Low Power Wide Area Networks for the Internet of Things

Framework, Performance Evaluation, and Challenges of LoRaWAN and NB-IoT

Samer Lahoud   Melhem El Helou

ESIB, Saint Joseph University of Beirut, Lebanon

ICT 2018, Saint-Malo, France

Tutorial Outcomes

- How do LPWAN complement traditional cellular and short-range wireless technologies?
- What are the fundamental mechanisms that enable to meet the LPWAN requirements?
- What are the major design choices made in the LoRaWAN and NB-IoT specifications?
- How do we evaluate the performance of a LoRaWAN and NB-IoT deployment in terms of coverage and capacity?
- What are the recent research directions for radio resource management in LoRaWAN and NB-IoT?

Feedback and Material

- Feedback form
- Presentation slides are available

Outline

Revisiting LPWAN Requirements

- Low device complexity and cost
- Reliability under extreme coverage conditions
- Low power consumption: long battery lifetime
- High capacity: support for massive number of low-rate devices
- Simplified network topology and deployment

Objectives and Approaches

- Develop a *clean-slate* technology that meets the LPWAN requirements
  $\Rightarrow$ LoRaWAN
- Adapt and leverage existing 4G technology to meet the LPWAN requirements
  $\Rightarrow$ NB-IoT

Low Device Complexity and Cost

Device Complexity and Cost

- Devices are mainly composed of:
  - a processing unit: usually a microcontroller with a limited amount of memory
  - a sensing unit: sensors and analog to digital converters
  - a radio unit: usually a transceiver capable of bidirectional communications
- The radio unit complexity and cost are primarily related to the complexity of:
  - digital baseband processing
  - radio-frequency (RF) analog processing

Digital Baseband Processing

- Reduce baseband processing complexity through:
  - limiting message size:
    - LoRaWAN: maximum application payload size between 51 and 222 bytes, depending on the spreading factor
    - NB-IoT: Downlink (DL) Transport Block Size (TBS) = 680 bits (R13), or 2536 bits (R14); Uplink (UL) TBS = 1000 bits (R13), or 2536 bits (R14)
  - using simple channel codes:
    - LoRaWAN: Hamming code
    - NB-IoT: LTE tail-biting convolution code (TBCC) in the DL; LTE turbo code, or repetition code in the UL
  - not using complex modulations or multiple-input multiple-output (MIMO) transmissions
    - LoRaWAN: LoRa
    - NB-IoT: QPSK in the DL; QPSK in the UL multi-tone; $\pi/4$-QPSK, or $\pi/2$-BPSK in the UL single-tone
  - supporting only half-duplex operation: no simultaneous transmission and reception

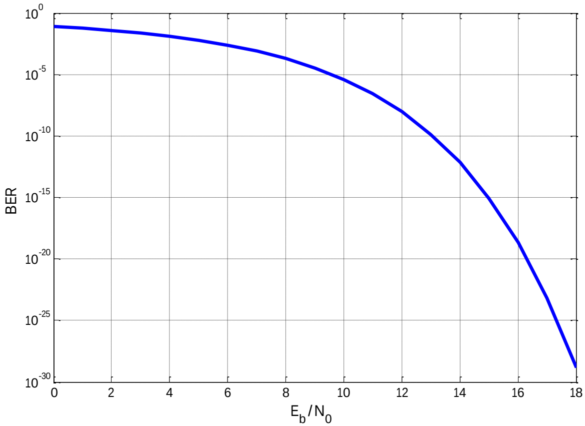
RF Processing

- Reduce RF processing complexity and cost through:
  - using one transmit-and-receive antenna
  - not using a duplexer (since only half-duplex operation is supported)
  - on-chip integrating power amplifier (since transmit power is limited)

Reliability Under Extreme Coverage Conditions

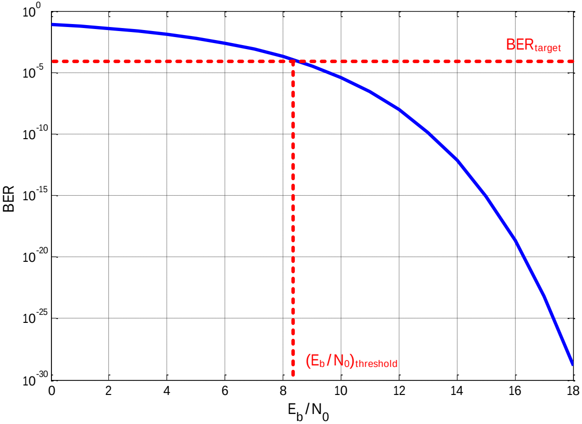
Radio Quality

- Reliability $\Rightarrow$ bit error rate (BER) $\leq$ target BER
- The energy per bit to noise power spectral density ratio (E_b/N_0) is defined as the ratio of the energy per bit (E_b) to the noise power spectral density (N_0)



Radio Quality

- Reliability $\Rightarrow$ bit error rate (BER) $\leq$ target BER
- The energy per bit to noise power spectral density ratio (E_b/N_0) is defined as the ratio of the energy per bit (E_b) to the noise power spectral density (N_0)



Radio Quality

$$BER \leq BER_{target} \Leftrightarrow \frac{E_b}{N_0} \geq \left(\frac{E_b}{N_0}\right)_{threshold}$$

- $(E_b/N_0)_{threshold}$ does not depend on the signal bandwidth and bit-rate
- The SNR, or equivalently the carrier-to-noise ratio (CNR or C/N), is defined as the ratio of the received signal power C to the power of the noise N within the bandwidth of the transmitted signal

$$SNR = \frac{C}{N} = \frac{E_b/T_b}{N_0 B} = \frac{E_b}{N_0}\frac{R_b}{B}$$

where B is the signal bandwidth in Hz, and R_b is the bit-rate in b/s.

Receiver Sensitivity

$$BER \leq BER_{target} \quad \Leftrightarrow \quad SNR \geq \underbrace{\left(\frac{E_b}{N_0}\right)_{threshold} \frac{R_b}{B}}_{SNR_{threshold}}$$

$$\Leftrightarrow \quad S\,(\text{dBm}) \geq \underbrace{SNR_{threshold}\,(\text{dB}) + N\,(\text{dBm})}_{\text{Receiver sensitivity}}$$

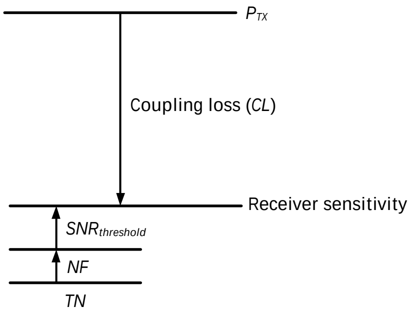
- N (dBm) is the background noise power at the receiver $= TN$ (dBm) $+NF$ (dB)
  - TN is the thermal noise caused by thermal agitation of charge carriers: $-174 + 10\log_{10}(B)$
  - NF is the noise figure caused by RF components

Maximum Coupling Loss

- The Maximum Coupling Loss (MCL) defines the maximum loss the system can cope with between a transmitter and a receiver:

$$MCL\ (\mathrm{dB}) = P_{Tx} - \underbrace{\left(SNR_{threshold} - 174 + 10\log_{10}(B) + NF\right)}_{\text{Receiver sensitivity}}$$

where P_{Tx} is the transmit power in dBm.



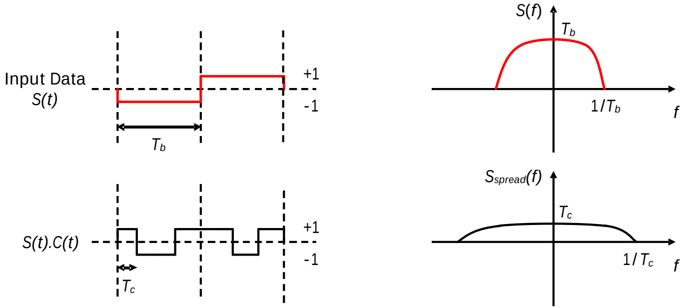
How to Improve Coverage?

- Coverage targets are usually specified in terms of *MCL*
- Increasing P_{Tx}, or lowering *NF*, leads to higher device complexity and cost $\Rightarrow$ inadequate solutions
- Reducing *B* leads to lower network capacity $\Rightarrow$ inadequate solution
- Reducing $SNR_{threshold}$
  - LoRaWAN: optimised radio modulation that uses spread spectrum $\Rightarrow$ LoRa
  - NB-IoT: repetitions and efficient HARQ retransmissions

Chirp Spread Spectrum in LoRaWAN

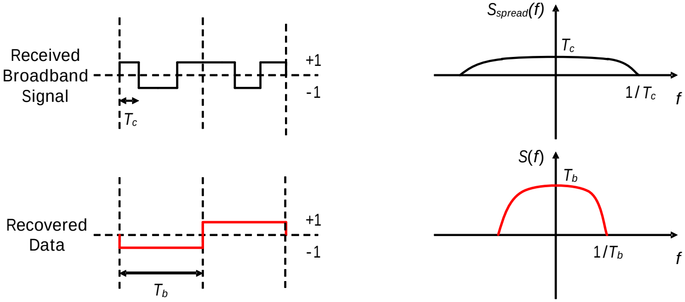
What is Spread Spectrum?

- Spread-spectrum techniques deliberately spread a signal in the frequency domain, resulting in a signal with a wider bandwidth
- Direct-sequence SS (DSSS), frequency-hopping SS (FHSS), time-hopping SS (THSS), and chirp SS (CSS) are forms of spread spectrum
- Spreading process in DSSS systems: at the transmitter, the input data $S(t)$ is multiplied with a spreading code $C(t)$



What is Spread Spectrum?

- De-spreading process in DSSS systems: at the receiver, $S(t)$ is re-covered by re-multiplying with the same spreading code $C(t)$



Why Spread Spectrum?

- Spread spectrum compensates for the SNR degradation

$$SNR = \frac{E_b}{N_0}\frac{R_b}{B} \Rightarrow \left(\frac{E_b}{N_0}\right)_{dB} = (SNR)_{dB} + G_p$$

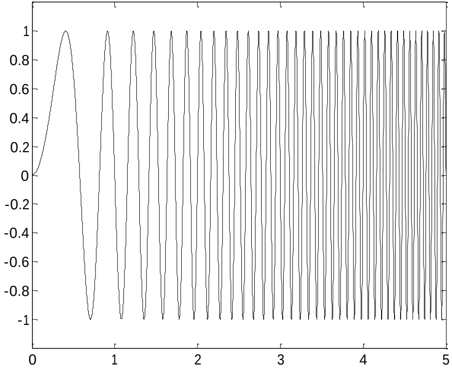
where G_p is the processing gain given by: $G_p = 10\log_{10}(T_b B)$

$$SNR_{threshold} = \left(\frac{E_b}{N_0}\right)_{threshold} - G_p$$

- The higher G_p is
  - the lower $SNR_{threshold}$ is $\Rightarrow$ larger radio coverage
  - the lower R_b is

Linear Chirp

- A linear chirp is a sinusoidal signal whose frequency linearly increases (*up-chirp*) or decreases (*down-chirp*) over time



A sinusoidal linear up-chirp in the time domain

Linear Chirp Theory

- A linear chirp waveform can be written as:

$$x(t) = a(t)sin(2\pi f_0 t + \pi \mu t^2 + \phi_0)$$

where $a(t)$ is the envelope of the chirp signal which is zero outside a time interval of length T, f_0 the initial frequency, μ the chirp rate, or chirpyness, and ϕ_0 the initial phase.

- The instantaneous frequency $f(t)$ is defined as:

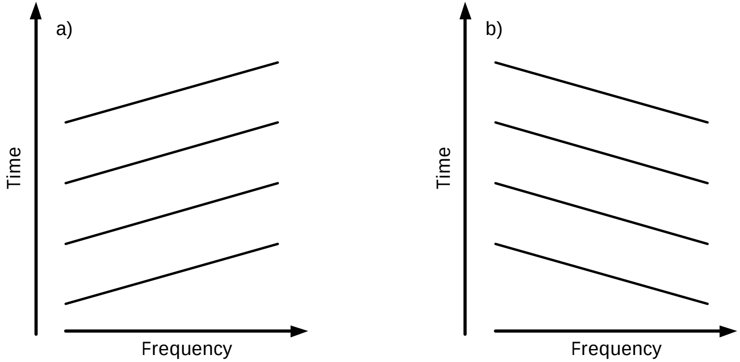
$$f(t) = \frac{1}{2\pi} \frac{d(2\pi f_0 t + \pi \mu t^2 + \phi_0)}{dt} = f_0 + \mu t$$

- The chirp rate μ represents the rate of change of the instantanous frequency:

$$\mu = \frac{df(t)}{dt}$$

Spectrograms of Linear Chirps

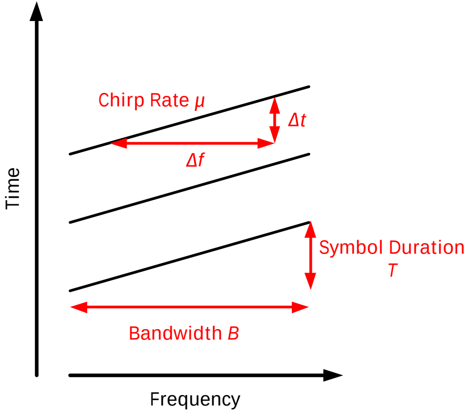
- $\mu > 0 \Rightarrow$ *up-chirps*, $\mu < 0 \Rightarrow$ *down-chirps*



Spectrograms of linear *up-chirp* (a) and *down-chirp* (b)

Bandwidth Spreading

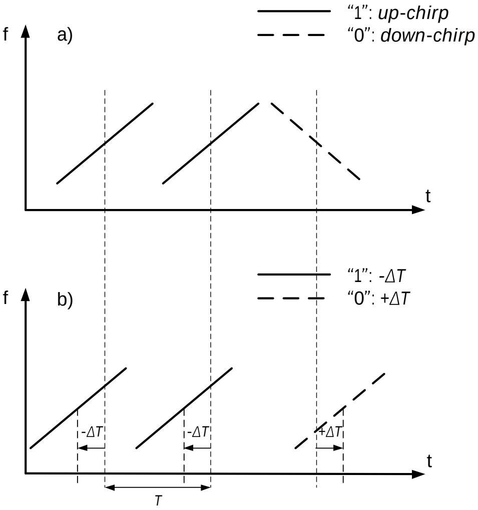
- The bandwidth B is defined as the range of the instantaneous frequency: $B = |\mu| T$
- The processing gain is given by the time-bandwidth product TB



What is Chirp Spread Spectrum?

- Chirp Spread Spectrum (CSS) is a spread spectrum technique that uses wideband linear frequency modulated chirps to encode information
- Encoding information using *up-chirp* and *down-chirp* signals:
  - Example: "1" $\Rightarrow$ transmit an *up-chirp*, "0" $\Rightarrow$ transmit a *down-chirp*
  - Chirps are transmitted in equidistant time steps
- Encoding information using only one chirp waveform with Pulse-Position Modulation (PPM):
  - M bits are encoded by transmitting a single *chirp* in one of 2^M possible time shifts $\Rightarrow$ bit-rate = M/T in b/s
  - Chirps are not transmitted in equidistant time steps
- At the receiver, the wanted information is re-covered through de-chirping

Example: Binary Orthogonal Keying (BOK) Schemes



a) BOK using *up*- and *down*- chirps b) BOK using PPM

Advantages of CSS

- CSS is robust to interference, multipath fading, and Doppler effect
- Time and frequency offsets between transmitter and receiver are equivalent, greatly reducing the complexity of the receiver design

Why CSS?

CSS provides a low-complexity, low-cost, low-power, yet robust alternative to the traditional SS techniques

Repetitions in NB-IoT

Signal Combination

- Users in extreme coverage conditions blindly repeat information (without any feedback from the receiver)
- The receiver accumulates the blindly transmitted signals and combines all the repetitions
- Repetitions compensate for the *SNR* degradation

$$(SNR)_N \ (\text{dB}) = \underbrace{10\log_{10}(N)}_{G_p} + (SNR)_1$$

where $(SNR)_N$ is the ideal *SNR* after combining *N* transmissions and $(SNR)_1$ is the *SNR* of a single transmission.

$$(SNR)_N \geq SNR_{threshold} \Rightarrow (SNR)_1 \geq \underbrace{SNR_{threshold} - 10\log_{10}(N)}_{\text{Reduced } SNR_{threshold}}$$

Real vs. Ideal Processing Gain

- In practice, combining two signals is rarely perfect: signal impairments will result in a lower overall processing gain
- For N between 2 and 16, the ideal gain can be achieved without any visible degradation[1]

[1] Simulations have been carried out for EC-GSM-IoT in O. Liberg et al., *Cellular Internet of Things - Technologies, Standards, and Performance*. Cambridge, MA, USA: American Press, 2017.

Low Power Consumption

Deep Sleep Mode

- Most of the IoT applications require infrequent transmission of small data volumes
- Idle devices enter in deep sleep mode. They:
  - shut down their transceiver
  - keep track of time and scheduled events via a low-power oscillator (that is kept running)
- Devices wake up from deep sleep to:
  - transmit data
  - open receive windows, or monitor paging channels

Battery Lifetime

- Battery lifetime is increased through:
  - optimizing device reachability:
    - LoRaWAN: Class A devices open two short DL receive windows only after an uplink transmission.
      Class B devices extend Class A by adding scheduled receive windows.
    - NB-IoT: devices monitor paging channels either periodically, or only after a mobile-originated data transfer (for a short period of time).
      extended Discontinuous Reception (eDRX) and *Power-Saving Mode (PSM)* support these operations.
  - reducing signaling messages when a device needs to transmit data
    - LoRaWAN: uncoordinated data transmission
    - NB-IoT: the device context is maintained during power-saving states, avoiding unnecessary signaling

High Capacity

Support for Massive Number of Low-Rate Devices

- Trading off data rate for coverage
- How to increase network capacity?
  - LoRaWAN uses multiple orthogonal spreading factors simultaneously on the same channel
  - NB-IoT uses single-tone transmissions in the UL when coupling loss is high

Why Single-Tone Transmissions?

- The channel capacity C is given by:

$$C = B\log_2(1+\frac{S}{N}) = B\log_2(1+\frac{S}{N_0B})$$

- When coupling loss is high, $\frac{S}{N_0B} \ll 1 \Rightarrow \ln(1+\frac{S}{N_0B}) \approx \frac{S}{N_0B}$.

$$\Rightarrow C = \frac{S}{N_0}\log_2(e)$$

C no longer depends on B

$\Rightarrow$ allocate a single tone (subcarrier) for devices in bad coverage to avoid resource wastage

Simplified Network Topology and Deployment

Network Topology and Deployment

- LoRaWAN has a simple network architecture and operates in license-free bands $\Rightarrow$ low-cost deployment
- NB-IoT reuses LTE frequency bands and infrastructure (through software upgrade) $\Rightarrow$ fast time-to-market

Outline

LoRa Radio Interface

What is LoRa?

Definition of LoRa

LoRa is a wireless modulation technique that uses Chirp Spread Spectrum (CSS) in combination with Pulse-Position Modulation (PPM).

- Processing gain given by $g_p = BT$
- Variable number of bits encoded in a symbol

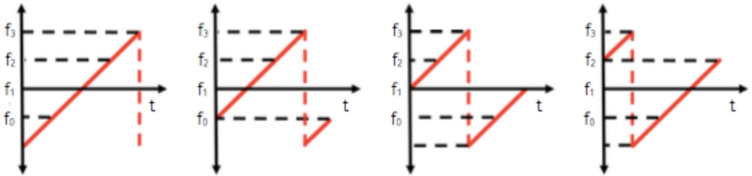
$$R_b = \frac{\log_2(g_p)}{T} = \log_2(g_p) \cdot \frac{B}{g_p}$$

- Spreading factor SF given by $\log_2(g_p)$

$$R_b = SF \cdot \frac{B}{2^{SF}}$$

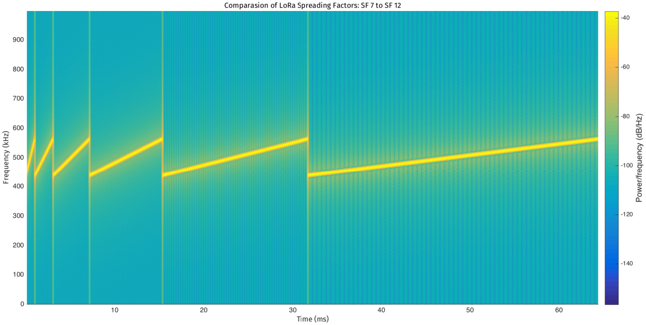
LoRa Symbols

- $\log_2(g_p)$ bits are encoded by transmitting a single *chirp* in g_p possible cyclic time shifts
- Example: $g_p = 4 \Rightarrow 2$ bits/symbol



LoRa Spreading Factors

- LoRa uses spreading factors from 7 to 12



LoRa Bit-Rate

- LoRa includes a variable error correction scheme that improves the robustness of the transmitted signal at the expense of redundancy
- Given a coding rate CR, the bit-rate is given by:

$$R_b = SF \cdot \frac{B}{2^{SF}} \cdot CR$$

- R_b can also be written as:

$$R_b = SF \cdot \frac{B}{2^{SF}} \cdot \frac{4}{4 + CR}$$

with $1 \leq CR \leq 4$, and $7 \leq SF \leq 12$

LoRa Radio Optimization

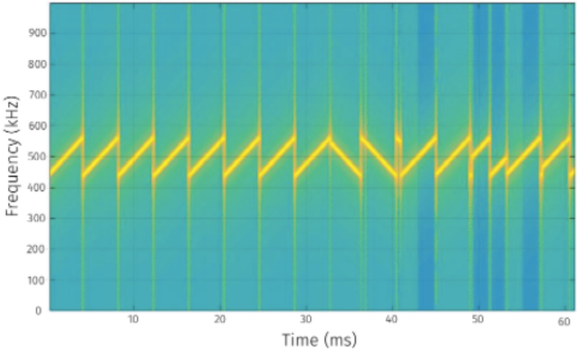
Spreading Factor	Bit Rate (kb/s)	Sensitivity (dBm)
6	9.375	-118
7	5.468	-123
8	3.125	-126
9	1.757	-129
10	0.976	-132
11	0.537	-134.5
12	0.293	-137

($CR = 1$ and $B = 125$ kHz)

- Higher spreading factors lead to lower sensitivity and larger coverage
- Lower spreading factors lead to higher data rates

LoRa Physical Layer

- LoRa transmission consists of:
  - 8 preamble (*up-chirp*) symbols
  - 2 synchronization (*down-chirp*) symbols
  - 5 modulated symbols (payload)



LoRa Channels

- Operates in license-free bands all around the world
  - 433, 868 (EU), 915 MHz
- EU 863-870MHz ISM Band
  - Default radiated transmit output power by devices: 14 dBm
  - Minimum set of three channels, maximum of 16 channels

Modulation	Bw [kHz]	Freq [MHz]	Data Rate	Nb Channels	Duty cycle
LoRa	125	868.10 868.30 868.50	DR0 to DR5 0.3-5 kbps	3	<1%

ETSI Limitations

- Restrictions on the maximum time the transmitter can be on or the maximum time a transmitter can transmit per hour
- Choice between
  - Duty-cycle limitation
  - Listen Before Talk Adaptive Frequency Agility (LBT AFA) transmissions management
- The current LoRaWAN specification exclusively uses duty-cycled limited transmissions to comply with the ETSI regulations

Duty Cycle Limitation

- The LoRaWAN enforces a per sub-band duty-cycle limitation
  - Each time a frame is transmitted in a given sub-band, the time of emission and the on-air duration of the frame are recorded for this sub-band
  - The same sub-band cannot be used again during the next T_{off} seconds where:

$$T_{off} = \frac{TimeOnAir}{DutyCyleSubband} - TimeOnAir$$

  - During the unavailable time of a given sub-band, the device may still be able to transmit on another sub-band
  - The device adapts its channel hopping sequence according to the sub-band availability

Example

A device just transmitted a 0.5 s long frame on one default channel. This channel is in a sub-band allowing 1% duty-cycle. Therefore this whole sub-band (868 – 868.6) will be unavailable for 49.5 s

From LoRa to LoRaWAN

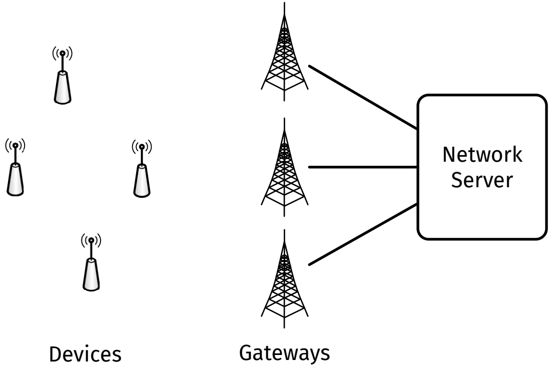
- LoRa
  - Modulation technique for LPWAN
- LoRaWAN
  - Uses LoRa modulation on physical layer
  - Proposes a MAC layer for access control
  - Specified by LoRa Alliance

LoRaWAN Timeline

- Cycleo first introduced LoRa in 2009
  - M2M communications
  - Large coverage
- Semtech acquired Cycleo in 2012 for 5 M$!
  - Patents filed in 2014
- LoRa Alliance initiated in 2014
  - Actility, Cisco, Bouygues, IBM, Orange, SK Telecom, KPN, ZTE, Semtech, La Poste, SoftBank, Swisscom, etc.
  - LoRaWAN 1.1 specification in 2018

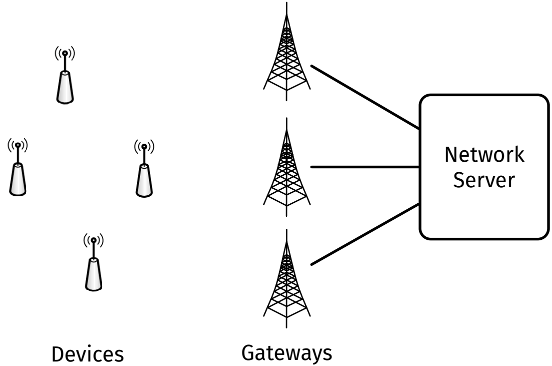
LoRaWAN Physical Architecture

End-Devices



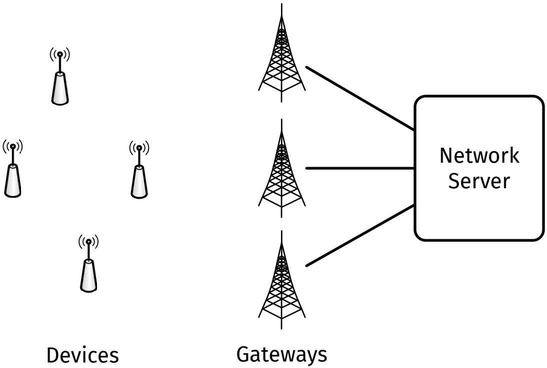
- End-devices are also called motes or devices
- Communicate to one or more gateways via a wireless interface using single hop LoRa or FSK

Gateways



- Gateways are also called concentrators or base stations
- Forward Frames between devices and network server
- Connected to the network server via IP interfaces

Network Server



- Network server is a central server located at the backend
- Provides mobility, frame control, and security functions
- Adapts data transmission rates

LoRaWAN General Characteristics

- LoRaWAN network architecture is typically laid out in a star-of-stars topology
- All end-point communication is generally bi-directional
  - Uplink communications are predominant
- Data rates ranging from 300 bps to 5.5 kbps
  - Two high-speed channels at 11 kbps and 50 kbps (FSK modulation)
  - Eight channels: bandwidth 125 kHz or 250 kHz
  - Support for adaptive data rate (power and spreading factor control)
- Secure bi-directional communication, mobility, and localization
  - Device authentication, message encryption, and frame counter

LoRaWAN Protocol Architecture

Uplink transmission

- Uncoordinated data transmission
  - Devices transmit without any coordination on a randomly chosen channel
  - Regulated maximum transmit duty cycle
  - Regulated maximum transmit duration (or dwell time)

LoRaWAN Access Method

LoRaWAN is an ALOHA-type protocol: transmission by the device is based on its own communication needs with a small variation based on a random time basis

Device Classes

- Class A
  - Each uplink transmission is followed by two short downlink receive windows
  - Adapted for applications that only require downlink communication from the server shortly after the end-device has sent an uplink transmission

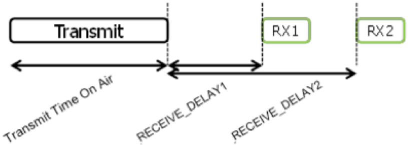
- Class B
  - In addition to class A, receive windows are opened at scheduled times
  - A time synchronized Beacon is sent by the gateway
- Class C
  - Nearly always open receive windows (unless transmitting)

Messages

- Uplink messages
  - Sent by devices to the NS
  - Relayed by one or multiple gateways
  - [Preamble, PHDR, PHDR_CRC, Payload, CRC]
- Downlink messages
  - Sent by the NS to only one device and is relayed by a single gateway
  - [Preamble, PHDR, PHDR_CRC, Payload]

Receive Windows for Class A Devices

- First receive window
  - Same channel (and data rate) as the uplink
- Second receive window
  - Predefined channel and data rate, and possibility to modify it by MAC commands



MAC Header

- Format
  - [MAC type, ..., Device Address, Frame Control, Frame Counter, Frame Options, Frame Port, Payload]
- Message Types
  - Join Request
  - Join Accept
  - Unconfirmed Data Up
  - Unconfirmed Data Down
  - Confirmed Data Up
  - Confirmed Data Down
  - RFU
  - Proprietary

ACK in Frame Control

- If the ACK (demanding acknowledge) sender is an end-device, the network will send the acknowledgement using one of the receive windows opened by the end-device after the send operation
- If the sender is a NS, the end-device transmits an acknowledgment at its own discretion, possibly piggybacked with the next Data message
- A message is retransmitted (predefined number of times) if an ACK is not received

Frame Counter

- Each device has two frame counters
  - Uplink frames, incremented by the device
  - Downlink frames, incremented by the NS

MAC Commands

- Commands are exchanged between devices and NS, not visible to the application layer
- Examples
  - Indicate the quality of reception of the device
  - Indicate the battery level of a device
  - Request the device to change data rate, transmit power, repetition rate or channel
  - Sets the maximum aggregated transmit duty-cycle of a device
  - Change to the frequency and the data rate set for the second receive window (RX2) following each uplink

Data Stored in Each device

- Device address
  - 7 bit network identifier
  - 25 bit network address arbitrarily assigned by the admin
- Application Identifier
  - 64 bits that uniquely identify the owner of the device (EUI-64)
- Session key
  - Used for integrity check and encryption/decryption of MAC only messages
- Application Session key
  - Used for integrity check and encryption/decryption of application data messages

Two Ways of Activation

- Over the air activation
  - Necessitates a globally unique end-device identifier (DevEUI), the application identifier (AppEUI), and an AES-128 key (AppKey)
  - Two MAC messages between NS and devices: Join and Accept
- Activation by Personalization
  - No MAC messages
  - The DevAddr and the two session keys NwkSKey and AppSKey are directly stored into the end-device

LoRa Radio Optimization

Spreading Factor	Bit Rate (kb/s)	Sensitivity (dBm)
7	5.468	-123
8	3.125	-126
9	1.757	-129
10	0.976	-132
11	0.537	-134.5
12	0.293	-137

(RC = 1 and B = 125 kHz)

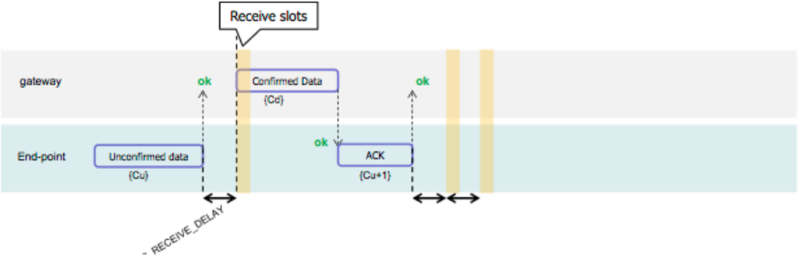
- Higher spreading factors lead to better sensitivity and larger coverage
- Lower spreading factors lead to higher data rates

Adaptive Data Rate

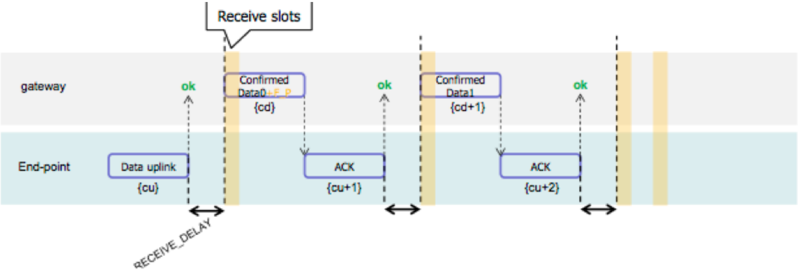
- Objectives
  - Increase battery life
  - Maximize network capacity
- Data rate validation
  - A device periodically sets the ADR acknowledgment bit and waits for an acknowledgment from the network
  - If an ACK is not received, the device switches to the next lower data rate that provides a longer radio range

Wrap-up Example (1/3)

Wrap-up Example (2/3)



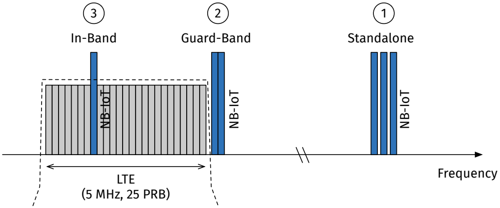
Wrap-up Example (3/3)



NB-IoT

What is NB-IoT?

- 3GPP radio technology standard: releases R13 and R14
- Reutilizes cellular telecommunication bands
  - Three possible operation modes: in-band, guard-band, or standalone
- Reutilizes cellular telecommunication infrastructure
- Shares common characteristics with LTE networks
  - Modulation technique
  - Access method
  - Core network functions



NB-IoT Characteristics

- Radio resource: 1 LTE Physical Resource Block (PRB)
  - Bandwidth: 180 kHz
  - Duration: 0.5 ms
- Multiple access: Downlink OFDMA, Uplink SC-FDMA
- Modulation scheme
  - Downlink: QPSK
  - Uplink: QPSK (multi-tone), $\pi/4$-QPSK, or $\pi/2$-BPSK (single-tone)
- Link budget: 164 dB
- Data rate: $\sim$ 250 kbps in the DL and $\sim$ 250 kbps in the UL (multi-tone)

Leveraging LTE mechanisms in NB-IoT

- Communication channels
  - Broadcast channel
  - Shared control channels (uplink and downlink)
- Access method
  - Cell acquisition and registration
  - Random access procedure
  - Scheduling of uplink and downlink transmissions
- Localization and mobility management (in idle mode)

NB-IoT Access Method

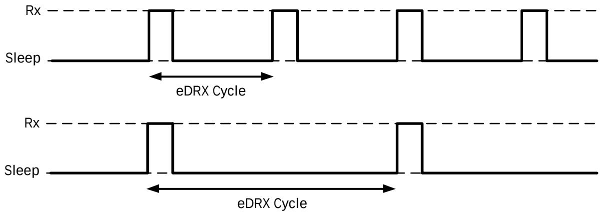
NB-IoT access is performed in two steps: random access then scheduled transmission

Optimizing LTE mechanisms for NB-IoT

- Coverage extension
  - Repeating the same transmission several times, available on all channels
  - Achieves extra coverage (up to 20 dB compared to GPRS)
- Energy saving
  - Monitoring paging channels either periodically, or only after a mobile-originated data transfer (for a short period of time).
    - extended Discontinuous Reception (eDRX)
    - Power-Saving Mode (PSM)

extended Discontinuous Reception (eDRX)

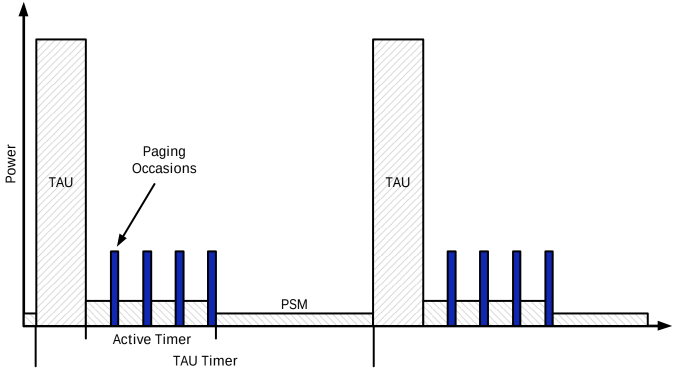
- How often an idle device monitors paging channels?
- An eDRX cycle is the time period between two paging occasions the device needs to monitor (up to 2 h, 54 min, and 46 s)
- In between these two occasions, the device is assumed to be in deep sleep mode
- eDRX cycle is negotiated on a per device basis



Two possible eDRX cycle configurations

Power-Saving Mode (PSM)

- In PSM, an idle device does not monitor paging channels ⇒ unreachability
- A device leaves PSM to send application data or a periodic tracking area update message



Operation in PSM including periodic TAU

Power-Saving Mode (PSM)

- After data transfer, the device monitors paging occasions until an active timer expires
- When the active timer expires, the device re-enters PSM and is unreachable until the next mobile-originated event
- The tracking area update period is configurable (up to a year)