

Convergecast Scheduling Problem in Case of Given Aggregation Tree

The complexity status and some special cases

Adil Erzin, and Artem Pyatkin
Sobolev Institute of Mathematics,
Siberian Branch, Russian Academy of Sciences;
Novosibirsk State University;
Novosibirsk, Russia
adilerzin@math.nsc.ru, artem@math.nsc.ru

Abstract—We consider the problem of conflict-free data aggregation in distributed wireless radio networks, that is known as a Convergecast Scheduling Problem. This problem is strongly NP-hard in general. However the complexity status of the problem remained open in the case when the aggregation tree was given. We have proved that in this case, the problem is also NP-hard. Additionally, we proposed several new approximate suboptimal solutions in the special cases of the problem on the grid graphs.

Keywords—conflict-free data aggregation; NP-completeness; approximation; wireless sensor network

I. INTRODUCTION

Some networks, for example, wireless sensor networks (WSN), use common radio frequency for data exchange. Thus, if in the receiving zone of a receiver there is more than one transmitter, then (due to interference phenomenon) the receiver cannot get any correct message. Such situation is called a *conflict*.

In WSN the data collected by sensors must be regularly delivered to the base station (the root node of the network). For this purpose a communication graph (CG) is used, which is usually constructed basing on the efficiency criterion of the transmission energy [1, 2]. As a result, not all nodes can be connected directly to the root. Some nodes could transmit data in transit via others. Although it is sufficient to use a spanning tree as a communication graph, in practice the power of transmitters generates quite arbitrary communication graph. And this means that conflicts are inevitable.

By the reasons of energy efficiency, each node sends the data only once during the round of data aggregation. Hence it is necessary to determine a rooted tree in the communication graph such that the data will be transmitted to the root along this tree. Such tree is called an *aggregation tree* (AT). Since each node sends data only once, a node must first collect the data packets from all its predecessors (children). Sending data takes one time slot for each vertex. Then we have a partial order on the set of arcs in the CG. Namely, the arcs of AT are directed to the root, while the other edges of the communication graph are undirected. In *Convergecast*

Scheduling Problem (CSP) it is necessary to find the shortest conflict-free aggregation schedule [3-11]. Its minimum length is called an *aggregation latency* or an *aggregation length of the schedule*. In CSP the aggregation tree has to be found. Several heuristic algorithms are known to generate AT [1-3, 6-8, 10, 11]. But even when AT is given (in this case we refer to the problem as CSP_T), the problem remains hard.

It is obvious that the problem CSP_T reduces to the mixed graph coloring problem (MGCP) [12], which is NP-hard [13] and can be formulated as follows. Let $G = (V, A \cup E)$ be a mixed graph with the set of nodes V , the set of arcs A , and the set of edges E . A k -coloring of a graph G is a function $f: V \rightarrow \{1, \dots, k\}$. The MGCP is finding the minimum k for which there exists a k -coloring such that if two vertices i and j are linked by the edge (i, j) , then $f(i) \neq f(j)$, and if there is an arc going from i to j then $f(i) < f(j)$.

Referring to MGCP, it is claimed in [3] that the CSP_T is also NP-hard. But this statement is not entirely correct. The matter is that in the problem CSP_T the mixed graph has a special structure induced by AT. In this paper we present correct proof of NP-hardness of CSP_T.

In some papers the problem CSP was considered on the special grid graphs [14-16]. Thus in [14] the aggregation is performed on the rectangular grid with the unit distance transmission. It is easy to prove that in this case the CSP is polynomially solvable. In [15, 16] attempts to use a grid structure for finding an AT were made.

In this paper we consider the cases of rectangular grid with non-unit distance transmission, and triangular and hexagonal grids as well. In some cases we propose an optimal schedule, in other – suboptimal one.

The paper is organized as follows. In the next section a formulation of the problem is given. A proof of NP-completeness is presented in the section 3. Section 4 contains the results for the special case when the communication graph is a rectangular grid. In the section 5 we consider the triangular grid, and in the section 6 a hexagonal grid is studied. Section 7 concludes the paper.

II. PROBLEM FORMULATION

CSP_T can be formulated as the following problem of tree coloring with bans.

Problem CSP_T. The aggregation tree $T = (V, A)$, the set of bans $M \subseteq A^2$ and the positive integer K are given. Is there a coloring of arcs $f: A \rightarrow \{1, \dots, K\}$ such that the following conditions are met:

1. If two arcs e_1 and e_2 have the same head vertex or $(e_1, e_2) \in M$, then $f(e_1) \neq f(e_2)$;
2. If the tree T contains a directed path from the head of the arc e_1 to the tail of the arc e_2 , then $f(e_1) < f(e_2)$?

If M is empty, then the problem can be solved easily in time $O(|V|)$. Otherwise, it is NP-complete, and we prove this in the next section.

III. NP-COMPLETENESS

Theorem 1. Problem CSP_T is NP-complete.

Proof. Let us define a rooted tree T_n as follows. Consider disjoint directed paths P_1, \dots, P_n , where i is the number of arcs in the path P_i . Identify the final vertices of these paths and denote the resulting tree by T_n .

Using induction on n , it is easy to prove the next

Property 1. Tree T_n can be colored with n colors, and moreover in any its n -coloring all pending arcs should be colored 1.

Let us construct a polynomial reduction of NP-complete vertex 3-coloring problem [17] to the CSP_T. Let $G = (V_0, E)$ be an instance of a graph on n nodes whose 3-colorability should be verified. Consider the tree T_n and add three nodes a_i, b_i and c_i for each leaf node i in the tree T_n . Connect each of them with the leaf i by the arcs $e(a_i), e(b_i)$ and $e(c_i)$ ingoing to i . Denote the constructed tree with new nodes and new arcs as T (see Fig. 1). Put $M = \{(e(a_i), e(a_j)) \mid (v_i, v_j) \in E\}$ and $K = n + 3$. Let us prove that for the described T, M and K the coloring in CSP_T exists if and only if the graph G is 3-colorable.

Suppose G is 3-colorable, i.e. the set of vertices V_0 can be partitioned into three independent sets. Then the set of arcs $\{e(a_1), \dots, e(a_n)\}$ also can be partitioned into three subsets A_1, A_2 and A_3 such that none of these subsets contains a ban in M . Denote by B_1, B_2, B_3 and C_1, C_2, C_3 the corresponding partitions of the sets of arcs $\{e(b_1), \dots, e(b_n)\}$ and $\{e(c_1), \dots, e(c_n)\}$ respectively. Let

$$f(e) = \begin{cases} 1, & \text{if } e \in A_1 \cup B_2 \cup C_3; \\ 2, & \text{if } e \in A_2 \cup B_1 \cup C_3; \\ 3, & \text{if } e \in A_3 \cup C_1 \cup C_2. \end{cases}$$

Let g be a coloring of the arcs in the tree T_n by n colors, that exists according to the Property 1. Put $f(e) = g(e) + 3$ for each arc e in T_n . Evidently, the coloring f satisfies both conditions of CSP_T.

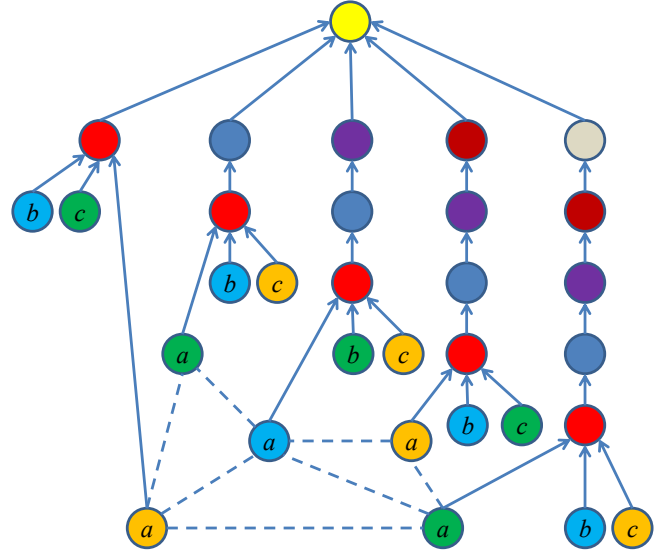


Fig. 1. Illustration to the proof of the Theorem.

Suppose now that the tree T is colorable by $K = n + 3$ colors. Let $e_i = (u, v)$ be a pending arc in the tree T_n . Then it has three preceding arcs $e(a_i) = (a_i, u)$, $e(b_i) = (b_i, u)$ and $e(c_i) = (c_i, u)$ in the tree T . Since all these arcs have the head u , their colors should be different. Therefore, the color of the arc e_i cannot be less than 4. Then it follows from the Property 1 that the color of all pendant arcs in the tree T_n is 4, and all arcs in the set $\{e(a_1), \dots, e(a_n)\}$ are colored by three colors. Since the set of arcs colored by the same color cannot contain any ban, the corresponding subsets of the nodes in graph G should be independent. Hence, the 3-coloring of G exists. Theorem 1 is proved.

Assume further that the AT is not specified and consider the special cases of CSP when the communication graph is a grid graph.

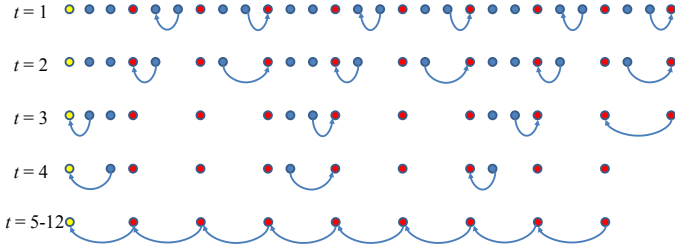
IV. CSP ON RECTANGULAR GRID

Denote by d the transmission distance. This means, in particular, that if a node x receives a message from a node y then no other node at distance at most d from x (in the rectangular metric l_1) may transmit information anywhere at the same time slot. Suppose $d \geq 1$. Moreover, suppose further (for simplicity) that $m = k \cdot d$ and $n = l \cdot d$, where k and l are positive integers.

Let us consider the case when all sensors are located on the line (Fig. 2) first. It is easy to prove the following

Lemma 1. If $d \geq 1$ is the transmission distance, then the minimal aggregation latency for the linear graph $n \times 1$ with the root in the origin is at most $L(n, d) = n/d + d$.

Let us consider the rectangular grid $n \times m$. Each node of the grid must send its packet to the root. In [14] a simple algorithm constructing an optimal aggregation schedule for the rectangular grid when transmission distance equals 1 is proposed.

Fig. 2. Aggregation in the linear graph ($d=3$).

Let transmission distance be $d \geq 2$ and (for simplicity) consider the grid with the root placed left-bottom (Fig. 3). On the one hand, if d is greater than 1, then each node can transfer the data to the receiver at (rectangular) distance up to d . On the other hand, transmission over long distances entails the appearance of new conflicts.

Trivial lower bound for the schedule length is $(n+m)/d$. The upper bound can be found using the next simple algorithm (when $d=1$ it matches with the algorithm in [14]). Send the data from all nodes with coordinates (x, m) , $x=0, \dots, n$, at maximal distance d . Then send the data from all nodes with coordinates $(x, m-1)$, $x=0, \dots, n$, at distance d . And so on. The last such transmission is sent from all nodes with coordinates (x, d) , $x=0, \dots, n$ (see Fig. 3a). The total number of these vertical transmissions is $m-d+1$.

Then we get a situation shown in Fig. 3b. At most $n-d+1$ time units are necessary for sending data horizontally from all nodes with coordinates (x, y) , $x=n, n-1, \dots, d$, $y=0, \dots, d-1$.

Then we get a situation shown in Fig. 3c. Two time units are sufficient for sending data from all nodes with coordinates $(d-1, y)$, $y=0, \dots, d-1$; three time units for sending data from all nodes with coordinates $(d-2, y)$, $y=0, \dots, d-1$ (Fig. 3d); and so on until we get situation shown in Fig. 3e. Additional $d-1$ time units are enough for delivering all the data to the root. The total number of the last transmissions is $(1+d)d/2 + d-2$.

So, the schedule length yielded by such (trivial) algorithm is $\bar{T}(n, m, d) = m + n + d^2/2 - d/2$. Then the ratio is

$$\frac{\bar{T}(n, m, d)}{(m+n)/d} = \frac{(2m+2n+d^2-d)d}{2(m+n)} = d + \frac{d^3-d^2}{2(m+n)},$$

which is greater than 2 when $d \geq 2$.

When $d \geq 2$ the aggregation time can be decreased if instead of the cells 1×1 the cells $d \times d$ are used. But it is not necessary to make all cells $d \times d$. It is sufficient to make the cells $1 \times d$ first and send the data packets from all nodes at one horizontal level (row) down during one time unit until all packets reach the lower row, and we get a linear graph.

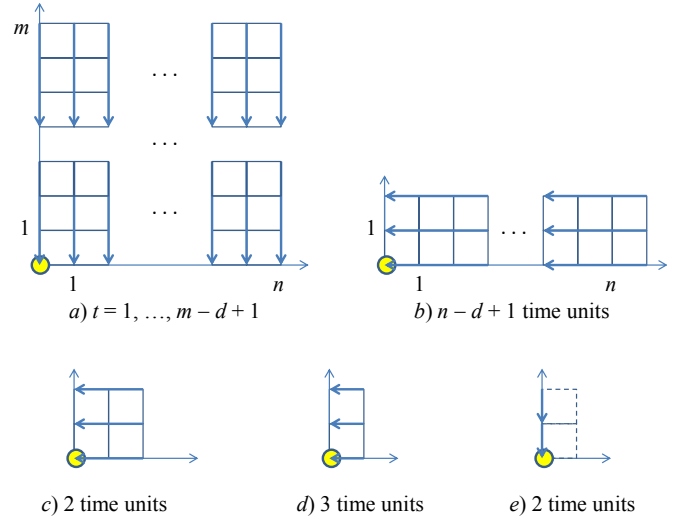


Fig. 3. Trivial aggregation on the grid graph when transmission distance equals 3; a) and e) vertical transmissions; b-d) horizontal transmissions.

Lemma 2. If $d=2$, then the minimal aggregation length of the schedule in the rectangular grid $n \times m$ (remind that m and n are even) with the root in the origin is at most $(n+m)/2 + 3$.

Proof. In order to accelerate the aggregation process on the grid, we need to send data from all nodes of one level at the maximal distance without conflicts. Consider the following algorithm for the case $d=2$.

Let us first change the size of cells from 1×1 to 1×2 . For this purpose during the first two time moments $t=1, 2$ all the nodes of level $m-1$ (with coordinates $(x, m-1)$, $x=0, \dots, n$) send up their packets to the corresponding nodes of level m . At the same time ($t=1, 2$) all the nodes of level 1 send packets up to the corresponding nodes of level 2. Moreover, at the moment $t=1$ all nodes of level $m-3$ send the data to the corresponding lower nodes at level $m-5$; and at the moment $t=2$ all nodes at level $m-5$ send the data to the lower nodes at level $m-7$ (Fig. 4a). As a result, we get three upper levels composed of the cells 1×2 (Fig. 4b).

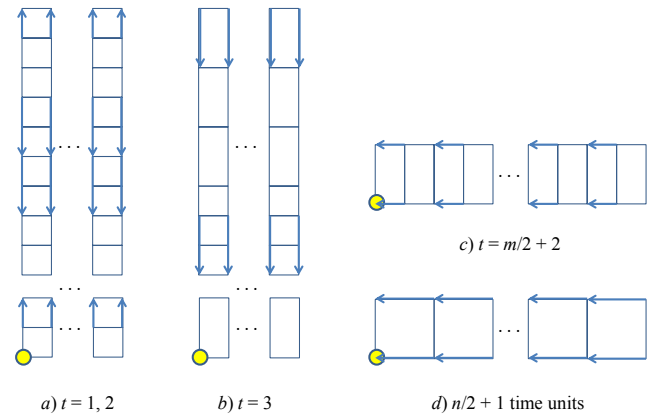


Fig. 4. Illustration to Algorithm A; a-b) vertical transmissions; c-d) horizontal transmissions.

Then all nodes of level m send the packets down to the corresponding nodes of level $m-2$. At the same time ($t=3$) the nodes of level $m-7$ send the data down at distance 2, making new cells 1×2 (Fig. 4b). Continue this process until we get the situation shown in Fig. 3c. It is time to make the 2×2 cells. For this purpose at the moment $t = m/2 + 2$ all the nodes at odd columns send the packets left to the root at distance 1 (Fig. 4c). Now all nodes at even columns can send successively from right to left the data at distance 2 (Fig. 4d).

Then the length of the constructed schedule is equal to $(n+m)/2 + 3$, and the proof is now complete.

Note, that the length of the constructed schedule is almost 2 times less than one yielded by the trivial algorithm for $d=2$.

Lemma 3. Let the grid have dimension $n \times d$, where $d \geq 2$ is the transmission distance. Then $T(d) = d + d^2/4 - 1$ time units are sufficient for transmitting the data packets vertically from all inner nodes (with coordinates (x, y) , $x = 0, \dots, n$, $y = 1, \dots, d-1$) to the nodes at the horizontal borders of the strip $n \times d$.

Proof. If transmission is performed down or up to the distance d , then all the elements of one row (with equal y -coordinate) can transmit packets simultaneously (conflict-free). If the transmission is carried out at a distance $d-1$, then two time rounds are necessary for all the elements of one row to transmit their packets (to avoid conflicts, every second node could transmit simultaneously). Similarly, if the data is sent at a distance $d-r$, it will take $r+1$ time units ($0 \leq r \leq d-1$).

So, we can organize transmission in the following way. During time rounds 1 and 2 all nodes with coordinates $(x, d-1)$, $x = 0, \dots, n$, send the packets down at a distance $d-1$. During the next two time rounds 3 and 4 all nodes $(x, 1)$, $x = 0, \dots, n$, send the packets up at a distance $d-1$. Then during the next 6 time rounds all nodes $(x, d-2)$, $x = 0, \dots, n$, send the packets down at a distance $d-2$, and all nodes $(x, 2)$, $x = 0, \dots, n$, send the packets up at a distance $d-2$. And so on. In Fig. 5 we show the sequence of transmissions when $d=8$.

If d is even ($d=2l$), then

$$T(d) = 2 \sum_{t=2}^l t + l + 1 = d + \frac{d^2}{4} - 1$$

time rounds are necessary to send the data vertically from all nodes with coordinates (x, y) , $x = 0, \dots, n$, $y = 1, \dots, d-1$.

If d is odd ($d=2l+1$), then

$$T(d) = 2 \sum_{t=2}^{l+1} t = d + \frac{d^2}{4} - \frac{5}{4}$$

time rounds are necessary to send the data vertically from all nodes with coordinates (x, y) , $x = 0, \dots, n$, $y = 1, \dots, d-1$.

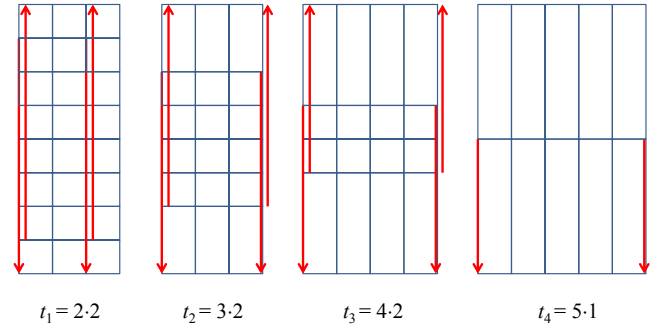


Fig. 5. Vertical aggregation ($d=8$).

In both cases, $T(d) \leq d + \frac{d^2}{4} - 1$, and lemma is proved.

Theorem 2. The minimum aggregation time in the rectangular grid $n \times m$ when the transmission distance is $d \geq 2$ is at most

$$\frac{m+n}{d} + 3d + \frac{d^2}{4} - 1.$$

Proof. Since all nodes of the same row can transmit vertically at a distance d simultaneously, it is easy to make during the d time rounds the strips of width d such that the distance between these strips is at least d . Here, to avoid any conflict, some nodes (on the same row) transmitted the data down and some upward. This results in the strips having height of d cells, with the distance is at least d between them. It follows from Lemma 3 that $T(d)$ time rounds are sufficient for sending the data packets from all inner nodes of such strips to the horizontal borders of the strips.

Then we get the cells $1 \times d$, and all nodes on the same rows $k \cdot d$, $k = 1, \dots, m/d$, can transfer the data at a distance d without conflict. As a result, after m/d time rounds all packets will arrive at the row 0, and we get a linear graph. So, the total aggregation time is

$$T(n, m, d) = d + T(d) + \frac{m}{d} + L(n, d) \leq 3d + \frac{d^2}{4} + \frac{m+n}{d} - 1,$$

and the theorem is proved.

Note that the ratio

$$\frac{T(n, m, d)}{(n+m)/d} \leq 1 + \frac{d^3 + 12d^2 - 4d}{4(n+m)}$$

tends to 1 under the unlimited growth of the dimension of the grid. Moreover, the length of the constructed schedule is

$$\frac{\bar{T}(n, m, d)}{T(n, m, d)} \geq d + \frac{d^3 - 14d^2 + 4d}{4m + 4n + d^3 + 12d^2 - 4d}$$

times less than the length of the schedule constructed by the trivial algorithm.

Suppose we can adjust the transmission range and choose the value of d in order to decrease the length of the schedule constructed as above (in the proof of the Theorem 2). The function $T(n, m, d)$ is convex with respect to d for any fixed dimension of the grid. Then it is possible to choose the best value for d , which can be used to decrease the length of the schedule constructed by our algorithm, by solving the equation

$$d^3 + 6d^2 - 2(m+n) = 0.$$

If, for example, $m = 100$ and $n = 100$, then the best d is equal to 5, and the length of the constructed schedule is at most 68. The trivial algorithm in this case yields the length of the schedule equal to $\bar{T}(100, 100, 5) = 210$.

V. CSP ON TRIANGULAR GRID

Let us briefly consider the triangular grid in the case when the transmissions distance $d = 1$ and the root is located in the middle of the grid (Fig. 6). The lower bound for the length of the aggregation schedule in this case is $R + 2$, where R is the radius of the grid graph (there are 3 nodes at distance R from the root and at distance $2R$ from each other).

The *level* of a node in the grid is the number of edges in the shortest path from this node to the root.

Proposition 1. If $d = 1$, then the aggregation length of the schedule in the triangular grid with radius R is at most $2R + 3$.

Proof. Let us first show that for packets transmission from level $k + 1$ to level k two time units are enough. Consider the AT in Fig. 6. Each node at level k has at most 2 adjacent nodes at level $k + 1$. Then the half of the vertices at level $k + 1$ can transmit packets simultaneously to the nodes at level k .

So, the data from all nodes at the levels $R, R - 1, \dots, 3$ can be delivered to the nodes at level 2 during $2(R - 2)$ time units.

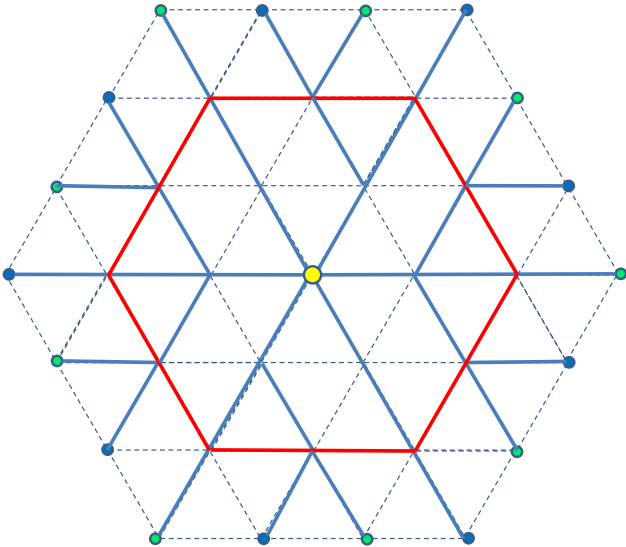


Fig. 6. Triangular grid and AT (blue bold lines).

The data from the nodes at levels 2 and 1 (inside the red hexagon in Fig. 6) can be delivered to the root during 7 time units. This can be proved by a simple brute-force. The proof is finished.

Then the ratio for the constructed schedule is

$$\frac{2(R-2)+7}{R+2} = 2 - \frac{1}{R+2}.$$

VI. CSP ON HEXAGONAL GRID

Levels in a hexagonal grid are defined similarly. Assume that there are q nodes at the level R , where R is the radius of the grid. We can partition the grid into the green and yellow zones and construct an aggregation sub-trees inside this zones (Fig. 7). Note that any two independent edges in one cell can be used for data transmission simultaneously.

Note that if the packets come from different zones to the neighbors of the root (red ones in Fig. 7) at the same time t , then the length of the schedule is $t + 3$.

If $q \geq 2$, then at most half of the nodes at the level R may transmit the packets simultaneously. Then at least R time units are necessary to deliver the data to the neighbors of the root.

The above arguments allows us stating the following

Proposition 2. If $q \geq 2$, then the minimum aggregation time in the hexagonal grid is at least R and at most $R + 3$.

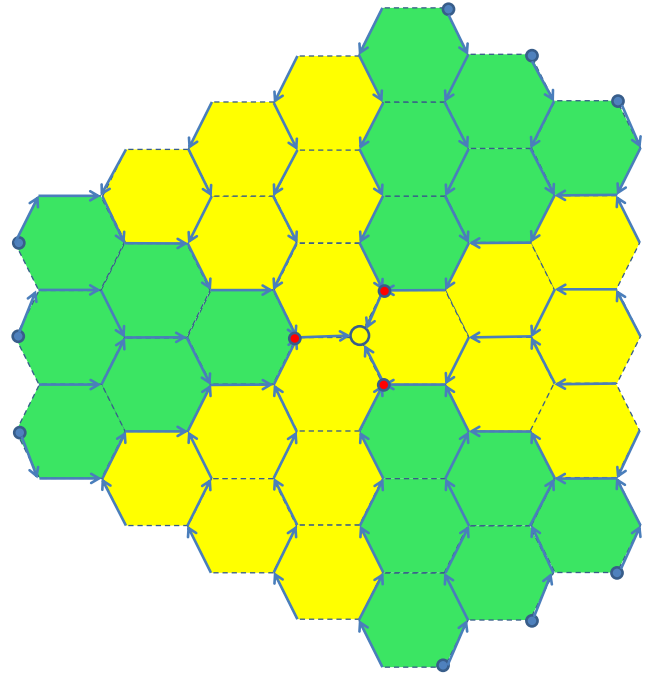


Fig. 7. Hexagonal grid and AT (blue bold arrows).

VII. CONCLUSION

In this paper, we have considered the problem of conflict-free data aggregation in the case of a given aggregation tree and proved NP-completeness of this problem. The complexity status of this problem was open heretofore.

We have considered also the special cases of the Convergecast Scheduling Problem on the rectangular, triangular and hexagonal grids and proposed the algorithms constructing the suboptimal conflict-free aggregation schedules. In [14] a simple algorithm was published, which yields an optimal schedule of data aggregation on a rectangular grid, when the transmission distance d is equal to 1. We have analyzed the case when the transmission distance d is greater than 1 and proposed the efficient procedures constructing the suboptimal solutions for $d \geq 2$. For an arbitrary transmission distance $d = \text{const} \geq 3$ the ratio

$$1 + \frac{12d^2 + d^3 - 4d}{4(n+m)}$$

tends to 1 under the unlimited growth of the dimension of the grid.

For the best of our knowledge, no efficient algorithm for the problem of data aggregation on triangular or hexagonal grids was known. We have proposed the simple polynomial algorithms constructing the solutions within the constant factor (suboptimal solutions).

Acknowledgment

This research is supported in part by Russian Foundation for Basic Research (grants 16-07-00552 and 15-01-00976) and the Ministry of Education and Science of the Republic of Kazakhstan (project No. 0115PK00550).

References

- [1] A.I. Erzin, R.V. Plotnikov, and Yu.V. Shamardin, "On some polynomially solvable cases and approximate algorithms in the optimal communication tree construction problem", *J. of Applied and Industrial Mathematics*, vol. 7, No. 2, pp. 142-152, 2013.
- [2] A.I. Erzin, and R.V. Plotnikov, "Using VNS for the optimal synthesis of the communication tree in wireless sensor networks", *Electronic Notes in Discrete Mathematics*, vol. 47, pp. 21-28, 2015.
- [3] E. de Souza, and I. Nikolaidis, "An exploration of aggregation convergecast scheduling," *Ad Hoc Networks*, vol. 11, pp. 2391-2407, June 2013.
- [4] O.D. Incel, A. Ghosh, B. Krishnamachari, and K. Chintalapudi, "Fast data collection in tree-based wireless sensor networks", *IEEE Trans. On Mobile Computing*, vol. 11, No. 1, pp. 86-99, January 2012.
- [5] P. Wang, Y. He, and L. Huang, "Near optimal scheduling of data aggregation in wireless sensor networks," *Ad Hoc Networks*, vol. 11, pp. 1287-1296, June 2013.
- [6] C.-T. Cheng, C.K. Tse, and F.C.M. Lau, "A delay-aware data collection network structure for wireless sensor networks", *IEEE Sensors J.*, vol. 11, No. 3, pp. 699-710, March 2011.
- [7] C.-T. Cheng, and C.K. Tse, "A delay-aware network structure for wireless sensor networks with consecutive data collection process", *IEEE Sensors J.*, vol. 13, No. 6, pp. 2413-2422, June 2013.
- [8] D. Li, J. Cao, and Q. Zhu, "Approximation algorithm for constructing data aggregation tree for wireless sensor networks", *Front. Comput. Sci. China*, vol. 3, No. 4, pp. 524-534, 2009.
- [9] S. Xiao, J. Huang, L. Pan, Y. Cheng, and J. Liu, "On centralized and distributed algorithms for minimizing data aggregation time in duty-cycled wireless sensor networks", *Wireless Netw.*, vol. 20, pp. 1729-1741, February 2014.
- [10] B. Malhotra, I. Nikolaidis, and M.A. Nascimento, "Aggregation convergecast scheduling in wireless sensor networks", *Wireless Netw.*, vol. 17, pp. 319-335, 2011.
- [11] R. Gandhi, M.M. Halldórsson, C. Konrad, G. Kortsarz, and H. Oh, "Radio aggregation scheduling", *LNCS*, book "Algorithms for sensor systems", vol. 9536, pp. 169-182, Jan. 2016.
- [12] P. Hansen, J. Kuplinsky, and D. de Werra, "Mixed graph colorings", *Mathematical Methods of Operations Research*, vol. 45, pp. 145-160, 1997.
- [13] R. Ries, "Coloring some classes of mixed graphs", *Discrete Applied Mathematics*, vol. 155, pp. 1-6, 2007.
- [14] J. Gagnon, and L. Narayanan, "Minimum latency aggregation scheduling in wireless sensor networks", *11th Int. Symp. On Algorithms and Experiments for Sensor Systems, Wireless Networks and Distributed Robotics*, Wroclaw, Poland, pp. 152-168, September 2014.
- [15] Y.-K. Chiang, N.-C. Wang, and C.-H. Hsieh, "A cycle-based data aggregation scheme for grid-based wireless sensor networks", *Sensors*, vol. 14, pp. 8447-8464, May 2014.
- [16] T. Yu, S. Liu, and Z. Zhang, "A cluster-based uneven grid data aggregation in wireless sensor networks", *J. of Computational Information Systems*, vol. 11, No. 19, pp. 7055-7062, 2015.
- [17] M.R. Garey, and D.S. Johnson, *Computers and intractability: A guide to the theory of NP-completeness*, W.H. Freeman, 1979.