

Article

# K-Means Spreading Factor Allocation for Large-Scale LoRa Networks

Muhammad Asad Ullah <sup>1</sup>, Junnaid Iqbal <sup>1</sup>, Arliones Hoeller <sup>1,2,3</sup>, Richard Demo Souza <sup>2</sup>, and Hirley Alves <sup>1,\*</sup>

<sup>1</sup> Centre for Wireless Communications, University of Oulu, 90014 Oulu, Finland; muhammad.asadullah@oulu.fi (M.A.U.); junnaid.iqbal@oulu.fi (J.I.)

<sup>2</sup> Department of Electrical and Electronics Engineering, Federal University of Santa Catarina, Florianópolis 88040-900, Brazil; richard.demo@ufsc.br

<sup>3</sup> Department of Telecommunications Engineering, Federal Institute for Education, Science, and Technology of Santa Catarina, São José 88103-310, Brazil; arliones.hoeller@ifsc.edu.br

\* Correspondence: Hirley.Alves@oulu.fi

Received: 30 September 2019; Accepted: 28 October 2019; Published: 30 October 2019



**Abstract:** Low-power wide-area networks (LPWANs) are emerging rapidly as a fundamental Internet of Things (IoT) technology because of their low-power consumption, long-range connectivity, and ability to support massive numbers of users. With its high growth rate, Long-Range (LoRa) is becoming the most adopted LPWAN technology. This research work contributes to the problem of LoRa spreading factor (SF) allocation by proposing an algorithm on the basis of K-means clustering. We assess the network performance considering the outage probabilities of a large-scale unconfirmed-mode class-A LoRa Wide Area Network (LoRaWAN) model, without retransmissions. The proposed algorithm allows for different user distribution over SFs, thus rendering SF allocation flexible. Such distribution translates into network parameters that are application dependent. Simulation results consider different network scenarios and realistic parameters to illustrate how the distance from the gateway and the number of nodes in each SF affects transmission reliability. Theoretical and simulation results show that our SF allocation approach improves the network's average coverage probability up to 5 percentage points when compared to the baseline model. Moreover, our results show a fairer network operation where the performance difference between the best- and worst-case nodes is significantly reduced. This happens because our method seeks to equalize the usage of each SF. We show that the worst-case performance in one deployment scenario can be enhanced by 1.53 times.

**Keywords:** stochastic geometry; resource allocation; Internet of Things

## 1. Introduction

The Internet of Things (IoT) is the integration of modern electronic devices, smart sensors, internet protocols, and wireless communications technologies. IoT applications are rapidly gaining popularity in many domains such as industrial operations, smart parking, augmented maps, healthcare, smart cars, and smart homes [1–5]. According to a Gartner Inc. report, there will be around 26 billion IoT devices deployed worldwide by 2020 [6]. In the Statista report, it is predicted that there will be over 75 billion IoT devices worldwide by 2025 [7].

In the modern era, the spectacular growth and transformation of wireless connectivity are driven by the IoT paradigm, with technologies having attributes of large-scale network infrastructure with low-cost sensors connected to the Internet. In this context, low-power wide-area networks (LPWANs) are quite popular in terms of prototypes, standards, and on the commercial level because of their

significance with respect to power efficiency along with long range [8,9]. Within this context, LoRa, SigFox, NB-IoT, Weightless, RPMA and DASH7 [10,11] are the most distinguished technologies.

This paper focuses on LoRa, which provides good performance in terms of reliability and energy consumption. The network architecture contains end-devices, gateways, and a network server (NS), forming a star topology. It operates at unlicensed frequency ISM (Industrial, Scientific, Medical) bands of 863–870 MHz and 915 MHz in Europe and the U.S., respectively [12,13]. In Europe, the duty cycle limitations range from 0.1% to 10%, following European Telecommunications Standards Institute (ETSI) standards. In addition, LoRa works on variable and adaptive data rates by using different spreading factors. This is achieved by the NS controlling the spreading factors (SFs) and bandwidth (BW) of the end-devices. Higher SFs allow larger coverage areas; however, as a drawback, they reduce the data rate and increase the time-on-air (ToA) of LoRa packets [14].

Notably, the gateway has the ability to receive data from multiple nodes at the same time because of the orthogonality of sub-bands and the quasi-orthogonality of different SFs. The LoRa MAC layer, known as LoRaWAN [15], is a type of ALOHA protocol controlled by the NS. LoRaWAN defines three classes of devices depending upon the application. Class A devices may wait for acknowledgments (ACK) only in their receiving windows during downlink transmission and consume the least power. Class B devices are able to open extra receiving windows at scheduled times, thus reducing downlink latency. Class C nodes consume the most energy because they leave the receiver enabled all the time, allowing for the lowest latency time [16].

For instance, extensive measurement campaigns show that the communication range of LoRa reaches up to 30 km over the water and more than 15 km on the ground [11]. LoRa is suitable for a wide range of telemetry applications (e.g., sensing and monitoring), which can be used in several industry verticals, such as smart grids and cities, and smart agriculture up to industrial IoT applications [17,18]. During the past few years, many studies have contributed by proposing new algorithms, systems models, analyses, and by designing new approaches for performance enhancement of LoRa networks. However, only a few considered resource allocation.

The major contribution of this work is the modeling of an approach for SF allocation for a large scale LoRa network based on K-means clustering and the analysis of connection, capture, and coverage probabilities. Instead of using constant steps of distance from the gateway to define SF areas [19,20], the proposed algorithm assigns a maximum range of individual SF regions, which allows for distinct user distribution. Then, we evaluate the performance of the proposed algorithm over the uplink of a large-scale LoRa network with a single gateway based on the model introduced in [19].

The remainder of this article is structured as follows. Section 2 discusses related work and a short overview of LPWAN technologies. Section 3 introduces the system model, and Section 3.1 details the outage probabilities of the baseline model, used to examine the performance of proposed SF allocation approach. The proposed algorithm is presented in Section 4. Simulation results are discussed in Section 5. Finally, Section 6 concludes the paper and proposes future work.

## 2. Related Work

Overviews of LoRa and LPWAN technologies are provided in [21,22]. Usually, LoRa operates with a bandwidth of 125 kHz, but it also allows for bandwidths of 250 kHz and 500 kHz. The wider bands promote resistance to fading, channel noise, Doppler effects, and long-term relative frequency [23]. Chirp spread spectrum (CSS) modulation, which enables high receiver sensitivity, makes LoRa more robust against the interference when compared to Sigfox, which employs ultra-narrowband (UNB) communication [24]. As a tradeoff, the use of wider bands for the transmission of narrowband signals makes less efficient use of the spectrum. A realistic SigFox communication model is implemented and tested in [25]; it evaluates the performance of a high-density large-scale wireless sensor network (WSN). From the obtained results, one can observe that the performance of the SigFox network significantly degrades by increasing the number of sensors, and some solutions are presented to improve the performance.

Unlike Sigfox, LoRa can be deployed locally, i.e., without the need for a cellular infrastructure, and has higher bit rates. By contrast, NB-IoT is an expensive technology having the pros of low latency and high quality of service (QoS) [26]. In [27], the authors compare different LPWAN technologies (Bluetooth, ZigBee, SigFox, and LoRa) and discuss LoRa with respect to code rate (CR), bandwidth (BW), and SF but without considering the influence of Rayleigh fading and path loss attenuation. Theoretical and simulation results show that SF, BW, and CR influence the ToA of a packet. Larger SFs and CRs result in higher ToA of LoRa packets. Conversely, ToA reduces with larger bandwidths.

The work in [28] proposes two different algorithms named EXPLoRa-SF and EXPLoRa-AT and shows in simulation results that these algorithms perform considerably better than the LoRaWAN adaptive rate strategy (ADR). EXPLoRa-AT delivers higher bit rates in the event of higher traffic loads, while EXPLoRa-SF allocates SFs at the different subgroups of end-devices depending on the received signal strength indicator (RSSI). The results demonstrate that the data extraction rate (DER) drops dramatically for higher SFs and larger numbers of end-devices. The authors, however, assume a short range and dense network in their analysis.

The EXPLoRa approach is further extended in [29], K-means is applied to identify the non-circular crowded region, and all the nodes inside that area are assumed to have same SF. On the other hand, in the proposed work the geometry of network is circular, with six annuluses representing the range of individual SFs. We have analyzed the scalability and the performance of the uplink LoRa model considering Rayleigh fading, connection  $H_1$ , capture  $Q_1$ , and coverage probabilities  $H_1Q_1$  in the presence of interfering signals using the same SF. The considerations of  $H_1$  and  $H_1Q_1$  are missing in [28,29]. Moreover, in our model, we consider a dense and wide network (radius of several kilometers) and analyze the performance by considering the maximum distance of individual SF boundaries from the gateway.

Another scientific study used K-means for the classification of end-devices into three groups based on traffic characteristics with different priorities. The grouping of end-devices was computed in terms of priority-based transmission instead of SF allocation [30].

In [31,32], SF distribution is mainly based on the power level of the signals that the gateway receives from the end-devices and gateway sensitivity, without considering the location of end-devices. As a drawback, SF allocation was disturbed because of high-density buildings, and 53.2% of the end-devices were forced to use SF12. Furthermore, in [28–31], only network-level simulators such as ns-3 and LoRaSim are used, which abstracts some characteristics of the physical layer that are incorporated in our analysis. Conversely, our study evaluated the performance of the proposed SF allocation algorithm considering the analytical model, realistic parameters, and averaging over  $10^5$  random deployment of the Poisson point process (PPP) by Monte Carlo computer simulations, which match with the theoretical results.

The tree-based spreading factor clustering algorithm (TSCA) for SF allocation in multihop LoRa networks is introduced in [33]. This approach offloads the data traffic in many sub-networks, which are linked to a sink node assigning a specific SF according to network clustering, thus enabling parallel frame transmission with multiple SFs. The authors show that TSCA increases the network performance in a network with rectangular geometry.

A single gateway uplink model considering path loss attenuation and Rayleigh fading is designed in [19], utilizing stochastic geometry to model network interference and then disconnection and collision probabilities. Such a model is further extended in [20], in which the authors propose a scheme that considers message replication and gateways with multiple receive antennas/decoders to attain time and spatial diversity. They demonstrate that the number of users and traffic density directly affects the performance of the LoRa network and that sending multiple message copies is beneficial for low-density networks. Both of these studies adopt equal radius SF allocation approaches. Unlike [19,20], our work considers K-means-based fair SF allocation of nodes in LoRa networks.

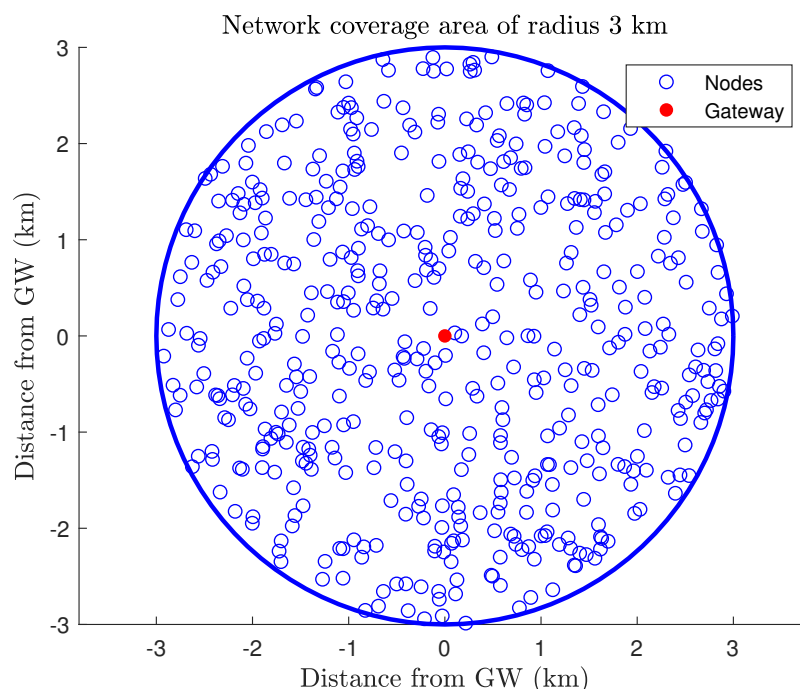
Recently, several studies have addressed the problems associated with automatic repeat request (ARQ) and contributed to downlink reliability in LoRaWAN applications. The sequential transmission

of downlink frames, saturation of duty cycle, and half-duplex nature of LoRa gateway radios are marked as the major shortcomings for the downlink transmission [34,35]. Furthermore, these works also highlight the significance of gateway selection algorithm to prevent traffic losses due to sequential transmission of downlink frames and duty cycle limitations.

One experimental study evaluates the performance of a LoRa network at a 125 kHz bandwidth and SF7 for a sailing monitoring model, and the measurements show a 60.49% packet loss at the maximum distance of 3284 m [36]. Another LoRaWAN-based indoor environment monitoring system composed of 331 sensor nodes is deployed at the University of Oulu, where the gateway is installed at a distance of  $\sim 180$  m and 24 m above the ground [37]. The measurements performed at SF7 show a maximum 11.33% packet error rate (PER), which can be due to co-spreading factor interference because all 331 end-devices use the same SF. As illustrated in [19], nodes using the same SF face co-spreading factor interference. The motivation behind our work is to propose a suitable SF allocation algorithm for a large-scale LoRa network to efficiently utilize the different data rates. To enhance SF allocation, we propose a novel algorithm, based on the machine learning technique called K-means clustering, for effectively allocating the SFs.

### 3. System Model

We consider  $\bar{N}$  uniformly distributed smart devices inside an uplink class-A LoRaWAN network without retransmissions, utilizing a single channel within radio range of  $R$  km and a circular area of  $V = \pi R^2$  around a single gateway. Figure 1 illustrates a deployment with  $\bar{N} = 500$  and  $R = 3$  km. The gateway is at the origin, and nodes are distributed uniformly in  $V = 28.26$  km<sup>2</sup>. Note that such model captures the characteristics of telemetry applications such as those in smart cities and smart buildings. For instance, the University of Oulu Smart Campus has a LoRaWAN network constantly monitoring several sensors such as temperature, luminosity, and CO<sub>2</sub> [37].



**Figure 1.** Uniform distribution of  $\bar{N} = 500$  nodes in a circular network area of radius  $R = 3$  km, with the gateway (GW) at the origin.

The LoRa modulation bit rate is defined as [14]

$$R_b = \frac{4}{4 + CR} \frac{BW}{2^{SF}}, \quad (1)$$

where  $\frac{4}{4+CR}$  is the effective coding rate, ranging from  $\frac{4}{5}$  to  $\frac{4}{8}$ , while  $CR$  denotes the LoRa coding rate configuration, varying from 1 to 4. In our work, we assume  $CR = 1$ , and the LoRa uplink channel aggregated bit rate is expressed as  $bitrate_U = \sum_{i=7}^{12} Rb_i = 12.17$  kbps. For instance, Table 1 shows the characteristics of 9 byte LoRa packets with explicit header and CRC modes enabled and  $BW = 125$  kHz.

**Table 1.** Characteristics of the LoRa uplink model containing packets of 9 bytes at  $BW = 125$  kHz.

SF (i)	Bit Rate (kbps) ( $Rb_i$ )	Receiver Sensitivity (dBm)	SNR (dB) ( $q_{SF}$ )	Range (km)
7	5.47	−123	−6	$l_0 - l_1$
8	3.13	−126	−9	$l_1 - l_2$
9	1.76	−129	−12	$l_2 - l_3$
10	0.98	−132	−15	$l_3 - l_4$
11	0.54	−134.5	−17.5	$l_4 - l_5$
12	0.29	−137	−20	$>l_5$

### 3.1. Uplink Outage Probability

The uplink transmission of nodes is based on the ALOHA protocol, and the probability of collision in ALOHA networks is high when many stations are connected [38]. In LoRa, simultaneous signals of different SFs are quasi-orthogonal because the inter-SF rejection gain varies from 16 to 36 dB [39]. Therefore, for the sake of simplicity, our work does not inspect inter-SF interference and focuses on co-SF interference only, which is stronger.

In this paper, the uplink model includes the influence of Rayleigh fading and path loss attenuation as the baseline model [19] for performance analysis, where  $g(d_k) = \frac{\lambda}{(4\pi d_k)^\eta}$  is the path loss attenuation function,  $\eta \geq 2$  is the path loss exponent,  $\lambda$  is the wavelength, and  $h_k$  is the fading in the link between the  $k$ -th node and the gateway. Let us consider the transmitted signal of a single LoRa node  $s_1(t)$  to examine the impact of co-SF interference originated due to simultaneous transmission of nodes with same SF. The mathematical expression of the received signal at the gateway can be expressed as

$$r_1(t) = g(d_1)h_1 * s_1(t) + \sum_{k=2}^N \chi_k^{SF}(t)g(d_k)h_k * s_k(t) + n(t), \quad (2)$$

where  $n(t)$  is additive white Gaussian noise with zero mean and variance  $\mathcal{N} = -174 + NF + 10\log_{10}(BW)$  dBm,  $NF$  is the noise figure of the receiver, and  $-174$  dBm/Hz is the thermal noise spectral density constant.

We consider that an outage of the received signal in an uplink channel can take place in the two scenarios [19]. First, if the signal-to-noise ratio (SNR) of the received packet is less than the SF specific threshold  $q_{SF}$ , then the node is considered disconnected. Second, if the signal-to-interference ratio (SIR) between the target-received packet and any other concurrent signals of the same SF and frequency channel is less than 6 dB, then it is considered as a collision.

#### 3.1.1. Outage Condition I

The distance of the end-device to the gateway in a wireless transmission domain is crucial. The instantaneous SNR can be expressed as  $SNR = \frac{\mathcal{P}_1 |h_1|^2 g(d_1)}{\mathcal{N}}$ , where  $\mathcal{P}_1$  is the transmit power of end-device 1 in mW and  $|h_1|^2$  is the squared envelop of the channel coefficient. Communication is

only possible when the SNR of the received signal at the gateway is less than the reception threshold  $q_{SF}$ . Thus, the first outage condition, the connection probability, is defined as [19]

$$H_1 = \exp\left(\frac{\mathcal{N}q_{SF}}{\mathcal{P}_1g(d_1)}\right), \quad (3)$$

where  $d_1$  (in meters) is the distance of the desired end-device from the gateway.

### 3.1.2. Outage Condition II

A collision in LoRa end-device transmission takes place if the SIR of the desired signal with respect to interference from the same SF and frequency channel is less than 6 dB, i.e., if the desired signal is at least four times stronger than the interference. We model this outage condition based on [19], where interference is approached by considering the strongest interfering device. According to [19], the highest interference comes from the end-device  $k^*$ .

The probability that no collision occurs or that the strongest interfering signal is at least 6 dB below the desired one, termed the capture probability, is

$$Q_1 = \mathbb{P}\left[\frac{|h_1|^2g(d_1)}{|h_{k^*}|^2g(d_{k^*})} \geq 4 \mid d_1\right] = \mathbb{E}_{|h_1|^2}\left[\mathbb{P}\left[X_{k^*} < \frac{|h_1|^2g(d_1)}{4} \mid |h_1|^2, d_1\right]\right]. \quad (4)$$

The probability above depends on the distribution of  $X_{k^*} = |h_{k^*}|^2g(d_{k^*})$ . The cumulative distribution function (CDF) of  $X_{k^*}$  is derived in [19] and is denoted as  $F_{X_{k^*}}$ . Thus,

$$Q_1 = \mathbb{E}_{|h_1|^2}\left[F_{X_{k^*}}\left(\frac{|h_1|^2g(d_1)}{4}\right)\right] = \int_0^\infty e^{-z}F_{X_{k^*}}\left(\frac{zg(d_1)}{4}\right) dz. \quad (5)$$

Moreover, in [19] the authors present an approximation for (5) that is only accurate at the edges of each annulus. This paper considers only the exact probability in (5).

### 3.1.3. Coverage Probability

The probability that defines whether a selected end-device is in coverage and can successfully communicate with the gateway is termed the coverage probability. It is the product of  $H_1$  and  $Q_1$ . The average coverage probability  $\varphi_c$  can be achieved by deconditioning the location of the individual node by averaging over the network coverage area  $V = \pi R^2$ , i.e., [19]

$$\varphi_c = \frac{2}{R^2} \int_0^R H_1(d_1)Q_1(d_1)d_1 dd_1. \quad (6)$$

The average coverage probability of a individual SF annulus is also inspected. It indicates the probability of an end-device at distance  $d_1$  in the annulus  $i$  by considering the connection and capture probabilities and is defined as [20]

$$\varphi_{c,i} = \frac{2}{(l_{i+1} - l_i)^2} \int_{l_i}^{l_{i+1}} H_1(d_1)Q_1(d_1)(d_1 - l_i) dd_1, \quad (7)$$

where  $l_{i+1}$  is the radius of the outer circle and  $l_i$  is the radius of the inner circle of the  $i$ th annulus.

## 4. Proposed SF Allocation Algorithm

In this paper, we propose an SF allocation algorithm, i.e., an algorithm to define the range of each SF annulus. Our solution uses the K-means machine learning algorithm [40], used in the process of vector quantization in data mining by clustering. It is a non-deterministic, numerical, and iterative approach. The main objective of the K-means algorithm is to find the minimum cost function, defined as the distance between each point in the data set and its nearest centroid. The distance

between the cluster centers and data elements typically assumes the Euclidean distance. K-means clustering method can efficiently achieve robust clustering results when dealing with large data sets. The K-means algorithm first arbitrarily chooses  $K$  points from the data set, which indicate the initial centroids. The remaining points are then clustered to the closest centroid, and the coordinates of centroids are recalculated, iteratively, until the cost function converges.

Consequently, it is important to choose the appropriate number of centroids during the initialization procedure because the area of each annulus  $\pi(l_{i+1}^2 - l_i^2)$  increases towards the higher SFs in a strategy based on equal distance steps per SF, which results in the growth of node density due to uniform distribution. That is why it is essential to select the sequences for K-means iterations that can provide larger values of  $K$  clusters for higher SFs. In order to avoid an extensive number of nodes in an individual SF, there should be a fair difference between the inner ( $l_i$ ) and outer ( $l_{i+1}$ ) radii of annulus. In the proposed work, the annulus area is directly dependent on the difference between the  $K$  clusters for two consecutive iterations.

In our approach, we use five iterations of K-means. We start by computing the boundaries of the outermost SF ring, SF12, and then proceed to define the inner boundaries for lower SFs. For each iteration,  $K$  clusters are selected to develop the centroids of end-devices in the LoRa network covered by a single gateway. Four different mathematical sequences listed in Table 2—a Fibonacci series, square numbers, Wythoff array, and arithmetic series—are used to assign the values of  $K$  for each iteration.

**Table 2.**  $K$  Cluster Values for K-Means Iterations.

Iteration Series	1st SF12	2nd SF11	3rd SF10	4th SF9	5th SF8
Fibonacci series	34	21	13	8	5
Square number	49	36	25	16	9
Arithmetic series	34	28	22	16	10
Wythoff array	37	32	24	16	11

In our work, K-means operates iteratively. Each iteration defines the set of nodes at the outer SF ring. In each iteration, the algorithm seeks the set of  $K$  centroids  $C$  that minimizes the average of the distances between any node and its closest centroid, i.e.,

$$C = \arg \min_{C_k \in C} \frac{1}{|ED_k|} \sum_{X_i \in ED_k} \text{dist}(C_k, X_i)^2, \quad (8)$$

where  $ED_k$  is the set of devices at the  $k$ -th iteration,  $X_i$  is a device in  $ED_k$ , and  $C_k$  is the closest centroid of  $X_i$ . The function  $\text{dist}(x_1, x_2)$  computes the Euclidean distance between  $x_1$  and  $x_2$ . This procedure returns the collection of  $K$  centroids of network nodes, whereas  $C = \{C_1, \dots, C_K\}$ . After computing the centroids, the algorithm determines the boundary of  $C$ , so that  $[C_x, C_y] = \text{boundary}(C)$ , which determines the 2-D vector of border points around the Cartesian coordinates of the centroids. Then, it separates the nodes that are inside of the centroid boundary, forming the set  $I = [I_x, I_y]$ , where  $I_x$  and  $I_y$  are vectors storing the coordinates of the inner nodes in each of the Cartesian dimensions. In the next step, the maximum absolute value of each dimension of  $I$  is calculated to set the radius as  $l_i = \frac{\max|I_x| + \max|I_y|}{2}$ , which defines the limit of the SF ring  $i$ . The procedure repeats to determine the boundaries of the remaining SF rings ( $l_5, \dots, l_1$ ).

The steps involved in the proposed SF allocation technique are described in Algorithm 1. It repeats the process five times to allocate nodes for SF12–SF8. At the end, the remaining nodes use SF7. Initially, the SF12 outer limit is set to the network radius. In each iteration, the number of clusters is assigned to  $K$  depending on the chosen mathematical series (as mentioned in Table 2). For the first iteration, the algorithm considers all of the  $\bar{N}$  nodes inside the set  $ED$ . In line 4, it computes the K-means of  $ED$ , which returns the centroids  $C = \{C_1, \dots, C_K\}$  by (8). Since nodes in the set  $I$  are inside the boundary of centroids, line 7 computes the inner limit of SF ring  $l_i$ . This process is repeated iteratively until  $l_1$  is

calculated for the allocation of SF8 and the remaining nodes are assigned to SF7 (line 11). Note that the set of nodes  $ED$  is updated at the end of each iteration by removing the nodes that were already allocated to an SF (line 9).

---

**Algorithm 1** K-Means-based SF Allocation
 

---

**Input:**  $ED := \bar{N}$  uniformly deployed nodes

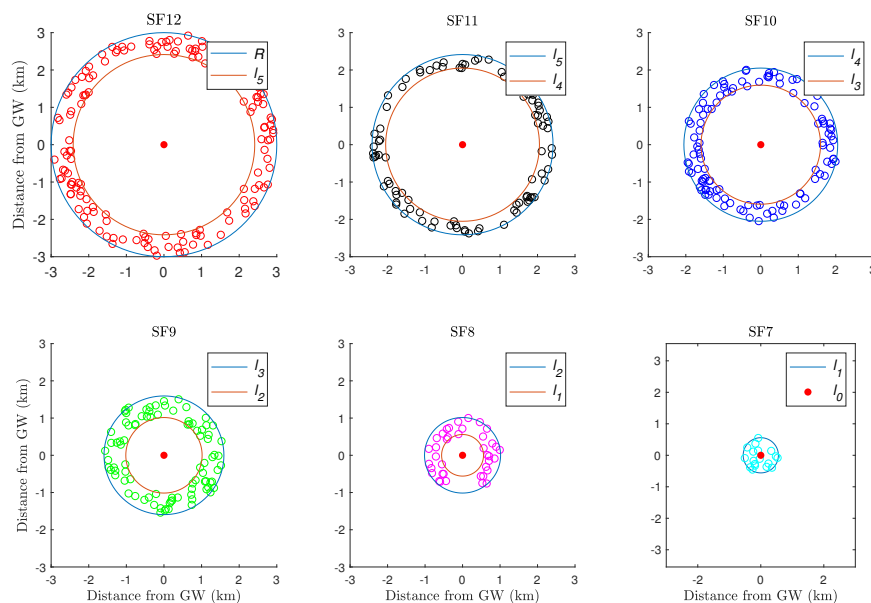
**Output:**  $L := \{l_0, l_1, \dots, l_6\}$

```

1:  $l_6 := R$ 
2: for  $i$  in  $\{5, \dots, 1\}$  do                                ▷ For each SF ring, starting from the outermost ring
3:    $K := \text{GetKfromSeries}(i)$                                 ▷ Set number of centroids for this iteration
4:    $C := \text{Kmeans}(ED, K)$                                     ▷ Compute the centroids
5:    $B := \text{boundary}(C)$                                        ▷ Compute the boundary of  $C$ 
6:    $I := \{x \in ED \mid x \in \text{conv}B\}$                         ▷ Select nodes that are inside the boundary  $B$ 
7:    $l_i := \frac{\max(|I_x|) + \max(|I_y|)}{2}$                        ▷ Compute the new SF ring limit
8:    $SF_{i+7} := \{x \in ED \mid x \notin \text{Ball}[(0,0), l_i]\}$     ▷ Allocate  $SF_{i+7}$  to nodes outside the circle of radius  $l_i$ 
9:    $ED := \{x \in ED \mid x \in \text{Ball}[(0,0), l_i]\}$           ▷ Remove nodes outside the circle of radius  $l_i$ 
10:  $l_0 := 0$ 
11:  $SF_7 := ED$                                                ▷ Allocate  $SF_7$  to remaining nodes
12: return  $L$ 
  
```

---

All of the iterations of the proposed algorithm for an example network are demonstrated in Figure 2. The radius of the network circular area is  $R = 3$  km, and therefore, the outer limit of SF12 is  $l_6 = R = 3$  km. The first iteration of the algorithm defines the inner boundary of SF12,  $l_5$ , as shown in Figure 2. After excluding the devices inside the SF12 ring from  $ED_k$ , the algorithm runs a new iteration and defines  $l_4$ , i.e., the inner boundary of the SF11 ring. The iterations continue until  $l_1$  is defined and the complete network geometry is obtained.



**Figure 2.** K-means iterations for SF allocation based on Fibonacci series and 500 nodes, where  $l_i$  and  $l_{i+1}$  are the inner and outer radii of  $i^{\text{th}}$  annulus, respectively.



Figure 3 depicts the clusters of nodes, centroids, and gateway at the origin for the last K-means iteration. The nodes outside  $l_1$  are allocated to SF8, and the remaining nodes are assigned to SF7. The SF distribution of  $\bar{N} = 500$  nodes based on the proposed approach considering the Fibonacci series for clustering is shown in Figure 4. The number of nodes in SF rings depends on the chosen series because of the distinct number of clusters for each sequence.

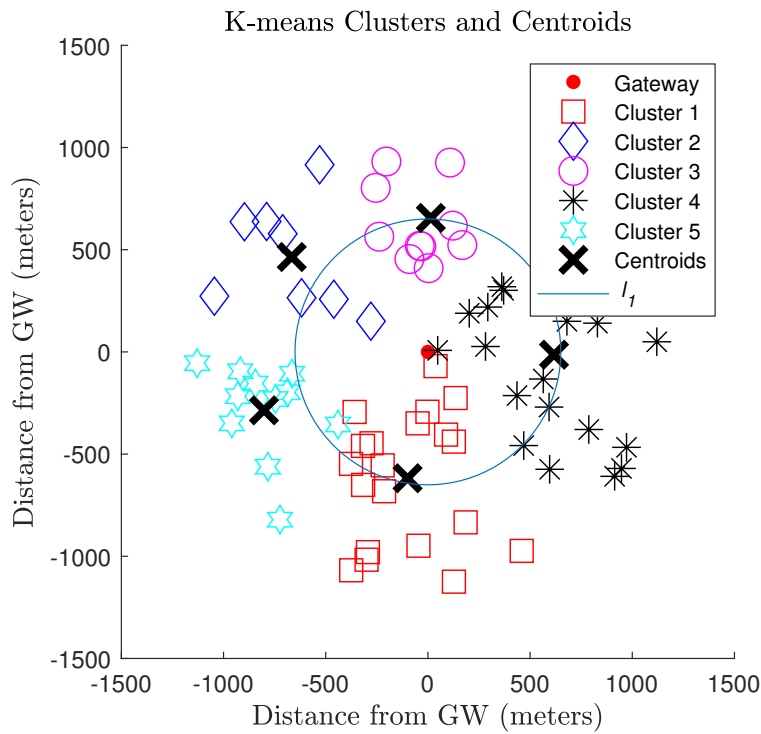


Figure 3. 5th iteration with Fibonacci series. Nodes outside  $l_1$  use SF8, those inside use SF7.

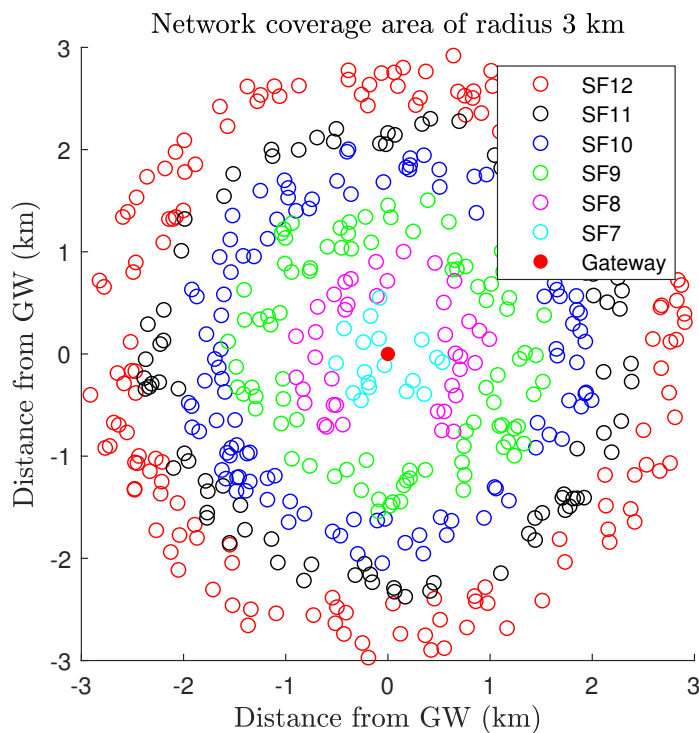


Figure 4. SF allocation of 500 nodes with Fibonacci series.

The area of each annulus also varies according to the mathematical series, which also affects the network performance. An important aspect to take into consideration is the selection of the number of clusters ( $K$ ); if the difference between the clusters of two consecutive iterations is too big, that will result in a large number of end-devices in that specific region and, as a consequence, the probability of collisions and of co-SF interference will be high. In the same way, SF7 will have a larger coverage area and more nodes if  $K$  is high for the last iteration.

## 5. Numerical Results and Discussion

In this section, we evaluate the scalability and performance of the proposed methodology by means of computer simulations. The results are based on a  $p_0 = 1\%$  duty cycle,  $BW = 125$  kHz,  $\eta = 2.75$  path loss exponent, 868 MHz European frequency band, and network radius of  $R = 3$  km. Table 3 summarizes the parameters considered for the results.

**Table 3.** System Parameters.

Parameter	Symbol	Value
Nodes	$\bar{N}$	300–700
Spreading Factor	SF	7–12
Bandwidth	$BW$	125 kHz
Carrier frequency	$f$	868 MHz
Noise figure	$NF$	6 dBm
Transmit power	$\mathcal{P}_1$	14 dBm
Duty cycle	$p_0$	1%
Path loss exponent	$\eta$	2.75

### 5.1. SF Allocation and Scalability Analysis

As discussed in Section 4, we used mathematical series to assign the numbers of clusters for K-means iterations. The Fibonacci series has the shortest ranges for SF7, 715 m for  $\bar{N} = 500$ . The distance between the SFs and the distribution of nodes can be changed by modifying the number of clusters ( $K$ ). On the other hand, for the same number of nodes, the Square series has a longer range for SF7, which is 1201 m, and it contains more end-devices. This type of configuration is due to the higher value of  $K$  clusters for iterations (see Table 2). While in the case of Fibonacci and equal-distance-based SF allocation in [19,20], the network has fewer nodes in SF7, and the number of nodes in each SF region increases considerably towards the higher SFs, as shown in Table 4. The average SF ranges for the Fibonacci series, square series, Wythoff array and arithmetic series are shown in Table 5.

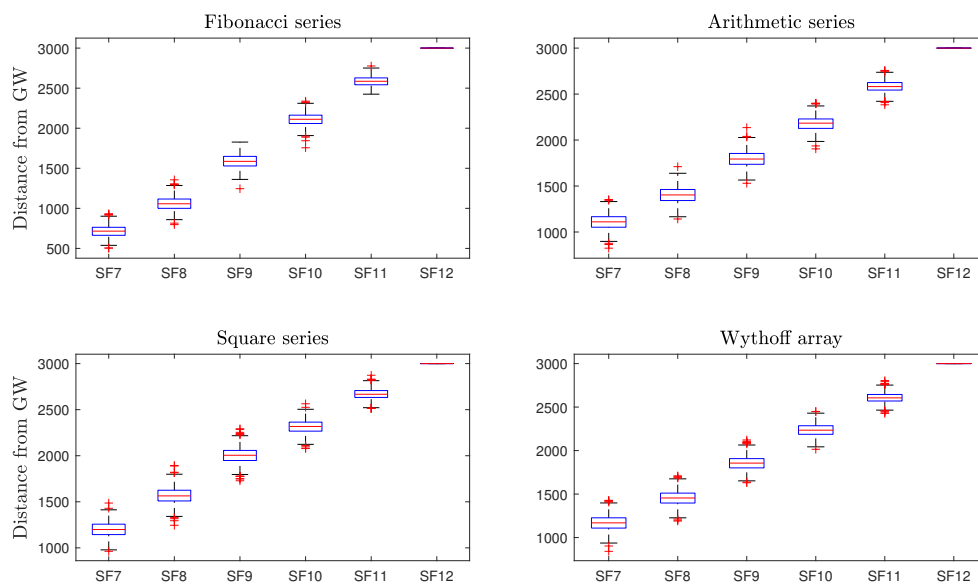
**Table 4.** Comparison of the proposed approach with the reference method, based on the number of nodes.

Series	$\bar{N}$	SF7	SF8	SF9	SF10	SF11	SF12
Fibonacci series	300	18	19	48	66	75	77
	500	29	35	79	107	124	129
	700	40	51	110	148	173	180
Arithmetic series	300	44	24	43	52	63	77
	500	69	42	70	87	107	129
	700	94	61	98	121	149	180
Square numbers	300	53	33	55	44	58	61
	500	81	57	87	75	99	105
	700	111	81	121	105	137	148
Wythoff array	300	49	24	45	52	60	73
	500	77	42	75	87	100	123
	700	103	62	103	121	140	173
Reference model	300	9	25	42	59	76	92
	500	14	42	70	98	126	152
	700	20	59	97	137	176	214

**Table 5.** Comparison of proposed approach with the reference method based on the distance of individual SF outer boundaries from the gateway (values in meters).

Series	$\bar{N}$	SF7	SF8	SF9	SF10	SF11	SF12
Fibonacci series	300	735	1049	1586	2112	2588	3000
	500	715	1060	1591	2112	2586	3000
	700	707	1071	1601	2112	2587	3000
Arithmetic series	300	1144	1412	1806	2194	2590	3000
	500	1110	1403	1795	2183	2584	3000
	700	1099	1409	1801	2188	2588	3000
Square numbers	300	1248	1591	2037	2336	2680	3000
	500	1201	1568	2004	2316	2670	3000
	700	1190	1568	2002	2313	2667	3000
Wythoff array	300	1209	1469	1872	2246	2613	3000
	500	1168	1453	1857	2237	2607	3000
	700	1150	1450	1851	2231	2604	3000
Reference model	300	500	1000	1500	2000	2500	3000
	500	500	1000	1500	2000	2500	3000
	700	500	1000	1500	2000	2500	3000

The boxplot is a standard process to quantify the variability of data on the basis of five parameters, i.e., the minimum, first quartile (25%), median, third quartile (75%), and maximum. The distance of the SFs boundary from the gateway for  $\bar{N} = 500$  is demonstrated in Figure 5 for each of the considered series.



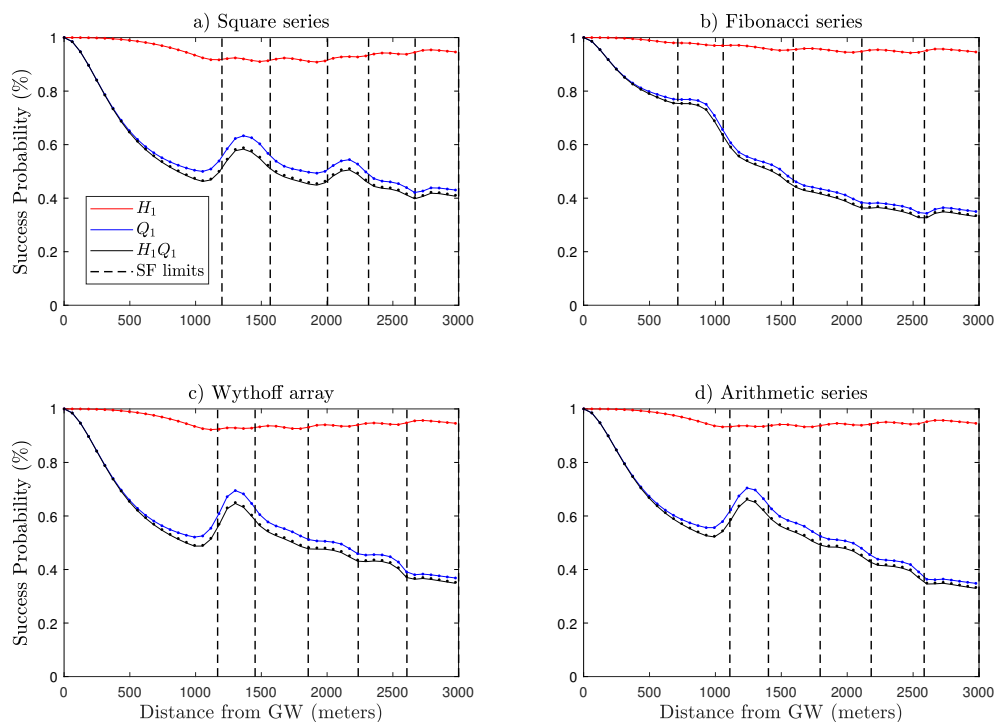
**Figure 5.** Boxplots representing the distance of SFs from the gateway. The red “+” signs indicate outliers. SF12 has a constant distance of 3 km for all of the series.

The median of every SF is identical to distances provided in Table 5. As depicted in the graphical results, SF7 and SF8 have a large disparity in range and number of nodes for different scenarios, while SF11 and SF12 have nearly close coverage areas for all scenarios. Furthermore, we can also clearly observe that variation in  $K$  clusters selection has a direct effect on SF allocation based on the proposed methodology. The number of nodes in each SF are illustrated in Table 4. Square-series-based networks demonstrate five times more nodes in SF7 as compared to the reference model.

The large values of  $K$  clusters for the last iterations result in longer radii that directly increase the region of SF7 and keep SF8 further away from the gateway. These different scenarios can be used according to different situations and requirements of LoRa applications. An approach based on the Fibonacci series is beneficial for applications where fewer nodes in lower SFs are required, while the Wythoff array, square series, and arithmetic series have wider regions for SF7, and thus can be used in setups where more nodes are required in SF7 inside the radio range of nearly 1200 m to provide highest data rate ( $R_b = 5.47$  kbps, see Table 1). Several studies examined the performance of LoRa networks and show that the success probability of data packets decreases for higher SFs. In our work, we consider the effect of changing the coverage range and varying the number of devices for individual SFs.

## 5.2. Performance Analysis

After the application of the proposed SF allocation algorithm, we investigated the performance of the resulting LoRa network. The theoretical results were verified by Monte Carlo simulations. In the figures, each marker represents the average over  $10^5$  random deployments of the Poisson point process (PPP) for a single gateway LoRa uplink model, considering an end-device at  $d_1$  meters from the gateway. In Figure 6, the solid lines demonstrate the theoretical results, while marker points of the same color illustrate the simulation outputs. The simulated results align with the theoretical ones. Within the context of the previously discussed mathematical sequences, we considered and examine the impact of different SF allocation scenarios on connection probability  $H_1$ , capture probability  $Q_1$ , and coverage probability  $H_1Q_1$  against the distance from the gateway.

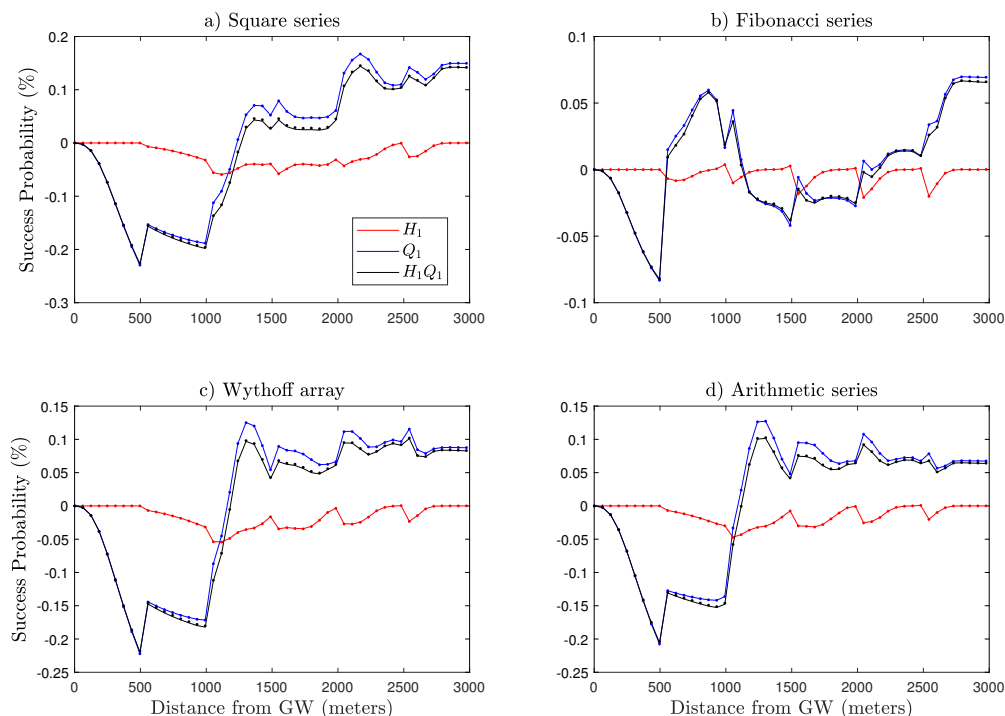


**Figure 6.** LoRa uplink performance with the proposed SF allocation algorithm for different series and  $N=500$  nodes. (a) Square series with  $K = \{49, 36, 25, 16, 9\}$ . (b) Fibonacci series with  $K = \{34, 21, 13, 8, 5\}$ . (c) Wythoff with  $K = \{37, 32, 24, 16, 11\}$ . (d) Arithmetic series with  $K = \{34, 28, 22, 16, 10\}$ .

As expected, the distance of the end-device from the gateway has considerable influence on connection probability  $H_1$ . In the case of the Fibonacci series, the model has a better success probability for lower SFs as compared to the square series, arithmetic series, and Wythoff array. This fact is

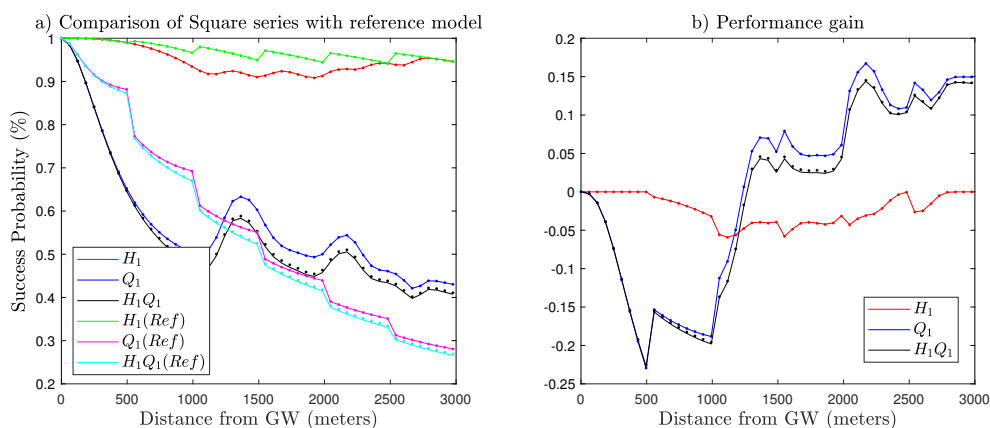
due to the difference in a range of individual SF boundaries for said SF allocation schemes and the distance-dependent SNR threshold  $q_{SF}$ . Furthermore, in the scenario of the square series, SF7 has large coverage areas of 1201 m (see Table 5), which affects path loss attenuation and the instantaneous SNR. The SNR threshold  $q_{SF}$ , however, remains the same at  $-6$  dB (see Table 1). As a consequence, the outage condition in (3) slightly degrades the network connection probability. Although a performance boost is illustrated during transitions of end-devices into the next SF because of the lower value of  $q_{SF}$ , the performance of the previous SF has a direct consequence on the next SF.

Moving towards the capture probability  $Q_1$ , unlike  $H_1$ , it considers co-SF interference.  $Q_1$  declines gradually with increasing SF, as illustrated in Figure 6. This trend is because of two major factors including ToA and the number of nodes in each annulus. ToA grows exponentially with SF, thus for the higher SFs, the wireless channel remains occupied for a long time slot, which increases the risk of collisions between simultaneously transmitted LoRa packets. In the same way, the number of end-devices in an individual annulus increases for higher SFs due to the uniform distribution of nodes in the circular coverage area, as demonstrated in Table 4. As a result, the network experiences co-SF interference that degrades the quality of transmission. For the cases with the square series, arithmetic series, and Wythoff array (Figure 6a,c,d), the model has a larger coverage area and more nodes in SF7, resulting in higher co-spreading factor interference, which is the major reason behind the lower performance of the network for these specific scenarios. Although we are sacrificing network quality for lower SFs with fewer nodes and high success probabilities, as presented in Figure 7, we improved the performance of the network for the higher SFs and regions with more nodes, where the network performance was weak in the baseline model from [19], which considers fixed distance steps from the gateway to define the SF allocation. Moreover, SF allocation based on the square series (Figure 6a) has better network performance, which happens because of improved gain in  $Q_1$  for higher SFs as compared to the Fibonacci series, arithmetic series, and Wythoff array.



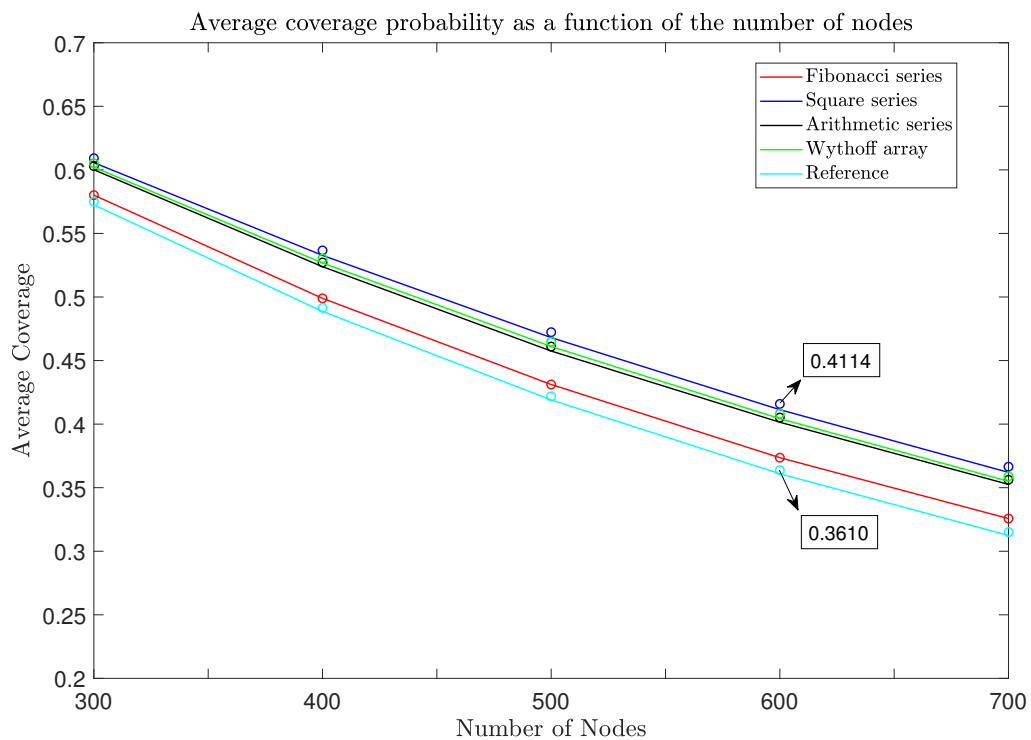
**Figure 7.** Comparison of performance gain of different series with respect to the baseline model for  $\bar{N} = 500$ , network size of  $R = 3$  km with  $p_0 = 1\%$ , and path loss exponent  $\eta = 2.75$ .

In Figure 8, we present the performance of the square series based on the proposed SF allocation algorithm in comparison with the baseline model. First, we observe that the capture probability  $Q_1$  and coverage probability  $H_1Q_1$  of the baseline model outperform the proposed algorithm in the region of radius  $R > 1000$  m from the gateway. Here, it is worth mentioning that there are fewer nodes in this region as compared to the remaining area of the network. The proposed algorithm surpasses the baseline model with a gain in capture probability  $Q_1$  (up to 53% for SF12 in Figure 8b). In addition, the nodes closer to the gateway always have better behavior in contrast with nodes far away, so that the network can tolerate a lower success probability with a boost in performance of farther end-devices. As expected, there was more change in success probability in the higher SFs region as compared to baseline model because of the fair distribution of SF by the proposed algorithm. Figure 8b shows the difference between the baseline model and the square series in terms of outage probabilities through the course of distance. The zero level on the y-axis (success probability) shows no difference, while there is a positive/negative gain either on the upper or lower side of that level. There is up to 16.73% growth in  $Q_1$  demonstrated by the end-devices present in higher SFs.



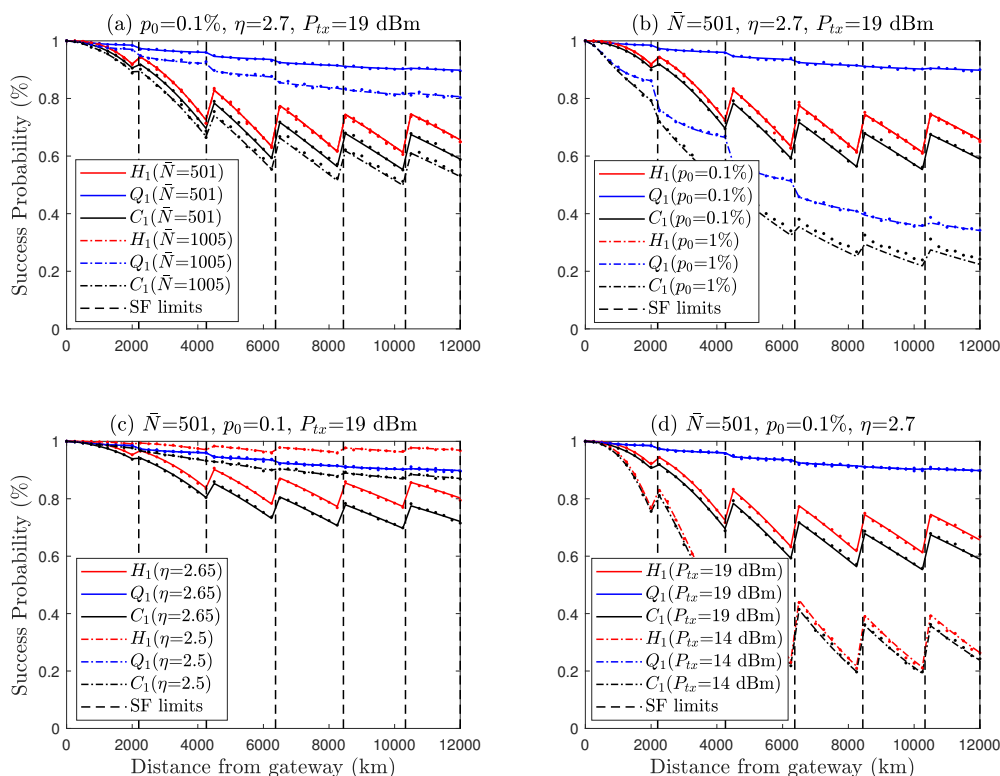
**Figure 8.** (a) Performance comparison of square series with the reference model for  $\bar{N} = 500$ . (b) Performance gain (success probability of proposed model – success probability of reference model). The proposed SF allocation approach sacrifices  $Q_1$ ,  $H_1Q_1$  for the lower SFs but achieves greater performance for the higher SFs.

Moreover, we also consider the evaluation of the average coverage probability of the networks for different numbers of nodes ( $\bar{N}$ ) ranging from 300 to 700, and results demonstrate that  $\varphi_c$  drops exponentially towards higher  $\bar{N}$ . Figure 9 depicts the average coverage probability of different numbers of end-devices. The SF allocation schemes deployed using the square series demonstrated a better performance gain than all other scenarios, including the reference model. It was followed by the Wythoff array, arithmetic series, and Fibonacci series, in that order. The proposed SF allocation scheme overcomes the performance of the baseline model by an overall growth of around 5% in its  $\varphi_c$ . For instance, taking the square series into account, at  $\bar{N} = 500$  there is a boost in the average coverage probability (6) from  $\varphi_c = 41.9\%$  to 46.81% compared with the reference model. On the other hand, SF allocation schemes deployed using the Fibonacci series showed the least-improved network coverage probability compared to all other user distribution series.



**Figure 9.** Comparison of the coverage probabilities of the LoRa uplink as a function of the number of nodes ranging from  $\bar{N} = 300$  to 700 nodes for the network size of  $R = 3$  km with  $p_0 = 1\%$  and path loss exponent  $\eta = 2.75$ .

We also investigated the performance of the proposed model taking into account the variation of different parameters on Fibonacci series. As seen in Figure 10a,b,  $H_1$  is agnostic to the number of nodes and duty cycle. It is assessed at 0.1% and 1%, which is within the duty cycle range specified by ETSI for LoRa applications [34,35]. Furthermore, node density and the duty cycle demonstrate a negative impact on  $Q_1$  because of co-SF interference caused by increasing medium usage from  $\bar{N} = 501$  nodes to  $\bar{N} = 1005$  nodes. Likewise, in Figure 10c, the path loss exponent  $\eta$  illustrates network connection degradation at 2.65 from 2.5 because  $H_1$  depends on the distance, while the the capture probability ( $Q_1$ ) is not dependent on the path loss exponent. In the case of one frequency channel, the transmit power  $\mathcal{P}_k$  of nodes can be up to 20 dBm (100 mW) [11,41]. In order to evaluate the effect of different transmit powers, in Figure 10d we raised the transmit power from 14 dBm to 19 dBm. The results demonstrate that the transmit power of 19 dBm causes a better connection probability as compared to 14 dBm. Nevertheless, variations of the path loss exponent and transmit power do not affect  $Q_1$  considerably because it is much more dependent on the number of nodes.



**Figure 10.** Performance LoRa uplink considering Fibonacci series K-means clustering, and the impact of different parameters on success probabilities. (a) Density of users increased from  $\bar{N} = 501$  to  $\bar{N} = 1005$ . (b) Duty cycle increased from  $p_0 = 0.1\%$  to  $p_0 = 1\%$ . (c) Path loss exponent from  $\eta = 2.5$  to  $\eta = 2.65$ . (d) Transmit power from ( $P_{tx} = 14$ ) dBm to  $P_{tx} = 19$  dBm.

### 5.3. Discussion

The architecture of LoRaWAN consists of end-devices, a gateway, network server (NetServer), and application server [15]. The NetServer is mainly responsible for the overall management of the network. The dynamic configuration of the SF by the NetServer is already possible in LoRaWAN during the network join procedure or through specific MAC commands. In fact, these features are used in LoRaWAN when the adaptive data rate (ADR) mechanism is active. For our approach to be implemented in practice, the NetServer could run our algorithm periodically or when the number of connected devices changes significantly, and then issue the required MAC commands to reconfigure the devices that need to change their SF. Therefore, as the current LoRaWAN specification is already able to dynamically allocate SFs, our proposal only changes the way the proper allocation is calculated at the NetServer, and is therefore feasible in practice.

## 6. Conclusions

This paper has presented a novel SF allocation technique for a large-scale LoRa network using the K-means clustering machine learning algorithm. The authors also analyzed the impact of the distance of end-devices from the gateway and the number of nodes in each SF on network performance. In this work, four different scenarios are considered, which have different distances for the SF boundaries and variations in the number of nodes in an individual SF. Such fair distribution results in a better average coverage probability in the higher SFs, while dealing with the maximum number of nodes. Numerical findings show that our SF allocation algorithm outperforms the reference model not only in terms of success probability but also in regards to fair resource distribution. The evaluated theoretical and simulation results are useful for an in-depth understanding of large and dense LoRa



networks. Our resource allocation method can handle dense and large circular coverage areas for LoRa sensors using distinct numbers of clusters instead of equal-radius-based SF allocation [19,20], while the techniques in [28,29,33] are designed for short-range networks. The studies [19,20,28,29,33] only highlighted the performance of networks for fixed parameters. In contrast to them, our work inspects different scenarios by obeying the restrictions of ETSI standards.

**Author Contributions:** Concept and methodology, H.A.; software, M.A.U. and J.I.; validation, A.H. and H.A.; Resources, A.H. and R.D.S.; writing original draft, M.A.U. and J.I.; writing revision, H.A., A.H., and R.D.S.; supervision, H.A.; Project Administration, H.A.

**Funding:** This research has been supported in Finland by the Academy of Finland: 6Genesis Flagship (Grant no318927) and EE-IoT (no319008); as well as the BusinessFinland MOSSAF project. This work has been supported in Brazil by CNPq, PrInt CAPES-UFSC “Automation 4.0”, and INESC P&D Brasil (Project F-LOCO, Energisa, ANEEL PD-00405-1804/2018).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Chen, S.; Xu, H.; Liu, D.; Hu, B.; Wang, H. A Vision of IoT: Applications, Challenges, and Opportunities With China Perspective. *IEEE Internet Things J.* **2014**, *1*, 349–359.
- Sisinni, E.; Saifullah, A.; Han, S.; Jennehag, U.; Gidlund, M. Industrial Internet of Things: Challenges, Opportunities, and Directions. *IEEE Trans. Ind. Inf.* **2018**, *14*, 4724–4734.
- Sotres, P.; Lanza, J.; Sánchez, L.; Santana, J.R.; López, C.; Muñoz, L. Breaking Vendors and City Locks through a Semantic-enabled Global Interoperable Internet-of-Things System: A Smart Parking Case. *Sensors* **2019**, *19*, 229.
- Rodrigues, J.J.P.C.; De Rezende Segundo, D.B.; Junqueira, H.A.; Sabino, M.H.; Prince, R.M.; Al-Muhtadi, J.; De Albuquerque, V.H.C. Enabling Technologies for the Internet of Health Things. *IEEE Access* **2018**, *6*, 13129–13141.
- Asadullah, M.; Ullah, K. Smart home automation system using Bluetooth technology. In Proceedings of the 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, Pakistan, 5–7 April 2017; pp. 1–6.
- Saha, H.N.; Mandal, A.; Sinha, A. Recent trends in the Internet-of-Things. In Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9–11 January 2017; pp.1–4.
- Internet of Things (IoT) Connected Devices Installed Base Worldwide from 2015 to 2025 (in Billions). Available online: <https://www.statista.com> (accessed on 16 October 2019).
- Yasmin, R.; Petäjäjärvi, J.; Mikhaylov, K.; Pouttu, A. On the integration of LoRaWAN with the 5G test network. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017; pp. 1–6.
- Mikhaylov, K.; Petrov, V.; Gupta, R.; Lema M.A.; Galinina, O.; Andreev, S.; Koucheryavy, Y.; Valkama, M.; Pouttu, A.; Dohler, M. Energy Efficiency of Multi-Radio Massive Machine-Type Communication (MR-MMTC): Applications, Challenges, and Solutions. *IEEE Commun. Mag.* **2019**, *57*, 100–106.
- Vangelista, L.; Zanella, A.; Zorzi, M. Long-range IoT technologies: The dawn of LoRa. In Proceedings of the Future Access Enablers for Ubiquitous and Intelligent Infrastructures, Cham, Switzerland, 23–25 September 2015; pp.51–58.
- Petajajarvi, J.; Mikhaylov, K.; Roivainen, A.; Hanninen, T.; Pettissalo, M. On the coverage of LPWANs: Range evaluation and channel attenuation model for LoRa technology. In Proceedings of the 2015 14th International Conference on ITS Telecommunications (ITST), Copenhagen, Denmark, 2–4 December 2015; pp. 55–59.
- Lauridsen, M.; Vejlgard, B.; Kovacs, I.Z.; Nguyen, H.; Mogensen, P. Interference Measurements in the European 868 MHz ISM Band with Focus on LoRa and SigFox. In Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, USA, 19–22 March 2017; pp. 1–6.
- Cattani, M.; Boano, C.A.; Romer, K. An experimental evaluation of the reliability of lora long-range low-power wireless communication. *J. Sensor Actuator Netw.* **2017**, *6*, 7.
- AN1200.22 LoRa™ Modulation Basics, Rev. 2. Semtech, 2015. Available online: <http://www.semtech.com/uploads/documents/an1200.22.pdf> (accessed on 31 July 2019).

15. LoRa Alliance. Available online: <http://www.lora-alliance.org> (accessed on 30 September 2019).
16. Phung, K.H.; Tran, H.; Nguyen, Q.; Huong, T.T.; Nguyen, T.L.; Nguyen, H. Analysis and assessment of LoRaWAN. In Proceedings of the 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), Ho Chi Minh City, Vietnam, 29–31 January 2018; pp. 241–246.
17. LoRa Technology Is Connecting Our Smart Planet. Available online: <https://www.semtech.com/lora/lora-applications> (accessed on 31 July 2019).
18. de Castro Tomé, M.; Nardelli, P.H.J.; Alves, H. Long-Range Low-Power Wireless Networks and Sampling Strategies in Electricity Metering. *IEEE Trans. Ind. Electron.* **2019**, *66*, 1629–1637.
19. Georgiou, O.; Raza, U. Low-Power Wide-Area Network Analysis: Can LoRa Scale? *IEEE Wirel. Commun. Lett.* **2017**, *6*, 162–165.
20. Hoeller, A.; Souza, R.D.; Alcaraz López, O.L.; Alves, H.; de Noronha Neto, M.; Brante, G. Analysis and Performance Optimization of LoRa Networks With Time and Antenna Diversity. *IEEE Access* **2018**, *6*, 32820–32829.
21. Augustin, A.; Yi, J.; Clausen, T.; Townsley, W. A Study of LoRa: Long-Range & Low-Power Networks for the Internet-of-Things. *Sensors* **2016**, *16*, 1466.
22. Haxhibeqiri, J.; De Poorter, E.; Moerman, I.; Hoebeke, J. A Survey of LoRaWAN for IoT: From Technology to Application. *Sensors* **2018**, *18*, 3995.
23. Mikhaylov, K.; Petaejaevaervi, J.; Haenninen, T. Analysis of Capacity and Scalability of the LoRa Low-Power Wide-Area Network Technology. In Proceedings of the 22th European Wireless Conference, Oulu, Finland, 18–20 May 2016; pp. 1–6.
24. Haxhibeqiri, J.; Karaagac, A.; Van den Abeele, F.; Joseph, W.; Moerman, I.; Hoebeke, J. LoRa indoor coverage and performance in an industrial environment: Case study. In Proceedings of the 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, 12–15 September 2017; pp. 1–8.
25. Lavric, A.; Petrariu, A.I.; Popa, V. Long Range SigFox Communication Protocol Scalability Analysis Under Large-Scale, High-Density Conditions. *IEEE Access* **2019**, *7*, 35816–35825.
26. Mekki, K.; Bajic, E.; Chaxel, F.; Meyer, F. A comparative study of LPWAN technologies for large-scale IoT deployment. *ICT Express* **2019**, *5*, 1–7.
27. Noreen, U.; Bounceur, A.; Clavier, L. A study of LoRa low-power and wide-area network technology. In Proceedings of the 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, Morocco, 22–24 May 2017; pp. 1–6.
28. Cuomo, F.; Campo, M.; Caponi, A.; Bianchi, G.; Rossini, G.; Pisani, P. EXPLoRa: Extending the performance of LoRa by suitable spreading factor allocations. In Proceedings of the 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Rome, Italy, 9–11 October 2017; pp. 1–8.
29. Cuomo, F.; Gámez, J.C.C.; Maurizio, A.; Scipione, L.; Campo, M.; Caponi, A.; Bianchi, G.; Rossini, G.; Pisani, P. Towards traffic-oriented spreading factor allocations in LoRaWAN systems. In Proceedings of the 2018 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), Capri, Italy, 20–22 June 2018; pp. 1–8.
30. Kim, D.Y.; Kim, S. Data Transmission Using K-Means Clustering in Low Power Wide Area Networks with Mobile Edge Cloud. *Wirel. Pers. Commun.* **2018**, *105*, 567–581.
31. Farhad, A.; Kim, D.; Pyun, J. Scalability of LoRaWAN in an Urban Environment: A Simulation Study. In Proceedings of the 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), Zagreb, Croatia, 2–5 July 2019; pp. 677–681.
32. Magrin, D.; Centenaro, M.; Vangelista, L. Performance evaluation of LoRa networks in a smart city scenario. In Proceedings of the 2017 IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017; pp. 1–7.
33. Zhu, G.; Liao, C.-H.; Sakdejayont, T.; Lai, I.-W.; Narusue, Y.; Morikawa, H. Improving the Capacity of a Mesh LoRa Network by Spreading-Factor-Based Network Clustering. *IEEE Access* **2019**, *7*, 21584–21596.
34. Vincenzo, V.D.; Heusse, M.; Tourancheau, B. Improving Downlink Scalability in LoRaWAN. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–7.

35. Hasegawa, Y.; Suzuki, K. A Multi-User ACK-Aggregation Method for Large-Scale Reliable LoRaWAN Service. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–7.
36. Li, L.; Ren, J.; Zhu, Q. On the application of LoRa LPWAN technology in sailing monitoring system. In Proceedings of the 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS), Jackson, WY, USA, 21–24 February 2017; pp. 77–80.
37. Yasmin, R.; Petäjäjärvi, J.; Mikhaylov, K.; Pouttu, A. Large and Dense LoRaWAN Deployment to Monitor Real Estate Conditions and Utilization Rate. In Proceedings of the 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, Italy, 9–12 September 2018.
38. Abramson, N. The Aloha System—Another alternative for computer communications. In Proceedings of the Fall Joint Computer Conference (AFIPS '70 (Fall)), New York, NY, USA, 17–19 November 1970; pp. 281–285.
39. Goursaud, C.; Gorce, J.M. Dedicated networks for IoT: PHY/MAC state of the art and challenges. *EAI Endorsed Trans. Internet Things* **2015**, *1*, doi: 10.4108/eai.26-10-2015.150597
40. Na, S.; Xumin, L.; Yong, G. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. In Proceedings of the 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, Jingtangshan, China, 2–4 April 2010; pp. 63–67.
41. Bouguera, T.; Diouris, J.F.; Chaillout, J.J.; Jaouadi, R.; Andrieux, G. Energy Consumption Model for Sensor Nodes Based on LoRa and LoRaWAN. *Sensors* **2018**, *18*, 2104.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).