Online Green Data Gathering from Geo-distributed IoT Networks via LEO Satellites

Huawei Huang*, Song Guo†, Weifa Liang‡, and Kun Wang§
*Academic Center for Computing and Media, Kyoto University, Japan
†Department of Computing, The Hong Kong Polytechnic University
‡Research School of Computer Science, Australian National University, Australia
§Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
Email: hwhuang@media.kyoto-u.ac.jp, song.guo@polyu.edu.hk, wliang@cs.anu.edu.au, kwang@njupt.edu.cn

***Abstract*—As the critical supplementary to terrestrial communication networks, the low-earth-orbit (LEO) satellite based communication networks regain growing attentions in recent few years. In this paper, we focus on data gathering for geo-distributed Internet-of-Things (IoT) networks via LEO satellites. Normally, the power supply in IoT data-gathering gateways is a bottleneck resource that constrains the network throughput. Thus, the challenge is how to upload data from IoT gateways to LEO satellites under dynamic uplinks in an energy-efficient way. To address this problem, we first formulate a novel optimization problem, and then propose an online algorithm for green data-uploading in geo-distributed IoT networks. In the proposed framework, we aim to jointly maximize the network throughput and minimize the energy consumption at gateways, while avoiding the buffer overflow at gateways. We finally evaluate the performance of the proposed algorithm through simulations using both real-world and synthetic traces. The simulation results demonstrate that the proposed approach can achieve high efficiency on the power consumption and significantly reduce queue backlogs compared with a benchmark using greedy policy.**

I. Introduction

The Internet-of-Things (IoT) networks have been widely applied to various applications, such as the remote surveillance systems used to monitor natural disasters, wild animals and environmental parameters of climate change, as well as the precision agriculture and other remote asset-management networks shown in Fig. 1.

Tremendous numbers of IoT devices and data-gathering gateways in the edge together constitute the data-sensing and capturing system. The data-sensing devices may have low cost and long battery lives based on the emerging Narrow-band IoT technology [1]. In the large-scale geo-distributed IoT networks, such as the oil & gas platforms located in remote deserts or oceans, data-sensing can be accomplished by well-connected ground IoT networks. However, the problem is how to timely and efficiently gather the data cached in the distributed IoT gateways, and then forward the data to data centers for further analytics to enable quick decision making.

For the urban IoT networks, some existing studies [2]–[4] use the cellular networks such as 3G, 4G or potentially the 5G technologies to establish the dedicated data gathering networks. For the offshore IoT networks, studies [5], [6] proposed to use the UAVs to gather data sensing from offshore ocean-observation devices. However, these approaches are technically impossible or prohibited by their operation costs for large-scale geo-distributed IoT networks. Recently, low-earth-orbit (LEO) satellite based constellation networks have been launched by a number of representative companies. For example, several LEO satellite based projects have been announced by *OneWeb*, *SpaceX*, and *Boeing*, aiming to provide global Internet-access services. Under the fully covered global access networks [7], [8], LEO satellites provide great opportunities to the geo-distributed IoT networks. However, the challenge is to design energy-efficient data gathering schemes to aggregate the data caching in IoT gateways under the LEO satellite based access networks. Adopting this new data gathering scheme is based on the following three aspects. First, the power supply for the large number of IoT gateways isolated in the remote locations is viewed as a bottleneck constraint [6]. Second, the uplinks from IoT gateways to LEO satellites are time-varying dynamic channels, which are particularly sensitive to the weather conditions. For example, as shown in Fig. 1, the weather conditions are usually different at different geo-distributed gateways. Transmitting the same volume of data under a bad channel condition consumes much higher energy than that under a good condition [9]. Finally, if the data cached in IoT gateways fail to be collected timely, the successive data stream will overflow the buffer space of gateways. The so-called *buffer overflow* problem [10]–[12] will incur data loss. Therefore, it is significant to make an optimal green

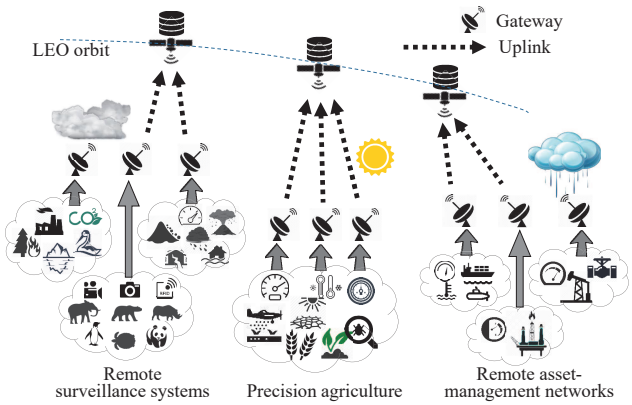


Fig. 1. Data gathering from geo-distributed remote IoT networks via LEO satellites. Weather conditions significantly affect the uplink channels.

scheduling for online data gathering for the geo-distributed IoT networks such that the total energy consumption is minimized, the throughput can be maximized, while the data overflow in gateways can be also avoided.

In this paper, a novel optimization problem based on this application scenario is formulated. Then, an online scheduling framework is developed using the Lyapunov optimization technique [13]. The main contributions of this paper are described as follows.

- To the best of our knowledge, this paper is the first to study the green online data gathering problem for the geo-distributed IoT networks by exploiting LEO based networks.
- To jointly minimize energy consumption and maximize throughput, and avoid buffer overflow problem meanwhile during data gathering from geo-distributed IoT networks, we devise a novel online algorithm for green data-uploading. The theoretic characteristics of this online algorithm are analyzed rigorously.
- Finally, based on the real-world traces of LEO constellation, the simulation results show that the proposed online algorithm achieves much higher efficiency of power consumption and lower queue backlogs than a greedy-based benchmark algorithm.

The rest of this paper is organized as follows. Section II reviews the related work. Section III specifies the system model and problem formulation. Section IV presents the online scheduling framework. Section V exhibits the performance evaluation. Finally, Section VI concludes this paper.

II. Related Work

In the perspective of data gathering for IoT networks, various approaches have been proposed for different scenarios [5], [6], [8], [14]–[21]. For example, Barbatei et al. [5] presented a UAV based prototype that can gather and relay data from the sensor nodes deployed in remote areas or floating on water surface. Zolich et al. [6] combined the UAV and the low-cost buoys hardware to implement a sensor data collection system, which has been used to gather the underwater sensor data in Norwegian subarctic fjord. To enable the IoT data collection processes for multiple parties, Cheng et al. [16] made use of a concurrent data collection tree to improve the collection effectiveness of IoT applications. A mobile satellite communication services company *Isatdata Pro* [8] exploited the LEO satellites to provide the global communication services for Machine-to-Machine (M2M) applications. This is very useful to relay the sensor data from remote assets such as oil, gas, maritime, commercial fishing and heave equipment sectors.

Several studies related to satellite based communication networks have been recently conducted. For example, Wu et al. [22] proposed a two-layer caching model for content delivery services in satellite-terrestrial networks. Jia et al. [23] studied data transmission and downloading by exploiting the inter-satellite links in the LEO satellite based communication networks.

Comparing with the existing studies, we particularly focus on the green online data gathering problem for the global distributed IoT networks using LEO satellites.

III. Network Model and Problem Formulation

A. System Model

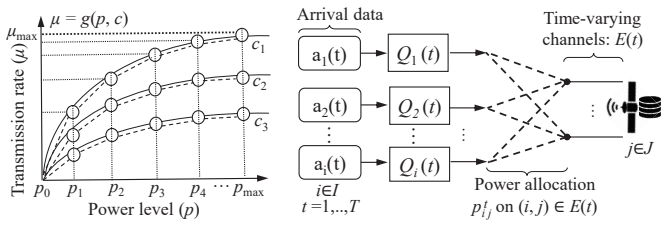
We consider a discrete system measured in time-slot with a period $t \in \{1, 2, ...T\}$, where T denotes the # of time slots. The length of each slot is denoted by δ, which ranges from hundreds of milliseconds to minutes [24]. We then focus on geo-distributed IoT networks $\mathcal{G} = \langle I \cup J, E(t) \rangle$, where I and J are a set of ground IoT data-gathering gateways and LEO satellites orbiting in specific planes, respectively. $E(t)$ is a set of time-varying uplinks in time slot t between the IoT data-gateways and LEO satellites. The gathered data can be temporally stored in satellites and transmitted to ground stations eventually. Note that, we only study data gathering through uplinks in this paper.

Since the LEO satellites are orbiting in their planes according to predefined parameters, the time-varying available uplinks between IoT gateways and satellites can be known in priori for every time slot. We use $(i, j) \in E(t)$ to denote an uplink channel between an IoT gateway $i \in I$ and a LEO satellite station $j \in J$, and let c_{ij}^t represent the channel state of (i, j) at time slot t. The time-varying channel state can be obtained by direct measurement [9] or by prediction [25]. We thus assume that the channel state can be pulled up by the centralized system controller at the beginning of a time slot in our system model.

Every satellite has a data-receiving rate capacity, which is denoted by C_j, $j \in J$. Furthermore, we consider that the transmission channels are working under the Time Division Multiple Access (TDMA) [26] mechanism, which has become a mature satellite communication technology. Thus, a transmission channel can be reused by different uplinks at different time slots.

We then describe the relationship between the power allocation and transmission rate on an uplink by referring a well-adopted concave rate-power curve $g(p, c)$ [9], [27] as shown in Fig. 2(a), where p and c denote the power allocation and the channel condition, respectively. In practice, the power-allocation parameter in a transmitter adopts linear piecewise power-rate curves with a pre-defined finite set of discrete operating *gears* [9], [27] denoted by $\overrightarrow{P} = [p_1, p_2, ..., p_{\max}]$, rather than the continuous concave function as shown in Figure 2(a). Thus, the transmission rate of an uplink is determined by two critical parameters, i.e., the power level allocated and the currently observed channel condition. It is worth noting that, the maximum transmission rate of each uplink is $\mu_{\max}$ under arbitrary channel conditions, i.e., $g(p, c) \leq \mu_{\max}, \forall p \in \overrightarrow{P}, \forall c \in \overrightarrow{C}$.

As shown in Fig. 2(b), IoT data stream arrives to each gateway $i \in I$ at each time slot t with a volume $a_i(t)$. Note that, we assume all the data-arrival rates at IoT gateways are within a positive peak value R_{max}. Let $Q_i(t)$ denote the



(a) Rate-power lines (dashed). (b) System model.

Fig. 2. The classical piecewise rate-power curve [9], [27] with parameters: power-supply level (p) and channel condition (c). c_1, c_2, and c_3 are different channel conditions, e.g., {good, medium, bad}. System model illustrates the decisions for the arrival IoT data: power allocation on selected uplinks.

time-varying backlog of the queue residing in gateway i. It can be seen that $Q_i(t)$ keeps growing if the data in gateway i cannot be successfully gathered by satellites, and finally triggers buffer overflow. Then, if a satellite uplink is launched for a gateway, it can be used to upload packets to the satellite immediately. Because of the relative motion between LEO satellites and gateways, we consider the *preemptive* model for each time-varying uplink. Specifically, an uplink is *preemptive* so that a data-uploading via this uplink in the current time slot can be taken over by another uploading in the next slot.

B. Problem Statement and Formulation

1) Variables: Given the system model described above, the crucial control decision we need to make is the power allocation in each uplink channel. Therefore, we define a real-valued variable $p_{ij}^t \in \overrightarrow{P}$ to represent the power allocation level on the uplink $(i,j) \in E(t)$ at time slot t.

2) Performance Metrics: For data gathering from the geo-distributed IoT networks, network throughput is the most critical performance metric, which should be devoted to improve. Denoted by $thr(t)$, we define the time-varying throughput at time slot t as

$$thr(t) = \sum_{(i,j)\in E(t)} g(p_{ij}^t, c_{ij}^t) \cdot \delta, \forall t. \tag{1}$$

As mentioned, the data-upload in an IoT gateway is constrained by its energy-budget. If the power allocation on uplink channels cannot be carefully scheduled, e.g., allocating too large power level to an uplink with bad channel condition, much energy is going to be wasted, thus reducing the network throughput. Therefore, the power consumption should be minimized when uploading data to satellites. Denoted by $power(t)$, the total power consumption on data-uploading throughout all IoT gateways at time slot t is calculated as

$$power(t) = \sum_{(i,j)\in E(t)} \delta \cdot p_{ij}^t, \forall t. \tag{2}$$

To maximize the throughput and minimize the power consumption simultaneously, we define a penalty function that positively associates with the numerical power consumption $power(t)$ and reversely associates with the numerical throughput $thr(t)$. The objective is to minimize a time-average penalty, which is denoted by $\overline{Pen}$, while all queue backlogs are keeping mean-rate stable. Note that, if a queue in gateway $i \in I$ is mean-rate stable [13], it satisfies $\lim_{t\to\infty} \frac{E\{Q_i(t)\}}{t} = 0$. We thus have the following penalty-minimization formulation.

$$\min \quad \overline{Pen} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [\beta \cdot power(t) - thr(t)] \tag{3}$$

$$\text{s.t.} \quad \sum_{i:(i,j)\in E(t)} g(p_{ij}^t, c_{ij}^t) \le C_j, \ \forall j \in J, \forall t. \tag{4}$$

$$Q_i(t) \text{ is mean-rate stable}, \ \forall i \in I. \tag{5}$$

$$\text{Variable: } p_{ij}^t \in \overrightarrow{P}, \forall (i,j) \in E(t), \forall t = 1, ..., T.$$

In the objective function (3), β indicates the weight of power consumption in the penalty function. Letting C_j denote the total data receiving rate capacity of LEO satellite $j \in J$ at any time slot, inequality (4) indicates that the total uploading data rate should not exceed the capability at each satellite when it is receiving data from ground IoT gateways. Finally, constraint (5) ensures to avoid the buffer overflow in queues.

IV. Online Algorithm Design

In this section, using the Lyapunov optimization theory [13], we strive for a near-optimal solution to the online green data gathering problem (3). Under the Lyapunov optimization framework, queue backlogs are extremely useful for designing dynamic algorithms that do not require a-priori knowledge of channel statistics.

A. Problem Transformation

1) Dynamics of Queues: Recall that the backlog $Q_i(t)$ represents the data size measured in bits in the queue of gateway $i \in I$. A small backlog indicates queue stability, while a large one implies high probability of buffer overflow. Initially, $Q_i(1) = 0, \forall i \in I$. Afterwards, the time-varying queue backlog of each IoT gateway evolves as follows.

$$Q_i(t+1) = \max\left[Q_i(t) - b_i(t), 0\right] + a_i(t), \forall i \in I, \tag{6}$$

where $b_i(t) = \delta g(p_{ij}^t, c_{ij}^t), (i,j) \in E(t)$, represents the total diminishing bits in backlog Q_i.

2) Virtual Queues: We then transform the original minimization problem (3) into a pure queue-stability problem based on Lyapunov optimization theory [13]. To make sure the constraint (4) still holds, we define a virtual queue X_j for each satellite $j \in J$ with the following update function.

$$X_j(t+1) = \max[X_j(t) + x_j(t), 0], \forall t = 1, ..., T, \tag{7}$$

where $x_j(t) = \sum_{i:(i,j)\in E(t)} g(p_{ij}^t, c_{ij}^t) - C_j$, $\forall j \in J$; $\forall t = 1, ..., T$. The initial backlog is $X_j(1) = 0$ for each virtual queue.

Insight: By summing $X_j(t)$ over time slots $t = 1, ..., T$, we have $\frac{X_j(T)}{T} - \frac{X_j(1)}{T} \ge \frac{1}{T}\sum_1^T x_j(t)$. With $X_j(1) = 0$, taking expectations on both sides and letting $T \to \infty$, we get $\lim_{T\to\infty} \sup \frac{\mathbb{E}\{X_j(T)\}}{T} \ge \lim_{T\to\infty} \sup \overline{x_j}(t)$, where $\overline{x_j}(t)$ is the time-average expectation of $x_j(t)$ over $t = 1, ..., T$. If $X_j(t)$ is mean-rate stable [9], we have $\lim_{T\to\infty} \sup \frac{\mathbb{E}\{X_j(T)\}}{T} = 0$, which indicates that

$\lim_{T\to\infty}\sup\overline{x_j}(t) \leq 0$. This implies that the desired constraints for $x_j(t)$ are met.

Then, combining all actual and virtual queues, we can obtain a concatenated vector $\Theta(t) = [\mathbf{Q}(t), \mathbf{X}(t)]$ with update equations (6) and (7). Next, a Lyapunov function of the geo-distributed data gathering system is defined as follows.

$$L(\Theta(t)) \triangleq \frac{1}{2}\sum_{i\in I} Q_i(t)^2 + \frac{1}{2}\sum_{j\in J} X_j(t)^2. \tag{8}$$

In fact, $L(\Theta(t))$ calculates a scalar volume of queue congestion [13] in the geo-distributed data gathering system. Normally, a Lyapunov function with a small value indicates short backlogs of both actual and virtual queues. Thus, the system could keep in a stable state.

3) Drift-plus-Penalty Expression: We then define a *one-slot conditional Lyapunov drift* [13], denoted by $\Delta(\Theta(t))$, which is calculated as

$$\Delta(\Theta(t)) = \mathbb{E}\{L(\Theta(t+1)) - L(\Theta(t))|\Theta(t)\}. \tag{9}$$

Insight: Given the current backlogs of the system $\Theta(t)$, the drift shown as Eq. (9) depicts the expectation of variation measured in Lyapunov function (8) over one time slot. Under the framework of Lyapunov optimization, the supremum bound of Lyapunov drift-plus-penalty expression is expected to be minimized in each time slot, aiming to retrieve the near-optimal decisions for our proposed original green data gathering problem.

Thus, the transformed problem is rewritten as the follows.

$$\begin{aligned} &\min\ \Delta(\Theta(t)) + V\mathbb{E}\{\beta\cdot power(t) - thr(t)|\Theta(t)\} \\ &\text{s.t.}\quad p_{ij}^t \in \overrightarrow{P}, \forall t = 1, ..., T. \end{aligned} \tag{10}$$

Note that, in the objective function (10), V is a tunable knob denoting the weight of penalty. It can be found that (10) reaffirms our three-fold goals for the online green data gathering from geo-distributed IoT networks: (1) to minimize the power consumption, (2) to maximize the throughput, and (3) to maintain the stability of the holistic system meanwhile.

We then have the following Lemma.

Lemma 1: Given that the data arrival rate $a_i(t)$, the time-varying available uplink set $E(t)$, the backlogs of both actual and virtual queues are observable at each slot t, for any value of $\Theta(t)$, the Lyapunov drift $\Delta(\Theta(t))$ of the geo-distributed IoT data gathering system under arbitrary control policies satisfies the following results:

$$\begin{aligned} \Delta(\Theta(t)) \leq & B + \sum_{i\in I} Q_i(t)\mathbb{E}\{a_i(t) - b_i(t)|\Theta(t)\} \\ & + \sum_{j\in J} X_j(t)\mathbb{E}\{x_j(t)|\Theta(t)\}, \end{aligned} \tag{11}$$

where $B = \frac{1}{2}|I|[R_{max}^2 + (|J| + \delta^2)\mu_{\max}^2] + \sum_{j\in J} C_j(\frac{1}{2}C_j - |I|\mu_{max})$ is a positive constant. Note that, $|.|$ represents the size of a set.

The proof of Lemma 1 is moved to our extended version [28] due to space limitation.

Based on Lemma 1, we then derive the upper bound of *drift-plus-penalty* expression for the geo-distributed data gathering system by combining (10) and (11) as follows.

$$\begin{aligned} \Delta(\Theta(t)) + V\mathbb{E}\{\beta\cdot power(t) - thr(t)|\Theta(t)\} \leq B & \\ +V\delta \sum_{(i,j)\in E(t)} [\beta p_{ij}^t - g(p_{ij}^t, c_{ij}^t)] & \quad (12) \\ + \sum_{i\in I} Q_i(t)\mathbb{E}\{a_i(t) - \delta g(p_{ij}^t, c_{ij}^t)|\Theta(t)\} & \quad (13) \\ + \sum_{j\in J} X_j(t)\mathbb{E}\{x_j(t)|\Theta(t)\}. & \quad (14) \end{aligned}$$

B. Online Scheduling Algorithm

Unlike existing offline solutions that make decisions based on the known data-arriving rates, we do not make such an impractical assumption. Instead, we design our online scheduling algorithm only depending on the observed queue backlogs in gateways. Driven by the upper bound of *drift-plus-penalty* expression derived in the end of last subsection, it can be seen that minimizing the objective in (10) is equivalent to minimizing expressions (12), (13) and (14) jointly. Thus, we have the following 2-phase power allocation algorithm for uplinks.

1) Phase-I, Power Allocation on uplinks: In each time slot, the power allocation decisions on uplinks are independent among different gateways. Therefore, the power allocation can be accomplished by the centralized system controller for each individual gateway without having to know the backlog information from other gateways. This is a very practical merit for the large-scale global geo-distributed IoT networks.

Letting (p, c) be short for the term (p_{ij}^t, c_{ij}^t), we have the following subproblem (15):

$$\begin{aligned} &\min\quad \Gamma(p, c) \\ &\text{s.t.}\quad p_{ij}^t \in \overrightarrow{P}, (i,j) \in E(t), i \in I, \forall t, \end{aligned} \tag{15}$$

where $\Gamma(p, c) = V[\beta p_{ij}^t - g(p, c)] + g(p, c)[X_j(t) - Q_i(t)]$.

It can be observed that the problem (15) is a simple linear programming. Partially differentiating $\Gamma(p, c)$ with respect to p and rearranging terms, we have

$$\frac{\partial\Gamma(p,c)}{\partial p} = V\beta + [X_j(t) - Q_i(t) - V]\frac{\partial g(p,c)}{\partial p}. \tag{16}$$

Note that, the term $\frac{\partial g(p,c)}{\partial p}$ in each discrete power supply level $p \in \overrightarrow{P}$ can be easily retrieved under the observed channel condition c. Letting p vary within the vector $\overrightarrow{P} = [p_1, p_2, ..., p_{\max}]$, a vector of derivative values can be obtained as follows.

$$\overrightarrow{\mathbb{D}} = [\frac{\partial\Gamma(p,c)}{\partial p_1}, \frac{\partial\Gamma(p,c)}{\partial p_2}, ..., \frac{\partial\Gamma(p,c)}{\partial p_{\max}}]. \tag{17}$$

Since $g(p, c)$ is a concave function, which determines that $\Gamma(p, c)$ is convex. By Eq. (16), we have the valley point (p^*, c_{ij}^t) of $\Gamma(p, c)$ such that

$$\frac{\partial g(p^*, c_{ij}^t)}{\partial p^*} = \frac{V\beta + X_j(t)}{Q_i(t) + V - X_j(t)}. \tag{18}$$

Finally, the power-allocation solution can be chosen from the given power-level vector as follows.

$$p_{ij}^t = \begin{cases} p_{\min}, & \text{if elements (ele.) in } \overrightarrow{\mathbb{D}} \text{ are non-negative;} \\ p_{\max}, & \text{if ele. in } \overrightarrow{\mathbb{D}} \text{ are non-positive;} \\ p^- \text{or } p^+: & \arg\min\{\Gamma(p^-, c_{ij}^t), \Gamma(p^+, c_{ij}^t)\}, \text{if ele.} \\ & \text{in } \overrightarrow{\mathbb{D}} \text{ vary from negative to positive,} \end{cases}$$

where p^- and p^+ are two successive discrete power levels such that $p^- \leq p^* \leq p^+$, where $p^-, p^+ \in \overrightarrow{P}$, and p^* is the optimal power level denoted by the valley point (p^*, c_{ij}^t).

2) Phase-II, Queue Update: In the end of each time slot, using the optimal solutions p_{ij}^t, the actual queues $\mathbf{Q}(t)$ and the virtual queues $\mathbf{X}(t)$ need to be updated by invoking Eq. (6) and Eq. (7), respectively.

V. Performance Evaluation

A. Simulation Settings

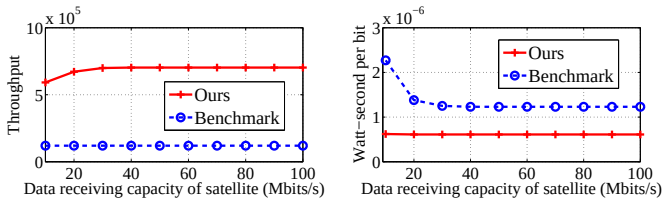
The performance of the proposed online green data gathering algorithm is evaluated using the well-known emulator Satellite Tool Kit (STK) [29], which is designed by AGI (Analytical Graphics, Inc.). Using STK, we retrieve the contact trace between LEO satellites and the terrestrial IoT gateways at different time slots. To strengthen the simulation, we build a LEO system based on the widely-adopted Globalstar constellation [23], [30], which is composed of 48 LEO satellites averagely distributed in 8 orbital planes.

In total 316 IoT gateways are deployed globally and averagely in the world map. We also generate the synthetic channel-state traces with three states (i.e., good, medium and bad) according to weather conditions of all locations obtained from the Internet. The one-day mission of LEO satellites starts from 12 July 2017 00:00:00 UTCG (Gregorian Coordinated Universal Time). The length of each time slot is set as 10 seconds. The contact trace between each satellite and each IoT gateways is retrieved in each time slot.

On the other hand, the bandwidth of each uplink channel is set to 1 megahertz (MHz). To calculate the data-receiving rate of uplinks, we adopt the classic rate-power function $g(p, c) = 1$ MHz$\cdot\log(1+\upsilon p)$ [27], where υ denotes the fading coefficient determined by the channel state. As the three-state condition model [25] adopted to depicts the satellite channels, υ is equal to 5.03, 3.46 and 1.0 corresponding to the *good*, *medium* and *bad* conditions. The power-level vector $\overrightarrow{P}$ is set to 11 gears averagely varying from 0 Watt to 1 Watt. We then generate the synthetic data-arrival traces for each IoT gateways with the predefined range, denoted by α_{LB} and α_{UB}, of the arrival data-volume in each time slot. In simulation, we set α_{LB} and α_{UB} to 10 Megabits (Mbits) and 100 Mbits, respectively.

B. Metrics and Benchmark

We evaluate the performance of the proposed online algorithm in three metrics: throughput, efficiency of power consumption and queue backlogs. Note that, the second metric is measured in the number of *watt·second* (w·s for short) spending on uploading per bit of data. The insight we design



(a) Throughput (Mbits). (b) Efficacy of power consumption.

Fig. 3. Performance comparison between our proposed online algorithm and the benchmark algorithm with 100 time slots. Note that, the efficiency of power consumption measured in *Watt·Second per bit* (*w·s/bit*).

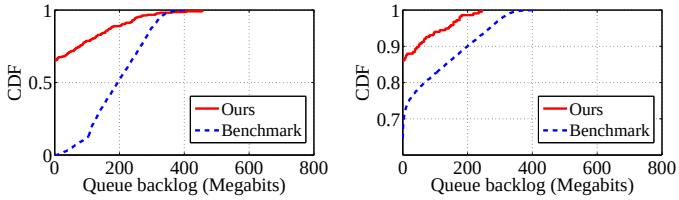
this metric is that a good algorithm probably yields both higher throughput and larger power consumption than a worse algorithm. Therefore, the most fair way to evaluate the performance of algorithms is the unit of power consumption used in the data-uploading through uplinks. Finally, the queue-backlog is the indicator of the system stability, thus backlog should be made as small as possible in each queue.

We then design a "big-backlog-first" greedy algorithm as the benchmark to compare the performance with the proposed one. The basic idea includes the following two steps. (a) Sort all the gateway queues in a non-increasing order by their queue backlogs. (b) Only the first few gateways can use the uplinks that are available currently. The data-receiving rate capacity of each satellite need to be conserved. In this greedy algorithm, to guarantee the fairness when allocating power levels to the uplinks for the gateways that are first to be served, we define a normalized control parameter, which is denoted by ζ, indicating the percentage of backlog that should be reduced in a queue through allocating power on the associated uplink.

C. Simulation Results

To evaluate the effect of the data-receiving rate capacity of each LEO satellite, β, V and ζ are set to 1, 1 and 10%, respectively. We then examine the throughput and the power consumption efficiency yielded by the proposed online algorithm by varying the data-receiving capacity of satellites from 10 to 100 Mbits/s. First, Fig. 3(a) illustrates the throughput performance under the proposed online algorithm and the benchmark algorithm. It can be seen that throughput shows as a non-decreasing function as the data receiving rate capacity of satellites grows, and the throughput of our online algorithm outperforms that of the benchmark algorithm significantly. Then, Fig. 3(b) demonstrates the efficiency of power consumption measured in terms of *w·s/bit* of data gathered by satellites. We can observe that the efficiency demonstrates as a non-increasing function of the data-receiving capacity under the benchmark algorithm. In contrast, our online algorithm achieves a significantly low *w·s/bit* measured in 10^{-7}. This implies that our algorithm has a much higher energy efficiency than the benchmark algorithm.

By fixing C_j as 10 Mbit/s, Fig. 4 shows the Cumulative Distribution Functions (CDFs) of the queue backlogs over all IoT gateways, at the 20th and 100th time slot, respectively. It can be seen that the average queue backlogs under the proposed algorithm are smaller than the ones indicated by the greedy algorithm. We also observe some interesting



(a) Backlogs in the 20th time slot. (b) Backlogs in the 100th time slot.

Fig. 4. Cumulative Distribution Function (CDF) of backlogs over all gateway queues at 2 different moments.

findings. The number of queues with empty-backlog increases drastically from 65% to 86% when system operates from the 20th to the 100th time slot, respectively. The reason is that the number of available time-varying uplinks increases during this period. Thus, the arrival IoT data caching in the queues of the connected IoT gateways can be uploaded quickly. In contrast, the backlogs in most of the queues keep growing under the benchmark greedy algorithm. Because only very few part of gateways can upload their data via the uplinks. This leads to that most of gateways have increasing backlogs.

In conclusion, the proposed online data gathering scheduling algorithm achieves larger throughput and higher power-consumption efficiency, and also yields significant smaller queue backlogs than those of the benchmark algorithm.

VI. Conclusion

In this paper, we studied how to upload IoT data from the geo-distributed networks by exploiting LEO based communication technology in an energy-efficient way. We proposed an online scheduling algorithm to address this problem. The simulation results show that the proposed algorithm can achieve much higher efficiency of power consumption, while maintaining significant lower queue backlogs in IoT gateways, compared with a greedy-based benchmark algorithm.

VII. Acknowledgement

This work is partially supported by Strategic Information and Communications R&D Promotion Programme (SCOPE No.162302008), MIC, Japan, and NSFC (61572262).

References

[1] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "Nb-iot system for m2m communication," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2016, pp. 1–5.

[2] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3gpp machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, 2011.

[3] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: The rising stars in the iot and smart city scenarios," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60–67, 2016.

[4] H. S. Dhillon, H. Huang, and H. Viswanathan, "Wide-area wireless communication challenges for the internet of things," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 168–174, 2017.

[5] R. Barbatei, A. Skavhaug, and T. A. Johansen, "Acquisition and relaying of data from a floating wireless sensor node using an unmanned aerial vehicle," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015, pp. 677–686.

[6] A. Zolich, J. A. Alfredsen, T. A. Johansen, and K. R. SkØien, "A communication bridge between underwater sensors and unmanned vehicles using a surface wireless sensor network-design and validation," in *OCEANS 2016-Shanghai*, 2016, pp. 1–9.

[7] M. De Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite communications supporting internet of remote things," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 113–123, 2016.

[8] [Online]. Available: https://www.inmarsat.com/service/isatdata-pro/

[9] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89–103, 2005.

[10] N. Accettura, M. R. Palattella, G. Boggia, L. A. Grieco, and M. Dohler, "Decentralized traffic aware scheduling for multi-hop low power lossy networks in the internet of things," in *IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2013, pp. 1–6.

[11] X. Luo, G. Liu, and H. Huang, "Buffer capacity-constrained epidemic routing model in mobile ad-hoc networks," in *IEEE 17th International Conference on Computational Science and Engineering (CSE)*, 2014, pp. 1443–1448.

[12] P. Gonizzi, G. Ferrari, V. Gay, and J. Leguay, "Data dissemination scheme for distributed storage for iot observation systems at large scale," *Information Fusion*, vol. 22, pp. 16–25, 2015.

[13] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[14] H. Huang, S. Guo, P. Li, and T. Miyazaki, "Deactivation-controlled epidemic routing in disruption tolerant networks with multiple sinks," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, 2014, pp. 4604–4609.

[15] H. Huang, S. Guo, P. Li, B. Ye, and I. Stojmenovic, "Joint optimization of rule placement and traffic engineering for qos provisioning in software defined network," *IEEE Transactions on Computers*, vol. 64, no. 12, pp. 3488–3499, 2015.

[16] C.-T. Cheng, N. Ganganath, and K.-Y. Fok, "Concurrent data collection trees for iot applications," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 793–799, 2017.

[17] K. Wang, Y. Wang, Y. Sun, S. Guo, and J. Wu, "Green industrial internet of things architecture: An energy-efficient perspective," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 48–54, 2016.

[18] H. Huang, S. Guo, J. Wu, and J. Li, "Green datapath for tcam-based software-defined networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 194–201, 2016.

[19] Y. Sun, X. Hu, X. Liu, X. He, and K. Wang, "A software-defined green framework for hybrid ev-charging networks," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 62–69, 2017.

[20] H. Huang, S. Guo, P. Li, and T. Miyazaki, "Stochastic analysis on the deactivation-controlled epidemic routing in dtns with multiple sinks," *Ad Hoc & Sensor Wireless Networks*, vol. 38, no. 1-4, pp. 143–167, 2017.

[21] H. Huang, S. Guo, and K. Wang, "Envision of wireless big data storage for low-earth-orbit satellite based cloud," *IEEE Wireless Communications*, pp. 1–6, 2018.

[22] H. Wu, J. Li, H. Lu, and P. Hong, "A two-layer caching model for content delivery services in satellite-terrestrial networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[23] X. Jia, T. Lv, F. He, and H. Huang, "Collaborative data downloading by using inter-satellite links in leo satellite networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1523–1532, 2017.

[24] D. Xu and X. Liu, "Geographic trough filling for internet datacenters," in *Proc. of IEEE Conference on Computer Communications (INFOCOM)*, 2012, pp. 2881–2885.

[25] J. Choi and V. Chan, "Predicting and adapting satellite channels with weather-induced impairments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 779–790, 2002.

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

[27] M. J. Neely, E. Modiano, and C. E. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Transactions on Networking (TON)*, vol. 11, no. 1, pp. 138–152, 2003.

[28] H. Huang, S. Guo, W. Liang, K. Wang, and A. Y. Zomaya, "Green online data-gathering from geo-distributed iot networks through low-earth-orbit satellites," in *IEEE Transactions on Wireless Communications (TWC), (under review)*, 2018, pp. 1–11.

[29] [Online]. Available: https://en.wikipedia.org/wiki/Systems_Tool_Kit/

[30] J. Wang, L. Li, and M. Zhou, "Topological dynamics characterization for leo satellite networks," *Computer Networks*, vol. 51, no. 1, pp. 43–53, 2007.