

Towards Energy Efficient LPWANs through Learning-based Multi-hop Routing

1st Sergio Barrachina-Muñoz
Wireless Networking
Universitat Pompeu Fabra
Barcelona, Spain
sergio.barrachina@upf.edu

2nd Toni Adame
Network Tech. and Strategies
Universitat Pompeu Fabra
Barcelona, Spain
toni.adame@upf.edu

3rd Albert Bel
Network Tech. and Strategies
Universitat Pompeu Fabra
Barcelona, Spain
albert.bel@upf.edu

4th Boris Bellalta
Wireless Networking
Universitat Pompeu Fabra
Barcelona, Spain
boris.bellalta@upf.edu

Abstract—Low-power wide area networks (LPWANs) have been identified as one of the top emerging wireless technologies due to their autonomy and wide range of applications. Yet, the limited energy resources of battery-powered sensor nodes is a top constraint, especially in single-hop topologies, where nodes located far from the base station must conduct uplink (UL) communications in high power levels. On this point, multi-hop routings in the UL are starting to gain attention due to their capability of reducing energy consumption by enabling transmissions to closer hops. Nonetheless, *a priori* identifying energy efficient multi-hop routings is not trivial due to the unpredictable factors affecting the communication links in large LPWAN areas. In this paper, we propose *epsilon multi-hop* (EMH), a simple reinforcement learning (RL) algorithm based on epsilon-greedy to enable reliable and low consumption LPWAN multi-hop topologies. Results from a real testbed show that multi-hop topologies based on EMH achieve significant energy savings with respect to the default single-hop approach, which are accentuated as the network operation progresses.

Index Terms—LPWAN, energy, routing, uplink, reinforcement learning, MAB

I. INTRODUCTION

Low-power wide area networks (LPWANs) are wireless networks conceived for providing extensive communication ranges, reducing the energy consumption of end devices (STAs), and diminishing the operational cost with respect to traditional cellular networks. As a result, they are envisioned to be a key communication technology for a vast variety of Internet of Things (IoT) applications. LPWANs reach such a low power operation and extensive coverage range by using the sub-1 GHz unlicensed, industrial, scientific and medical (ISM) frequency band, high processing gains, narrow bandwidths, and by sporadically transmitting packets at low data rates, which allows achieving very low sensitivities.

Most LPWAN solutions like LoRaWAN or SIGFOX are based on star topologies, where STAs directly transmit to the base station or gateway (GW), making them to heavily rely on transceiver's capabilities (e.g., available transmission powers, antenna gains or data rates). While this approach

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence (MDM-2015-0502), by the Spanish Government through the project TEC2016-79510-P, and by the Catalan Government through the projects 2017SGR 1188 and 2017SGR 1739. The work done by S. Barrachina-Muñoz is supported by a FI grant from the Generalitat de Catalunya.

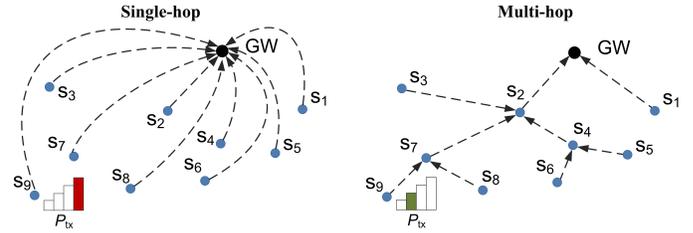


Fig. 1: Single-hop vs. multi-hop topology. Note the power level reduction of STA s_9 when multi-hop is adopted.

facilitates network designs and provides a robust centralized management, it usually leads to a shortening of the lifetime of STAs located far from the GW since they are most likely required to transmit using high power levels. In addition, the inclusion of STAs with limited transmission power is greatly compromised due to this range constraint. Moreover, long-range single-hop topologies lead to interference and packet collisions among uncoordinated devices, which importantly affects the reliability and scalability of networks with a large number of nodes [1], [2].

Although multi-hop energy savings have shown its potential for wireless sensor networks [3], there is scarce literature on LPWANs on this topic [4], [5]. These works reveal that multi-hop topologies can importantly extend LPWAN's lifetime by providing significant energy-savings to the STAs located farthest from the GW. In this regard, authors in [6] present a novel LPWAN protocol stack enabling multi-hop communication in the uplink (HARE) that is able to achieve important energy savings in a real testbed. One of the key challenges of multi-hop routing, however, is how to find both energy efficient and reliable links in a distributed way. Such difficulty results from the lack of global information required to make proper decisions. Instead, by exploiting the traditional centralized management of LPWANs, where a global view of the network is available at the GW (e.g., number of nodes, packet error rate, delay, etc.), the system is able to foresee if multi-hop routing strategies can outperform single-hop, and if so, reconfigure the network accordingly. Fig. 1 shows the single-hop and a possible multi-hop topology on the same network deployment.

Indeed, it is difficult and fuzzy to determine *a priori* the

most energy efficient multi-hop routing for a given LPWAN. The main reason is that performance depends on too many factors such as the operation modes of the nodes (at microprocessor and radio modules) and the network deployment (e.g., location of the nodes, running applications, or environmental conditions). Hence, while deterministic rule-based solutions are not accurate for identifying energy efficient routings, machine learning approaches are appealing for such a task; especially when considering the scalability issues involved in huge LPWANs. Notwithstanding, a learning-based routing algorithm, if not properly set, could also entail significant extra consumption, and be consequently counterproductive since highly energy consuming topologies may be occasionally adopted [5]. Therefore, the trade-off between the energy savings achieved with the most efficient routing and the cost of learning should not be overlooked.

As in any exploration/exploitation problem, reinforcement learning (RL) methods are appropriate due to their ability to cope online with such a tradeoff: *i*) selecting the routing providing the best-known results (i.e., exploiting), or *ii*) broadening the gathered knowledge about the performance of unexplored routings (i.e., exploring). In this paper, we present results from a real testbed for assessing the energy savings achieved with *epsilon multi-hop* (EMH), a Multi-Armed Bandits (MABs) ϵ -greedy-based algorithm for learning energy efficient uplink routings in LPWANs. Namely, we assess the performance of EMH in an LPWAN operating under the HARE protocol stack, probing significant energy savings with respect to the single-hop approach.

II. OVERVIEW OF LPWANs

A. Communication challenges

Most emergent LPWAN technologies only rely on the capabilities of low communication layers to achieve large single-hop coverage areas, disregarding multi-hop schemes already existing in other wireless networks. Star topologies are therefore predominant in LPWANs, where one central element (i.e., the GW) is the single responsible for configuring and managing the whole network. While simple and robust, this approach does not seem the most appropriate to face the following challenges [7]:

- **Scalability and reliability:** since the propagation ranges are much higher, LPWANs cause interference at a much larger scale, creating bottlenecks in highly dense scenarios. Besides, most existing channel access mechanisms of LPWAN technologies resort to the use of ALOHA, which does not require much coordination between the AP and STAs [8]. However, as the number of devices attempting to access the channel increases, so does the collision probability.
- **Flexibility:** current LPWANs are deployed, operated and managed in a completely uncoordinated manner, hindering new application purposes and/or possible network reconfigurations/upgrades.
- **Energy efficiency:** battery-powered LPWAN devices are currently lacking strategies beyond the PHY layer to

extend their lifetimes, such as adaptive power control or advanced low duty cycle techniques combined with grouping strategies.

- **Quality of Service (QoS):** since channel access is still randomized to some extent, no real guarantees in terms of QoS can be offered in LPWANs.

B. Current LPWAN technologies

While numerous LPWAN technologies have emerged in the last years, only some of them are able to combine long-range links and heterogeneous network topologies:

- **HARE**, unlike other LPWAN technologies, is able to adopt uplink multi-hop communications without affecting data transmission reliability and achieving a notable energy consumption reduction [6]. Multi-hop paths also involve intermediate STAs, which must be awake during the periodic association stages to execute the own distance-vector routing protocol.
- **LoRaBlink** incorporates multi-hop bi-directional communication enabling sensing and actuation [9]. Messages from nodes to the sink are directly flooded.
- **D7AP** networks consist of gateways and endpoints, and can optionally contain sub-controllers, thus also enabling tree topologies [10]. While gateways are permanently listening for packets, sub-controllers are allowed to sleep and are mainly used to relay packets. Lastly, endpoints can transmit asynchronously and wake up periodically to listen to possible incoming data.
- **IEEE 802.11ah** includes in its specification a two-hop mode by using relays [11]. Consequently, when transmitting to a closer relay instead to the AP, STAs reduce the transmission power level and use higher data rates, thus also shortening the transmission time, and consequently, the energy consumption.

III. EMH: LEARNING-BASED UL MULTI-HOP ROUTING

In this section, we argue the need of learning proper routings for obtaining significant energy savings in real LPWANs and present the novel EMH approach. To do so, let us define some terms regarding the UL routing for the sake of facilitating further explanation.

Any routing in a network can be represented by a vector \vec{r} of size $n - 1$, where n is the number of nodes (GW and STAs) in the network, and whose element $\vec{r}(s)$ is the network address of the parent of STA s . The GW address is always set to 0 for simplicity. In Fig. 1 it is shown an UL multi-hop routing example with the corresponding routing vector $\vec{r} = (0, 0, 2, 2, 4, 4, 2, 7, 7)$. For instance, the parent of s_5 is s_4 , i.e., $\vec{r}(5) = 4$. The set $\mathcal{R} = \{\vec{r}\}$ is composed of all the possible UL routings that can be given in the network. Like in common LPWANs, we assume that every STA is capable (if required) to successfully communicate with the GW in a single-hop manner.

A. The need of learning

Due to the fact that the network performance depends on multiple factors such as the deployment of nodes, protocol stack, hardware, or environment conditions, it is hard and fuzzy to predefine proper UL routings beforehand. Therefore, even though experimentally tuning routing parameters can importantly enhance the energy savings in distributed approaches [6], in most of the cases its performance is sub-optimal. Moreover, such tuning approach comes at the cost of flexibility since the resulting configuration is deeply tied to the targeted scenario.

In this regard, the problem of identifying the optimal routing can be modeled as a finite-horizon multi-armed bandit (MAB) problem due to its exploration/exploitation nature (i.e., the trade-off between exploring new knowledge or exploiting gathered knowledge) and the need of maximizing the lifetime of battery constrained STAs. While over-exploring routings prevents from maximizing the short-term reward in terms of energy savings, exploiting only partial knowledge prevents from identifying the optimal routing and maximizing the long-term reward accordingly. RL for wireless network has been previously covered in a number of papers like [12]–[14].

We propose EMH as a centralized learning-based routing approach that enables the GW to stochastically compute the routing table according to a MAB's ϵ -greedy procedure. The goal of EMH is to minimize the energy consumption of the bottleneck STA (i.e., the STA that consumes the most) by exploring different UL routings. While simple to implement, this algorithm effectively serves to evaluate the impact of the different explored routings on the network's lifetime in a real-time manner.

B. The EMH approach

The well-known ϵ -greedy method sets the randomness in action selection through a parameter ϵ that determines the probability of exploring a new action (already explored or not) rather than exploiting a previously explored one [15]. The simplicity of ϵ -greedy together with the fact that no memorization of exploration specific data (e.g., counters or confidence bounds) is required [16] are its main advantages with respect to other MAB methods. However, a substantial disadvantage of ϵ -greedy is the complexity of determining the optimal initial value and the updating function of ϵ .

In the EMH algorithm, a principal variation is included with respect to the regular ϵ -greedy method: each action is explored just once.¹ Namely, since LPWAN deployments are characteristically static, we assume that the average energy consumed by the STAs in a certain UL routing does not significantly vary over time. Thus, we are able to set a deterministic (experimental) payoff to every routing by exploring it just once.

With respect to the testbed presented below in this work, the reward or payoff (p) provided by any possible action or routing

Algorithm 1: Implementation of EMH in HARE. $\mathcal{U}(\mathcal{A}')$ is a distribution that randomly chooses any unexplored routing in \mathcal{A}' uniformly at random.

```

1 Input:
2  $K$  #Number of averaging cycles
3 Initialize:
4  $t := 0$ 
5  $\hat{p}(\vec{r}) := 0$  for  $\forall r \in \mathcal{R}$ 
6  $\epsilon := \epsilon_0$ 
7 while active do
8   #New iteration
9    $\vec{\gamma}_t \leftarrow \text{estimate\_rssi}()$  #RSSI from each STA
10   $\mathcal{A}_t \leftarrow \{\vec{r} \in \mathcal{R} \mid \vec{\gamma}_t(s) \geq \vec{\gamma}_t(s') \text{ for } \forall(s, s')\}$  #Constraint
11   $\mathcal{A}'_t \leftarrow \{\vec{r} \in \mathcal{A} \mid \hat{p}(\vec{r}) = 0\}$  #Unexplored routings
12   $\vec{r}_t \begin{cases} \text{Explore: } \vec{r} \sim \mathcal{U}(\mathcal{A}'_t), & \text{with prob. } \epsilon \\ \text{Exploit: } \operatorname{argmax}_{\vec{r} \in (\mathcal{A}_t \setminus \mathcal{A}'_t)} \hat{p}(\vec{r}), & \text{otherwise} \end{cases}$ 
13   $\bar{e}_b(\vec{r}_t) \leftarrow \max_s \frac{1}{K} \sum_{k=1}^K e_{s,k}(\vec{r}_t)$ 
14   $\hat{p}(\vec{r}_t) \leftarrow 1/\bar{e}_b(\vec{r}_t)$ 
15   $\epsilon \leftarrow \epsilon_0/\sqrt{t}$ 
16   $t \leftarrow t + 1$ 
17 end

```

in \mathcal{R} may vary according to the channel condition (e.g., people crossing by or changing weather conditions). Therefore, in order to ensure enough accuracy of the payoff estimate (\hat{p}), we average the reward of each explored routing by measuring its corresponding energy consumption K times. The pseudocode containing the main steps of EMH is depicted in Algorithm 1.

1) *Estimating the single-hop RSSI:* the number of existing UL routings for networks of n nodes is given by Cayley's formula.² Specifically, $|\mathcal{R}| = n^{(n-2)}$. Hence, $|\mathcal{R}|$ grows extremely rapidly for large networks. In this regard, exploring routings without any predefined discrimination criteria could have a negative impact on the EMH performance. For instance, if considering a network deployment like the one shown in Fig. 1, an alternative routing with $\vec{r}(1) = 9$ would be most likely sub-optimal since link s_1 - s_9 probably suffers worst channel conditions than link s_1 -GW. Consequently, s_1 will most likely suffer from higher energy consumption with respect to the original routing with $\vec{r}(1) = 0$.

In order to avoid exploring such *naive* routings, we apply a received signal strength indication (RSSI) constraint stating that any children-parent link (s, s') can only be performed if the RSSI received at the GW from the children is less or equal than the one received from the parent, i.e., $\vec{\gamma}(s) \leq \vec{\gamma}(s')$. Thus, we are able to significantly reduce the number of possible routings from \mathcal{R} to $\mathcal{A} \subseteq \mathcal{R}$ by excluding those that do not comply with the *RSSI constraint*. Note that we consider RSSI values rather than distances since channel conditions also depend on other deployment factors.

Accordingly, a preliminary step is conducted in each ϵ -greedy iteration t . Basically, the GW estimates $\vec{\gamma}_t$ in the association phase, when STAs transmit directly to the GW (i.e., in a single-hop manner) asking for being associated to the network. With such a metric, the algorithm is able to discern

¹Note that exploring one routing takes several energy consumption measures since they are averaged for improving the estimation accuracy.

²In graph theory, Cayley's formula states that for every positive integer n , the number of trees on n labeled vertices is $n^{(n-2)}$.

what routings comply with the *RSSI constraint* and identify the set \mathcal{A} . EMH re-estimates $\bar{\gamma}_t$ in each iteration just in case the channel conditions have significantly changed since the network initialization.

2) *Exploring or exploiting*: once $\bar{\gamma}_t$ is estimated, the algorithm decides whether to explore an unexplored routing or exploiting the best-known one according to the ϵ parameter. Specifically, with probability $(1 - \epsilon)$ the algorithm picks the most energy efficient routing from the set of explored ones. That is, the already explored routing providing the highest estimated reward \hat{p} . Instead, with probability ϵ , the algorithm picks uniformly at random an unexplored routing in \mathcal{A}_t^i .

3) *Estimating the payoff*: after routing \bar{r}_t is selected, the GW starts collecting the energy consumption measures of the STAs during K HARE operation cycles. An important trade-off exists in this regard: the larger K , the more accurate the estimated payoffs, but the longer the time to identify the most energy efficient routing. The latter also entails the risk of exploring high consuming routings during larger periods of time. After every cycle k , STA s estimates the energy consumed during the cycle ($e_{s,k}$), generates a payload including the corresponding value, aggregates the payloads received from its children (if any), and transmits the packet/s with the payload/s to its parent. Once the K -cycles data collection phase finishes, the GW is able to determine the average energy consumed by every STA. Then, it sets the payoff estimate \hat{p} corresponding to the current routing \bar{r}_t as the inverse of the bottleneck node's average consumption.

4) *Updating the ϵ value*: once the payoff corresponding to the current routing \bar{r}_t is estimated, the algorithm updates ϵ . In our experiments, we use a time-dependent exploration rate $\epsilon_t = \epsilon_0/\sqrt{t}$ with $\epsilon_0 = 1$ as suggested in [17]. This ϵ setting entails substantially exploring in early stages and frequently exploiting afterwards, which is convenient for avoiding payoff local minimums. After updating the ϵ value, a new iteration begins with the single-hop RSSI estimation.

IV. EVALUATION

In this section, we evaluate the performance in terms of energy savings of the single-hop (SH) and EMH approaches. We first describe the testbed used for conducting the experiments along with the STA's energy consumption model. Then, we compare the performance of the aforementioned approaches.

A. Testbed

1) *Deployment*: the performance evaluation of EMH and SH was performed in an indoor testbed located in one office building from Universitat Pompeu Fabra facilities. 9 Zolertia RE-Mote development boards/nodes [18] acting as STAs were deployed throughout the offices and the main corridor, maintaining their location for all the experiments performed (see Fig. 2).³ Another Zolertia RE-Mote played the role of GW and was connected to a PC for logs generation. All devices ran Contiki 3.0 OS [19] as operating system and HARE

³More details such as the generated logs of the experiment are available at https://github.com/sergiobarra/data_repos/tree/master/barrachina2018towards.

TABLE I: Current values of the Zolertia RE-Mote platform at the different operational states.

	Operational state	Current
Microprocessor ARM Cortex-M3	Processing (CPU)	$I_{\text{CPU}} = 13 \text{ mA}$
	Low power mode (LPM)	$I_{\text{LPM}} = 0.4 \text{ }\mu\text{A}$
Radio Module TI CC1200 868 MHz 2-GFSK, 50 kbps	Receiving (RX)	$I_{\text{RX}} = 19 \text{ mA}$
	Transmitting (TX)	$I_{\text{TX}} = 39 - 61 \text{ mA}$
	Sleeping (SL)	$I_{\text{SL}} = 0.12 \text{ }\mu\text{A}$

as wireless communication protocol stack like the testbed from [6].⁴ The selected radio duty cycle (RDC) sublayer was X-MAC, which defines sleeping periods for receivers and strobed preambles for transmitters [20]. The largest single-hop distance from the GW to STA #9 was 45 meters. All operational tests were conducted considering no mobility. All STAs were powered by an 800 mAh battery except the GW, which was permanently powered by the PC. STAs transmitted packets of 43 bytes every 2-minute cycle in a time division multiple access (TDMA) basis for group of contenders. In addition, the GW broadcast the routing at each iteration period so the STAs were able to identify their next-hop (or parent) and keep it until a new iteration was started.

Although LPWANs are characteristically composed of a large number of STAs located in outdoor scenarios, the presented testbed is sufficient to conduct a proof of concept providing significant results.⁵ In fact, since RSSI levels perceived by the GW are the main parameters used by the ϵ -greedy approach, the actual position of the STAs and the channel conditions are always mapped to such parameters. That is, EMH is transparent to the actual LPWAN deployment.

2) *Energy estimation*: the total energy (e) consumed by an STA is employed by two main elements: the microprocessor (e_μ) and the radio power module (e_r). Specifically, $e_\mu = V_{\text{DD}}(t_{\text{CPU}}I_{\text{CPU}} + t_{\text{LPM}}I_{\text{LPM}})$ and $e_r = V_{\text{DD}}(t_{\text{RX}}I_{\text{RX}} + t_{\text{TX}}I_{\text{TX}} + t_{\text{SL}}I_{\text{SL}})$, where V_{DD} is the supply voltage. The duration and current consumption corresponding to the operational states of the microprocessor and the radio module are t and I , respectively. Table I lists these states and the values of current consumption corresponding to the Zolertia RE-Mote. Notice that I_{TX} value grows according to higher P_{TX} transmission power levels (with a P_{TX} operational range going from -16 to 14 dBm). We use the `energest()` function from Contiki to estimate the time an STA spends in each of the possible operational states for $K = 10$ averaging measures.

B. Results

In order to assess the energy efficiency of SH and EMH, we use two main metrics: the cycle bottleneck energy in iteration t , i.e., $e_b(t)$, and the cumulated bottleneck energy until t , i.e., $\mathcal{E}(t)$. While the former refers to the energy consumed by the STA that has consumed the most in iteration t , the latter refers to the cumulated energy consumed by the STA that has

⁴To the best of our knowledge, HARE is the only well tested LPWAN protocol stack specifically designed for UL multi-hop communications.

⁵Note that larger LPWAN deployments have not been considered since they are expensive and hard to monitor. Nonetheless, the HARE protocol stack operation was already validated in outdoor environments [6].

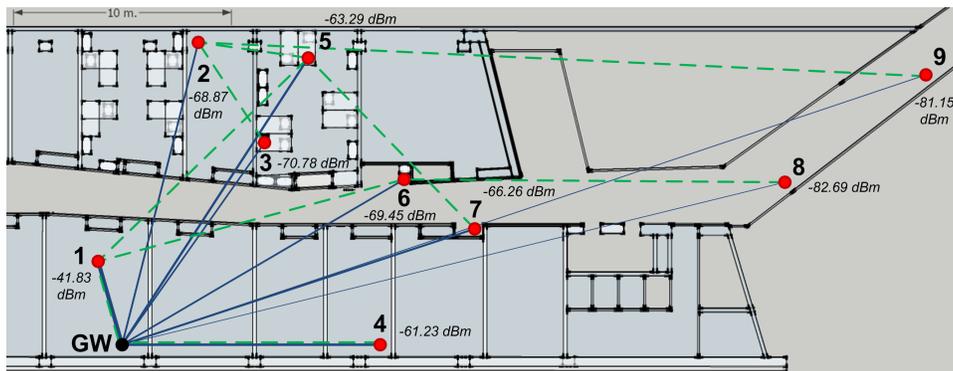


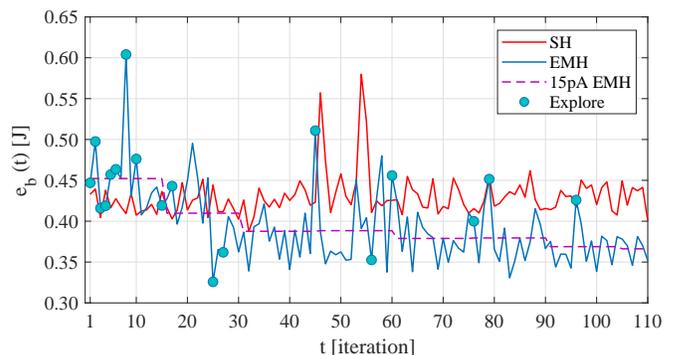
Fig. 2: Testbed deployment. The weights of the SH edges correspond to the RSSI level perceived by the GW in the first association phase. The most energy efficient routing explored by EMH, $\vec{r} = (0, 5, 2, 0, 1, 1, 5, 6, 2)$, is drawn in dashed lines.

historically consumed the most since iteration 1. Therefore, the metric $e_b(t)$ serves to assess the performance in terms of energy efficiency of the routing being applied in iteration t and assigning its corresponding reward. Instead, $\mathcal{E}(t)$ allows us to estimate the lifetime of the network since it is directly related to the remaining energy in the STA that has historically consumed the most. Note that, regardless of the considered packet transmission frequency (1 packet every 2 minutes cycle in this setup), a similar lifetime value in terms of iterations would be obtained since STAs consume very little when being in sleeping mode ($0.12 \mu\text{A}$). That is why we represent time in the x-axis of the plots in Fig. 3 in iteration units.

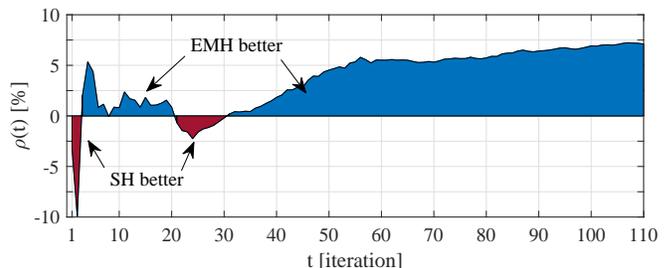
1) *Cycle bottleneck energy*: while in EMH unexplored routings are stochastically picked in exploration iterations according to probability ϵ , in SH the same routing is used throughout all the experiment, i.e., $\vec{r}_t = \vec{0}, \forall t$. Accordingly, as shown in Fig. 3a, the cycle bottleneck tends to decrease as the experiment evolves when implementing EMH.

Regarding the payoff of each action, we can see significant variability for the same routing both in SH and EMH. The main cause is the dynamic nature of the indoor communication channel, which affects the routing performance. Specifically, once a communication link (s, s') is assigned between a pair of nodes s and s' , if channels conditions are not appropriate, several transmissions may be required to successfully deliver a packet due to noise and interference. In this regard, retransmissions entail important extra energy consumption in HARE networks due to the fact that both the transmitter and receiver must wake up again for retrying the communication. Such a phenomenon is more frequent in SH because of its inherent higher collision probability. Besides, in this type of routing, the bottleneck STA is normally located far from the GW because of its lower signal-to-interference-plus-noise ratio (SINR).

Nonetheless, by gathering K energy measures per each routing, the GW is capable of estimating the corresponding payoff with sufficient accuracy to decide whether the routing is efficient or not. In fact, the tendency curve plotted in Fig. 3a shows the considerable energy reduction of the bottleneck STA with respect to SH, and its trend to slowly decay as more efficient routings are explored. We believe that in outdoor



a Average bottleneck energy for SH and MH. The curve 15pA-EMH refers to the average value corresponding to 15 consecutive measures of e_b .



b Saving ratio evolution. We highlight the periods when SH and EMH more energy-efficient in red and blue, respectively.

Fig. 3: Performance of SH vs. EMH.

scenarios, since channel conditions are less dynamic, the K value could be decreased while achieving better accuracy, thus outperforming the energy savings of the presented testbed.

In Fig. 2 it is shown the most energy efficient routing that the LPWAN explored during $T = 110$ iterations. We note a clear multi-hop-like topology where packets are transmitted to intermediate STAs, which entails higher SINR values and corresponding reliability. Besides, the staggered wakeup pattern of HARE in multi-hop topologies allows reducing the channel contention and packet losses due to interference accordingly.

However, since most of the energy consumed by the STAs is

due to the operation in RX state, parent nodes tend to consume more energy as they need to wait for and decode packets from children. Hence, a balanced routing as the presented in this proof of concept is required. Besides, the dynamics and interrelations among STAs and the channel make it very difficult to determine beforehand whether a routing is energy efficient or not. In fact, in our experiments, we noticed that some routings following multi-hop approaches were clearly not energy efficient because, even though the average node consumption was low, one of the STAs consumed a lot compared to the rest. That is why learning is critical, especially for LPWANs with a huge number of STAs, where lots of different routings can be potentially established.

2) *Historic bottleneck energy*: the metric the metric that best maps the lifetime of the network is $\mathcal{E}(t)$ because of its direct relation with the remaining battery energy of the bottleneck STA. That is, any routing approach can be assessed in terms of energy saving by measuring its corresponding $\mathcal{E}(t)$, which is defined by

$$\mathcal{E}(t) = \max_s \left(\sum_{t'=1}^t \frac{1}{K} \sum_{k=1}^K e_{s,k}(\vec{r}_{t'}) \right).$$

In order to compare the SH and EMH routing approaches, we show in Fig. 3b the saving ratio between the energies consumed by their historic bottlenecks at iteration t , i.e.,

$$\rho(t) = \frac{\mathcal{E}_{SH}(t) - \mathcal{E}_{EMH}(t)}{\mathcal{E}_{SH}(t)}.$$

At the beginning of the experiments, due to the small amount of iterations performed and the frequent explorations, the routing heavily influences the historic bottleneck of EMH, and the saving ratio ρ fluctuates accordingly. Instead, when the LPWAN is running for about 30 iterations, we note a more stationary behavior, where EMH clearly outperforms SH in terms of energy saving. Specifically, we achieve about 7 % of saving in 110 iterations, which keeps growing over time.

V. CONCLUSIONS

Lowering energy consumption is critical for LPWANs due to their aim of supporting applications based on unattended, battery-powered devices. In this regard, multi-hop routings in the UL are starting to gain attention in the field. However, it is hazardous and sometimes counterproductive to predefine static routings prior to the deployment of an LPWAN.

In this paper we have proposed EMH, a centralized reinforcement learning (RL) algorithm for finding energy efficient routings in an exploration/exploitation approach. That is, while the network is normally operating, unexplored routings are stochastically chosen and assessed according to the bottleneck energy payoff function. Results from a HARE testbed with real LPWAN devices show that EMH achieves important energy savings with respect to single-hop topologies.

Finally, we envision that the use of centralized learning-based multi-hop routing will result in high energy savings in massive LPWANs (with up to thousands STAs) for two main reasons. On the one hand, multi-hop approaches are able to

reduce the single-hop bottleneck energy consumed by those STAs located far from the GW, which are more likely to suffer from low SINR and other medium access issues like hidden and exposed node problems. On the other hand, with simple learning-based routing algorithms like EMH, we are able to find energy efficient routings that diminish the contention among STAs and build more reliable communication links.

REFERENCES

- [1] D. Bankov, E. Khorov, and A. Lyakhov. On the limits of lorawan channel access. In *Engineering and Telecommunication (EnT), 2016 International Conference on*, pages 10–14. IEEE, 2016.
- [2] Orestis G. and U. Raza. Low power wide area network analysis: Can LoRa scale? *IEEE Wireless Communications Letters*, 6(2):162–165, 2017.
- [3] M. Kakitani, G. Brante, R. Souza, and A. Munaretto. Comparing the energy efficiency of single-hop, multi-hop and incremental decode-and-forward in multi-relay wireless sensor networks. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pages 970–974. IEEE, 2011.
- [4] S. Barrachina-Muñoz, B. Bellalta, T. Adame, and A. Bel. Multi-hop communication in the uplink for LPWANs. *Computer Networks*, 2017.
- [5] S. Barrachina-Muñoz and B. Bellalta. Learning optimal routing for the uplink in LPWANs using similarity-enhanced epsilon-greedy. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on*, pages 1–5. IEEE, 2017.
- [6] T. Adame Vázquez, S. Barrachina-Muñoz, B. Bellalta, and A. Bel. HARE: supporting efficient uplink multi-hop communications in self-organizing LPWANs. *Sensors*, 18(1):115, 2018.
- [7] U. Raza, P. Kulkarni, and M. Sooriyabandara. Low power wide area networks: An overview. *IEEE Communications Surveys & Tutorials*, 19(2):855–873, 2017.
- [8] A. Laya, C.s Kalalas, F. Vazquez-Gallego, L. Alonso, and J. Alonso-Zarate. Goodbye, ALOHA! *IEEE Access*, 4:2029–2044, 2016.
- [9] Martin Bor, John Edward Vidler, and Utz Roedig. LoRa for the Internet of Things. 2016.
- [10] Wael Ayoub, Fabienne Nouvel, Abed Ellatif Samhat, Jean-Christophe Prevotet, and Mohamad Mroue. Overview and measurement of mobility in DASH7. In *2018 25th International Conference on Telecommunications (ICT)*, pages 532–536. IEEE, 2018.
- [11] T. Adame, A. Bel, B. Bellalta, J. Barcelo, and M. Oliver. IEEE 802.11 ah: the WiFi approach for M2M communications. *Wireless Communications, IEEE*, 21(6):144–152, 2014.
- [12] K. Yau, P. Komisarczuk, and P. Teal. Reinforcement learning for context awareness and intelligence in wireless networks: Review, new features and open issues. *Journal of Network and Computer Applications*, 35(1):253–267, 2012.
- [13] F. Wilhelmi, C. Cano, G. Neu, B. Bellalta, A. Jonsson, and S. Barrachina-Muñoz. Collaborative spatial reuse in wireless networks via selfish multi-armed bandits. *arXiv preprint arXiv:1710.11403*, 2017.
- [14] F. Wilhelmi, B. Bellalta, C. Cano, and A. Jonsson. Implications of decentralized Q-learning resource allocation in wireless networks. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on*, pages 1–5. IEEE, 2017.
- [15] C. Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge England, 1989.
- [16] M. Tokic and G. Palm. Value-Difference Based Exploration: Adaptive Control between Epsilon-Greedy and Softmax. In *KI*, pages 335–346. Springer, 2011.
- [17] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.
- [18] A. Lignan. Zolertia RE-Mote platform. Technical report, Zolertia, 2016. Available online: <https://github.com/Zolertia/Resources/raw/master/RE-Mote/Hardware/Revision>(accessed 09/09/2018).
- [19] A. Dunkels, B. Gronvall, and T. Voigt. Contiki-a lightweight and flexible operating system for tiny networked sensors. In *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, pages 455–462. IEEE, 2004.
- [20] M. Buettner, G. Yee, E. Anderson, and R. Han. X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 307–320. ACM, 2006.