Structural and functional analysis of the LoRa protocol for telemetry information exchange in wireless sensor networks

Zhelyazkov Y. and Torlakov I.

***Abstract*—The rapid deployment of the Internet of Things (IoT) in various areas of household and industry is based on the construction of low-power wide-area wireless networks, taking into account the general requirements for them as automated control systems. Low-power networks allow to provide low-cost energy-efficient wireless communication for telemetry and control. One of the most important technologies in constructing such networks is the LoRa protocol, which consists of a physical layer protocol and a set of open base layer protocols. In this paper, a structural and functional analysis of approaches for implementing wireless sensor networks based on LoRa is performed, taking into account the specifics and logic of functioning at different layers in the exchange of telemetry information.**

***Keywords*—Internet of Things, LoRa, wireless communication, sensor networks**

I. Introduction

A wireless sensor network is a set of different wireless nodes that are sensor nodes in nature. They collect information from the external environment, then process it and communicate with other neighboring devices in the network. WSNs (wireless sensor networks) consist of multiple applications using different types of sensors that place different demands in terms of QoS and reliability. Depending on the application-specific traffic distribution methods, they can be grouped into four categories, namely: continuous sensing applications; query-driven applications; event-driven applications; and hybrid applications. The widespread application of WSNs today has affected all areas of human professions. Despite its advantages, WSNs have several challenges such as resource scarcity, memory limitations, and computational capabilities. One of the most important technologies used in wireless sensor networks is the energy efficient LoRa network technology, which helps to manage the data collection from user devices and their analysis for management decision making [1,2,3,6].

II. LoRa Protocol Features

LoRa technology combines the LoRa modulation method developed and patented by Semtech Corporation [1] and the open LoRaWAN protocol [2].

The term LoRa itself is an abbreviation for Long Range, which stands for a radio modulation format that provides a longer link distance compared to direct Frequency Shift Keying (FSK) modulation. The longer range is achieved by a combination of the following solutions: using a spectral broadening method called Chirp Spread Spectrum (CSS), or its analogue Linear Frequency Modulation (LFM), and applying noise-resistant coding combined with masking and interleaving. A useful effect of using LoRa modulation is the high sensitivity at which the receiver can receive signals below the noise level. The transmitter can therefore transmit at lower power. The LoRa protocol is a reliable, energy-efficient and low-speed system that is one of the main ones for use in the Internet of Things (IoT) [3,4,5].

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not number text heads-the template will do that for you.

For LoRa modulation the key parameters are:

- Bandwidth B, which is the difference between the minimum and maximum frequency of the signal;
- The expansion factor SF. This is a parameter for the number of bits encoded in one symbol of the LoRa signal;
- Coding rate CR, which sets the parameters of the noise code;
- Preamble length and value of the special timing signal. The number of preamble characters varies depending on the operating modes of the terminal devices and the sensitivity of the base station receiver in LoRaWAN networks. The sync signal value can indicate the type of network (public or private) as well as the specific value for the FSK channel;
- Presence or absence of a separate allocated message header. If an explicit header is used, it determines the size of the data field, the value of the CR parameter, and the presence or absence of a CRC (Cyclic redundancy check) checksum. The implicit header does not transmit these data fields;
- Presence or absence of rate control mode up to SF Low Data Rate Optimisation;
- Presence or absence of a 16-bit CRC checksum for the data field.

A. Modulation

When LoRa modulation is used, information is transmitted in symbols whose duration is $Ts = M/B$, and whose duration is $2^{SF}/B$ seconds. Within a symbol, the signal frequency ranges from $f_0 - B/2$ to $f_0 + B/2$, where f_0 is the center frequency of the channel. These are the so-called Chirp symbols. Chirp symbols have a constant rate of change of modulation frequency for each symbol value. When the frequency at the symbol duration reaches the end of the range, it “returns” to the other side of the frequency band [1]. The Chirp symbol signal is determined by the following formula:

$$s(t) = exp\left(j2\pi \int_0^t \left[(\beta x + \gamma_n) mod_B - \frac{B}{2}\right] dx\right) \quad (1)$$

where γ_n is the frequency offset obtained as follows

$$\gamma_n = m_n \Delta_f = \frac{m_n}{T_S} \quad (2)$$

where m_n is the value of the data symbol $m_n \in \{0,1,\dots,M-1\}$, n is the symbol index, and Δf is the step frequency between shifts. The step frequency in the case of LoRa is chosen to be equal to the transmission rate B/M. Furthermore, β represents the slope of the Chirp symbol, which is a function of the bandwidth and symbol time, as follows:

$$\beta = \frac{f_{high} - f_{low}}{T_S} = \frac{BW}{T_S} \quad (3)$$

where f_{low} and f_{high} are the lower and upper cut-off frequencies of the Chirp symbol, respectively. The dependence of the signal frequency change on time (spectrogram) is illustrated in Figure 1, which shows that the symbol energy is distributed across the entire band. The figure also shows that as the value of *SF* increases, the energy of the symbol increases. This is due to the slower data transmission rate and higher energy consumption.

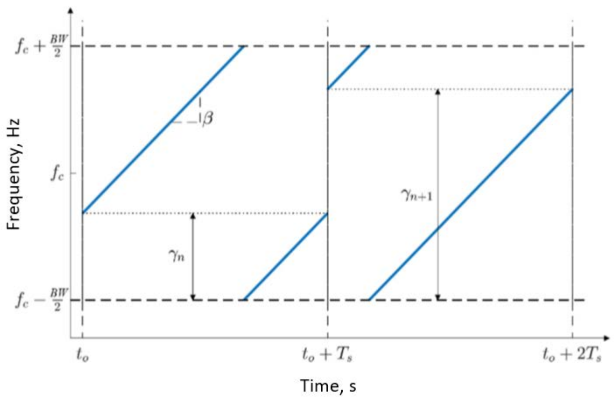

Fig. 1. Visualization of the spectrogram of two LoRa Chirp symbols

The description of LoRa modulation shows that two types of errors are possible during demodulation:

1. First type demodulation errors - the Chirp value at the demodulator output differing from the transmitted value by +/- 1 position.

2. Type 2 demodulation errors - the Chirp value at the output of the demodulator differs from the transmitted value by more than 1 position.

The likelihood of errors occurring depends on the noise level and at low noise levels errors of the first type are most likely.

In order to correct the first type of demodulation errors, the Hemming code sequences in combination with the Gray code is used to encode the information. Gray's code provides a difference of one bit between adjacent characters and Hemming's code corrects these single errors. To correct errors of the second type, so-called interleaving is used, which ensures the transmission of a code word in different Chirp symbols [1,3].

B. Noise-resistant coding

The characteristics of the noise-resistant code used in the LoRa protocol are described in Table 1. The following symbols are used in the table: n is the number of bits in the code word; k is the number of information bits; r is the number of control bits; d is the minimum code distance.

Parameters of the noise-resistant code of the LoRa protocol

TABLE I
THE CHARACTERISTICS OF THE NOISE-RESISTANT CODING

CR	Parameters				Code description
	n	k	r	d	
4	8	4	4	4	Haming code with parity check added
3	7	4	3	3	Haming code
2	6	4	2	2	Code derived from the Hamming code by removing one control bit
1	5	5	1	2	Parity check code

The encoding procedure is performed in two stages. First, multiplication by the generator matrix is performed. In the second stage, the information and parity bits are reordered as follows for a speed of 4/8 generating matrix. When other speeds are used, the check bits are excluded. The last columns are deleted from the matrix. And for code speed with parity checking (CR = 1), a matrix with ones in the last column is actually used. In fact, two-stage encoding can be combined into one using another generating matrix that contains permutations at the output of the multiplier[2,5].

C. Masking

The masking sequence is formed in such a way that a pseudo-random sequence of symbols is created at the output of the block. For this procedure, the so-called whitening is used. The whitening method should increase the number of zeros. In the LoRa protocol, whitening is performed by adding the pseudo-random sequence to the data blocks using modulo 2 (XOR). It is more logical to call this method masking. Masking is not applied to the header.

The masking block works as follows. First, the input sequence is divided into blocks. An individual masking sequence is added to each block, which is represented as a matrix for the entire message. The masking sequence is obtained by unfolding this matrix into a row and performing an XOR operation with the encoder output. The number of rows in the matrix cannot be precisely determined from existing research, but the masking principle tells us that this sequence must be larger than the size of the information data.

D. Interleaving

Code word interleaving follows the masking block. During transmission, a symbol may be distorted by noise or fading in the channel, leading to multiple bit errors after demodulation. Since all errors are caused by distortion of the entire Chirp symbol, they are highly correlated. The problem with a channel with such errors is that they are not well handled by error correction codes designed to correct random errors in the code word.

The goal is to break this correlation between error bits by distributing errors across several code words. This is possible by distributing the bits that make up the code word among several Chirp symbols. This operation increases the efficiency of the error-correcting code by increasing the waiting time caused by the need to receive several symbols before it is possible to recover one complete code word.

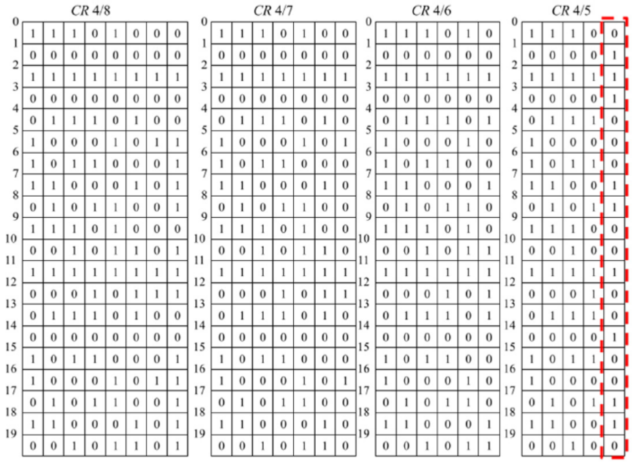

Fig. 2. Masking matrices for different coding rates

The error-correcting coding method used by the LoRa protocol is quite simple. It allows for the correction of a single error in a data block, and errors in a given channel will be grouped within a single Chirp symbol. This process

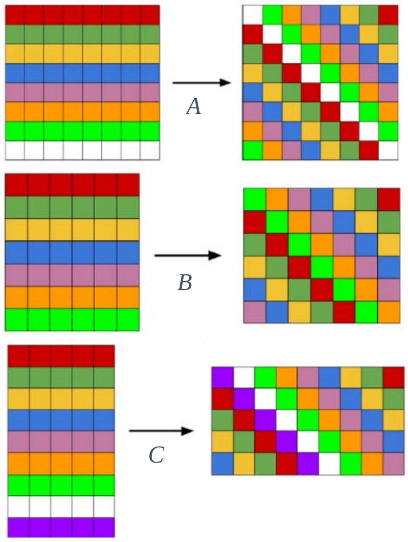

Fig. 3. Interleaving scheme for SF=8, CR=4/8 (A); SF=7, CR=4/6 (B) and SF=9, CR=4/5 (C)

interleaving is performed according to a specific rule described in the Semtech patent. The main idea is that after error-correcting coding, the control and information bits must be evenly distributed across the nearest Chirp symbols.

The result of interleaving will be the distribution of characters of a single code word across different Chirp characters according to the following rule: there should be no Chirp characters that store only the information part or only the control part of the code.

Figure (3) shows the code words before and after interleaving.

E. Grey Code

Grey coding is the mapping of a symbol in any numerical representation to a binary sequence. The distinctive feature of the resulting binary sequence is that neighboring symbols in the original representation differ by only one bit in the Grey code representation. This feature of Grey codes makes them very useful in many wireless communication modulations, where it is more likely that a symbol will be misinterpreted as a neighboring symbol than as another symbol. Grey codes are also used in many digital communication systems, such as digital television and digital radio [1,5].

Traditionally, the forward Gray code is used on the transmitting side, and the reverse Gray code is used on the receiving side. As noted above, in this work, Gray code combined with error-correcting coding allows for the correction of type 1 demodulation errors (the value of the transmitted and received FM symbol differs by 1 unit). In the LoRa system, the reverse code is used on the transmitting side and the forward code on the receiving side.

F. Communication

A typical LoRaWAN network consists of a network server, gateways, end devices (sensors), application servers

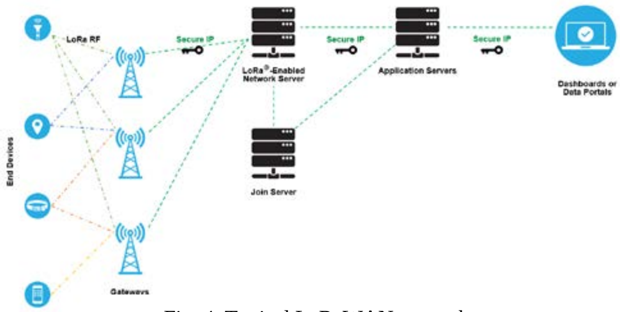

Fig. 4. Typical LoRaWAN network implementation

and join servers. Gateways have a fast and reliable connection to the server (via Ethernet or cellular network), and gateways connect to sensors via a wireless channel using LoRa technology. The server is the network coordinator, and the gateways, act as repeaters between the sensors and the server - after receiving a frame via a wireless connection, the gateway encapsulates the frame in an IP packet and transmits it to the server and, similarly, transmits packets received from the server to the sensors. Application servers are responsible for securely handling, managing and interpreting sensor application data. They also generate all the application-layer downlink payloads to the connected end devices. The join server contains the information required to process uplink join-request frames and generate the downlink join-accept frames. It signals to the network server which application server should be connected to the end-device, and performs the network and application session encryption key derivations. Fig. 4 shows a typical LoRaWAN network implementation from end to end.

The system uses Forward Error Correction (FEC) and operates in the sub-gigahertz frequency range: 169, 433 and 915 MHz in the United States, and 868 MHz in Europe. The most commonly used operating frequencies are 868 and 915 MHz.. LoRa uses a cyclically unique transmission option, which limits the speed of message creation. However, by supporting multiple channels, LoRa allows end nodes to participate in data exchange procedures by changing the carrier frequency while observing the duty cycle limit in each channel. The choice of data transmission speed is a compromise between coverage area and data volume; messages with different data transmission speeds do not interfere with each other. LoRa data rates range from 0.3 to 50 kbps. To maximize the battery life of end devices and the overall network throughput, the LoRa network infrastructure can manage the data rate for each device individually through adaptive data rates [1,6].

LoRaWAN has three different classes of end devices to meet different needs.

1. Class A - end devices with the lowest power consumption. Communication is always initiated by end devices and is completely asynchronous. Each data transmission can be performed at any time and is followed by two short windows that allow network control commands to be received if necessary.

2. Class B – end devices with deterministic downlink delay. In addition to the initiated data reception windows of class A, the devices synchronize with the network at scheduled times. The delay is programmable up to 128 seconds for different applications, and the additional energy consumption is quite low.

3. Class C – end devices with the lowest latency. The device constantly 'listens' to the network at all times until it sends information. The network server can initiate data transfer at any time, there is no delay. However, this class is suitable for devices with constant power supply.

Asynchronous data transmission means that data is transmitted only after the device establishes a connection (in the case of Class A devices) or at any time (for Class C devices). Scheduled communication saves battery power, as the transceiver is in sleep mode most of the time. It is possible to temporarily switch between Class A and Class C modes for periodic tasks, such as updating the built-in software. Base stations listen to the network, receiving signals from the interface module and transmitting them to the network server. The network server authenticates devices and frames on the network and manages message flows, transmitting data to application servers.

The spectrum expansion factor SF determines the ratio between the data transfer rate and the bandwidth. Messages transmitted at the maximum value of the factor are less sensitive to negative effects such as attenuation, but their transfer takes longer. At low SF values, the transmission speed increases, but the signal is subject to interference and the width of the spectrum used decreases. Orthogonal symbol coding allows signals with different spectrum expansion factor values to be transmitted simultaneously in a single channel. LoRaWAN specifies six SF values (from 7 to 12).

In the Gmsk channel, the LoRa modem can generate interference of up to 19.5 dB thanks to Gaussian filtering. In other words, such a modem has the ability to receive and demodulate signals whose level is 19.5 dB below the interference (or noise) level, despite the fact that for most systems with frequency shift keying (FSK) modulation, a signal power of at least 8-10 dB above the level is required for correct demodulation.

LoRaWAN end devices can change the adaptive data rate (ADR), SF factor, and bandwidth between messages and can select the appropriate data transfer rate without intervention from the system operator [1].

Adaptive data rate can be described as a method in which the actual data rate is adjusted to ensure reliable packet delivery, maintain optimal performance lines, and provide the necessary range for network loads. For example, nodes located near the LoRa gateway will operate at a high data rate. Therefore, such nodes will have less active transmission time over the radio channel and lower output power. Low transmission rates and high transmitter output power are only possible for the most remote network devices. ADR algorithms can introduce the necessary infrastructure networks, thereby compensating for various losses along the signaling path.

III. Conclusion

The article presents a systematic overview of the specific features of the implementation of the LoRa protocol and analyzing the principles of operation of the physical layer.

LoRa is one of the most widely used low-power technologies for wireless sensor networks for Internet of Things (IoT) implementation. It offers communication over longer distances compared to other technologies such as Zigbee and Bluetooth, but at a lower data rate.

Due to their low power consumption, LoRa modules are suitable for applications such as battery-powered sensor nodes or photovoltaic panels, where the amount of data to be transmitted is not large and low power consumption is prioritized over high transmission speed. Available protocols such as LoRaWAN ensure the compatibility of LoRa networks with the global Internet.

It will allow LoRa device developers to offer their solutions for improving data processing efficiency on the receiving end without changing the operation of the physical layer on the transmitting end.

References

[1] LoRaWAN 1.1 Specification, Oct. 2017. http://lora-alliance.org/lorawan-for-developers.

[2] F. Adelantado, X. Vilajosana, P. Tuset-Peiró, B. Martinez, J. Melià-Seguí, and T. Watteyne, "Understanding the Limits of LoRaWAN," IEEE Communications Magazine, vol. 55, no. 9, pp. 34–40, Sept. 2017

[3] J. P. S. Sundaram, W. Du, and Z. Zhao, "A Survey on LoRa Networking: Research Problems, Current Solutions and Open Issues," IEEE Communications Surveys & Tutorials, vol. 22, no. 1, pp. 371–388, 2020.

[4] D. Magrin, M. Capuzzo, and A. Zanella, "A Thorough Study of LoRaWAN Performance Under Different Parameter Settings," arXiv:1906.05083, 2019.

[5] F. H. Khan and M. Portmann, "Experimental Evaluation of LoRaWAN in NS-3," arXiv:2009.09503, 2020.

[6] R. Sánchez-Iborra, J. Sánchez-Gómez, J. Ballesta-Viñas, M.-D. Cano, and A. F. Skarmeta, "Performance Evaluation of LoRa Considering Scenario Conditions," Sensors, vol. 18, no. 3, Art. 772, Mar. 2018.