IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 7, 2018, accepted May 16, 2018, date of publication June 6, 2018, date of current version June 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2841846

A Bounded Heuristic for Collection-Based Routing in Wireless Sensor Networks

GABRIELA SCHÜTZ[1,2], NOELIA CORREIA[1,3], JAIME A. MARTINS[1], (Member, IEEE), ANDRIY MAZAYEV[1], AND ALVARO BARRADAS[1,3]

[1]Center for Electronic, Optoelectronic and Telecommunications, University of Algarve, 8005-139 Faro, Portugal
[2]Institute of Engineering, University of Algarve, 8005-139 Faro, Portugal
[3]Faculty of Science and Technology, University of Algarve, 8005-139 Faro, Portugal

Corresponding author: Gabriela Schütz (gschutz@ualg.pt)

This work was supported by the FCT from Portugal within the CEOT research center under Grant UID/MULTI/00631/2013.

ABSTRACT Wireless sensor networks are used to monitor and control physical phenomena and to provide interaction between clients and the physical environment. Clients have been typically users or user applications, but next generation wireless sensor networks will also work in machine-to-machine scenarios where some nodes can be interested in some other nodes' data. These scenarios may run the risk of becoming overloaded with messaging, a pernicious fact in particular for constrained networks where both bandwidth and power supply are limited. Resource collections can be used in wireless sensor networks to improve bandwidth usage and to reduce energy consumption, reducing the overall number of notification packets and wrapping overhead, required for the delivery of sensor data. This article proposes a heuristic algorithm for the planning of both routing and collections, in wireless sensor networks. Results show that collections are always worthwhile, and that the heuristic is able to find feasible and cost effective solutions, approaching its lower bound.

INDEX TERMS Constrained networks, energy efficiency, heuristic algorithms, wireless sensor networks.

I. INTRODUCTION

Wireless sensor nodes are used to monitor and control physical phenomena, enabling interactions between clients (e.g., users or applications) and the surrounding environment [1]. In machine-to-machine communication scenarios, nodes can also be interested in each other's data, meaning that both producer and consumer entities will exist. In such environments, the delivery of data notifications should be done while trying to increase the network lifetime [2], [3].

Recently, interfaces for the creation/update of resources (i.e., Collections) in *Constrained RESTful Environments* (CoRE) were proposed [4]. The CoRE working group, within the Internet Engineering Task Force (IETF), focus on realizing the *Representational State Transfer* (REST) architecture in a suitable form for constrained nodes, allowing for applications to discover resources hosted by constrained servers [5], [6]. In [4], a Collection is defined as a resource representing one or more related resources, and its state can be observed using CoAP/Observe, proposed in [7] and [8], similarly to individual resources. This means that the overall number of notification packets and wrapping information is reduced when compared with the transmission of the same data, by individual notification packets. Collections can be explored to use bandwidth more efficiently, reducing energy consumption. This article proposes a bounded heuristic algorithm for the routing of notification packets, while considering the creation of Collections.

The remainder of this article is organized as follows: Section II discusses the optimization problem, while Section III presents the bounds and claims to get feasible solutions. Section IV presents the heuristic algorithm. In Section V the results are discussed, while in Section VI some conclusions are drawn.

II. ASSUMPTIONS AND OPTIMIZATION PROBLEM

Let us assume a constrained network with nodes detecting an event or change in the environment, producing a measurement, and/or interested in measurements done by others. Such network is represented as an undirected graph $\mathcal{G}(\mathcal{N}, \mathcal{E}, \mathcal{S})$, where $\mathcal{N}$ and $\mathcal{E}$ denote the set of nodes and edges, respectively, and $\mathcal{S}$ denotes the overall set of measured subjects (e.g., temperature measurement). Each node $n \in \mathcal{N}$ is assumed to produce a single subject $s \in \mathcal{S}$, and δ_s^n is used to indicate whether node n produces subject s or not. An edge $e = (n_i, n_j) \in \mathcal{E}$ is a wireless direct communication channel between nodes n_i and n_j.

The set of consumer nodes, which are interested in specific measurements done by others, is denoted by $\mathcal{M}$, $\mathcal{M} \subset \mathcal{N}$. A consumer can be any inner network node, also performing any task that requires measurement, or a gateway responsible for the forwarding of measurements towards clients/applications. A consumer node $m \in \mathcal{M}$ has a set of subject interests, denoted by $\mathcal{I}(m)$. To meet the interests of consumer nodes, subject data must flow throughout the network graph (from producers to consumers). Subjects can flow either individually, each inside its own notification message, or inside Collection notification messages, which aggregate a set of subjects. The way measurements flow can change at intermediate nodes, as they can build new Collections from arriving subjects. Regarding Collections, the following condition holds:

Condition 1 (Collection Flow): A k-subject Collection arriving to a node can either be consumed by the node and/or forwarded. In the latter case, the Collection must flow out undivided, either alone or inside another Collection.

Regarding the cost associated with a flow f that remains unchanged from node n_i towards node n_j, this will be:

$$\Delta(f, n_i, n_j) = (k \times \alpha + \beta) \times \text{hops}(n_i, n_j), \tag{1}$$

where k is the number of subjects, $\text{hops}(n_i, n_j)$ is the number of hops from n_i to n_j, and $\alpha + \beta = 1$. Variable α is a per subject data overhead, while β is the overhead associated with wrapping all data. Therefore, a subject flowing individually will have a cost of $\text{hops}(n_i, n_j)$, while a k-subject Collection will have a cost of $(k \times \alpha + \beta) \times \text{hops}(n_i, n_j)$, meaning that there is a benefit when subjects are aggregated into Collections (a reduction in overall wrapping information, for the same data being transferred). In the following discussion, without loss of generality, $\alpha = \beta = 0.5$. Note that n_i can either be the source/producer of a subject, or an intermediate node; while n_j can either be the destination/consumer node, or an intermediate node.

With the previous assumptions in mind, the problem of routing subject measurements in constrained wireless sensor networks, called *Collection-based Routing in Wireless Sensor Networks* (CR-WSN), is defined as:

Definition 1 (CR-WSN Problem): Given a constrained network with a set of nodes $\mathcal{N}$, where each node $n \in \mathcal{N}$ produces a single subject $s \in \mathcal{S}$, and each consumer node $m \in \mathcal{M}$, $\mathcal{M} \subset \mathcal{N}$, has a set of subject interests $\mathcal{I}(m)$, determine the multi-hop routing of subjects throughout the network, building (if necessary) Collections so that the overall cost is minimized. This must be done satisfying consumer interests and Collection Flow condition.

III. FEASIBLE SOLUTIONS ON CR-WSN PROBLEM

The following sections discuss some ways of obtaining feasible solutions for the CR-WSN problem, which are later used by the proposed heuristic.

A. AN UPPER BOUND

Under the previous assumptions and the undirected graph $\mathcal{G}(\mathcal{N}, \mathcal{E}, \mathcal{S})$ defined, the following can be stated:

- If a consumer $m \in \mathcal{M}$ has subject $s \in \mathcal{S}$ in its interests, and it is flowing individually from one of its producers $n \in \mathcal{N}$ towards m, then the shortest path is the one providing the lowest cost.
- For multiple consumer nodes interested in subject $s \in \mathcal{S}$, the set of shortest paths (between producers and consumers of s) sharing the largest number of arcs will be the set of shortest paths providing the lowest cost.
- Considering multiple subjects, from multiple consumers, the shortest paths from producers to consumers that share more arcs will be good candidates for the insertion of Collections (at sections where arcs are highly shared), reducing costs.
- Network graphs (emerging from the shortest paths required to meet all consumer interests) with a high number of shared arcs have more potential for the replacement of individual subject flows by Collections, leading to better solutions for the CR-WSN problem.

Assuming a consumer $m \in \mathcal{M}$ and a subject interest $s \in \mathcal{I}(m)$, the set of paths from producers of s towards m will be:

$$\begin{aligned}\mathcal{P}_s^m = \{p = (e_i, \ldots, e_j) : e_j = (n_k, m) \in \mathcal{E}, n_k \in \mathcal{N} \wedge \\ \wedge e_i = (n_i, n_j) \in \mathcal{E}, \delta_s^{n_i} = 1, \forall n_i, n_j \in \mathcal{N}\} \\ \forall m \in \mathcal{M}, \quad s \in \mathcal{I}(m),\end{aligned} \tag{2}$$

and the shortest paths $\mathcal{P}_s^{*,m}$ are given by:

$$\mathcal{P}_s^{*,m} = \text{argmin}_{p \in \mathcal{P}_s^m}(\text{hops}(p)) \forall m \in \mathcal{M}, s \in \mathcal{I}(m) \tag{3}$$

where $\text{hops}(p)$ is the number of hops in path p, and $\mathcal{P}_s^{*,m} \subset \mathcal{P}_s^m$.

For multiple consumer nodes interested in a subject $s \in \mathcal{S}$, and assuming only the shortest paths, the set of arcs used for transmission of subject s will be:

$$\mathcal{P}_s^* = \bigcup_{m \in \mathcal{M}: s \in \mathcal{I}(m)} \mathcal{P}_s^{*,m} \quad \forall s \in \bar{\mathcal{S}} \tag{4}$$

where $\bar{\mathcal{S}} = \bigcup_{m \in \mathcal{M}} \mathcal{I}(m)$ is the overall consumer node interests, $\bar{\mathcal{S}} \subseteq \mathcal{S}$. Duplicate arcs are merged.

An upper bound on the cost for the CR-WSN problem will, therefore, be:

$$\zeta^{\text{ub}} = \sum_{s \in |\bar{\mathcal{S}}|} |\mathcal{P}_s^*|. \tag{5}$$

That is, the upper bound is based on the shortest paths in $\mathcal{P}^* = \{\mathcal{P}_1^*, \ldots, \mathcal{P}_{|\bar{\mathcal{S}}|}^*\}$, and no Collections still exist.

B. IMPROVED FEASIBLE SOLUTION

The set $\mathcal{P}^* = \{\mathcal{P}_1^*, \ldots, \mathcal{P}_{|\bar{\mathcal{S}}|}^*\}$ provides a feasible solution for the CR-WSN, but multiple combinations of shortest paths may exist. Also, slightly longer paths can reduce cost if Collections are introduced. This means that there is potential for improvement of such feasible solution. Let us define a directed graph $\bar{\mathcal{G}}(\bar{\mathcal{N}}, \mathcal{A}, \bar{\mathcal{S}})$ where $\bar{\mathcal{N}}$ includes nodes used at $\mathcal{P}^*$, $\mathcal{A}$ includes the arcs in $\mathcal{P}^*$, and $\bar{\mathcal{S}}$ includes subjects of interest to consumers, previously defined. For such a feasible solution the following claims, regarding solution improvement (cost reduction), can be stated:

Claim 1 (Triangles): If (n_i, n_k), (n_j, n_k) and (n_j, n_i) exist, and the last two arcs both carry subject s, then (n_j, n_k) can be eliminated (a triangle exists). This causes either the reduction of cost, if (n_i, n_k) already carries s, or keeps the cost unchanged, if s currently does not flow at (n_i, n_k) (needs to be inserted now in (n_i, n_k)).

Claim 2 (Redundancies): If a node has more than one incident arc for the same subject, one of the arcs can be removed. If the source node of such arc has no arcs towards other nodes, or other subjects flowing into it, then it can be removed together with its arcs.

Claim 3 (Creation of Collection): The graph that emerges after aplying Claims 1 and 2 is in condition to have some individual subject flows replaced by Collection flows, for minimizing cost, while not violating Condition 1. If q arcs from n_i to n_j carry different subjects, then a Collection can be built at n_i including:

- q subjects, if these are of interest to all nodes reachable from n_j; or n_j has no successors (consumer node);
- r subjects, $r < q$, if only the r subjects are of interest to all nodes reachable from n_j.

Claim 4 (Partial Cost): For k-subjects to flow from node n_i towards n_j (with the flow starting at node n_i and kept unchanged until node n_j is reached), the lowest cost is achieved if a Collection of such k-subjects is built. Their individual flow, or including them in a larger Collection, has a higher cost. More specifically:

- $\text{Cost}(f_1) = (k \times \alpha + \beta) \times \text{hops}(n_i, n_j)$, if a Collection of k-subjects is built;
- $\text{Cost}(f_2) = k \times \text{hops}(n_i, n_j)$, for the individual flow of the k-subjects;
- $\text{Cost}(f_3) = (k' \times \alpha + \beta) \times \text{hops}(n_i, n_j)$, if the k-subjects flow inside the Collection of k'-subjects, $k' > k$.

$\text{Cost}(f_1) < \text{Cost}(f_2)$ and $\text{Cost}(f_1) < \text{Cost}(f_3)$, where f_1, f_2 and f_3 are flows kept unchanged from source to destination.

The cost of the feasible solution obtained after applying Claims 1, 2, 3 and 4, while not violating Condition 1, will be:

$$\zeta^* = \sum_{f \in \mathcal{F}} \text{Cost}(f) \tag{6}$$

where $\mathcal{F}$ is the overall set of flows. The heuristic algorithm discussed below finds a solution, providing the lowest cost feasible solution for the shortest paths based graph $\mathcal{P}^*$. A lower bound on ζ^* can be obtained as follows:

Claim 5 (Lower Bound): When relaxing the Collection Flow condition, a lower bound can be obtained for the lowest cost feasible solution, on the shortest paths based graph.

The lower bound at Claim 5, which will be denoted by ζ^*_{lb}, can be used to evaluate the quality of the heuristic algorithm (values obtained for ζ^*).

IV. HEURISTIC ALGORITHM

For subjects to flow towards consumers in WSNs using, if necessary, Collections (CR-WSN problem) several issues must be solved: *i*) determine which Collections to build, their content and where to place them; *ii*) multi-hop routing. This is an NP-hard combinatorial optimization problem and, therefore, a reasonable way to find a solution is to solve it heuristically. Although there are efficient algorithms to find a shortest path, such approaches do not incorporate building Collections. Also, in WSNs there are usually multiple shortest paths, which could lead to different final solutions. The heuristic algorithm proposed in this section takes these issues into consideration while including the characteristics and reasoning discussed in the previous sections. The heuristic includes two steps:

Step I: Build an initial feasible solution. This solution is based on Section III-A.

Step II: Improve the initial solution. Improvements are based on the Claims presented in Section III-B.

These steps are discussed in more detail below, for which the following extra notation is required:

γ_s^n	Has value *one* if subject $s \in \bar{\mathcal{S}}$ arrives and/or departs from node $n \in \bar{\mathcal{N}}$; *zero* otherwise.
γ_s^a	Has value *one* if subject $s \in \bar{\mathcal{S}}$ flows through arc $a \in \mathcal{A}$; *zero* otherwise.
$\mathcal{R}(n)$	Set of subjects $\{s \in \bar{\mathcal{S}} : \delta_s^n = 1 \vee \gamma_s^n = 1\}$, $\forall n \in \bar{\mathcal{N}}$.
$\mathcal{R}(a)$	Set of subjects $\{s \in \bar{\mathcal{S}} : \gamma_s^a = 1\}$, $\forall a \in \mathcal{A}$.
$\mathcal{R}_{col}(n)$	Set of Collections built at node $n \in \bar{\mathcal{N}}$.
$\mathcal{R}_{col}(a)$	Set of Collections flowing through arc $a \in \mathcal{A}$.
$\Gamma^+_{n_i}$	Set of direct successors of $n_i \in \bar{\mathcal{N}}$, $\{n_j \in \bar{\mathcal{N}} : (n_i, n_j) \in \mathcal{A}\}$.
$\Gamma^-_{n_i}$	Set of direct predecessors of $n_i \in \bar{\mathcal{N}}$, $\{n_j \in \bar{\mathcal{N}} : (n_j, n_i) \in \mathcal{A}\}$.
Γ_{n_i}	Set of nodes reachable from $n_i \in \bar{\mathcal{N}}$, $\{n_j \in \bar{\mathcal{N}} : \exists$ path from n_i to $n_j\}$.
ζ^*	Cost associated with the best feasible solution.

A. STEP I: BUILD AN INITIAL FEASIBLE SOLUTION

An initial feasible solution is built to be used as input to Step II. This procedure, shown in Algorithm 1, is based on Section III-A where shortest paths are used to meet consumer interests. No Collections are considered during this first step.

B. STEP II: IMPROVE INITIAL FEASIBLE SOLUTION

This step improves the feasible solution provided by Step I. Such improvement is based on Claims presented in Section III-B, and the possibility of building Collections for

Algorithm 1 Building Initial Feasible Solution

Input: $\mathcal{N}$, $\mathcal{E}$, $\mathcal{M}$, $\mathcal{I}(m)$, $\bar{\mathcal{S}}$, δ_s^n, $\forall n \in \mathcal{N}$, $\forall s \in \bar{\mathcal{S}}$;
Output: $\bar{\mathcal{N}}$, $\mathcal{A}$, $\mathcal{P}^*$, $\mathcal{R}(n)$, $\mathcal{R}(a)$ *and* ζ^{ub}, $\forall n \in \bar{\mathcal{N}}$, $\forall a \in \mathcal{A}$;

Initialize $\bar{\mathcal{N}}$, $\mathcal{A}$, $\mathcal{R}(a = (n_i, n_j))$ and $\mathcal{R}(n_i)$ to empty sets, $\forall n_i, n_j \in \mathcal{N}$;
for $m \in \mathcal{M}$ **do**
  for $s \in \mathcal{I}(m)$ **do**
    Determine $\mathcal{P}_s^{*,m}$ using Eq. (3);
  end
end
for $s \in \bar{\mathcal{S}}$ **do**
  Determine $\mathcal{P}_s^*$ using Eq. (4);
  for $a = (n_i, n_j) \in \mathcal{P}_s^*$ **do**
    $\mathcal{A} \leftarrow \{a\}$;
    $\bar{\mathcal{N}} \leftarrow \{n_i, n_j\}$;
    $\mathcal{R}(a) \leftarrow \{s\}$;
    $\mathcal{R}(n_i) \leftarrow \{s\}$;
    $\mathcal{R}(n_j) \leftarrow \{s\}$;
  end
  $\mathcal{P}^* \leftarrow \mathcal{P}_s^*$;
end
Compute ζ^{ub} using Eq. (5);

---

cost improvement is included. This procedure is shown in Algorithm 2.

V. ANALYSIS OF RESULTS

A. SCENARIO SETUP

Randomly generated graphs, using the algorithm in [9], were used to analyse the performance of the heuristic algorithm. Graphs of 25, 50 and 100 nodes were tested. Two networks were generated, one denser (with more links) than the other, for each dimension. The subjects produced by each node (a single subject), consumer nodes and their interests were also randomly generated. The diversity of the subjects produced by the network is equal to 20% of the number of nodes, meaning that some nodes will produce the same subjects. Regarding consumer nodes, tests were done for 20%, 40% and 60% of the number of nodes. The number of subject interests requested by consumer nodes varies from 2 to 4. For each such scenario, 10 graph instances were generated and averages of such results are plotted. Table 1 summarizes the network parameters.

B. RESULTS

Table 2 summarizes the average result values obtained for: (*i*) ζ^{ub}, the upper bound on the CR-WSN problem; (*ii*) ζ^*, the best feasible solution found by the heuristic algorithm for the CR-WSN problem; (*iii*) ζ_{lb}^*, the lower bound; (*iv*) the average number of Collection in ζ^*'s and ζ_{lb}'s solutions, and their ratio $\frac{\zeta^*\text{'s Avg No. of Collections}}{\zeta_{\mathrm{lb}}\text{'s Avg No. of Collections}}$. These are analysed next.

Algorithm 2 Improving the Initial Feasible Solution

Input: $\bar{\mathcal{N}}$, $\mathcal{A}$, $\mathcal{P}^*$, $\mathcal{R}(n)$, $\mathcal{R}(a)$ *and* ζ^{ub}, *provided by Step I*;
Output: Improved $\bar{\mathcal{N}}$, $\mathcal{A}$, $\mathcal{P}^*$, $\mathcal{R}(n)$, $\mathcal{R}(a)$, $\mathcal{R}_{col}(n)$, $\mathcal{R}_{col}(a)$, ζ_{lb}^* *and* ζ^*, $\forall n \in \bar{\mathcal{N}}$, $\forall a \in \mathcal{A}$;

$\zeta^* = \zeta^{\mathrm{ub}}$;
Initialize $\mathcal{R}_{col}(n)$ and $\mathcal{R}_{col}(a)$ to empty sets, $\forall n, \in \bar{\mathcal{N}}$, $\forall a \in \mathcal{A}$;
for $n_k \in \bar{\mathcal{N}}$ **do**
  for $(n_i, n_j) \in \Gamma_{n_k}^-$ **do**
    Apply Claims 1 and 2 to n_i, n_j and n_k, as discussed in Section III-B;
    Update $\mathcal{R}(n_i) \vee \mathcal{R}(n_j)$;
    Update $\mathcal{R}(a = (n_i, n_k)) \vee \mathcal{R}(a = (n_j, n_k))$;
    Update $\bar{\mathcal{N}}$, $\mathcal{A}$, $\mathcal{P}^*$;
  end
end
Compute ζ_{lb}^* using Claim 5;
for $m \in \mathcal{M} : \Gamma_m^+ = \emptyset$ **do**
  for $n_i \in \Gamma_m^-$ **do**
    Apply Claim 3 to n_i, m and $a = (n_i, m)$ to build a Collection, if Condition 1 is fulfilled;
    Update $\mathcal{R}_{col}(n_i)$, $\mathcal{R}_{col}(m)$, $\mathcal{R}_{col}(a)$ and ζ^*;
  end
end
for $a = (n_i, n_j) \in \mathcal{A} : |\mathcal{R}(a)| > 1$ **do**
  for $f \in \{n_i, n_j, \ldots, n_q\}$, $n_q \in \Gamma_{n_j}$ **do**
    Apply Claim 4 to f to build a Collection for flow f, if possible, as discussed in Section III-B;
    Update $\mathcal{R}_{col}(n)$, $\forall n \in f$;
    Update $\mathcal{R}_{col}((n_k, n_l))$, $\forall (n_k, n_l) \subset f$;
    Update ζ^*;
  end
end

---

1) RESULTS ON COST VALUES

For a more rigorous analysis of the heuristic performance, the gap between ζ^{ub} and ζ^*, and between ζ_{lb} and ζ^*, were computed as follows:

$$\mathrm{Gap}^{\mathrm{ub}} = \frac{\zeta^{\mathrm{ub}} - \zeta^*}{\zeta^{\mathrm{ub}}} * 100 \tag{7}$$

$$\mathrm{Gap}^{\mathrm{lb}} = \frac{|\zeta_{\mathrm{lb}} - \zeta^*|}{\zeta_{\mathrm{lb}}} * 100. \tag{8}$$

Figures 1 and 2 plot the average of gap values obtained for all networks scenarios under analysis. It is possible to observe that the heuristic algorithm is able to significantly improve the ζ^{ub} values, with improvement percentages ranging from 1.5% to 14%, due to the creation of Collections. Therefore, better feasible solutions for the CR-WSN problem can be obtained using collections.

TABLE 1. Network parameters.

Net	$\lvert\mathcal{N}\rvert$	Den*	$\lvert\mathcal{S}\rvert$	$\lvert\mathcal{M}\rvert$	$\frac{1}{\lvert\mathcal{M}\rvert}\sum_{m\in\mathcal{M}}\lvert\mathcal{I}(m)\rvert$	Ref
A	25	36	5	5	2.80	A5
				10	2.70	A10
				15	2.73	A15
B	25	22	5	5	2.70	B5
				10	2.69	B10
				15	2.71	B15
C	50	21	10	10	2.70	C10
				20	2.68	C20
				30	2.67	C30
D	50	13	10	10	2.67	D10
				20	2.65	D20
				30	2.68	D30
E	100	11	20	20	2.65	E20
				40	2.68	E40
				60	2.67	E60
F	100	8	20	20	2.67	F20
				40	2.67	F40
				60	2.66	F60

* An n-node network has density equal to $\frac{\#edges}{[n\times(n-1)]/2}\times 100$.

TABLE 2. Average result values.

Ref	ζ^{UB}	ζ^*	ζ^*_{LB}	ζ^*'s Avg No. of Coll.	ζ^*_{LB}'s Avg No. of Coll.	Ratio
A5	16.00	14.95	14.90	2.20	2.10	1.05
A10	31.10	29.10	28.50	2.90	4.10	0.71
A15	46.70	43.50	42.75	5.80	6.90	0.84
B5	17.10	15.75	15.55	2.40	2.70	0.89
B10	33.40	30.95	30.35	4.20	5.40	0.78
B15	50.10	45.85	44.20	6.80	10.00	0.68
C10	39.50	37.05	36.50	4.10	5.10	0.80
C20	73.00	68.20	67.30	8.70	10.70	0.81
C30	104.90	98.15	95.45	11.90	17.10	0.70
D10	46.80	42.95	42.35	6.10	7.60	0.80
D20	85.10	79.25	75.65	10.10	16.70	0.60
D30	119.10	110.75	104.60	14.20	23.30	0.61
E20	95.70	89.50	87.95	10.70	13.50	0.79
E40	184.50	174.35	169.35	17.30	25.90	0.67
E60	268.00	251.75	241.95	26.90	43.30	0.62
F20	122.50	114.20	110.70	12.90	20.40	0.63
F40	230.60	213.40	202.55	24.90	44.10	0.56
F60	326.70	302.00	279.70	34.90	72.90	0.48

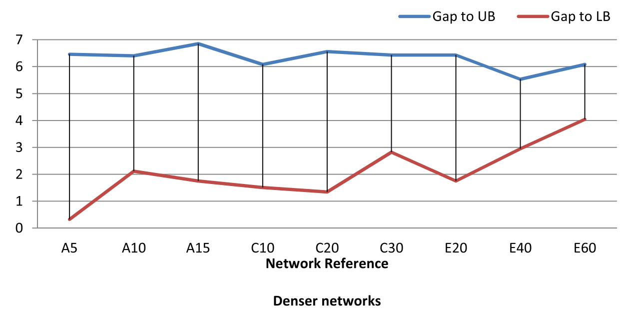

FIGURE 1. Average gap values: ζ^{ub} vs ζ^* (GapUB) and ζ^*_{lb} vs ζ^* (GapLB) for all network scenarios with denser networks.

The heuristic algorithm values are also very close to the lower bound ζ_{lb}, meaning that it is able to achieve a high efficiency under the $\mathcal{P}^*$ graph based on the shortest paths.

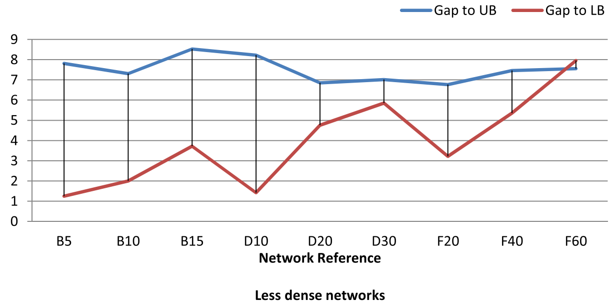

FIGURE 2. Average gap values: ζ^{ub} vs ζ^* (GapUB) and ζ^*_{lb} vs ζ^* (GapLB) for all network scenarios with less dense networks.

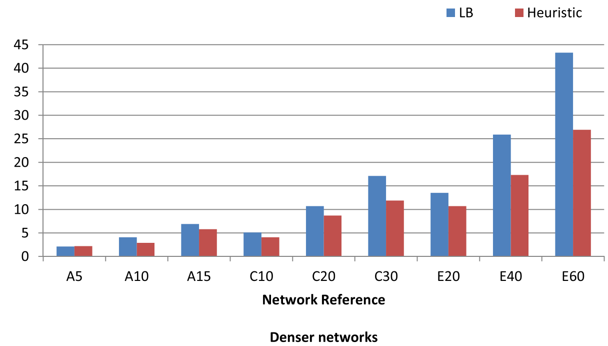

FIGURE 3. Average number of collections at ζ^* (heuristic feasible solution) and ζ_{lb} (lower bound on ζ^*), for all network scenarios with denser networks.

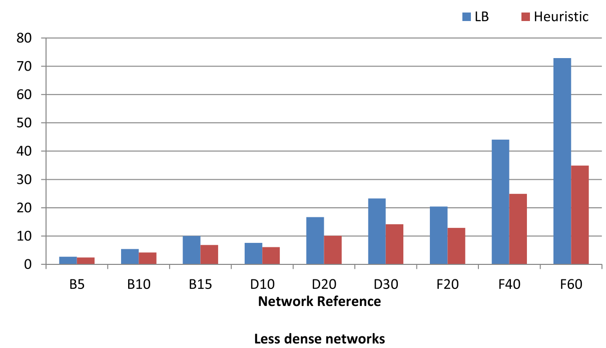

FIGURE 4. Average number of collections at ζ^* (heuristic feasible solution) and ζ_{lb} (lower bound on ζ^*), for all network scenarios with less dense networks.

The average gap values between ζ_{lb} and ζ^* range from 0% to 4.04% (denser networks) or to 7.95% (less dense networks). Generally, the best values were obtained for smaller networks and for smaller number of consumers. The worst values were obtained for larger and less dense networks. Nevertheless, Collections are always worthwhile and the heuristic has a good performance, given the complexity of trying to find an optimal solution for the CR-WSN problem. That is, the heuristic results are close to the lower bound. Consequently, they correspond to the optimal solution, for the considered shortest paths networks, or they are within 6 or 10 percent of the optimum, respectively, for denser or less dense networks.

2) RESULTS BY NUMBER OF COLLECTIONS

Figures 3 and 4 plot the average number of Collections found at feasible heuristic solutions, ζ^*, and the average number of Collections when Condition 1 (Collection Flow condition) is relaxed, ζ^*_{lb}. The solutions obtained by ζ^*_{lb} include, in general, more Collections than the ones obtained by ζ^*, as expected. The exception is for A5, where less chance to perform Collections exists due to the parameters involved (see Table 1). The largest differences occur for larger and less dense networks with more consumers. The ratios of the average number of collections obtained at feasible heuristic solutions to the lower bound are always bigger at denser networks (except for A10 and C10). Therefore, in less dense networks, there are more arcs where Collection Flow condition is not fulfilled, leading to solutions farther from the lower bound.

In our results (except for A and B) the average number of subjects per Collection is always smaller at solutions obtained by the heuristic (ranged from 2.0 to 2.22 subjects), when compared with ζ^*_{lb} solutions (ranged from 2.0 to 2.38 subjects).

VI. CONCLUSION

This article presents a heuristic algorithm for the routing of sensor data while planning for the creation of Collections. Such Collections might be built to reduce the overall number of network notifications, reducing the overall overhead information for a certain amount of data to be delivered, leading to better use of bandwidth and higher energy savings. Results show that Collections are always worthwhile, and that the heuristic obtains optimal or near optimal solutions (within 10 percent of the optimum), for the considered shortest paths networks.

REFERENCES

[1] P. Chatterjee, S. C. Ghosh, and N. Das, "Load balanced coverage with graded node deployment in wireless sensor networks," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 3, no. 2, pp. 100–112, Apr./Jun. 2017.

[2] N. Correia, A. Mazayev, G. Schütz, J. Martins, and A. Barradas, "Resource design in constrained networks for network lifetime increase," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1611–1623, Oct. 2017.

[3] H. Wang, N. Agoulmine, M. Ma, and Y. Jin, "Network lifetime optimization in wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1127–1137, Sep. 2010.

[4] Z. Shelby, M. Vial, M. Koster, C. Groves, J. Zhu, and B. Silverajan, *Reusable Interface Definitions for Constrained Restful Environments*, document Internet-Draft draft-ietf-core-interfaces-10, Working Draft, IETF Secretariat, Sep. 2017. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-core-interfaces-10.txt

[5] Z. Shelby, *Constrained RESTful Environments (CoRE) Link Format*, document RFC 6690, Aug. 2012. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6690.txt

[6] Z. Shelby, M. Koster, C. Bormann, P. Van der Stok, and C. Amsuess, *Core Resource Directory*, document Internet-Draft draft-ietf-core-resource-directory-12, Working Draft, IETF Secretariat, Oct. 2017. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-core-resource-directory-12.txt

[7] Z. Shelby, K. Hartke, and C. Bormann, *The Constrained Application Protocol (CoAP)*, document RFC 7252, Jun. 2014. [Online]. Available: http://www.rfc-editor.org/rfc/rfc7252.txt

[8] K. Hartke, *Observing Resources in the Constrained Application Protocol (CoAP)*, document RFC 7641, Sep. 2015.

[9] F. A. Onat and I. Stojmenovic, "Generating random graphs for wireless actuator networks," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2007, pp. 1–12.

GABRIELA SCHÜTZ received the five-year degree in mathematics and M.Sc. degree in statistics and operations research from the Faculty of Sciences, University of Lisbon, in 1985 and 1989, respectively, and the Ph.D. degree in mathematics-operations research from the University of Algarve, in 2000. She was a Lecturer with the Department of Mathematics, University of Aveiro, from 1985 to 1986, and also with the Department of Mathematics, Institute of Economics and Management, Technical University of Lisbon, from 1986 to 1990; an Invited Professor with the Polytechnic Institute of Portalegre, from 1990 to 1991; and an Adjunta Professor with the School of Technology, University of Algarve, from 1991 to 2001. Since 2001, she has been a Coordenadora Professor with the School of Engineering, University of Algarve. Her research activity is focused on combinatorial optimization, mainly optimization problems in networks, pickup and delivery problems, and assignment and scheduling problems. She is a member of the Center for Electronic, Optoelectronic and Telecommunication, University of Algarve.

NOELIA CORREIA received the B.Sc. and M.Sc. degrees in computer science and Ph.D. degree in optical networks (computer science) from the University of Algarve, Faro, Portugal, in 1995 and 1998, respectively, and the Ph.D. degree in optical networks (computer science) from the University of Algarve, in collaboration with University College London, London, U.K., in 2005. She is currently a Lecturer with the Faculty of Science and Technology, University of Algarve. Her research interests include the application of optimization techniques to several network design problems, in the optical, wireless, and sensor networks fields, and development of algorithms. She is a founding member of the Center for Electronics, Optoelectronics and Telecommunications, University of Algarve, a research center supported by the Portuguese Foundation for Science and Technology. She is also the Networks and Systems Group Coordinator.

JAIME A. MARTINS (M'87) received the five-year degree in electronics and telecommunications, four-year degree in clinical psychology, and Ph.D. degree in computer science (in 3-D face and object recognition) from the University of Algarve, in 2001, 2007, and 2014, respectively. He is currently a full-time Researcher with the Networks and Systems Group, Center for Electronics, Optoelectronics and Telecommunications, a research center supported by the Portuguese Foundation for Science and Technology. His major interests span several emerging IT areas, from the Internet of Things to computer/human vision, and cognitive neuroscience.

ANDRIY MAZAYEV received the B.Sc. and M.Sc. degrees in computer science from the University of Algarve, Faro, Portugal, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree in wireless sensor networks. He is currently a collaborator of the Networks and Systems Group, Center for Electronics, Optoelectronics and Telecommunications. His major research interests include the Internet of Things, semantic web, web protocols, ubiquitous computation, and wireless sensor networks in general.

ALVARO BARRADAS received the four-year first degree in computer science and Ph.D. degree in electronics engineering and computing from the University of Algarve, Faro, Portugal, in 1995 and 2009, respectively. His major field of study has concerned load balancing for optical networks. He is currently an Assistant Professor with the Faculty of Science and Technology, University of Algarve. His previous professional experience includes two years working on systems test for computer science and telecommunications consulting companies in Lisbon and Dusseldorf. His current research interest includes the study and simulation of hybrid networks, where QoS provisioning is challenging. He is a member of the Center of Electronics Optoelectronics and Telecommunications, University of Algarve, a research center supported by the Portuguese Foundation for Science and Technology.