A Wireless Sensor Network for Outdoor Environmental Monitoring

Andre Xavier

1Department of Electrical and Computer Engineering, Instituto Superior Tecnico, Universidade de Lisboa

With recent advances in wireless communication technology, new radio transceivers have surfaced with more range and less energy consumption. With them Low Power Wide Area Networks are becoming more frequent and less constrained. This thesis aims to propose a Low Power Wide Area Network deployment for environmental parameter sampling consisting of 576 nodes over a 36 square kilometer area.

In order to achieve this goal, wireless communication technologies and hardware for the sensing nodes were surveyed. LoRa was selected for the connectivity. A design for the sensing nodes is proposed consisting of the sensors, transceivers and battery. Medium Access Control protocols for LoRa technology were surveyed and a new protocol is proposed. A simulator was developed for estimating the performance of the protocol and of the proposed network.

This thesis focuses on the development of the design of the sensing nodes, the development of the simulator and the development of the protocol. The performance of the simulator is assessed and is shown to be in accordance with other existing simulators. The presented results from simulations show that the proposed protocol increases the packet extraction rate and, in some cases, increases the lifetime of the network for the proposed deployment when compared to the alternatives.

Index Terms—LoRa, MAC, LPWAN, Environmental Sensors

I. INTRODUCTION

With recent advances in different fields such as telecommunications, informatics and electronics [1] the Internet of Things and Ambient Intelligence paradigm has the potential to turn into an aspect of central relevance for people’s day to day life and business, as well as to support disruptive changes to society.

Supported on increasingly cheaper and versatile embedded computing platforms as well lower power consuming communication modules, computation is becoming ubiquitous and pervasive.

Applications for distributed sensor networks often stem from scientific, social or business needs such has habitat monitoring [2], drought monitoring [3], fire detection [4], precision farming [5], to list a few. However, the motivation behind this thesis comes from the belief, if it were possible to relax some of the constraints on WSNs, the ubiquity discussed above would be one step closer to the user. This might lead to scientific, social and business applications enhancing our experiences as human beings. Applications of WSNs such as creating finer (e.g. 250 meter) grid weather models and predictions or optimal routing of airborne delivery drones taking into account wind speed measurements. Both applications could be supported by a distributed sensor network of weather stations.

A. Proposed Work

In this thesis the proposed work is the project of a private long-range network of wireless nodes capable of sampling the requirements below, on a periodic time basis of five minutes, encompassing an area of 36 squared Kilometers, defined as a 6 by 6 Kilometer square. This area is approximately equivalent to Lisbon delimited by 2l Circular, Eixo Norte-Sul and the Tagus river, with Instituto Superior Tecnico in a relatively centered position ( Figure 2 ).

Considering a deployment in a grid fashion where nodes distance 250 meters from each other, this yields a network with a density of 16 nodes per squared kilometer, with a total of 576 nodes.

The environment related parameters to be sampled are:

- Temperature
- Atmospheric Pressure
- Relative Humidity
- Wind Speed
- Wind Direction
- Occurrence of Rain
- UV rating

II. BACKGROUND

When a Sensor Network is deployed, an observer means to monitor the behavior of some phenomena under some requirements. A sensor performs local sampling of the phenomenon and the network is in charge of gathering and/or reporting the sampled information to the observer, who may not be aware of the network infrastructure and its sensing
nodes. In [6] networks are evaluated according to their energy efficiency and system lifetime, Latency, Accuracy, Fault-tolerance and Scalability. It is suggested that, for each network, considering its requirements, the previous listing could change, since an evaluation of a network is naturally dependent on the application for which it was developed. Nevertheless, the topics listed are representative of what is generally expected of a WSN.

A. Communication technologies

In [7], Keith E. Nolan et.al explore state of the art solutions of Low Power Wide Area Networks for the Internet of things. Namely Sigfox and LoRa, both operating in licence-free frequency bands since the transition from analog to digital television, clearing the sub-1 GHz spectrum range. Both implemented resorting to star topologies, the two technologies claim ranges up to tens of kilometers from the end-stations to the gateways. In the case of SigFox ranges up to 500 kilometers have been reported, transmitting on 25 mW power. Both technologies are based on end stations - the sensing devices and gateways that relay information to the internet. This results in battery being depleted first in stations further away from gateway nodes.

1) Sigfox

Sigfox is a French company who owns an ultra-narrowband communications system based on antennas on towers and receives data in a star topology fashion, like cellular technology. Transmissions are in the 868MHz band, and data is uploaded in 12 bytes payloads and downloaded in 8 bytes per payload, with 26 bytes overhead in both cases. Only up to 140 messages may be uploaded daily and 4 downloaded. This means only 4 very important packages, such as ACKs can be received by the nodes each day.

Despite the lack of available acknowledge messages for each and every upload, each message is sent 3 times in 3 different frequencies to avoid multipath or noise or message loss for any other reason.

After discussing the project with the portuguese operator of SigFox, Narrownet, this certification would cost something like 4000 euros. The cost of certifying a custom developed module such as the one manufactured by TI would be around 26,000 euros, according to Narrownet. Furthermore, also according to Narrownet, the cost of the packet of 140 daily upload messages and 4 download messages would cost, for 10 end-station nodes, 2 euros per month. For 10,000 nodes, this cost would come down to 2 euros per year.

Regarding coverage Sigfox assures it on many countries. Portugal has been the fourth country with nation-wide coverage.

2) Wi-Fi

In an urban context it is reasonable to assume there are many Wi-Fi networks in operation, possibly able to support the desired deployment arrangement with virtually no infrastructure cost apart from the sensing nodes. Moreover, recently developed power-efficient Wi-Fi promising years of battery lifetime are a strong candidate for communicating data over IP from sensors to the network’s router. As sensors may turn off their transceivers to extend battery lifetime. A beacon is sent by the gateway of the network informing there are messages available for the node. The node periodically wakes up and finds out that there are messages for him, thanks to the beacon, initiating the process of receiving data from the gateway.

3) LoRa

In Semtech’s application note [8] some basic concepts of LoRa modulation are explained and the advantages of such a Physical Layer stack are emphasized.

Lora is a PHY layer implementation and it may interoperate with existing network architectures, while at the same time relying on these networks to provide its service, e.g. the Gateways relay received Lora packets using TCP/IP to a network server.

In information theory, the Shannon-Hartley theorem states that the maximum channel capacity \( C \) (bits per second) that can be transmitted over a communication channel of a specified bandwidth \( B \) (hertz) in the presence of noise, assuming a bounded transmission power is given by equation 1 below, in which \( S \) stands for the average received signal power over the bandwidth measured in watts, and \( N \) is the average power of the noise interference over the bandwidth also measured in watts:

\[
C = B \log_2(1 + S/N) \text{bits/sec.} \tag{1}
\]

Manipulating 1, taking into account that in spread spectrum applications \( S/N \ll 1 \), the proportion \( C/B \propto S/N \) emerges, which is equivalent to:

\[
N/S \propto B/C \tag{2}
\]

which shows that in order to transmit error free information in a channel of fixed noise-to-signal ratio, only the transmitted signal bandwidth need be increased, or equivalently, in order to cope with increasing noise on the channel one must only compensate the bandwidth accordingly.

Lora modulation’s bit rate is given by equation 3

\[
R_b = SF \times \left[ \frac{\frac{4}{2^CR}}{\frac{2^{SF}}{BW}} \right] \text{bits/sec} \tag{3}
\]

where \( SF \) stands for the spreading factor, ranging from 7 to 12, \( CR \) stands for the coding rate, ranging from 1 to 4. This coding rate follows from Lora’s inherent Forward Error Correction scheme that improves the robustness of the transmitted signals at the expense of a small overhead in the transmitted packet. This FEC scheme adds robustness in the presence of thermal noise, but the real achievement here is in the presence of bursts of interference.

There are also exist two operational modes - Header Mode and Low Data Rate Optimisation Mode.

A symbol’s duration \( T_{sym} = \frac{2^{SF}}{BW} \) is the time it takes to send \( 2^{SF} \) chips at the chip rate defined by the bandwidth.

A Lora Packet may have the following format:

<table>
<thead>
<tr>
<th>Preamble</th>
<th>Header + CRC</th>
<th>Payload</th>
<th>Payload CRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>nP symbols</td>
<td>Header mode ( CR = 4/8 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Common to all modem configurations is a sequence of preamble, whose duration is \( T_{\text{preamble}} = (nP + 4.25)T_{\text{sym}} \), with \( nP \) user defined. The number of symbols of a packet is given by equation 4 below

\[
N_{\text{sym}} = 8 + \max(\text{ceil}\left(\frac{8PL - 4SF + 44 - 20H}{4(SF - 2DE)}\right) \times (CR + 4), 0)
\]  

(4)

where:
- \( PL \) is the number of payload bytes
- \( SF \) is the spreading factor
- \( H = 0 \) when the Header mode is enabled and 1 when the Header is not part of the packet
- \( DE = 1 \) when the low data rate optimization is enabled
- \( CR \) is the coding rate from 1 to 4.

In order to get the time on air for a packet it is only needed to add the time on air of both the preamble and payload, by multiplying the number of symbols by the duration of a symbol \( T_{\text{sym}} \), as is detailed in equation 5.

\[
T_{\text{packet}} = \left[ nP + 4.25 + 8 + \max(\text{ceil}\left(\frac{8PL - 4SF + 44 - 20H}{4(SF - 2DE)}\right) \times (CR + 4), 0) \right] \times \frac{2SF}{BW}
\]  

(5)

<table>
<thead>
<tr>
<th>SF</th>
<th>CR = 1</th>
<th>CR = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>125</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>125</td>
<td>25</td>
</tr>
</tbody>
</table>

From Table I it is possible to extract the maximum capacity a single transceiver can handle. With a coding rate of one, the maximum number of transmitted packets per hour is 310773, with spreading factor 7 and bandwidth 500 KHz. It should also be noted that the maximum number of packets a Base Station can receive is upper bound by this value. However, the most common baseband processors encourage the use of bandwidths of 125 KHz, and in this case, the maximum capacity is 77693 packets per hour. Furthermore, the CR’s impact on airtime is also apparent. In fact, the increase in airtime introduced by changing CR from 1 to 4, for a 15 byte payload with SF 7 is of 15 milliseconds, which is approximately 33% of the original packets airtime. This value drops to 25% for SF 12.

A LoRa collision occurs when two packets arrive at a radio on the same frequency ( or within the frequency offsets mentioned above ) and have the same spreading factor. Once this has occurred, timing and reception power become determinant factors for the correct decoding of at least one of the packets, as shown in [9]. The header section of the packet is optional. In a packet with explicit header, 3 bytes are prepended to the payload. The purpose of this header is to introduce the payload to the receiver, by letting it know in advance the length and coding rate of the ensuing message as well as the presence of a CRC trailing the packet.

In [10] researchers perform several measurements on different scenarios to test the coverage of potential deployments. On all experiments, the Base Station was fixed and its antenna was 24 meters above sea level. The scenarios considered reception in a moving car and reception on a moving boat. The nodes were configured with spreading factor 12 and the bandwidth was set to 125 KHz. Their results showed, for the car scenario are presented in Table II.

<table>
<thead>
<tr>
<th>Range</th>
<th>Packet loss ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2 Km</td>
<td>12%</td>
</tr>
<tr>
<td>2-5 Km</td>
<td>15%</td>
</tr>
<tr>
<td>5-10 Km</td>
<td>33%</td>
</tr>
<tr>
<td>10-15 Km</td>
<td>74%</td>
</tr>
</tbody>
</table>

From these measurements, the authors also propose a Channel Attenuation Model, given by equation 6 where \( EPL \) stands for Expected Path Loss, \( n \) is the path loss exponent, \( B \) is the path loss found at the reference distance and \( d0 \) is a reference distance of 1 Km.

\[
EPL(d) = B + 10n\log_{10}(d/d0)
\]  

(6)

The results the researchers obtained allowed them to compute for the path loss exponent a value of 2.32 and for the reference path loss a value of 128.95 dBm.

Other important results can be extracted regarding the capacity of the Lora technology from [9]. In Figures 2 and 3 the results were produced with a LoRa network simulator, considering that random 20 bytes packets were sent every 16.7 minutes. The experiments go through a multitude of configuration parameters regarding node transmissions. These are summarized in Table III.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SN1</th>
<th>SN2</th>
<th>SN3</th>
<th>SN4</th>
<th>SN5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission Power</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>min. power</td>
</tr>
<tr>
<td>Spreading factor</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>min. airtime</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>125</td>
<td>500</td>
<td>125</td>
<td>min. airtime</td>
<td></td>
</tr>
<tr>
<td>CR</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Their findings show that dynamic settings for transmission parameters greatly increase the data extraction rates obtained when compared to the LoraWAN specifications, which can only support 64 nodes with the default transmission parameters, which are similar to the parameters SN1 in Table III.

4) Short Range Devices in the EU

From the regulation tables in [11], Lora is contemplated in the Annex 1 h (863-870MHz), where there are 7 different subbands, most of them limited to 25mW, subject to a Duty Cycle ≤ 0.1% or Listen Before Talk (LBT) + Adaptive
B. MAC protocols

1) LoraWAN

LoRaWAN is a MAC specification intended for wireless battery operated devices in a wide network, targeting requirements of the Internet of Things such as secure bi-directional communication, mobility and localization services. The topology is a star of stars where gateways forward messages to a network server. These gateways are usually powerful devices with powerful radios, listening on at least three channels and are capable of listening and decoding multiple concurrent transmissions.

Regarding the bands, it should be noted that there are only 3 common 125kHz channels for the 868 MHz band, 868.1, 868.3, 868.5 MHz, that must be supported by all devices and networks, and that all gateways must always be receiving on. Through these channels all devices may join the network. During the join procedure the network may instruct the devices to add channels to its set. They are all inside subband h1.4.

Lorawan specification in Europe may be summarized:

- Frequency band is 867-869 MHz
- 10 Channels, 8 of which are multi data rate from 250bps to 5.5kbps, one is a high data Rate LoRa channel at 11kbps and one FSK Channel at 50kbps.
- Bandwidths are 125/250kHz for uplink and 125kHz for downlink
- maximum transmission power is +14dBm
- Transmissions can only occur once every 16.7 minutes.

2) LoRaBlink

In reference [13] researchers argue that Lora transceivers may be useful in developing multi-hop bi-directional communication.

The proposed MAC, LoraBlink syncs time among nodes to define slotted channel access. Nodes transmit concurrently within slots, hoping at least one of the messages will be successfully received. Messages are distributed from the sink to nodes using flooding and from the nodes to the sink using directed flooding.

The network has a low density and low traffic volume and contains a limited number of nodes, assuming there is only one sink used for packet reception. The messages sent have a payload of 10 bytes, the transmission parameters are Spreading Factor 12, Bandwidth 125 KHz and transmission power of 17 dBm, which is a very power hungry configuration. The epoch length, meaning the period of beacon transmissions was set on 15 minutes, and each node randomly transmits one packet on one data slot each epoch. The reported maximum node lifetime was 2 years considering batteries with 5.4 mAh.

3) Symphony Link MAC Protocol

Symphony Link is a wireless solution for enterprise and industrial customers who need to securely connect their IoT devices to the cloud, built on the same physical layer as LoRaWAN. Symphony Link is a protocol developed by Link Labs and claims to merge the range of LoRa technology with higher levels of performance. For instance, the duty cycle limitation on LoRa technology imposed by ISM regulations is avoided in a Symphony Link solution with Frequency Hopping Listen Before Talk with adaptive frequency agility band. This
allows for applications where there is a need to send a lot of data at a time. It also allows them to introduce repeaters on a low power wide area network. This was not possible in LoRaWAN architecture because of Duty Cycle limits. The repeaters in their product range are less expensive than outdoor gateways so a network’s range can be expanded without a large infrastructure investment. It acknowledges all messages up and downlink.

C. Environment Monitoring Networks

In [17] the authors examine the Lora radio transmission technology and the LoraWAN network aiming to evaluate the feasibility of a network to transmit data from weather stations. They propose a specification for extending the range of the network making use of intermediate nodes with repeater functionality. A pure ALOHA scheme is considered, employing LoRaWAN class A device specification.

Nodes only have a normal transceiver, one channel only. The Base Station considered is Kerlink’s LoRa IoT Station, with 8 channels of capacity. The intermediate nodes repeating traffic are set on one of the LoRaWAN obligatory frequencies, listening on spreading factor 8. Between the nodes and the gateway, spreading factor 7 is fixed.

The data transmission frequency is set to 5 minutes, with a random factor of 30 seconds. The payload size transmitted by the nodes is 16 bytes, with 2 bytes for node ID’s, 2 bytes for frame counters, 2 bytes for temperature readings, 2 bytes for rain measurements, 2 bytes for Pressure measurements, 2 bytes for windspeed, 1 byte for wind direction, 1 byte for humidity and 2 byte CRC. This payload is contained in the LoRaWAN MAC frame, which adds at least 12 bytes to transmitted packet size. This yields a minimum total of 28 bytes per transmission.

In [2] an in-depth study of habitat monitoring via WSN’s is provided. The system’s requirements are presented and cover the hardware design of the nodes, the architecture of the network and the remote access to data and its management. A network of 32 nodes is deployed, sensing temperature, luminosity, barometric pressure, humidity, and also taking infrared measurements from the nesting places of seagulls.

In [4], an implementation with ten nodes with 802.15.4 compliant devices was developed, as well as a web interface for monitoring the data from the nodes. A scenario in which a fire may have been detected will trigger a variation in the sampling rate of the nodes, increasing the data throughput. The sampled data was temperature, humidity and luminosity, each with a size of 2 bytes in the payload. The total transmitted packet had 73 bytes, which represents a considerable overhead. Simulations were undertaken using Cooja, which is developed with java to emulate wireless sensor network deployments.

III. APPROACH/DESIGN

The nodes would be composed of a sensing module equipped with one or more sensors, a communication module for connectivity with the network server and a battery accumulator with possible energy harvesting technology. Taking into account the Communication subsection in the State-of-the-art section the best communication solutions for implementing wireless connectivity and transmitting the collected data are...
Wi-Fi and LoRaWan. This naturally leads to a network with end-stations equipped with one of the two communication technologies in end-stations, since together they offer a flexible solution, with practically no operating costs and low module costs, not too power hungry transmissions and relatively simple deployment.

The de facto standard for LoRa devices networking, the LoRaWAN network protocol, despite being flexible and open for different implementations, adds considerable overhead for each packet transmitted, as it is designed to suit a wide number of applications, and thus is found somewhat inefficient for a sensing application with such little data payloads.

A. End Station Design

In [14] researchers propose a cheap wind speed and direction measurement system, composed of an array of ultrasonic transmitters and receivers displayed in a cross fashion, 12 centimeters apart which can be operated on PWM outputs. This design works by measuring delays between emitted and received pulses. To measure pressure, temperature and humidity, the MS8607 implements the desired measurement, and is interfaceable via I2C. The rain measurement can be performed by means of detecting a connection to a resistor that in series with an open circuit designed like a comb that closes the circuit in the presence of water. UV sensing is performed by means of the ML8511 chip that outputs an analog voltage mappable to UV intensity. Regarding the communication modules to be employed, the considered Wi-Fi chip is MOD-WIFI-ESP8266 [16] which receives UART commands. As for the Lora connectivity the considered module is RN2483 since Semtech does not sell their transceivers SX1272/SX1276 to the general public. Microchip implemented this on an autonomous module, which handles voltage scaling internally and also implements UART communication. Its voltage can range from 2.1V to 3.6V. The alternative would be Hope RF’s RFM95 which is cheaper but not already certified, and as such would be unsuitable for a large scale deployment.

Regarding microcontroller units, the best choice for this project is to go with the ARM STM32L0 series since it achieves low power consumptions across its seven low power modes as well as during Run Mode by using dynamic voltage scaling. It also features some tamper detection features and many different manners for waking up.

To achieve low power consumption during sleeping periods, 3 High-Side Active High switches should be used: for switching the communications section of the board, for switching the sensor section, and finally for switching a voltage divider so as to allowing being able to measure the battery’s voltage without constantly draining current.

The minimum operating voltage for the ML8511 sensor is 2.7 Volts, for the MS8607 is 1.5 Volts, for the Lora Module RN2483 is 2.1 Volts and for the Wi-Fi module is 3.0 Volts. Thus the voltage regulator circuitry is dimensioned to output 3.0 Volts, regardless of the input voltage (so long as this is in the input range).

The proposed payload to be extracted should be as small as possible, considering that any increase in time one air

<table>
<thead>
<tr>
<th>Sampled Parameters</th>
<th>Resolution on the Payload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>unsigned 7 bits [-40,88]C</td>
</tr>
<tr>
<td>Pressure</td>
<td>signed 8 bits [758,1268]mBar, even numbers only</td>
</tr>
<tr>
<td>Humidity</td>
<td>unsigned 4 bits [0,16] maps to 0-100%</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>unsigned 4 bits, beaufort scale</td>
</tr>
<tr>
<td>Wind Direction</td>
<td>unsigned 3 bits</td>
</tr>
<tr>
<td>Rain</td>
<td>1 bit</td>
</tr>
<tr>
<td>UV</td>
<td>4 bits [0,15] mW/cm²</td>
</tr>
</tbody>
</table>

B. Base Stations

Base Stations or Sinks for LoRa Networks are usually massive concentrators capable of receiving on multiple frequencies and different spreading factors concurrently. The SX1301 is one of such chips, which, when implemented with two SX1257 is capable of demodulating eight concurrent non colliding packets. It listens on every spreading factor on the specified frequencies, but is only capable of demodulating concurrent messages if they have 125 KHz bandwidths.

A cheaper alternative is to employ a normal SX1272/SX1276 chip. This would require either fixing transmission parameters for the network (every transmission is on the highest SF with the highest transmission power), or scheduling transmissions to happen at fixed time slots, with parameters negotiated by the nodes. The second option promises a longer lifetime since for the first, higher SFs are required so as to assert that no message is lost because of lack of range or power. Furthermore, the first option will have more collisions, since higher spreading factors have higher times on air.

This alternative may seem like it lessens the versatility of the network by constraining the capacity and functionality (no more receptions with any spreading factor, and only one center frequency instead of eight) but in fact it allows for easier expansion of the network, since if more nodes are added to the network after the deployment, and capacity becomes scarce, another single channel chip may be added to the sink, effectively doubling the capacity. The MAC protocol must support informing the nodes on how many frequencies are being listened to, and on what spreading factors are packets supposed to be sent.

The Hardware architecture could consist of any platform supporting internet connectivity through Wi-Fi or cable solutions and a SPI interface for the radios.
C. Proposed MAC

To avoid high collision rates reported by the ALOHA implementations in the state of the art section, the MAC considered can reserve time slots for transmission for each node.

Bandwidth 125 is considered for every message, since this leads to less spectrum occupation. Also coding rates are fixed at the smallest possible to change airtime. This should, however, be subject of a more complex analysis depending on the deployments exhibited error ratios. If more redundancy is needed, this coding rate should be increased.

Epochs and frames are defined, where each Epoch will be divided into frames, and one or more frames are assigned to each node at the time they connect to the Base Station, depending on the length of the Epoch. The node, after being acknowledged, receives a reading of the reception strength of his connection packet, and calculates the link margin to the base station. After this, another message is sent, this time specifying what the spreading factor will be for his reserved time slots. This is required since the proposed MAC ideally makes use of normal single channel radios which are much cheaper than the SX1301 multi channel solution, but may achieve similar data extraction rates.

The spreading factors allowed each epoch are specified via a mask in which each of the six least significant bits correspond to a spreading factor being considered (if the bit is 1) or not.

The protocol supports base stations having more than one LoRa radio as referred. This effectively multiplies its capacity and allows for scheduling enough time frames to serve all nodes’ transmissions. In this case, the only change is that there will be the same number of epochs as there are radios, all concurrently happening and the reserved frames for each radio are shifted by the number of nodes divided by the number of radios. So, for example a deployment of 100 nodes with a base station with 4 radios will begin with a frame for node 1, node 25, node 50 and node 75, one in each epoch. The nodes then select the transmission frame that is allocated to them and that happens sooner. It should be noted that the reservation of frames introduces a certain delay between the production and collection of data.

The protocol also supports multi channel receivers such as those based on the SX1301 chip in which case frame reservation is dropped and the access scheme is slotted aloha. In this later case, nodes can employ LBT + AFA when transmitting and as such higher data sampling rates can be achieved, since the regulations are more lenient for nodes politely accessing the spectrum.

1) Security

Nodes and Base Stations share a common Network Key. The trailing 4 bytes of this network key are used for authenticating node’s messages, specifically login requests and data messages. This is done resorting to Blake2s [15] algorithm which computes a 1 Byte message authentication code over the contents of each message. Blake2s was selected for its small footprint, requiring only 168 bytes to be stored in node memory and for the speed required for computation. Also a determining factor was that Blake2s can output message authentication codes as small as one byte, which is good for keeping the message length small.

The Base Stations’ messages however, not being constrained by energy requirements, employ AES-128 encryption in Galois/Counter mode, effectively encrypting and authenticating messages. This encryption mode is featured on the STM32 specific libraries, although not implemented for the L0, only for the F0 series. This means that its employability would require porting the code from one platform to the other, but everything suggests it is possible. So for acknowledgments, ADR messages and beacons, base stations prepend the nonce used in generating the cipher and the generated authentication tag that was computed over the contents of the message.

Frame counters were discarded since they attempt to mitigate replay attacks, but in doing so open the chance for an attacker to invalidate all messages from a node. To do this, as reported in [18], all an attacker has to do is wait for a counter overflow, record that maximum value, and then, since when an overflow is reached LORAWAN specifies that the frame counter returns to 0, an attacker can simply retransmit the packet with the largest frame counter, invalidating all messages of the node, until its counter resets again.

To sum up, the structure of the packets to be sent follows the format in Table V for the Beacon, Acknowledgment, and ADR messages, and the format in Table VI for the Login, SF disclosure and Data messages.

<table>
<thead>
<tr>
<th>Table V</th>
<th>Structure of encrypted messages of the proposed MAC protocol.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonce</td>
</tr>
<tr>
<td></td>
<td>16 Bytes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table VI</th>
<th>Structure of authenticated messages of the proposed MAC protocol.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Payload</td>
</tr>
<tr>
<td></td>
<td>—</td>
</tr>
</tbody>
</table>

2) Adaptive Data Rates

The proposed MAC aims to maximize network lifetime by selecting optimal transmission parameters for each node dynamically, meaning that if any node’s location changes, its transmission parameters should change accordingly. For this an Adaptive Data Rate scheme different from the proposed in the LoRaWAN specification is presented where nodes locally compute the link’s margin and select the best transmission parameters for themselves.

This is achieved resorting to messages sent by the Base Station disclosing its sensitivity thresholds for each spreading factor for the 125KHz bandwidth setting. This ADR message is disclosed prior to every Epoch start. In order to effectively compute the existing link’s margin, nodes are also required to know the reception power of their messages at the Base Station. To achieve this, every connection request is acknowledged and in this acknowledgment the received packet power is present. Nodes retransmit connection requests randomly every few epochs, in order to receive updated reception powers
at the base station in order to adjust for potential changes in the link’s quality.

By default, these messages for ADR computation are sent with the maximum transmission power, of 14 dBm. As such, when nodes receive the acknowledgment of their connection requests, considering they already have the base stations target reception sensitivities, they can compute the margin and decrease the transmission power for all the frame’s spreading factor slots if it is possible. When a transmission is to be performed, all the slots are considered and a routine sorts the slot’s desirability resorting to an heuristic that takes into account a data message’s airtime and transmission power, as output by the ADR process, for each slot’s spreading factor.

3) Connections

The connection procedure at the nodes starts with a discovery epoch which consists of listening to the Default Network Link for a certain period of time, which is hardcoded at the time of the Nodes’ programming. After the discovery epoch, nodes choose to connect to the base station whose beacon they received with most power. This is indicative of the quality of the link.

After this, they take into account the beacon’s contents and select a random frame for transmitting a login request, with 14 dBm transmission power and spreading factor 12, for best reception chances at the base station. In the end of the epoch the nodes which sent connection requests expect an acknowledgment packet in which figures their ID, their frame, and the reception power of their connection request at the base station. If this occurs, nodes then compute the link’s margin and the transmission power setting for each spreading factor. They then select the configuration that yields less energy consumption for transmission and produce a SF disclose message to inform the base station of their choice.

Nodes remain disconnected if their ID is not in the acknowledgment packet or if they fail to catch the next beacon from their base station. In either case, they repeat this connection process.

D. Proposed Network

The proposed Network follows, for the node deployment, the strategy outlined in the Proposed Work section, which is to deploy a node every 250 meters, in a grid fashion, totalling 576 nodes over an area of 16 $Km^2$. The single base station is in the center of the area. The nodes are transmitting with a period of 5 minutes between data samples. If the worst transmission parameters were set as default for all nodes, the time on air each hour would be of 14.895 seconds, which is well within the regulated 36 seconds per hour.

The epoch duration is set to one hour since smaller epochs introduce more energy consumption from receiving downlinks. All spreading factors are considered. This yields a frame duration of, taking into account the worst transmission time for an uplink message of 991.232 milliseconds (calculated with equation 5) plus the slot margin of 250 milliseconds, 1241.232 milliseconds. Since there are 576 nodes, if only one single channel radio is used at the base station, the total duration required for all nodes to send a message is approximately 715 seconds. Since each node must transmit a message every 300 seconds, the base station will require at least 3 single channel radios in order to serve time frames with a periodicity such as to meet the time requirements of all nodes. All nodes will have a frame reserved for them every 238.31 seconds, representing the worst case scenario delay for data collection.

IV. Simulator

In order to access the Proposed Mac’s applicability for the network considered in the proposed work section, a network simulator was developed since it is unfeasible to test it with a deployment. When simulating, relevance is given to figures such as collision numbers and Data Extraction Rate, energy consumption and network lifetime. These metrics can then be compared and the simulation’s logs can be checked to assess in which conditions were packets lost or collided, and check for bottlenecks regarding network lifetime.

To develop a simulator that supports testing a MAC protocol, the work in [9] and lorasim were leveraged to build a lora radio module for SimPy, giving more detail to the specifics of LoRa Radios and providing an abstraction away from the specifics of the physical medium.

SimPy is a python framework for process-based discrete-event simulation. Processes are python generator methods and can be used to model active components or time dependant processes. Events are the means for process interaction with the environment and between themselves. Following this, nodes and sinks are processes, and their interactions are also modelled as processes, triggering events such as transmissions and receptions.

The simulation consists of a main script called with parameters specifying the duration of the simulation, a file with the topology of the network and link related parameters, the node sampling period to be considered and the desired logging level.

A simulation follows the following steps:

**Algorithm 1: Simulation flow**

1. load simulation configuration file
2. set path-loss model parameters
3. make stations according to the configuration file
4. run until all nodes are connected
5. reset counters for all nodes and base stations
6. run for the specified duration of simulation
7. print report

The MAC protocol is chosen at the time of starting a simulation by passing it as a parameter to the main script. Depending on the choice, the script imports the corresponding class from the corresponding file, and instantiates all base stations and nodes accordingly.

The simulator developed for the work on this dissertation is slow when compared to lorasim. This is natural given the increase in complexity of the behaviour of the nodes, but is mostly due to the increase in the complexity implementation of the transmissions in this simulator. This is done, as referred, by poking nodes in the vicinity and giving them a copy of the
packet. In contrast, in lorasim, only base stations get packets from nodes. In fact, no node ever considers receiving a packet. This is fine when all we are trying to model are the medium’s capacity for transmissions with different parameters as is the case in [9]. However, for the work at hand, nodes needed to consider receiving messages, otherwise the Listen Before Talk mechanism, and message acknowledgments or other downlink messages would not be simulatable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>lorasim</th>
<th>Proposed simulator</th>
</tr>
</thead>
<tbody>
<tr>
<td>DER</td>
<td>60.7</td>
<td>60.62</td>
</tr>
<tr>
<td>Collisions</td>
<td>214</td>
<td>215</td>
</tr>
<tr>
<td>Sent</td>
<td>544</td>
<td>544</td>
</tr>
<tr>
<td>Energy for transmissions (J)</td>
<td>123</td>
<td>107</td>
</tr>
<tr>
<td>Runtime ( seconds )</td>
<td>0.38</td>
<td>0.6</td>
</tr>
<tr>
<td>Peak memory ( KB )</td>
<td>46436</td>
<td>59281</td>
</tr>
</tbody>
</table>

**TABLE VII**

**RESULT COMPARISON BETWEEN lorasim AND THE PROPOSED SIMULATOR. DER STANDS FOR DATA EXTRACTION RATE AND IS CALCULATED AS THE RATIO OF RECEIVED PACKETS OVER SENT PACKETS.**

From the table above it is possible to extract some comparisons between the two simulators and to notice some discrepancies regarding energy consumptions. For the values of colliding packets, sent packets and data extraction rates, results are very similar. Regarding runtime and memory usage, the proposed simulator performs considerably worse.

The difference in energy is negligible. However, inspecting lorasim’s script, it is possible to find that, for the transmission power setting of 14 dBm, the consumption is 44 milliAmperes, while in the proposed simulator this is calculated resorting to a linear regression done over the values disclosed in the RN2483 module’s datasheet and its value is 38.2. The ratio of the discrepancy in energy (1.1518) is practically the same as the ratio between the two different values of current - hence power - consumption (1.1495).

**V. SIMULATIONS**

The simulations were performed considering the topology described in the proposed work section.

The simulation’s size was set to 10 Km by 10 Km, with the base station centered at position 5000,5000. The nodes were set from 2000, 2000 to 7750, 7750 in a grid fashion as proposed.

The parameters for path-loss were the parameters calculated in [10] for the car experiment in a city environment which were: path-loss exponent of 2.32, reference distance of 1 Km, and reference path loss of 128.95.

The base station in the topology.txt file was set to a single spreading factor sink, first with two radios and then with three radios as suggested in the Design Section.

In the third experiment, 4 base stations were considered, instead of the single base station taken into account thus far. Each base station had 3 single channel radios, although this could have been relaxed, since each of them would get approximately one quarter of the nodes, and as such could provide enough capacity for their nodes. Their locations were 3000,3000; 3000, 7000; 7000, 3000; 7000, 7000.

The duration for the discovery epoch had to be increased in the node’s mac run methods.

The LoraWAN mock considered was not an implementation of the protocol, but a modification of the previous simulation. Base stations are listening on all 8 channels a normal LoraWAN concentrator would possess, capable of decoding messages on every spreading factor.

The data packets will have prepended to them 12 Byte headers, which represent the minimum header size for LoraWAN networks. Furthermore, after each transmission there are two received windows, the first with the duration required to detect a preamble of a packet with the same transmission configurations as the packet sent, and the second with a duration for the largest spreading factor’s preamble duration. This mimicks LoraWAN type B devices.

In the first experiment no ADR is used, and only Spreading Factor 12 is allowed. This is similar to out of the box LoraWAN implementations. The ADR functionality was used in the second experiment. An assumption was made that the result would be the same as the proposed ADR method in LORAWAN’s specification. In the third experiment all Spreading Factors are allowed, and ADR is employed.

**VI. DISCUSSION**

Experiment 1 from the table IX shows that the nodes produced more messages than they could transmit since there is far less sent data than there is pushed data, indicating that not enough time frames were being reserved for each nodes.

Experiment 2 simulated the proposed MAC with enough radios to support the capacity needs from the nodes. The collision’s number is in both simulations as expected, 0. The Data Extraction Rate for experiment 3 however was not 100% due to the already discussed fact that messages are not sent immediately after being produced.

Experiment 3 shows the results of a deployment with more than one base station, showing the simulator’s functionality and reporting a significant increase in lifetime and very little delay between data production and transmission. The former
is due to the fact that nodes are on average closer to the base station and as such the link margin is greater, and transmission power and airtime is lower, while the later is due to the fact that the capacity of the deployed base stations largely exceeds the capacity required by the nodes.

Regarding lifetime, comparing Experiment 2 with the LoRaWAN’s mock results, it is possible to extract that the proposed MAC and the proposed network architecture exhibit a shorter lifetime, when considering the use of the multi SF radios at the base station with all spreading factors allowed and with ADR functionality (Experiment 3). This is due to the fact that the SX1301 chip has considerably better sensitivity figures, and as such when it is employed at the base station, the link’s margin is better. This translates into lower transmission power requirements for the nodes. However, without the ADR functionality, considering that the nodes must employ the transmission parameters with the lowest energy consumption performance, the results are much worse than the proposed MAC’s.

Despite this the proposed MAC, although with a more complex solution, is capable of handling the requirements of the proposed Network while requiring considerably less spectrum bandwidth than the solutions employing SX1301 concentrators, such as LoraWAN. These require 1600 KHz bandwidth, while the proposed MAC handles the same capacity with only 437.5 KHz of spectrum.

VII. CONCLUSION

A weather monitoring system capable of monitoring temperature, atmospheric pressure, relative humidity, wind speed, wind direction, occurrence of rain and UV rating, spread over a large area with a high density of nodes, tailored for low-cost and long duration of the deployment was intended for development in this project.

Considering the hardware surveyed and the proposed design, the communication technologies considered, and the simulations performed, the proposed network was concluded to fulfill its main objectives, despite introducing a delay in data collection and sacrificing some lifetime when compared to other existing solutions for overall cost of the deployment.

It is also concluded that the simulator developed for this thesis was fit for assessing the network’s and the proposed mac’s performance, although its implementation falls short in terms of the models employed when compared with existing simulators, and also in terms of both code quality and functionality given that some experiments require changes in the code (LBT, ADR, multiple base stations) so as to run smoothly.

ACKNOWLEDGMENT

I must begin by thanking my parents for their support and guidance through University and during these theses. From them I learned integrity and perseverance, character and self-control and self-awareness, to put up with discomfort and unwelcome truths. I would not be finishing this degree nor this dissertation if not for them.

To the TLMOTO team, I thank the last third of my experience in University, I thank the opportunity to learn beyond the classroom, from teamwork to endurance, technical and social skills and also independence. Thank you Diogo Neves, Francisco Jesus, Andr Miguel and Gerardo Lima for your companionship during this stage. To the Department of Electrical and Computer Engineering at IST I must thank for the lessons and for such interesting and demanding courses.

Finally I must thank my girlfriend Marta for her kind support during the good and the bad moments, for motivation during the long hours of work and for helping see the nearing end of this project.

REFERENCES

[5] Sherine M. Abd El-kader, Basma M. Mohammad El-Basioni, Precision farming solution in Egypt using the wireless sensor network technology in Electronics Research Institute, Computers and Systems Dept., Cairo, Egypt Received 13 December 2012; revised 18 June 2013; accepted 26 June 2013