Query range problem in wireless sensor networks

Bing Han, Jimmy Leblet, Gwendal Simon

Query Range Problem in Wireless Sensor Networks

Bing Han, Jimmy Leblet and Gwendal Simon

***Abstract*—Wireless sensor networks with multiple users extracting data directly from nearby sensors have many potential applications. An important problem in such a network is how to allocate the multi-hop query range for each user such that a certain global optimality is achieved. We introduce this problem and show it is NP-complete in its generic form. Distributed heuristic is proposed and evaluated with simulations. Interesting behaviors of the network when optimized with different global objectives are observed from the simulation results.**

***Index Terms*—wireless sensor networks, optimization, NP-complete**

I. Introduction

Wireless Sensor Network (WSN) is believed to be helpful during emergency situation [1]. A typical example for WSN emergency application considers a WSN deployed in a fire site where some firemen are in operation, each fireman being equipped with a device which is able to gather crucial information from the WSN, *e.g.* explosives nearby or surviver found. Firemen send requests to and collect data from the sensors around them by multi-hop wireless communication. Since firemen are generally more interested in what happens nearby, it is beneficial to interact *directly* with sensors around them, instead of via an infrastructure. Sensors spend their bandwidth for queries, either for sending their own data or for forwarding data from other sensors. Many works have dealt with such a context, for instance routing mechanisms supporting efficient data collection to multiple users [2].

On one hand, it is more important for each fireman to know what is happening around himself rather than to have a global knowledge, while on the other hand, each fireman would prefer to enlarge his local view thus to increase his personal awareness. Obviously, if a sensor is queried by many users, it may experience congestion. Packets dropped due to congestion not only waste energy but also generate blind spots in the queried area of related firemen. Thus, a natural requirement is that each user sets a proper query range to both avoid congestion and achieve a global optimality at the same time. We introduce here this *query range problem* with two commonly considered optimization objectives: fairness and maximization.

A basic requirement of the query range problem is congestion control which has been studied in WSN in recent years. Most proposed solutions have focused on the transport layer [3], and dealt with providing fairness for sources (sensors) *i.e.* allocating for each sensor a fair amount of bandwidth [4]. In contrast, we investigate both maximality and fairness objectives *in favor of users*, *i.e.* allocating for each user a proper query range to achieve global optimality. We emphasize that it is not necessary to cover every sensor with at least one query, instead, data from sensors within a reasonable query range should not be dropped due to congestion. As a consequence, some sensors far from any fireman do not have to generate data. The idea behind this is obvious: only when a fireman is near to a certain position, the data from this position is meaningful, under the considered application scenario. Similar query range problem has been studied in our previous works [5] with continuous query radius and in [6] with hop-based queries in a ZigBee tree based WSN. Note that the continuous version of this problem is considerably easier than the hop-based version. Besides, special properties of ZigBee tree structure has been exploited to estimate traffic in [6]. While in this paper, we investigate hop-based query range problem in a generic network where there is no obvious way to estimate traffic other than measuring it. We also prove NP-completeness of the query range problem and highlight the impact of different optimization objectives on the network.

We confine our study with the following assumptions: (i) Each fireman tends to set its query range as large as possible. The query range is measured by hop numbers such that a k hop query range will cover all k hop neighbors of the user. (ii) Shortest path routing is assumed and users do not forward data for sensors. (iii) The requested data reporting rate should at least be equal to a predefined threshold in order to detect certain real time events and no in-network data aggregation or compression is employed.

The rest of the paper is organized as follows. We formulate the query range problem in Section II, prove its NP-completeness in Section III. Then a distributed heuristic is proposed in Section IV. Simulation results are presented in Section V and we conclude the paper in Section VI.

II. Model and Problem Formulation

We consider a set $\mathcal{V}$ of n sensors and a set $\mathcal{M}$ of m users. The communication graph is defined as $G = (\mathcal{V} \cup \mathcal{M}, E)$ with $\mathcal{V} \cup \mathcal{M}$ the node set and E the link set. There is a link between two nodes of $\mathcal{V} \cup \mathcal{M}$ if they are within the wireless transmission range of each other. The available bandwidth r_i of a sensor is assumed to be the highest achievable shared bandwidth seen by the application. Accordingly, the congestion state is identified when data transmission and reception rate sum up to more than available bandwidth.

A *query* is initiated by a user p and sent to all sensors within $u(p)$ hops around p. $u(p)$ will also be referred to as the query range of user p. Queried sensors will generate data at a certain constant rate in response to the query of p. A *configuration* is the query ranges chosen by all users, noted as $C = \{u(p) :$

B. Han is with the Institut TELECOM; TELECOM Bretagne and TELECOM ParisTech, France; State Key Laboratory of Networking & Switching Technology, Beijing University of Posts and Telecommunications, China.

J. Leblet and G. Simon are with the Institut TELECOM; TELECOM Bretagne, France.

$\forall p \in \mathcal{M}\}$. A configuration is *feasible* when the bandwidth expenditure on each sensor is less than its available bandwidth. Our goal is to determine the set of feasible configurations which optimize a given global objective.

This problem can be formulated into a well-known Multi-dimension Multi-choice Knapsack Problem (MMKP) with generalized optimization objectives in the following way. Let each possible query range of user p correspond to an item to be selected and the value of the items is the query range, then we have $u(p) \in [1, d_G]$ where d_G is the diameter of G. Since each user sets its query range to a particular value at a certain time, the items could be seen as grouped into m classes each corresponding to a certain user. A binary variable x_{pu} is then associated with user p where $x_{pu} = 1$ indicates that query range u is selected and $x_{pu} = 0$ otherwise, with $u \in [1, d_G]$. The bandwidth consumption of a sensor i by a query of user p when p takes each of its possible query range level will be referred to as a vector $\{r_{ip1}, r_{ip2}, \ldots, r_{ipd_G}\}$ and is mapped to the ith dimension of weight of items in class p. Finally, each sensor forms a constraint dimension with its available bandwidth r_i.

The general MMKP is formulated as follows:

$$\begin{aligned} &\text{Achieve:} && \text{General Objective} && \\ &\text{Subject to:} && \textstyle\sum_{p\in\mathcal{M}} \sum_{u=1}^{d_G} r_{ipu} x_{pu} \le r_i,\ i \in \mathcal{V} && (1) \\ &&& \textstyle\sum_{u=1}^{d_G} x_{pu} = 1,\ p \in \mathcal{M} && (2) \\ &&& x_{pu} \in \{0,1\},\ p \in \mathcal{M},\ u \in [1, d_G] && (3) \end{aligned}$$

The general objective could be either maximizing the total query range of users $\sum_{p\in\mathcal{M}} u(p)$ (MNU), or give each user a lexicographical max-min fair (MMF) query range. For a complete definition of MMKP and MMF, see [7] and [8]. In the following, we refer to the problems formulated above as MMKP-MNU and MMKP-MMF. Both problems are intuitively hard to solve as they are related with MMKP. Actually, it is well known that the traditional MMKP and MMF problems are NP-complete. But it is not sufficient to prove the NP-hardness of MMKP-MNU and MMKP-MMF, because the traffic pattern imposed by the queries brings strong correlation between weights and profits and between weights across multiple dimensions. Therefore, both problems are strict sub-cases of MMKP. As a result, a formal proof is necessary.

III. NP-Completeness of MMKP-MNU

We prove the corresponding decision problem of MMKP-MNU to be NP-complete. For the MMKP-MMF problem, a similar result can be obtained by techniques used in [9].

Decision Problem of MMKP-MNU

INSTANCE: A graph $G = (\mathcal{V} \cup \mathcal{M}, E)$ such that $E \cap (\mathcal{M} \times \mathcal{M}) = \emptyset$, a weight capacity function $w : \mathcal{V} \to \mathbb{R}^+$, for each node $v \in \mathcal{V}$ and for each point $m \in \mathcal{M}$ a path $p(v, m)$ in $G_m = G[\mathcal{V} \cup \{m\}]$ of length $d(v, m)$ and a positive integer $K \in \mathbb{N}$.

QUESTION: Is there a radius function $r : \mathcal{M} \to \mathbb{N}$ such that:

(i) $\sum_{m\in\mathcal{M}} r(m) \ge K$, *i.e.* the sum of radius is at least K,

(ii) for every $v \in \mathcal{V}$ we have that the number of data forwarding path which go through the node v is lower or equal to the weights of this node, *i.e.* $|\{(x, m) \in \mathcal{V} \times \mathcal{M} : d(x, m) \le r(m) \text{ and } v \in p(x, m)\}| \le w(v)$.

Theorem 1: MMKP-MNU is NP-Complete.

Proof: We reduce MMKP-MNU to independent set problem [10]. Given a graph $G = (V, E)$ and a positive integer $K \le |V|$, let $\mathcal{M} = V$, $\mathcal{V} = E \cup \{\alpha\}$ and let $G' = (\mathcal{V} \cup \mathcal{M}, E')$ where $E' = \{\{x, e\} : x \in V, e \in E : x \in e\} \cup \{\{e, \alpha\} : e \in E\}$. G' is the incidence graph of G which we add a vertex α connected to every edge of G. Notice that we have $E' \cap (\mathcal{M} \times \mathcal{M}) = \emptyset$. We define the weight capacity w as $w(v) = 1$ for every $v \in \mathcal{V}$. For any edge $e \in E$ and for any vertex $u \in V$, we define the path $p(e, u) = [e, u]$ if $u \in e$, otherwise we take any shortest path from e to u in $G'[\mathcal{V} \cup \{u\}]$ as $p(e, u)$. Note that we have $\alpha \in p(e, u)$ if and only if $u \notin e$. Now, for any $x \in V$, we take an arbitrary $e \in E$ such that $x \in e$ and we define $p(\alpha, x)$ as $[\alpha, e, x]$. By construction, and as G has no isolated vertex, for any vertex $v \in V$ we have $d(\alpha, x) = 2$ and, for any edge $e \in E$ and for any vertex $v \in V$, it holds $d(e, v) = 1$ if and only if $x \in e$. We claim that G has an independent set of size at least K if and only if there exists a radius function $r : \mathcal{M} \to \mathbb{N}$ which fulfills the conditions (i) and (ii). ■

As a consequence of Theorem 1, no optimal query configuration can be obtained in large-scale dynamic networks within reasonable time limit. Possible countermeasures are either to use an inexact algorithm or to design a specific network structure on top of which the problem becomes solvable in polynomial time. We investigate the former approach and propose a distributed heuristic in this paper.

IV. Distributed Heuristic

We outline the basic idea of the distributed heuristic. Each sensor monitors the traffic it is handling and decides if a congestion will appear. If sensor i identifies a potential congestion state, it solves a local MMKP based on: (1) traffic measurement corresponding to each user whose query covers i; (2) bandwidth limitation of i; (3) the optimization objective, either MNU or MMF. Obviously, this local MMKP has only one dimension (which is the bandwidth of i) and very few number of classes (which are the users querying i). So an approximate solution can be quickly obtained by a sensor. Then the results are sent to the users and the users apply certain strategy to adapt their query range. A simple strategy could be to use suggested range only when it is smaller than its current query range, and try to increase periodically. Many existing algorithms could be employed in solving the local MMKP with MNU, but considering the computation capability of the sensor nodes, simple inexact algorithms are preferred. Based on the same understanding, a greedy algorithm could be used for MMKP with MMF, *i.e.* increase query range for all related users by one level each round and stop when bandwidth limitation is reached.

V. Simulation evaluation

We implemented and evaluated the heuristic in ns2. Local MMKP with MNU is solved with GLPK [11] solver and the MMF version is solved with the greedy algorithm mentioned

TABLE I
AVERAGE QUERY RANGE.

m	MNU			MMF		
	l_{DIS}	l_{OPT}	l_{UP}	l_{DIS}	l_{OPT}	l_{UP}
10	11.47	11.93	12.46	10.44	11.4	12.37
20	10.55	11.81	12.10	9.44	10.52	11.73
50	8.67	–	11.21	7.96	–	10.59
90	7.88	–	10.59	7.23	–	10.14

above. Note that it is obviously infeasible to use GLPK on a real sensor node, we use it in simulations only for convenience. Simulations are carried out in a network with 1000 sensors and $10, 20, 50$ and 90 users. Both sensors and users are uniformly deployed in a $800\text{m} \times 800\text{m}$ square and the transmission range of each node is set to 30m. Shortest paths to users are setup at the same time when the queries are disseminated.

The distributed heuristic is first evaluated against a branch-and-bound exact algorithm. The effectiveness of the heuristic could be verified from Table I where heuristic solution, exact solution and linear programming upper bound are presented.

Next we investigate how users are distributed in the query range domain. As shown in Fig. 1, the height of the bar in the figure represents the ratio of users that have the corresponding query range. As expected, MMF results in a more concentrate distribution while MNU let more users enjoy a maximal query range. More interestingly, as the number of users grows, the distribution tends to concentrate first instead of shifting leftwards and this holds for both MNU and MMF although it is more obvious for the latter. This phenomenon implies that the network under local query model tries to allocate medium query range for users when the number of users grows.

The distribution of sensors according to the number of users querying them are demonstrated in Fig. 2. We can observe that the distribution is more concentrate for MNU than for MMF, which is indicated by higher and narrower curves. When MMF is applied in favor of users, it gives fairness to each user but results in unfair query coverage on the sensors. In contrast, MNU optimization gives maximum overall query range at the cost that users are served unfairly, but the resulted configuration covers the sensors more evenly. Note that all sensors are not covered *e.g.* over 20% of sensors have no query in 10-user case. Fig. 2 mainly reveals that query range problem have unique characteristics other than commonly studied full coverage or k-connectivity problems.

VI. CONCLUSION

We have introduced the query range problem which deals with the cooperation of users and sensors when setting their query ranges in order to optimize certain global objective and to avoid congesting the sensors. This problem reveals, in its theoretical aspect, that heuristic algorithms exploiting special underlying network structure would be of interest due to the NP-completeness of the problem; while in its practical aspect, that optimization in favor of users in WSN has great importance when multiple users present in the network as in the fireman scenario that has motivated this study.

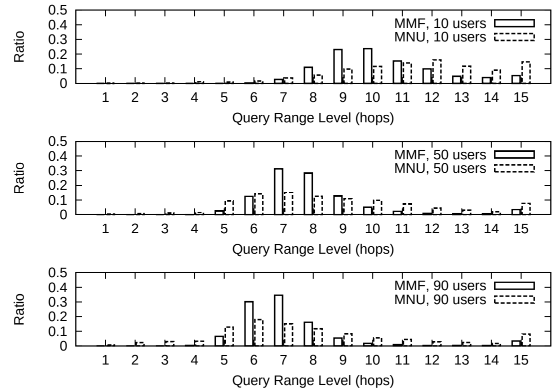

Fig. 1. Distribution of users at each query range level.

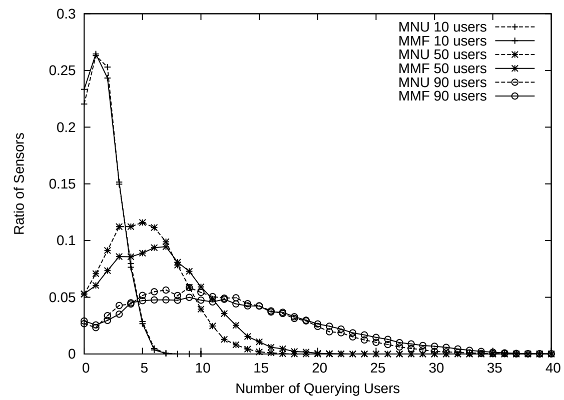

Fig. 2. Sensor distribution on the number of querying users.

REFERENCES

[1] K. Lorincz, D. Malan, T. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, M. Welsh, and S. Moulton, "Sensor networks for emergency response: challenges and opportunities," *Pervasive Computing, IEEE*, vol. 3, no. 4, pp. 16–23, Oct.-Dec. 2004.

[2] K. il Hwang, J. In, and D. S. Eom, "Distributed dynamic shared tree for minimum energy data aggregation of multiple mobile sinks in wireless sensor networks," in *Proc. of EWSN Conf.*, ser. LNCS, vol. 3868, Jan. 2006, pp. 132–147.

[3] C. Wang, K. Sohraby, B. Li, M. Daneshmand, and Y. Hu, "A survey of transport protocols for wireless sensor networks," *IEEE Network*, vol. 20, no. 3, pp. 34–40, 2006.

[4] S. Rangwala, R. Gummadi, R. Govindan, and K. Psounis, "Interference-aware fair rate control in wireless sensor networks," in *ACM SIGCOMM*, 2006.

[5] B. Han and G. Simon, "Fair capacity sharing among multiple sinks in wireless sensor networks," in *Proc. of the IEEE MASS Conf.*, Oct. 2007, pp. 1–9.

[6] ——, "Optimizing multi-hop queries in zigbee based multi-sink sensor networks," in *Proc. of Int. Conf. on Distributed Computing and Networking (ICDCN'09)*, to appear.

[7] S. Khan, K. F. Li, E. G. Manning, and M. M. Akbar, "Solving the knapsack problem for adaptive multimedia systems." *Stud. Inform. Univ.*, vol. 2, no. 1, pp. 157–178, 2002.

[8] W. Ogryczak, M. Pioro, and A. Tomaszewski, "Telecommunications network design and max-min optimization problem," *Journal of Telecom. and Information Tech.*, vol. 3, pp. 43–56, 2005.

[9] S. Sarkar and L. Tassiulas, "Fair allocation of discrete bandwidth layers in multicast networks," in *Proc. of the IEEE INFOCOM Conf.*, vol. 3, Mar. 2000, pp. 1491–1500.

[10] M. R. Garey, D. S. Johnson, and L. Stockmeyer, "Some simplified np-complete problems," in *Proc. of ACM Symp. on Theory of computing (STOC)*, 1974, pp. 47–63.

[11] http://www.gnu.org/software/glpk/.