

## HARDNESS AND APPROXIMATION OF GATHERING IN STATIC RADIO NETWORKS\*

JEAN-CLAUDE BERMOND<sup>a</sup>, JEROME GALTIER<sup>b</sup>,  
RALF KLASING<sup>c</sup>, NELSON MORALES<sup>a</sup> and STEPHANE PERENNES<sup>a</sup>  
<sup>a</sup>{bermond, morales, speren}@sophia.inria.fr  
*MASCOTTE project I3S/CNRS/Université de Nice/Sophia-Antipolis, France*  
<sup>b</sup>jerome.galtier@francetelecom.fr  
*France Telecom Research and Development*  
<sup>c</sup>klasing@labri.fr  
*LaBRI - Université Bordeaux I - CNRS*

Received December 15, 2005

Revised May 09, 2006

Communicated by Michel Cosnard

### ABSTRACT

In this paper, we address the problem of gathering information in a specific node (or *sink*) of a radio network, where interference constraints are present. We take into account the fact that, when a node transmits, it produces interference in an area bigger than the area in which its message can actually be received. The network is modeled by a graph; a node is able to transmit one unit of information to the set of vertices at distance at most  $d_T$  in the graph, but when doing so it generates interference that does not allow nodes at distance up to  $d_I$  ( $d_I \geq d_T$ ) to listen to other transmissions. Time is synchronous and divided into time-steps in each of which a round (set of non-interfering radio transmissions) is performed. We give general lower bounds on the number of rounds required to gather into a sink of a general graph, and present an algorithm working on any graph, with an approximation factor of 4. We also show that the problem of finding an optimal strategy for gathering is NP-HARD, for any values of  $d_I$  and  $d_T$ . If  $d_I > d_T$ , we show that the problem remains hard when restricted to the uniform case where each vertex in the network has exactly one piece of information to communicate to the sink.

*Keywords:* Gathering, approximation algorithms, hardness, radio networks, interference constraints.

## 1. Introduction

### 1.1. Background and motivation

In radio networks a set of radio devices communicate by using radio transmissions which, depending on the technology used, are subject to different interference constraints (see for instance [9,17,22]). This means that only certain transmissions can be performed

---

\*This work has been supported by the European projects AEOLUS and COST Action TIST 293 (GRAAL), and is part of the CRC CORSO with France Telecom R&D. Nelson Morales is funded by a CONICYT/INRIA doctoral grant. An extended abstract of this research was presented at FAWN 2006, Pisa, Italy.

simultaneously, therefore the devices have to act in a cooperative manner in order to achieve an effective flow of information in the network. In this context, we study a problem proposed by FRANCE TELECOM, about “how to provide Internet to villages” (see [8]).

The houses of the village are equipped with radio devices and they aim to access the *rest of the world* via Internet. For that purpose they have to send (and receive) information via a gateway where there is a *central* antenna. This creates a special *many-to-one* information flow demand in which the access to the gateway must be provided. Therefore, we will consider a specific traffic pattern, similar to a single commodity flow with a distinguished node representing the gateway, called *sink* and denoted  $t$ .

Unlike in wired networks, when a node  $u$  transmits a message it does not use a resource as simple as some capacity on a link; instead it produces a signal that may prevent other transmissions to occur. The set of possible concurrent transmissions follows from a complex  $n$ -ary interference relation which properly models the idea that the noise intensity must be small enough compared to the signal intensity. In order to get tractable models, a widely used simplification consists of associating to each node a *transmission area* in which it can transmit a message and an *interference area* in which it produces a strong noise. Then, the communication from a node  $u$  to a node  $v$  is possible if  $v$  is in the transmission area of  $u$ , and no third node transmitting has  $v$  in its interference area. Note that, by doing so, we replace the  $n$ -ary relation with a binary relation: two (possible) transmissions (that we will denote calls) can be performed concurrently when they do not interfere.

## 1.2. Modeling aspects

One possible way of modeling would be to represent the houses (radio devices) as nodes in the plane with Euclidean distance (the areas of transmission and interference being disks). Here, we choose to model the network by an undirected graph  $G = (V, E)$ , where  $V$  is the set of devices in the network and to use as distance the distance between nodes in the graph. Firstly, it simplifies the analysis and enables us to give tractable gathering algorithms. Secondly, for some graphs like grids or hexagonal grids the distance in the graphs is a good approximation for the Euclidean distance. Finally, some nodes which are close to each other in the plane might not be able to communicate due to different reasons like obstacles, hills, social relations, security. So, there is an edge if two houses are neighbors and able to communicate.

We model the transmission area and the interference area as balls in the graph by introducing two parameters:  $d_T$ , the transmission radius and  $d_I$ , the interference radius and we suppose that  $d_I \geq d_T$ . The transmission area (resp. interference area) is then the ball of radius  $d_T$  (resp.  $d_I$ ).

The information transmitted by a node becomes available to all the nodes that are in its transmission area if they are listening, and if they are not in the interference area of a third (transmitting) node. We will denote the fact that node  $s$  (like sender) is transmitting a message to node  $r$  (like receiver) by saying there is a *call*  $(s, r)$ . We will say that two calls  $(s, r)$  and  $(s', r')$  with  $s \neq s'$  are *compatible* if  $s$  does not interfere with  $r'$  and  $s'$  does not interfere with  $r$ .

Figure 1 shows a set of 3 calls, which are represented by the arrows over the edges of the graph. If  $d_I = d_T = 1$ , all these calls are compatible. However, if  $d_I = 2, d_T = 1$ , vertex  $b$  is under the interference of vertex  $e$ , and vertex  $f$  is under the interference of vertices  $a$  and  $c$ . In this case, a round could either consist of one single call  $((a, b)$  or  $(c, d)$

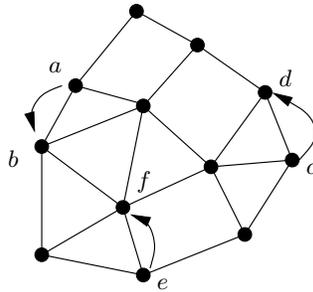


Figure 1: Interfering/compatible calls.

or  $(e, f)$ ), or of the two calls  $(a, b), (c, d)$ .

Under this model, the problem raised by France Telecom consists of gathering information from each node of the network into the central node (the sink  $t$ ). We will suppose that each node has to transmit an integer ( $\geq 0$ ) number of units of information.

Our measure of efficiency is the time (i.e., the number of rounds) needed to achieve gathering, hence our objective is to study the *minimum time gathering problem*. Figure 2 shows an optimal gathering protocol using 18 rounds for a path with 7 vertices (each having one piece of information), with  $d_T = 1, d_I = 2$  and sink  $t = 0$ .

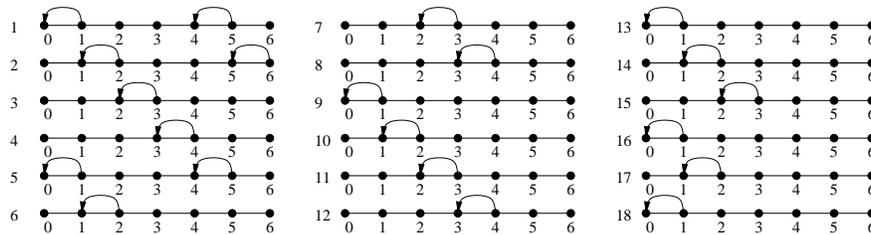


Figure 2: A gathering protocol in the path when  $d_T = 1, d_I = 2$  and every vertex has one message to send to the sink  $t = 0$ .

Note that we may as well study the inverse problem (personalized broadcast) for which we need to send personalized information from the central node to each node. But, like in many other communication models, we can simply reverse the order of the communication steps and the direction of the calls and state that gathering and personalized broadcast are formally equivalent. Due to the perfect equivalence, all the results (algorithms, complexity, bounds) that we give are also valid for personalized broadcast.

### 1.3. Related work

Basic communication problems for the dissemination of information (like gathering, broadcasting, gossiping) have been widely studied in classical interconnection networks (see the book [20]). The fact that a node cannot both send and receive in the same round is known as the *half-duplex 1-port* model and the unit message constraint is studied for

example in [4,5,6].

The broadcasting and gossiping problems in radio networks with  $d_T = d_I = 1$  are studied in [11,15,18] and [13,16,12] respectively. Note that in a broadcast the same information has to be transmitted to all the other nodes and therefore flooding techniques can be used.

Notice that our model with  $d_I = d_T = 1$  is more constrained than the half-duplex model. Indeed, consider  $a, b, c, d$  to be four different vertices. In the half-duplex model, the transmission  $(a, b)$  and  $(c, d)$  can be performed simultaneously, but for instance if  $a$  and  $d$  are neighbors, then this is not allowed in our model.

With respect to the gathering problem, the uniform case with  $d_T = 1$  and any  $d_I$  is studied in depth for the case of the path in [2] and the two-dimensional square grid, for which optimal solutions are provided in [7], whereas in [3], the case  $d_T \geq 2$  and any  $d_I \geq d_T$  is studied for the same topologies. Another related model can be found in [21], where the authors study the case in which steady-state flow demands  $f(u, v)$  between each pair of nodes  $(u, v)$  have to be satisfied.

#### 1.4. Results and Structure of the paper

The paper is organized as follows. In Section 2, the model and the gathering problem are formalized through a number of definitions. In Section 3, we provide a general lower bound and a protocol (valid for any graph and any quantity of information) which allows us to prove that our protocol achieves an approximation ratio of 4, independently of  $d_I$  and  $d_T$ .

In Section 4, we show that the problem is NP-HARD, for any values of  $d_I$  and  $d_T$ . When  $d_I > d_T$ , we show that this remains true when the traffic demand is uniform, i.e., such that exactly one unit of information from each vertex in the network has to be gathered into the sink.

Finally, we present the conclusions in Section 5.

## 2. The model: definitions and notation

In the whole paper, we are given a graph  $G = (V, E)$  with  $n$  vertices and with a distinguished vertex  $t \in V$ , called the *sink*, and two integers  $d_I, d_T \in \mathbb{N}$ , such that  $d_I \geq d_T > 0$ , where  $d_I$  is the *interference distance* and  $d_T$  is the *transmission distance*.

The distance between two vertices  $u$  and  $v$  is the length of the shortest path from  $u$  to  $v$  and is denoted  $d_G(u, v)$ . For  $u \in V$  and  $h \in \mathbb{N}$ , we define the *neighborhood of radius  $h$  or  $h$ -neighborhood of  $u$*  as  $\Gamma_G^h(u) = \{v \in V : d_G(u, v) \leq h\}$ . When the context is clear we will omit the index  $G$  and write shortly  $d(u, v)$  instead of  $d_G(u, v)$  and  $\Gamma^h(u)$  instead of  $\Gamma_G^h(u)$ .

In the *gathering problem*, every node  $u \in V$  has  $w(u)$  pieces of information (called shortly *messages*) which have to reach the sink  $t$ , where  $w(u)$  is a non-negative integer. The particular case  $w(u) = 1$  is referred to as the *uniform case*.

A *call* is a couple  $(s, r)$ ,  $s, r \in V$  with  $0 < d(s, r) \leq d_T$  where  $s$  is the *sender* and  $r$  the *receiver*. The call  $(s, r)$  *interferes* with the call  $(s', r')$  if  $d(s, r') \leq d_I$ . We say that the two calls  $(s, r)$  and  $(s', r')$  are *compatible* if they do not interfere, that is both  $d(s, r') > d_I$  and  $d(s', r) > d_I$ .

A *round* is a set of compatible calls. During a round, a sender transmits a new message

if there is one available.

A *gathering protocol* is an ordered sequence of rounds that allows to gather the information of the nodes in the sink.

We will often specify protocols by giving simply the sequence of rounds, without specifying which message is sent, indeed that is irrelevant as long as each vertex can forward something new. Also, observe that when gathering it is not useful to have multiples copies of a message in different vertices: it suffices to keep the copy that arrives first to the sink. This allows us to consider simply calls of the type  $(s, r)$ , meaning that the sender can select a unique receiver between the potential ones.

Our objective is to find gathering protocols minimizing the number of rounds needed to gather all the messages into the sink. The minimum number of rounds will be called the *gathering number* and denoted shortly  $g(G, w, t)$  (although it formally depends on  $d_T$  and  $d_I$  and the function  $w$  and should be denoted  $g_{d_I, d_T}(G, w, t)$ ).

Note that in any gathering protocol there is a bottleneck near the sink as there is a critical section, where during one round only one message near the sink can move towards the sink. We will define that precisely below. First, let us rule out a trivial case.

**Trivial case.** When  $V$  itself is a critical section, that is when any two calls in  $V$  interfere. Hence, in that case there is at most one call per round and to transmit a message of  $u$  to the sink  $t$  we need at least  $\left\lceil \frac{d(u, t)}{d_T} \right\rceil$  rounds and so in that case  $g(G, w, t) = \sum_{u \in V} w(u) \left\lceil \frac{d(u, t)}{d_T} \right\rceil$ .  $\square$

In what follows, we will suppose that we are not in the trivial case. We define a *critical section of the sink*  $t$  as an  $h$ -neighborhood of  $t$ ,  $\Gamma^h(t)$ , such that any two vertices in  $\Gamma^h(t)$  cannot receive in the same round; said otherwise, there cannot exist two compatible calls  $(s, r)$  and  $(s', r')$  with both  $r$  and  $r'$  in  $\Gamma^h(t)$ . We define the *critical radius*  $r_C = r_C(G, t)$  as the greatest integer  $h$  such that  $\Gamma^h(t)$  is a critical section.

**Example.** Consider a path  $P_n$  with  $n$  vertices  $0, 1, \dots, n-1$ . If  $n \leq d_I + 2$ , we are in the trivial case where  $V$  is a critical section. So we suppose that  $n \geq d_I + 3$ . The computation of  $r_C$  will depend on the position of the sink. If the sink is at one end, say vertex  $0$ , we have  $r_C = d_I + 1$ ; indeed the  $h$ -neighborhood of the sink  $0$  consists of the vertices  $0, 1, \dots, h$  and so if  $h \leq d_I + 1$  it is a critical section (as the sink is not a sender); but for  $h > d_I + 1$  both  $0$  and  $d_I + 2$  can receive, the two calls  $(1, 0)$  and  $(d_I + 1, d_I + 2)$  for example being compatible. If  $n > d_I + d_T + 1$  and if the sink is  $\left\lceil \frac{d_I + d_T}{2} \right\rceil$ , the two calls  $(0, d_T)$  and  $(d_I + d_T + 1, d_I + 1)$  are compatible and therefore a simple computation shows that  $r_C(P_n, \left\lceil \frac{d_I + d_T}{2} \right\rceil) \leq \frac{d_I - d_T}{2}$ .  $\square$

The next lemma shows that there is equality. (Recall that we are not in the trivial case.)

**Lemma 1**

$$\left\lfloor \frac{d_I - d_T}{2} \right\rfloor \leq r_C \leq d_I + 1.$$

**Proof.** For the first inequality, suppose that  $(s, r), (s', r')$  are two calls such that  $r, r' \in \Gamma^{\left\lfloor \frac{d_I - d_T}{2} \right\rfloor}(t)$ . Then  $d(s, r') \leq d(s, r) + d(r, r') \leq d_T + 2 \left\lfloor \frac{d_I - d_T}{2} \right\rfloor \leq d_T + d_I - d_T = d_I$ . Therefore these calls interfere. For the second inequality, suppose that  $r_C \geq d_I + 2$ , then

the two calls  $(s, t)$  with  $d(s, t) = 1$  and  $(s', r')$  with  $s'r'$  an edge of  $G$  and  $d(r', t) = d_I + 2$  are compatible and therefore  $r_C \leq d_I + 1$ .

Note that the bounds are attained as shown by the example of the path.  $\square$

### 3. Constant approximation algorithm for arbitrary graphs

In this section, we first introduce two general lower bounds for the gathering time in general graphs (Lemmas 3 and 4). Then, we present an algorithm for gathering in general graphs and show that this algorithm achieves an approximation ratio of 4 (Theorem 1).

#### 3.1. Lower bounds

Recall that we suppose we are not in the trivial case. For a vertex  $u$ , let us denote  $p(u, t)$  the minimum number of calls, with their receiver inside the critical section that are necessary to take a message originated at  $u$  to the sink.

**Lemma 2** *If  $u \in \Gamma^{r_C}(t)$ :  $p(u, t) = \lceil \frac{d(u, t)}{d_T} \rceil$  and if  $u \notin \Gamma^{r_C}(t)$ :  $p(u, t) = \lceil \frac{1+r_C}{d_T} \rceil$ .*

**Proof.** For a given message the distance to the sink of the vertex containing this message can decrease during a call by at most  $d_T$ . Therefore, if  $u$  is in the critical section, we need at least  $\lceil \frac{d(u, t)}{d_T} \rceil$  calls (rounds) in order that one message of  $u$  reaches the sink. If  $u$  is outside the critical section, the first call with a receiver inside the critical section has its sender at distance at least  $r_C + 1$  from the sink, and so we need at least  $\lceil \frac{1+r_C}{d_T} \rceil$  calls (rounds) to take a message from  $u$  to the sink.  $\square$

#### Lemma 3

$$g(G, w, t) \geq \sum_{u \in V} w(u)p(u, t)$$

**Proof.** For any vertex  $u$ , the minimum number of calls having their receiver in the critical section and needed to transmit a message of  $u$  is  $p(u, t)w(u)$ . Hence  $\sum_{u \in V} w(u)p(u, t)$  such calls have to be performed. By definition of the critical section, all these calls have to be done in different rounds.  $\square$

Let us denote  $\delta(G, w, t)$  (or simply  $\delta$ ) the *eccentricity* of  $t$  in the graph  $G$  with weights  $w$ , that is the maximum distance  $d(u, t)$  with  $w(u) > 0$ . When the messages are concentrated far from the sink, the next bound may provide a better bound.

**Lemma 4** *For any  $1 + r_C + d_T \leq a \leq \delta$ ,*

$$g(G, w, t) \geq \left\lceil \frac{a - (d_T + r_C)}{d_T} \right\rceil + \sum_{u \in V: d(u, t) \geq a} w(u)p(u, t)$$

**Proof.** The bound follows from the observation that no message from a vertex at a distance greater than  $a$  can reach  $\Gamma^{r_C + d_T}(t)$  before  $\lceil \frac{a - (d_T + r_C)}{d_T} \rceil$  rounds and that  $\Gamma^{r_C}(t) \subset \Gamma^a(t)$ .  $\square$

### 3.2. A general protocol

We can derive a protocol that matches the above lower bounds up to a factor of 4. The idea is to pipeline towards the sink by partitioning the graph into intervals. The lengths of the intervals are chosen such that messages can advance in parallel without interfering with each other.

**Theorem 1** *There exists a 4-approximation for the gathering problem.*

**Proof.** The proof is divided in five parts.

**(i) Partition of the set of distances.** Let  $K = \left\lceil \frac{d_I + d_T + 1}{d_T} \right\rceil$ .

If there is a vertex  $u$  such that  $d(u, t) \leq r_C + d_T$  and  $w(u) > 0$ , we define  $M = \max\{d(u, t) : d(u, t) \leq r_C + d_T, w(u) > 0\}$ . If not, we set  $M = r_C + d_T$ . In any case, let  $K_0 = \left\lceil \frac{M}{d_T} \right\rceil$ , and  $L = 1 + \left\lceil \frac{\delta - K_0 d_T}{K d_T} \right\rceil$ .

We partition the set of distances to the sink  $[1, \delta]$  (recall that  $\delta$  is the eccentricity of  $t$ ) into  $L$  intervals. Interval  $I_0 = [1, K_0 d_T]$ , and for  $i = 1, \dots, L - 1$ , interval  $I_i = [K_0 d_T + 1 + (i - 1)K d_T, K_0 d_T + iK d_T]$ .

Now, we split  $I_i$  into areas of length  $d_T$ , so  $I_0$  is split into  $K_0$  areas  $I_0^j = [K_0 d_T + 1 - j d_T, K_0 d_T - (j - 1) d_T]$ ,  $j = 1, \dots, K_0$ ; and  $I_i, i = 1, \dots, L - 1$  is split into  $K$  areas  $I_i^j = [K_0 d_T + 1 + iK d_T - j d_T, K_0 d_T + iK d_T - (j - 1) d_T]$ ,  $j = 1, \dots, K$ . (Notice that to simplify the description of the protocol we have labeled the areas from right to left, i.e., in order of decreasing distance to the sink.)

Figure 3 shows a partition with  $K = 4, K_0 = 3, d_T = 2$ .

We denote the set of vertices whose distance is in  $I_i$  (respectively  $I_i^j$ ) by  $V_i$  (respectively  $V_i^j$ ) and  $w_i$  (respectively  $w_i^j$ ) denotes the sum of the weights of the vertices in  $V_i$  (respectively  $V_i^j$ ). That is  $w_i = \sum_{u: d(u, t) \in I_i} w(u)$ ,  $w_i^j = \sum_{u: d(u, t) \in I_i^j} w(u)$ .

**(ii) The protocol.** The protocol works in steps. Each step consists of  $K$  rounds  $R_j, j = 1, \dots, K$  (except the last step).

During round  $R_j$ , the protocol selects in each interval  $I_i$  a vertex  $w_i^j$  in  $V_i^j$  with an available message to transmit (if such a vertex exists). Vertex  $w_i^j$  calls the closest vertex in the preceding area, i.e., if  $d(w_i^j, t) = K_0 d_T + 1 + iK d_T - j d_T + \alpha$  for some  $0 \leq \alpha < d_T$ , then  $w_i^j$  calls a vertex  $v$  such that  $d(v, t) = K_0 d_T + iK d_T - j d_T$ . This means that if  $i = 0$  and  $j < K_0$  (or  $i > 0$  and  $j < K$ ) then  $v \in V_i^{j+1}$ , if  $i > 0, j = K$  then  $v \in V_{i-1}^1$ , and if  $i = 0, j = K_0$  then  $v = t$ . (Figure 3 shows one step for  $K = 4$ .)

Let us show that, for fixed  $j$ ,  $R_j$  is a round (i.e. a set of compatible calls). Indeed, consider two calls  $(s, r) \neq (s', r')$  in the same round  $R_j$ . Then  $d(s, t) = K_0 d_T + 1 + iK d_T - j d_T + \alpha$ , for some  $i \geq 0, 0 \leq \alpha < d_T$ , and  $d(r', t) = K_0 + i'K d_T - j d_T$  for some  $i' \neq i$  (as  $r \neq r'$ ). Therefore  $d(s, r') = |(i' - i)K d_T - 1 - \alpha| \geq d_I + d_T - \alpha \geq d_I + 1$ .

Finally, observe that after a step of  $K$  rounds, the protocol ensures that if a vertex of  $V_i$  contains a message, then the last vertex of  $V_{i-1}$  has received a new message.

**(iii) A corresponding buffer process.** Our protocol is similar to the following buffer process, where buffer  $i$  will correspond to interval  $I_i$ , a packet in buffer  $i$  to a message in a vertex of  $V_i$ , and a step of the buffer process to a step of  $K$  rounds in our protocol.

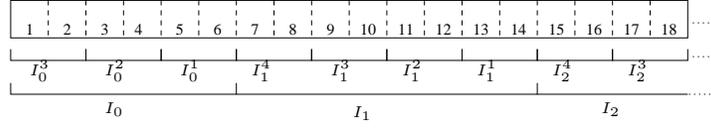


Figure 3: Partitioning of distance intervals for  $K = 4, K_0 = 3, d_T = 2$ .

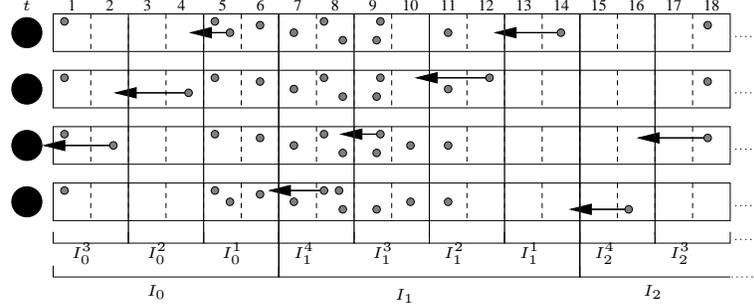


Figure 4: A step of the algorithm, consisting of  $K = 4$  rounds. Here messages are represented as small balls. Notice that messages in the same cell are at the same distance from the sink, but they can be in different vertices.

We have  $L$  buffers,  $i = 0, \dots, L - 1$  organized from left to right on a line. Initially, buffer  $i$  contains  $w_i$  packets. At each time step, any buffer with at least one packet sends a packet to its left neighbor (buffer 0 consumes one packet = it transmits to the outside).

**(iv) Analysis of the buffer process.** We wish to determine the minimum number of steps  $T(w_0, w_1, \dots, w_{L-1})$  needed to transmit all the packets to the outside.

Let us say that buffer  $j$  is non-critical if it gets empty while there still exist packets on its right (that is, a buffer  $j' > j$  still has messages). Remark that once a buffer gets empty, it will never contain more than 1 packet. From this, if  $j$  is non-critical, then  $T(w_j, w_{j+1}, \dots, w_{L-1}) = 1 + T(w_{j+1}, w_{j+2}, \dots, w_{L-1})$ . Otherwise if buffer  $j$  is critical  $T(w_j, w_{j+1}, \dots, w_{L-1}) = \sum_{i \geq j} w_i$ . Therefore, if  $i_0$  is the leftmost critical buffer we get:  $T(w_0, w_1, \dots, w_{L-1}) = i_0 + T(w_{i_0}, w_{i_0+1}, \dots, w_{L-1}) = i_0 + \sum_{i \geq i_0} w_i$ .

**(v) Analysis of the protocol.** If we apply the buffer scheme directly we get for our protocol a number of rounds equal to  $K \cdot T(w_0, w_1, \dots, w_{L-1})$  which gives only an approximation ratio tending to 4 when  $\delta$  (or  $L$ ) goes to  $\infty$ . To obtain the approximation ratio of 4 we need to be more precise about what happens in the end on interval  $I_0$ , and sometimes to do some preprocessing.

In what follows, let  $K' = \left\lceil \frac{1+r_C}{d_T} \right\rceil$  and recall that  $K = \left\lceil \frac{d_I+d_T+1}{d_T} \right\rceil$ . Because  $r_C \leq d_I + 1$  (Lemma 1), we have that  $1 \leq K' \leq K$ . Also, we have that  $K/K' \leq 4$ : indeed because  $r_C \geq \lfloor \frac{d_I-d_T}{2} \rfloor$ ,  $K/K' \leq K / \left\lceil \frac{d_I-d_T+1}{2d_T} \right\rceil$ , but if  $d_I \leq 3d_T - 1$ , then  $K \leq 4$  and

if  $d_I = qd_T + r$ ,  $q \geq 3$ ,  $0 \leq r < d_T$ , then  $K = q + 2$  and  $\left\lceil \frac{d_I - d_T + 1}{2d_T} \right\rceil = \left\lceil \frac{q}{2} \right\rceil \geq \frac{q}{2}$ , but  $q \geq 3$ , so  $q + 2 \leq 2q$ . Finally, because  $1 \leq K_0 \leq \left\lceil \frac{rc + d_T}{d_T} \right\rceil$ , we get  $K_0/K' \leq 2$ .

To analyze the protocol, we distinguish two cases :

**(v.1)  $w_0 > 0$  and ( $i_0 = 0$  or  $1$ ):** In this case the protocol performs  $\sum_{i \geq 1} w_i$  times a step of  $K$  rounds (which corresponds to the buffer process emptying all the buffers but buffer 0). Because  $i_0 \leq 1$ , this requires  $K \sum_{i \geq 1} w_i$  rounds.

Notice that because  $w_0 > 0$  there exists initially a vertex  $u_0$  such that  $d(u_0, t) \leq M$  (i.e.,  $u_0 \in V_0^1$ ) with a message  $m$  and we can, without loss of generality, assume that this message has been selected in the first  $K_0 (\leq K)$  rounds and therefore it has reached the sink during the first step of  $K$  rounds. If during the rest of the protocol we choose to transmit, when possible, the new messages that arrived in the preceding round, then all the messages with distance in  $I_i, i \geq 1$  have reached the sink, except for one message  $m'$  which has arrived at a vertex at a distance  $K_0 d_T \in I_0^1$  and consequently this message will reach the sink in  $K_0$  rounds.

If  $i_0 = 1$ , then  $m'$  is the only message that has not been gathered into the sink. If  $i_0 = 0$ , then apart from  $m'$  there remain in  $V_0$  the initial messages of vertices of  $V_0$  (excepting  $m$ , which is already gathered into the sink). To gather these messages (including  $m'$ ) into the sink the protocol requires at most  $\sum_{u \in V_0 - V_0^1} w(u) \left\lceil \frac{d(u, t)}{d_T} \right\rceil + K_0(w_0^1 - 1) + K_0$ , so altogether the number of rounds used by the protocol is

$$UB = \sum_{u \in V_0 - V_0^1} w(u) \left\lceil \frac{d(u, t)}{d_T} \right\rceil + K_0 w_0^1 + K \sum_{i \geq 1} w_i.$$

But from Lemma 3 we know that any protocol needs

$$LB = \sum_{u \in V_0 - V_0^1} w(u) \left\lceil \frac{d(u, t)}{d_T} \right\rceil + K' w_0^1 + K' \sum_{i \geq 1} w_i,$$

and because  $K_0/K' \leq 2$  and  $K/K' \leq 4$ , we have  $UB/LB \leq 4$ .

**(v.2)  $w_0 = 0$  or ( $w_0 > 0$  and  $i_0 \geq 2$ ):** First notice that if  $w_0 = 0$  then  $V_0$  is not critical and  $i_0 \geq 1$ .

In this case we have to perform a preprocessing phase. Let  $u_{i_0}$  be the first vertex in  $V_{i_0}$  that has a message ( $w(u_{i_0}) > 0$ ). We move this message to the preceding interval  $V_{i_0-1}$  and, more precisely, to the area  $V_{i_0-1}^1$ . If  $u_{i_0} \in V_{i_0}^j$ , then this requires  $p = K + 1 - j$  rounds.

Then we apply our protocol. As all the buffers (intervals) with  $i < i_0$  are non-critical, after  $\sum_{i \geq i_0} w_i - 1 + i_0 - 1$  steps we are left with exactly one message in  $V_0$  and this message is exactly at a distance  $K_0 d_T$  from  $t$  (the  $-1$  in  $\sum_{i \geq i_0} w_i - 1$  comes from the fact that we moved one message out of interval  $V_{i_0}$  during the preprocessing phase). To gather this very last message into the sink takes  $K_0$  rounds, therefore the protocol uses a total number of rounds

$$UB = p + K \left( \sum_{i \geq i_0} w_i + i_0 - 2 \right) + K_0.$$

For a lower bound we use Lemma 4 with  $a = d(u_{i_0}, t) = (K_0 + K(i_0 - 1) + K - j)d_T + \alpha$ , for some  $0 \leq \alpha < d_T$ . First, let us prove that  $a \geq r_C + d_T$ . Indeed,  $r_C \leq d_I + 1$  and  $r_C + d_T \leq Kd_T$ , therefore: either  $i_0 \geq 2$  and  $a > Kd_T \geq r_C + d_T$ ; or  $i_0 = 1$  and then  $w_0 = 0$ , so  $K_0 = \left\lceil \frac{r_C + d_T}{d_T} \right\rceil$  and  $a > K_0 d_T \geq r_C d_T$ .

Finally,  $a - (r_C + d_T) \geq a - Kd_T = (K_0 + K(i_0 - 2) + K - j)d_T + \alpha$  and  $\left\lceil \frac{a - (r_C + d_T)}{d_T} \right\rceil \geq K_0 + K(i_0 - 2) + K - j + 1 = K_0 + K(i_0 - 2) + p$ . Therefore, from Lemma 4,

$$LB \geq K_0 + K(i_0 - 2) + p + K' \sum_{i \geq i_0} w_i.$$

The result follows from the fact that  $K/K' \leq 4 \Rightarrow UB/LB \leq 4$ .  $\square$

**Remark 1** Notice that the exact approximation ratio given by the proof is  $K/K'$  (as  $K_0 \leq K$ ), which depends on  $d_I, d_T$  and  $r_C$ , 4 being an upper bound independent of these parameters.

For example, if  $d_I = qd_T + r, 0 \leq r < d_T$  and  $q \neq 2$ , we have  $K/K' \leq 3$  ( $q = 1$  and  $K \leq 3$  or  $q \geq 3$  and  $K = q + 2, K' \geq \lceil \frac{q}{2} \rceil$ ), and more generally  $K/K' \leq 2 + \varepsilon$  with  $\varepsilon \rightarrow 0$  when  $d_I/d_T \rightarrow \infty$ .

Similarly, if  $r_C = d_I + 1$  (case of the path with the sink being an end vertex)  $K' \geq q + 1$  and therefore  $K/K' \leq \frac{q+2}{q+1} \leq \frac{3}{2}$ , and furthermore if  $d_I = qd_T - 1$  (i.e.,  $d_I = -1 \pmod{d_T}$ ) then the approximation ratio is 1 and the protocol is optimal (see [2] for further results on the path when  $d_T = 1$  and [3] for more results on the path in the general case).

On the negative side, 4 is the best value that we can obtain with the above protocol. Consider a star with  $B \in \mathbb{N}$  branches, each of length  $n/B$  and assume that the sink is in the center and each vertex has one message to transmit. If  $d_I = 5, d_T = 2$ , we can assume that the protocol gathers each branch sequentially, using a number of rounds of order  $4n$ . However, in [3] we present a protocol that requires an order of  $\frac{B+1}{B}n$  rounds to gather in this case. The ratio  $4\frac{B}{B+1}$  can be made as close to 4 as wanted.

**Note added in print:** It has been shown recently [10] that any protocol that uses only shortest paths to the sink (like ours) cannot achieve an approximation ratio smaller than 4 in the general case.

#### 4. Hardness results

In this section we show that MINIMUM GATHERING TIME is NP-HARD. Our reductions depend on the values of  $d_I$  and  $d_T$ . If  $d_I = d_T$ , we show in Lemma 5 that 3SAT can be reduced to gathering, and henceforth the problem is NP-HARD in this case. When  $d_I > d_T$ , we reduce MINIMUM VERTEX COLORING (Lemma 6) and MAXIMUM INDEPENDENT SET (Lemma 7) to a uniform instance of MINIMUM GATHERING TIME, thus showing that the problem remains hard in the uniform case.

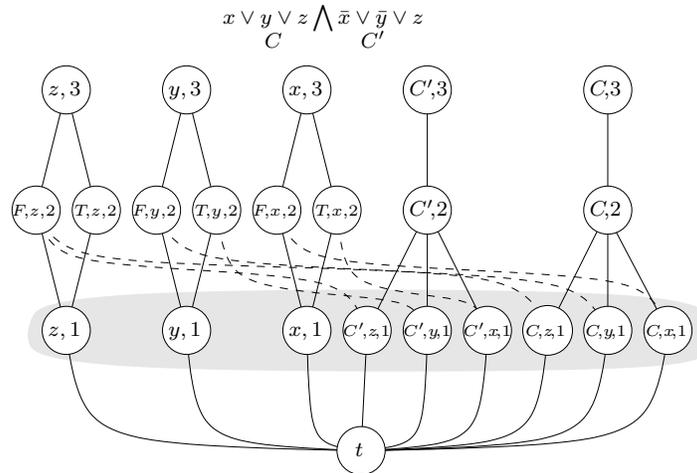


Figure 5: An example of the reduction with  $m = 2$  clauses and  $n = 3$  variables. Vertices in the grey zone are connected as the complete graph.

#### 4.1. $d_I = d_T$

When the transmission distance and the interference distance coincide, we show that it is possible to reduce 3SAT [14] to MINIMUM GATHERING TIME, meaning that the latter problem is NP-HARD.

**Lemma 5** *If  $d_I = d_T$ , then 3SAT can be reduced to MINIMUM GATHERING TIME in polynomial time.*

**Proof.** Let us first concentrate on the case  $d_I = d_T = 1$ , as the argument for the remaining case is very similar.

We reduce 3SAT to the gathering problem. We consider a 3SAT instance with  $m$  clauses  $C \in \mathcal{C}$ , and  $n$  variables  $x \in X$ . The nodes of the network are defined as follows:

- We start with layer 3 with  $n + m$  nodes associated to variables and clauses, they are denoted  $(C, 3), C \in \mathcal{C}$  and  $(x, 3), x \in X$ . We set  $w(C, 3) = w(x, 3) = 1$  for any  $C \in \mathcal{C}, x \in X$ .
- In layer 2, we find one node  $(C, 2)$  per clause  $C \in \mathcal{C}$  and two nodes  $(T, x, 2), (F, x, 2)$  per variable  $x \in X$  [ $(T, x, 2)$  (resp.  $(F, x, 2)$ ) being associated to setting variable  $x$  to true (resp. false)]. These vertices have weight zero.
- In layer 1, we find three nodes per clause. If a clause  $C \in \mathcal{C}$  uses the three variables  $x_1, x_2, x_3$  we add the nodes  $(C, x_1, 1), (C, x_2, 1), (C, x_3, 1)$ . We also create one node  $(x, 1)$  per variable  $x \in X$ . These vertices also have weight zero.
- In layer 0, we find the sink of the gathering.

Nodes are interconnected as follows :

- From layer 3 to layer 2 we find edges between  $(x, 3)$  and the two nodes  $(T, x, 2)$  and  $(F, x, 2)$ ; and between  $(C, 3)$  and  $(C, 2)$ .
- From layer 2 to layer 1 we find edges between  $(T, x, 2)$  (resp.  $(F, x, 2)$ ) and any node  $(C, x, 1)$  such that  $x$  appears as negated (resp. non-negated) in the clause  $C$ . We connect  $(T, x, 2)$  and  $(F, x, 2)$  to  $(x, 1)$ . We also connect  $(C, 2)$  to the three nodes  $(C, x, 1)$  where  $x$  is a variable of  $C$ .
- Nodes in layer 1 are connected as a complete graph.
- Nodes in layer 1 are connected to the sink.

Refer to Figure 5 for a concrete example.

We will show that the 3SAT formula is satisfiable if and only if the gathering time is  $n + m + 2$ .

First, notice that as it is not possible to have a reception during the first 2 rounds, any protocol requires at least  $n + m + 2$  rounds.

Now, we consider a protocol using  $n + m + 2$  rounds. As during the first 2 rounds the sink cannot receive any message, after round 2 the sink must receive during any round. Since vertices in layer 1 are connected using a complete graph, this means that no vertices in layer 1 can receive after the second round. It follows that any protocol lasting  $n + m + 2$  rounds must gather all the messages in layer 1 after 2 rounds. The protocol must then during round 1 perform the calls  $\forall C \in \mathcal{C}, ((C, 3), (C, 2))$  and for  $x \in X$  either  $((x, 3), (T, x, 2))$  or  $((x, 3), (F, x, 2))$ .

Consider an assignment of the variables such that the formula is satisfied. We define a protocol that gathers in  $n + m + 2$  rounds as follows.

In round 1, if  $x$  is true (resp. false) then  $(x, 3)$  calls  $(T, x, 2)$  (resp.  $(F, x, 2)$ ); also  $(C, 3)$  calls  $(C, 2)$ , for each clause  $C$ .

In round 2,  $(C, 2)$  calls a vertex  $(C, x, 1)$  such that the setting of  $x$  satisfies clause  $C$  (such a variable exists because the formula is satisfied). That is: we choose one variable  $x$  such that  $x$  is in  $C$  and  $x$  is set to true, or  $\bar{x}$  is in  $C$  and  $x$  is set to false. The node  $(T, x, 2)$  or  $(F, x, 2)$  which has the message of  $(x, 3)$  calls  $(x, 1)$ .

The calls in round 2 are compatible. Indeed, if  $x$  is not negated (resp. negated) in  $C$ , then  $(C, x, 1)$  is not linked to  $(T, x, 2)$  (resp.  $(F, x, 2)$ ).

In the remaining  $n + m$  rounds, the messages are gathered one by one into the sink.

Conversely, consider a protocol with  $n + m + 2$  rounds. If during round 1 of the protocol  $(x, 3)$  calls  $(T, x, 2)$  then we set  $x$  to true, otherwise (i.e.  $(x, 3)$  calls  $(F, x, 2)$ ) we set  $x$  to false.

In round 2, for any  $C$ ,  $(C, 2)$  calls  $(C, x, 1)$  for some  $x$  in  $C$ . This call is compatible with the call of the same round received by  $(x, 1)$ . If the sender is  $(T, x, 2)$ , then  $(T, x, 2)$  is not linked to  $(C, x, 1)$  and  $x$  appears as non negated in  $C$ . But  $x$  is true and so  $C$  is satisfied. Similarly, if  $(F, x, 2)$  calls  $(x, 1)$ , then  $(F, x, 2)$  is not adjacent to  $(C, x, 1)$  and therefore  $\bar{x}$  appears in  $C$ , but  $x$  is false and also  $C$  is satisfied.

For the case in which  $d_I = d_T \geq 2$ , we replace the inter-layer edges of the construction with paths of length  $d_T$ , but keep the weight function  $w(u) = 1$  if  $u$  is in the last layer and  $w(u) = 0$  if not and the complete graph in layer 1 (we keep 4 layers and do not name the vertices in inter-layer paths). The lower bound of  $n + m + 2$  rounds for the gathering time follows exactly as before.

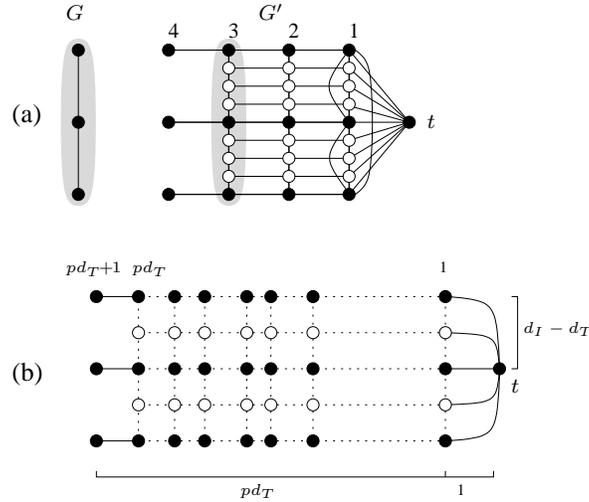


Figure 6: The reduction from MINIMUM VERTEX COLORING. (a)  $d_I = 5, d_T = 1$  (i.e.  $p = 3, r = 0$ ). (b) The general construction for  $r > 0$ . (Solid lines represent edges, dotted lines are paths.)

Now, because the vertices in layer 1 are connected as a complete graph, receptions in the sink and receptions in a vertex at a distance  $1 < i \leq d_T$  from the sink must occur in rounds that are disjoint from each other, therefore if a protocol achieves gathering in  $n + m + 2$  rounds, the protocol must transmit all the messages from layer 3 to vertices in layer 1 (or closer to the sink) in exactly 2 rounds. This forces the protocol to perform only calls of length  $d_T$  (i.e., calls from layer 3 to layer 2 and from layer 2 to layer 1). But the protocol has to make  $n + m$  calls from layer 2 to layer 1 at the same time and this is possible if and only if the formula is satisfiable.  $\square$

#### 4.2. $d_I = (2p - 1)d_T + r, p \geq 2, 0 \leq r < d_T$ , or $p = 1, 0 < r < d_T$

For a given  $G$ , we will construct a gathering instance  $(G', w, t)$  such that the protocols for this instance are all made of two parts, one trivial part in which the messages move one by one towards the sink, and one complex part in which the messages cross a bottleneck using non-interfering calls. The second part induces a coloring for  $G$ .

**Lemma 6** *Let  $d_I, d_T \in \mathbb{N}$  such that  $d_I = (2p - 1)d_T + r, p \geq 2, 0 \leq r < d_T$ , or  $p = 1, 0 < r < d_T$ . Any instance  $G$  of MINIMUM VERTEX COLORING can be mapped in polynomial time to a uniform instance  $(G', w, t)$  of MINIMUM GATHERING TIME and  $G$  can be colored with  $c$  colors if and only if  $(G', w, t)$  admits a gathering protocol of time  $g(G, w, t) = c + f(G, d_I, d_T)$ , where  $f$  is a simple polynomial function.*

**Proof.** The main idea is summarized in Figure 6. Given  $d_I = (2p - 1)d_T + r, p \geq 2, 0 \leq r < d_T$ , or  $p = 1, 0 < r < d_T$ , and a graph  $G = (V, E), n = |V|$ , the gathering instance  $(G', w, t)$  is constructed as follows. The network  $G' = (V', E')$  is organized in  $pd_T + 1$  layers.

- Every  $v \in V$  is replaced by  $pd_T + 1$  copies ( $v_i, i = 1, \dots, pd_T + 1$ ).  
(For  $1 \leq i \leq pd_T + 1$ , the vertices  $\{v_i | v \in V\}$  are called the  $i$ -th layer of  $G'$ .)  
An extra vertex  $t$  is added (as the sink of the gathering instance).

- (A) Every  $v_i, i \geq 2, v \in V$  is connected to  $v_{i-1}$ .  
The vertices  $v_1, v \in V$  are connected to the sink  $t$ .
- (B) If  $uv \in E$ , then  $v_i, i = 1, \dots, pd_T$  is connected with  $u_i$  through a path of length  $d_I - d_T$  (recall that  $d_I > d_T$ ). This introduces for each edge  $d_I - d_T - 1$  new vertices in layer  $i$ :  $uv_{i,j}, j = 1, 2, \dots, d_I - d_T - 1$  and we connect  $uv_{i,j}$  with  $uv_{i-1,j}$  for  $i \geq 2$ .
- (C) If  $r = 0$ , then the vertices  $v_1, v \in V$  are connected as a complete graph. (Notice this regards to the copies of the original vertices only, and not those added as paths between them in (B).)
- Each node (except for the sink  $t$ ) has one message to send.

We remark that the vertices in layer  $i$  are exactly at distance  $i$  from the sink, and the distance between a node in layer  $i$  and a node in layer  $j$  is at most  $i + j$  (and at most  $i + j - 1$  when  $r = 0$ ).

Given a  $c$ -coloring  $C$  of  $G$ , that is a partition of  $V$  into  $c$  independent sets  $U_1, U_2, \dots, U_c$ , we construct a gathering protocol  $R_C$  for  $(G', w, t)$  as follows. The analysis is different, depending on  $p$ .

**If  $p \geq 2$ :** The protocol  $R_C$  performs the  $c$  rounds  $\{(v_{pd_T+1}, v_{pd_T})\}_{v \in U_i, i = 1, 2, \dots, c}$ , and then  $R_C$  gathers the messages one by one into the sink in  $\sum_{x \in V'} \min(p, \lceil d(x, t)/d_T \rceil)$  rounds. Hence,

$$|R_C| \leq c + \sum_{x \in V'} \min(p, \lceil d(x, t)/d_T \rceil).$$

We only have to check that a round  $\{(v_{pd_T+1}, v_{pd_T})\}_{v \in U_i}$  is valid. Because  $p \geq 2$ , the distance from  $u_{pd_T+1}$  to  $v_{pd_T}$  (for  $uv \notin E$ ) is  $2pd_T + 1$  (when  $r > 0$ ) and  $2pd_T$  (when  $r = 0$ ) which is at least  $d_I + 1$  (note that no interference can be induced by following twice the paths of length  $d_I - d_T$ , as  $2(d_I - d_T) + 1 > d_I$ ).

**If  $p = 1$ :** In this case, we have  $r > 0$  and  $d_I = d_T + r$ , the protocol  $R_C$  performs the rounds  $\{(v_{d_T+1}, v_r)\}_{v \in U_i, i = 1, 2, \dots, c}$ , before gathering the messages one by one towards the sink. In this case, the distance from  $v_{d_T+1}$  to  $u_r$  is  $d_I + 1$  going through the sink and at least  $d_T + 1 - r + 2r = d_I + 1$  using two paths (of length  $d_I - d_T$ ) representing edges of  $G$ . It follows that

$$|R_C| \leq c + |V'| = c + \sum_{x \in V'} \min(p, \lceil d(x, t)/d_T \rceil).$$

Conversely, given a gathering protocol  $R$  for  $(G', w, t)$ , we construct a coloring  $C_R$  for  $G$  as follows. We study the possible calls:

- (I) Assume that a vertex in a layer  $\leq (p-1)d_T$  receives. Then, the sender must be in a layer at most  $\leq pd_T$  and this induces interference on all the layers  $\leq (p-1)d_T + r$ . Hence, at most one node in a layer  $\leq (p-1)d_T$  can receive during a round.
- (II) Assume that a vertex in layer  $pd_T + 1$  sends a message, then all layers  $\leq (p-1)d_T + \max(r-1, 0)$  are in its interference range. (Note, when  $r = 0$ , the edges of the complete graph introduced in (C) shorten the distance by one.) This prevents reception in layers  $\leq (p-1)d_T$ .

It follows that calls of kind I and II interfere. Remark also that for any two nodes  $u_{pd_T+1}, v_{pd_T+1}$  such that  $uv \in E$ , one has for any node  $x$ ,  $d(x, v_{pd_T+1}) \leq d(x, u_{pd_T+1}) + d_I - d_T$  (due to the paths (B)). It follows that if  $uv \in E$ ,  $u_{pd_T+1}$  and  $v_{pd_T+1}$  cannot both be sending during to the same round. Rounds with calls of kind II henceforth correspond to independent sets in the graph  $G$ . (More precisely, vertices in layer  $pd_T + 1$  who are sending in the same round, form an independent set in  $G$ .)

Now, any message originating from a node  $x$  at distance  $d(x, t)$  from the sink induces at least  $\min(p, \lceil d(x, t)/d_T \rceil)$  receptions in layers at most  $(p-1)d_T$ . This implies any protocol must have  $\sum_{x \in V'} \min(p, \lceil d(x, t)/d_T \rceil)$  rounds with calls of kind I.

Because all the nodes of layer  $pd_T + 1$  must send once, any protocol  $R$  induces a coloring  $C_R$  of  $G$  (corresponding to rounds with calls of kind II), we have

$$|R| \geq |C_R| + \sum_{x \in V'} \min(p, \lceil d(x, t)/d_T \rceil).$$

Hence, the lemma holds with  $f(G, d_I, d_T) = \sum_{x \in V'} \min(p, \lceil d(x, t)/d_T \rceil) = pn + \frac{p(p+1)}{2}(n + e(d_I - d_T - 1))d_T$ , where  $e = |E|$ .  $\square$

#### 4.3. $d_I = 2pd_T + r, p \geq 1, 0 \leq r < d_T$

We will show that, for an appropriate network, MAXIMUM INDEPENDENT SET (maximum stable set) and MINIMUM GATHERING TIME are almost equivalent, the only difference being in the cost function. Given a graph  $G$ , we denote  $\alpha(G)$  the maximum size of an independent set of  $G$  (stability number).

For a given  $G$  and  $N \in \mathbb{N}$ , we will construct a gathering instance  $(G_N, w, t)$  such that the protocols for this instance are all made of two parts, one trivial part in which the messages move one by one towards the sink, and one complex part in which the messages cross a bottleneck using non interfering calls. The second part induces an independent set for  $G$ .

**Lemma 7** *Let  $d_I, d_T \in \mathbb{N}$  such that  $d_I = 2pd_T + r, p \geq 1, 0 \leq r < d_T$ . For any  $N \in \mathbb{N}$ , any instance  $G$  of MAXIMUM INDEPENDENT SET can be mapped in polynomial time to a uniform instance  $(G_N, w, t)$  of MINIMUM GATHERING TIME such that  $G$  has an independent set  $I$  of size  $|I|$  if and only if  $g(G_N, w, t) = N/|I| + pN + o(N)$ .*

**Proof.** The main idea is summarized in Figure 7. Given  $d_I = 2pd_T + r, 0 \leq r < d_T$ , a graph  $G = (V, E)$  and a positive integer  $N$  we construct a uniform gathering instance  $(G_N, w, t)$  as follows

- For any vertex  $v \in V$ , make  $(2p+1)d_T - 1$  copies of it, named  $v_1, \dots, v_{(2p+1)d_T-1}$ . Call the set  $\{v_i : v \in V\}$  the layer  $i$ .

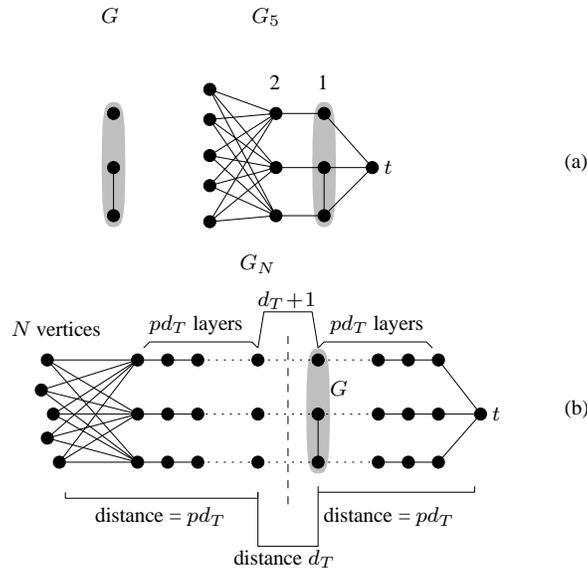


Figure 7: The reduction from MAXIMUM INDEPENDENT SET. (a)  $d_I = 2, d_T = 1, N = 5$ . (b) General construction. The total number of layers is  $(2p + 1)d_T - 1$ .

- Connect  $v_i$  with  $v_{i+1}$ .
- Add a sink  $t$  and connect all the vertices in layer 1 with it.
- Connect  $v_{p d_T}$  and  $u_{p d_T}$  with a path of length  $d_I - d_T$  if  $uv \in E$ . This introduces  $(d_I - d_T - 1)e$  extra vertices. Notice that these extra vertices do not belong to layer  $p d_T$ , nor are they at a distance  $p d_T$  from the sink.
- Finally, take  $N$  independent vertices  $s^k : k = 1, \dots, N$  and connect  $s^k$  with all the vertices in the layer  $(2p + 1)d_T - 1$ .

Because it is a uniform instance, we set  $w(u) = 1$  for any vertex  $u \in G_N$ . The rest of the proof is divided in 2 parts.

**Constructing a gathering protocol from an independent set.** Let  $I = \{b^j\}_{j=0}^{|I|-1}$  be an independent set of  $G$ . We can construct a gathering protocol  $R_I$  for  $(G_N, w, t)$  as follows:

1. First, gather into the sink the messages of all vertices except for  $s^k$  and  $b_{i d_T}^j, i = 1, \dots, 2p$ . Do so making only one call per round and through shortest paths to the sink.
2. While there are at least  $|I|$  messages in the vertices  $s^1, \dots, s^N$  that have not been transmitted, repeat:
  - (a) Choose  $|I|$  vertices  $s^{k_j}$  with one message to transmit and for  $j = 0, \dots, |I| - 1$  do simultaneously the calls  $(s^{k_j}, b_{2 p d_T}^j)$  and  $(b_{d_T}^{j+1 \bmod |I|}, t)$ .

- (b) For  $i = 0, \dots, p-2$ , and  $j = 0, \dots, |I| - 1$  do simultaneously the calls  $(b_{(p-i)d_T}^j, b_{(p-1-i)d_T}^j)$  and  $(b_{(p+2+i)d_T}^{j+1 \bmod |I|}, b_{(p+1+i)d_T}^{j+1 \bmod |I|})$ .
- (c) Perform the round  $\{(b_{(p+1)d_T}^j, b_{pd_T}^j)\}_{j=0}^{|I|-1}$ .

3. Transmit the remaining messages to the sink by performing rounds consisting of only one call and using shortest paths.

We observe that 2.b) is a valid round. Indeed, for any  $j = 0, \dots, |I| - 1$ , the shortest path from the sender  $b_{(p-i)d_T}^j$  to the receiver  $b_{(p+1+i)d_T}^{j+1 \bmod |I|}$  (resp. the sender  $b_{(p+2+i)d_T}^{j+1 \bmod |I|}$  to the receiver  $b_{(p-1-i)d_T}^j$ ), either goes through the sink and then its length is  $(2p+1)d_T > d_I$ , or it uses two paths of length  $d_I - d_T$  between vertices in layer  $pd_T$ , but then the distance is at least  $2(d_I - d_T) + d_T = d_I + d_I - d_T > d_I$ . The analysis for 2.a) and 2.c) are exactly the same.

We also notice that steps 1 and 3 require a number of rounds that does not depend on  $N$ , so altogether they use  $o(N)$  rounds. Step 2 is executed  $N/|I| + o(N)$  times and consists of  $1 + p|I|$  rounds, thus it consists of  $(1 + p|I|)N/|I| + o(N)$  rounds. Therefore, we conclude that for any independent set  $I$ , there exists a gathering protocol satisfying  $|R_I| = pN + \frac{N}{|I|} + o(N)$ . Hence, if  $I^*$  is such that  $|I^*| = \alpha(G)$ , then  $g(G_N, w, t) \leq |R_{I^*}|$  and we obtain

$$g(G_N, w, t) \leq pN + \frac{N}{\alpha(G)} + o(N).$$

**Extracting an independent set from a gathering protocol.** Now we consider the inverse process. First, observe that two calls  $(u_i, v_j)$  and  $(u_{i'}, v_{j'})$  such that  $i \leq pd_T$  and  $j' \leq (p-1)d_T$  interfere. Indeed,  $d_{G_N}(u_i, v_{j'}) \leq d_{G_N}(u_i, v_j) + d_{G_N}(v_j, t) + d_{G_N}(t, v_{j'}) \leq d_T + i + j' \leq d_T + pd_T + (p-1)d_T = 2pd_T \leq d_I$ .

Say that a call  $(u_i, v_j)$  is of type A if  $j \leq (p-1)d_T$ . Because each of the  $N$  messages coming from vertices  $s^k$  require at least  $p$  calls of type A to reach the sink, then any protocol performs at least  $pN$  calls of type A. But from the previous paragraph, any two calls of type A interfere with each other, so the  $pN$  calls of type A are performed in  $pN$  rounds, all different from each other.

Next, define a call  $(u_i, v_j)$  to be of type B if  $(p-1)d_T < j \leq pd_T < i$ . Clearly a call of type B is not of type A, but also any call of type B interferes with any call of type A (again see the first paragraph) meaning that a round cannot contain calls of both types.

Now, we lower bound the number of rounds that contains calls of type B. First, notice that two calls  $(u_i, v_j), (u_{i'}, v_{j'})$  of type B can be performed at the same time only if  $u$  and  $u'$  are independent in  $G$ . Indeed, if  $uu' \in E$ , then  $u_{pd_T}$  and  $u'_{pd_T}$  are connected with a path of length  $d_I - d_T$  and we have  $d_{G_N}(u_i, v_{j'}) = i - pd_T + d_I - d_T + pd_T - j' = d_I - d_T + i - j'$  and similarly  $d_{G_N}(u_{i'}, v_j) = d_I - d_T + i' - j$ . Hence,  $d_{G_N}(u_i, v_{j'}) + d_{G_N}(u_{i'}, v_j) = 2d_I - 2d_T + i - j + i' - j' \leq 2d_I - 2d_T + 2d_T = 2d_I$ , thus at least one of the two distances is  $\leq d_I$ . Therefore, at most  $\alpha(G)$  calls of type B are performed per round, but any protocol must perform at least one call of type B for each of the  $N$  messages of vertices  $s^k$ , therefore the protocol needs at least  $N/\alpha(G)$  rounds to perform all the calls of type B.

We have shown that any gathering protocol  $R$  satisfies  $|R| \geq pN + \frac{N}{\alpha(G)}$ . In particular, if  $R$  is optimum we have  $|R| = g(G_N, w, t)$ , and then

$$g(G_N, w, t) \geq pN + \frac{N}{\alpha(G)},$$

and the result follows.  $\square$

#### 4.4. Main result

**Theorem 2** MINIMUM GATHERING TIME is NP-HARD for any values of  $d_I, d_T$ . Moreover, if  $d_I > d_T$  the problem remains hard even for the uniform case where  $w = 1$ .

**Proof.** For  $d_I = d_T$ , the result follows from Lemma 5.

For  $d_I > d_T$ , we use Lemma 6 and Lemma 7. More precisely, by Lemma 6, a polynomial algorithm for MINIMUM GATHERING TIME would imply an approximation for MINIMUM VERTEX COLORING and similarly, by Lemma 7, a polynomial algorithm for MINIMUM GATHERING TIME would imply an approximation for MAXIMUM INDEPENDENT SET, but both problems are NP-HARD to approximate (see for example [1,19]).  $\square$

## 5. Conclusions

In this paper, we investigated the MINIMUM GATHERING TIME PROBLEM for a given graph  $G$  and sink  $t$  with interference distance  $d_I$  and transmission distance  $d_T$ . We proved that the problem is NP-HARD, even when the values of  $d_I, d_T$  are fixed. We also proposed a constant approximation algorithm. Some complexity issues remain open: Is the problem hard in the uniform case when  $d_I = d_T$ ? Does there exist a PTAS or a  $(1 + \varepsilon)$ -approximation algorithm for general graphs for small values of  $\varepsilon > 0$ ? In the case of particular topologies like trees or paths we can find an approximation close to 1, for example in the case of paths it is possible to give approximations up to an additive constant depending on  $d_I, d_T$  [3], but it is unclear if the problem is polynomial or not. A more practical question would be to study more dynamic cases (e.g. using on-line algorithms) or to derive algorithms that would not assume a global control but rely on local decisions (distributed algorithms).

## References

- [1] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs and non-approximability - towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.
- [2] J.-C. Bermond, R. Corrêa, and J. Yu. Gathering algorithms on paths under interference constraints. *6th Conference on Algorithms and Complexity, May 29-31, Rome*, 2006.
- [3] J.-C. Bermond, J. Galtier, R. Klasing, N. Morales, and S. Perennes. Gathering in specific radio networks. In *8èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel'06)*, Trégastel, May 2006.
- [4] J.-C. Bermond, L. Gargano, and S. Perennes. Optimal sequential gossiping by short messages. *Discrete Applied Mathematics*, 86:145–155, 1998.
- [5] J.-C. Bermond, L. Gargano, A.A. Rescigno, and U. Vaccaro. Fast gossiping by short messages. *SIAM Journal on Computing*, 27(4):917–941, 1998.

- [6] J.-C. Bermond, T. Kodate, and S. Perennes. Gossiping in Cayley graphs by packets. In *Proceedings of 8-th Franco-Japanese, 4-th Franco-Chinese Conference on Combinatorics and Computer Science*, volume 1120 of *Lecture Notes on Computer Science*, pages 301–315. Springer, 1995.
- [7] J.-C. Bermond and J. Peters. Efficient gathering in radio grids with interference. In *Septièmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel'05)*, pages 103–106, Presqu'île de Giens, May 2005.
- [8] P. Bertin, J.-F. Bresse, and B.L. Sage. Accès haut débit en zone rurale: une solution “ad hoc”. *France Telecom R&D*, 22:16–18, 2005.
- [9] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal of Selected Areas of Communication*, 18:535–547, 2000.
- [10] V. Bonifaci, P. Korteweg, A. Marchetti-Spaccamela, and L. Stougie. An approximation algorithm for the wireless gathering problem. In *10th Scandinavian Workshop on Algorithm Theory (SWAT)*, 2006.
- [11] I. Chlamtac and O. Weinstein. The wave expansion approach to broadcasting in multihop radio networks. *IEEE Transaction on Communications*, 39(3):426–433, 1991.
- [12] M. Christersson, L. Gasieniec, and A. Lingas. Gossiping with bounded size messages in ad-hoc radio networks. In *Proceedings of 29th International Colloquium on Automata, Languages and Programming (ICALP'02)*, volume 2380 of *Lecture Notes in Computer Science*, pages 377–389. Springer-Verlag, 2002.
- [13] M. Chrobak, L. Gasieniec, and W. Rytter. Fast broadcasting and gossiping in radio networks. *Journal of Algorithms*, 43(2):177–189, 2002.
- [14] S.A. Cook. The complexity of theorem proving procedures. In *Proc. 3rd Annual ACM Symposium on Theory of Computing*, pages 151–158, 1971.
- [15] M.L. Elkin and G. Kortsarz. Logarithmic inapproximability of the radio broadcast problem. *Journal of Algorithms*, 52(1):8–25, 2004.
- [16] I. Gaber and Y. Mansour. Centralized broadcast in multihop radio networks. *Journal of Algorithms*, 46(1):1–20, 2003.
- [17] J. Galtier. Optimizing the IEEE 802.11b performance using slow congestion window decrease. In *Proc. 16th ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*, pages 165–176, 2004.
- [18] L. Gasieniec and I. Potapov. Gossiping with unit messages in known radio networks. In *Proceedings of the IFIP 17th World Computer Congress - TC1 Stream / 2nd IFIP International Conference on Theoretical Computer Science*, pages 193–205. Kluwer, B.V., 2002.
- [19] J. Hastad. Clique is hard to approximate within  $n^{1-\epsilon}$ . *Acta Mathematica*, 182:105–142, 1999.
- [20] J. Hromkovic, R. Klasing, A. Pelc, P. Ruzicka, and W. Unger. *Dissemination of Information in Communication Networks: Broadcasting, Gossiping, Leader Election, and Fault-Tolerance*. Springer Monograph. Springer-Verlag, 2005.
- [21] R. Klasing, N. Morales, and S. Perennes. On the complexity of bandwidth allocation in radio networks with steady traffic demands. Technical report, INRIA Research Report RR-5432 and I3S Research Report I3S/RR-2004-40-FR, 2004.
- [22] P. Mühlethaler. *802.11 et les réseaux sans fil*. Eyrolles, 2002.