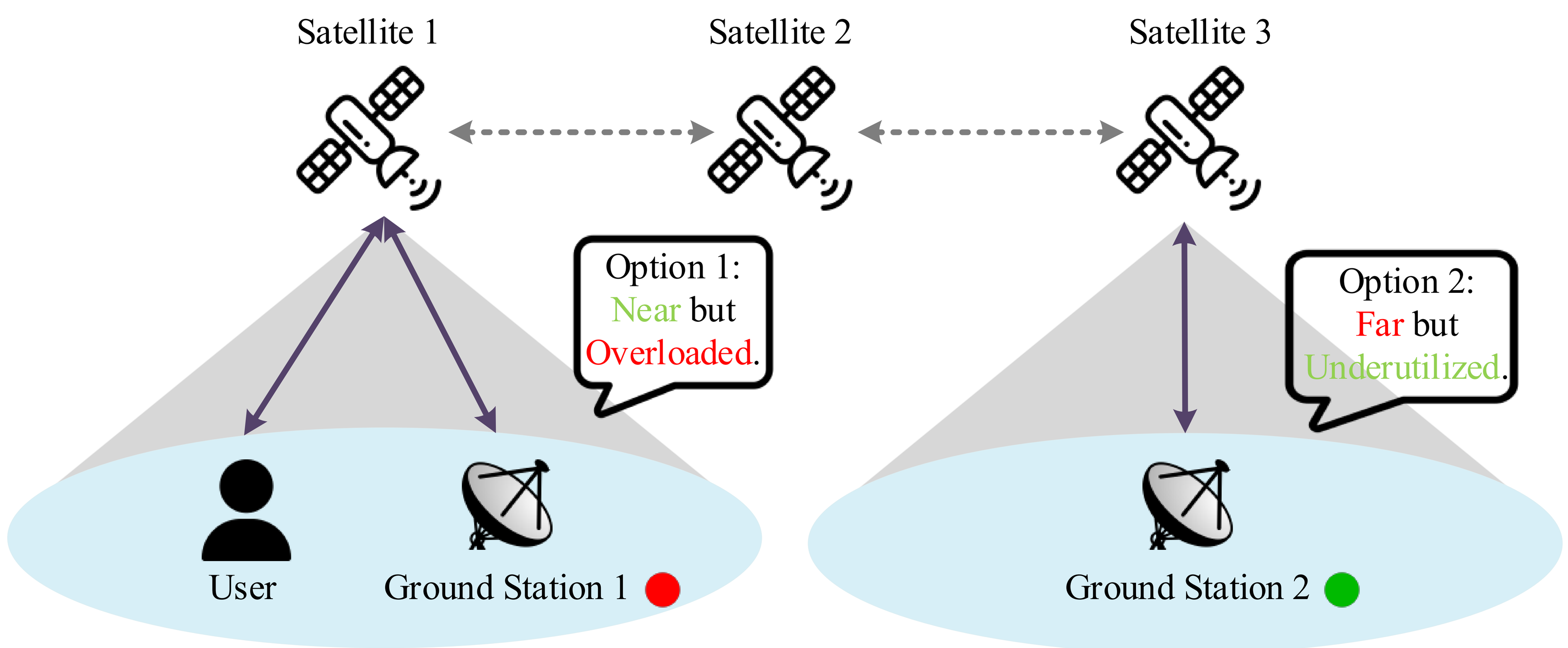
Rethinking Nearest-Is-Best: Towards Load-aware Ground Station Assignment for LEO Satellite Networks

Songshi Dou[1], Jinxian Wu[2], A-Long Jin[3], Zehua Guo[2], and Kwan L. Yeung[1]

[1]*The University of Hong Kong*, [2]*Beijing Institute of Technology*, [3]*Xi'an Jiaotong-Liverpool University*

① Introduction

- ***The Rise of LEO Mega-Constellations***
  - **Global Coverage**: LEO constellations like Space's Starlink and OneWeb are deploying thousands of satellites to reshape global broadband.
  - **"Bent-pipe" Model**: User terminal → overhead satellite → local Ground Station (GS); the satellite acts as a simple relay.
  - **Dependency**: Depends on a dense GS network to reach the terrestrial Internet.
  - **Current Routing Standard**: Industry-standard "Nearest-GS" picks the geographically closest GS to minimize propagation delay.



Illustrative example: the nearest GS is saturated; routing via ISL to a distant GS restores balance.

② Motivation & Challenges

- ***The Ground Segment Bottleneck***
  - **Uneven Traffic Demand**: Traffic is spatially non-uniform — dense populations form hotspots; oceans/rural areas are nearly idle.
  - **Capacity Constraints**:
    - Each GS has a fixed aggregate GSL (Ground-to-Satellite Link) capacity.
    - When high user demand is funneled to a single "nearest" GS, this capacity is quickly saturated.
  - **The Consequence**: Severe load imbalance — some GSs hit >100% utilization, causing packet loss and service degradation.
- ***The Opportunity: ISL-Enabled Flexibility***
  - **Dynamic Mesh in Space**: Modern satellites (*e.g.*, Starlink v1.5+) feature Inter-Satellite Links (ISLs), allowing data to hop between satellites before reaching the ground.
  - **Offloading Potential**: Route traffic to distant, underutilized GSs to bypass local congestion.
  - **Key Challenge**: Offloading increases latency. We must balance *load distribution* against *latency penalties*.

③ Solution – Load-aware Assignment

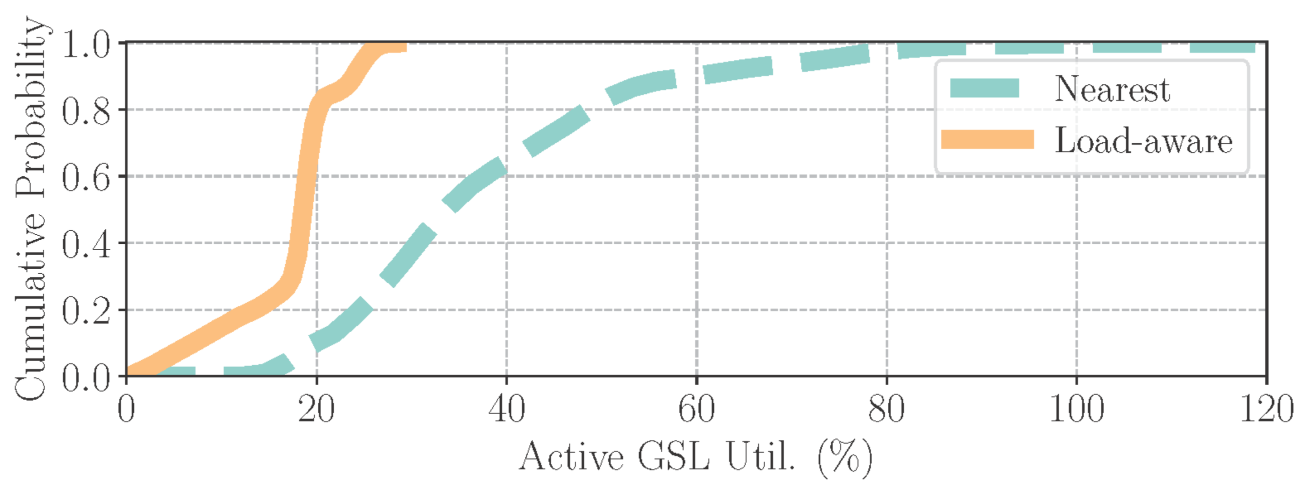
- ***Optimization Framework (MILP)***
  - **Objective**: MILP that minimizes peak GSL utilization across all GSs at every time slot.

$$
\begin{aligned}
\min_{x,\omega}\quad & \omega^t \\
\text{s.t.}\quad & x_{ij}^t \le \sum_{k=1}^{K} \left(\alpha_{ik}^t \cdot \beta_{jk}^t\right), && \forall i,j,t,\\
& \sum_{j=1}^{M} x_{ij}^t = 1, && \forall i,t,\\
& \sum_{j=1}^{M} \left(x_{ij}^t \cdot D_{ij}^t\right) \le \delta \cdot \Omega_i^t, && \forall i,t, \qquad (P_t)\\
& \sum_{i=1}^{N} \left(x_{ij}^t \cdot R_i^t\right) \le \omega_j^t \cdot Cap_j, && \forall j,t,\\
& \omega_j^t \le \omega^t, && \forall j,t,\\
& x_{ij}^t \in \{0,1\},\ \omega^t \in [0,1],
\end{aligned}
$$

- ***Critical Constraints***
  - **GS Assignment Constraint**:
    - A user–GS pair is feasible only if a path exists.
  - **User Request Constraint**:
    - Each user is assigned to exactly one GS per time slot.
  - **User Latency Constraint**:
    - Assignment respects a latency threshold: end-to-end delay ≤ α × baseline nearest-GS delay, capping the latency penalty.
  - **GSL Utilization Constraint**:
    - Aggregate user demand at each GS cannot exceed its downlink capacity.

④ Results

- ***Simulation Parameters***
  - **Constellation**: Modeled after Starlink Phase 1 with 1,584 satellites at 550 km altitude.
  - **Ground Segment**: 200+ GSs, 4 Gbps GSL capacity each.
  - **Traffic Model**: Demand scaled by real-world population density.
- ***Performance Analysis***
  - Baseline (*Nearest-GS*):
    - Shows a "long tail" distribution in GSL utilization.
    - A significant portion of GSs are severely overloaded, with utilization reaching nearly **120%**.
  - Proposed (*Load-aware*):
    - Balances traffic across the whole GS network.
    - The CDF curve is tightly clustered, with all active GSL utilization maintained **below 30%**.



CDF of GSL utilization across active GSs and time. Load-aware keeps every GS below 30%; Nearest-GS saturates near 120%.