

Dedicated networks for IoT : PHY / MAC state of the art and challenges

Claire Goursaud, Jean-Marie Gorce

► To cite this version:

Claire Goursaud, Jean-Marie Gorce. Dedicated networks for IoT : PHY / MAC state of the art and challenges. EAI endorsed transactions on Internet of Things, European Alliance for Innovation, 2015, <10.4108/eai.26-10-2015.150597>. <hal-01231221>

HAL Id: hal-01231221

<https://hal.archives-ouvertes.fr/hal-01231221>

Submitted on 19 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dedicated networks for IoT : PHY / MAC state of the art and challenges

C. Goursaud^{1,*}, J.M. Gorce¹

¹Univ Lyon, INSA Lyon, Inria, CITI, F-69621 Villeurbanne, France

Abstract

This paper focuses on the emerging transmission technologies dedicated to IoT networks. We first analyze the classical cellular network technologies when taking into account the IoT requirements, and point out the need of dedicated technologies for IoT. Then, we present the PHY and MAC layers of the technologies that are already deployed, or likely to be deployed: UNB by SigFox, CSS by LoRaTM, Weighless, and RPMA by Ingenu. We then compare their performances to highlight their pros and cons. Finally, we discuss on the open research challenges that still need to be addressed.

Received on 22 September 2015; accepted on 21 October 2015; published on 26 October 2015

Keywords: Internet of Things, M2M, PHY and MAC layer, Long Range transmissions.

Copyright © 2015 C. Goursaud, J.M. Gorce, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.26-10-2015.150597

1. Introduction

Internet of Things (IoT) is going to take a major place in the telecommunications market as announced in technical and public medias. The paradigm of IoT relies on the deployment of billions of objects having the capability of transmitting information about their context and environment and to create a real-time, secured and efficient interaction between the real and the virtual worlds, pushing them to evolve from the state of cousins to the state of Siamese twins. IoT revealed to be a key technology for solving societal issues such as digital cities, intelligent transportation, green environment monitoring or medical care and elderly person monitoring.

IoT has strong connections with machine-to-machine (M2M), and sometimes in literature, both refer to the same idea. From our point of view, IoT covers a broader scope including as well the technology and the applications. On the opposite, M2M refers to the technologies that allow machines or objects to communicate.

Whatever, from the technical point of view, the main challenge of this new paradigm is to let a very huge number of machine type devices (MTDs) be connected to the Internet at a low cost, with a limited

infrastructure and featuring a very long life time with very small battery or energy needs.

In this global picture, we may consider different technical issues. M2M has been first defined to connect MTDs in their vicinity. The proposed solutions extensively rely on the research results produced over the last 20 years for ad-hoc and wireless sensor networks. Starting 20 years ago from theoretical concepts, this very active research area went up to the definition of full standards (802.15.4, 802.15.6, Zigbee, Bluetooth) which already found a market.

More recently, the IoT paradigm has been extended to the problem of connecting all these MTDs to the Internet, and through Internet to anyone or anything. The massive connection of objects spread over the world is a challenge that has some similarities with the paradigm of cellular networks which aimed at connecting people. This similarity attracted the interest of mobile network providers, to exploit such attractive potential market and IoT has been identified as a target for the future 5G [1], while several proposals already exist to adapt the 4G technology to IoT [2]. Nevertheless the IoT paradigm may present some very specific features that cannot be easily integrated with the constraints of cellular networks. In many applications, the individual targeted throughput is very low and the capacity is not a relevant criterion. On the opposite, the latency, the energy efficiency or the reliability are more

*Corresponding author. Email: claire.goursaud@insa-lyon.fr

critical. Except for cars or few other mobile objects, IoT may rely mostly on static nodes. But the dynamic of the problem comes from the fact that these nodes may transmit a packet with a very low probability (e.g. once a week, once a month, or even once a year !!!). Keeping these nodes continuously connected would be not efficient and an important issue is to allow a fast and reliable bursty connection. In section 2 we first summarize the recent results attempting to use 4G networks for serving IoT nodes at a large scale and we also discuss the challenges for 5G. We then enumerate the features that may claim for the design of new Physical (PHY) layers. Section 3 and section 4 detail the Ultra-narrow band (UNB) and the Chirp Spread Spectrum (CSS) technologies. Section 5 briefly describes two additional technologies. We compare the performances of the four technologies in section 6. Furthermore, we discuss and provide perspectives in section 7.

2. PHY/MAC design for massive IoT services

Because the massive connection of MTDs is commonly recognized as an important challenge [1], different technologies and strategies are currently under investigation. The most straightforward approach relies on exploiting the current cellular networks to absorb this new traffic as summarized in [2]. The first idea relies on exploiting the second generation (2G) technology (e.g. GSM) which is progressively freed, the voice traffic being progressively offloaded toward 3G and 4G networks. If 2G networks present interesting features in a mid-term perspective [3], some technical limitations cannot be alleviated: the granularity offered by the native GSM cannot serve a million of MTDs, the system is not energetically efficient and the signaling is oversized for such small packets. Further, if the provider would keep the 2G system active only for MTDs, it may be anticipated that using these radio resources with a more appropriate technology would offer a more efficient perspective.

While UMTS is assumed less efficient for MTDs for energy and coverage range reasons, LTE appeared more appealing [2]. However several issues have to be risen. Above some classical considerations such as energy consumption or coverage limitation, the most challenging issues are related to the medium access protocols which are inappropriate for massive access. More specifically, the RACH (Random Access Channel) channel used for contention access for mobile users would be overloaded by massive MTDs request. Although different subtleties have been suggested to share optimally the RACH channel [4], the most efficient way relies on providing MTDs and UEs (User Equipment) with separate access channels. Due to the different nature of the data flows and their QoS

objectives, the separation seems inevitable but the right question is to find the best level at which the separation should occur. Therefore, using a separate network for MTDs is appealing even if additional deployment costs have to be supported. This drawback is balanced by the efficiency gain a dedicated PHY layer may bring to MTDs. More precisely the challenge is to design jointly a PHY/MAC protocol being able to manage a high density of nodes with bursty transmissions, with a reduced signaling overhead, a long range capability and very low energy requirements.

Different technologies have been already investigated to address in this way this IoT paradigm. The following sections investigate these recent technologies and present a comparison of their main features.

3. UNB

In 2004, Walker [5, 6] first proposed the use of VMSK (Very Minimum Shift Keying) to compress data transmission in the smallest possible band. However, in practice, this modulation technique did a step forward but didn't reach the claimed ultra narrow frequency occupancy [7]. Nonetheless, the french company SigFox has successfully attempted to develop, patent, and deploy a new technology for UNB [8]. UNB standardization is now ongoing.

3.1. UNB PHY Principle

The UNB PHY layer used in SigFox's network is very simple. Binary data are broadcast with a BPSK modulation at a very low rate $R_b = 100$ bps. The transmitted signal thus occupies a band of about $b = 100$ Hz.

The novelty comes from the fact that the multiple transmissions are performed at a carrier frequency chosen in a much larger band B (typically 192 kHz in the 868 MHz (resp 915 MHz) ISM band in Europe (resp in the America)). As narrow band signals may suffer from flat fading, Frequency Hopping (FH) inside B is supported to introduce diversity and improve the reliability.

A key parameter of such system is the oscillator precision, which induces an offset between the targeted frequency and the actual one. Indeed, the very low signal bandwidth used in UNB implies a high sensitivity to the oscillators precision. Objectively, the system is referred as UNB when the frequency uncertainty is higher than the signal bandwidth. In this case, contrarily to NB systems, it is not possible to obtain non-overlapping frequency channels with reasonable guard intervals [11]. This leads to a new paradigm for the multiple access scheme, described in the next section.

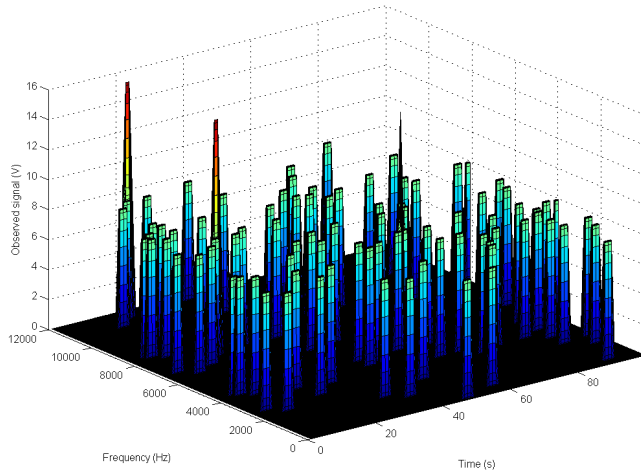


Figure 1. Example of temporal & spectral repartitions of nodes.

3.2. UNB associated MAC

The inherent MAC associated to UNB is RFTDMA (Random Frequency and Time Division Multiple Access). Nodes access to the wireless medium randomly both in time and frequency domain, and without any contention-based protocol. This corresponds to an Aloha-based protocol without preliminary sensing of the channel occupancy. Nonetheless, contrary to classical Aloha transmissions, the carrier frequencies are chosen in B inside a continuous interval, instead of a predefined discrete set. To illustrate the system behavior, an example is drawn in Fig.1, representing the signal generated by 100 nodes and observed by a receiver in the time-frequency plan.

The benefits of RFTDMA are:

- no energy-consumption for medium sensing. Indeed, due to the very large size of the cells, sensing the channel at the node location would not ensure collision-free signals at the base station.
- no need for time-synchronization over the network. Thus beacon packets can be eliminated.
- no constraint on the oscillator precision. As any frequency can be chosen in the whole band, even the cheapest oscillators can be used without performance degradation.

Nevertheless, the uncontrolled medium access leads to interference or packet collisions between the active users. This will be discussed in section 3.3.

From the receiver's point of view, as shown in Fig.1, the monitored bandwidth B is filled by a combination of narrow-banded signals randomly located in time and frequency. Their demodulation relies on efficient SDR

(Software Design Radio) algorithms designed to analyze the total band, to detect transmitted signals and to retrieve sent data. This is done with a FFT block applied to the received signal followed by an adaptive detector which aims at identifying the spectral signatures of the transmitted UNB signals. So any uncontrolled frequency shift at any transmitter is not problematic if this shift is fix during the short message duration. For each detected transmission, the appropriate frequency band is filtered and demodulated with a standard BPSK demodulator.

To improve reliability, each message can be sent up to 3 times on different frequencies.

The base station then responds exactly on the same frequency which allows to keep the node reception algorithm simpler as the band analysis is not necessary.

Finally, the transmitted packet is made of :

- a preamble of 4 bytes,
- a frame synchronization part of 2 bytes,
- a device identifier of 4 bytes,
- a payload of up to 12 bytes,
- a Hash code to authenticate the packet in SigFox network (variable length),
- Cyclic Redundancy Check (CRC) syndroms of 2 bytes for security and error detection.

3.3. UNB Analysis

As explained in section 1, three main criteria are of major importance in IoT networks which are now used to evaluate the UNB-based technology.

Communication range. The range usually depends on the channel characteristics, interference and the receiver noise level. At a first glance, we consider a Line of Sight (LoS) transmission and interference free conditions. Thus, the range is first constrained by the noise level given by:

$$N_{dBm} = -174 + NF + 10 \cdot \log_{10}(B), \quad (1)$$

with NF the noise figure of the receiver.

In the reference UNB case described above, the noise floor is thus $-154 + NF$ dBm.

Therefore, in free-space conditions and assuming the antenna gains balance roughly the noise factor loss, the signal to noise ratio is :

$$SNR = P_r - N = T_x + 132 - 20 \log\left(\frac{r}{\lambda}\right), \quad (2)$$

where r is the range and λ the wavelength.

Now considering a SNR threshold of 8 dB and a link margin of 4 dB, the received power is required to be $P_r \geq -142$ dBm.

Thus, an emission power of 14 dBm (maximum radiation power allowed by ETS 300-220 regulation) may ensure thousands kilometers range in free-space. In practice, about a 63 kms range is observed for terrestrial communications. Thus, UNB is suitable for long distance transmissions (up to the horizon).

Interference Sensitivity. Interference occurs when some packets are transmitted simultaneously, which is inherent to RFTDMA. The few interference situations observed in Fig.1 are characterized by the peaks above the average. The characteristics of this interference evaluated in [10] are summarized below.

Consider a multiple access channel with $N = 2$ active transmitters. The received signal can be expressed as:

$$r(t) = \sum_{i=1}^2 s_i(t) \cdot g(f_i, t) \otimes h_i(t) + n(t) \quad (3)$$

where $s_i(t), \forall i \in \{1, 2\}$ are the BPSK symbols sent by the active user i , $g(f_i, t)$ the impulse response of the emission FIR filter (centered at f_i); $h_i(t)$ is the path-loss of the corresponding link, and $n(t)$ is an additive white Gaussian noise with zero mean, and variance σ^2 .

For the sake of simplicity, consider the case where $h_i(t) = \delta(t), \forall i \in [1, \dots, k+1]$. This corresponds to the worst case where both users are at the same distance from the base station and experience the same flat channel. Without loss of generality, we consider that the desired user is #1. The signal used for data recovery is thus:

$$r'(t) = r(t) \otimes g(f_1, t) \quad (4)$$

$$= \sum_{i=1}^{k+1} s_i(t) \cdot g(f_i, t) \otimes g(f_1, t) + n(t) \otimes g(f_1, t). \quad (5)$$

The rejection coefficient relies on the product of the pulse-shapping filter 2 and the matching filter 1:

$$P_i(t, f_1, f_2) = |G(f_1, t) \cdot G(f_2, t)|. \quad (6)$$

More precisely, the rejection coefficient $\beta(\delta_f, t)$ is a function of the frequency shift between the 2 active users $\delta_f = |f_1 - f_2|$, and is normalized by the power of the desired signal:

$$\beta(\delta_f, t) = \frac{P_i(t, f_1, f_2)}{P_i(t, f_1, f_1)} \quad (7)$$

In Fig.2, the rejection coefficient is plotted as a function of the frequency difference (6), for a given point in time. We can observe 2 main areas, with transitions around ± 100 Hz. In the central area, i.e. for low δ_f in the range $[-100, 100]$ Hz, the interference level is important (up to 0 dB when $\delta_f = 0$). Contrarily, in the outer area, i.e. for high δ_f , the interference level

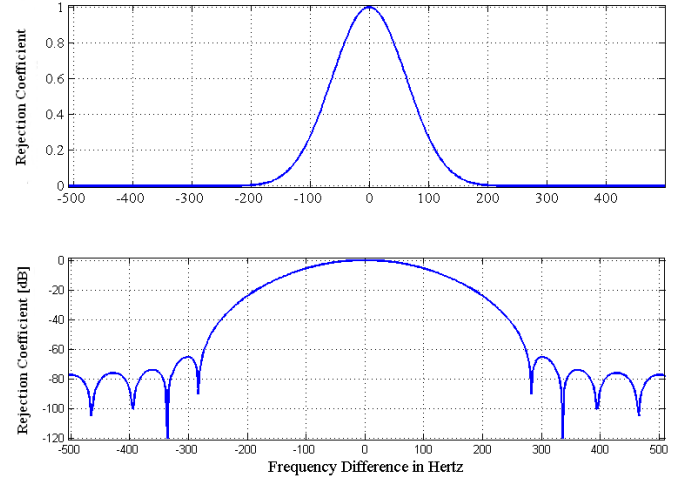


Figure 2. Behavior of the rejection coefficient in linear and logarithmic scale vs frequency difference δ_f .

is low, and mainly distributed around -90 dB. As, the considered band is much larger than 200 Hz (at least 12 kHz), the interferers are positioned with a highest probability in the second area, thus not impacting the desired user detection.

Energy Consumption. Typical emission consumption varies from 20mA to 70mA, while it drops to almost 0 when inactive. The actual energy consumption depends on the message size as well as the emission power. For the considered frequencies, the nodes are allowed to transmit up to 14 dBm in Europe, and up to 21.7 dBm in America. Thus, the energy consumption can be kept very low to comply with the battery-powered IoT nodes.

4. CSS

In this section, we present the general properties of CSS modulation with a deeper focus on a specific technique. The use of CSS for IoT networks was initially developed and patented by Cycléo, a french compagny acquired by Semtech in 2012. The later includes this modulation in LoRaTM devices, along with FSK modulation. Standardization of Low Power Wide Area Network (LPWAN) using such technology is targeted by the LoRaTM Alliance [12].

CSS is a rather classical technique in radar systems but was proposed for the first time for communication systems by Winkler in 1962 [13] and barely used since. For readers unfamiliar with this technique, the following paragraph provides its main features.

4.1. CSS Principle

In contrast to UNB, a CSS transmission occupies a bandwidth much larger than what is actually needed

for the considered data rate. It is thus a subcategory of Direct-Sequence Spread Spectrum (DSSS), which takes advantage of the controlled frequency diversity to recover data from weak signals (even under the noise level). Thus, compared to narrow band transmissions, DSSS permits to alleviate the constraint on the receiver's sensitivity and increase the communication range, but at the cost of a reduced data rate. Thus, DSSS is compliant with IoT networks needs.

Usually, in DSSS, data are spread with a sequence. Each symbol duration is divided into F small chips (with F the spreading factor), whose level is in a finite set. The sequence of chips used by the transmitter is known by the receiver, which searches such pattern in the signal.

However, in CSS modulation, the spreading effect is obtained through a continuously varying carrier frequency [14]. In CSS case, chips do not correspond to physical realizations anymore, but are considered by analogy with coded DSSS. Nevertheless, the spreading factor still characterizes the increase in band occupation. LoRaTM defines the spreading factor SF as follows [15]:

$$2^{SF} = \frac{B}{R_s} = B \cdot T \quad (8)$$

where B is the spread bandwidth, R_s the symbol rate, and $T = \frac{1}{R_s}$ the chirp duration.

So, the original basic element of CSS modulation is the chirp. Its waveform is written as follow :

$$c(t) = \begin{cases} \exp(j\phi(t)) & \text{if } -\frac{T}{2} \leq t \leq \frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

with $\phi(t)$ the chirp phase.

The instantaneous frequency is given by :

$$f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} \quad (10)$$

For linear chirps, such as the ones used by LoRaTM, $f(t)$ is defined by

$$f_{lc}(t) = f_c + \mu \cdot \frac{B}{T} \cdot t \quad (11)$$

with f_c the central carrier frequency. If $\mu = 1$, an up-chirp is obtained, while $\mu = -1$ corresponds to a down-chirp. One may note that for CSS modulation, B corresponds to the spectral occupancy, as well as the difference between the maximum and minimum instantaneous frequency during the chirp [16]. Fig.3 shows an example of an up-chirp. We can verify the linearity of the frequency (a), the quadratic evolution of the phase (b), and the varying period of the in-phase and quadrature components (c).

At the receiver, the following properties are used:

- the multiplication of an up-chirp with an up-chirp (resp down-chirp with down-chirp) leads to

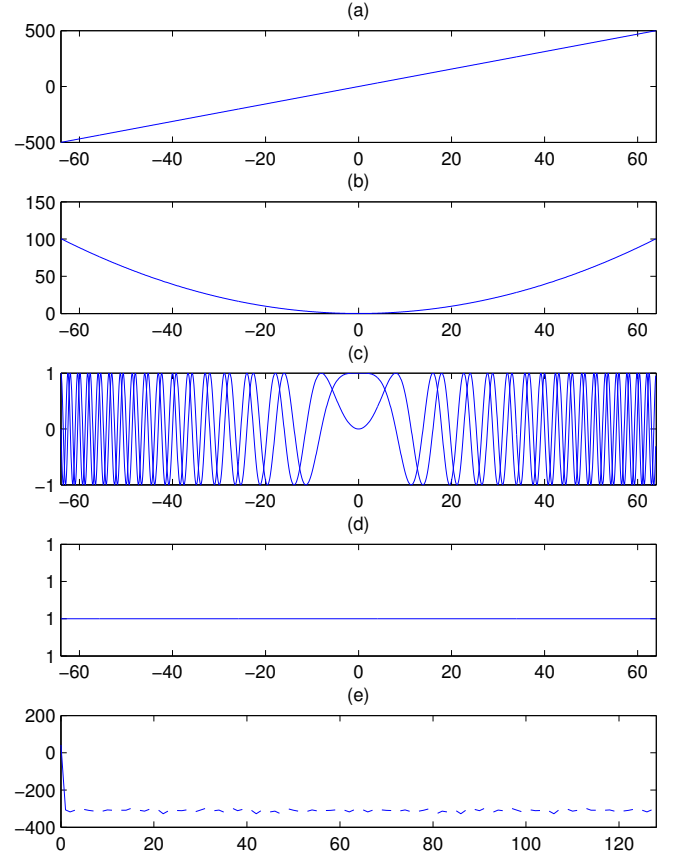


Figure 3. Up raw chirp : (a) frequency evolution during the chirp, (b) the corresponding instantaneous phase, (c) varying period of the phase and quadrature components, (d) the output of the matched filter and (e) its FFT.

an up-chirp (resp down-chirp), as instantaneous frequencies are added,

- the multiplication of an up-chirp with the corresponding down-chirp (i.e. the same instantaneous frequency expression with opposite μ , also called conjugate chirp) leads to a narrow peak at twice the carrier frequency.

Thus, the matched receiver for a linear chirp is performed by multiplication with the conjugate chirp, and leads, as shown in Fig.3(d) to a constant. The output signal is then analyzed to identify the presence or absence of the narrow interference peak (which is at 0 in Fig.3(e)). This basic CSS modulation permits to send one bit per chirp.

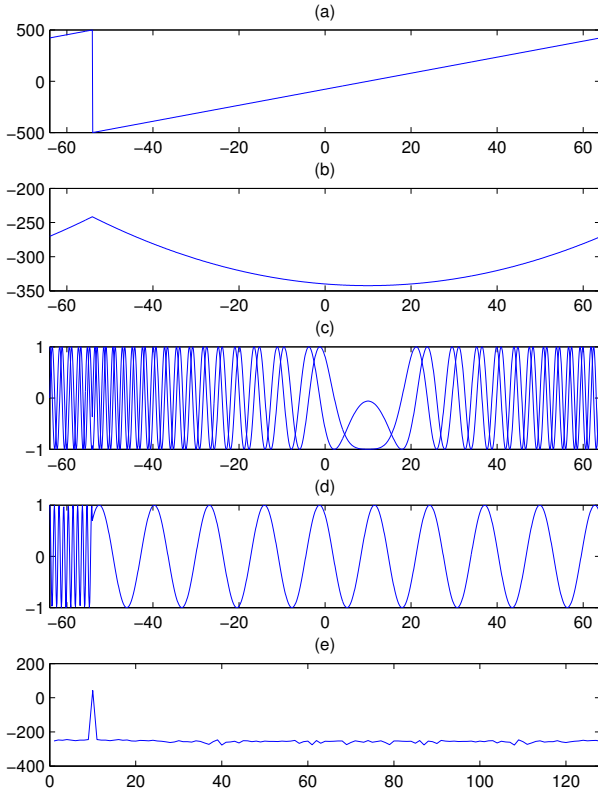


Figure 4. Coded chirp : $SF = 2^7$, raw chirp shifted by 10

4.2. CSS Adaptation for LoRaTM

The CSS modulation used in LoRaTM networks is more evolved and constrained to fit with IoT requirements [17].

First of all, a single LoRaTM chirp may code up to $SF = 12$ bits. To do so, during one chirp period, a specific frequency trajectory is defined for each of the 2^{SF} symbols. This is done by shifting the frequency ramp based on the symbol value, as illustrated in Fig.4. Thus each coded chirp is obtained by a cyclic shift of the reference chirp. This introduces a sharp edge in the instantaneous frequency trajectory, occurring in the example at the chip number 10.

The new expression for the instantaneous frequency of the coded chirp is :

$$f_{cc}(t) = \begin{cases} f_c + \mu \cdot \frac{B}{T} \cdot \left(t - \frac{k}{B}\right) + B & \text{if } -\frac{T}{2} \leq t \leq \frac{k}{B} \\ f_c + \mu \cdot \frac{B}{T} \cdot \left(t - \frac{k}{B}\right) & \text{if } \frac{k}{B} \leq t \leq \frac{T}{2} \end{cases}, \quad (12)$$

with k the number of shifted chips.

Thus, as seen in Fig.4(a-c), the raw chirp pattern is cyclic shifted of 10 chips toward the right.

At the receiver, the multiplication of the received signal with the raw down-chirp (supposedly perfectly synchronized) modulated at a carrier frequency f_d , the instantaneous frequency becomes:

$$f_p(t) = \begin{cases} f_c + f_d - \mu \cdot \frac{k}{T} + B & \text{if } -\frac{T}{2} \leq t \leq \frac{k}{B} \\ f_c + f_d - \mu \cdot \frac{k}{T} & \text{if } \frac{k}{B} \leq t \leq \frac{T}{2} \end{cases}. \quad (13)$$

The product signal is thus made of two periods, each having a constant frequency (Fig.4(d)). The transition between them occurs at the time index corresponding to the value of the coded chirp. In both periods, the frequency linearly depends on the offset and the difference between the 2 levels is B Hz. Therefore sampling the signal at the chip rate, i.e., at B Hz, the instantaneous frequency becomes continuous over the whole chirp and, after subtracting the carrier $f_c + f_d$, is proportional to the shift k . Thus, the FFT of the sampled signal provides a flat response with the peak shifted by the chip shifting value (Fig.4(e)).

To do so, two constraints must be verified :

- the phase of the transmitted signal must be continuous, especially at the transition. Besides, the instantaneous phase must be identical at the beginning and at the end of the symbol. This ensures an exploitable FFT result. This feature is controlled by the transmitter.
- all previous results are obtained when considering an ideal time and frequency synchronization between the transmitter and the receiver. However, any imprecision in time or frequency will be perceived as an additive offset compared to the nominal frequency. Consequently, before the transmission of coded chirps, a preamble made of raw chirps has to be sent, to estimate this offset and determine the reference frequency. One may note that Nanoscale has evaluated that an uncertainty up to 40 ppm is supported [17], which enables the use of cheap devices. Once the synchronization is done, the decoder evaluates the offset of the coded symbols with respect to the reference frequency. For sake of simplicity, from now on, we suppose that the reference is known, and neglect $f_c + f_d$, to get the baseband signal.

In short, the whole modulation process defined by LoRaTM follows. First, a preamble is sent for the intrinsic offset estimation. Bits are divided into words of SF bits. There are 2^{SF} code words, mapped to the 2^{SF} possible offsets. Gray indexing is used to reduce the bit error rate, when a symbol is erroneous.

The typical bandwidth values are 125, 250 and 500 kHz in the HF ISM 900 MHz band, while they are 7.8, 10.4, 15.6, 20.8, 31.2, 41.7 and 62.5 kHz in the LF 160 and 480 MHz band. Besides, the spreading factor SF

can vary from 7 to 12. The chip rate remains unchanged, and equals to B , so the chirp duration is impacted when modifying the spreading factor. A high spreading factor corresponds to a long chirp. Therefore, the raw data rate can be calculated as follows :

$$R_b = SF \cdot \frac{BW}{2^{SF}}. \quad (14)$$

It thus varies from 22 bit/s ($B = 7.8$ kHz and $SF = 12$) to 27 kbit/s ($B = 500$ kHz and $SF = 7$). These degrees of freedom permit to adapt the rate and frequency occupancy to the transmission conditions. In particular, devices at the edge of the gateway range are likely to be assigned with a high SF, while closer devices use lower SFs.

Finally, for a high amount of data to transmit, FH is used. The frequency pattern is predefined, and known by the receiver.

4.3. MAC for CSS networks

The LoRaTM Alliance is currently defining the LoRaWAN protocol which precises the MAC protocol envisioned for LoRaTM. An end device is driven by the LoRaTM MAC master if its Adaptive Data Rate (ADR) mode is enabled. In this case, the LoRaTM MAC has the ability to control the spreading factor, the bandwidth occupation and the RF output power of each node in order to both maximize the node battery life and the network overall capacity. This also permits to transmit with the highest possible rate, thus reducing the time occupancy. Consequently, closer nodes to the gateway benefit from a higher data rate than the nodes at the cell edge.

Data transmission can be bi-directional, even if the uplink is expected to be dominant. The specification thus defines 3 categories of end-devices :

- Class A : the communication is initiated by the end-device. The uplink transmission triggers two short downlink receive windows. The uplink transmission slot is scheduled when needed by the end-device on a random time basis (ALOHA based protocol). This Class A is more suitable for energy constrained nodes.
- Class B: this class adds scheduled receive windows to the Class A random ones. The required time synchronization between the gateway and the end-device is obtained thanks to a time synchronized Beacon sent by the gateway.
- Class C: the end-device is always available for reception, except when transmitting. A low latency is granted, at the cost of a higher energy consumption.

In class A and class B, nodes initiate the uplink transmission. In this case, whenever it wants to transmit data, a node pseudo-randomly chooses a channel from the list of available channels, and performs a Listen Before Talk (LBT) before actually emitting its data. Transmission is performed with the fixed default SF and B values. However, if a node needs to check its connection to the network, a short frame (17 bytes) is sent first at $SF = 8$, then $SF = 10$ and finally $SF = 12$. In this case, the end-device gets a feedback on the link margin of the received test frame.

To improve reliability, a retransmission scheme is available. The number of retransmissions is defined by the end-device, but the LoRaTM MAC master can also modify it.

Finally, the transmitted PHY packet is composed by :

- a preamble of at least 12 raw chirps
- a header (optional) of 4 to 8 symbols
- a payload of 2 up to 255 bytes,
- the CRC calculated for the payload (2 bytes).

4.4. CSS Performance Analysis

Communication range. As in the UNB section, we first evaluate the noise floor. For CSS spread over a frequency band with width B , the noise level is

$$N_{dBm} = -174 + NF + 10 \cdot \log_{10}(B) \quad (15)$$

with $125\text{kHz} \leq B \leq 500\text{kHz}$ for the 900 MHz band.

Thus, in free-space, the link budget is :

$$SNR = T_x + 132 - 20 \log\left(\frac{r}{\lambda}\right) + CG \quad (16)$$

with CG the coding gain due to the spreading. This coding gain is estimated by $CG = 2,5 \cdot SF$ [18].

The theoretical maximal range is about hundreds kilometers but was evaluated to be 22 km in practice. Thus, long ranges are achieved.

Interference Sensitivity. There are two sources of interference: from non-LoRaTM signals, and LoRaTM signals. For the first category, it was estimated [18] that a single tone pulse is not a problem if it is less than 5 dB (resp 19,5 dB) above the desired signal for $SF = 7$ (resp $SF = 12$) with an error correcting scheme of 4/6.

For the second category, one may note that two devices can not use the same SF, on the same frequency at the same time. Indeed, the detection is a linear process. Thus, with two devices transmitting, the FFT output would provide the summation of each FFT, leading to 2 indiscernible peaks. The receiver would not be able to identify which offset to take into account. Nevertheless, one transmission over the two can be

interferer desired	7	8	9	10	11	12
7	-6	16	18	19	19	20
8	24	-6	20	22	22	22
9	27	27	-6	23	25	25
10	30	30	30	-6	26	28
11	33	33	33	33	-6	29
12	36	36	36	36	36	-6

Table 1. Cochannel rejection (dB) for all combinations of spreading factor for the desired and interferer user

successful if one signal is received at least 6 dB above the other (Table 1).

Finally, we have computed and reported in Table 1, the co-channel rejection when considering all couples of SF. We can observe that two devices using different spreading factors can transmit their data simultaneously, as long as none is received with a power significantly higher. We can also note that the rejection coefficient increases with the spreading factors. Thus, the high SF usually assigned to distant nodes for the noise sensitivity also permits to overcome the impact of closer devices which are likely to be received with a higher power level.

Energy Consumption. As for SigFox, the emission power may vary up to 14 dBm. Besides, the circuit consumption depends on the device state. During idle or sleep mode, the supply current is negligible (around 1 μ A), while reception (resp. transmission) consumes up to 11 mA (resp. 125 mA). A complete simulator, which takes into account all specificities of the device behavior is available at [19].

5. Additional technologies

In this section, we present two others emerging communications systems dedicated to IoT networks. Contrarily to UNB and CSS, they are based on more usual technologies, which are adapted to IoT needs.

5.1. Weightless

The third communication technology is proposed by Weightless. Weightless SIG [20] is an organization which aims at providing Weightless standards for IoT networks. These standards are based on the technology initially developed by the British company Neul [21], recently acquired by Huawei in fall 2014. At the time of this article publication, 3 different standards are proposed : Weightless-N, Weightless-P and Weightless-W. Each standard targets different use-cases, but each of them complies with the low-cost, large range and low power consumption as required for IoT.

Weightless-N. This standard is based on narrow band technology, with a differential binary phase shift keying (DBPSK) digital modulation scheme, and is mainly based on Nwave technology [22]. The announced data rates are 30 – 100 kbps. Transmissions are performed in the sub-GHz ISM bands 868 MHz. This leads to a range of 5 kms even in challenging urban environments. A FH algorithm is used to counteract interference and fading. This standard is intended for devices that need one-way communications at a very low cost.

Weightless-P. This second standard improves the first one by allowing two-way communications. This permits to enhance the reliability by using acknowledgment protocols. A multiple access is performed with FDMA+TDMA in 12.5kHz narrow band channels. Thus the BSs are time-synchronized to schedule the transmissions in the slots. However, the range is slightly reduced to 2 km in urban environment.

Weightless-W. The last standard permits to take advantage of the available white-space in the spectrum use. Use of TV white spaces is the primary idea of the founding members of Weightless. The modulation modes vary from 16-QAM to DBPSK, in the 470 – 790 MHz TV band. Besides, spreading is also enabled (with a spreading factor up to 1024) to dynamically adapt the rate and the range to the actual needs. One may note that the BS transmission power is 20 dB higher than the end-device's one. To balance the link budget, the end-device uses a channel with a bandwidth 64 smaller than the BS one. Time Division Duplexing is used to provide uplink and downlink pairing, as spectrum is not guaranteed in the TV white space. The indoor (resp. outdoor) range is up to 5 km (resp. 10 km).

Complementary information on these standards is available only for members of the Weightless SIG.

5.2. RPMA

RPMA (Random Phase Multiple Access) was developed by On-Ramp Wireless. This American company was founded in 2008, to provide connectivity to oil and gas actors. In september 2015, it was renamed Ingenu, and targets to extend its technology to the IoT and M2M market [23].

RPMA is based on DSSS. Transmissions are made in the 2.4 GHz ISM band. Data are first encoded (1/2 rate) and interleaved. The resulting stream is then D-BPSK modulated, before being spread by a Gold Code. The signal is then randomly delayed before transmission. One may note that additional blocks ensure time and frequency synchronization between the BS and the end-devices [24], but are not detailed in this paper.

The spreading factor of the Gold Codes is 2^k with $2 \leq k \leq 13$. Each time the spreading factor is doubled, the processing gain increases of 3 dB. This permits to

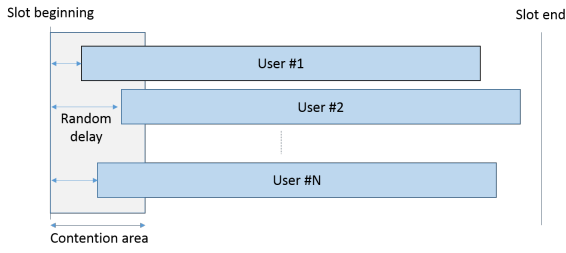


Figure 5. RPMA Multiple Access Principle

adapt the data rate to the propagation conditions. For the uplink, or the downlink broadcast transmission, a unique Gold code is used. On the contrary, for unicast downlink transmission, the Gold code is built with the end-device ID, such that no other end-device is able to decode the data.

Uplink and downlink are performed in an half-duplex way, with a downlink period of 2s followed by an uplink period of 2s. This permits to dynamically adapt the spreading factor to the channel conditions depending on the received power. One may note that a smaller SF can be used in downlink compared to uplink, as the BS is not energy-constrained and can transmit at a higher power level.

Random multiple access is performed by delaying the signal to transmit at each end-device as illustrated in Fig.5. The slot is first divided into N_s subslots such that $N_s = \frac{8192}{2^k}$, with 2^k the used spreading factor. For $k \leq 13$, the transmitter selects one (or several to increase the effective data rate) subslot called *access slot*. Within the subslot (gray area in Fig.5), transmission is delayed by a random number $d \in [0; 2^k - 1]$. Ingenu estimates that up to 1000 uplink users can be served in each slot.

Communication range. By using eq.15 and eq.16 with $CG = 3 \cdot k$, we can evaluate that for $B = 10^6$, and $k = 13$, the minimal received power is -145 dBm, which corresponds to 200 kms in free space. In [26], the range was estimated to 10 kms with Okumura-Hata model.

Interference Sensitivity. The random delay permits to shift the time of arrival of the different signals. As Gold codes have low auto-correlation, each arrival can be decoded as long as at most one end-devices selected the considered delay. The multiple access is thus a slotted ALOHA protocol.

6. Technologies comparison

We present in Table 2 some important features of the presented emerging technologies for IoT applications.

We can first note that the sub-GHz bands are favored. Indeed, the pathloss is smaller for low frequencies. Thus, higher ranges are obtained with the same emission power.

On the contrary, there is no consensus on the spectrum use. Indeed, one half of the technologies aim at minimizing the bandwidth (Narrow Band (NB) or Ultra NB) to reduce the probability of interference, while the second half spread the information on the available channel to take advantage of spectral diversity, and to benefit from adaptable data rate (with varying spreading factor).

In addition, the channel width is also varying. One may note that for Sigfox, 192 kHz is the observed channel, while the signals width is 100 Hz. For the other technologies, the two values coincide.

For a fair comparison, we compare the raw data rate, i.e., the actual bit rate transmitted on the medium independently of the nature of the traffic (preamble, data, code syndrome, ...). Numerical values were obtained from [25] for Ingenu, and from [20] for Weightless.

For Sigfox, (6) corresponds to the maximum number of users guarantying that there is no collision on a given user for at least 90% transmissions. In addition, for LoRaTM, this value is constrained by the number of available codes which is 6. Finally, for RPMA, Ingenu claims 1000 simultaneous users [25]. Note that such value mainly corresponds to high SF codes. Unfortunately, for Weightless, such information is neither estimated on public documents, nor computable with the available data.

(7) was evaluated by normalizing the overall data rate for all active users, with the spectrum occupancy. We can note that in spite of the wide range of raw data rate (1:300), the spectral efficiency is almost the same (1:2).

Furthermore, cell size is also reported, constraint for which SigFox and LoRaTM are the most efficient ones.

Besides, we can observe that all technologies allow two-ways communications, and that Weightless permits to select either one-way or two-way transmissions.

Last but not least, the receiver sensitivity to a frequency offset is discussed, as cheap (thus imprecise) oscillators are favored for dense deployment. Ingenu algorithm to compensate for this offset tolerates a deviation up to 10 ppm. In addition, the intrinsic behavior of LoRaTM permits to correctly transmit even with a 40ppm oscillator. Finally, SigFox has no limitation and can decode the signals as long as they are in the monitored band.

7. Summary and challenges

IoT is a relatively new market which may lead to the deployment of billions of connected devices. The associated networking paradigm is quite different from those of cellular systems. Despite that in some specific applications (e.g. for transmitting videos) MTDs may need high rate communications, it may be anticipated that a majority of these MTDs will require only a





						
			N	P	W	
(1) Band	868/915 MHz	868/915 MHz	Sub-GHz	Sub-GHz	470 – 790 MHz	2.4 GHz
ISM	Yes	Yes	Yes	Yes	No	Yes
(2) PHY	UNB	CSS	NB	NB	DSSS	RPMA
(3) Spreading factor	NA	$2^7 - 2^{12}$	NA	NA	$1 - 2^{10}$	$2^4 - 2^{13}$
(4) Typical channel bandwidth	192 kHz	500 – 125 kHz	NAD	12.5 kHz	6 – 8 MHz	1 MHz
(5) Raw rate (kbps)	0.1	27 – 0.37	30 – 100	0.2 – 100	1 – 10000	0.06 – 30
(6) Simultaneous active users in (4) for $OP = 10^{-1}$	100	6	NAD	NAD	NAD	1000
(7) Raw spectral efficiency (b/s.Hz)	0.05	0.12	NAD	NAD	NAD	0.1
(8) Range (km)	63	22	5	2	10	10 – 2
(9) Downlink	Yes	Yes	No	Yes	Yes	Yes
(10) Doppler sensitivity	Unconstrained	Up to 40 ppm	NAD	NAD	NAD	Up to 10 ppm

Table 2. Technologies summary and comparison (NA= Non Applicable; NAD= Non Available Data)

bursty access, with very small packets but under high QoS constraints such as high reliability or low latency. Therefore, two strategies are competing to absorb this new traffic. The first strategy relies on broadening the capabilities of cellular technologies especially with the preparation of the future 5G at horizon 2020. In this case, 5G intends to jointly manage M2M and mobile H2H traffics with appropriate dynamic sharing rules. The advantage of this strategy is to foster the development of IoT by avoiding the development of a specific access network and minimize infrastructure costs.

The second strategy relies on the deployment of a new network technology which may better comply with the specific features of IoT. This objective needs a cross-layer design of the PHY and MAC layers. The winning technology will be surely the one which will be able to maximize the manageable MTDs density, to minimize

the architecture cost and to ensure low latency and high reliability. This optimization will rely on joint PHY/MAC strategies which may permit to use original distributed coding approaches as proposed for instance in [28]. In addition the energy consumption of MTDs will play an important role for ensuring long life-time and small battery requirements.

For the best of our knowledge the competition is still open between the actual technologies.

To foster the analysis of these technologies and to evaluate the possible gains, a suitable general theoretical model is needed. Some works exist, such as the recent work proposed in [27] which exploits stochastic geometry and related models previously used for cellular networks. However, it is important to emphasize that the classical Shannon capacity established in asymptotic regime does not hold for short packets communications. If the underlying theoretical

model is quite clear for the uplink situation and relies on the Multiple Access Channel (MAC, see e.g. [29]), the specific bursty nature related to short packets imposes to develop a new framework allowing to establish the achievable region balancing reliability, latency, MTDs density and energy.

References

- [1] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5g," *Communications Magazine, IEEE*, vol. 52, pp. 74–80, February 2014.
- [2] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of m2m massive access in wireless cellular networks," *Digital Communications and Networks*, vol. 1, no. 1, pp. 1–19, 2015.
- [3] P. Jain, P. Hedman, and H. Zisimopoulos, "Machine type communications in 3gpp systems," *Communications Magazine, IEEE*, vol. 50, pp. 28–35, November 2012.
- [4] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of lte and lte-a suitable for m2m communications? a survey of alternatives," *Communications Surveys & Tutorials, IEEE*, vol. 16, no. 1, pp. 4–16, 2014.
- [5] WALKER, H.R. (2004) *Ultra narrow band modulation* in 2004 IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication, pp.19-22, 26-27 Apr 2004
- [6] SHIKAI ZHANG (2013) *Spectrum analyses of UNB modulation formats* in 2013 3rd International Conference on Consumer Electronics, Communications and Networks (CECNet), pp.594-597, 20-22 Nov. 2013
- [7] PHIL KARN (2013) *The VMSK Delusion* <http://www.ka9q.net/vmsk/>, accessed 2015-08-28
- [8] www.sigfox.com accessed 2015-09-14
- [9] www.sigfox.com/static/media/Files/Documentation/SIGFOX_Whitepaper.pdf accessed 2015-09-14
- [10] MINH-TIEN DO, CLAIRE GOURSAUD, AND JEAN-MARIE GORCE, *Interference Modelling and Analysis of Random FDMA scheme in Ultra Narrowband Networks*, in AICT 2014
- [11] MINH-TIEN DO, CLAIRE GOURSAUD, AND JEAN-MARIE GORCE, *On the benefits of random FDMA schemes in ultra narrow band networks*, in IEEE 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), pp. 672-677, 2014
- [12] www.lora-alliance.org/ accessed 2015-09-14
- [13] M. WINKLER, *Chirp signals for communications*, in IEEE WESCON Convention Record, vol. Pt. 7, 1962
- [14] A. SPRINGER, W. GUGLER, M. HUEMER, L. REINDL, C. C. W. RUPPEL, AND R. WEIGEL,, *Spread spectrum communications using chirp signals*, in IEEE Proc. Eurocomm, pp.166 - 170, 2000
- [15] SEMTECH CORPORATION, *AN1200.22 LoRa™ Modulation Basics*, May 2015, Rev2, available at www.semtech.com/images/datasheet/an1200.22.pdf accessed 2015-09-14
- [16] PATRICK FLANDRIN, *Time-Frequency/Time-Scale Analysis*, ISBN 0-12-259670-9, in Academic Press, San Diego, 1999
- [17] NANOSCALE LABS, *Communications system*, Patent US8406275, submitted 2010-03-09, published 2013-03-26
- [18] SEMTECH CORPORATION, *SX1272/73 - 860 MHz to 1020 MHz Low Power Long Range Transceiver Datasheet*, March 2015, rev3, available www.semtech.com/images/datasheet/sx1272.pdf accessed 2015-09-14
- [19] <http://www.semtech.com/apps/filedown/download.php?file=SX1272LoRaCalculatorSetup1%271.zip>, accessed 2015-09-10
- [20] www.weightless.org accessed 2015-09-14
- [21] www.neul.com/neul/ accessed 2015-09-14
- [22] www.nwave.io/ accessed 2015-09-14
- [23] www.ingenu.com/ accessed 2015-09-14
- [24] ON-RAMP WIRELESS, INC., *Random phase multiple access communication interface system and method*, Patent WO 2009117284 A3, submitted 2009-03-11, published 2009-12-17
- [25] ONRAMP WHITE PAPER *RPMA Technology for the Internet of Things*, 010-0064-00 Rev. B, May 11, 2015
- [26] PROJECT IEEE P802.15 WORKING GROUP FOR WIRELESS PERSONAL AREA NETWORKS (WPANs) *On-Ramp Wireless Dynamic Direct Sequence Spread Spectrum (D-SSSS) Proposal for 802.15.4g*, IEEE 802.15-09-0307-00-004g, May 2009
- [27] H. S. Dhillon, H. Huang, H. Viswanathan, R. Valenzuela, et al., "Fundamentals of throughput maximization with random arrivals for m2m communications," *Communications, IEEE Transactions on*, vol. 62, no. 11, pp. 4094–4109, 2014.
- [28] M. Shirvanimoghaddam, Y. Li, M. Dohler, B. Vucetic, and S. Feng, "Probabilistic rateless multiple access for machine-to-machine communication," *Wireless Communications, IEEE Transactions on*, 2015.
- [29] T.M. Cover, J.A. Thomas, "Elements of information theory." John Wiley & Sons, 2012.