# Data Dissemination Problem in Wireless Networks


**Ivo Kubjas and Vitaly Skachek**

Institute of Computer Science, University of Tartu

ivokub@ut.ee, vitaly@ut.ee


## 1. Data Dissemination Problem

**Generalization of the index coding and the data exchange problem to networks with arbitrary topology and requests.**

**Example 1** *There are five nodes, which in total possess three bits of information $x_1, x_2, x_3$. If node 1 transmits $x_1+x_2$ and node 2 transmits $x_2+x_3$, then the requests of all nodes will be satisfied with only two transmissions.*

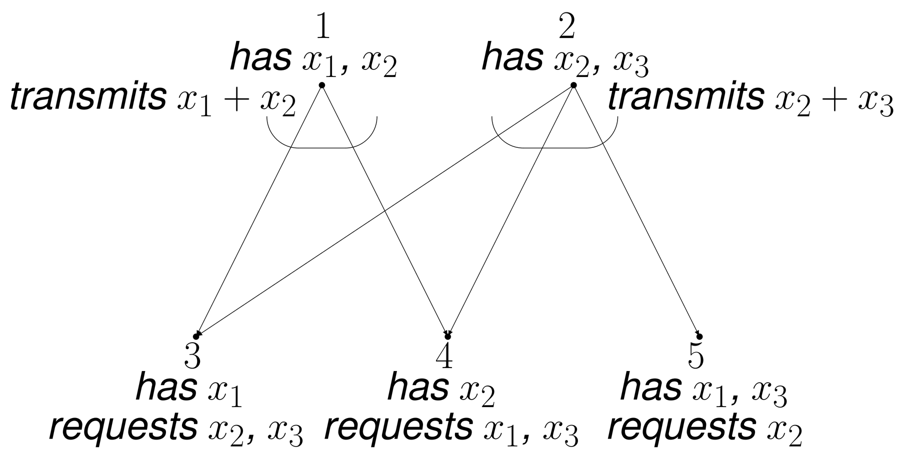


***Figure 1:*** *Example network*

## 2. Problem setup

$\mathcal{G}(\mathcal{V}, \mathcal{E})$ is a directed graph with the vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$. $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$ is the information vector. Each node $\ell \in \mathcal{V}$ possesses as side information the symbols $x_j, j \in \mathcal{P}_\ell \subseteq [n]$ and requests the symbols $x_i, i \in \mathcal{T}_\ell \subseteq [n] \backslash \mathcal{P}_\ell$. For each $\ell \in \mathcal{V}$, let $\mathcal{N}_{in}(\ell)$ be the set of incoming edges. Denote by $(\boldsymbol{A})_{i,j}$ the entry in the $i$-th row and $j$-th column of the matrix $\boldsymbol{A}$. For $\boldsymbol{v} \in \mathbb{F}^n$, $\mathrm{diag}(\boldsymbol{v})$ returns a matrix whose diagonal entries are the elements of $\boldsymbol{v}$ and other values are zeroes.

- $\boldsymbol{E}$ is the all-one square matrix;
- $\boldsymbol{I}$ is the identity matrix;
- $\boldsymbol{D}$ is the transposed adjacency matrix of the graph.

**The goal is to find a schedule which minimizes the number of transmissions such that each node obtains the requested symbols.**

**Definition 1** *The network based on the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is said to be $r$-solvable if for any combination of the sets $\mathcal{P}_\ell$ and $\mathcal{T}_\ell$, $\ell \in \mathcal{V}$, $r$ communications rounds are sufficient to satisfy all the node requests, but $r-1$ rounds are not sufficient. If the network is not $r$-solvable for any $r \in \mathbb{N}$, then we say that it is* not solvable.

## 3. Possession and query matrices

For each node $\ell \in \mathcal{V}$, define a matrix family $\mathbb{A}_\ell$ as

$$(\mathbb{A}_\ell)_{i,j} = \begin{cases} \text{“}\star\text{”} & \text{if } j \in \mathcal{P}_\ell \\ 0 & \text{otherwise} \end{cases} .$$

Each entry "$\star$" can be replaced by an arbitrary element in $\mathbb{F}$. In this way, $\mathbb{A}_\ell$ denotes a family of matrices over $\mathbb{F}$.
The matrix family $\mathbb{A}$ is defined as

$$\mathbb{A} = \begin{bmatrix} \mathbb{A}_1 \\ \mathbb{A}_2 \\ \vdots \\ \mathbb{A}_k \end{bmatrix} .$$

Given $\boldsymbol{A} \in \mathbb{A}$, the $j$-th $n \times n$ sub-matrix of $\boldsymbol{A}$ will be denoted as $\boldsymbol{A}_j$. We will also use the notation $\boldsymbol{A}_{\mathcal{N}_{in}(\ell)}$ to denote the $dn \times n$ matrix

$$\boldsymbol{A}_{\mathcal{N}_{in}(\ell)} = \begin{bmatrix} \boldsymbol{A}_{i_1} \\ \boldsymbol{A}_{i_2} \\ \vdots \\ \boldsymbol{A}_{i_d} \end{bmatrix} ,$$

where $\mathcal{N}_{in}(\ell) = \{i_1, i_2, \cdots, i_d\}$, and $d$ is an in-degree of $\ell$ in $\mathcal{G}$.
For each $\ell \in \mathcal{V}$, we define an $n \times n$ *information matrix* $\boldsymbol{P}_\ell = (\boldsymbol{P}_\ell)_{i \in [n], j \in [n]}$,

$$(\boldsymbol{P}_\ell)_{i,j} = \begin{cases} 1 & \text{if } i = j \text{ and } i \in \mathcal{P}_\ell \\ 0 & \text{otherwise} \end{cases} .$$

Similarly, for each $\ell \in \mathcal{V}$, we define an $n \times n$ *query matrix* $\boldsymbol{T}_\ell = (\boldsymbol{T}_\ell)_{i \in [n], j \in [n]}$,

$$(\boldsymbol{T}_\ell)_{i,j} = \begin{cases} 1 & \text{if } i = j \text{ and } i \in \mathcal{T}_\ell \\ 0 & \text{otherwise} \end{cases} .$$

**Example 1 (cont.)** *The possession sets are*

$$\mathcal{P}_1 = \{1,2\},\ \mathcal{P}_2 = \{2,3\},\ \mathcal{P}_3 = \{1\},\ \mathcal{P}_4 = \{2\},\ \mathcal{P}_5 = \{1,3\},$$

*and query sets are*

$$\mathcal{T}_1 = \emptyset,\ \mathcal{T}_2 = \emptyset,\ \mathcal{T}_3 = \{2,3\},\ \mathcal{T}_4 = \{1,3\},\ \mathcal{T}_5 = \{2\}.$$

*The matrix families $\mathbb{A}_\ell$ are*

$$\mathbb{A}_1 = \begin{bmatrix} \star & \star & 0 \\ \star & \star & 0 \\ \star & \star & 0 \end{bmatrix},\ \mathbb{A}_2 = \begin{bmatrix} 0 & \star & \star \\ 0 & \star & \star \\ 0 & \star & \star \end{bmatrix},\ \mathbb{A}_3 = \begin{bmatrix} \star & 0 & 0 \\ \star & 0 & 0 \\ \star & 0 & 0 \end{bmatrix},$$
$$\mathbb{A}_4 = \begin{bmatrix} 0 & \star & 0 \\ 0 & \star & 0 \\ 0 & \star & 0 \end{bmatrix},\ \mathbb{A}_5 = \begin{bmatrix} \star & 0 & \star \\ \star & 0 & \star \\ \star & 0 & \star \end{bmatrix}.$$

*The information matrices $\boldsymbol{P}_\ell$ corresponding to the matrix families $\mathbb{A}_\ell$ are*

$$\boldsymbol{P}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ \boldsymbol{P}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \boldsymbol{P}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\boldsymbol{P}_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ \boldsymbol{P}_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

*The query matrices $\boldsymbol{T}_\ell$ corresponding to the query sets $\mathcal{T}_\ell$ are*

$$\boldsymbol{T}_1 = \boldsymbol{0},\ \boldsymbol{T}_2 = \boldsymbol{0},$$
$$\boldsymbol{T}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \boldsymbol{T}_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \boldsymbol{T}_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

## 4. Optimal schedule for 1-solvable networks

**Theorem 1** *The minimal number of transmissions needed to satisfy the demands of all the nodes in $\mathcal{V}$ in one round of communications is*

$$\tau = \min_{\boldsymbol{A} \in \mathbb{A}} \left\{ \sum_{\ell \in \mathcal{V}} \mathrm{rank}\left(\boldsymbol{A}_\ell\right) \right\} , \quad (1)$$

*where for all $\ell \in \mathcal{V}$*

$$\mathrm{rowspace}\left( \begin{bmatrix} \boldsymbol{A}_{\mathcal{N}_{in}(\ell)} \\ \hline \boldsymbol{P}_\ell \end{bmatrix} \right) \supseteq \mathrm{rowspace}(\boldsymbol{T}_\ell) . \quad (2)$$

*If the matrix $\boldsymbol{A} \in \mathbb{A}$ does not exist then there is no algorithm that satisfies all the requests in one round.*

**Example 1 (cont.)** *The matrix $\mathbb{A}$ is*

$$\mathbb{A} = \left[\begin{array}{ccc|ccc|ccc|ccc|ccc} \star & \star & \star & 0 & 0 & 0 & \star & \star & \star & 0 & 0 & 0 & \star & \star & \star \\ \star & \star & \star & \star & \star & \star & 0 & 0 & 0 & \star & \star & \star & 0 & 0 & 0 \\ 0 & 0 & 0 & \star & \star & \star & 0 & 0 & 0 & 0 & 0 & 0 & \star & \star & \star \end{array}\right]^T .$$

*By iterating over all matrices in the family,*

$$\boldsymbol{A} = \left[\begin{array}{ccc|ccc|ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]^T$$

*achieves the minimum in Equation (1) such that the condition in Equation (2) holds for all $\ell \in \mathcal{V}$.*
*The induced vector spaces for the left-hand and right-hand side of Equation (2) for this $\boldsymbol{A}$ are*

| $\ell$ | *left-hand side* | *right-hand side* |
|---|---|---|
| 1 | $\mathrm{span}\{(1,0,0),(0,1,0)\}$ | $\{(0,0,0)\}$ |
| 2 | $\mathrm{span}\{(0,1,0),(0,0,1)\}$ | $\{(0,0,0)\}$ |
| 3 | $\mathrm{span}\{(1,0,0),(1,1,0),(0,1,1)\}$ | $\mathrm{span}\{(0,1,0),(0,0,1)\}$ |
| 4 | $\mathrm{span}\{(0,1,0),(1,1,0),(0,1,1)\}$ | $\mathrm{span}\{(1,0,0),(0,0,1)\}$ |
| 5 | $\mathrm{span}\{(1,0,0),(0,0,1),(0,1,1)\}$ | $\mathrm{span}\{(0,1,0)\}$ |

*Node $\ell \in \mathcal{V}$ transmits the vector $\boldsymbol{A}_\ell \cdot \mathbf{x}^T$ whenever the value is not $\mathbf{0}^T$:*

$$\boldsymbol{A}_1 \cdot \mathbf{x}^T = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ 0 \\ 0 \end{bmatrix},$$
$$\boldsymbol{A}_2 \cdot \mathbf{x}^T = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 + x_3 \\ 0 \\ 0 \end{bmatrix}.$$

## 5. Schedule for $r_0$-solvable networks

**Theorem 2** *There exists an iterated data exchange protocol with $r$ rounds, for any $r \geq r_0$, and $\tau$ transmissions, where*

$$\tau = \sum_{i=1}^{r} \left( \min_{\boldsymbol{A}^{(i)} \in (\boldsymbol{D}^{i-1} \otimes \boldsymbol{E}) \cdot \mathbb{A}} \left\{ \sum_{j=1}^{k} \mathrm{rank}\left(\boldsymbol{A}_j^{(i)}\right) \right\} \right) \quad (3)$$

*for matrices $\boldsymbol{A}^{(i)}$ which are subject to*

$$\forall j \in [k] \ : \ \mathrm{rank}\left( \left[ \frac{\left(\mathrm{diag}\left(\boldsymbol{D}^{[j]}\right) \otimes \boldsymbol{I}\right) \cdot \boldsymbol{A}^{(i)}}{\Gamma_j((\boldsymbol{D}^{i-1} \otimes \boldsymbol{E}) \cdot \mathbb{A})} \right] \right)$$
$$= \text{max-rank}\left( \left(\mathrm{diag}(\boldsymbol{e}_j) \otimes \boldsymbol{I}\right) \cdot \left(\boldsymbol{D}^i \otimes \boldsymbol{E}\right) \cdot \mathbb{A} \right) , \quad (4)$$

*where the matrices $\boldsymbol{I}$ and $\boldsymbol{E}$ are both $n \times n$.*

## 6. Experimental results

**Proposition 3** *For a node $\ell \in \mathcal{V}$, and for $i \in [n]$, denote by $\mathsf{d}_\ell(x_i)$ the length of the shortest path from a set of vertices having $x_i$ in their possession to $\ell$. Let $\mathsf{d}_\ell = \sum_{i \in \mathcal{T}_\ell} \mathsf{d}_\ell(x_i)$ and*

$$\mathsf{d}_{\max} = \max_{\ell \in \mathcal{V}} \mathsf{d}_\ell . \quad (5)$$

*Then, the minimum number of transmissions in any algorithm for data dissemination problem is at least $\mathsf{d}_{\max}$.*

We generate the adjacency matrix of the graph randomly, while fixing the diameter and the number of vertices in the graph. We also randomly generate the possession matrix of the network.
For each randomly chosen network with four nodes and four items, we compute the number of transmissions guaranteed by Theorem 2 and the lower bound on the number of transmissions in Proposition 3. The distribution of the ratios in the tested cases is shown in the following tables.

| Range | $[1, 1.2)$ | $[1.2, 1.4)$ | $[1.4, 1.6)$ | $[1.6, 1.8)$ | $[1.8, 2.0)$ | $[2.0, \infty)$ |
|---|---|---|---|---|---|---|
| % | 54 | 22 | 6 | 4 | 0 | 14 |

*The efficiency of the algorithm for graphs of diameter 2*

| Range | $[1, 1.2)$ | $[1.2, 1.4)$ | $[1.4, 1.6)$ | $[1.6, 1.8)$ | $[1.8, 2.0)$ | $[2.0, \infty)$ |
|---|---|---|---|---|---|---|
| % | 30 | 18 | 24 | 0 | 6 | 22 |

*The efficiency of the algorithm for graphs of diameter 3*

## References


[1] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol. Index coding with side information. *Information Theory, IEEE Transactions on*, 57(3):1479–1494, March 2011.

[2] J. F. Geelen. Maximum rank matrix completion. *Linear Algebra and its Applications*, 288:211–217, 1999.

[3] M. Gonen and M. Langberg. Coded cooperative data exchange problem for general topologies. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT) 2012*, pages 2606–2610.

[4] S. Y. El Rouayheb, A. Sprintson, and P. Sadeghi. On coding for cooperative data exchange. *CoRR*, abs/1002.1465, 2010.



Investing in your future
European Union
European Social Fund
norway grants
Eesti Teadusagentuur
Estonian Research Council
European Union
Regional Development Fund
Investing in your future
IT Akadeemia
toetab Skype™