

Load Balancing and Performance Optimization in wM-Bus Smart Meter Networks

Andrea Pimpinella, Alessandro E. C. Redondi, Matteo Cesana
DEIB, Politecnico di Milano
Milan, Italy
Email: {name.surname}@polimi.it

Abstract—Smart meters networks are rapidly becoming a reality in many developed countries. In this paper, we focus on the optimization of a network of smart meters operated by the wM-Bus protocol, which is the de-facto metering standard in Europe. In such a scenario, data concentrators receive data from smart meters using the wM-Bus protocol and relay it to a central server using a legacy mobile cellular backhauling technology such as GSM/GPRS. Due to the massive amount of data produced by meters installed in urban scenarios and the association-less nature of the wM-Bus protocol, data concentrators may be overloaded with many duplicate measurement packets, causing capacity problems on the backhauling links and computational overload at the central server. To solve these issues, we propose a data-driven optimization framework to populate forwarding whitelists at each data concentrator so that (i) load is balanced among the different concentrators and (ii) the overall performance of the network in terms of packet reception rate and received signal strength are maximized. We also propose a heuristic algorithm to generate near optimal forwarding whitelists in acceptable computing time. Extensive experiments are performed on a real scenario consisting of a city wide gas meter network deployed in northern Italy. Results show that the proposed heuristic is able to produce whitelists that reduce the average backhauling traffic as much as 80%, with a corresponding network quality within 4% of the one computed by the optimal solution.

I. INTRODUCTION

The last few years have witnessed a rapid growth of systems and technologies related to the Internet of Things, in particular for what concerns applications related to Smart Cities. In this context, the field of Smart Metering is emerging as one of the most concrete examples of how connected things may bring improvements to the everyday life of both citizens and municipality administrators. As a matter of fact, smart meters roll-outs plans are already active in several countries. To mention a few significative examples, Germany's electricity grid operator will install about 7.5 million smart energy meters across Germany between 2018 and 2020 as part of the transition from centralized energy generation to distributed energy resources [1]. Italy is also following this trend: the Italian Authority for Energy, Gas and Water has recently opted for a mandatory roll-out according to which services operators will provide nearly 40 million energy customers and 22 million gas and water customers with a smart meter network over the next few years.

Smart meters have unique requirements in terms of bandwidth, energy consumption and costs. On the one hand, they do not require high data rate as the number of measurements has

very coarse time granularity (e.g. 2/4 measurements per day). On the other hand, smart meters are generally battery operated and required by national regulations to have a minimum of 10-15 years of autonomy. At the same time, their cost should be limited in order to make it convenient for services operators to deploy them in place of the traditional meters. These requirements clearly have an impact on the choice of the communication technology to be used for smart meters. Solutions based on GSM/GPRS, where the meters are directly able to communicate their measurements to a central server have the drawback of being costly and power eager. A solution which respects the aforementioned requirements and has become the de-facto standard for many smart meter deployments in Europe is the wireless M-Bus (wM-Bus) protocol.

In the reference architecture of a city-wide wM-Bus network, depicted in Figure 1, hundreds of thousands of meters communicate their measurements to a small number (tens or hundreds) of data concentrators (DC) which are deployed in order to provide network coverage. Due to the harsh propagation conditions of urban environments, data concentrators are generally deployed in elevated positions (e.g. light poles or other structures) and use GSM/GPRS as backhauling technology for relaying the meter data to a central server, as other wireless or wired backhaul technologies with higher throughputs are either too expensive or impractical to set up. Due to the association-less nature of the wM-Bus protocol, data transmitted by one meter can be received at multiple data concentrators. This redundancy has a twofold harmful effect on the entire system: (i) on the uplink, it reduces the capacity of the already limited GSM/GPRS backhauling links and it causes computational overload at the DCs and at the central server; (ii) on the downlink, it creates an ambiguity about which DC should deliver system commands to each meter (e.g. increase/decrease measurement rate, close the main gas/water valve for insolvent customers).

In principle, a careful network planning could avoid the redundancy caused by duplicate packets: this could be obtained by minimizing the overlap between the coverage areas of different DCs or with frequency reuse techniques. However such planning strategies are seldom used, because both DCs and meters are deployed incrementally where needed rather than in a global single deployment process. For this reason, we focus in this paper on the creation of *whitelists* to be used on each DC for deciding which meter should be managed (by

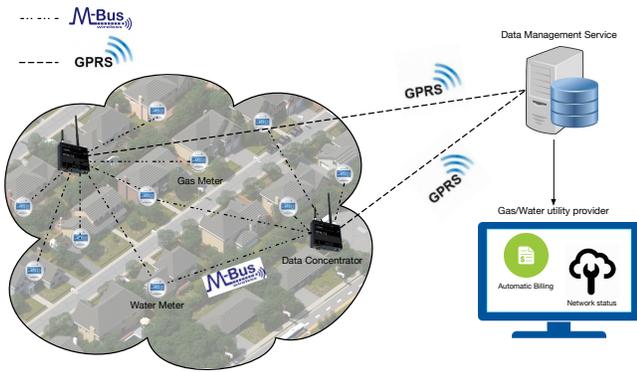


Fig. 1: Reference architecture: smart meters of different utilities (gas/water) communicate their measurements to multiple data concentrators using the wM-Bus protocol. Data concentrators relay such information to a central data management service through GPRS communication for automatic billings and management procedures

forwarding its data or sending commands to it) or not. We propose an optimization framework to create whitelists with the following features:

- 1) they balance as much as possible the amount of meters that each DC manages;
- 2) they provide the best associations between meters and DCs, in terms of link quality;
- 3) they can be easily updated in order to cope with the installation of new meters or with changes in the RF propagation environment;

The problem is tackled using a data-driven approach to obtain information on the current propagation conditions between meters and DCs. Such data is then fed into an optimization problem that produces balanced whitelists, also maximizing the link quality between meters and DCs. A heuristic is also proposed to provide quick updates even on big instances of the problem, such as the ones arising from real deployments in urban scenarios.

The remainder of this paper is organized as follows: Section II briefly summarize the wM-Bus protocol; Section III introduces the optimization framework, gives the ILP formulation and focuses on the proposed heuristic algorithm. Results obtained on a real smart meter deployment are shown in Section IV. Finally, Section V reports on the related literature and Section VI concludes the paper.

II. WIRELESS M-BUS PROTOCOL

The wM-Bus protocol was introduced in 2005 as a European standard (EN 13757-4:2005), specifying the communication primitives between utility meters and data loggers at the physical and data link layer. The presence of a European standard to regulate how such communications should be performed allows to create an ecosystem where utility providers, meter manufacturers and data management companies can coexist to create better services for the end users as well as business

opportunities. The current version of the standard (EN 13757-4:2013) specifies several modes of operation for smart meters. Each mode is characterized by a specific carrier frequency (around 434MHz, 868MHz or 169MHz), data rate, duty cycle and whether it supports unidirectional or bidirectional communication¹. Table I summarizes the available communication modes. The choice of the particular mode to use is left to each country regulation: Italy and France, two of the main smart meters markets in Europe, use narrowband mode N at 169 MHz. The advantages of using a narrowband communication consist mainly in the limited noise resulting at the receiver, which greatly increases the available link budget and allows to decode frames even at a sensitivity as low as -115 dBm. The reduction in data rate compared to modes operating at higher carrier frequencies is not critical since the amount of data to be transmitted by each meter is very low. At the same time, using a lower carrier frequency allows for a greater transmission range, which can reach 1Km even in urban scenarios [2]. At the data link layer, the standard specifies two different frame formats (A and B): for uplink communications from meters to DCs, both formats provide an 8 byte sender address field, but due to the association-less nature of the protocol, no destination address field is contemplated². A frame transmitted by a meter can be therefore accepted by more than one DC, and relayed on the backhauling infrastructure more than once. Therefore, without proper filtering mechanisms implemented at DCs, the efficiency of the entire network can be greatly affected. In the following sections we target precisely this issue.

III. OPTIMIZATION FRAMEWORK

A. Integer programming formulation

Keeping in mind the reference architecture of Figure 1, we consider a network of C data concentrators and M meters uniformly deployed over a urban area. Let \mathcal{C} be the set of the concentrators while \mathcal{M} the set of meters. We define the *Range Set* \mathcal{R}_c for the c -th concentrator as the set of in-range meters, containing all those meters from which concentrator c has received at least one frame in a specified time window. Similarly, let \mathcal{R}_m for the m -th meter be the set of in-range concentrators. Let $x_{m,c}$ be a binary association variable which is equal to 1 if the m -th meter is associated with the c -th concentrator and 0 otherwise. The whitelist of the c -th concentrator therefore contains all those meters m for which $x_{m,c} = 1$. Note that, by definition, $x_{m,c} = 0$ if $m \notin \mathcal{R}_c$, that is a meter cannot be in the whitelist of a concentrator if it is not in its Range Set. Finally, let $q_{m,c}$ a measure of the link quality existing between meter m and concentrator c . Note that $q_{m,c} = 0$ if $m \notin \mathcal{R}_c$. Under these definitions, we

¹Each mode is named with a letter and a number. The letter specifies the mode and the number specifies unidirectional (=1) or bidirectional (=2) communication

²The EN 13757-4 standard actually specifies an optional manual association between a meter and a DC through a three-way handshake process. However, no specifications are provided on how the DC should be selected or what happens in case multiple DCs receive the association request.

TABLE I: wM-Bus communication modes

Mode	Frequency	Max. Duty Cycle	Description
S (Stationary)	868 MHz	0.02 - 1%	Meters transmit few samples per day
T (Frequent Tx)	868 MHz	0.1 - 1%	Meters transmit short frames every few seconds
R (Frequent Rx)	868 MHz	1%	Meters listen every few seconds for reception of a wake up message and starts transmission. Used for readout of several meters at once.
C (Compact)	868 MHz	0.1%	Optimized version of mode T
N (Narrowband)	169 MHz	10%	Long range, narrowband system
F (Frequent Tx/Rx)	433 MHz	10%	Frequent bidirectional communication

propose a nonlinear integer programming (NIP) formulation for obtaining balanced whitelists while maximizing the overall quality of the wireless links:

$$\max_{x_{m,c}} J = \alpha \cdot \sum_{c \in \mathcal{C}} \log(L_c) + \beta \cdot \sum_{m \in \mathcal{M}} \sum_{c \in \mathcal{C}} x_{m,c} \cdot q_{m,c} \quad (1)$$

$$\text{s.t.} \quad \sum_{c \in \mathcal{C}} x_{m,c} = 1 \quad \forall m \in \mathcal{M} \quad (2)$$

$$L_c = \sum_{m \in \mathcal{M}} x_{m,c} \quad \forall c \in \mathcal{C} \quad (3)$$

$$x_{m,c} = 0 \quad \forall c \in \mathcal{C}, m \notin R_c \quad (4)$$

The objective function (1) is a weighted sum of two contributions. The first term aims at balancing the amount of meters L_c in each concentrator's whitelist. Balancedness is attained through the use of the logarithmic function, which is known to provide fairness and thus it is a common choice in load balancing problems [3], [4]. The second term of the objective function maximizes the total quality of the wireless links resulting after the whitelist creations. Moreover, the objective function can be shaped according to a user's preference through setting proper values for parameters α and β . Constraint (2) ensures that one meter is associated to the whitelist of only one concentrator, constraint (3) defines the load of each concentrator as the sum of associated meters and finally constraint (4) enforces that a meter can be inserted in the whitelist of a concentrator only if it is in that concentrator Range Set.

B. Heuristic algorithm

The optimization problem (1)-(4) can be considered a Generalized Proportional Fairness (GPF) problem, whose solution is known to be NP-Hard [3]–[5]. At the same time, meters and gateways are often deployed incrementally rather than in a single solution, thus requiring to often recompute the aforementioned whitelists. Therefore, it is important to provide a way for quickly obtaining a good solution to the optimization problem, even though non optimal. Here we introduce a heuristic algorithm able to do so. The basic idea of the heuristic is to separately optimize the two terms of the objective function, taking first the perspective of the DCs and secondly that of the meters, alternating between the two until a stop condition is reached.

The heuristic starts by creating an initial solution \mathbf{S}_0 , obtained by sorting the DCs in ascending order of the cardinality of the corresponding \mathcal{R}_c and then assigning meters to DCs in a round robin fashion. At each turn of the round robin schedule,

the c -th concentrator gets assigned the meter among the non-assigned ones for which $q_{m,c}$ is maximized. Note that \mathbf{S}_0 is balanced by construction, as each concentrator is assigned the same number of meters (but one).

The heuristic then runs for *max_iter* iterations, each one consisting of two phases. In the first phase, meters are sorted in descending order of the associated *Maximum Metric Gain* G_m , defined as it follows:

$$G_m = \left| \sum_{c \in \mathcal{C}} x_{m,c} \cdot q_{m,c} - q_{m,c_m} \right| \quad (5)$$

$$\text{where} \quad (6)$$

$$c_m = \operatorname{argmax}_{c \in \mathcal{R}_m} q_{m,c} \quad (7)$$

As one can see, G_m captures the maximum improvement of the link quality of a meter if it is removed from its current whitelist and inserted in the whitelist of the best concentrator quality-wise, c_m . The algorithm then proceeds in performing such exchanges one at a time. However, while each exchange improves the second term of the objective function (1), it does so without taking balancedness (i.e., the first term of eq. (1)) into account. Therefore, to avoid potential heavy load unbalances among DCs, after each exchange the algorithm evaluates the term $\sum_{c \in \mathcal{C}} \log(L_c)$ and compares it with its initial value L_0 , according to \mathbf{S}_0 . Meter exchanges are interrupted if the decrease in the sum of the logarithms is greater than a predefined threshold ξ_{\max} , that is when:

$$\frac{L_0 - \sum_{c \in \mathcal{C}} \log(L_c)}{L_0} > \xi_{\max}. \quad (8)$$

When this happens, all meters that have been moved are inserted into a *Blocking List*: such meters will be prevented to be exchanged again in future iterations. In the second phase, the heuristic re-balances the loads through redistributing the non-blocked meters to concentrators with the same rationale used in producing the initial solution \mathbf{S}_0 . In details, concentrators are sorted in ascending order of $|\mathcal{R}_c|$ (this time without considering the meters in the Blocking List) and non-blocked meters are assigned one at a time in a round robin fashion, each time to the concentrator maximizing the correspondent $q_{m,c}$. We proceed with this phase until all the non-blocked meters have been redistributed, producing a better balancedness of the loads. This concludes the i -th iteration of the algorithm. At each iteration, a partial solution \mathbf{S}_i and its corresponding objective function value $J(\mathbf{S}_i)$ is evaluated. When *max_iter* iterations are executed, the best solution \mathbf{S} is

returned. A formal description of the whole algorithm is given in **Algorithm 1**.

Algorithm 1 Heuristic algorithm

Input: max_iter ; ξ ; $q_{m,c}, \mathcal{R}_c, \mathcal{R}_m \forall m \in \mathcal{M}, \forall c \in \mathcal{C}$

Output: Association Indicators $x_{m,c}$ (whitelists)

{**Step 0**}

$x_{m,c} = 0, \forall m \in \mathcal{M}, \forall c \in \mathcal{C}$

Sort the DC in ascending order on \mathcal{R}_c cardinality

Set $i = 1$

while ($\exists m \mid \sum_{c \in \mathcal{R}_m} x_{m,c} = 0$) **do**

 Pick from the list concentrator c_i

$m' = \underset{m}{\operatorname{argmax}} q_{m,c_i}$

$x_{(m',c_i)} = 1$

$i = i + 1 \bmod \mathcal{C}$

end while

Set $\mathbf{S}_0 = \{x_{m,c} \mid \forall m \in \mathcal{M}, \forall c \in \mathcal{C}\}$

Set $\mathbf{S} = \mathbf{S}_0$

Compute $L_0 = \sum_{c \in \mathcal{C}} \log(L_c)$

Set *Blocking List* = \emptyset

for ($i = 1 : 1 : max_iter$) **do**

{**Phase 1**}

 Set $k = 1$

 Compute $G_m \forall m \in \mathcal{M}$

 Sort meters in descending order of G_m

while ($\frac{L_0 - \sum_{c \in \mathcal{C}} \log(L_c)}{L_0} < \xi_{max}$) and $k < \mathcal{M}$ **do**

 Pick from the list meter m_k

if $m_k \notin \textit{Blocking List}$ **then**

$x_{(m_k,c)} = 0, x_{(m_k,c_{m_k})} = 1$

 Insert m_k in the *Blocking List*

end if

$k = k + 1$

end while

{**Phase 2**}

$x_{(m,c)} = 0 \forall m \notin \textit{Blocking List}, \forall c \in \mathcal{C}$

 Sort the DC in ascending order on R_c cardinality, ignoring the meters in the *Blocking List*

 Set $k = 1$

while ($\exists m \mid \sum_{c \in \mathcal{R}_m} x_{m,c} = 0$) **do**

 Pick from the list concentrator c_k

$m' = \underset{m}{\operatorname{argmax}} q_{m,c_k}$

$x_{(m',c_k)} = 1$

$k = k + 1 \bmod \mathcal{C}$

end while

 Set $\mathbf{S}_i = \{x_{m,c} \mid \forall m \in \mathcal{M}, \forall c \in \mathcal{C}\}$

if $J(\mathbf{S}_i) > J(\mathbf{S})$ **then**

$\mathbf{S} = \{\mathbf{S}_i\}$

end if

end for

C. Data collection

The NIP formulation and the proposed heuristic algorithm require the knowledge of the link quality $q_{m,c}$ between each

meter and a DC. We observe that, beside the application data (i.e., the meter measurements themselves), data frames received at the different DCs provide a priceless source of information for estimating the link quality $q_{m,c}$. We propose to perform local processing of the received frames at each DC with the objective of retaining important information from meters while at the same time reducing the amount of data to be transmitted at the data management server. In addition to the aforementioned whitelist, each DC maintains a local database for storing meters related information. For each meter m , a concentrator stores the number of received messages from the meter $n_{m,c}$ and the average Received Signal Strength (RSS) $\mu_{m,c}$. Upon reception of a frame from a meter m with RSS s , a data concentrator c performs the following operations:

- 1) Check if m is already present in the local database. If not, create a new entry for that meter and set $n_{m,c} = 0, \mu_{m,c} = 0$.
- 2) Update the statistics relative to m in online fashion:

$$n_{m,c} = n_{m,c} + 1 \quad (9)$$

$$\mu_{m,c} = \frac{(n_{m,c} - 1) \cdot \mu_{m,c} + s}{n_{m,c}} \quad (10)$$

- 3) Check if m is present in the current whitelist. In case it is, forward the frame to the data management server. Otherwise, discard the frame.

At the end of a predefined time window (e.g., 2 weeks), the data concentrator transmits a copy of the local database to the data management server and clears all information. Such an approach allows to limit both the memory requirements on the data concentrator and the amount of resources needed for transmitting such information. Note that since the transmission rate of meters is known a priori (e.g., 2/4 frames per day), observing the number of received messages $n_{m,c}$ over a predefined period of time can be considered as a good estimator of the packet error rate (PER) $p_{m,c}$. Being f the transmission rate of a meter (in frames per day) and w the length in days of the time window, the PER can be estimated as:

$$p_{m,c} = 1 - \frac{n_{m,c}}{f \cdot w} \quad (11)$$

Normally, the PER is limited between 0 and 1. However, if a node is subject to clock-related issues (e.g., transmits more messages than what set per day), the PER value may also be lower than 0. In this case, simply checking the PER value may identify malfunctioning meters that may deplete their batteries sooner than expected. Finally, the indicators $p_{m,c}$ and $\mu_{m,c}$ can be conveniently fused to estimate the link quality $q_{m,c}$:

$$q_{m,c} = u \cdot \mu_{m,c} - v \cdot p_{m,c} \quad (12)$$

where u and v are tunable weighting coefficients which can be adjusted according to a user's preference.

Also, considering that variation ranges of $\mu_{m,c}$ and $p_{m,c}$ can be very different ($\mu_{m,c}$ typically varies between -10 dbm and -120 dbm while $p_{m,c}$ is normally limited between 0 and 1), for numerical stability they are normalized to unity, such that both vary within 0 and 1.

TABLE II: Numerical Results. S,M and L instances have respectively: 1436, 10828 and 82616 number of variables. OF stands for Objective Function, ET stands for Execution Time

		Optimum		Blind Optimum		Heuristic		Heur vs Opt
		OF	ET (min)	OF	Gap from Opt (%)	OF	ET (min)	Gap (%)
S	1	229,94	256,2	211,37	8,1	223,91	0,18	2,6
	2	414,25	66	358,40	13,5	400,49	0,17	3,3
	3	387,19	285,6	342,53	11,5	373,43	0,19	3,6
M	1	1448,63	>480	1319,71	8,9	1400,46	1,46	3,3
	2	1855,52	>480	1593,26	14,1	1794,36	1,45	3,3
	3	1763,15	>480	1550,54	12,1	1700,76	1,46	3,5
L	1	9446,19	>480	8437,46	10,7	9173,90	36	2,9
	2	12006,6	>480	10470,82	12,8	11636,55	33	3,1
	3	11173,9	>480	9860,81	11,8	10770,54	35	3,6

IV. EXPERIMENTS

A. General Description and Experiments Setup

We consider three instances of a real smart meter network deployed over a urban area in the north of Italy. The network is composed of about 10000 smart gas meters and 201 concentrators, operating the wM-Bus protocol using mode N. The backhaul link from each concentrator to the network server is based on GPRS technology. DCs collect channel information as detailed in Section III-C and transmit it every 14 days using the MQTT protocol to a central server. There, a centralized MySQL database is used to store all received information.

From this scenario, we created three network instances by sampling spatially three windows of different sizes, containing $M = [373, 1931, 9512]$ meters, always considering all available concentrators. Windows sizes have been empirically set with the rationale of considering problems with an increasing number of variables such to compare the quality of the obtained solutions and the corresponding execution times. We refer to these instances as Small (**S**), Medium (**M**) and Large (**L**) in Table II.

The experiments are characterized by four degrees of freedom which regard on the one hand the channel quality metric definition (u, v in (12)) and on the other hand the weights of the utility objective function (α, β in (1)). Believing that neither optimal channel conditions with unbalanced loads among the DCs nor perfect load balancedness with poor channel conditions can be considered desired network states for the application under study, we set α and β such that the tradeoff between the two system aspects we consider is balanced. In practice, α is set to $\sum_{c \in \mathcal{C}} |\mathcal{R}_c| / 10^4$ while β is set to 1. For what concerns the channel quality part of the objective function, we investigate three cases, considering i) only $\mu_{m,c}$ ($(u, v) = (1, 0)$), ii) only $p_{m,c}$ ($(u, v) = (0, 1)$) and iii) a convex combination of the two ($(u, v) = (0.3, 0.7)$). These three cases are referred to in Table II as: *Load+RSS* (**S1,M1,L1**), *Load+PER* (**S2,M2,L2**) and *Load+RSS&PER* (**S3,M3,L3**).

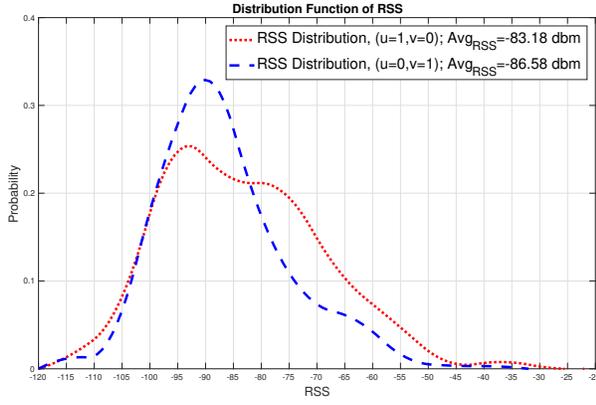
A further investigated case is referred to in Table II as *Blind Optimum*, where we suppose that channel quality information

is not available and the channel quality metric $q_{m,c}$ is replaced in the optimization problem (1)-(4) by $-d_{m,c}$, where $d_{m,c}$ is the physical distance between the m -th meter and the c -th concentrator. This case reflects the scenario of the very first network deployment, where whitelists are computed assuming that the shorter the meter-DCs distance, the higher the channel quality. For what concerns other settings, the NIP is formulated with AMPL and solved with the BONMIN 1.3 solver on a Intel Core 2 Duo processor with P9700 chipset, 2.8 GHz CPU and 4 GB RAM running on the platform Microsoft Windows 7. We choose to limit the maximum running time of the NIP to 8 hours; after this, the solver returns the best solution found so far, if the problem is feasible. The heuristic algorithm is implemented in Matlab, with max_iter set to 1000 and ξ_{max} to 5%.

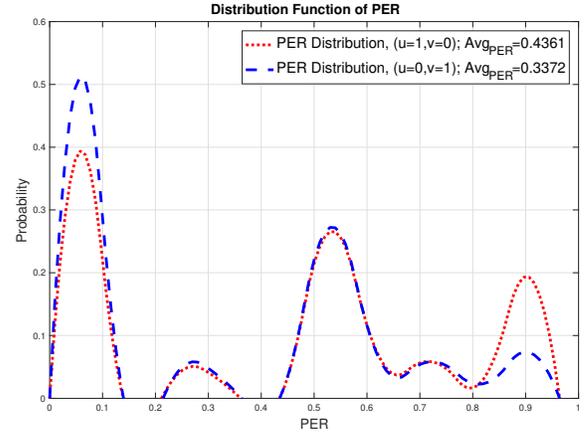
B. NIP Output

We first analyse the results obtained by solving the NIP formulation to optimality. Figure 2a and 2b show the distributions of the RSS and PER resulting after the computation of the optimal whitelists, considering different quality metric parameters. When $u = 1$ and $v = 0$ the channel metric optimizes the RSS, while for $v = 1$ and $u = 0$ the problem focuses on the PER. As one can see, in the former case (red curves) the average RSS is 3dBm higher than in the latter case, whereas the average PER is almost 10% better when explicitly targeted (blue curves). Looking at the results reported in Table II for the small instance, we observe that targeting only the PER is computationally cheaper than targeting only the RSS (or both) by almost four times. For larger instances however, due to the combinatorial structure of the ILP, execution times are greater than 8 hours. This cannot be acceptable considering that the network is expanded incrementally according to specific needs and thus requiring quick and easy updates of the association patterns. Table II shows that both the **M** and the **L** instances have led the NIP to exit for time limit, investigating just a small subgroup of the solution space; only the **S** instance returns the optimum association in a reasonable, still too large, time frame.

Moreover, as shown in Table II, the results obtained from



(a) RSS Distribution, M instance



(b) PER Distribution, M instance

Fig. 2: Sensibility of the ILP to weights u and v

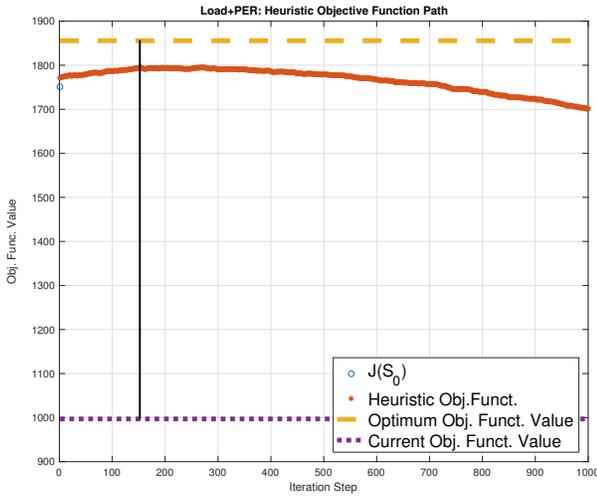


Fig. 3: Obj. Function path of heuristic approach and gap with the optimum (M instance, $(u = 0, w = 1)$)

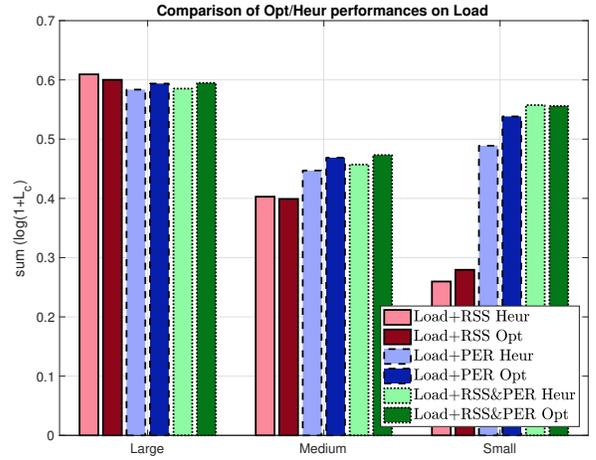


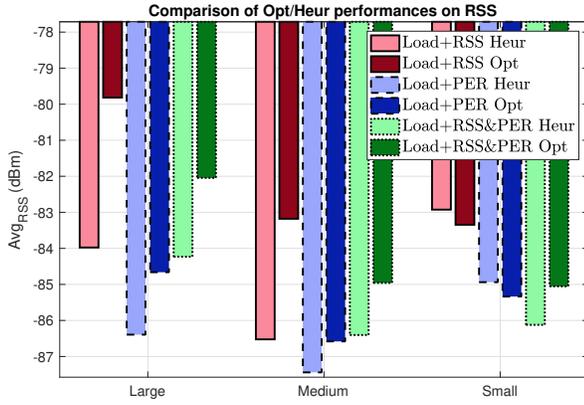
Fig. 4: First term of J in Optimum and Heuristic cases, light and dark bars respectively

the Blind case are not close to the optimal ones: the gaps are greater than 8%, approaching in some cases 14%. This is expected, considering the harsh propagation conditions usual of urban environments. In such a scenario, the meter-DC couple with the shorter distance is not necessarily the best association.

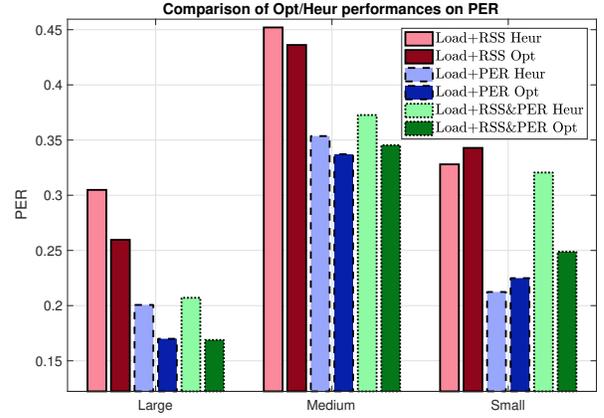
C. Heuristic vs Optimal Solution

The heuristic algorithm is meant to give acceptable sub-optimal solutions in shorter execution times compared to those of the NIP. The algorithm tries to improve as much as possible the objective function value from the starting point of the S_0 solution. Figure 3 plots the evolution of the objective function value during the execution of the heuristic for one of the considered case study. As reported in Table II, the

solutions obtained by the proposed heuristic are very close to the optimal ones, even after relatively shorter execution times: gaps are limited to less than 4% of the optimal solution. It is worth looking at the first term of the objective function and at the final quality metrics resulting from the computed whitelists, which contribute to define the second term of the objective function. Figure 4 shows that the load term of J in the heuristic case is close to the optimum one, being even greater in some cases: what happens is that when the heuristic load term dominates the optimal one then the opposite happens for the quality metric term and vice versa, the sum of the two being always greater in the optimum case. Figure 5a illustrates the average RSS resulting from the Optimum and Heuristic in the three scenarios, while Figure 5b does the same for what concerns the PER. As one can see, when the focus is on RSS (red bars), the average RSS is between 1 and 2 db higher



(a)



(b)

Fig. 5: Average RSS (a) and PER (b) resulting from the Optimum (dark colors) and Heuristic solutions (light colors)

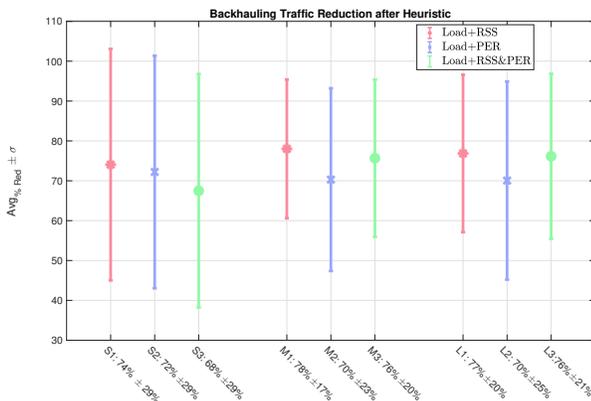


Fig. 6: Mean per-GW Backhauling Traffic Reduction

in Optimum than in Heuristic in **M** and **L** instances while the Heuristic gives slightly higher RSS in **S** instance. When PER is targeted (blue bars), in **L** instance the Heuristic gives almost 15% higher PER than the Optimum while the gap in **M** instance is reduced to 6%. The situation is reversed in **S** instance, where the Heuristic gives 9% better PER with respect to the Optimum. Again, where Heuristic performance are better from a quality metric perspective, the Optimum does better on the load side. We anyway underline that the tradeoff between the two terms of the objective function depends on the choice regarding the setting of the parameters α and β , which can be tuned according to the user's preference.

D. Reduction of the Backhauling Load

One of the main motivation for the adoption of whitelists, disregarding the strategy implemented to create them, is to limit the amount of traffic to be forwarded from data concentrators on their GPRS-based backhauling links to the central server, as well as the computational load the server is subject to. Figure 6 reports the average and standard deviation of the

traffic reduction that can be obtained on each gateway after the activation of whitelists in the three considered network scenarios and according to the three different combinations of u and v parameters, with respect to a whitelists-less situation. As one can see, the average per-concentrator reduction is between 65% and 80% regardless of the considered scenario, with peaks of more than 100% (i.e., the backhauling traffic is halved). At the central server, such reduction is even more important, as it is multiplied by the number of concentrators.

V. RELATED WORK

Several works in the last few years have given attention to the wireless M-Bus protocol, either from a performance evaluation point of view [6] or from a system design perspective [7]. In [2], [8] and later in [9], the application of wM-Bus protocol to smart water grids is analyzed. Authors focus in particular on energy requirements of wM-Bus compliant smart meters, and reason on the possibility of using water kinetic energy as an energy harvesting source, also taking into consideration the transmission range and the collision probability of each meter. Using both simulations and experimental tests with a system prototype, authors conclude that the wM-Bus protocol is an adequate tradeoff for the applicative context under study. In [10], authors analyze statistical properties of frames transmitted by wM-Bus gas/water meters and received at different DCs during one year. The analysis reveals that regulating the transmission power solely according to the distance from the nearest DC is not enough to provide adequate transmission quality, and suggests to use the monthly average value of the RSS as a predictor to tune the transmission power for the next observation period. In this work, on the one hand we confirm that distance-based approaches do not capture channel quality in all its aspects and on the other hand we investigate a new channel quality metric, evaluating its efficiency to drive the network towards better system performances.

The problem of balancing the number of users connected to each gateway/base station is of great importance in the context

of both mobile cellular and wireless sensor networks. In cellular networks, the problem of unbalanced loads among cells deteriorating system-wide performances is of great concern and is usually considered in the context of resources allocation to new users approaching the network and looking for a base station to associate. In [3] and later in [4], authors propose an ILP formulation which aims at finding the best associations between users and base stations in order to achieve load balancing among base stations and some level of fairness among the bandwidth allocated to users. Authors underline that while linear utility functions results in throughput-optimal non-balanced solutions, concave objective functions, such as the logarithm, naturally achieve fairness among users. An interesting option is given also in [11], where a weighted utility objective function is proposed, such that weighting coefficients on a load balance index and network throughput are properly tuned according to the desired level of tradeoff between the two. In sensor networks, where equipments are usually static and requirements in terms of throughput are much coarser, the problem of balancing the load among a-priori chosen DCs is tackled by properly clustering the sensor nodes such that each sensor belongs to one and only one cluster and communicates with the central server only via the head of that cluster, as presented in [12]–[15]. Load balancing among cluster heads is achieved by either minimizing the variance of the number of sensors connected to each cluster head [12] or by taking a min-max approach [13]. In [15] a distance based, load aware clustering algorithm is presented where gateways are sorted at each iteration in ascending order on the number of so far assigned sensors (min-heap tree), which are then associated to the nearest gateways following the aforementioned list. A similar algorithm but working from sensors' side is shown in [14], where sensors are sorted in ascending order on the number of in-range gateways and then assigned to the best gateway according to the chosen metric. We note that there is no evidence in literature about a joint use of the above perspectives and this work shows a possible approach in this context. Moreover, beside the focus on load balancing, our work focuses also on system performances, introducing new aspects different than the distance in order to achieve better results.

VI. CONCLUSION

In this paper we propose a data-driven optimization framework to jointly balance the loads and maximize the channel quality of a smart meter network. This is crucial for the reduction of the traffic on the limited-capacity backhauling links, the reduction of the computational load at the network gateways towards the central server and the establishment of a safer and less-ambiguous association-based communication paradigm between meters and gateways. We show that the information contained in the data frames sent by the sensors can be stored at DC side and exploited to lead the system to better performances in terms of Packet Error Rate, Received Signal Strength and load balancedness. Then, we propose

a heuristic approach which is able to give good solutions in proportionately smaller execution times, which is crucial relatively to the need of smart metering networks. In fact, considering that usually such networks face more than one deployment phases, they strongly need to quickly update their network association states if the case requires it and so computational times cannot be too long. The gaps between the Optimum and the Heuristic objective function values show that heuristic solutions are close enough to optimal ones, being therefore a good approximation. Future works could regard, among the others, the definitions of finer channel quality metrics, the formulation of quicker and more-performing heuristics and the consideration of different utility functions.

REFERENCES

- [1] "Gesetz zur digitalisierung der energiewende," *Deutscher Bundestag*, Sep 2016.
- [2] S. Spinsante, M. Pizzichini, M. Mencarelli, S. Squartini, and E. Gambi, "Evaluation of the wireless m-bus standard for future smart water grids," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*. IEEE, 2013, pp. 1382–1387.
- [3] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*.
- [4] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.
- [5] E. Rakotomanana and F. Gagnon, "Fair load balancing in heterogeneous cellular networks," in *Ubiquitous Wireless Broadband (ICUWB), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–5.
- [6] P. Masek, K. Zeman, Z. Kuder, J. Hosek, S. Andreev, R. Fujdiak, and F. Kropfl, "Wireless m-bus: an attractive m2m technology for 5g-grade home automation," in *International Internet of Things Summit*. Springer, 2015, pp. 144–156.
- [7] Ž. Mihajlović, A. Milankov, L. Živković, and M. Tolić, "Implementation of wireless m-bus concentrator/gateway for remote reading of smart gas meters," in *Telecommunications Forum (TELFOR), 2016 24th*. IEEE, 2016, pp. 1–4.
- [8] S. Squartini, L. Gabrielli, M. Mencarelli, M. Pizzichini, S. Spinsante, and F. Piazza, "Wireless m-bus sensor nodes in smart water grids: the energy issue," in *Intelligent Control and Information Processing (ICICIP), 2013 Fourth International Conference on*. IEEE, 2013, pp. 614–619.
- [9] S. Spinsante, S. Squartini, L. Gabrielli, M. Pizzichini, E. Gambi, and F. Piazza, "Wireless m-bus sensor networks for smart water grids: analysis and results," *International Journal of Distributed Sensor Networks*, vol. 10, no. 6, p. 579271, 2014.
- [10] M. Carratù, M. Ferro, V. Paciello, A. Pietrosanto, and P. Sommella, "Performance analysis of wm-bus networks for smart metering," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7849–7856, 2017.
- [11] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu, and X. You, "Dynamic load balancing and throughput optimization in 3gpp lte networks," in *Proceedings of the 6th international wireless communications and mobile computing conference*. ACM, 2010, pp. 939–943.
- [12] G. Gupta and M. Younis, "Load-balanced clustering of wireless sensor networks," in *Communications, 2003. ICC'03. IEEE International Conference on*, vol. 3. IEEE, 2003, pp. 1848–1852.
- [13] C. P. Low, C. Fang, J. M. Ng, and Y. H. Ang, "Load-balanced clustering algorithms for wireless sensor networks," in *Communications, 2007. ICC'07. IEEE International Conference on*. IEEE, 2007, pp. 3485–3490.
- [14] P. Kuila and P. K. Jana, "Improved load balanced clustering algorithm for wireless sensor networks," in *International Conference on Advanced Computing, Networking and Security*. Springer, 2011, pp. 399–404.
- [15] —, "Energy efficient load-balanced clustering algorithm for wireless sensor networks," *Procedia Technology*, vol. 6, pp. 771–777, 2012.