On the Scalability and Coverage of LoRa Mesh for Monitoring Linear Infrastructure

Yu Chen, *Member, IEEE,* Guo Shi, *Graduate Student Member, IEEE,* Yusuf Sambo, *Senior Member, IEEE,* Oluwakayode Onireti, *Senior Member, IEEE,* and Muhammad Imran, *Fellow, IEEE*

***Abstract*—This paper investigates the scalability and coverage of long range (LoRa) mesh networks deployed for monitoring linear infrastructure, such as railways and pipelines. Although the monitored infrastructure is linear, the network topology is a mesh. The study addresses the challenges posed by the scalability of such networks, considering constraints imposed by duty cycle regulations in license-free frequency bands, particularly within the European Union where a 1% duty cycle limitation exists. A deployment strategy is proposed to optimize the placement of LoRa mesh nodes along linear infrastructure. A comprehensive analysis of scalability and coverage yields their bounds, along with the condition necessary to achieve the upper bounds. To enhance scalability and coverage, a novel routing algorithm is proposed considering the number of hops and the received signal strength indicator. Additionally, a LoRa mesh simulator, LoRaMeshSim, is developed to validate the system analysis and evaluate the effectiveness of the proposed routing algorithm. The findings provide valuable insights into the practical deployment of LoRa mesh networks for monitoring linear infrastructure.**

***Index Terms*—LoRa mesh, scalability, coverage, routing algorithm, linear infrastructure monitoring.**

I. Introduction

THE efficient monitoring of linear infrastructure, which refers to extensive yet narrow installations that span considerable distances and follow a linear or elongated path, such as pipelines, roads, railways, mines, and international borders, plays a pivotal role in ensuring the safety, reliability, and optimal performance of these critical systems [1]. Moreover, enabled by remote monitoring, predictive maintenance has become a promising method for maintaining extensive infrastructure given the high cost of traditional periodic maintenance approaches. However, substantial parts of these infrastructures are located in remote areas without terrestrial network coverage which makes deployment slow and expensive [2]. Given the linear nature of these infrastructures, wireless networks with star topology are not well-suited for such monitoring. This is because deploying numerous gateways with backbone connections is necessary alongside the linear infrastructure, resulting in high deployment and management costs. Instead, mesh topology offers a promising method for linear infrastructure monitoring due to their multi-hop communication technologies [3].

The advent of low power wide area network technologies, exemplified by long range (LoRa) [4] communication, has revolutionized the field of remote monitoring. Given the low-power, long-range, and low-cost capabilities, LoRa is well-suited for applications in remote areas such as smart agriculture [5] and smart farming [6]. However, the star topology limits the use of LoRa in extensive linear infrastructure monitoring applications. In this context, the implementation of LoRa mesh [7] networks has emerged as a notable advancement. Unlike traditional single-hop LoRa setups, LoRa mesh offers extended coverage and improved reliability through its self-organizing capabilities, making it an appealing choice for large-scale linear infrastructure monitoring deployments. More significantly, LoRa mesh nodes can be deployed alongside linear infrastructure and form multi-hop communication paths, eliminating the need for densely deploying gateways and significantly reducing the deployment cost.

Monitoring extensive linear infrastructure requires numerous devices connected to a single network, which poses the network scalability challenge. In LoRa and LoRa mesh, scalability is constrained by 1) internal signal collisions, and 2) duty cycle regulations. The duty cycle of LoRa refers to the fraction of time a device is allowed to transmit in a given period, which is typically regulated to prevent network congestion and interference. For the first constraint, internal signal collisions refer to the collision caused by concurrent signal transmission from multiple nodes of the network. With the increase in the number of nodes, the possibility of internal signal collisions increases, thereby constraining network scalability [8] [9] [10]. This limitation can be eliminated or mitigated by properly scheduling data collection and transmission. In addition, signals from external devices may also collide with internal signals, as LoRa and LoRa mesh operate in license-free frequency bands. To mitigate interference among networks, regulators impose restrictions on the maximum duty cycle, i.e., the second constraint. For example, in Europe, LoRa operates in the frequency bands 433.05-434.79 MHz or 863-870 MHz, where European Telecommunications Standards Institute (ETSI) limits the duty cycle under 0.1%, 1%, or 10% [11] depending on the sub-band. To avoid network congestion, LoRa Alliance specifies 1% as the maximum duty cycle for LoRa in the bands where ETSI requires 10% [12]. These

(Corresponding author: Guo Shi).
Yu Chen, Yusuf Sambo, Oluwakayode Onireti, and Muhammad Imran are with the James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, UK (e-mail: {yu.chen.2, yusuf.sambo, oluwakayode.onireti, muhammad.imran}@glasgow.ac.uk).
Guo Shi is with both the Department of Electronic and Electrical Engineering and the Department of Management Science, University of Strathclyde, Glasgow, G4 0QU, UK (e-mail: guo.shi@strath.ac.uk).

duty cycle regulations impose limitations on the scalability of LoRa-based networks [13]. However, existing research on the scalability of LoRa mesh focuses on the constraint of internal signal collisions, ignoring the constraint of duty cycle regulations. On the other hand, since the coverage extension ratio of LoRa mesh, compared to LoRa, is related to the number of nodes that can be deployed in the network [14], duty cycle regulations also constrain the coverage extension ratio. Investigating the coverage extension ratio is also crucial for monitoring linear infrastructure, given the considerable length of the infrastructure.

In [15], we have integrated LoRa mesh with a 5G network for trackside weather monitoring, and proposed a cloud-edge-terminal collaborative architecture to reduce the network data volume. In this paper, we investigate the challenges associated with the scalability and coverage of LoRa mesh for linear monitoring.

Firstly, we propose a deployment strategy for a LoRa mesh network designed for monitoring linear infrastructure. The strategy not only ensures scalability and extensive coverage but also reduces deployment complexity by introducing deployment adaptability, allowing flexibility in the distances between proximate nodes. This eliminates the need for evenly spaced nodes, simplifying deployment in challenging terrains.

Secondly, we analyze the scalability and coverage of the LoRa mesh network. In terms of the constraints on scalability and coverage, we focus on duty cycle regulations instead of internal signal collisions. In terms of coverage, we focus on the coverage extension ratio of the LoRa mesh compared to single-hop LoRa instead of absolute coverage. We derive upper and lower bounds for scalability and coverage extension ratio. We observe that the values of these metrics depend on the routing algorithm.

Thirdly, we propose a novel routing algorithm that considers the minimum number of hops and the received signal strength indicator (RSSI) to achieve maximum scalability and coverage.

Lastly, with the lack of a suitable LoRa mesh simulator for the research community, we develop a simulator called LoRaMeshSim to verify the system analysis and the proposed routing algorithm.

The main contributions of this paper are listed below:

- A deployment strategy is proposed for LoRa mesh in the context of linear infrastructure monitoring. This strategy reduces deployment complexity by enabling variable node-to-node distances, simplifying placement in uneven terrain and constrained environments while ensuring the scalability and extended coverage of the network.
- We derive the upper and lower bounds of the scalability and coverage extension ratio and present the conditions for reaching the upper bounds.
- A novel routing algorithm that achieves the upper bounds of the network's scalability and coverage extension ratio is proposed.
- A LoRa mesh simulator is developed to verify the system analysis and the proposed routing algorithm. This simulator is expected to make a valuable contribution to the field of LoRa mesh.

The remainder of this paper is organized as follows: Section II provides an overview of related work in the field of linear infrastructure monitoring and LoRa-based technologies. Section III presents the system model of the LoRa mesh network and introduces the proposed deployment strategy for LoRa mesh. Section IV details the system analysis on scalability and coverage. In Section V, the novel routing algorithm is introduced. Section VI describes the validation using the LoRa mesh simulator. Finally, Section VII concludes the paper.

II. RELATED WORK

There has been extensive research on sensing and communication technologies for monitoring linear infrastructures, including roads, railway tracks, and pipelines.

Firstly, in the context of road monitoring, Benedetto *et al.* [1] provided a general overview of innovative remote sensing techniques, such as synthetic aperture radar, light detection and ranging, and mobile laser scanner, for assessing road safety. Guirado *et al.* [16] utilized a 4G-based unmanned aerial system to monitor road traffic and linear infrastructure. Liaqat *et al.* [17] evaluated gossip routing in LoRa-based networks for linear infrastructure monitoring.

Secondly, in the context of railway monitoring, Paula *et al.* [18] examined the evolution of railway communication technologies from the global system for mobile communications-railway to the long term evolution-railway. Their findings indicate a shift from traditional infrastructure to intelligent train management systems, which offer a range of Internet of things (IoT)-enabled services including predictive maintenance, intelligent train control, and enhanced infrastructure capabilities. Jo *et al.* [19] emphasized the need to integrate IoT with condition-based maintenance for effective railway monitoring. They illustrated that LoRa stands out as the optimal solution, considering factors such as coverage, power consumption, and implementation complexity.

Lastly, in the context of pipeline monitoring, Stoianov *et al.* [20] presented a prototype for water pipe monitoring based on a wireless sensor network, aiming to achieve real-time detection of leaks and blockages. Kurnaz *et al.* [21] investigated the potential of satellite remote sensing methods for linear system monitoring to assess the safety and security of the oil and gas transportation system. Narayana *et al.* [22] employed LoRa technology to monitor an oil pipeline and assessed its performance through simulation. Although various technologies have been employed to monitor linear infrastructure, challenges related to cost and coverage in remote areas remain unresolved.

As a promising technology to provide wide coverage in remote areas at low cost, LoRa mesh has received much attention from academia and industry. Lundell *et al.* [23] introduced a routing protocol designed for LoRa mesh networks, validating its effectiveness through both laboratory and field testing. Berto *et al.* [24] implemented a LoRa mesh network using the RadioHead packet radio library [25], a popular open-source LoRa mesh library tailored for embedded microprocessors with limited resources. Huh and Kim [14] proposed a LoRa mesh protocol, exploring its applications in

fire pipe freeze monitoring, street light smart control, and toxic gas monitoring. Ebi *et al.* [7] employed a LoRa mesh network for monitoring underground infrastructure, demonstrating its advantages over LoRa and LoRaWAN in terms of coverage and packet delivery reliability through field tests. Hong *et al.* [8] suggested a hierarchical-based energy-efficient routing protocol for LoRa mesh networks, showing superior energy efficiency and reduced transmission delay compared to traditional ad hoc on-demand distance vector routing methods. Tian *et al.* [10] developed LoRaHop, an add-on protocol compatible with LoRaWAN, showcasing its ability to extend coverage through a multi-hop mesh network. Their outdoor testbed evaluations demonstrated that LoRaHop enhances the coverage of a LoRaWAN network with improved reliability and reduced power consumption. Chen *et al.* [15] designed a LoRa mesh server to enhance network management by integrating LoRa mesh into a 5G network. The integrated network was applied to monitor the weather alongside railway tracks, a typical linear infrastructure. However, the scalability and coverage of the network were not analyzed.

For the practical deployment of mesh networks, substantial research has focused on analyzing their coverage and scalability. The coverage and reliability of packet delivery in LoRa mesh networks are evaluated in [7], [10], and [26] through real experiments. Shrestha and Xing [27] conducted a comparison of mesh topology in wireless sensor networks with others, such as star, tree, and cluster, focusing on key performance indicators including reliability, energy efficiency, scalability, and data latency. Despite substantial research on the scalability of LoRa, such as [28], [29], and [30], conducted since 2016, research specifically addressing the scalability of LoRa mesh has just started. We only found one paper focusing on addressing the scalability of LoRa mesh networks. In the paper [9], the authors proposed a self-organizing communication protocol, called CottonCandy, to mitigate internal packet collisions, achieving the scalability of up to a hundred nodes. However, the paper only focuses on the constraint of internal signal collisions, ignoring the constraint of duty cycle regulations.

III. SYSTEM MODEL

As shown in Fig. 1, a linear infrastructure can be modeled as a straight line with LoRa mesh nodes deployed alongside it. A LoRa mesh gateway is deployed in the middle of the line with N sensor nodes at both sides of the gateway. The i^{th} sensor node from left to right is denoted as s_i for $i \in [1, N] \cap \mathbb{Z}$ and s_{i+1} for $i \in [N+1, 2N] \cap \mathbb{Z}$, respectively. To simplify the notation in the following analysis, we denote the gateway as s_{N+1} even though it does not have sensors. Sensor nodes collect data and transmit it to the gateway by encapsulating it as LoRa mesh packets. The gateway has access to the Internet via different technologies such as Ethernet, WiFi, satellite communication, and cellular networks [15], [31]. When receiving LoRa mesh packets, the gateway uploads data to the cloud via the Internet for further processing.

Unlike single-hop LoRa networks, most of the sensor nodes are outside of the coverage of the gateway, thus requiring other sensor nodes to relay packets. The selection of a relayer node depends on the routing algorithm, while the number of relayer candidates for each node is determined by its distance from other nodes. We assume that nodes are placed in a two-dimensional plane with the gateway located at coordinates (0,0). The distance between s_i and s_{i+1} is denoted as l_i for $i \in [1, 2N] \cap \mathbb{Z}$. Then, each sensor node s_i is placed at coordinates $(\sum_{j=i}^{j=N} -l_j, 0)$ for $i \in [1, N] \cap \mathbb{Z}$ and coordinates $(\sum_{j=N+1}^{j=i-1} l_j, 0)$ for $i \in [N+2, 2N+1] \cap \mathbb{Z}$. The maximum communication distance of single-hop LoRa depends on the frequency band, the transmit power, the spreading factor (SF), and the terrain [32]. In this work, we consider a fixed value, r for the distance since our focus is on the coverage extension ratio of LoRa mesh compared to LoRa, instead of the absolute coverage. Traditional research methods of linear infrastructure monitoring usually assume that sensors are deployed evenly [33] [34], i.e., all the l_i equal to a fixed value. In a real deployment, achieving evenly spaced nodes is challenging due to practical constraints such as terrain and available deployment locations. To address these challenges, we propose a more adaptable deployment strategy by introducing flexibility in the node-to-node distances denoted by l_i, thereby simplifying deployment logistics and enhancing the network's adaptability to real-world conditions. Specifically, we define a positive integer, denoted as the distance factor ϕ. With this parameter, we specify flexible deployment distances as follows:

$$l_i \in \left(\frac{r}{\phi + 1}, \frac{r}{\phi} \right) \text{ for } i \in [1, 2N] \cap \mathbb{Z}, \tag{1}$$

The parameter ϕ plays a pivotal role in defining the spacing between two proximate nodes, thus directly shaping the overall deployment density of the network. Since the system model is symmetric, we will only consider the sensor nodes on the left side of the gateway in the following analysis. The results will also be applicable to the sensor nodes on the right side of the gateway. Let $l_{i,j}$ denote the distance between s_i and s_j, i.e., $l_{i,j} = \sum_{\mu=i}^{\mu=j-1} l_\mu$, where $i, j \in [1, N+1] \cap \mathbb{Z}$. Then, according to (1),

$$l_{i,j} \begin{cases} < r, & \text{for } j - i \in [1, \phi] \\ > r, & \text{for } j - i > \phi. \end{cases} \tag{2}$$

Thus, $\forall i, j \in [1, N+1] \cap \mathbb{Z}$, s_i and s_j can directly communicate with each other only when $|j - i| \in [1, \phi]$. For instance, if $\phi = 1$, the distance between two proximate nodes can be any value between $r/2$ and r. s_1 can directly communicate with s_2 as $l_{1,2} = l_1 < r$. s_1 cannot directly communicate with s_3 as $l_{1,3} = l_1 + l_2 > r$. A direct link exists between two nodes if they can communicate directly with each other. All direct links and the network topology are illustrated in Fig. 2. By comparing Fig. 2 with Fig. 1, it becomes clear that while the monitored infrastructure is linear, the network topology is a mesh.

The sensor nodes that can directly communicate with the gateway are referred to as adjacent nodes. According to (2), the number of adjacent nodes on the left side of the gateway, consisting of s_i for $i \in [N - \phi + 1, N] \cap \mathbb{Z}$, equals the distance factor, ϕ. All the sensor nodes outside of the coverage of the gateway, i.e., s_i for $i \in [1, N - \phi] \cap \mathbb{Z}$, are referred to as remote nodes and need adjacent nodes to relay data packets. Based

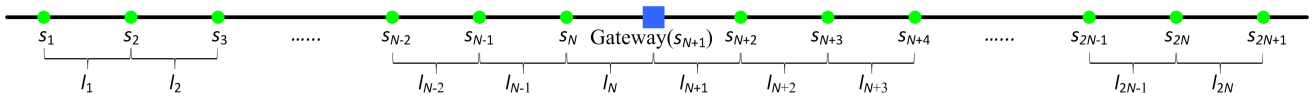

Fig. 1. System model of LoRa mesh network on linear infrastructure.

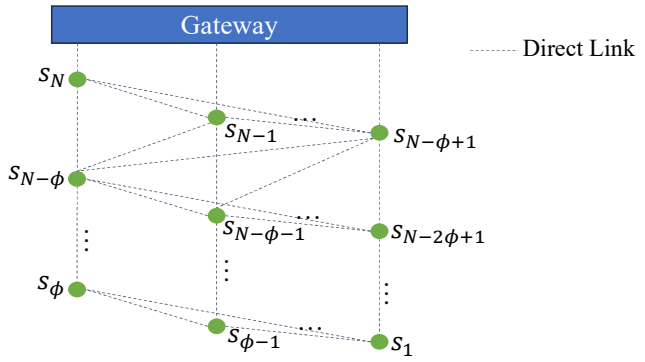

Fig. 2. Network topology and direct links.

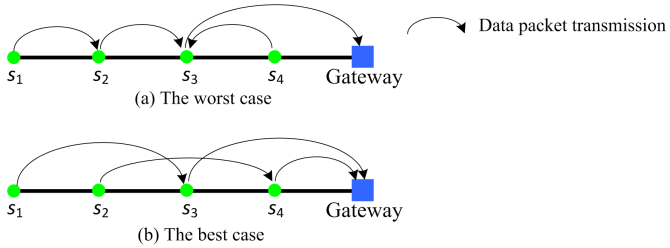

Fig. 3. Examples of the worst and best cases in terms of duty cycle with $\phi = 2$ and $N = 4$

on (2), the sub-network composed of $s_1, s_2, ..., s_{N+1}$ can be denoted as a graph $G = (S, E)$, where $S = s_1, s_2, ..., s_{N+1}$ and $E = \{(s_i, s_j) \mid s_i, s_j \in S \text{ and } j - i \in [1, \phi]\}$.

IV. SYSTEM ANALYSIS

In this section, we will analyze the scalability and coverage extension ratio.

A. Scalability

In this subsection, we will investigate the impact of duty cycle regulations on the scalability of the LoRa mesh network—specifically, examining the maximum number of sensor nodes that the network can support under the set duty cycle requirement.

Mesh networks require an acknowledgment message for each data message as each node needs to know the validity of its routes to the destination nodes. We assume that all the data messages have the same packet length denoted as L_d bytes, all the acknowledgment messages have the same packet length denoted as L_a bytes, all the sensor nodes transmit data at the same packet rate denoted as p packets per second per node, and each node selects only one route to transmit a data packet to the gateway. With these assumptions, the duty cycle of sensor node s_i can be expressed as

$$d_i = (n_i + 1)pt_d + n_i pt_a, \tag{3}$$

where n_i is the number of sensor nodes whose data packets are relayed by s_i. Here, the term 'relayed' includes not only the initial relay but also subsequent relays, given that a data packet may need to be relayed multiple times before reaching the gateway. t_d and t_a are the time on air of each data packet and each acknowledgment packet, respectively, which can be calculated by [35]

$$t_d = (\Omega + 4.25 + 8 + n_d^p)\frac{2^{SF}}{BW}, \tag{4}$$

$$t_a = (\Omega + 4.25 + 8 + n_a^p)\frac{2^{SF}}{BW}, \tag{5}$$

where

$$n_d^p = \max\left(\left\lceil \frac{8L_d - 4SF + 28 + 16 - 20H}{4(SF - 2\Psi)} \right\rceil (CR + 4), 0\right), \tag{6}$$

$$n_a^p = \max\left(\left\lceil \frac{8L_a - 4SF + 28 + 16 - 20H}{4(SF - 2\Psi)} \right\rceil (CR + 4), 0\right), \tag{7}$$

where $\lceil \cdot \rceil$ is the ceiling function. The explanation of the variables in (4), (5), (6), and (7) is listed:

- Ω is the number of programmed preamble symbols.
- SF is the spread factor from 7 to 12.
- BW is the channel bandwidth.
- $H = 1$ indicates the LoRa header is explicit and 0 otherwise.
- $\Psi = 1$ indicates low data rate optimize can be enabled and 0 otherwise.
- CR is the coding rate from 1 to 4.

According to the duty cycle regulation, $\forall i \in [1, N] \cap \mathbb{Z}, d_i \leq D$, where D is the maximum duty cycle requirement. Combining (3), to comply with the regulation, the following must hold

$$n_i \leq \frac{D - pt_d}{p(t_d + t_a)}, \forall i \in [1, N] \cap \mathbb{Z}. \tag{8}$$

To relieve the relaying burden, we prevent sensor nodes on one side of the gateway from relaying packets from the sensor nodes on the other side by allocating different identification addresses to them. Thus, $n_i \leq N - 1$. Since d_i is positively correlated with n_i according to (3), $\exists i \in [1, N] \cap \mathbb{Z}, n_i = N - 1$ is the worst case where all the packets on the left side are relayed by one sensor node. An example is provided to illustrate the worst case. As shown in Fig. 3(a), when $\phi = 2$ and $N = 4$, s_3 and s_4 can communicate with the gateway directly while s_1 and s_2 cannot. In the worst case, either s_3 or s_4 would relay packets for all the other sensor nodes on the left side. If s_3 is the relayer, it must handle packets from s_1, s_2 and s_4. Consequently, $n_3 = 3 = N - 1$.

On the other hand, comparing the duty cycle of adjacent nodes and remote nodes, the following lemma holds for the sensor nodes in the LoRa mesh network:

Lemma 1: Among the sensor nodes, the node with the highest duty cycle, when compared to all other sensor nodes, is an adjacent node.

Proof: See Appendix. Based on Lemma 1, the best case is that the ϕ adjacent nodes evenly share the relay tasks for the $N-\phi$ remote nodes, which can be formulated as $n_i = (N-\phi)/\phi = N/\phi - 1$ for $i \in [N-\phi+1, N] \cap \mathbb{Z}$. Thus, in the best case, $n_i \leq N/\phi - 1$ for $\forall i \in [1, N] \cap \mathbb{Z}$. We use the same network setting from the worst-case example ($\phi = 2$ and $N = 4$) to illustrate the best-case scenario. As shown in Fig. 3(b), if s_3 relays packets for s_1 and s_4 relays packets for s_2, s_3 and s_4 evenly share the relay tasks. In this case, $n_1 = n_2 = 0$, $n_3 = n_4 = 1$, and $n_1, n_2, n_3, n_4 \leq N/\phi - 1$. Given the worst and best cases, we can derive

$$\begin{cases} \forall i \in [1, N] \cap \mathbb{Z}, & n_i \leq \frac{N}{\theta} - 1 \\ \exists i \in [1, N] \cap \mathbb{Z}, & n_i = \frac{N}{\theta} - 1, \end{cases} \tag{9}$$

where $\theta \in [1, \phi]$. $\theta = 1$ indicates that the network is in the worst status while $\theta = \phi$ indicates the best status in terms of duty cycle. The value of θ is determined by the routing algorithm which will be discussed in Section V. If N is configurable, to satisfy (8), the following must hold

$$\begin{aligned} \frac{N}{\theta} - 1 &\leq \frac{D - pt_d}{p(t_d + t_a)}, \\ N &\leq \theta \frac{D + pt_a}{p(t_d + t_a)}, \theta \in [1, \phi]. \end{aligned} \tag{10}$$

In terms of the gateway, although it does not send any data packets, it sends out an acknowledgment packet for each data packet from any sensor node. Thus, its duty cycle $d_{\mathrm{G}} = 2Npt_a$. As the gateway also complies with the duty cycle regulation,

$$\begin{aligned} 2Npt_a &\leq D, \\ N &\leq \frac{D}{2pt_a}. \end{aligned} \tag{11}$$

Combining (10) and (11),

$$N \leq \min\left(\frac{D}{2pt_a}, \theta \frac{D + pt_a}{p(t_d + t_a)}\right), \theta \in [1, \phi]. \tag{12}$$

As N is an integer, we derive the maximum number of sensor nodes on the left side of the gateway as

$$N_{\max} = \min\left(\left\lfloor \frac{D}{2pt_a} \right\rfloor, \left\lfloor \theta \frac{D + pt_a}{p(t_d + t_a)} \right\rfloor\right), \theta \in [1, \phi], \tag{13}$$

where $\lfloor \cdot \rfloor$ is the floor function. Since the network is symmetric, the maximum total number of sensor nodes is $2N_{\max}$. As indicated in (13), ϕ determines the upper bound of $N_{\max}$, and when $\theta = \phi$, the network can reach this upper bound.

B. Coverage Extension Ratio

In addition to scalability, coverage is another important aspect of the LoRa mesh network. For the linear LoRa mesh network, we use the coverage extension ratio as the metric which is defined as

$$c = \frac{l_{1,N+1}}{r} = \frac{\sum_{i=1}^{N} l_i}{r}. \tag{14}$$

As $l_i < r/\phi$ for $\forall i \in [1, N] \cap \mathbb{Z}$,

$$c < \frac{\sum_{i=1}^{N} \frac{r}{\phi}}{r} = \frac{N\frac{r}{\phi}}{r} = \frac{N}{\phi} \leq \frac{N_{\max}}{\phi}. \tag{15}$$

Thus, combining (13) and (15),

$$c < \min\left(\frac{1}{\phi}\left\lfloor \frac{D}{2pt_a} \right\rfloor, \frac{1}{\phi}\left\lfloor \theta \frac{D + pt_a}{p(t_d + t_a)} \right\rfloor\right), \theta \in [1, \phi]. \tag{16}$$

Similar to scalability, ϕ determines the upper bound of c, and when $\theta = \phi$, the network can reach this upper bound.

V. ROUTING ALGORITHM

In this section, we will first derive the network topology required for achieving the upper bounds of $N_{\max}$ and c. Subsequently, we will present a routing algorithm designed to achieve this network topology.

A. Required Topology

As indicated by (13) and (16), both the maximum number of sensor nodes and the maximum coverage extension ratio are functions of θ. The upper bounds for the number of sensor nodes and coverage extension ratio can be obtained when $\theta = \phi$. As we analyzed, $\theta = \phi$ when ϕ adjacent nodes evenly share the relay tasks for $N - \phi$ remote nodes. Routing algorithms determine the relay relationship and enable self-organization in the topology of a spanning tree of the graph G [9]. If a node in a tree is designated as the root, the nodes in the tree are hierarchical based on the length of the path from the node to the root, which is termed as the level. The root is at level zero, while nodes directly connected to the root are at level one, nodes connected to them are at level two, and so on. We will discuss the topology required for $\theta = \phi$ level by level.

First, as illustrated in Fig. 4 (a), the gateway is designated as the root. At level one, the ϕ adjacent nodes, i.e., s_i for $i \in [N-\phi+1, N] \cap \mathbb{Z}$, serve as the children of the gateway and each adjacent node becomes the root of a sub-tree with N/ϕ nodes. This arrangement is necessary because each adjacent node needs to relay data packets from $N/\phi - 1$ remote nodes to achieve $\theta = \phi$.

Second, at level two, as illustrated in Fig. 4 (b), the candidates include s_i for $i \in [N - 2\phi + 1, N - \phi] \cap \mathbb{Z}$, determined based on their distances to the nodes at level one. Each node at level one must select at least one node as its child until the number of its descendants reaches $N/\phi - 1$. Consequently, among these candidates, $s_{N-\phi}$ must serve as the child of s_N as only $s_{N-\phi}$ can communicate with s_N directly. Once $s_{N-\phi}$ is selected, the process repeats, and $s_{N-\phi-1}$ becomes the child of s_{N-1} as only $s_{N-\phi-1}$ in unselected candidates can communicate with s_{N-1} directly. This derivation continues iteratively, assigning each candidate a parent node, i.e., s_i as the parent node of $s_{i-\phi}$ for $i \in [N - \phi + 1, N] \cap \mathbb{Z}$. The result is illustrated in Fig. 4 (b) where the solid lines denote

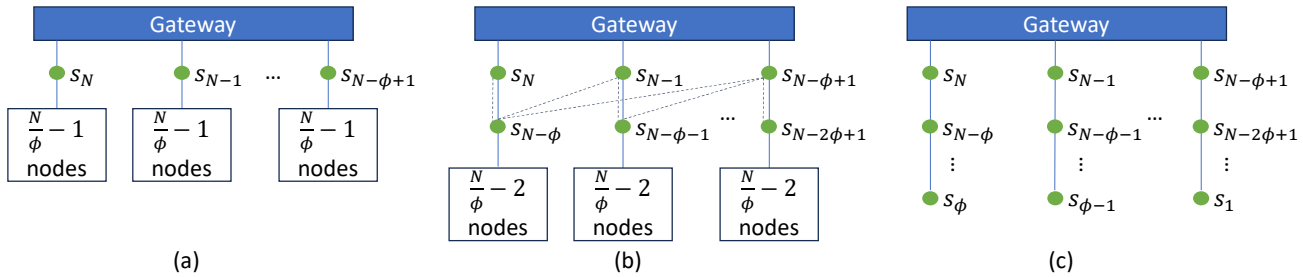

Fig. 4. Network topology derivation process and routing discovery process

selection and the dashed lines denote that the two nodes can communicate with each other directly.

Finally, by iteratively progressing from level one to level two, and continuing this process until reaching from level $N/\phi-1$ to level N/ϕ, we can derive the complete topology of the spanning tree, denoted as $T=(S,E_t)$ where $S=s_1,s_2,...,s_{N+1}$ and $E_t=E'_t\cup E''_t$. $E'_t=\{(s_i,s_j)\mid s_i,s_j\in S$ and $j-i=\phi$ for $j\leq N\}$ and $E''_t=\{(s_i,s_j)\mid s_i,s_j\in S$ and $j-i\leq\phi$ for $j=N+1\}$. As illustrated in the Fig. 4 (c), for $i\in[N-\phi+1,N]\cap\mathbb{Z}$, s_i relays data packets from $N/\phi-1$ remote nodes, which consists of $s_{i-j\phi}$ for $j\in[1,N/\phi-1]\cap\mathbb{Z}$. Thus, the topology achieves the requirement of $\theta=\phi$. Furthermore, it is evident from the derivation process that the topology is uniquely determined, signifying that there exists only one possible topology, i.e., T, for $\theta=\phi$.

B. Routing

To achieve the topology of T, we propose a routing algorithm based on RSSI and the number of hops. Since T is a spanning tree, each sensor node only has one parent node towards the gateway. So, the aim of the routing algorithm is to equip each sensor node with a routing table that contains only one piece of information, i.e., the identifier of the next-hop node towards the gateway. The routing process consists of two stages, i.e., routing request and routing discovery.

At stage one, when a node, called an initializer, wants to join the network or its routing table is no longer valid (known by acknowledged packets), it broadcasts a routing request to the network. Any node in the network receiving this request unicasts it to the gateway using the routing table of the node. Since multiple nodes may receive the broadcasted request, the gateway would receive the request from the same initializer multiple times. However, the gateway only needs to respond once by initializing a routing discovery process.

At stage two, after receiving a routing request, the gateway broadcasts a routing discovery message with the identifier of the initializer. When a sensor node receives the message, it checks whether a routing discovery message for the same initializer has been received. If yes and the number of hops of the new message is bigger than the previous one, it discards the message. Otherwise, it updates its routing table using the identifier of the sender of the message, rebroadcasts the message, and records the message for future route checking purpose. However, if the number of hops of the new routing discovery message is equal to the previous one, the node compares their RSSI and chooses the worst one. The worst RSSI in this scenario would typically imply the longest distance between the transmitter and the receiver as RSSI is negatively correlated with the communication distance [36].

We use Fig. 4 again to illustrate the broadcasting process of the routing discovery message as it is similar to the network topology derivation process. As shown in Fig. 4 (a), when the gateway broadcasts the routing discovery message, adjacent nodes receive it and update their routing tables. Then, as shown in Fig. 4 (b), the adjacent nodes rebroadcast the message, and the nodes at level two receive them. Some nodes at level two receive the message multiple times with the same number of hops, but s_i chooses $s_{i+\phi}$ for $i\in[N-2\phi+1,N-\phi]\cap\mathbb{Z}$ based on the criterion of worst RSSI. In addition to the nodes at level two, nodes at the lower levels also receive the rebroadcasted message, but they discard it as the number of hops is bigger than the message they received before. By doing so, the nodes at level two have the correct routing tables specified by the topology of T. Finally, by iteratively progressing from level one to level two, the whole network shapes the topology of T.

Although signal collisions are out of the scope of this paper, we must address signal collisions arising from broadcasts during the routing process, as they significantly affect the success rate of routing. When a node broadcasts a message, multiple nodes may receive it at the same time. If they relay it immediately after receiving it, any node within the coverage of two or more transmitters cannot receive any messages due to signal collisions. In the proposed routing algorithm, there are two kinds of broadcasts, i.e., initial routing request broadcast and routing discovery broadcast. We propose to address their signal collisions using random delay. When a node receives a broadcasted routing request message, it delays unicasting the message by it_r, where t_r denotes the time on air of the routing request message and i is randomly chosen in $[0,X]\cap\mathbb{Z}$. When a node receives a routing discovery message, it delays rebroadcasting the message by it_c, where t_c denotes the time on air of the routing discovery message and i is randomly chosen in $[0,Y]\cap\mathbb{Z}$. The determination of the values of the two integers, X and Y, referred to as the maximum delay integer for routing request and routing discovery, respectively, will be discussed in Section V through experiments. The routing algorithm is outlined in Algorithm 1.

Algorithm 1 Routing Algorithm

```
IF the node is initializer
  WHEN joining the network or routing table no longer valid
    WHEN routing discovery not received
      Broadcast routing request
      Wait for routing discovery message
    Update routing table
ELSE IF the node is the gateway
  WHEN routing request received
    IF not received from the initializer before
      Record the message
      Delay by (X + N)t_r
      Broadcast a routing discovery
ELSE
  WHEN routing request received
    IF it is a broadcast
      Delay by it_r, i randomly chosen in [0, X]
    Unicast to the next hop
  WHEN routing discovery received
    IF received before
      IF new hops < old hops
        Delay by it_c, i randomly chosen in [0, Y]
        Update routing table
        Record the message
        The hops of the message + 1
        Rebroadcast
      ELSE IF new hops = old hops
        IF new RSSI < old RSSI
          Delay by it_c, i randomly chosen in [0, Y]
          Update routing table
          Record the message
          The hops of the message + 1
          Rebroadcast
```

VI. VERIFICATION

Simulation is important for analyzing LoRa mesh performance and developing routing algorithms. However, despite the existence of several LoRa mesh simulators, such as NS-3 [37] [38] and LoRaSim [28] [39], there is currently no dedicated LoRa mesh simulator available. Thus, to validate our analysis and the proposed routing algorithm and to support the research and development of LoRa mesh, we develop a LoRa mesh simulator called LoRaMeshSim[1] based on LoRaSim and SimPy (a discrete event simulator) [40]. In LoRaMeshSim, in accordance with the deployment strategy formulated in (1), N independent random numbers are generated for l_i where $i \in [1, N] \cap \mathbb{Z}$, utilizing a uniform distribution with the open interval $(\frac{r}{\phi+1}, \frac{r}{\phi})$. All the nodes including sensor nodes and the gateway are activated at the same time. After activation, each sensor node independently collects data continuously at intervals that follow an exponential distribution with a rate parameter equal to the packet rate, p. The data is then transmitted to the gateway in the form of a data packet if the sensor node has a valid routing table. Otherwise, the sensor node initializes a routing process by broadcasting a routing request to the network. Although we do not consider signal collisions (except the ones caused by broadcast) in this paper, they are simulated in the LoRaMeshSim. Specifically, in the simulator, a node cannot transmit or receive a packet if it is transmitting or receiving another packet. When multiple signals persist simultaneously in the surroundings of a node for a while, it cannot receive any of them. By doing so, a more practical LoRa mesh network is simulated. When signal collisions occur, the data packet will not be retransmitted in the following experiments. We will use LoRaMeshSim to validate the following aspects of the network in this section: 1) routing success rate and 2) scalability and coverage extension ratio. The values of the key parameters used in the simulations are listed in Table I. To assist readers in calculating N_{max} and c in the following experiments, we drive t_a and t_d from other parameters based on equations (4)-(7). Multiple values will be used for L_d and ϕ. Specifically, $t_d = 71.936, 97.536, 112.896, 158.976,$ and 246.016 correspond to $L_d = 30, 50, 60, 90,$ and 150, respectively.

[1]Available at https://github.com/YuChenUoG/LoRaMeshSim

TABLE I
THE VALUES OF THE KEY PARAMETERS IN THE SIMULATIONS.

Parameter	Value	Derived	Fixed
simulation time	24 (hours)	×	✓
L_a	5 (bytes)	×	✓
Ω	8	×	✓
SF	7	×	✓
BW	125 (kHz)	×	✓
H	0	×	✓
Ψ	0	×	✓
CR	1	×	✓
L_d	30, 50, 60, 90, 150 (bytes)	×	×
ϕ	1, 2, 3, 4, 5	×	×
t_a	30.976 (ms)	✓	✓
t_d	71.936, 97.536, 112.896, 158.976, 246.016 (ms)	✓	×

A. Routing Success Rate Investigation

The routing process consists of the routing request and the routing discovery. To investigate their success rates affected by the signal collisions arising from packet broadcast, sensor nodes are not required to collect data or send data packets.

To investigate the routing discovery success rate, after placing all the nodes, we trigger the routing discovery process from the gateway assuming that the gateway receives a routing request successfully. Then, we determine if the routing discovery is successful by comparing the routing tables of all the nodes with the ideal topology of T. We vary values for ϕ, N, and maximum delay integer for routing discovery Y, and repeat the simulation 10,000 times for each set of these parameters to obtain the success rate. As shown in Fig. 5, when ϕ is fixed at 2 and N takes values in the set $\{2, 6, 10, 14, 18\}$, the success rate increases either with the increase of Y or with the decrease of N. When $N = 2$, the success rate is always 100% regardless of the rebroadcast delay. The reason is that rebroadcast is not required as all the sensor nodes can receive the broadcasted routing discovery message from the gateway directly when $N = 2$ and $\phi = 2$. As shown in Fig. 6, when N is fixed at 8 and ϕ takes values in the set

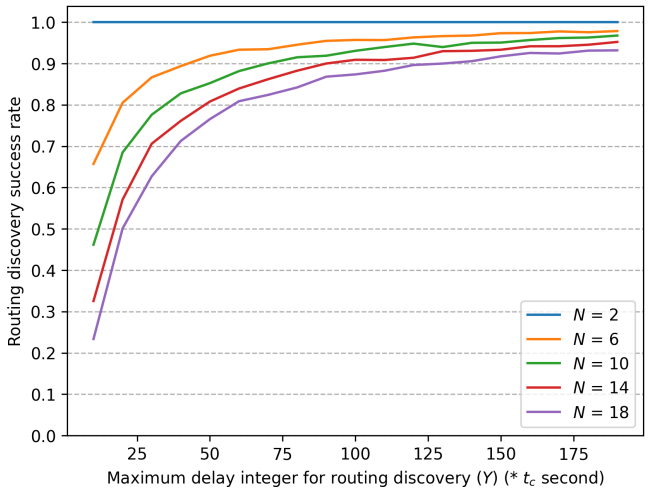

Fig. 5. Routing discovery success rate with different N when $\phi = 2$

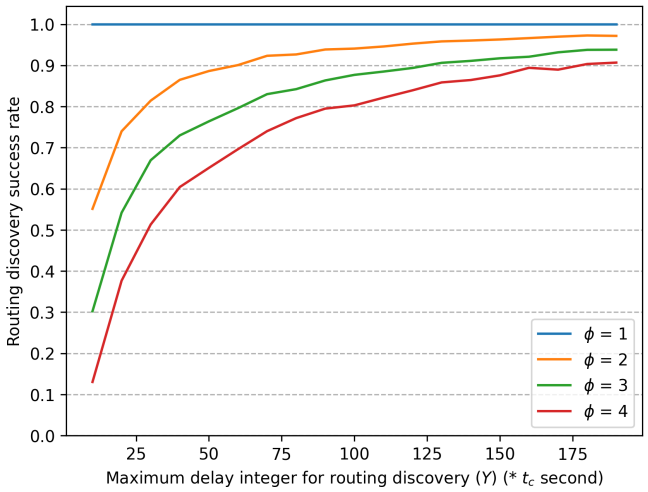

Fig. 6. Routing discovery success rate with different ϕ when $N = 8$

$\{1, 2, 3, 4\}$, the success rate increases either with the increase of Y or with the decrease of ϕ. When $\phi = 1$, the success rate is always 100% regardless of the rebroadcast delay. The reason is that only one node broadcasts or rebroadcasts a routing discovery message at a time when $\phi = 1$. This is because only one sensor node receives each broadcasted or rebroadcasted routing discovery message, except for another sensor node that has previously received a discovery message with fewer hops and thus discards the current one. In both Fig. 5 and 6, with the increase of Y, the success rate exceeds 90%, a high success rate. It can be observed that the worst case is when $\phi = 4$ and $N = 8$. In this case, the minimum Y to achieve the success rate of 90% is around 175 corresponding to a delay of $175t_c$ seconds. The routing discovery packet contains four pieces of information: the initializer identifier, the sender identifier, the number of hops, and the message identifier. If adopting the size allocation method of RadioHead, each piece of information occupies one-byte space. Then, based on (4) and (6), $t_c = 0.036$ seconds. Thus, $175t_c = 6.318$ seconds which is the maximum delay to obtain a high routing discovery success rate in the worst case.

To trigger a routing discovery process, the gateway needs to receive at least one routing request message. Thus, a routing request process is considered successful if the gateway receives at least one routing request message from the initializer after the initialization of the routing request process. To investigate the routing request success rate, after placing all the nodes, we delete a node and trigger a routing discovery process from the gateway for the left nodes. Note that the topology of the network cannot be the topology of T since one node is deleted. Thus, in this case, the routing discovery process is considered successful if the routing tables of all the sensor nodes follow the rules of minimum hops and worst RSSI which we discussed in Section IV. After the routing discovery process, the deleted sensor node is placed back at its previous location and initializes a routing request process. Then, we observe the gateway to determine if the routing request process is successful. In this simulation, N is fixed at 20 since the

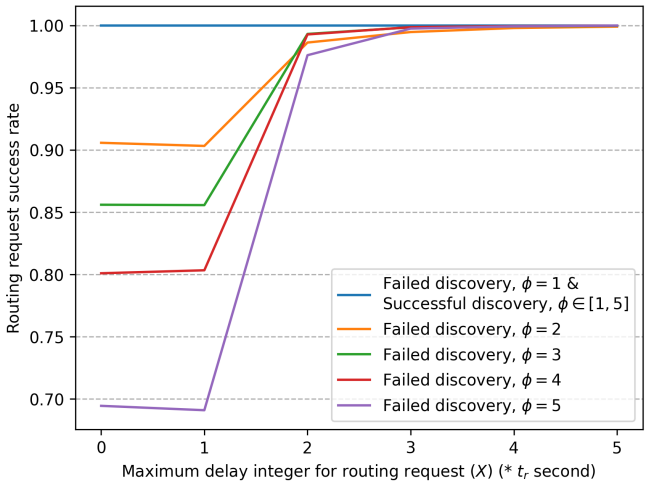

Fig. 7. Routing request success rate with different ϕ when $N = 20$

signal collision arising from the routing request broadcast is only relevant to the nodes receiving the broadcast request and is independent of the total number of sensor nodes. We vary values for ϕ and X, and repeat the simulation 50,000 times for each set of these parameters to obtain the routing request success rate. The results are classified into two groups according to if the routing discovery is successful. As shown in Fig. 7, if the routing discovery is successful, the routing request success rate is 100% regardless of the maximum delay integer for routing request X and ϕ. If the routing discovery fails, the routing request success rate decreases with the increase of ϕ. For $\phi = 1$, the routing request success rate is 100% regardless of the random delay value. For $\phi \in \{2, 3, 4, 5\}$, the routing request success rate increases with the increase of X. When $X \geq 3$, the routing request success rate approaches 100% regardless of ϕ and the routing discovery result. Thus, a very small random delay can resolve the signal collision arising from the routing request broadcast.

Note that we only consider the case with successful routing discovery in the following subsection.

B. Scalability and Coverage Extension Ratio Verification

In this subsection, we will evaluate the duty cycle of the LoRa mesh network, and verify the performance of the proposed routing algorithm and the bounds of scalability and coverage extension ratio.

1) Duty Cycle Evaluation: Before demonstrating the scalability and coverage extension ratio of the network, we utilize LoRaMeshSim to analyse the duty cycle of each node and evaluate how it affects the scalability and coverage extension ratio. The parameters ϕ and L_d are fixed at 2 and 50, respectively. We vary values for N and conduct the simulation only once, with a duration of 24 hours for each value of N. We calculate the duty cycle of each node and the results are shown in Fig. 8. The simulation validates Lemma 1, i.e., one of the adjacent nodes has the maximum duty cycle in the sensor nodes. Moreover, either the gateway or an adjacent node has the maximum duty cycle in the network. Since the proposed routing algorithm evenly allocates the relay burdens to the two adjacent nodes, their duty cycles are close to each other. With the increase of N, the maximum duty cycle of the network increases. It does not exceed 1.0% until N exceeds 14. Thus, to comply with the duty cycle regulation, the maximum number of sensor nodes is 14, aligning with the theoretical $N_{\max}$ given in (13). By contrast, using Andrei Broder and David Alduous algorithm [41] [42], a random spanning tree of the graph G is generated as the topology of the network to replace the proposed routing algorithm. As shown in Fig. 9, we repeat the simulation with the topology of a random spanning tree. In this case, Lemma 1 still holds. However, the gap between the duty cycles of the two adjacent nodes significantly expands due to the unevenly allocated relay burden. Consequently, there is a significant increase in the duty cycle of an adjacent node, resulting in a significant increase in the maximum duty cycle of the network. To comply with the duty cycle regulation, the maximum number of sensor nodes drops to 7. In the following simulations on scalability, we determine $N_{\max}$ by conducting an experiment with $N = 2$ and then increasing N until the maximum duty cycle exceeds 1%.

To illustrate the coverage extension ratio, we repeat the simulation of Fig. 8 100 times, given that the sensor nodes are randomly located. To reduce the number of repeats required to find the maximum coverage extension ratio, c, we employ beta distribution with parameters $\alpha = 0.005$ and $\beta = 0.018$, instead of uniform distribution, to place the sensor nodes. As shown in Fig. 10, c varies in a range for a specific N. The maximum c increases with the increase of N and the maximum duty cycle does not exceed 1.0% until N exceeds 14. Thus, the maximum c of the network is the maximum c when $N = 14$. The value of the maximum c aligns with the theoretical maximum c given in (16). In the following simulations on the coverage extension ratio, we will use the same beta distribution for sensor node placement to reduce the number of repeated experiments. Moreover, unlike in this simulation, we determine the maximum c by first identifying $N_{\max}$ and then repeating the experiment only for $N = N_{\max}$.

2) Routing Performance Verification: In the simulation, we will verify that the proposed routing algorithm enables the

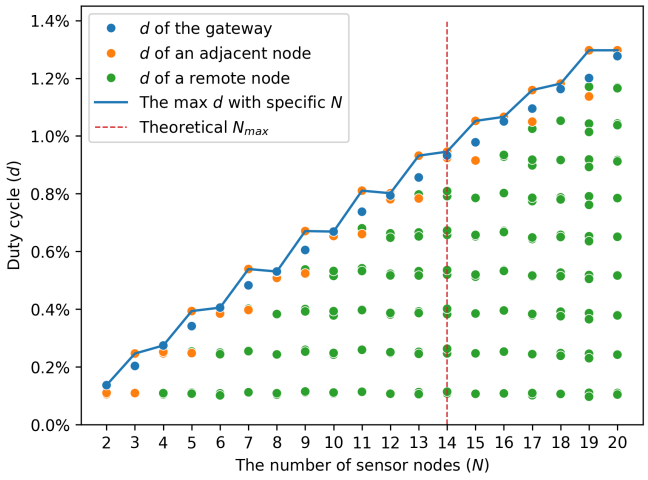

Fig. 8. Duty cycle and the maximum number of sensor nodes of the proposed routing algorithm

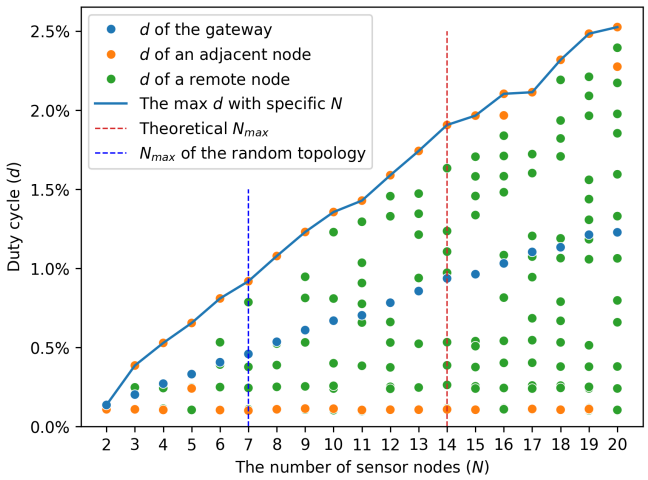

Fig. 9. Duty cycle and the maximum number of sensor nodes of a random spanning tree

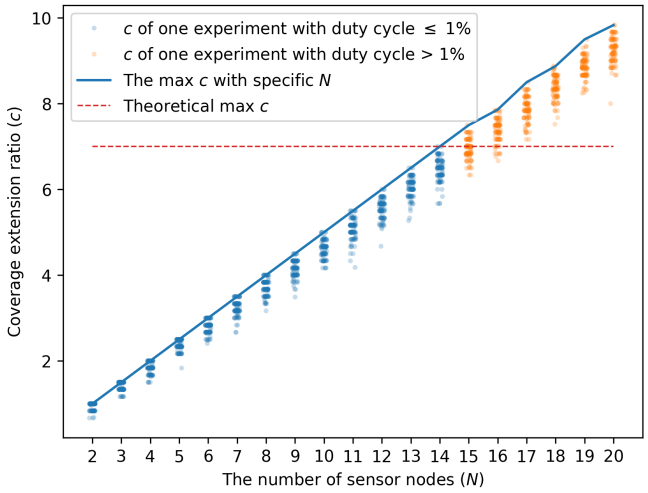

Fig. 10. Duty cycle and the coverage extension ratio of the proposed routing algorithm

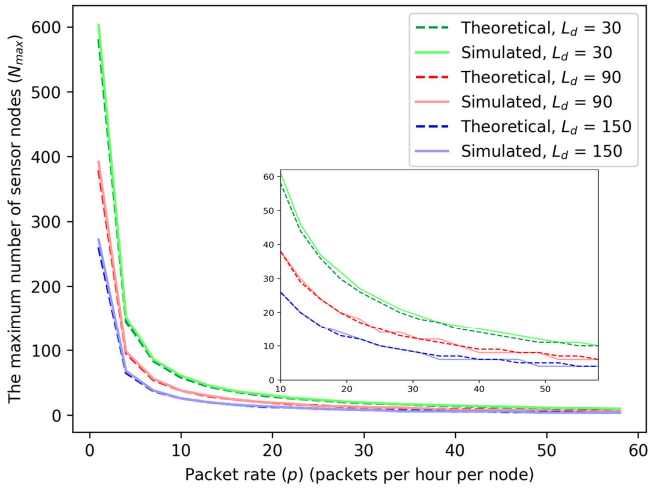

Fig. 11. The maximum number of sensor nodes with different L_d when $\phi = 2$

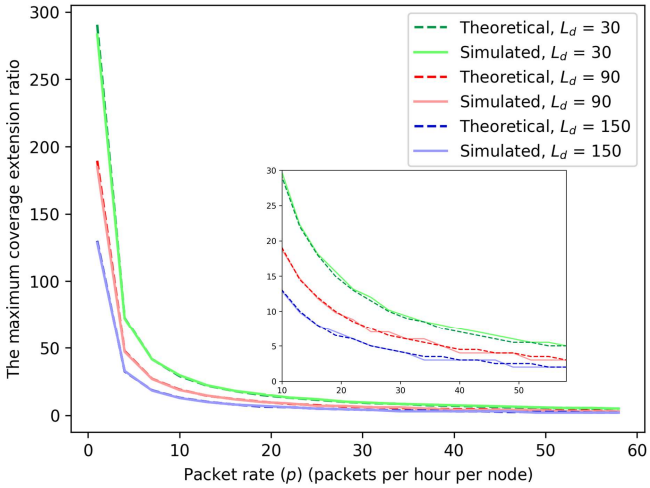

Fig. 13. The maximum coverage extension ratio with different L_d when $\phi = 2$

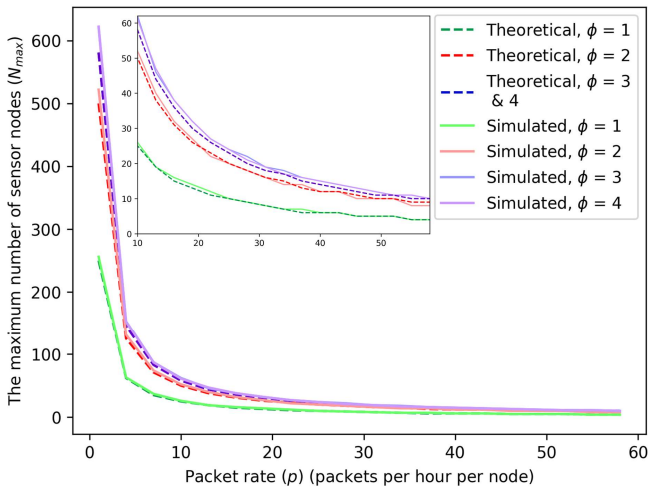

Fig. 12. The maximum number of sensor nodes with different ϕ when $L_d = 60$

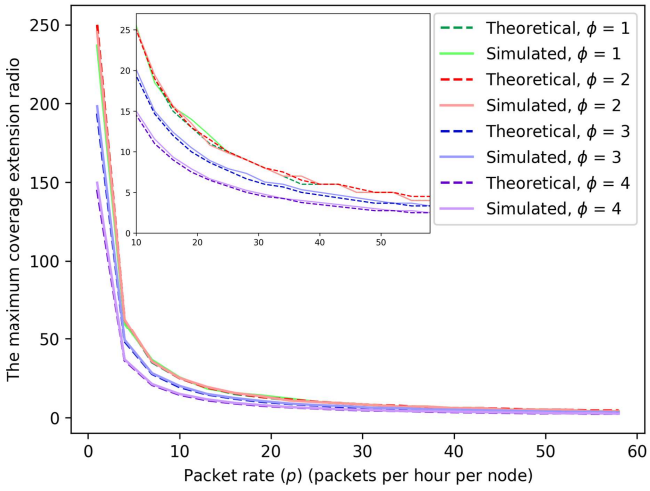

Fig. 14. The maximum coverage extension ratio with different ϕ when $L_d = 60$

network to reach the upper bounds of N and c. To obtain simulated $N_{\max}$, we vary values for p, L_d, and ϕ, and conduct simulations by increasing N for each set of these parameters. As shown in Fig. 11, when ϕ is fixed at 2, $N_{\max}$ decreases either with the increase of p or with the increase of L_d. There is a sharp drop of $N_{\max}$ when p increases from 1 to 10. For better display in simulations, the unit of p is set as packets per hour per node, which differs from packets per second per node in Section IV.

Given that the differences in $N_{\max}$ between different L_d values become smaller when p exceeds 10, we zoom in on this part for a clearer display. As shown in Fig. 12, when L_d is fixed at 60, $N_{\max}$ decreases with the increase of p, following the same trend of Fig. 11. $N_{\max}$ increases with the increase of ϕ until ϕ reaches 4. The theoretical $N_{\max}$ when $\phi = 3$ exactly matches the one when $\phi = 4$. Thus, to obtain the maximum $N_{\max}$ when $L_d = 60$, ϕ can be any integer no less than 3.

After obtaining the simulated $N_{\max}$, we use them to search for the simulated maximum coverage extension ratio, c. As shown in Fig. 13, when ϕ is fixed at 2, the maximum c decreases either with the increase of p or with the increase of L_d. As shown in Fig. 14, when L_d is fixed at 60, the maximum c decreases with the increase of p, following the same trend of Fig. 13. The maximum c increases with the decrease of ϕ until ϕ drops to 1. The theoretical maximum c when $\phi = 1$ closely matches the one when $\phi = 2$, with any differences attributed to the floor function in (13). Thus, to obtain the maximum c when $L_d = 60$, ϕ can be either 1 or 2.

From these figures, it is evident that the simulated results closely match the theoretical results, proving that the network attains the upper bounds of N and c. To measure the difference, we employ the mean absolute percentage error (MAPE) metric, which is defined as:

$$MAPE = \frac{1}{\delta} \sum_{i=1}^{\delta} \left| \frac{A_i - F_i}{A_i} \right| \tag{17}$$

where A_i is the theoretical value, F_i is the simulated value, and δ is the number of values. MAPE between theoretical values

TABLE II
MAPE BETWEEN THEORETICAL VALUES AND SIMULATED VALUES.

	Fig. 11	Fig. 12	Fig. 13	Fig. 14
$L_d = 30$	5.0%	/	3.6%	/
$L_d = 90$	4.7%	/	4.4%	/
$L_d = 150$	5.4%	/	5.1%	/
$\phi = 1$	/	2.7%	/	2.8%
$\phi = 2$	/	4.0%	/	3.2%
$\phi = 3$	/	6.0%	/	4.8%
$\phi = 4$	/	5.4%	/	4.5%

and simulated values in Fig. 11, 12, 13, and 14 is listed in Table II. The error arises from two aspects: 1) The simulation duration of 24 hours may not be sufficient to eliminate the randomness in the data collection of each sensor node. 2) Signal collisions are not considered when deriving theoretical $N_{\max}$ and c, but they are simulated in LoRaMeshSim. Signal collisions can result in the simulated outcomes exceeding the theoretical values. This occurs because, when signal collisions happen, packets are lost due to the absence of a retransmission mechanism in the experiments. Consequently, nodes that would otherwise relay the lost packets no longer need to do so, reducing their duty cycle. This reduced duty cycle, in turn, leads to an increase in the simulated $N_{\max}$ and c, potentially surpassing the theoretical values.

3) Bounds Verification: In (13), $N_{\max}$ has a upper bound and a lower bound when $\theta = \phi$ and 1, respectively. This implies that the values of $N_{\max}$ are between the two bounds regardless of the routing algorithms. To verify it, a straightforward method is to exhaust the possible topologies, i.e., all the spanning trees of graph G. However, the number of spanning trees grows exponentially with the increase of N. For example, according to Kirchhoff's theorem [43], when $\phi = 2$, the number of spanning trees of graph G is 6,765 and 102,334,155, respectively, for $N = 10$ and $N = 20$. On the other hand, to determine the maximum duty cycle for each spanning tree, it is necessary to set the simulation time to at least several hours—a task that takes seconds for code execution. Thus, the exhaustive search method is impossible to realize. Alternatively, we propose a random search method using the Andrei Broder and David Alduous random spanning tree generation algorithm. As shown in the Algorithm 2, to obtain the upper bound, we incrementally search from $N = 2$ until we no longer find a maximum duty cycle less than 1%. To obtain the lower bound, we decrementally search from N equal to the upper bound until we no longer find a maximum duty cycle greater than 1%. To reduce the execution time, the simulation time is reduced from 24 hours to 5 hours and 12 hours for the upper bound and lower bound, respectively. The parameters ϕ and L_d are fixed at 2 and 90, respectively, for this experiment and the subsequent experiment. As shown in Fig. 15, the simulated bounds match the theoretical bounds with tolerable errors listed in Table III. Moreover, we also conduct simulations with the proposed routing algorithm and RadioHead. In the simulation of RadioHead, collisions of routing broadcasts are intentionally disregarded to minimize interference from other factors. As shown in Fig. 15, $N_{\max}$ of RadioHead lies on the lower bound due to uneven relay task assignments, resulting in a bottleneck effect where certain adjacent nodes become overloaded. This behavior significantly reduces the scalability of the network. By contrast, the proposed routing algorithm balances relay tasks among nodes, achieving the theoretical upper bound of $N_{\max}$. This demonstrates the superior efficiency of our method, particularly in environments where balanced resource utilization is critical. Furthermore, the random search method generates diverse topologies using the random spanning tree generation algorithm. Although it cannot exhaust all possible topologies due to computational limitations, it covers a wide range of routing algorithms. The results in Fig. 15 demonstrate that the proposed routing algorithm outperforms all others in terms of the $N_{\max}$.

Algorithm 2 Random Search for the Bounds of $N_{\max}$

IF search for upper bound
  INPUT: search times = 2000
  INITIALIZATION: $N = 2, i = 0$
    WHEN $i <$ search times
      Generate a random spanning tree
      Conduct a simulation using the tree
      If the maximum duty cycle $\leq 1\%$
        $N+ = 1, i = 0$
      ELSE
        $i+ = 1$
  OUTPUT: $N - 1$
ELSE IF search for lower bound
  INPUT: search times = 1000
  INITIALIZATION: $N =$ the upper bound, $i = 0$
    WHEN $i <$ search times AND $N > 0$
      Generate a random spanning tree
      Conduct a simulation using the tree
      If the maximum duty cycle $> 1\%$
        $N- = 1, i = 0$
      ELSE
        $i+ = 1$
  OUTPUT: N

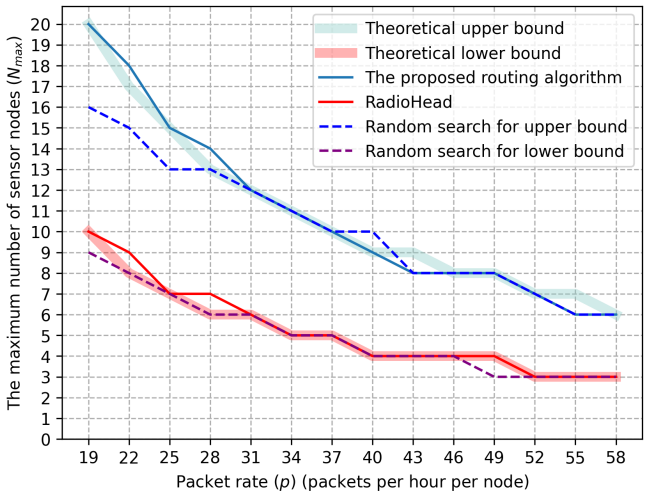

Fig. 15. The bounds of $N_{\max}$ with $\phi = 2$ and $L_d = 90$

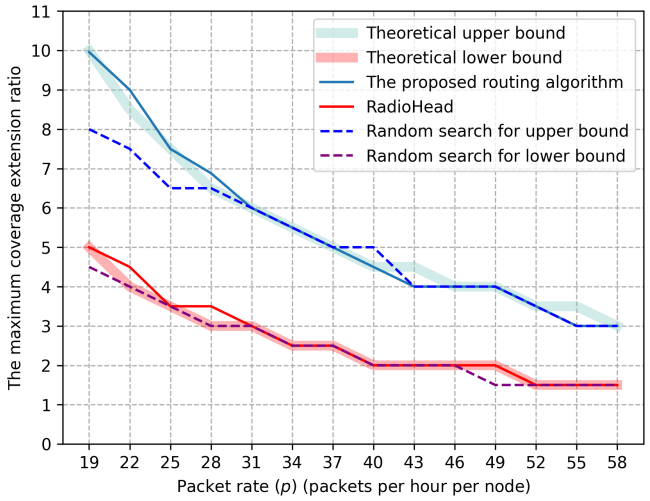

Fig. 16. The bounds of the maximum c with $\phi = 2$ and $L_d = 90$

TABLE III
MAPE BETWEEN THEORETICAL BOUNDS AND SIMULATED BOUNDS.

		Theoretical Upper Bound	Theoretical Lower Bound
$N_{\max}$	Proposed Routing Algorithm	2.7837%	/
	Random Search for Upper bound	5.8290%	/
	RadioHead	/	2.0833%
	Random Search for lower bound	/	2.5000%
maximum c	Proposed Routing Algorithm	2.6813%	/
	Random Search for Upper bound	5.8293%	/
	RadioHead	/	2.0833%
	Random Search for lower bound	/	2.5000%

After verifying the bounds of $N_{\max}$, we use the simulated $N_{\max}$ to verify the bounds of the maximum c indicated in (16). To obtain the lower bound of the maximum c, we set the simulated lower bound of $N_{\max}$ as N and repeat the simulation 100 times with random spanning trees as the topology. For the upper bound, we set the simulated upper bound of $N_{\max}$ as N and repeat the simulations with random spanning trees until a topology is found with a maximum duty cycle less than 1%. Employing the found topology, we repeat the simulation 100 times to obtain the upper bound of the maximum c. Moreover, we also conduct simulations with the proposed routing algorithm and RadioHead in terms of the maximum c. As shown in Fig. 16, the simulated lower bound matches the theoretical lower bound and RadioHead also lies on the theoretical lower bound due to uneven relay task assignments. By contrast, the simulated upper bound and the proposed routing algorithm match the theoretical upper bound. This highlights the superior performance of the proposed algorithm over RadioHead, particularly in maximizing network coverage. Furthermore, the random search method, which generates diverse topologies, validates the superior performance of the proposed algorithm over other routing algorithms in terms of coverage.

The errors between theoretical bounds and simulated bounds are listed in Table III. This demonstrates the correctness of the theoretical bounds for $N_{\max}$ and the maximum c, while highlighting the advantage of the proposed routing algorithm over others in terms of both $N_{\max}$ and the maximum c.

VII. CONCLUSION AND FUTURE WORK

This paper contributes to the understanding of deploying LoRa mesh networks for monitoring linear infrastructure, offering insights into scalability and coverage extension. The proposed deployment strategy optimizes the placement of nodes, ensuring effective coverage along linear infrastructure while adhering to duty cycle regulations. The system analysis also reveals the impact of duty cycle limitations on network scalability and coverage extension ratio. Moreover, their bounds are derived with the condition to reach the upper bounds. The proposed routing algorithm demonstrates its efficacy in achieving maximum scalability and coverage extension ratios. The developed LoRa mesh simulator validates the proposed system analysis and routing algorithm, providing a valuable tool for further research and practical implementation. This work facilitates the design and deployment of robust LoRa mesh networks for monitoring linear infrastructure, addressing the challenges associated with scalability and coverage extension.

In addition to duty cycle regulations, internal signal collisions also impact the scalability and coverage extension of LoRa mesh networks. In our future work, we aim to analyze this impact and explore solutions to eliminate or mitigate it. Furthermore, while this study focuses on linear infrastructure monitoring—a prevalent application scenario—we plan to extend our findings to broader use cases, such as smart buildings, smart cities, and smart farming.

APPENDIX
PROOF OF LEMMA 1

The lemma can be proved using a proof of contradiction. Assume that a remote node has the maximum duty cycle in the sensor nodes, denoted as $\max(d_i)$. According to (3),

$$\max(d_i) = [\max(n_i) + 1]pt_d + \max(n_i)pt_a, \qquad (18)$$

where $\max(n_i)$ denotes the number of sensor nodes whose data packets are relayed by the remote node. The data packets of the remote node have to be relayed by an adjacent node to reach the gateway. All the nodes whose data packets are relayed by the remote nodes also require the adjacent node to relay data packets subsequently. Additionally, the adjacent node also relays the data packets generated in the remote node. Thus, the adjacent node relays data packets for at least $\max(n_i) + 1$ nodes and its duty cycle is not less than $[\max(n_i) + 2]pt_d + [\max(n_i) + 1]pt_a$ which is bigger than $\max(d_i)$. This contradicts our assumption, so Lemma 1 holds.

REFERENCES

[1] P. D'Aranno, A. Di Benedetto, M. Fiani, and M. Marsella, "Remote sensing technologies for linear infrastructure monitoring," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 461–468, 2019.

[2] "Mobile signal strength measurement data from Network Rail's engineering trains," Ofcom, London, UK. Accessed: Jan. 16, 2024. [Online]. Available: https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-2019/data-downloads

[3] R. M. Liaqat, P. Branch, and J. But, "LoRa based linear network applications, design considerations and open challenges: A review," in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, 2022, pp. 913–917.

[4] J. P. S. Sundaram, W. Du, and Z. Zhao, "A survey on LoRa networking: Research problems, current solutions, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 371–388, 2019.

[5] A. Pagano, D. Croce, I. Tinnirello, and G. Vitale, "A survey on LoRa for smart agriculture: Current trends and future perspectives," *IEEE Internet of Things Journal*, vol. 10, no. 4, pp. 3664–3679, 2022.

[6] B. Citoni, F. Fioranelli, M. A. Imran, and Q. H. Abbasi, "Internet of things and LoRaWAN-enabled future smart farming," *IEEE Internet of Things Magazine*, vol. 2, no. 4, pp. 14–19, 2019.

[7] C. Ebi, F. Schaltegger, A. Rüst, and F. Blumensaat, "Synchronous LoRa mesh network to monitor processes in underground infrastructure," *IEEE access*, vol. 7, pp. 57 663–57 677, 2019.

[8] S. Hong, F. Yao, Y. Ding, and S.-H. Yang, "A hierarchy-based energy-efficient routing protocol for LoRa-mesh network," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22 836–22 849, 2022.

[9] D. Wu and J. Liebeherr, "A low-cost low-power LoRa mesh network for large-scale environmental sensing," *IEEE Internet of Things Journal*, vol. 10, no. 19, pp. 16700–16714, 2023.

[10] P. Tian, C. A. Boano, X. Ma, and J. Wei, "LoRaHop: Multi-hop support for LoRaWAN uplink and downlink messaging," *IEEE Internet of Things Journal*, 2023.

[11] "ETSI EN 300 220-2 v3.2.1 (2018-06)," ETSI, Sophia Antipolis, France, Rep. REN/ERM-TG28-535, 2018. Accessed: Jan. 22, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_en/300200_300299/30022002/03.02.01_60/en_30022002v030201p.pdf

[12] LoRa Alliance, "LoRaWAN 1.1 regional parameters," LoRa Alliance, 2017.

[13] A. Mahmood, E. Sisinni, L. Guntupalli, R. Rondón, S. A. Hassan, and M. Gidlund, "Scalability analysis of a LoRa network under imperfect orthogonality," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1425–1436, 2018.

[14] H. Huh and J. Y. Kim, "LoRa-based mesh network for IoT applications," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*. IEEE, 2019, pp. 524–527.

[15] Y. Chen, G. Shi, M. Al-Quraan, Y. Sambo, O. Onireti, and M. Imran, "LoRa mesh-5G integrated network for trackside smart weather monitoring," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 6, pp. 8903–8914, 2024.

[16] R. Guirado, J.-C. Padró, A. Zoroa, J. Olivert, A. Bukva, and P. Cavestany, "Stratotrans: Unmanned aerial system (UAS) 4G communication framework applied on the monitoring of road traffic and linear infrastructure," *Drones*, vol. 5, no. 1, p. 10, 2021.

[17] R. M. Liaqat, P. Branch, and J. But, "Design considerations and performance evaluation of gossip routing in LoRa-based linear networks," *Future Internet*, vol. 15, no. 11, p. 366, 2023.

[18] P. Fraga-Lamas, T. M. Fernández-Caramés, and L. Castedo, "Towards the Internet of smart trains: A review on industrial IoT-connected railways," *Sensors*, vol. 17, no. 6, p. 1457, 2017.

[19] O. Jo, Y.-K. Kim, and J. Kim, "Internet of things for smart railway: feasibility and applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 482–490, 2017.

[20] I. Stoianov, L. Nachman, S. Madden, and T. Tokmouline, "Pipeneta wireless sensor network for pipeline monitoring," in *Proceedings of the 6th international conference on Information processing in sensor networks*, 2007, pp. 264–273.

[21] S. Kurnaz, G. Babayeva, R. B. Rustamov, and E. Aleskerov, "Monitoring of the linear infrastructure: Environmental and social impacts," in *2009 4th International Conference on Recent Advances in Space Technologies*. IEEE, 2009, pp. 273–276.

[22] C. L. Narayana, R. Singh, and A. Gehlot, "Performance evaluation of LoRa based sensor node and gateway architecture for oil pipeline management," *Int. J. Electr. Comput. Eng*, vol. 12, no. 1, pp. 974–982, 2022.

[23] D. Lundell, A. Hedberg, C. Nyberg, and E. Fitzgerald, "A routing protocol for LoRa mesh networks," in *2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 2018, pp. 14–19.

[24] R. Berto, P. Napoletano, and M. Savi, "A LoRa-based mesh network for peer-to-peer long-range communication," *Sensors*, vol. 21, no. 13, p. 4314, 2021.

[25] *RadioHead.* (2023). AirSpayce. Accessed: Jan. 16, 2024. [Online]. Available: https://www.airspayce.com/mikem/arduino/RadioHead/

[26] H. C. Lee and K. H. Ke, "Monitoring of large-area IoT sensors using a LoRa wireless mesh network system: Design and evaluation," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 9, pp. 2177–2187, 2018.

[27] A. Shrestha and L. Xing, "A performance comparison of different topologies for wireless sensor networks," in *2007 IEEE Conference on Technologies for Homeland Security*. IEEE, 2007, pp. 280–285.

[28] M. C. Bor, U. Roedig, T. Voigt, and J. M. Alonso, "Do LoRa low-power wide-area networks scale?" in *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2016, pp. 59–67.

[29] O. Georgiou and U. Raza, "Low power wide area network analysis: Can LoRa scale?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, 2017.

[30] Y. Bouazizi, F. Benkhelifa, H. ElSawy, and J. A. McCann, "On the scalability of duty-cycled LoRa networks with imperfect sf orthogonality," *IEEE Wireless Communications Letters*, vol. 11, no. 11, pp. 2310–2314, 2022.

[31] Y. Chen, Y. A. Sambo, O. Onireti, S. Ansari, and M. A. Imran, "LoRaWAN-5G integrated network with collaborative RAN and converged core network," in *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2022, pp. 1–5.

[32] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo, "On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology," in *14th international conference on its telecommunications (ITST)*. IEEE, 2015, pp. 55–59.

[33] V. J. Hodge, S. O'Keefe, M. Weeks, and A. Moulds, "Wireless sensor networks for condition monitoring in the railway industry: A survey," *IEEE Transactions on intelligent transportation systems*, vol. 16, no. 3, pp. 1088–1106, 2014.

[34] A. M. Obeid, F. Karray, M. W. Jmal, M. Abid, S. Manzoor Qasim, and M. S. BenSaleh, "Towards realisation of wireless sensor network-based water pipeline monitoring systems: a comprehensive review of techniques and platforms," *IET science, measurement & technology*, vol. 10, no. 5, pp. 420–426, 2016.

[35] T. Bouguera, J. F. Diouris, J. J. Chaillout, R. Jaouadi, and G. Andrieux, "Energy consumption model for sensor nodes based on LoRa and LoRaWAN," *Sensors*, vol. 18, no. 7, p. 2104, 2018.

[36] K. H. Lam, C. C. Cheung, and W. C. Lee, "RSSI-based LoRa localization systems for large-scale indoor and outdoor environments," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 11 778–11 791, 2019.

[37] NS-3 network simulator [Online]. Available: https://www.nsnam.org/, Accessed on: Jan. 16, 2024.

[38] F. Van den Abeele, J. Haxhibeqiri, I. Moerman, and J. Hoebeke, "Scalability analysis of large-scale LoRaWAN networks in NS-3," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2186–2198, 2017.

[39] T. Voigt and M. Bor, "LoRaSim," Version 0.2.1, Jun. 10, 2017. [Online]. Available: https://mcbor.github.io/lorasim/, Accessed on: Jan. 16, 2024.

[40] SimPy - Event discrete simulation for Python [Online]. Available: https://simpy.readthedocs.io/, Accessed on: Jan. 16, 2024.

[41] A. Z. Broder, "Generating random spanning trees," in *FOCS*, vol. 89, 1989, pp. 442–447.

[42] D. J. Aldous, "The random walk construction of uniform spanning trees and uniform labelled trees," *SIAM Journal on Discrete Mathematics*, vol. 3, no. 4, pp. 450–465, 1990.

[43] D. B. West, *Introduction to graph theory*. Upper Saddle River: Prentice hall, 2001, vol. 2.

Yu Chen (Member, IEEE) received his B.S. and M.S. degrees from the University of Electronic Science and Technology of China in 2017 and 2020, respectively. He earned his Ph.D. in Electronics and Electrical Engineering from the University of Glasgow in 2024. Currently, he is a Postdoctoral Research Associate in the James Watt School of Engineering at the University of Glasgow. Previously, he served as a Research Assistant at the University of Glasgow, where he led a design team for the European Space Agency-funded Satellites for Digitalization of Railways (SODOR) project. His research interests include self-organizing wireless networks, open radio access network, hybrid network management, and compressive sampling.

Guo Shi (Graduate Student Member, IEEE) received the B.S. degree in Industrial Engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2018, and the M.S. degree in Industrial Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2021. Currently, she is a Ph.D. student in the Department of Management Science at the University of Strathclyde, Glasgow, U.K. Her research interests include data-driven decision-making, machine learning, maintenance modeling, prognostic and health management.

Yusuf A. Sambo (Senior Member, IEEE) received the MSc degree (distinction) in Mobile and Satellite Communications, in 2011 and the Ph.D. degree in electronic engineering in 2016 from the Institute for Communication Systems (ICS, formally known as CCSR), University of Surrey. He was a Postdoctoral Research Associate and Lecturer at the James Watt School of Engineering, University of Glasgow, from 2017 and 2019, respectively. He is currently a Senior Lecturer in Communication Systems and the 5G Testbed Manager for the Scotland 5G Centre-funded testbed at the University of Glasgow. Prior to joining Glasgow, he was a Lecturer in Telecommunications Engineering at Baze University, Abuja between 2016 and 2017. His interests include self-organizing networks, cell-free massive MIMO, EMF exposure, and green communications.

Oluwakayode Onireti (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from the University of Ilorin, Ilorin, Nigeria, in 2005, and the M.Sc. degree (Hons.) in mobile and satellite communications, and the Ph.D. degree in electronics engineering from the University of Surrey, Guildford, U.K., in 2009 and 2012, respectively. He is currently a Senior Lecturer with the University of Glasgow, Glasgow, U.K. He has been actively involved in projects, such as ROCKET, EARTH, Greencom, QSON, DARE, and Energy proportional EnodeB for LTEAdvanced and Beyond. His main research interests include self-organizing cellular networks, energy efficient networks, wireless blockchain networks, millimeter wave communications, and cooperative communications.

Muhammad A. Imran (Fellow, IEEE) received the M.Sc. (Hons.) and Ph.D. degrees from Imperial College London, London, U.K., in 2002 and 2007, respectively. He is currently the Dean of University of Glasgow UESTC, Head of Autonomous Systems and Connectivity Division, and a Professor of communication systems with the James Watt School of Engineering, University of Glasgow, Glasgow, U.K. He is an Affiliate Professor with The University of Oklahoma, Norman, OK, USA, and 5G Innovation Centre, University of Surrey, Guildford, U.K. He is leading research with the University of Glasgow for Scotland 5G Centre. He has more than 20 years of combined academic and industry experience with several leading roles in multi-million pound funded projects, working primarily in the research areas of cellular communication systems. He was the recipient of an Award of Excellence in recognition of his academic achievements, conferred by the President of Pakistan. He was also the recipient of the IEEE Comsoc's Fred Ellersick Award 2014, FEPS Learning and Teaching Award 2014, Sentinel of Science Award 2016, and ten patents.