

Article

Contention-Aware Adaptive Data Rate for Throughput Optimization in LoRaWAN

Sungryul Kim  and Younghwan Yoo * 

School of Electrical and Computer Engineering, Pusan National University, Busan 46241, Korea;
xmfhxm12@pusan.ac.kr

* Correspondence: ymomo@pusan.ac.kr; Tel.: +82-51-510-3633

Received: 31 March 2018; Accepted: 24 May 2018; Published: 25 May 2018



Abstract: In Long Range Wide Area Network (LoRaWAN), the data rate of the devices can be adjusted to optimize the throughput by changing the spreading factor. However, the adaptive data rate has to be carefully utilized because the collision probability, which directly affects the throughput, is changed according to the use of spreading factors. Namely, the greater the number of devices using the same spreading factor, the greater the probability of collision, resulting in a decrease of total throughput. Nevertheless, in the current system, the only criteria to determine the data rate to be adjusted is a link quality. Therefore, contention-aware adaptive data rate should be designed for the throughput optimization. Here, the number of devices which can use a specific data rate is restricted, and accordingly the optimization problem can be regarded as constrained optimization. To find an optimal solution, we adopt the gradient projection method. By adjusting the data rate based on the retrieved set of optimal data rate, the system performance can be significantly improved. The numerical results demonstrate that the proposed method outperforms the comparisons regardless of the number of devices and is close to the theoretical upper bound of throughput.

Keywords: adaptive data rate; contention-aware; gradient projection method; LoRaWAN; throughput optimization

1. Introduction

During the past decade we became witnesses of the evolution of the Internet of Things (IoT) technology. The IoT provides a useful service platform by connecting various things such as devices, objects, animals and plants [1]. Recently, the value of small data has been garnering attention in the IoT industry [2,3]. The things that transmit a small amount of data such as temperature, humidity, weight, and location are classified as small things, and the network composed of such small things is called the Internet of Small Things (IoST). Since the exchanged data size in IoST is very small, a low-cost and low-performance processor is able to sufficiently meet the application requirements. In addition, the small things sparsely transmit the sensing data, and consequently the device can operate for several years with just one battery. These advantages encourage developers and companies to freely release their IoT-based services.

The IoST network can be divided into two categories, local area network and wide area network. Local area network, which has a short communication range, provides a relatively high data rate, e.g., Radio Frequency Identification (RFID) system, Bluetooth Low Energy (BLE) beacon, and Near Field Communication (NFC). Meanwhile, Low-Power Wide Area Network (LPWAN) is characterized by the long-range communication and the low data rate. The gateway is able to use the sensing data from the end-devices distributed in a city, thus the LPWAN technology is suitable for the purpose of collecting small amounts of data scattered over a large area. For instance, metering such as electricity and gas, remote management of livestock, smart traffic management, and waste

collection scheduling. Moreover, this can be adopted for the networking of Unmanned Aerial Vehicle (UAV)-based system, and those collaborative operations will create innovations [4,5]. Accordingly, many standards related to the LPWAN have been introduced, and LoRaWAN and SIGFOX are considered as the leading group [6,7].

The LoRaWAN provides the higher data rate as compared with SIGFOX while consuming the lower amount of energy. Furthermore, in contrast to the SIGFOX adopting Gaussian Frequency-Shift Keying (GFSK) and Binary Phase Shift Keying (BPSK) for the modulation [8], the LoRaWAN adopts Chirp Spread Spectrum (CSS), and it makes LoRa signal more robust to the interference and Doppler shifting [9]. Meanwhile, in the LoRaWAN, the signal length is determined by Spreading Factor (SF) indicating the number of encoded bits per symbol. In detail, the bigger SF is, the longer the signal length becomes. Since the minimum Signal-to-Noise Ratio (SNR) required for successful decoding increases as the signal length increases, the devices should use the larger SF as it increases communication range. One notable thing is that the data rate is also determined by the SF . In other words, the gateway can control the data rate of each device by adjusting SF , i.e., Adaptive Data Rate (ADR) is permitted [6]. The current LoRaWAN standard gives a simple guideline; the gateway is able to change the data rate according to the link quality retrieved from SNR. However, the detailed descriptions for the ADR have not yet been concretely established.

This paper proposes a contention-aware adaptive data rate to maximize the total throughput in the LoRaWAN. For higher throughput, it seems to be reasonable that all end-devices just use the smallest SF among available SFs . However, the more end-devices that use the same SF , the greater the probability of collision, resulting in decreasing throughput [10]. This is due to the fact that the signals generated by different SFs can be separated even though they overlap at the gateway owing to its orthogonality. In contrast, when the signals generated by the same SF are overlapped, the collision happens. Therefore, not only link quality, but also the impact of changed SFs on the contention has to be considered in ADR. For this purpose, we define the total throughput as an objective function with respect to the number of devices using specific SFs . Accordingly, finding a set of the number of devices per SF that maximizes the objective function is the goal of this paper. That is, we derive the theoretical optimal throughput that can be achieved by adjusting only the data rate of the devices. Here, the available SF of each device is limited, depending on the communication range. Therefore, the maximization problem is categorized into the constrained optimization. To solve the problem, we adopt the gradient projection method that finds the optimal solution while keeping the constraints. As a result, owing to the mitigation of the contention, our method can achieve optimal throughput in the given environments. For the performance evaluation, we assume two comparable ADR schemes based on the guidance of the LoRaWAN. In the first scheme, all devices change their data rate to the fastest data rate by using the smallest possible SF . This called the naive approach. This can be considered to be very similar to the method described in the current standard. The second approach is that the number of devices using a specific SF is made equal, named as the uniform approach. This can alleviate the contention problem by inducing the devices to use the SFs evenly. The numerical results show that our method outperforms the comparisons.

The contribution of this paper is twofold as follows.

- To the best of our knowledge, this is the first ADR considering the contention problem in the LoRaWAN.
- For the optimization, we adopt the gradient projection method. This allows the gateway to easily find the optimal set under the constraints regardless of the number of end-devices deployed in the network.

The rest of this paper is organized as follows. Section 2 briefly introduces basic knowledge of the LoRaWAN. Section 3 describes the contention problem caused by the adjustment of data rate. Section 4 presents the proposed method, contention-aware ADR and how to find the optimal solution with the gradient projection method. The numerical results show that the proposed method supports higher

throughput as compared with the comparisons and it is close to the upper-bound of the performance. Finally, we conclude our paper with a summary and future works.

2. Background Knowledge

This section discusses overall features of the LoRaWAN. First, we describe the modulation in the LoRaWAN, CSS, and illustrate how to adaptively adjust the data rate by changing the SF . In addition, pure ALOHA protocol and Frequency Hopping Spread Spectrum (FHSS) are briefly explained to understand how the LoRaWAN handles the multiple access. Meanwhile, the LoRaWAN standards are slightly different from region to region, and we consider EU863-870 (European standard) which operates in the Industrial Scientific Medical (ISM) band, 863–870 MHz.

2.1. Chirp Spread Spectrum

In the LoRaWAN, CSS modulation is used to generate the signal. The frequency of the CSS signal either linearly increases or decreases over the whole bandwidth, and this makes the signal robust against the interference and the Doppler shifting [11,12]. The number of bits encoded per symbol is decided by the spreading factor ranging from 7 to 12 in CSS modulation. Let R_s and R_b denote the symbol rate and bit rate respectively, and then the SF can be defined as below:

$$SF = \frac{R_b}{R_s}. \quad (1)$$

In addition, the symbol length T_s is given as

$$T_s = \frac{2^{SF}}{BW}, \quad (2)$$

where BW indicates the system bandwidth. The symbol rate is inverse of the symbol length, consequently (1) can be modified as follows.

$$R_b(SF) = SF \times R_s = \frac{SF}{2^{SF}} BW. \quad (3)$$

The data rates and symbol lengths according to the used SF are presented in Table 1.

Table 1. Data rate according to the SFs .

SF	R_b (bps)	R_b (bps) FEC with 4/5	T_s (ms)
$SF7$	6835.94	5468	1.024
$SF8$	3906.25	3125	2.048
$SF9$	2197.27	1757	4.096
$SF10$	1220.70	976	8.192
$SF11$	671.39	537	16.384
$SF12$	366.21	292	32.768

Figure 1 shows the spectrogram of the symbol in the case of $SF8$, $SF10$, and $SF12$ respectively. As shown in the figure, the symbol length, i.e., Time on Air (ToA) is proportional to SF . Typically, as the symbol duration increases, the required SNR for decoding decreases. Therefore, the device near the gateway is able to use all kinds of SF owing to the loose constraint of SNR, but the device far from the gateway is forced to use only large SFs . If the gateway estimates that a certain device is able to use the smaller SF as compared with the current one, it tries to adjust the data rate to maximize the total throughput, and this is referred to as ADR in the LoRa system.

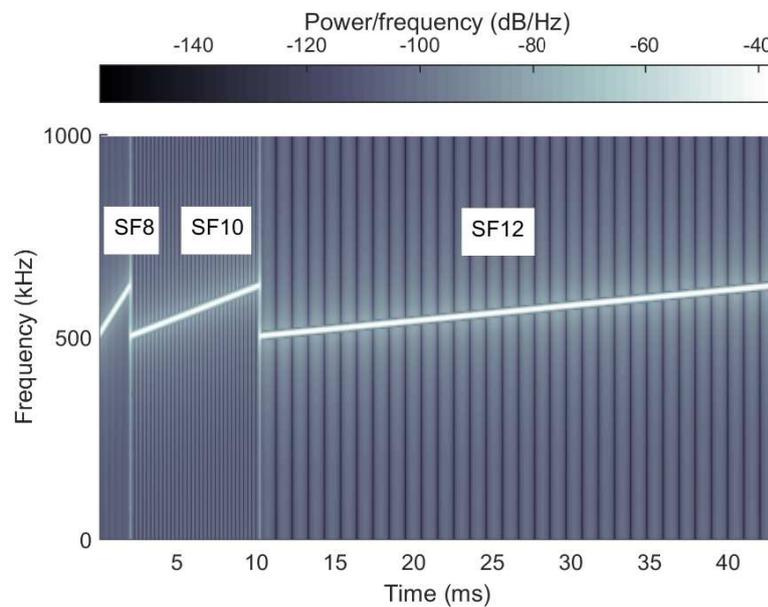


Figure 1. The spectrogram of the symbols generated by SF8, SF10 and SF12.

2.2. Adaptive Data Rate

The LoRaWAN allows the devices to use any of the possible data rates. By adjusting the data rate of each device, the throughput of the LoRaWAN can be optimized. As mentioned above, the data rate is determined by the used *SF*, consequently adjusting the data rate to select one of *SFs* to be used among available *SFs*.

ADR is initiated by the gateway. The gateway retrieves available minimum *SF* based on the SNR. After that, the gateway informs the device of the target *SF* via a MAC command named LinkADRReq. One notable thing is that this mechanism is applicable only to the devices that set ADR bit to 1, otherwise the ADR cannot be applied regardless of the link quality [6].

Meanwhile, after adjusting the data rate, it has to be verified whether the changed *SF* is appropriate because it is possible that the gateway may excessively recommend too small of *SF* for the data rate enhancement. To verify that the gateway still successfully receives the transmitted signal generated by the adjusted *SF*, the device increments a counter, referred to as ADR_ACK_CNT by one for each transmission, and this counter is reset whenever the device receives the response from the gateway. If ADR_ACK_CNT reaches the pre-defined threshold, ADR_ACK_LIMIT, due to the consecutive communication failures, the device sets ADRACKReq on the MAC command field in the next transmission. If there is no response from the gateway within a time duration specified at ADR_ACK_DELAY, the changed *SF* is considered as too small, and thus the devices increment *SF* for more reliable communication. This procedure is repeated until the device finds an appropriate *SF* or the adjusted *SF* is equal to the default one.

In EU863-870, the default setting of ADR_ACK_LIMIT is 64. This means that once the data rate is incorrectly adjusted, it takes a long time to be corrected at the proper data rate. Therefore, the data rate should be carefully adjusted from the perspective of the entire network rather than from the perspective of the individual devices.

2.3. Medium Access Control in the LoRaWAN

The strategies to control the multiple medium access can be explained with pure ALOHA, FHSS, and signal orthogonality. Due to the philosophy of the LoRaWAN, i.e., low performance and low power, the complicate MAC protocols are not preferred [5]. Instead of some advanced MAC protocols, the LoRaWAN adopts Pure ALOHA, i.e., broadcasting and reply mechanism [13]. The slotted

ALOHA is also inappropriate because it requires time synchronization consuming additional battery and computational resources. For the more efficient medium access, the Listen Before Talk (LBT) mechanism is applicable, but it is ignored in this paper to focus on the impact of ADR on the throughput optimization.

Along with the pure ALOHA, the LoRaWAN uses FHSS to mitigate the collision problem [14]. Therefore, each device attempts to send their packet on a channel randomly selected every transmission. Even though the multiple signals arrive at the gateway at the same time, they can be successfully decoded as long as the used channels are different each other. According to EU863-870, a total of six channels are available. Three channels working on 868.10, 868.30, 868.50 MHz are regarded as default channels and all devices are obliged to support those channels. In addition, the gateway should always be listening on the channels. The remaining channels working on 868.10, 868.30, 868.50 MHz are mainly used by the devices to broadcast the JoinReq message.

One notable point is that chirp rate describing how fast the frequency changes during the symbol duration can also be utilized to avoid packet collision. The signal generated with different chirp rates are near orthogonal each other, thus they can be separated, even though the arrival times and the used channels are the same [15]. The chirp rate is defined as

$$k = \frac{BW^2}{2SF}. \quad (4)$$

As shown in the above equation, for the different chirp rates, we can change the bandwidth or SF . Actually, the bandwidth of all channels is fixed at 125 kHz, consequently the only method to get a unique chirp rate is to adjust SF , which determines the data rate.

Figure 2 illustrates multiple access control in the LoRaWAN. For simplicity, we assume that the number of available SFs is two, $SF7$ and $SF12$. In case 1, although two signals generated with $SF12$ arrive at the same time, they can be successfully decoded because of different channel usage. In case 2, owing to the orthogonality, the collision can be avoided, even though multiple devices use the same channel. However, the collision occurs in case 3 since the two signals generated with the same SF ($SF7$) arrive on the same channel. In conclusion, a collision occurs only when the multiple signals generated with the same SF overlap on the same channel.

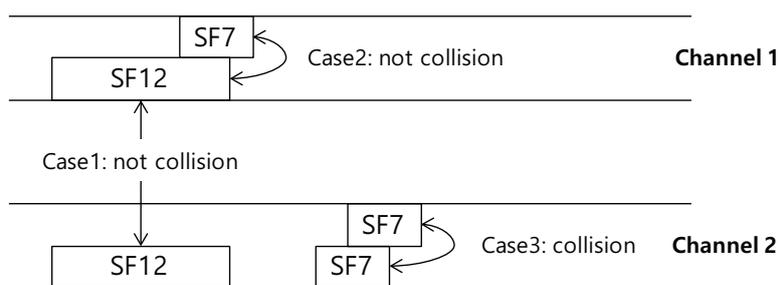


Figure 2. Multiple access control in the pure-ALOHA based the LoRaWAN.

3. Impact of ADR on the Contention

As mentioned above, the SF is used not only to adjust the data rate, but also for the multiple access control. Even if the devices are allowed to use the higher data rates, it is useless if excessive collisions occur. Therefore, we have to investigate the impact of ADR on the collision probability and corresponding throughput.

As long as the used SF or channel are different, collision does not occur, consequently the LoRaWAN can be regarded as a set of $N_c \times N_s$ pure ALOHA-based sub-networks which independently operate each other, where N_c and N_s is the number of available channels and SFs respectively. Let N denote the total number of devices, and n_i denote the number of devices using $SF i$. We assume that

the devices select a channel with the same probability, and then the number of devices which attempt to transmit a packet using SF i and channel j can be approximated to

$$n_{i,j} \approx \frac{n_i}{N_c}. \quad (5)$$

ToA for sending a packet of length L can be calculated as below:

$$t_i = \frac{L}{R_b(i)}. \quad (6)$$

Under the assumption that the transmission probability of all devices is equal to P_{tx} , the average number of packets arriving on channel j within the time interval t_i is denoted by $\lambda_{i,j}$, and it is calculated by

$$\lambda_{i,j} = P_{tx} n_{i,j}. \quad (7)$$

Assuming that the distribution of packet arrival follows the Poisson process, the probability that k packets arrive within t_i can be represented by

$$P(k; \lambda_{i,j}) = \frac{(t_i \lambda_{i,j})^k}{k!} e^{-t_i \lambda_{i,j}}. \quad (8)$$

For the successful transmission, only one device attempts to send a packet when the channel is idle state, and the others must keep silent during the vulnerable period, i.e., $2t_i$. Therefore, in ALOHA-based communication, the success probability P_s can be written as

$$P_s(i, j) = e^{-2t_i \lambda_{i,j}} = e^{-2G(i,j)}. \quad (9)$$

where $G(i, j)$, which is calculated by $t_i \lambda_{i,j}$, is the traffic load on the sub-network defined by SF i and channel j . Based on (9), we can induce the throughput of a sub-network defined by the SF i as below:

$$S_i = \sum_{j=1}^{N_c} G(i, j) P_s(i, j). \quad (10)$$

Finally, the total throughput can be calculated roughly as the superposition of independent sub-networks, that is,

$$S = \sum_{i=1}^{N_s} S_i. \quad (11)$$

To give an insight into the relationship between the number of devices using the same SF and the throughput, we define the ratio of devices using SF i as α_i (calculated by $\frac{n_i}{N}$), and then the distribution of used SFs can be presented by a combination of α_i . Assuming that SF7, SF8 and SF9 are available, we examine three cases of $(\alpha_7, \alpha_8, \alpha_9)$; *Case1*: (0.33, 0.33, 0.33), *Case2*: (0.7, 0.2, 0.1) and *Case3*: (0.5, 0.3, 0.2) for $(\alpha_7, \alpha_8, \alpha_9)$. *Case1* considers that all SFs are used uniformly. In *Case2*, a majority device might use the highest data rate by using SF7. Meanwhile, in *Case3*, most traffic load is adequately distributed to the spreading factor domains. From the results shown in Figure 3, we can notice two important facts. First, it is not always effective for a majority device to simply use the highest data rate by applying SF7. The throughput presented in case 2 is less than the others until the number of devices is over 1000. This implies that although the devices are able to enhance their data rate using the small SF, this may need to be refrained from in some cases. Second, there is no sure way of providing the highest throughput regardless of the number of devices. *Case1* is the best policy when the number of devices is small, and *Case3* outperforms others when the number of devices is about 1000, and finally *Case2* is the best when the number of devices is large. Therefore, the data rate adaptation has to be dynamically applied depending on the given network state.

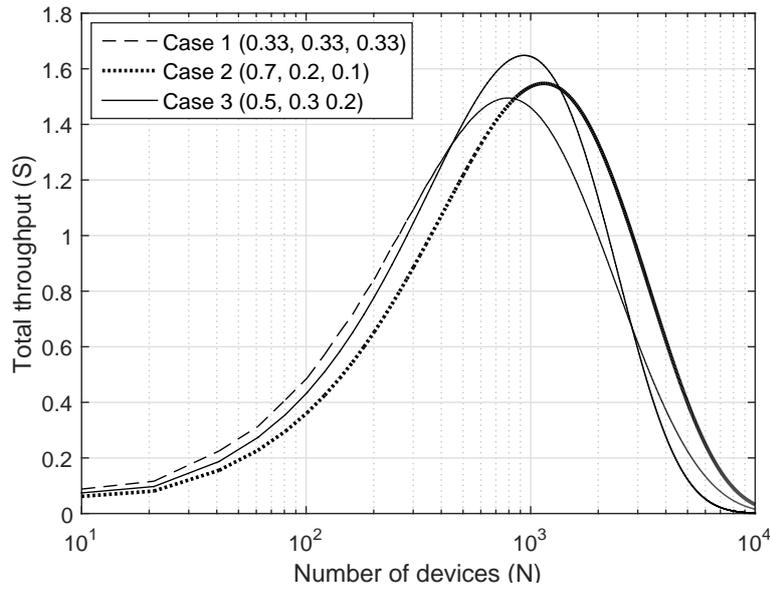


Figure 3. The comparison of the throughput according to the SF distribution.

4. Contention-Aware Adaptive Data Rate

The maximization of the throughput can be translated into an optimization problem. In particular, the smallest available SF of each device is restricted according to the communication distance, consequently this problem can be categorized into constrained optimization. In this section, we define the throughput as an objective function with respect to the number of devices using SF i , and present how to find the optimal solution maximizing the total throughput based on the gradient projection method.

4.1. Objective Function

Let $\mathbf{X} = \{n_7, n_8, \dots, n_{12}\}$ denote a set of the number of devices using SF i , the total throughput presented in (11) can be modified as a function $f(\mathbf{X})$ using (9), that is,

$$S = f(\mathbf{X}) = \sum_{j=1}^{N_c} t_j P_{tx} n_{7,j} e^{-2t_j P_{tx} n_{7,j}} + \dots + t_{12} P_{tx} n_{12,j} e^{-2t_{12} P_{tx} n_{12,j}} \quad (12)$$

$$\approx N_c (t_7 P_{tx} n_7 e^{-2t_7 P_{tx} n_7} + \dots + t_{12} P_{tx} n_{12} e^{-2t_{12} P_{tx} n_{12}}).$$

The $f(\mathbf{X})$ is the objective function, and finding a matrix \mathbf{X} that maximizes the function output is the goal of our study. This can be presented as below:

$$\underset{\mathbf{X} \in \chi}{\text{maximize}} f(\mathbf{X}) \quad (13)$$

where χ is all feasible sets of α_i .

The searching spaces are bounded by several constraints. Most of all, the sum of all n_i should be equal to the total number of devices, i.e., $\sum_{i=1}^{N_s} n_i = N$. In addition, n_i is bounded because some devices cannot use any kind of SF. For instance, a device near the gateway is allowed to SF7. However, a device deployed at the edge is forced to use just large SFs such as SF11 or SF12. To understand the constraints in the optimization, we suppose that a total of four devices are deployed in the LoRaWAN. Device 1 and 2 are located close to the gateway, device 3 is at the middle, and device 4 is at the edge, and the smallest available SF in each device is presented in Table 2. This table might be maintained in the gateway. By counting the devices with the available smallest SF i , we can obtain the upper bounds

as shown in Table 2. This means that the number of devices using the SF7 cannot be in excess of 2. In the same manner, the number of devices using the SF8 cannot exceed 3. Over those constraints, the optimal solution should be found. Let u_i denote the upper bound of n_i , and then all the constraints in the optimization can be represented by

$$g_1(\mathbf{X}) = n_7 + n_8 + n_9 + n_{10} + n_{11} + n_{12} = N$$

$$g_i(\mathbf{X}) = \sum_{j=7}^{i+5} n_j \leq u_i, \quad 2 \leq i \leq 6 \tag{14}$$

Here, g_1 is called equality constraint and the others (from g_2 to g_6) is called inequality constraint. As a result, the total throughput can be maximized by deriving \mathbf{X} that maximizes the objective function while satisfying the constraints expressed in (14).

Table 2. An example of how to set the upper bound of the number of devices which use a specific SF.

(a) Investigation of an available smallest SF in each device				
Device number	1	2	3	4
The available smallest SF	7	7	8	9
(b) The maximum number of devices permitted to use the SF				
SF	7	8		
The list of devices possible to use the SF	{ 1,2 }	{ 1,2,3 }	{ 1,2,3,4 }	
The upper bound(counting the components in each list)	2	3	4	

4.2. Gradient Projection Method

This paper adopts the gradient projection method to solve the constrained optimization problem [16]. This algorithm is improved from the gradient descent method that finds the optimal value by iteratively updating the current solution in a small gradient direction. Both algorithms are based on the fact that the optimal is found at the point where the gradient of the objective function is zero. In particular, in gradient projection method, the next solution is always examined to determine if it violates the given constraints. If it is determined that the constraint will be violated, the next solution is projected into the constraint space. Figure 4 geometrically illustrates the process finding an optimal in the gradient projection method. The solution \mathbf{X}_t found at the t th time step has to be updated in the direction in which the optimal marked as star exists. However, \mathbf{X}_{t+1} is out of the constraint spaces, i.e., constraint violation occurs. To find the optimal on the feasible space, \mathbf{X}_{t+1} is projected into the constraint space by using the projection matrix \mathbf{Pr} . In conclusion, \mathbf{X}_t is updated to the next projected solution, \mathbf{X}'_{t+1} .

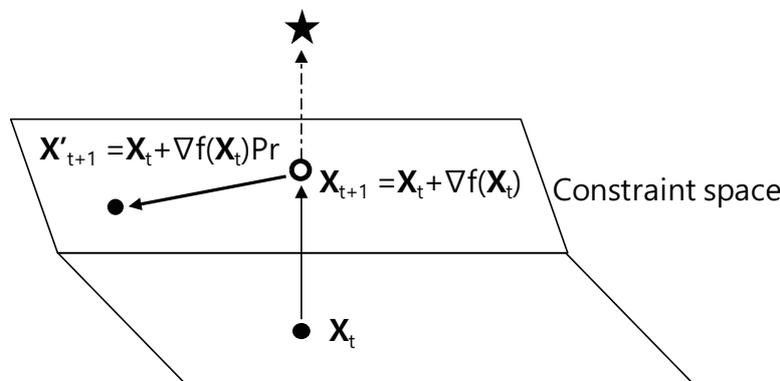


Figure 4. The comparison of the throughput according to the SF distribution.

To determine the direction of updating, partial derivatives of the objective function with respect to n_i should be derived as follows:

$$\frac{\partial f(\mathbf{X})}{\partial n_i} = N_c(t_i P_{tx} e^{-2t_i P_{tx} n_i} - 2t_i^2 P_{tx}^2 n_i e^{-2t_i P_{tx} n_i}). \quad (15)$$

The partial derivative can be expressed in form of a matrix, that is,

$$\mathbf{D}(\mathbf{X}) = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{n_7} \\ \dots \\ \frac{\partial f(\mathbf{X})}{n_{12}} \end{bmatrix}. \quad (16)$$

In the gradient descent method, the gradient derived from the partial derivative is directly applied to update the current solution. However, in the gradient projection method, it is validated that the updated solution still satisfies all constraints. For this purpose, the constraint equations expressed in (14) is modified to matrix form as follows:

$$\begin{bmatrix} g_1(\mathbf{X}) \\ g_2(\mathbf{X}) \\ g_3(\mathbf{X}) \\ g_4(\mathbf{X}) \\ g_5(\mathbf{X}) \\ g_6(\mathbf{X}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} n_7 \\ n_8 \\ n_9 \\ n_{10} \\ n_{11} \\ n_{12} \end{bmatrix} = \begin{bmatrix} N \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \\ u_{11} \end{bmatrix}. \quad (17)$$

For simplicity, it can be rewritten as

$$\mathbf{G}(\mathbf{X}) = \mathbf{A}\mathbf{X}^T - \mathbf{U}. \quad (18)$$

The constraint types can be divided into tight constraint and loose constraint, and they can be distinguished by observing $\mathbf{G}(\mathbf{X})$. If the k th component of $\mathbf{G}(\mathbf{X})$ is zero or a positive value, it is regarded as a tight constraint. It should be noted that the equality constraint is always treated as a tight constraint. In addition, except for the tight constraints, the remainder is considered a loose constraint. For the projection, the projection matrix is defined as

$$\mathbf{Pr} = \mathbf{I} - \mathbf{A}'_T [\mathbf{A}_T \mathbf{A}'_T]^{-1} \mathbf{A}_T \quad (19)$$

where \mathbf{I} is the identity matrix and \mathbf{A}_T is the tight constraint matrix containing all row elements of \mathbf{A} that are found to be tight constraint. If the tight constraints are founded, the corresponding gradients have to be projected into constraint space, namely,

$$\mathbf{D}'(\mathbf{X}) = \mathbf{D}(\mathbf{X})\mathbf{Pr}, \quad (20)$$

After projecting the gradients, the forwarding step size should be decided. It is similar to the step size used in the gradient descent method. The step size should also be bounded by the constraints, and it can be induced by calculating a matrix \mathbf{H} as follows:

$$\mathbf{H} = -\mathbf{G}(\mathbf{X}) ./ \mathbf{A}_L \quad (21)$$

where the notation $./$ means element-wise division and \mathbf{A}_L is a matrix containing the loose constraints. Between the smallest positive element in \mathbf{H} and 1, the smaller value is selected for the step size λ to be used in the update; it is expressed by

$$\lambda = \min(\mathbf{H}_{min}^+, 1). \quad (22)$$

It is noted that in the case of minimization problem, λ is determined by the smallest negative in \mathbf{H} . Finally, the current solution \mathbf{X}_t is updated as follows:

$$\mathbf{X}_{t+1} = \mathbf{X}_t + \lambda \mathbf{D}'(\mathbf{X}). \quad (23)$$

The updating rule shown in (23) is continuously repeated until it meets the termination condition. Typically, the algorithm stops when the norm of the gradients is less than the pre-defined threshold θ_d . The termination condition can be presented by

$$\|\mathbf{D}(\mathbf{X})\| \leq \theta_d. \quad (24)$$

When the current solution satisfies (24), the algorithm stops and the solution found at that time is regarded as the optimal solution. This solution indicates how many devices must use a particular SF for the optimal throughput. The gateway adjusts the data rate of each device based on the retrieved solution. The pseudo code of the algorithm is shown in Algorithm 1.

Algorithm 1 Gradient Projection Method to find an optimal distribution of SFs.

Require:

```

1:  $N$ : Total number of devices
2:  $\mathbf{U}$ : A matrix containing the upper bound of the number of devices can be used  $SFi$ 
3:  $\theta_d$ : Threshold for norm of gradients
4:  $\mathbf{A}$ : Constraint matrix
5:  $f(\mathbf{X})$ : Objective function
6:
7: procedure INITIALIZATION
8:    $\mathbf{D}(\mathbf{X}) \leftarrow f(\mathbf{X})'$ , ▷ get partial derivative
9:    $\mathbf{X}_0 \leftarrow$  initial value ▷ setting the starting point
10: end procedure
11: procedure FINDING OPTIMAL SOLUTION
12:    $t \leftarrow 0$ 
13:   while  $i \neq N_r$  do ▷ Stop if the number of iterations exceeds the threshold
14:     if  $\|\mathbf{D}(\mathbf{X})\| \leq \theta_d$  then break ▷ Stop if the norm of gradient is sufficiently small
15:     end if
16:      $\mathbf{A}_T \leftarrow$  tight constraints
17:      $\mathbf{A}_L \leftarrow$  loose constraints
18:     if  $\mathbf{A}_T \neq \text{empty}$  then
19:        $\mathbf{Pr} \leftarrow \mathbf{I} - \mathbf{A}_T' [\mathbf{A}_T \mathbf{A}_T']^{-1} \mathbf{A}_T$  ▷ Calculate projection matrix
20:
21:        $\mathbf{D}'(\mathbf{X}_i) \leftarrow \mathbf{Pr} \mathbf{D}(\mathbf{X}_i)$ 
22:
23:     end if
24:     if  $\mathbf{A}_L \neq \text{empty}$  then
25:        $\mathbf{H} \leftarrow -\mathbf{G}(\mathbf{X}) ./ \mathbf{A}_L$ 
26:
27:        $\lambda \leftarrow \min(\text{positive minimum element in } \mathbf{H}, 1)$ 
28:     end if
29:      $\mathbf{X}_{t+1} \leftarrow \mathbf{X}_t + \lambda \mathbf{D}'(\mathbf{X})$ 
30:   end while
31:   return final  $\mathbf{X}$  ▷ optimal solution
32: end procedure

```

Here, it is noteworthy that the duty-cycle should be obligated not only in the uplink but also in the downlink. In other words, it takes a considerable time to properly adjust the data rate of all devices, especially in the case where a huge number of devices are deployed. Therefore, how long the network will take to update the entire SF should be discussed. Let α_n denote the ratio of devices that must update their SF , and then the gateway has to transmit $\alpha_n N$ packets to deliver the changed SF information through the downlink. In general, the downlink is available in one out of two reception windows, just after the uplink transmission. Therefore, the required time for the SF update is related to not only the duty-cycle in the downlink, but also the arrival rate of the devices.

We assume that all devices have the same arrival rate, λ , accordingly the N packets arrive at the gateway within $1/\lambda$ sec on average. For the sake of simplicity, the collisions are ignored. If the transmission of a gateway is not limited by the duty-cycle, the update can be completed within $1/\lambda$ sec. However, some devices cannot receive the update information due to the duty-cycle, consequently an SF update for those devices is conducted for next $1/\lambda$ sec. Meanwhile, $\alpha_n N$ packets can be transmitted every $(1/d - 1)\alpha_n N T_{dl}$ sec from the gateway, where d is the duty-cycle of the gateway and T_{dl} is the ToA of the downlink packet. As a result, we can roughly retrieve a lower bound of the time required for SF update using the following equation:

$$m(1/d - 1)\alpha_n N T_{dl} = k/\lambda, \quad (25)$$

where m and k are arbitrary variables ($m, k > 0$). By finding the minimum value of m and k satisfying the above equation, the minimum time for SF update can be calculated. We roughly explain the concept here, but it is sufficient to give an insight into the time required to complete the proposed method. The more detailed analysis and efficient solution for the SF update via downlink should be investigated in future works.

5. Numerical Results

5.1. Evaluation Environment

This section presents the numerical results of the proposed method. We calculated the total throughput with respect to the total number of devices ranging from 0 to 10,000. We assumed that the gateway knows the smallest available SF of each device by measuring the SNRs. Originally, ADR in the LoRaWAN was designed for static devices, so SNR variation is assumed to be small. Meanwhile, the number of devices using the large SFs should be relatively small in the total throughput optimization. Therefore, we supposed that just three kinds of SFs , i.e., $SF7$, $SF8$, and $SF9$ are used, for a clear explanation. It is reasonable because the devices far away from a gateway are likely to communicate with other nearby gateways. The channel occupation time per device is limited to the duty-cycle because LoRa works at the unlicensed band [17]. In EU863-870, the duty-cycle is 1%, accordingly we roughly decide the transmission probability of all the devices to be 0.01.

We denote α_i^+ as the ratio of devices with the smallest available SF i . The motivation of our study is to optimize the throughput in environments where a large number of devices can use the same SF , thus the biased deployment must be considered. We set $(\alpha_7^+, \alpha_8^+, \alpha_9^+)$ to $(0.7, 0.2, 0.1)$. This can be understood that maximally 70% of the devices are allowed to use $SF7$, and 90% of the devices are allowed to use $SF8$ and so on. In the LoRaWAN, the maximum packet length can also be adjusted according to the used SF , but we assumed that all the devices transmit the same length of packet. For the comparison, we considered another two methods, uniform approach and naive approach. Uniform approach means that SFs are evenly assigned to the devices, i.e., setting $(\alpha_7, \alpha_8, \alpha_9)$ to $(0.33, 0.33, 0.33)$. Intuitively, this scheme adequately distributes the traffic load to all SFs . Meanwhile, the naive approach describes a method in which the devices directly use the smallest available SFs , i.e., $(0.7, 0.2, 0.1)$. This method is intended to achieve high throughput by allowing all the devices to use the highest available data rate. It is also very similar to the behavior described in the current

standard. In contrast to the comparisons using the fixed ratio, the proposed method dynamically derives $(\alpha_7, \alpha_8, \alpha_9)$ according to the given network environments to maximize the throughput.

The optimization algorithm should be appropriately tuned according to the characteristic of the problem. Here, we focus on two design issues, scaling and local maxima problem. The gradient descent-based algorithms iteratively find the optimal solution by observing the current gradients until the observed gradients are almost zero. In this situation, scaling is very important. In detail, the value indicating the throughput is less than 10, but the number of devices ranges from 0 to 10,000. Therefore, the gradients are almost zero, regardless of the solution, accordingly the iterative process fails to find an optimal solution due to the early stopping. To deal with this problem, we adopt learning rate η , and the throughput multiplied by η is used instead of the original throughput shown in (12). In addition, we employed η as one tenth of the number of devices. This is due to the fact that the larger the number of devices, the longer it takes to find the solution due to the enlarged searching space. Therefore, the learning rate should increase in order to derive the optimal within the allowed time. Meanwhile, gradient descent-based algorithms inevitably face the local maxima problem. To find a global maximum, we simply use multiple initial points, and select the best solution among the multiple sub-optimal solutions independently retrieved from the initial points. In the evaluation, we use eight different initial points. For the summary, The used parameters are presented in Table 3.

Table 3. Parameter setting.

Parameter	Value
Number of devices	0 to 10,000
Device type	Static
Available SF	SF7, SF8, SF9
Packet size	50, 100 byte
Number of channels	3, 6
Transmission probability	0.01
Number of initial points in the gradient projection method	8
Gradient threshold(θ_d)	10^{-4}
Maximum iteration number	10^6

5.2. Performance Evaluation

Figure 5 shows the total throughput according to the number of devices, assuming that the number of channels is 3. As shown in the figure, until the throughput reaches the peaks, they gradually increase. In addition, when the number of devices is small, the uniform approach is better than the naive approach owing to the less contention. However, as the number of devices increases, the naive approach is better than the uniform approach. This is because the smaller the SF used, the lower the rate of increase in the number of devices to the throughput reduction rate. Therefore, in the case that enormous devices are deployed at the network, it is recommended that a majority of devices uses as small of SFs as possible. Meanwhile, the proposed method provides higher performance than the others in all cases. Especially when the number of nodes is between 4000 and 6000, which leads to saturation throughput in the proposed method; the throughput is close to the theoretical upper bound. This result implies that the contention problem caused by the adjusted data rate must be appropriately handled. In addition, the ADR policy should be dynamically changed according to the network state.

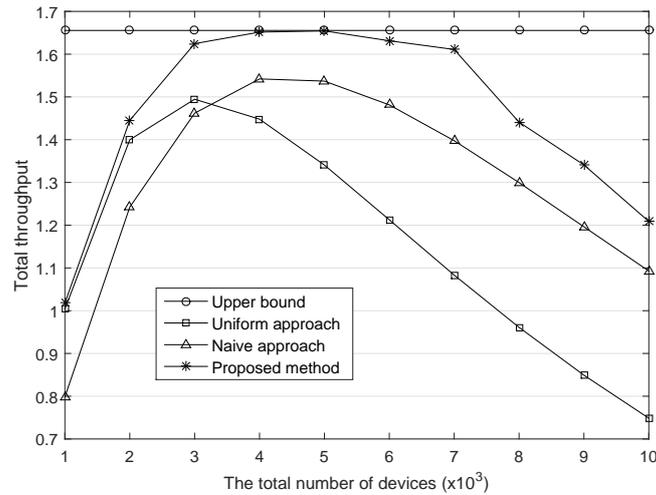


Figure 5. The comparison of the total throughput. The number of channel, N_c is 3.

As the number of available channels increases, the networks can accommodate more devices owing to the less contention resulting from the channel diversity. Figure 6 shows the throughput in the same environment as the previous evaluation except for the number of channels; it changed from 3 to 6. Since the network capacity is proportional to the number of channels, the upper bound of throughput becomes double as compared with the previous one. It is noteworthy that the performance differences between the proposed method and the comparisons are more obvious when the number of devices is large. This means that the proposed method enhances the scalability in the LoRaWAN that considers a city scale area as a service coverage [18].

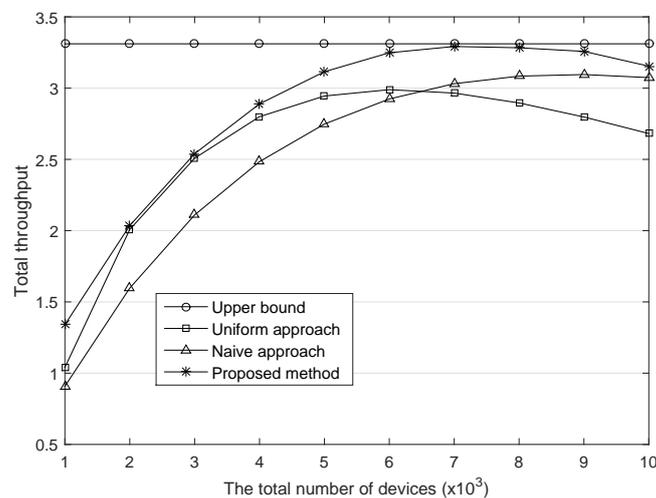


Figure 6. The comparison of the total throughput. The number of channel, N_c is 6.

In the LoRaWAN, the maximum MAC payload length is allowed from 59 to 230 bytes according to the used SF , and the smaller the factor is, the greater the payload length is allowed. Basically, the LoRaWAN technology is suitable for the exchange of small size data, but it may increase according to applications [19]. The ToA is proportional to the packet length, thus the traffic load on the network increases to exchange the large size packet, even with the same transmission probability. We changed the packet length from 50 bytes to 100 bytes and calculated the throughput as shown in Figure 7. Due to the growth of traffic load, the number of devices inducing the peak throughput is

diminished. This reveals that the packet length can be a consideration in the throughput optimization. One suggestion is that it is a good strategy that reduces the packet length leading to less contention. In addition, based on the collaborative operation with those strategies, our method is able to flexibly adjust the data rate for the optimal throughput in the given network state.

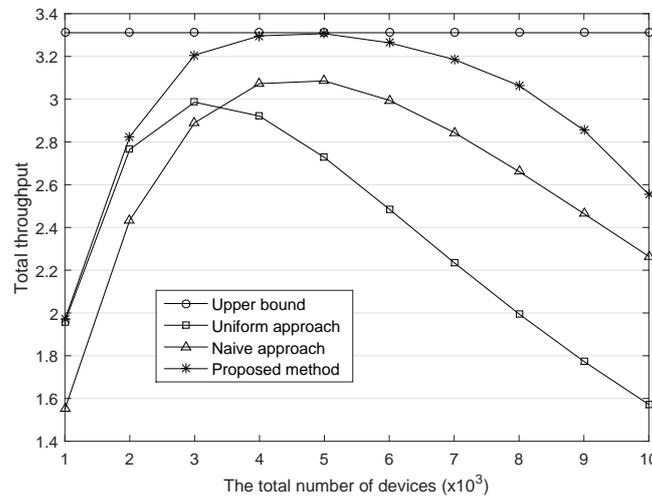


Figure 7. The comparison of the total throughput. The packet length is 100 byte.

This paper assumes that only one gateway is dedicated to all devices. However, in terms of an empirical system, multiple gateways exist in the LoRaWAN and each device communicates with one of them. Therefore, the devices are more likely to join to the nearby gateway, and this may cause the more biased distribution of SF usage. To consider this environment, we set $(\alpha_7^+, \alpha_8^+, \alpha_9^+)$ to $(0.8, 0.1, 0.1)$ and evaluated our method. As shown in Figure 8, the throughput gaps between the proposed method and the others are greater as compared with the previous results. The average difference of the throughput between the proposed method and the naive approach is 0.37 in the previous environment, whereas it is 0.62 in this environment. Moreover, most of the presented throughput is close to the upper bound. This result indicates that in the case where multiple gateways are evenly deployed in the network, the advantage of the proposed method will be further emphasized.

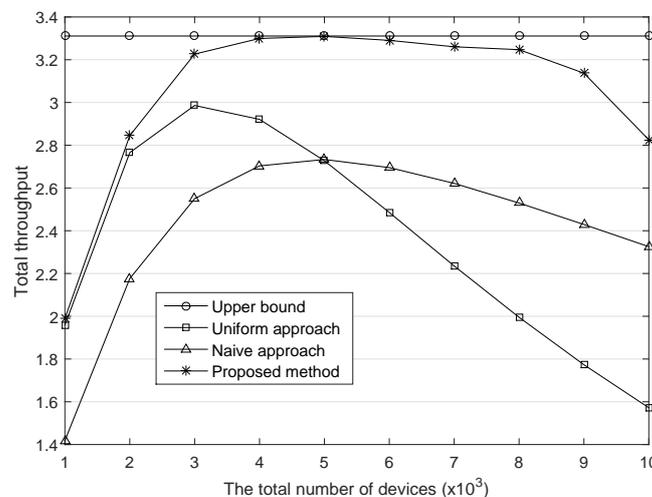


Figure 8. The comparison of the total throughput. The upper bounds of ratio of smallest SF {0.8, 0.1, 0.1}.

6. Conclusions

The ADR provides a chance to optimize the total throughput in the LoRaWAN. Nevertheless, if most devices use the same data rate without consideration of the contention problem, the throughput may be reduced. This paper proposes contention-aware ADR to get an optimal throughput, and we find the optimal set via the gradient projection method. In particular, when a large number of devices have similar link quality, namely in the case of biased usage of the spreading factors, the proposed method can achieve considerably higher throughput than the current system owing to the load balancing effect.

The contributions provided in this study are the first step for the optimization of the LoRaWAN. In this paper, we focus on the numerical result achieved by the ADR. Therefore, practical issues should be considered in the next step, such as duty-cycle of downlink and consideration of all kinds of SF in the evaluation. Meanwhile, the main goal of this paper is throughput optimization. Therefore, the data rate is adjusted in the direction of increasing the number of devices using small spreading factors. Although this policy increases overall throughput, the transmission success ratio of the devices decreases. Accordingly, for applications in which reliability is more important than the throughput, our approach has to be supplemented. For future works, we will develop an optimization technique that takes into account the success ratio at the same time. In other words, our solution will be extended to the multi-objective optimization problem.

Author Contributions: S.K. investigated the theoretical basis for this work, implemented the simulator for the experiments as well as numerical analysis, and wrote the manuscript. Y.Y. supervised the entire work and helped to write the manuscript, and revised the manuscript.

Funding: This work (2016R1A2B4016588) was supported by Mid-career Researcher Program through NRF grant funded by the MEST.

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

References

1. Da Xu, L.; He, W.; Seto, M.; Li, S. Internet of Things in Industries: A Survey. *IEEE Trans. Ind. Inform.* **2014**, *10*, 2233–2243.
2. Xiong, X.; Zheng, K.; Xu, R.; Xiang, W.; Chatzimisios, P. Low Power Wide Area Machine-to-Machine Networks: Key Techniques and Prototype. *IEEE Commun. Mag.* **2015**, *53*, 64–71. [[CrossRef](#)]
3. Bardyn, J.P.; Melly, T.; Seller, O.; Sornin, N. IoT: The era of LPWAN is starting now. In Proceedings of the 2016 European Solid-State Circuits Conference (ESSCIRC), Lausanne, Switzerland, 17–19 October 2016.
4. Neumann, P.; Montavont, J.; Noël, T. Indoor Deployment of Low-Power Wide Area Networks (LPWAN): A LoRaWAN Case Study. In Proceedings of the IEEE Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), New York, NY, USA, 17–19 October 2016.
5. Raza, U.; Kulkarni, P.; Sooriyabandara, M. Low power wide area networks: An overview. *IEEE Commun. Surv. Tutor.* **2017**, *10*, 855–873. [[CrossRef](#)]
6. Sornin, N.; Luis, M.; Eirich, T.; Kramp, T. LoRa Specification 1.0. Lora Alliance Standard Specification. Available online: https://lora-alliance.org/sites/default/files/2018-04/lorawantm_specification_v1.1.pdf (access on 11 October 2017)
7. Parada, R.; Cárdenes-Tacoronte, D.; Monzo, C.; Melià-Seguí, J. Internet of THings Area Coverage Analyzer (ITHACA) for Complex Topographical Scenarios. *Symmetry* **2017**, *9*, 237, doi:10.3390/sym9100237. [[CrossRef](#)]
8. Chen, J.; Hu, K.; Wang, Q.; Sun, Y.; Shi, Z.; He, S. Narrowband internet of things: Implementations and applications. *IEEE Internet Things J.* **2017**, *4*, 2309–2314. [[CrossRef](#)]
9. Goursaud, C.; Gorce, J. M. Dedicated Networks for IoT: PHY/MAC State of the Art and Challenges. *EAI Endors. Trans. Internet Things* **2014**, *15*, 2233–2243.
10. Adelantado, F.; Vilajosana, X.; Tuset-Peiro, P.; Martinez, B.; Melia-Segui, J.; Watteyne, T. Understanding the Limits of LoRaWAN. *IEEE Commun. Mag.* **2017**, *55*, 34–40. [[CrossRef](#)]
11. Berni, A.J.; Gregg, W. On the Utility of Chirp Modulation for Digital Signaling. *IEEE Trans. Commun.* **1973**, *21*, 748–751. [[CrossRef](#)]

12. Reynders, B; Pollin, S. Chirp Spread Spectrum as a Modulation Technique for Long Range Communication. In Proceedings of the Symposium on Communications and Vehicular Technologies (SCVT), Mons, Belgium, 22 November 2016.
13. Morrow, R.K.; Lehnert, J.S. Packet Throughput in Slotted ALOHA DS/SSMA Radio Systems with Random Signature Sequences. *IEEE Trans. Commun.* **1992**, *10*, 1223–1230. [[CrossRef](#)]
14. Kumar, S.; Raghavan, V.S.; Deng, J. Medium Access Control Protocols for Ad Hoc Wireless Networks: A Survey. *Ad Hoc Netw.* **2006**, *4*, 326–358. [[CrossRef](#)]
15. Semtech Corporation. LoRa Modulation Basics. Tech. Rep. AN1200.22. Available online: <http://www.semtech.com/images/datasheet/an1200.22.pdf> (accessed on 2 May 2015).
16. Rosen, J.B. The Gradient Projection Method for Nonlinear Programming. Part I. Linear Constraints. *J. Soc. Appl. Math.* **1960**, *8*, 181–217.
17. Federal Communications Commission. *FCC Part 15—Radio Frequency Devices, Code of Federal Regulation 47 CFR Ch. 1 (10-1-15 Edition)*; Federal Communications Commission: Washington, DC, USA, 2010.
18. Centenaro, M.; Vangelista, L.; Zanella, A.; Zorzi, M. Long-Range Communications in Unlicensed Bands: The Rising Stars in the IoT and Smart City Scenarios. *J. Soc. Appl. Math.* **2016**, *23*, 60–67. [[CrossRef](#)]
19. Zanella, A.; Bui, N.; Castellani, A.; Vangelista, L.; Zorzi, M. Internet of Things for Smart Cities. *IEEE J. Internet Things* **2014**, *53*, 22–32.



© 2018 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/>).