SateRIoT: High-performance Ground-Space Networking for Rural IoT

Yidong Ren[1], Amalinda Gamage[3], Li Liu[1], Mo Li[2,3*]
Shigang Chen[4], Younsuk Dong[1], Zhichao Cao[1*]
[1]Michigan State University [2]Hong Kong University of Science and Technology
[3]Nanyang Technological University [4]University of Florida

ABSTRACT

Rural Internet of Things (IoT) systems connect sensors and actuators in remote areas, serving crucial roles in agriculture and environmental monitoring. Given the absence of networking infrastructure for backhaul in these regions, satellite IoT techniques offer a cost-effective solution for connectivity. However, current satellite IoT architectures often struggle to deliver high performance due to temporal and spatial link challenges. This paper presents SateRIoT, a new network architecture with temporal link estimation and spatial link sharing that fully exploits the capability of space low-cost low-earth-orbit (LEO) IoT satellites and ground low-power wide area (LPWA) IoT techniques in rural areas. First, we introduce a bursty link model that predicts the number of transmittable packets within a transmission window, reducing energy waste from failed uplink transmissions. Moreover, we enhance the model by selecting informative features and optimizing the window length. Additionally, we develop a multi-hop flooding protocol that enables gateways to buffer and share data packets across the network while incorporating a priority data queue to avoid duplicate transmissions. We implement SateRIoT with commercial-off-the-shelf (COTS) IoT satellite and LoRa radios, then evaluate its performance based on real deployment and real-world collected traces. The results show that SateRIoT can consume 3.3× less energy consumption for an individual gateway. Moreover, SateRIoT offers up to a 5.6× reduction in latency for a single packet and a 1.9× enhancement in throughput.

CCS CONCEPTS

• **Networks** → **Network reliability**; **Network dynamics**; **Network mobility**.

*Zhichao Cao and Mo Li are the co-corresponding authors.

ACM MobiCom '24, September 30–October 4, 2024, Washington D.C., DC, USA

ACM ISBN 979-8-4007-0489-5/24/09…$15.00
https://doi.org/10.1145/3636534.3690659

KEYWORDS

Rural IoT, LPWAN, Satellite Networks

ACM Reference Format:
Yidong Ren[1], Amalinda Gamage[3], Li Liu[1], Mo Li[2,3*], Shigang Chen[4], Younsuk Dong[1], Zhichao Cao[1]. 2024. SateRIoT: High-performance Ground-Space Networking for Rural IoT. In *International Conference On Mobile Computing And Networking (ACM MobiCom '24), September 30–October 4, 2024, Washington D.C., DC, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3636534.3690659

1 INTRODUCTION

The Internet of Things (IoT) is revolutionizing the way we understand and interact with the physical world, enhancing precise agriculture [15, 34, 40], environment monitoring [12, 24, 58], and forest fire prevention [17, 50] among others. Low-power wide-area (LPWA) IoT techniques (e.g., LoRa [1, 26, 49], NB-IoT [60], LTE-M [46])[1] are desirable to fit the scale of rural areas. Their networking architecture consists of gateways with backhaul links allowing internet access and many sensor nodes served by those gateways. However, the lack of urban-area networking infrastructures (e.g., 5G, LTE, wired networks) in rural areas presents a cost-benefit concern regarding the huge expenses of establishing new backhaul links.

Recently, LEO satellites [28, 57] demonstrate global Internet access for ground satellite radios with space links. A satellite radio sends its packets to LEO satellites over space up-links, and then these packets are forwarded to several territorial ground stations, which are Internet-connected, through space down-links. Exploring existing LEO satellites to enable direct-to-satellite (DtS) IoT is a cheaper alternative to establishing new backhaul links on the ground. For example, SpaceX's [47] subsidiary Swarm [53][2], a commercial-off-the-shelf (COTS) DtS-IoT provider, provides global connectivity for $5 per link per month.

SWARM uses ultra-small LEO satellites whose orbit altitude is either 462 kilometers or 510 kilometers. The connections between ground satellite radios and the satellites can be

[1]We implement our prototype with LoRa so use LoRa to represent general LPWA IoT techniques in the rest of the paper.
[2]We implement our prototype with SWARM so use SWARM to represent general DtS-IoT techniques in the rest of the paper.

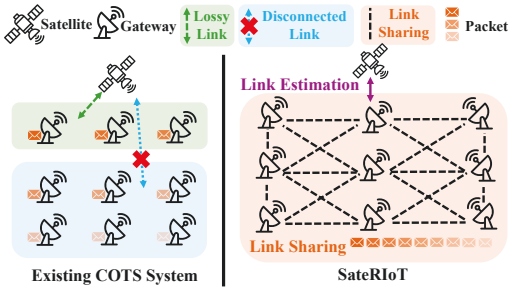


Figure 1: An illustration of link-aware SateRIoT with multi-hop data sharing.

affected by many factors [3, 10, 18, 19, 32], e.g., weather, satellite orientation, relative antenna position between the radios and the satellites, etc. According to our measurement study (§ 2.3), the available connections are sparse during a SWARM satellite passing by. In 80% cases, the total connection time is less than 20% of the pass duration. Moreover, the data rate of SWARM-M138 [52] radio is as low as 1 kbps. The sparse connection time and low link capacity motivate us to rethink the ground-space IoT architecture that can fully utilize the LEO IoT satellite network to achieve optimal IoT backhaul throughput while keeping efficient energy consumption and low end-to-end delay. As shown in the left part of Figure 1, existing works [8, 20, 31, 51] simply equip a satellite radio on each LoRa gateway so that the LoRa gateway forwards its collected sensory data to the Internet when a satellite passes by. However, this intuitive ground-space IoT architecture fails to achieve high performance due to the following temporal-spatial link challenges.

Challenge 1: Lossy up-links during short-window data pulling. When a gateway is connected to a satellite, the up-links could be lossy, degrading the energy efficiency of the gateway with a fixed number of packet transmissions. Specifically, deploying power infrastructure in remote areas such as farms and forests is challenging, complicating the powering of LoRa gateways [9, 27, 38]. The existing IoT gateway utilizes small solar panels for cost-effectiveness and ease of deployment. However, due to insufficient sunlight, these small panels become unsustainable in variable weather conditions such as rain, snow, cloudy days, and nighttime. Therefore, energy efficiency is important for satellite IoT gateways. The energy consumption of SWARM-M138 [52] radio reaches 7.97 mJ/bit in its Tx mode with 1 kbps, which is 408× larger than 0.0195 mJ/bit LoRa SF10 976 bps.

A SWARM satellite transmits data beacons to pull data from ground SWARM radios. When a SWARM radio receives this beacon, confirming that a connection is established, it performs 6 subsequent mandatory transmission attempts (i.e., a fixed number in SWARM protocol), irrespective of how many are received. However, in such a short-term transmission window, we observe that a SWARM radio can still fail to deliver its packets to the satellite despite having received the data beacon successfully. With the temporally lossy up-link, the enforced repeated transmissions cause significant degradation of LoRa gateways' energy efficiency.

Challenge 2: Diverse connection time among spatial LoRa gateways. Due to the spatial diversity of LoRa gateways, although they all experience sparse connections, we observe the connection time is quite diverse among them. In our measurement study, when a satellite passes over four LoRa gateways, different LoRa gateways can successfully transmit their data to the satellite at different times. In most cases, only one LoRa gateway is available at a time (§ 2.5). If a LoRa gateway only transmits the data received by itself, the LoRa gateway with sufficient connection time and transmission capabilities will keep idle when all its own data has been forwarded. Considering the buffered data on other gateways with insufficient connection time, the overall network throughput is lowered.

To address these challenges, in this paper, we propose SateRIoT, a new network architecture with temporal link estimation and spatial link sharing that fully exploits the capability of SWARM backhaul connections, thus enabling high-performance ground-space IoT for rural areas. We estimate the temporal link behavior to determine the optimal number of data packets to save energy and allow gateways to share their idle connection time with others to maximize network throughput and lower the end-to-end delay. As shown in the right part of Figure 1, besides self-generated data, each LoRa gateway buffers data shared from all other gateways. When the top-middle LoRa gateway finds that k packets should be transmitted in the current time window, it will immediately transmit k packets from either itself or others. The design of SateRIoT includes three key problems to guarantee high network performance while reducing extra computation and maintenance costs.

Firstly, the quality of a SWARM link is determined by many factors (e.g., atmosphere, satellite orbit) [3, 10, 18, 19, 32], creating a huge feature space that exposes the challenge of balancing between accuracy and agility. The accuracy cannot be guaranteed if we only use simple features like weather and satellite orbit information as inputs. If more link information is counted while a SWARM satellite is passing, we have limited time for the computation of link modeling and traffic scheduling. In SateRIoT, we choose to guarantee accuracy first while optimizing its agility. Specifically, we design a bursty link model to depict the temporal link behavior. We estimate the number of successful packet transmissions in an adaptive transmission window in which the link quality is consistent. To find the optimal trigger timing that balances between sufficient feature collection and limited computation time, we propose an Acknowledgement(ACK)-triggered lightweight ensemble learning model.

Second, selecting proper model parameters is not trivial to achieve optimal performance in practice. On one hand, we need to determine a small group of specific features from many available ones as the model inputs that derive accurate link estimation. Meanwhile, we need to define an appropriate time length for each transmission window adaptively. Due to the lossy link nature, a long transmission window makes it hard to keep link consistency, degrading link estimation accuracy. However, due to link sharing among multiple LoRa gateways, short transmission windows lead to frequent model computation and network data status synchronization, causing increased energy consumption. Additionally, to ensure a transmission window can be configured in practice, it must align with the timing of the SWARM data-pulling window. To address the issues, SateRIoT proposes meticulously utilizing features from the physical layer, COTS satellite IoT protocols, and environmental information. For the optimal window length, we aim to find the longest window with consistent link quality to save energy. We adopt an empirical method to identify the optimal configuration.

Third, in link sharing, a LoRa gateway needs to collect data from others. However, when and how to collect the data is challenging. Since the connection time is sparse, after a LoRa gateway sets up a transmission window, we have no time to use energy-efficient tree-based ad-hoc routing protocols for packet pulling from others [16, 59]. Since the energy consumption of a LoRa radio is far less than that of the SWARM satellite radio, we enable link sharing with a multi-hop flooding protocol, which utilizes broadcast to achieve network-wide data consistency quickly. Specifically, when a gateway collects new data from its sensor nodes, it initiates a flooding process, where other gateways forward the packet until all receive a copy. In this way, each LoRa gateway collects up-to-date data from others in a timely manner. In addition, carrier sense backoff is adopted to prevent packet collisions during concurrent LoRa flooding, improving network consistency and energy efficiency. On the other hand, it is possible for gateways to establish their transmission windows simultaneously. Although collisions can be mitigated by the LR-FHSS modulation [35, 61] in SWARM, duplicate packets may still occur in link sharing. To prevent duplicates, we prioritize self-generated data and shuffle relayed data. We also design beacons to lock transmitting packets and synchronize queues with the flooding service.

We implement SateRIoT on four gateways embedded with COTS SWARM M138 [52] and LoRa SX1262 [42] radio chips. We deploy these four gateways in several university farms that cover 9 km^2 areas to evaluate the performance of SateRIoT. Moreover, we have conducted extensive trace-driven emulation experiments for a large network scale with 12 gateways. Our results indicate SateRIoT delivers comparable throughput using up to 3.3× less energy for each gateway. Additionally, SateRIoT reduces packet delay by up to 5.6× and boosts throughput by 1.9×. Our contribution can be summarized as follows:

- We empirically measure the characteristics of LEO IoT satellite links with real-world deployment and observe three insights of temporal-spatial link behavior, which demonstrate the barriers of directly adopting the DtS-IoT to achieve high-performance rural IoT.
- We design SateRIoT, a new networking architecture consisting of well-designed satellite link estimation and sharing modules to enable network-wide data sharing and link-aware data transmission, enhancing network performance.
- We fully implement the proposed SateRIoT design with COTS LoRa gateways and SWARM radios. We extensively evaluate its efficiency with real-world deployment and trace-driven emulation. The results demonstrate that SateRIoT achieves equivalent throughput with 3.3× less energy for individual satellite devices. Additionally, our scheduling scheme reduces packet delay by up to 5.6× and enhances throughput by 1.9×.

2 PRELIMINARY AND OBSERVATION

2.1 LEO Satellite for IoT

LEO satellites operate between 500 km and 2,000 km above Earth, offering advantages such as shorter latency and more frequent revisits compared to satellites in higher orbits. LEO satellite communication can provide high-speed broadband internet access globally, like Starlink [48] and Oneweb [33]. Additionally, LEO satellites can also facilitate low-cost IoT in rural areas with limited network infrastructure.

Satellite IoT has become an emerging field with selective commercial corporations striving to provide global IoT connectivity without resorting to expensive high-data-rate satellite communication systems. These entities have launched satellites to form extensive constellations, enabling the use of ground-based small satellite radios equipped with advanced long-range communication capabilities. For example, Sateliot [41] proposes to integrate NB-IoT for satellite communications into both current and forthcoming 5G infrastructures. Lacuna Space [22] delivers an uplink service utilizing a LoRa-inspired physical layer. SpaceX's IoT solution [47] SWARM Technologies [53] collaborates with Semtech [43] to employ long-range frequency hopping spread spectrum (LR-FHSS) modulation within the VHF bands. This operates with an uplink range of 148-150 MHz and a downlink range of 137-138MHz, aiming for exceptionally cost-effective satellite communication at a 60 USD data fee per year. In this study, we use SWARM satellite radio to evaluate the efficacy of COTS satellite IoT systems and subsequently craft our design based on these insights. SWARM with 168 satellites

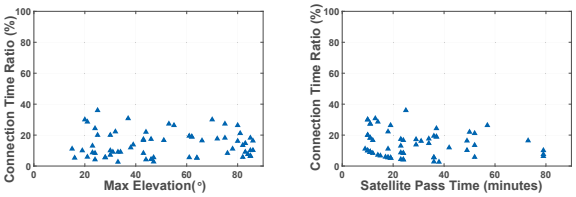


Figure 2: Connection time ratio.

differs quite from traditional high-data-rate LEO satellite networks (e.g., Starlink with more than 10000 satellites) in terms of protocols, satellite size, and constellation scale. SWARM links exhibit unique characteristics such as a more sparse satellite distribution, shorter connection times, lower frequency usage, and reduced network costs [3, 53].

2.2 Measurement Experiments

To evaluate the network performance of the COTS satellite IoT system, we carried out outdoor measurement experiments on a rural farm. We employed 4 SWARM Eva Kits [54], each fitted with an M138 modem [52], and powered them using 18-24V DC solar panels. Over a span of 120 hours, we gathered data from 12 unique locations within a large farm, collecting approximately 3,000 packets and spanning an area of 9km^2 in rural farmland. We use infinite traffic to ensure we can capture the link status of every satellite passing by.

2.3 Up-link Connection Time

The propagation environment for SWARM satellites operating in the VHF bands introduces unique challenges associated with the temporal and spatial link variability [3]. For instance, atmospheric elements such as precipitation rates and cloud cover can cause time-dependent attenuation. This attenuation can vary over brief periods, making a substantial impact on link quality. Meteorological changes, especially seasonal variations, exacerbate these dynamics [19]. Due to unpredictable conditions, a fixed satellite link pattern cannot be relied upon. On the other hand, the link's spatial variability is attributed to differences in terrain and topography. These geographical features can result in shadowing and multipath effects, which differ based on the elevation angles and the terrain's contours [18, 19]. Additionally, satellite orientation is critical, especially in low Earth orbit, due to its high angle sensitivity. As a satellite progresses along its orbit, its relative positioning with respect to the ground satellite radio also changes. The elevation angle, defined as the angle between the ground's horizontal plane and the satellite's line-of-sight, influences the atmospheric path length and, subsequently, the link attenuation [21, 57].

Observation 1: Disconnected links During a SWARM satellite passes by, we divide the passing duration into several time windows with 1 minute. Given a time window, if a packet is successfully transmitted, we mark this time window as a connected time window. We define the connection time ratio as equal to the ratio between total connected time windows and total time windows in the passing duration to quantify the link disconnected property. According to the collected data traces, Figures 2 shows the results based on different maximum elevation angles and satellite passing duration. The results show that the connection time ratio remains consistently low. This trend holds irrespective of the maximum elevation or the duration of satellite passage. The average and median values are 14.1% and 11.9%, respectively. The highest observed ratio doesn't exceed 40%, and in 80% of the satellite passes, this ratio is below 20%. Such findings indicate the satellite links are disconnected, increasing the transmission latency.

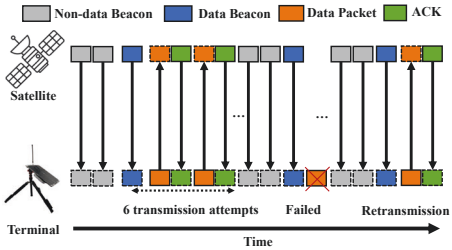


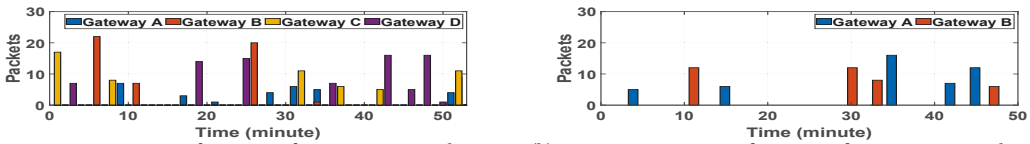
Figure 3: COTS SWARM IoT Communication Protocol.

2.4 SWARM Communication Protocol

Figure 3 depicts the communication protocol within the Swarm satellite IoT system. There are 3 basic packet types: satellite **non-data/data beacon**, **data packet**, and **ACK** (acknowledgment). Dotted edges represent the reception side and solid edges represent the transmission side. As a satellite coverage period begins, all the satellite radios awaken from their sleep mode. These satellite radios then start monitoring space, anticipating the satellite beacon. The received satellite beacon is the unsolicited message from the overhead satellite that can be classified into two types: non-data and data beacon. The data beacon informs satellite radios of the satellite's presence and its readiness to accept messages from the ground. Receiving the satellite data beacon is a prerequisite for initiating data packet transmission to the satellite. Non-data beacons are used for downlink traffic and will not trigger the satellite radio to transmit data. Multiple non-data beacons appear before data beacons. Both beacon types include information on the Received Signal Strength Indicator (RSSI), Signal-to-noise ratio (SNR), frequency, timestamp, and satellite ID.

According to the collected data traces, during times without transmissions, the non-data beacons were sent every 60 s. During each data transmission period, the satellite will increase the frequency of non-data beacons at the beginning and send a data beacon afterward. The time offset between two data beacons uniformly varies from 20 s to 52 s. All the data is transmitted between two data beacons. The median and average time gap between two adjacent data beacons during gateway transmission is 33 s and 35.2 s respectively.

When data packets are successfully received by the satellite, an ACK is promptly sent back to the satellite radio, as



(a) Data transmission performance of 4 gateways in each minute. (b) Data transmission performance of 2 gateways in each minute.

Figure 4: Comparisons of transmitted packet number at each minute between two gateways at different distances.

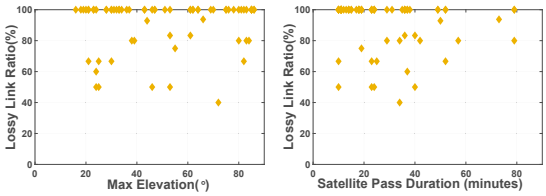


Figure 5: Lossy link ratio.

illustrated by the green block in Figure 3. The ACK signifies a successful data packet transmission. It includes RSSI, SNR, the sent packet ID, and frequency information. All packets delivered to the cloud correspond with acknowledgments received at satellite radio. If an ACK is not detected within a receiving window, it indicates a failed packet delivery. There are two main reasons why a packet may fail to reach the satellite: The first is when satellite radio chooses not to transmit data because of a high background noise level with RSSI exceeding -88 dBm. The second is when satellite radio transmits packets, but the satellite cannot detect them, either due to poor signal quality or interference.

Observation 2: Lossy Links. Satellite radio has a transmission buffer with a queue system called *satellite radio queue*, and messages are generally dispatched based on their entry sequence into this queue. Any unsuccessful packets are then re-queued for another attempt in the next transmission window accompanied by the following data beacon. In practice, the energy efficiency of satellite radio transmission performance is often degraded by failed transmissions. This inefficiency is further exacerbated by retransmissions, particularly when the communication channel deteriorates. "Lossy link" indicates the link condition when a downlink data beacon can be received while the uplink data packets may suffer loss in the following transmission windows. Using a software-defined radio, we detected 6 signal pulses on the uplink spectrum. COTS SWARM fixes the uplink transmission of 6 consecutive packets as the default setting to match its temporal dynamics. COTS radios set this fixed value because more attempts may result in high chance of transmission failures while fewer attempts might waste good transmission opportunities.

During the same satellite pass, the lossy link ratio is defined as the proportion of 1-minute transmission windows experiencing lossy links to the total number of transmission windows during the pass duration. According to our collected data trace, Figure 5 illustrates the lossy links ratio under different maximum elevation and pass duration settings. We can observe that 71% satellite radios experience 100% lossy links ratio, leading to significant energy wastage, especially in satellite communication radios with high energy consumption. This motivates us to predict transmission capability to reduce energy waste.

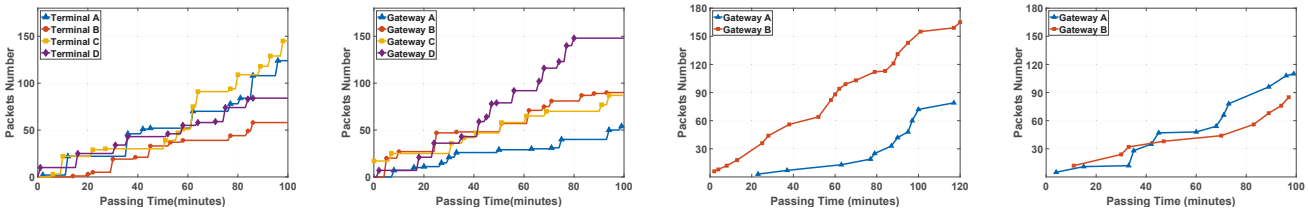
2.5 Geo-spatial Distributed Radios

As mentioned in Section 2.3, too many factors can influence the satellite communication link. The link of different geospatial locations can be variable and dynamic. To better understand the link dynamics across multiple locations, we compared the performance of multiple satellite radios operating concurrently to measure the maximum transmission capability as illustrated in Figure 6 and Figure 4.

Observation 3: Dynamic spatial and temporal links. We analyzed the performance during each minute to further understand the packet transmission dynamics, as illustrated in Figure 4. The bars represent the number of packets successfully sent to the satellite during a 1-minute time window. In Figure 4(a), the satellite radios of 4 gateways transmit packets alternately over time, with primarily non-overlapping periods. In most cases, only one gateway can connect to the satellite for one minute. Figure 4(b) further reveals that the two gateways' sending periods do not overlap at all, occurring at different times separately. This suggests that transmission capabilities vary by location and time.

Moreover, we analyzed the overall performance of 4 gateways arranged in a square formation, where each gateway was equidistant from its neighbors by either 500 m or 800 m with the same satellite pass duration. The slopes change over time for 4 gateways, demonstrating the temporal diversity. In Figure 6(a), A, C, D exhibit varying leadership in overall transmission rates. Notably, B and D experience minimal data transmission for about 15 minutes, and C surges ahead after a 62-minute delay, maintaining its lead thereafter. As the observation period progresses, the performance gap among A, D, and B significantly increases, with A and B achieving 1.9× and 1.5× the packets of D. Conversely, Figure 6(b) shows D's progressive increase, with B and C alternatively leading to a similar trend, ultimately achieving comparable throughput. A maintains low data rates, managing only 54 packets over 100 minutes, 2.8× less than D, which significantly falls short of transmission requirements.

We also compare the performance of two terminals with 2300 m distance. Figure 6(c) illustrates that A begins its data transmission 20 minutes after B. And B progressively extends its lead over time. As a result, by the end of a 120-minute period, B has achieved approximately 2× the throughput of A. Conversely, Figure 6(d) presents a scenario where the



(a) 4 gateways with 500 m distance (b) 4 gateways with 800 m distance (c) 4 gateways with 2300 m distance (d) 4 gateways with 2300 m distance

Figure 6: Comparisons of throughput with the same satellite pass among multiple gateways at different locations.

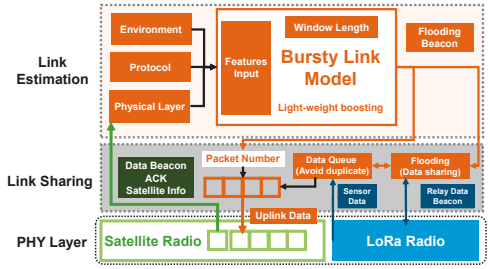


Figure 7: The network architecture in SateRIoT.

performance disparity between the gateways narrows over the same distance. Although B maintains its lead initially, A catches up by the 43-minute, matching B's cumulative data transmission. Subsequently, A overtakes B and sustains this advantage up to the 100-minute, achieving 1.3× throughput compared to B. Collectively, these illustrations underscore the dynamic performances of satellite radio under varying temporal and spatial scenarios. Temporal and spatial link dynamic creates varying transmission capabilities among satellite radios. Some satellite radio may struggle to send packets, while others have surplus capacity after completing the delivery of packets within their own covered ground area. Thus, the network throughput is degraded.

3 SYSTEM DESIGN

3.1 System In A Nutshell

In SateRIoT, a LoRa gateway is equipped with a SWARM radio to access the Internet through SWARM satellites. On the other hand, it can use its LoRa radio to communicate with other LoRa gateways and collect sensory data from sensor nodes. All LoRa gateways operate in a distributed manner. Figure 7 illustrates the network architecture design of SateRIoT, which consists of three layers: link estimation, link sharing, and physical layer communication radio.

In the link estimation layer, the core design is the **bursty link model** (§ 3.2). First, we build a *lightweight model structure* (§ 3.2.1). The temporal burstiness of a link determines the short stable duration of the temporal links for data packet transmission. Based on our understanding of SWARM protocols, we design an ACK-triggered scheme to estimate the link, balancing model accuracy and agility to make it practical for real-time link estimation. Second, *features* (§ 3.2.2) are selected to be the most informative for accurate temporal link estimation. These features include information from the physical layer, environmental characteristics, and COTS satellite protocols. Third, we empirically select the longest *window length* (§ 3.2.3) to maintain consistent link quality while keeping the computation and system overhead low.

To combine the capabilities of multi-gateways, we design **link sharing** module to improve the overall network performance by exploring link spatial diversity. *the multi-hop flooding protocol* (§ 3.3.1) uses the LoRa radio to enable efficient network-wide data sharing of sensory data or flooding beacon. *Priority data queue* (§ 3.3.2) avoid duplicate transmission by setting priority order among data packets. The data packets in the priority data queue are managed by both the link model and the flooding protocol.

3.2 Bursty Link Modeling and Estimation

According to the observation of temporal lossy link, it suggests the up-link is short-term bursty, which means the link behavior (e.g., success or failure) is only stable in a varied short-time window. We design a bursty link model that estimates how many packets can be successfully transmitted in a stable transmission window. To guarantee the current transmission window is bursty for successful transmission and collect sufficient information to estimate the number of successful transmissions, instead of using satellite beacons to trigger a link estimation, we trigger the link estimation after an ACK is received, indicating a packet has been received at the satellite side. Moreover, we can collect additional critical features from the packet ACK.

Specifically, based on our observation in §2.4, a gateway attempts data packet transmission once it has received a data beacon from a passing-by satellite. To estimate the current link quality, the gateway only adds a probe packet to its satellite radio queue. Once the gateway receives the acknowledgment of the probe packet from a satellite. We initiate a link estimation process to determine how many data packets should be set as pending status in a transmission window, whose time length is predetermined as T_{tx}, including multiple data beacons. From our measurement study, both the PDR and SNR remain stable across these consecutive data beacons. We leverage the short-term stability to perform link estimation for a transmission window to lower the link estimation overhead. When all pending data packets are transmitted once or the transmission window expires, regardless of whether all ACKs are received, we dequeue the

left data packets from the satellite radio queue and insert a probe packet into the satellite radio queue again to trigger the link estimation of the next transmission window.

3.2.1 Light-weight model structure. To predict the versatile link within a given transmission window, we use boosting model architecture to build an effective, lightweight machine-learning algorithm for link estimation and scheduled packet prediction. Figure 8 illustrates our link estimation model architecture and transmission capability prediction. We utilize the four kinds of features as the input X and try to predict 1 to k implement classifications where $k = 6 \times w$, w represents data beacon amount given the duration of the current transmission window T_{tx}. The model $F_0(x)$ is initialized to the logarithm of class priors, i.e., for all x, $F_0(x) = [\log p_1, \log p_2, \ldots, \log p_N]^T$, where p_k is the proportion of class k in the training set. The tree index in our boosting model is t from 1 to N:

(1) For each class k, compute the pseudo-residuals:

$$r_{ik}^t = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F_k(x_i)}\right]_{F(x)=F_{t-1}(x)}$$

L is the multi-class logarithmic loss. For a dataset of N samples, the total multi-class logarithmic loss is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log(p_{ik})$$

y_{ik} indicates whether sample i is in class k (1 if true, 0 otherwise). p_{ik} is the model's predicted probability that sample i belongs to class k.

(2) For each class k, fit a new tree $f_{tk}(x)$ with r_{ik}^t.

$$f_{tk}(x) = \arg\min_f \sum_{i=1}^{M} L\left(r_{ik}^t, f(x_i)\right)$$

(3) Update the model: For each class k, find the coefficient γ_{tk} that minimizes the overall loss, and update

$$\gamma_t k = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, F_{t-1}(x_i) + \gamma f_t(x_i))$$

$$F_N(x) = F_0(x) + \sum_{t=1}^{N}\sum_{k=1}^{6\times w}\gamma_{tk} f_{tk}(x)$$

During the training process, we adjust data or feature sample ratios to randomly selectively use data and features for each tree rather than employing the entire set. This random feature and data diversity makes the final model more robust, enabling it to better adapt to various data distributions. We can effectively reduce the risk of overfitting the training data, thereby enhancing the model's generalization ability. Our model balances uneven real data sets by giving more weight to underrepresented categories during training. We also use random sampling to oversample/desample the data with different labels and balance the class distribution.

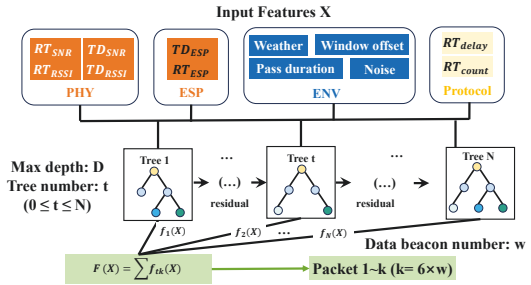


Figure 8: Boosting based link estimation model.

3.2.2 Link features. The input of link estimation model.

Physical layer features: The successful transmission of the probe packet indicates a satellite data beacon and an acknowledgement have been received by the gateway. RT_{rssi} and RT_{snr} indicate the RSSI and SNR of the received satellite data beacon. TD_{rssi} and TD_{snr} indicate RSSI and SNR of the acknowledgment of the probe packet. The RSSI and SNR values represent the physical propagation property of the current down-link from the satellite to the gateway.

Expected Signal Power (ESP) feature: RSSI and SNR can be distorted by environmental noises. To focus on satellite signals, we combine RSSI and SNR to generate an ESP [7, 39] value, which indicates the signal attenuation along the propagation path. The ESP feature is calculated as follows:

$$\text{ESP} = \text{RSSI} + \text{SNR} - 10\log_{10}(1 + 10^{0.1\cdot\text{SNR}}) \quad (1)$$

Then we can get RT_{esp} and TD_{esp} with physical layer features.

Environmental features: We include several environmental features as follows: *Noise:* RSSI of background noise. This is measured by SWARM radio when no satellite appears. *Elevation*: The maximum elevation during a satellite pass can be found on the SWARM website [53] given the location of the gateway. *Passing duration*: The passing duration of a satellite is announced on SWARM website [53] according to the location of the gateway. *Transmission Window Offset*: The time offset between the current estimated transmission window and the time that the satellite starts to pass the area. The relative antenna position between the gateway and the satellite is different at different time offsets. *Weather*: Based on the weather released by local weather station, we use four-levels quantization to define the weather values from sunny to drizzle.

Protocol features: RT_{delay}*:* The time delay between the received data beacon and acknowledgement of the probe packet. RT_{count}*:* The number of satellite non-data beacons during a 30-seconds period before the acknowledgement of the probe packet.

3.2.3 Link window length. The predicted window length of our link estimation model is significant. Predicting long windows can lead to low link estimation accuracy since only 2% of the minutes have data transmission modes lasting more than 2 minutes. Consequently, selecting overly long observation windows can significantly degrade link estimation

accuracy. Conversely, shorter windows necessitate more frequent transmission of LoRa beacons, which increases both energy consumption and computational overhead. Therefore, optimizing the window length is critical to achieving a balance between estimation accuracy and system efficiency. To maintain consistency and continuity with the transmission patterns, the window length must align with the natural data transmission cycles. Higher link quality and more reliable data transmission are indicated by consecutive data beacons. Inappropriate window lengths can disrupt these continuous transmission periods, resulting in a marked decrease in estimation accuracy. Thus, careful selection and tuning of window lengths are essential for improving link quality assessments and ensuring efficient network operation.

3.3 Link Sharing

3.3.1 Multi-hop Packet Flooding Protocol. The multi-hop flooding protocol provides a primitive networking method to enable network-wide packet sharing. The basic idea is that a LoRa gateway initiates a flooding process once it has a packet to share. There are two types of packets: data packets and flooding beacons, with the latter used to manage the data queue. A LoRa gateway immediately forwards the packet once it receives one from another LoRa gateway. A flooding packet starts from the original gateway. In the first round, the packet will be delivered to the closest gateways. Next, these gateways will keep relaying this packet to their next-hop adjacent gateways. Eventually, the packet can be delivered to all gateways [2, 30]. For IoT data collection systems in rural areas, flooding protocol is easy to deploy and implement in the low-cost IoT gateways without coordination overhead, avoiding complex network traffic patterns.

Carrier-sense based Collision Avoidance. It is possible that several LoRa gateways initialize multiple flooding processes in a short period. Additionally, several LoRa gateways may receive the same packet and start to forward simultaneously. Without noticing others' packet transmission, the potential packet collision could degrade the reliability of the data sharing. It will be worse in dense network deployment with larger flooding scale. To solve this, We adopt channel activity detector (CAD) on COTS LoRa radio chips [14] to enable low-cost carrier-sense-based collision avoidance. Before a LoRa gateway forwards a received packet, it will wait for a random initial backoff, then repeat carrier sense until the channel is clean.

Flooding Beacons. The link model of a gateway triggers two types of beacons to synchronize the data packets among all gateways with the flooding protocol. Firstly, when the gateway determines which packets will be transmitted in a transmission window, it sends out a beacon including information on these packets. Secondly, when the transmission window ends, the gateway sends a beacon indicating which packets have been successfully transmitted.

Network Consistency. In case a LoRa gateway misses a packet from other gateways due to LoRa link dynamics [39]. Each LoRa gateway maintains the status of its local buffered packets and broadcasts the status with the schedule of a Trickle timer [25]. If a gateway receives others' status and finds an inconsistency with its local status, it will request other gateways to send the missing packet. The trickle timer will be reset when a gateway receives a request. This strategy offers both a systematic approach to maintaining information consistency and a proactive method to resolve potential mismatches in a distributed environment. In this way, all LoRa gateways consistently buffer all packets and receive the flooding beacons from others for later global packet transmission scheduling.

3.3.2 Priority Data Queue Management. **Priority and data enqueue:** The priority data queue structure includes a self-generated data queue and a relay data queue. The self-generated data queue has higher priority than the relay data queue. Since self-generated data packets are unique across different gateways, duplicate packets will be prohibited when multiple gateways simultaneously send data packets to a satellite with non-collided frequency hopping. We do order shuffling for the relay data queue. Namely, when a gateways received a relay data packet, it will insert the data packet to a random position in the relay data queue. In this way, when multiple gateways concurrently transmit data packet, the number of duplicate packets will be further reduced since they will forward the relayed data in different order.

Packet holding and releasing When a gateway receives a flooding beacon from others indicating the data packets they will transmit in the coming transmission window, the gateway will check its queue. For those identical data packets, the gateway will hold them for a holding period that equals two transmission windows $2T_{tx}$. The holding packets have no chance scheduled if a transmission window is coming. In the holding period, if the gateway receives another flooding beacon (§3.3.1) indicating the holding packets are successfully transmitted, it will dequeue them directly. When the holding period expires, the gateway will release them. In this way, we actively reduce potential duplicate transmissions. For each gateway, when a transmission window ends, it will dequeue packets whose ACKs have been received.

4 IMPLEMENTATION

Our methods are fully compatible with LoRaWAN and COTS satellite devices without additional hardware or centralized coordination. The data for the SWARMIoT only costs 5 USD each month [53], much less than the cost required for network infrastructure construction in rural areas.

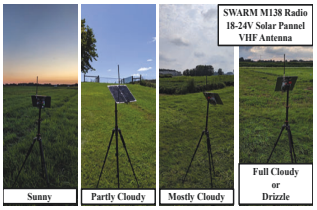


Figure 9: Outdoor deployment of satellite IoT radio with different weather.

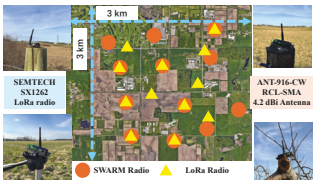


Figure 10: Outdoor deployment of flooding experiments and satellite radio locations.

Ourdoor SWARM satellite radio deployment: We employed 4 Swarm Eva Kits [54], each fitted with an M138 Modem [52], and powered them using 18-24V DC solar panels. Figure 9 shows the outdoor Swarm ground gateway deployment scenarios to collect data in four different weather conditions (e.g., sunny, partly cloudy, mostly cloudy, fully cloudy or drizzle, rain). Two similar weather conditions can coexist during the same time period in a close area. The locations of SWARM devices are shown as orange circles in Figure 10 around 9 km^2 rural area farmland. For the satellite radio, the length of a packet is 192 bytes that can contain multiple sensory data packets.

Outdoor LoRa radio deployment: Figure 10 depicts the outdoor deployment of LoRa radios in 9 km^2 farm zones. The yellow triangles represent locations. We deploy LoRa devices in elevated positions with open space, such as high fences, small trees, or on tripods at heights ranging from 1.5 m to 2.5 m. We use 4.2 dBi gain LPWA antennas [6] and 12 SX1262 [42] radios controlled by ESP32 MCUs [11] operating on US915 ISM bands with SF12 and 125kHz bandwidth. We uses LMAC-1 [14] to enable CSMA for LoRa transmission.

Link estimation model: The predicted transmission window is set to 120 s when $w = 4$, which is the maximum duration observed with continuous data beacons with consistent link quality (§ 3.2.3). The model generates 330 trees, each with a maximum depth of 5. We use the objective function 'multi:softprob' to show each class's probability distribution and the Mean Squared Error (MSE) as the loss function. We design our model based on XGBoost [5] and SMOTE [4]. The inference time per sample is 0.373 milliseconds on a Raspberry Pi 4 Model B [36]. Given the lightweight nature of the model, it is particularly well-suited for deployment on resource-constrained devices. We can anticipate exceptionally rapid inference times when the model is deployed at LoRa gateways.

5 EVALUATION

Performance metrics: To evaluate the overall performance for uplink transmission, we focus on the cumulative number of delivered packets as **Throughput** and the time from packet generation to successfully delivered for each packet as **Latency**. For the link estimation model's performance, we rely on **Accuracy** and **Mean Absolute Error (MAE)** as metrics. We measure the performance of individual gateway using **Energy Efficiency**, represented as actual transmission attempts for each packet. For multi-hop flooding protocol performance, we employ **Latency** as the key indicator to present the packet delay among multiple gateways.

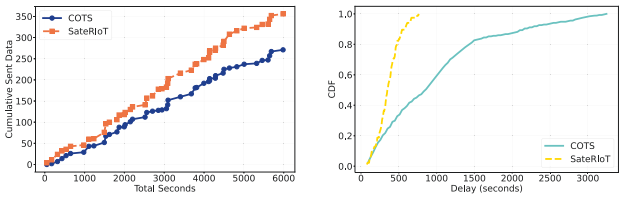
Baseline method: 1. COTS: We use the existing COTS protocol and network architecture of Direct-to-satellite IoT devices as our baseline. Each gateway directly sends all the packets from their satellite communication radio's transmission buffer without any link estimation or link sharing. This is evaluated and compared in energy efficiency evaluation (§5.2.2) and uplink data transmission (§5.1). **2. LDB** link estimation model: For link prediction, we only use the latest data beacon (LDB) to guarantee agility at first. The input features include RSSI, SNR, ESP of data beacon, noise, weather, satellite elevation, pass duration, and transmission window offset. **3. ENV** link estimation model: We build a linear regression model to predict the total delivered data amount during one satellite pass duration with input from weather, satellite elevation, pass duration and location index. We use the ENV model to compare energy efficiency performance (§5.2.2). **4. SateRIoT-w: We use different link window lengths as our baselines. We select the w (as mentioned in Section 3.2.2) values of 1,2,3,5 data beacon numbers to compare the different SateRIoT scheme.**

Default settings: The default packet generation frequency of each gateway is 1 packet per minute. The packet includes multiple sensory data frames from its covered area. The default weather value is 2, partly cloudy.

5.1 Overall Performance

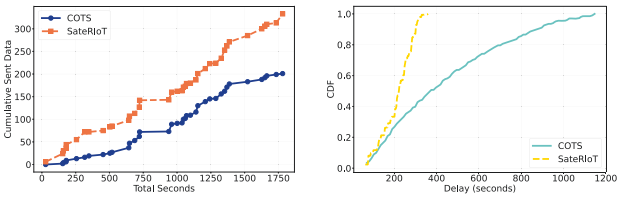
In this section, we evaluate the overall performance in throughput and latency with SateRIoT and COTS protocol. The delay and possible duplicate caused by link sharing are considered in calculating the latency and throughput according to the results from §5.3. Experiment A uses real data collected simultaneously from 4 gateways in different outdoor locations. Experiment B employs trace-driven datasets from 12 gateways to conduct emulation, involving 4 gateways transmitting concurrently during 3 different time periods with similar satellite orbits and weather conditions. Experiment C simulates 12 gateways, with ground-truth data generated by our accurate link estimation model.

A. Real datasets experiments We use data collected outdoors over 100 minutes from 4 gateways in §2.5 as ground-truths. We use our link estimation model and real collected information as input to predict the link with real and emulate the throughput and latency. Figure 11(a) depicts the throughput variations over time for all four gateways. As time progresses, the throughput disparity between SateRIoT and the COTS method widens. After 100 minutes (6000 seconds),



(a) Real Datasets Throughput (b) Real Datasets Latency

Figure 11: Performance in real data of 4 gateways



(a) Throughput of trace-driven data (b) Latency of trace-driven data

Figure 12: Performance of trace-driven 12 gateways.

SateRIoT can achieve 1.31× cumulative sent data packet compared to COTS protocol. Additionally, Figure 11(a) details the latency from packet generated time to delivery time. It is obvious that SateRIoT offers substantially shorter latency compared to the COTS approach. Specifically, around 80% of the packets can be transmitted in less than 8 minutes, compared to the COTS method, which requires up to 24 minutes to achieve the same level of packet transmission. This results in SateRIoT being 3× faster. All the packet in SateRIoT reaches a delay of less than 750 s while COTS extends to 3250 s. This indicates that the maximal latency of COTS can be up to 4.3× longer than SateRIoT. These findings affirm the superiority of SateRIoT in terms of throughput and latency in real-world scenarios. The overall performance can be further improved when more gateways join the uplink transmission.

B. Trace-driven experimental settings: To evaluate more gateways performance with SateRIoT, We enlarge the gateways number to 12. We use 4 satellite radio devices as a group. Each deployed 3 times to emulate scenarios where gateways in 12 locations attempt to schedule data transmissions during the same satellite pass. We select 3 satellite traces that were closely matched in terms of elevation and pass duration. The maximum elevations and pass duration are 71° with 31 minutes, 67° with 29 minutes, and 69∘ with 30 minutes, respectively. In each satellite trace, we deploy 4 gateways to collect real data sets from 4 distinct locations. The locations of the gateways vary across the 3 satellite passes, ensuring coverage of 12 unique locations in total. This allows us to effectively emulate a scenario where 12 gateways concurrently connect to the same satellite.

Results: Figure 12(a) shows that SateRIoT transmits 332 packets, which is 1.65× more data than the COTS baseline. Over time the data transmission gap between them widens because the baseline focuses on individual transmissions, missing the opportunity to use the connection time fully. SateRIoT efficiently schedules data and controls traffic, maximizing the use of the transmission window across all gateways with a reliable connection. Figure 12(b) reveals that SateRIoT delivers all packets to the satellite within 357s, while the COTS protocol takes up to 1147s. Therefore, SateRIoT is up to 3.4× faster, sending 80% of its packets in 280s compared to 700s for the COTS.

C. Generative datasets: Data availability for situations where elevation, pass duration, and weather conditions are extremely similar is severely restricted. In order to broaden the scope of our emulation experiments to encompass a wider range of satellite orbits and weather conditions, we deployed 2 or 4 satellite radios at 12 different locations, each tracking multiple satellite passes. This allows us to generate datasets for simulating uplink data transmission among multiple satellite radios during the same satellite pass. Figure 13 provides a clear overview of the interrelationships within the original features directly recorded from collected real datasets. We can easily observe that RT_{rssi}, RT_{snr}, TD_{rssi}, TD_{snr} exhibit strong positive correlations, while RT_{delay} and RT_{count} display notable negative correlations. Additionally, pass duration, transmission window offset, and weather also demonstrate noteworthy positive correlations with each other.

Based on the above observations, we establish a process to determine feature values using their inherent physical significance and a correlation heatmap. Initially, we randomly pick the maximum elevation from real datasets and then choose a pass duration that aligns with this elevation. We randomly generate varying frequencies of connection time windows and set start times from collected real datasets, considering the same maximum elevation and pass duration. Second, We randomly select a value for RT_{snr} from the real datasets. Then we randomly choose a TD_{snr} from the subset of real datasets that match the chosen RT_{snr}. Following this rule, based on determined RT_{snr} and TD_{snr}, we randomly select RT_{rssi} and TD_{rssi} values. Third, we randomly chose RT_{delay} from collected datasets, followed by a random RT_{count} based on consistent RT_{delay}. In addition, we randomly set background noise RSSI from -106 dB to -87 dB ranging as real noise level.

Generative datasets experiments: We conduct further experiments to validate performance in various satellite scenarios with the generative datasets. We set the maximum satellite elevations to 81°, 66° and 28° with corresponding pass durations of 45, 30, and 14 minutes at 12 different locations. Figure 14(a), Figure 14(b), and Figure 14(c) demonstrate that SateRIoT can achieve 1.6×, 1.7× and 1.9× the data amount of baseline during one satellite pass. the data volume of the baseline within a single satellite pass, respectively. Meanwhile, Figure 16(a), Figure 16(b), and Figure 16(c) highlight that SateRIoT can achieve packet delivery latency that

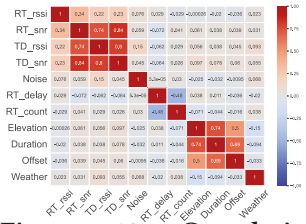
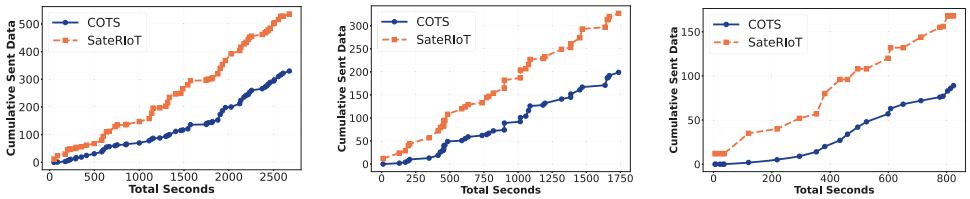


Figure 13: Correlation heatmap of input features.



(a) Throughput of 12 gateways with max elevation 81° in 45 min. (b) Throughput of 12 gateways with max elevation 66° in 30 min. (c) Throughput of 12 gateways with max elevation 28° in 14 min.

Figure 14: The throughput performance of generative data sets.

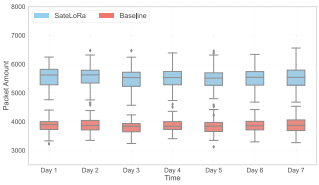
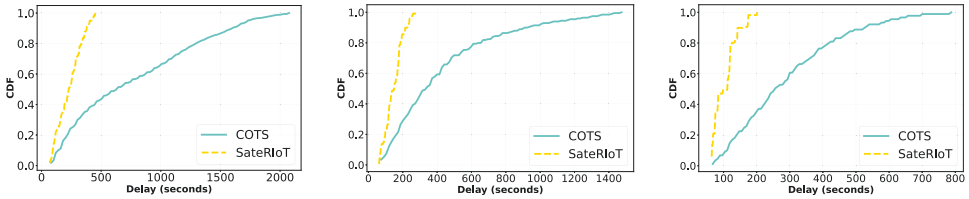


Figure 15: Throughput of large-scale simulation.



(a) Latency of 12 gateways with max elevation 81° in 45 min. (b) Latency of 12 gateways with max elevation 66° in 30 min. (c) Latency of 12 gateways with max elevation 28° in 14 min.

Figure 16: The latency performance of generative data sets.

is 4.6×, 5.6× and 3.93× lower than the COTS protocol at most, respectively.

Large-scale simulation: We further simulate long-term application in 7 days. We select elevation and pass duration from the real satellite orbit and emulate the performance of multiple satellite passes using the generative datasets model. We repeat the 7-day experiment 100 times to obtain large-scale simulation results to discover the performance gain of SateRIoT further. The boxplot figures as shown in Figure 15. The accumulative packet volume of SateRIoT consistently exceeds the COTS baseline over the 7 days, maintaining a steady range approximately from 1.5 to 2.0 ×.

5.2 Link Estimation Model

In this section, we compare the performance among different link estimation trigger schemes and different link estimation transmission window lengths. The ground truths for link estimation during transmission windows are derived from local logs of real outdoor experiments. Then, we use the link estimation model and input feature derived from the logs to predict transmission link capability by comparing it with the real link. MAE is derived by comparing predicted values from the link model to ground truths. We use w(windows) 1,2,3,4,5 to represent the data beacon amount during one transmission window.

5.2.1 Accuracy. We split our datasets randomly with 20% of the total as test datasets. The predicted accuracy of window 4 can be up to 95.07% ACC with only 0.11 MAE. The feature importance of TD ESP and RT ESP features are 0.16 and 0.13, respectively. This affirms the significance of our feature engineering. The results affirm the reliability of our link estimation model, which can enhance the performance of individual gateways and optimize overall data transmission.

Comparison among different models: The comparison results are shown in Figure 17. the accuracy is only 44.44% and 2.12 MAE for the LDB model. This suggests that the link information from the data packet acknowledgment is necessary to guarantee accuracy. The ACCs of window 1,2,3,5 are 90.88%, 91.35%, 80.00%, 76.98% respectively. The MAEs of window 1,2,3,5 are 0.16, 0.23, 0.70 and 0.93 respectively. The performance for SateRIoT-1 and SateRIoT-2 are similar to SateRIoT-4, whereas SateRIoT-3 performs poorly as it may interrupt continuous data pattern that should have been captured in window 4, leading to inaccurate predictions of the last data beacon with 18 classifications. When the window length exceeds 4, performance declines because only 2% of transmission periods last longer than 120 seconds. This highlights the importance of selecting an optimal window length to balance accuracy and efficiency. Based on empirical testing, we have selected w=4 as our optimal window length.

Performance with different maximum elevation and locations: To verify the performance in different scenarios, we evaluated our model's performance in 12 locations and multiple satellite elevation values. Figure 18(a) displays the distribution of ACC and MAE for various groups of maximum elevation angles, categorized into "10-30", "30-50", "50-70", and "70-90" degrees. All groups achieve more than 91% accuracy and a very low MAE of less than 0.3. This suggests that our link estimation model is versatile and functions effectively across different elevation angles and for satellites with varying orbits. Figure 18(b) presents the distribution of accuracy and MAE for link estimation performance across 12 locations within a farm. Notably, our model achieves nearly 100% accuracy and zero MAE at 8 locations. Three locations surpass 87% accuracy, and one location achieves 82%, with all MAEs remaining below 0.7. This deviation is within an acceptable range and has minimal impact on the estimations

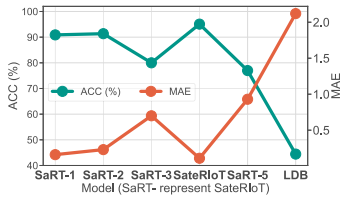


Figure 17: ACC and MAE of different link models.

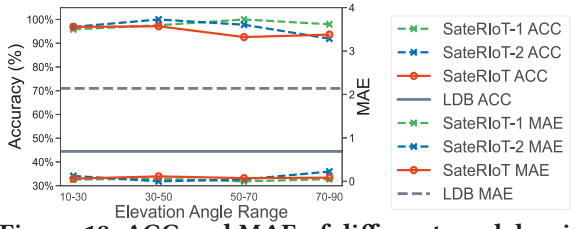


Figure 18: ACC and MAE of different models with various elevation and locations

and gateway performance. These results highlight the robustness and reliability of our bursty link model across different elevations and geographic locations.

Remark: Overall, the versatility of our model across varying elevation angles, combined with its robustness in different geographic locations, demonstrates its efficacy in diverse operational conditions. Balancing window length and triggering post-ACK receipt are key strategies to maximize link estimation accuracy and system efficiency in LoRa networks. To improve adaptability and scalability, it can be further trained in more complex environments.

5.2.2 Energy Efficiency. **Energy experimental settings:** We conducted experiments using real datasets mentioned in §2.5 as test datasets. As outlined in Section 2.4, the COTS method attempts to transmit 6 times after the gateway receives each satellite data beacon. It continues attempting transmissions whenever a satellite data beacon arrives. **ENV** stops the transmission if the cumulative packet amount reaches its predicted value during the satellite pass duration. **LDB** predicts the data packet transmission capability after one data beacon arrives and schedules predicted attempts. SateRIoT and SateRIoT-n stop after successfully transmitting a predicted number of packets. Such an approach helps reduce energy consumption by avoiding unreliable links. We predict and calculate the average attempts for each packet as an energy efficiency metric during one satellite pass duration. Then, we count the energy efficiency distributions from multiple satellite passes to compare the performance.

Results: Figure 19 compares the energy efficiency distributions between SateRIoT and the baseline methods. SateRIoT consistently requires fewer than 1.65 attempts per packet across all scenarios, in contrast to LDB, ENV, and COTS, which require up to 3.5, 6, and 6 attempts per packet, respectively. For 80% of the cases, SateRIoT's model prediction keeps energy efficiency under 1.25, while LDB and ENV, along with COTS, reach 1.6 and 1.85, respectively. Over 50% of the cases with SateRIoT transmit packets without extra energy waste, whereas LDB achieves this in 32%. For each packet at one gateway, averagely, SateRIoT attains up to 3.3× or 28.15 J lower energy usage than the COTS protocol during a single satellite pass. We can also observe that the model with different window lengths exhibits a similar trend in energy efficiency, but SateRIoT performs slightly better than other SateRIoT-n solutions thanks to the higher accuracy. This demonstrates that SateRIoT, unlike the more generalized link prediction methods like LDB or the COTS protocol, utilizes precise link estimation for energy-efficient transmissions. The overall energy saving will increase as the number of gateways grows.

5.3 Multi-hop Flooding Protocol

Beacon flooding experiments: We conducted experiments to evaluate the performance of flooding protocol in terms of reliability and latency. Initially, we designated one gateway in active mode and initiated its beacon broadcast according to its unique transmission plan to all other gateway using our CSMA-enabled multi-hop protocol. Then, we recorded the received time for all packets. This process was repeated 10 times for each gateway, designating each one as the active gateway in turn. For a setup of 12 gateways, we obtained latency data from a total of 30 experimental runs.

One active gateway results: Figure 20(a) depicts the overall latency distribution for all the LoRa radios. The PDR is 100%. Approximately 70% of the standby gateways can receive the flooding beacon from the active gateway within a latency of less than 2 s. Furthermore, around 94% of the standby gateways can capture the flooding beacon from the active gateway within a 5-second window. This indicates that in 94% of cases, beacon dissemination from two active gateways with a time offset of larger than 5 s can prevent packet duplication. In our real data sets measured in § 2.5, instances where a concurrent time offset of less than 5 s occur in less than 2% of all effective transmission slots. This latency level is sufficiently low to facilitate timely data transmission.

Concurrent flooding beacons experiments: We seek to understand the performance when two gateways broadcast their beacons concurrently with a random time offset ranging from 0 s to 5 s. We documented both the arrival time of the first beacon and the arrival times of both beacons at each gateway. Moreover, to examine the extreme scenarios where three gateways become active in close succession, we orchestrated an experiment where three beacons were broadcasted in rapid succession with a random time offset ranging from 0 to 5 seconds, either between the first and second beacon or between the second and third beacon. We record the arrival times of three beacons for each gateway. Such instances accounted for less than 0.1% of our emulation data. This three-beacon experiment was also repeated 120 times.

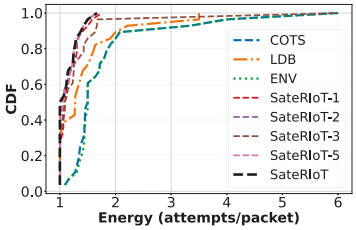
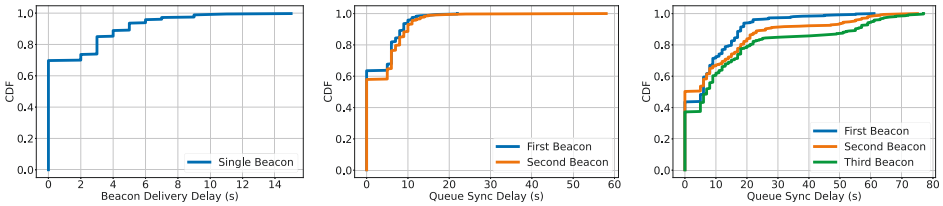


Figure 19: CDF comparison of energy efficiency and link waste.



(a) One active gateway (b) Two concurrent active gateways (c) Three concurrent active gateways

Figure 20: The latency distribution of flooding beacons in multi-hop protocol.

Latency Results: Figure 20(b) illustrates that approximately 90% of the gateways can receive both beacons under 9 s, which may result in a maximum of 4 duplicate packets. Additionally, around 60% of the gateways can receive both beacons within 5 s. Figure 20(c) indicates that around 60% beacons can be received with a delay of up to 20 s. around 60% beacons can be received with a delay of up to 9 s. When comparing three figures in Figure 20, it's evident that the time required to reach approximately 80% of gateways approximately triples with the addition of each beacon.

Remark: The effective performance of the CSMA-enabled multi-hop flooding protocol enables SateRIoT to operate efficiently. Its impressive time tolerance within 5 s ensures that concurrent cases are infrequent and hardly impact overall data transmission or create duplicate issues.

Overall Energy Analysis: SateRIoT achieves a comparable throughput for each individual gateway while maintaining approximately half the energy consumption of the baseline method. The propagation path between sensor nodes and satellites is usually hundreds of kilometers long, dramatically increasing the energy consumption to transmit the same amount of data. For example, the power consumption of SWARM-M138 [52] modem is 12.24 J in Tx mode, while Semtech SX1262 LoRa radio [42] only consumes from 0.002 J to 0.045 J [44] with a 192-byte packet from SF7 to SF12, which is 6120× to 272× less. In the busiest situations with the most power consumption mode with SF12, one satellite transmits one packet; this requires data sharing with LoRa radio transmission at most 12 times. The energy consumption of the LoRa packet is much lower than the power consumption of satellite radio.

6 RELATED WORK

Satellite Networking: L2D2 [57] presents a distributed scheduling system to reduce latency for downlink data transmission with hybrid ground stations. Umbra [56] proposes a withholding scheduling scheme for backhaul from satellites to large ground stations to overcome the uneven queueing effect. Li et al. [29] evaluate that the orbit of COTS satellites is unpredictable because of collision avoidance for Starlink [48]. However, SateRIoT does not heavily rely on orbit parameters but instead on coarse elevation and duration information. STARRYNET [23] builds an open-source experimental framework to simulate complicated network behaviors. SpaceCore [28] designs a stateless architecture to enable 5G deployment at satellites. Serval [55] enables near-real-time insights for latency-sensitive imagery applications by emerging computational capabilities on the satellites and ground stations. Compared to these works, SateRIoT focuses on the low-cost satellite IoT in rural areas.

Satellite-based IoT: The research work on satellite IoT mainly focuses on network modeling and physical layer design. Fraire et al. [13] present a sparse direct-to-satellite constellation design combined with LoRa. Zhang et al.[62] utilize Bernoulli–Rician message to enable channel estimation and user activity detection. Qian et al. [37] proposed a symmetry chirp spread modulation. Spectrumize [45] utilizes satellite movement-induced Doppler shift as a unique identifier for detection and decoding. In contrast, SateRIoT designs and implements a network backhaul for rural area IoT by space link.

7 CONCLUSION

To conclude, we introduce SateRIoT, a novel IoT backhaul architecture that merges LPWA technology on the ground with cost-effective IoT LEO satellites in space to support efficient rural area networking. First, we design a bursty link model to predict packet transmission capacity, reducing unnecessary data transmission. Next, we refine the model by selecting key features and optimizing the window length. To achieve link sharing, we develop a multi-hop flooding protocol to maintain data sharing among all gateways and use a priority-based queue structure to avoid duplicate transmissions. We implement SateRIoT with COTS satellite IoT and LoRa radios and evaluate the performance on real deployment and real-world collected traces. The results show that SateRIoT can achieve 3.3× less energy consumption for an individual gateway. For overall performance, SateRIoT reduces latency for packet delivery up to 5.6× and improves overall throughput by 1.9×.

ACKNOWLEDGEMENT

We sincerely thank the anonymous reviewers and our shepherd for their valuable feedback. This work was partially supported by US NSF under grant CAREER-2338976, NeTS-2312674, 2312676. and Hong Kong GRF 16204224.

REFERENCES

[1] LoRa Alliance. Retrieved Mar 15, 2023. A technical overview of LoRa and LoRaWAN. https://lora-alliance.org/resource_hub/what-is-lorawan/.

[2] Zhichao Cao, Jiliang Wang, Daibo Liu, Qiang Ma, Xin Miao, and Xufei Mao. 2020. Chase++: Fountain-enabled fast flooding in asynchronous duty cycle networks. *IEEE/ACM Transactions on Networking* 29, 1 (2020), 410–422.

[3] Marco Centenaro, Cristina E Costa, Fabrizio Granelli, Claudio Sacchi, and Lorenzo Vangelista. 2021. A survey on technologies, standards and open challenges in satellite IoT. *IEEE Communications Surveys & Tutorials* 23, 3 (2021), 1693–1720.

[4] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[5] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of ACM SIGKDD*.

[6] TE Connectivity. [n. d.]. ANT-916-CW-RCS-SMA. https://www.te.com/usa-en/product-ANT-916-CW-RCL-SMA.html?q=CW-RCL&source=header. Retrived Mar 15, 2024.

[7] Silvia Demetri, Marco Zúñiga, Gian Pietro Picco, Fernando Kuipers, Lorenzo Bruzzone, and Thomas Telkamp. 2019. Automated estimation of link quality for LoRa: A remote sensing approach. In *Proceedings of IPSN*.

[8] DRYAD. [n. d.]. Fighting Wildfires with Long-Range, Low-Power Technologies. https://https://www.dryad.net/post/fighting-wildfires-with-lorawan. Retrieved Mar 15, 2024.

[9] Jialuo Du, Yidong Ren, Zhui Zhu, Chenning Li, Zhichao Cao, Qiang Ma, and Yunhao Liu. 2023. SRLoRa: Neural-enhanced LoRa Weak Signal Decoding with Multi-gateway Super Resolution. In *Proceedings of ACM MobiHoc*.

[10] Bruce R Elbert. 2008. *Introduction to satellite communication*. Artech house.

[11] ESPRESSIF. [n. d.]. ESP32. https://www.espressif.com/en/products/socs/esp32. Retrived Mar 15, 2023.

[12] Shifeng Fang, Li Da Xu, Yunqiang Zhu, Jiaerheng Ahati, Huan Pei, Jianwu Yan, and Zhihui Liu. 2014. An integrated system for regional environmental monitoring and management based on internet of things. *IEEE Transactions on Industrial Informatics* 10, 2 (2014), 1596–1605.

[13] Juan A Fraire, Santiago Henn, Fabio Dovis, Roberto Garello, and Giorgio Taricco. 2020. Sparse satellite constellation design for LoRa-based direct-to-satellite Internet of Things. In *Proceedings of IEEE GLOBECOM*.

[14] Amalinda Gamage, Jansen Christian Liando, Chaojie Gu, Rui Tan, and Mo Li. 2020. LMAC: efficient carrier-sense multiple access for LoRa. In *Proceedings of ACM MobiCom*.

[15] Maolin Gan, Yimeng Liu, Li Liu, Chenshu Wu, Younsuk Dong, Huacheng Zeng, and Zhichao Cao. 2023. Poster: mmLeaf: Versatile Leaf Wetness Detection via mmWave Sensing. In *Proceedings of ACM MobiSys*.

[16] Omprakash Gnawali, Rodrigo Fonseca, Kyle Jamieson, David Moss, and Philip Levis. 2009. Collection tree protocol. In *Proceedings of ACM Sensys*.

[17] MOUNIR Grari, MIMOUN Yandouzi, IDRISS Idrissi, MOHAMMED Boukabous, OMAR Moussaoui, MOSTAFA AZIZI, and MIMOUN MOUSSAOUI. 2022. Using IoT and ML for Forest Fire Detection, Monitoring, and Prediction: a Literature Review. *Journal of Theoretical and Applied Information Technology* 100, 19 (2022), 5445–5461.

[18] Louis J Ippolito. 2012. *Radiowave propagation in satellite communications*. Springer Science & Business Media.

[19] Louis J Ippolito Jr. 2017. *Satellite communications systems engineering: atmospheric effects, satellite link design and system performance*. John Wiley & Sons.

[20] Iridium. [n. d.]. Urb-IoT: LoRaWAN Outdoor Gateway. https://www.iridium.com/products/urb-iot/. Retrieved Mar 15, 2024.

[21] Oltjon Kodheli, Eva Lagunas, Nicola Maturo, Shree Krishna Sharma, Bhavani Shankar, Jesus Fabian Mendoza Montoya, Juan Carlos Merlano Duncan, Danilo Spano, Symeon Chatzinotas, Steven Kisseleff, et al. 2020. Satellite communications in the new space era: A survey and future challenges. *IEEE Communications Surveys & Tutorials* 23, 1 (2020), 70–109.

[22] Lacuna. [n. d.]. Lacuna Space. https://lacuna.space/. Retrieved Mar 15, 2023.

[23] Zeqi Lai, Hewu Li, Yangtao Deng, Qian Wu, Jun Liu, Yuanjie Li, Jihao Li, Lixin Liu, Weisen Liu, and Jianping Wu. 2023. {StarryNet}: Empowering Researchers to Evaluate Futuristic Integrated Space and Terrestrial Networks. In *Proceedings of USNEIX NSDI)*.

[24] Mihai T Lazarescu. 2013. Design of a WSN platform for long-term environmental monitoring for IoT applications. *IEEE Journal on emerging and selected topics in circuits and systems* 3, 1 (2013), 45–54.

[25] Philip Levis, Neil Patel, David Culler, and Scott Shenker. 2004. Trickle: A self-regulating algorithm for code propagation and maintenance in wireless sensor networks. In *Proceedings of USENIX/ACM NSDI*.

[26] Chenning Li and Zhichao Cao. 2022. Lora networking techniques for large-scale and long-term iot: A down-to-top survey. *Comput. Surveys* 55, 3 (2022), 1–36.

[27] Chenning Li, Yidong Ren, Shuai Tong, Shakhrul Iman Siam, Mi Zhang, Jiliang Wang, Yunhao Liu, and Zhichao Cao. 2024. ChirpTransformer: Versatile LoRa Encoding for Low-power Wide-area IoT. In *Proceedings of ACM MobiSys*.

[28] Yuanjie Li, Hewu Li, Wei Liu, Lixin Liu, Yimei Chen, Jianping Wu, Qian Wu, Jun Liu, and Zeqi Lai. 2022. A case for stateless mobile core network functions in space. In *Proceedings of ACM SIGCOMM*.

[29] Yuanjie Li, Hewu Li, Wei Liu, Lixin Liu, Wei Zhao, Yimei Chen, Jianping Wu, Qian Wu, Jun Liu, Zeqi Lai, et al. 2023. A networking perspective on starlink's self-driving leo mega-constellation. In *Proceedings of ACM MobiCom*.

[30] Zhenjiang Li, Mo Li, Jiliang Wang, and Zhichao Cao. 2011. Ubiquitous data collection for mobile users in wireless sensor networks. In *Proceedings of IEEE INFOCOM*.

[31] LONESTAR. [n. d.]. LoRaWAN Satellite Gateway. https://www.lonestartracking.com/lorawan-satellite-gateway/. Retrieved Mar 15, 2024.

[32] Gérard Maral, Michel Bousquet, and Zhili Sun. 2020. *Satellite communications systems: systems, techniques and technology*. John Wiley & Sons.

[33] OneWeb. [n. d.]. OneWeb Constellation. https://oneweb.net/. Retrieved Mar 15, 2023.

[34] Antonino Pagano, Daniele Croce, Ilenia Tinnirello, and Gianpaolo Vitale. 2023. A Survey on LoRa for Smart Agriculture: Current Trends and Future Perspectives. *IEEE Internet of Things Journal* 10, 4 (2023), 3664–3679.

[35] Pablo Ilabaca Parra, Samuel Montejo-Sánchez, Juan A Fraire, Richard Demo Souza, and Sandra Céspedes. 2022. Network size estimation for direct-to-satellite iot. *IEEE Internet of Things Journal* 10, 7 (2022), 6111–6125.

[36] Raspberry Pi. [n. d.]. Raspberry Pi 4. https://www.raspberrypi.com/products/raspberry-pi-4-model-b/. Retrived Mar 15, 2023.

[37] Yubi Qian, Lu Ma, and Xuwen Liang. 2018. Symmetry chirp spread spectrum modulation used in LEO satellite Internet of Things. *IEEE Communications Letters* 22, 11 (2018), 2230–2233.

[38] Yidong Ren, Puyu Cai, Jinyan Jiang, Jialuo Du, and Zhichao Cao. 2023. Prism: High-throughput LoRa backscatter with non-linear chirps. In *Proceedings of IEEE INFOCOM*.

[39] Yidong Ren, Li Liu, Chenning Li, Zhichao Cao, and Shigang Chen. 2022. Is LoRaWAN Really Wide? Fine-grained LoRa Link-level Measurement in An Urban Environment. In *Proceedings of IEEE ICNP*.

[40] Yidong Ren, Wei Sun, Jialuo Du, Huaili Zeng, Younsuk Dong, Mi Zhang, Shigang Chen, Yunhao Liu, Tianxing Li, and Zhichao Cao. 2024. Demeter: Reliable Cross-soil LPWAN with Low-cost Signal Polarization Alignment. In *Proceedings of ACM MobiCom*.

[41] Sateliot. [n. d.]. Sateliot. https://sateliot.space/en/. Retrieved Mar 15, 2023.

[42] Semtech. [n. d.]. Product Details SX1262. https://www.semtech.com/products/wireless-rf/lora-connect/sx1262. Retrived Mar 15, 2023.

[43] Semtech. [n. d.]. Semtech Smart Agriculture. https://www.semtech.com/lora/lora-applications/smart-agriculture. Retrieved Mar 15, 2023.

[44] Semtech. Retrieved Sep 11, 2023. LoRa Calculator. https://lora-developers.semtech.com/build/tools/calculator/.

[45] Vaibhav Singh, Tusher Chakraborty, Suraj Jog, Om Chabra, Deepak Vasisht, and Ranveer Chandra. 2024. Spectrumize: Spectrum-efficient Satellite Networks for the Internet of Things. In *Proceedings of USENIX NSDI*.

[46] André Sørensen, Hua Wang, Maxime Jérôme Remy, Nicolaj Kjettrup, René Brandborg Sørensen, Jimmy Jessen Nielsen, Petar Popovski, and Germán Corrales Madueño. 2022. Modeling and experimental validation for battery lifetime estimation in nb-iot and lte-m. *IEEE Internet of Things Journal* 9, 12 (2022), 9804–9819.

[47] SpaceX. [n. d.]. Space X. https://www.spacex.com/. Retrieved Mar 15, 2023.

[48] Starlink. [n. d.]. Space X Starlink. https://www.starlink.com/. Retrieved Mar 15, 2023.

[49] Zehua Sun, Huanqi Yang, Kai Liu, Zhimeng Yin, Zhenjiang Li, and Weitao Xu. 2022. Recent advances in LoRa: A comprehensive survey. *ACM Transactions on Sensor Networks* 18, 4 (2022), 1–44.

[50] Akey Sungheetha, Rajesh Sharma, et al. 2020. Real time monitoring and fire detection using internet of things and cloud based drones. *Journal of Soft Computing Paradigm (JSCP)* 2, 03 (2020), 168–174.

[51] SWARM. [n. d.]. CASE STUDY | DRYAD NETWORKS. https://swarm.space/wp-content/uploads/2022/02/Swarm-Environmental-Case-Study-Dryad.pdf. Retrieved Mar 15, 2024.

[52] SWARM. [n. d.]. Product Overview Swarm M138 Modem. https://swarm.space/wp-content/uploads/2022/09/Swarm-M138-Specifications.pdf. Retrieved Mar 15, 2023.

[53] SWARM. [n. d.]. Space X Swarm - Low cost, global satellite connectivity for IoT. https://swarm.space/. Retrieved Mar 15, 2023.

[54] SWARM. [n. d.]. SWARM EVA KIT. https://swarm.space/product/swarm-eval-kit/. Retrieved Mar 15, 2023.

[55] Bill Tao, Om Chabra, Ishani Janveja, Indranil Gupta, and Deepak Vasisht. 2024. Known Knowns and Unknowns: Near-realtime Earth Observation Via Query Bifurcation in Serval. In *Proceedings of USENIX NSDI*.

[56] Bill Tao, Maleeha Masood, Indranil Gupta, and Deepak Vasisht. 2023. Transmitting, Fast and Slow: Scheduling satellite traffic through space and time. In *Proceedings of ACM MobiCom*.

[57] Deepak Vasisht, Jayanth Shenoy, and Ranveer Chandra. 2021. L2D2: Low latency distributed downlink for LEO satellites. In *Proceedings of ACM SIGCOMM*.

[58] Ruihao Wang, Yimeng Liu, and Rolf Müller. 2022. Detection of passageways in natural foliage using biomimetic sonar. *Bioinspiration & Biomimetics* 17, 5 (2022), 056009.

[59] Tim Winter, Pascal Thubert, Anders Brandt, Jonathan Hui, Richard Kelsey, Philip Levis, Kris Pister, Rene Struik, Jean-Philippe Vasseur, and Roger Alexander. 2012. *RPL: IPv6 routing protocol for low-power and lossy networks*. Technical Report.

[60] Deliang Yang, Xianghui Zhang, Xuan Huang, Liqian Shen, Jun Huang, Xiangmao Chang, and Guoliang Xing. 2020. Understanding power consumption of nb-iot in the wild: tool and large-scale measurement. In *Proceedings of ACM MobiCom*.

[61] Fanhao Zhang, Fu Yu, Xiaolong Zheng, Liang Liu, and Huadong Ma. 2023. DFH: Improving the Reliability of LR-FHSS via Dynamic Frequency Hopping. In *Proceedings of IEEE ICNP*.

[62] Zhaoji Zhang, Ying Li, Chongwen Huang, Qinghua Guo, Lei Liu, Chau Yuen, and Yong Liang Guan. 2020. User activity detection and channel estimation for grant-free random access in LEO satellite-enabled Internet of Things. *IEEE Internet of Things journal* 7, 9 (2020), 8811–8825.