Broadcasting, Multicasting and Gossiping in Trees under the All-Port Line Model

Johanne Cohen*
Laboratoire de Recherche en Informatique
Université Paris Sud
91405 Orsay cedex, France
Email:cohen@lri.fr

Abstract

This paper is devoted to multi-point communication problems under the all-port *line model*. The line model assumes long distance calls between non neighboring processors. In this sense, the line model is strongly related to circuit-switched networks, wormhole routing, optical networks supporting wavelength division multiplexing, ATM switching, and networks supporting connected mode routing protocols. Since tree-networks are basic tools for the management of multi-point applications in both parallel systems and computer networks, we propose polynomial algorithms to derive optimal or near optimal broadcast, multicast and gossip protocols in trees.

*Additional support by the DRET of the DGA.

1 Introduction

Assume that every node of a network has some piece of information. Broadcasting is the information dissemination problem that consists, for one node of a network, to send its piece of information to all the other nodes of a communication network. Multicasting is the information dissemination problem in which the source sends its piece of information to an arbitrary group of destinations. In general, this group is different from the whole set of vertices. Gossiping is a simultaneous broadcast from every node of the network. Due to their complexity, theses three communication primitives are often provided at the software level. Most of communication libraries available on parallel systems (as MPI [25]) provide access to such communication procedures. More generally, theses three communication patterns are fundamental primitives used in many algorithms for the programming, and for the control of parallel and distributed systems. For example, they are used for barrier synchronization or cache coherence [30], for parallel search algorithm [10], and for linear algebra algorithms [12]. Finally, they are basic tools for the management of multi-point applications in computer networks [11].

In most of modern distributed memory parallel computers, and in many point-to-point LAN [8] or WAN [29], nodes communicate together using various types of switching technics such as *virtual cut-through* routing [19] (including *direct-connect* on several Intel's machines), *circuit-switching*, *wormhole* routing [24], wavelength division multiplexing [5], and ATM switching [18]. At a very abstract level, all theses switching modes perform as follows: when a node x sends a message to a non neighboring node y, a path is created between x and y in order to directly connect these two nodes. The message from x is then transmitted along this path. Intermediate nodes do not receive the message that goes through them (apart for path-based or other multicast technics [22, 23] that are not considered here).

In [13], Farley introduced the model called *Line Model* which satisfies the following: **(i)** a *call* involves exactly two nodes (these two nodes can be at distance more than 1), **(ii)** any two paths corresponding to simultaneous calls are edge-disjoint. Furthermore, Farley assumed that nodes satisfy the $1-$*port* hypothesis, that is: **(iii)** a node can take part in one call at a time. However, since nodes of many modern parallel and distributed systems can in fact send and receive through many ports simultaneously [22], and since the same holds for point-to-point networks [8], in this paper, we will replaced the $1-$port hypothesis by the all-port hypothesis, that is: **(iv)** a node can take part in many calls at a time. When two nodes are involved in the same call, they can exchange all the informations they are aware of (*full-duplex* mode).

In a point-to-point network, the probability of successful multicast delivery may decrease as the distance between sender and group members increases [8]. Moreover, the probability of damage, duplication, or misordering of multicast packets in a computer networks is very low, but not necessarily zero [8]. So, minimizing the number of "rounds" of a multicast (or broadcast) protocol allows the multicast scheme to increase the success of the delivery.

A *round* is the set of all calls carried out simultaneously. The complexity of our communication schemes will be measured by the number of communication rounds required to complete these schemes. For a given graph $G = (V, E)$, and for any arbitrary node u in G, we denote by $b(G, u)$ (resp. $g(G)$) the minimum number of rounds for broadcasting from the source node u (resp. for gossiping) in the graph G. Similarly, for any subset $X \subset V$, we denote by $m(G, X, u)$ the minimum number of rounds to multicast from u in X.

Furthermore, the usual architecture used for multicasting in computer networks is a shortest path directed tree between each sender and the group of destinations [8, 1]. The multicast packet passes through the edges of the tree. Thus, it is worth to study the complexity of broadcast, multicast, and gossip in tree-networks. This is the main purpose of this paper. In a more general setting, we will first consider undirected tree, and then we will show how to apply our broadcast and multicast protocols to directed trees. Of course, gossiping can be considered in undirected trees only.

In [13], Farley has proved that, in the $1-$port model, broadcast from any node in any $n-$node network can be performed in $\lceil \log_2 n \rceil$ rounds. His proof makes use of routing along the edge of a spanning tree of the network. Farley's theorem has been recently extended by Cohen, Fraigniaud, König and Raspaud [7] who showed that one can make the protocol as furtive as possible at each round, although the problem turns to be NP-complete when considering the whole protocol. Birchler, Esfahanian and Torng [4] derived independently similar results. In [3], the same authors have also studied the multicast problem in directed tree networks under the 1-port model. Their protocols minimize the total number of used links. The gossip problem is still open for arbitrary networks, although some results have been derived by Laforest [21] for tree-networks. Actually, the complexity of gossiping in the $1-$port line model in arbitrary network is not known. (It lies between $\lceil \log_2 n \rceil$ and $2\lceil \log_2 n \rceil - 1$, and both bounds are tight.)

In the literature relative to broadcasting and gossiping, one can also find many works under the all-port model. However, most of these works assume that the network has a specific topology as hypercube [15, 16], or torus [9, 26]. It is shown in [6] that broadcast and gossip are both NP-complete problems for arbitrary networks. This paper deals with arbitrary trees.

Our contribution

Theorem 1 *There exists an $O(n^2)$-time algorithm which returns, for any $n-$node tree, an optimal broadcast protocol in the all-port line model.*

Theorem 1 can be extended to directed tree (the edges are oriented from the source toward the leaves) and to the multicast problem:

Corollary 1 *There exists an $O(n^2)$-time algorithm which returns, for any undirected $n-$node tree, an optimal multicast protocol in the all-port line model for any source and any destination set.*

Corollary 2 *There exists an $O(n^2)$-time algorithm which returns, for any directed $n-$node tree rooted in u, an optimal multicast protocol from u in the all-port line model for any destination set.*

Moreover, we have shown that:

Theorem 2 *There exists an $O(n^3)$-time algorithm which returns, for any $n-$node tree, a near optimal gossip protocol in the all-port line model. (The algorithm is optimal up to an additive factor of* 1.*)*

Although we conjecture that the gossip problem can be polynomialy solved optimally for undirected tree-networks, we were not able to prove this fact, and we let it as an open problem.

The next section deals with broadcasting and multicasting in tree, whereas Section 3 deals with gossiping.

2 Broadcasting and multicasting in tree-networks

In this section, we describe a polynomial time algorithm to compute an optimal communication scheme for broadcasting from any arbitrary source node u of a tree T. In Section 2, T is considered as rooted in u. First, we will describe the algorithm, and then we will prove that the communication scheme generated by this algorithm is optimal in terms of rounds.

As in [17], for every node $v \neq u$ in T, we denote by T_v and $\overline{T}_v$ the two trees obtained by deleting the edge e containing vertex v, and such that the source u and vertex v do not belong to the same subtree. We assume that T_v contains vertex v. Moreover, we associate node v to the edge e by the function $\alpha : V \backslash \{u\} \rightarrow E$, such that $\alpha(v) = e$ if v is incident to e in T, and if the two trees obtained by deleting edge e are T_v and $\overline{T}_v$. Actually, $\alpha(v)$ is the first edge starting from v on the shortest path from v to u in T. For any graph G, we denote by $\Gamma_G(v)$ the set of neighboring vertices of v in G.

2.1 A broadcasting algorithm for undirected trees

2.1.1 Description of the algorithm

Our construction is recursive from the leaves to the root u of T. The leaves are at level 0. The level of a node is 1 plus the maximum of the levels of its children in T.

For any node v, our algorithm constructs a broadcast scheme $\mathcal{A}_v$ in T_v from an arbitrary node in $\overline{T_v}$. The communication scheme $\mathcal{A}_v$ is trivial if v is a leaf of T. Given a communication scheme $\mathcal{A}_v$ where v is a leaf, all communication schemes $\mathcal{A}_w$ are constructed for each node w at level 1. And so on. In other words, assuming that, for all nodes y of level at most ℓ, the communication scheme $\mathcal{A}_y$ is known, our algorithm constructs a communication scheme $\mathcal{A}_v$ where v is at level $\ell+1$ in T. The construction is based on all communication schemes $\mathcal{A}_y$ where y is a child of v in T_v. There will be a merging of not yet specified calls as $\{x \to ?\}$ (i.e., x calls some not yet specified vertex), and $\{? \to x\}$ (i.e., x is called by some not yet specified vertex).

To simplify the notation, the rounds are counted from the end of the broadcast scheme: round 1 is the last round, round 2 is the penultimate round, and so on. Let $\mathcal{A}_v[i]$ be the set of calls of the broadcast protocol $\mathcal{A}_v$ at the round i. Three cases are considered below: v is a leaf, v is an internal node, or v is the root u.

- v is a leaf. A single round is enough for v to receive the piece of information from a node in $\overline{T_v}$. Note then that the broadcast protocol generated by our algorithm imposes that all leaves are informed at the last round of the protocol (i.e., round 1). Formally, $\mathcal{A}_v$ is composed of a single call $\{? \to v\}$ at round 1. The sign "?" means that the sender of the call is unknown at this moment of the construction. It will be specified later. One only knows that node "?" is outside of T_v.
- v is vertex of level $\ell \geq 1$. Let us assume that for every vertex w of level less than ℓ, the protocol $\mathcal{A}_w$ is known. In particular, all protocols $\mathcal{A}_y$, where y is a child of v, are known. We merge these protocols $\mathcal{A}_y$ into a scheme $\mathcal{A}_v$ such that, at any round i, $\mathcal{A}_v[i] = \bigcup_{y \in \Gamma_{T_v}(v)} \mathcal{A}_y[i]$. For that purpose, every pair of calls of type $\{? \to x\}$ and of type $\{z \to ?\}$ in $\mathcal{A}_v$ is replaced by an unique call $\{z \to x\}$. After that, there is either no more call with unspecified end points, or there are either only calls of type $\{? \to x\}$, or of type $\{x \to ?\}$.

  Now, we have to check whether protocol $\mathcal{A}_v$ respects the constraints of the all-port line model. We also have to determine a possible round τ_v when v could be informed. We force that v is informed as soon as possible if at least one node in T_v can also call a node outside of T_v at the same round. We force that v is informed as late as possible otherwise. Let us formalize this strategy.

  Let t be the maximum number of rounds of the broadcast schemes $\mathcal{A}_y$, where y is a child of v. At the round i, we compute an integer value $\rho[i]$ as follows: $\rho[i]$ is the difference between the number of calls $\{? \to x\}$, and the number of calls $\{x \to ?\}$ in $\mathcal{A}_v[i]$. The value $\rho[i]$ indicates whether the broadcast communication scheme $\mathcal{A}_v$ respects the communication constraints. Indeed, there are five cases:

  - If $\rho[i] < -1$, then there are at least two calls $\{x \to ?\}$ in $\mathcal{A}_v$: it means that at least two nodes call a node in $\overline{T_v}$. Thus, one of theses calls can be used to inform node v ($\tau_v = i$), and the other call informs a node in $\overline{T_v}$ (which will be specified later). After this round, and until the end of the protocol, node v can give a call to each subtree T_y, and to $\overline{T_v}$. $\mathcal{A}_v$ satisfies the communication constraints.
  - If $\rho[i] = -1$, then there is an unique call of type $\{x \to ?\}$ from T_v to $\overline{T}_v$. The sender of this call can inform v or a node in $\overline{T}_v$ ($\tau_v \leq i$).
  - If $\rho[i] = 0$, then edge e is not used by the protocol $\mathcal{A}_v$ at this round. Thus node v has the possibility to receive the piece of information from a node in $\overline{T}_v$ ($\tau_v \leq i$).
  - If $\rho[i] = 1$, then a single call passes through edge e in order to inform a node in T_v. At this moment of the construction, the sender is not specified. At this round, the

communication constraints are respected. (The value τ_v is not modified because at this rounds, vertex v cannot be informed.)

  - If $\rho[i] > 1$, then at least two calls are of type $\{? \rightarrow x\}$. It means that the senders of these calls are in subtree $\overline{T_v}$, and they must inform nodes in T_v. Thus, these calls pass through the edge e (where $\alpha(v) = e$). Therefore, if it was not possible to inform node v before this round, then the communication constraint (iii) is not respected. We will have to check later whether there exists a possibility to inform v before this round. If it is not possible, one round must be added in order to inform v, and to satisfy the communication constraints.

  Now, we determine when v is informed.

  - If the previous process does not find a round when v can be informed ($\rho[i] \geq 1$ for every round i), then the broadcasting scheme $\mathcal{A}_v$ in T_v is performed in $t+1$ rounds as follows. The first node informed in T_v is v, that is $\mathcal{A}_v[t+1] = \{? \rightarrow v\}$. Then, at every forthcoming round of $\mathcal{A}_v$, node v can send the piece of information to a node in $\overline{T_v}$, and to a node in each subtree T_y. At the round i, $i \leq t$, a call $\{v \rightarrow ?\}$ is inserted into the scheme $\mathcal{A}_v[i]$, and every call of type $\{? \rightarrow x\}$ is replaced by $\{v \rightarrow x\}$.
  - If there exists a round τ_v during which node v can be informed, then we update the protocol $\mathcal{A}_v$ as previously.

- v is the source node ($v = u$). All the protocols $\mathcal{A}_y$, where y is a child of u, are merged into a scheme $\mathcal{A}_u$. Moreover, at any round, u can give a call to each subtree. We complete the broadcasting scheme $\mathcal{A}$ as follows:
  - Every call $\{? \rightarrow x\}$ is replaced by a call $\{u \rightarrow x\}$;
  - Every call $\{x \rightarrow ?\}$ is removed because such calls indicate that, at this round, x can take part as sender in a call outside of T, and that x already knows the piece of information.

This construction is formally described in Algorithm 1.

Time complexity of Algorithm 1. For every node v in T, the cost of the construction of the protocol $\mathcal{A}_v$ depends on the cost of the union of the protocols (i.e., the number of calls). Assuming that n is the number of vertices in T, the number of calls is equal to $n-1$. Therefore the time complexity of the construction of the protocol $\mathcal{A}_v$ is equal to $O(n)$. And hence the time complexity of our algorithm is $O(n^2)$.

2.1.2 Proof of optimality

This entire section is devoted to the proof of theorem 1, that is, we will prove the optimality of Algorithm 1. Given a broadcast scheme $\mathcal{X}$ from u in T, we denote by $\mathcal{X}^{(v)}$ the part of $\mathcal{X}$ such that the scheme $\mathcal{X}^{(v)}$ contains only calls which have a node of T_v as sender or receiver. Moreover, if a node in $\overline{T_v}$ takes part in a call of $\mathcal{X}^{(v)}$, then this node is represented by the sign "?" in $\mathcal{X}^{(v)}$. We express a broadcast protocol $\mathcal{X}^{(v)}$ by a couple $\mathcal{X}^{(v)} = (t, L)$ where t denotes the number of rounds of $\mathcal{X}^{(v)}$, and L is a list which indicates when and how edge $e = \alpha(v)$ is used:

- $L[i] = -1$ means that, at the round i, a call whose sender is in T_v passes through edge e.
- $L[i] = 1$, means that, at the round i, a call whose sender is in $\overline{T_v}$ passes through edge e.
- $L[i] = 0$ means that, at the round i, no call passes through edge e.

Moreover, we define an order $\preceq$ on the broadcast protocols. $\mathcal{X}^{(v)} \preceq \mathcal{Y}^{(v)}$ where $\mathcal{X}^{(v)} = (t, L)$, and $\mathcal{Y}^{(v)} = (t', L')$, if and only if one of theses three properties is satisfied:

- $t < t'$;

Algorithm 1 **Broadcast from u in T**

/∗ ***Initialization phase*** ∗/

Let h be the height of tree T rooted in u.

For each leaf v of T do

  /∗ ***The character ? means that the sender of this call is unknown at this step.*** ∗/

  $A_v[1] := \{? \to v\}$

End For

For $\ell := 1$ to $h-1$ do

  For each node v of level ℓ do

  /∗ *Construction of $\mathcal{A}_v$* ∗/

    Let t be the maximum number of rounds of $\mathcal{A}_y$ where $y \in \Gamma_{T_v}(v)$

    $i := t$ **and** $\tau_v := -1$

    $not\_end := true$

    While $(not\_end = true)$ **and** $(i > 0)$ **do**

      $\rho[i] = \left|\left\{\{? \to x\} \ s.t. \ \{? \to x\} \in \mathcal{A}_v[i]\right\}\right| - \left|\left\{\{y \to ?\} \ s.t. \ \{? \to x\} \in \mathcal{A}_v[i]\right\}\right|$

      case

        •$\rho[i] < -1$**: then** $\tau_v := i$ **and** $not\_end := false$

        /∗ ***There exist at least two calls $\{y \to ?\}$, one can inform v, the other can call a node in $\overline{T_v}$*** ∗/

        •$\rho[i] = -1$ **or** $\rho[i] = 0$**: then** $\tau_v := i$

        •$\rho[i] = 1$**: then** $noinstruction$

        •$\rho[i] > +1$**: then** $not\_end := false$

        /∗ ***There exist at least two calls of type $\{? \to y\}$, that must cross the edge connecting the two subtrees T_v and $\overline{T_v}$.*** ∗/

      Endcase

      i:=i-1;

    EndWhile

    /∗ ***Insertion of the call informing v*** ∗/

    if $(\tau_v = -1)$ **then** $\tau_v := t + 1$

    if $(\tau_v = t + 1)$ **or** $(\rho[\tau_v] = 0)$ **then** $\mathcal{A}_v[\tau_v] := \mathcal{A}_v[\tau_v] \cup \{? \to v\}$

      else replace a call $\{x \to ?\}$ **by a call** $\{x \to v\}$

    For each round $i := \tau_v - 1$ **down to** 1 **do**

      $A_v[i] := A_v[i] \cup \{v \to ?\}$

      every call $\{? \to x\}$ **is replaced by a call** $\{v \to x\}$

    End For

  End For

End For

/∗ *case $v = u$* ∗/

For each round $i := \tau_v - 1$ **down to** 1 **do**

  $A_u[i] := \cup_{y \in \Gamma_T(u)} A_y[i]$

  every call $\{? \to x\}$ **is replaced by a call** $\{u \to x\}$

End For

- $t = t'$, and there exists an integer k ($k \leq t$) such that $L[k] < L'[k]$, and $L[i] = L'[i]$ for all i, $k < i \leq t$;
- $t = t'$, and $L[i] = L'[i]$ for all i, $1 \leq i \leq t$.

Given two broadcast protocols $\mathcal{X}$ and $\mathcal{X}'$ from node u in tree T, $\mathcal{X}$ and $\mathcal{X}'$ are said to be *pseudo equivalent in* T_v if all calls of $\mathcal{X}'$ in which a node of a subtree of T_v is involved, are calls of the broadcast protocol $\mathcal{X}$, and conversely (i.e. for each child y of v in T_v, we have the equality $\mathcal{X}^{(y)} = \mathcal{X}'^{(y)}$). Note that the fact that $\mathcal{X}$ and $\mathcal{X}'$ are *pseudo equivalent in* T_v do not imply that $\mathcal{X}^{(v)} = \mathcal{X}'^{(v)}$ because v is not considered in the pseudo equivalence.

Lemma 1 *Let v be a vertex of T. Assume that v has d children v_i, $1 \leq i \leq d$, in T_v. Let $\mathcal{X}$, and $\mathcal{Y}$ be two broadcast schemes from u in T such that $\forall i \in \{1, \ldots, d\}, \mathcal{X}^{(v_i)} \preceq \mathcal{Y}^{(v_i)}$. There exists a broadcast scheme $\mathcal{X}'$ which is pseudo equivalent to $\mathcal{X}$ in T_v, and such that $\mathcal{X}'^{(v)} \preceq \mathcal{Y}^{(v)}$.*

Proof. Recall that the rounds are counted from the end of the broadcast scheme: round 1 is the last round, round 2 is the penultimate round, and so on. Let $t_\mathcal{X}$ (resp. $t_\mathcal{Y}$) be the maximum taken over all i of the number of rounds of schemes $\mathcal{X}^{(v_i)}$ (resp. $\mathcal{Y}^{(v_i)}$) where v_i is a child of v. We construct a protocol $\mathcal{X}'$ from $\mathcal{X}$. Let $\mathcal{X}'^{(v)} = (t_{\mathcal{X}'}, L_{\mathcal{X}'})$. Our construction is decomposed into two cases: $t_\mathcal{X} < t_\mathcal{Y}$, and $t_\mathcal{X} = t_\mathcal{Y}$.

Case 1. First, assume that $t_\mathcal{X} < t_\mathcal{Y}$. Moreover, we assume that the first call that informs a node of T_v in the protocol $\mathcal{X}$ is a call $\{w \rightarrow x\}$ where w and x are in $\overline{T_v}$ and T_v respectively. We transform the broadcast scheme $\mathcal{X}$ into a scheme $\mathcal{X}'$ as follows:

- all calls of the round k of $\mathcal{X}$ which are not in $\mathcal{X}^{(v)}$ are inserted in the protocol $\mathcal{X}'$ at the round $(k + 1)$.
- The first node informed in T_v is v in scheme $\mathcal{X}'$, that is, during the round $(t_\mathcal{X} + 1)$, node w sends the piece of information to node v. It implies that $L_{\mathcal{X}'}[t_\mathcal{X} + 1] = 1$.
- Finally, the scheme $\mathcal{X}'^{(v)}$ is defined as a copy of $\mathcal{X}^{(v)}$ where every call that has a vertex in $\overline{T_v}$ as sender, or as receiver, is modified in such a way that vertex v becomes the sender. Moreover, at the round j, $1 \leq j \leq t_\mathcal{X}$, node v call a node in $\overline{T_x}$, that is a call $\{v \rightarrow ?\}$ is inserted. By definition, it implies that, for all $i \in \{1, \ldots, t_\mathcal{X}\}$, $L_{\mathcal{X}'}[i] = -1$.

So, we have $\mathcal{X}'^{(v)} = (t_\mathcal{X} + 1, L_{\mathcal{X}'})$. Thus, by definition, we have $\mathcal{X}'^{(v)} \preceq \mathcal{Y}^{(v)}$.

Case 2. Now, assume that $t_\mathcal{X} = t_\mathcal{Y}$. For any node w in T, let us denote $\mathcal{X}^{(w)}$ and $\mathcal{Y}^{(w)}$ by (t_w, L_w) and (t'_w, L'_w) respectively. Due to the hypotheses of the lemma and to the definition of order $\preceq$, there exists an integer k such that, for all i, $k < i \leq t_\mathcal{X}$, $\rho_\mathcal{X}[i] = \rho_\mathcal{Y}[i]$ where $\rho_\mathcal{X}[i] = \sum_{y \in \Gamma_{T_v}(x)} L_y[i]$ and $\rho_\mathcal{Y}[i] = \sum_{y \in \Gamma_{T_v}(x)} L'_y[i]$ and $\rho_\mathcal{X}[k] < \rho_\mathcal{Y}[k]$.

We denote by $r_\mathcal{Y}$ the round at which node v is informed in $\mathcal{Y}$. We construct $\mathcal{X}'$ such that, (1) at the first round of the broadcasting, node u calls the father $v' \in \overline{T_v}$ of v in T, and (2) $\mathcal{X}'$ is a copy of $\mathcal{X}$ apart few modifications. We modify $\mathcal{X}'^{(v)}$ in a way depending on two cases: $k < r_\mathcal{Y} \leq t_\mathcal{X}$, and $r_\mathcal{Y} \leq k$.

- Assume that node v is informed at the round $r_\mathcal{Y}$ in $\mathcal{Y}$ where $r_\mathcal{Y} \leq k$. By definition, we have $\rho_\mathcal{X}[k] < \rho_\mathcal{Y}[k] < 1$. In the protocol $\mathcal{X}'$, we require that v is informed at the round k. This is done as follows. If $\rho_\mathcal{X}[k] \leq -1$ (resp. $\rho_\mathcal{X}[k] = 0$) , then a node of T_v (resp. node v') informs v. In any case, it is easy to see that $L_v[k] \leq L'_v[k]$. Thus, $\mathcal{X}'^{(v)} \preceq \mathcal{Y}^{(v)}$.
- Assume that node v is informed at the round $r_\mathcal{Y}$ in protocol $\mathcal{Y}$ where $k < r_\mathcal{Y} \leq t_\mathcal{X}$. By definition, we have $\rho_\mathcal{Y}[r_\mathcal{Y}] \leq 0$. As $\rho_\mathcal{Y}[r_\mathcal{Y}] = \rho_\mathcal{X}[r_\mathcal{Y}]$, we can modify $\mathcal{X}'^{(v)}$, such that during the round $r_\mathcal{Y}$, node v can be informed in $\mathcal{X}'$ by a node in T_v (resp. the node v' in $\overline{T_v}$) if v is informed in $\mathcal{Y}$ by a node in T_v (resp. in $\overline{T_v}$): we have $L_v[k] = L'_v[k]$. Afterwards, we transform the scheme $\mathcal{X}$ into $\mathcal{X}'$ in the same way we did in the case $t_\mathcal{X} < t_\mathcal{Y}$, and we have $\mathcal{X}'^{(v)} \preceq \mathcal{Y}^{(v)}$.

In both cases, $\mathcal{X}'^{(v)} \preceq \mathcal{Y}^{(v)}$, and the proof is completed. □

From this previous lemma, we can deduce Lemma 2, that is:

Lemma 2 *The broadcast scheme generated by the algorithm 1 is optimal in term of rounds.*

Proof. Recall that the rounds are counted from the end of the broadcast scheme. Let $\ell(v)$ be the level of node v in the tree T rooted in u. Let us denote by $\mathcal{A}$ the broadcast scheme generated by Algorithm 1, and by $\mathcal{A}_v$ all the intermediate schemes in nodes $v \neq u$. The proof is based on the order $\preceq$, and on the parameter $\ell(v)$. We will show property stating that, for any node v in the tree T, and for any broadcast scheme $\mathcal{X}$ from the source node u in T, we have $\mathcal{A}_v \preceq \mathcal{X}^{(v)}$. We prove this by induction on the level.

As the basis for our induction, let us consider the case where $\ell(v) = 0$ (i.e, v is a leaf in T). To broadcast from node u in the subtree T_v, we need one single call such that the vertex v is the receiver. And, assuming the broadcast scheme $\mathcal{A}_v$ is defined by the couple (t_v, L_v), we get $t_v = 1$, and $L_v[1] = 1$. Thus, the property holds for $\ell(v) = 0$.

Assume now that the property is true for any node w such that $\ell(w) < i$. Assume that $\ell(v) = i$. By induction, for any broadcast protocol $\mathcal{X}$, we have $\mathcal{A}_y \preceq \mathcal{X}^{(y)}$ where y is a child of v in T_v. Thanks to Lemma 1, there exists a broadcast protocol $\mathcal{A}'$ that is pseudo equivalent to $\mathcal{A}$ in T_v, and such that $\mathcal{A}'^{(v)} \preceq \mathcal{X}^{(v)}$ for any broadcast protocol $\mathcal{X}$. Let us show that $\mathcal{A}_v \preceq \mathcal{A}'^{(v)}$. Let us assume, for a purpose of contradiction, that there exists a broadcast scheme $\mathcal{A}'$ that is pseudo equivalent to $\mathcal{A}$ in T_v, and such that $\mathcal{A}'^{(v)} \prec \mathcal{A}_v$. Moreover, let us denote the broadcast schemes $\mathcal{A}_v$ and $\mathcal{A}'^{(v)}$ by (t_v, L_v) and (t'_v, L'_v) respectively. Similarly, for any child y of v, $\mathcal{A}_y$ is denoted by (t_y, L_y).

Let $\rho[i]$ and t be equal to $\sum_{y\in\Gamma(v)} L_y[i]$ and $\max_{y\in\Gamma_{T_v}(v)} t_y$, respectively. We consider two cases.

Case 1. First, we assume that the protocol $\mathcal{A}'^{(v)}$ requires $t+1$ rounds to broadcast in T_v from a node of $\overline{T_v}$. In the worst case, the scheme $\mathcal{A}_v$ requires $t+1$ rounds too. By construction of scheme $\mathcal{A}_v$, the first node informed in T_v is v, and then node v can send the piece of information to a node of $\overline{T_v}$ at every forthcoming round. Thus, $L_v[t+1] = 1$, and, for any $i \in \{1, \ldots, t\}$, $L_v[i] = -1$. Therefore, we have $\mathcal{A}_v \preceq \mathcal{A}'^{(v)}$, and there is a contradiction. Thus, in order the inequality $\mathcal{A}'^{(v)} \prec \mathcal{A}_v$ to be satisfied, $\mathcal{A}'$ must require exactly t rounds to broadcast in tree T_v from a node of $\overline{T_v}$.

Case 2. Now, we assume that the protocol $\mathcal{A}'^{(v)}$ requires t rounds to broadcast in T_v from a node of $\overline{T_v}$. We focus on the round τ'_v at which node v is informed in $\mathcal{A}'^{(v)}$. Because of communication constraints, during the round i, $\tau'_v \leq i \leq t$, we have $\rho[i] \leq 1$ since, otherwise, at least two calls would pass simultaneously through the edge $e = \alpha(v)$.

2.1 if there exists an integer i, $\tau'_v \leq i \leq t$, such that $\rho[i] < -1$, then Algorithm 1 imposes that v was informed at round i in $\mathcal{A}_v$. It implies that the scheme $\mathcal{A}_v$ informs v at round i, and therefore, for all j, $1 \leq j \leq i$, we have $L_v[j] = -1$. So, we have $\mathcal{A}_v \preceq \mathcal{A}'^{(v)}$, and there is a contradiction.

2.2 If there does not exist an integer i, $\tau'_v \leq i \leq t$, such that $\rho[i] < -1$, then, in particular, we have $0 \geq \rho[\tau'_v] \geq -1$. It means that Algorithm 1 detects the possibility that v can be informed at the round τ'_v in $\mathcal{A}_v$. If the scheme $\mathcal{A}_v$ does not inform node v at round τ'_v, then it informs node v at the round j where $j < \tau'_v$. It implies that $L_v[\tau'_v] = \rho[\tau'_v]$, and $L'_v[\tau'_v] = \rho[\tau'_v] + 1$. For example, if $\rho[\tau'_v] = -1$ then an internal call of T_v informs v. Since $L_v[\tau'_v] < L'_v[\tau'_v]$, we have $\mathcal{A}_v \prec \mathcal{A}'^{(v)}$, and there is a contradiction. If the scheme $\mathcal{A}$ informs node v at the round τ'_v, then $\mathcal{A}_v \preceq \mathcal{A}'_v$, and there is again a contradiction.

All the cases investigated above give rise to a contradiction. So for any node v in the tree T, and for any broadcast scheme $\mathcal{X}$ from the source node u in T, we have $\mathcal{A}_v \preceq \mathcal{X}^{(v)}$. Therefore, the lemma holds. □

2.2 Extension to multicast problems and to directed tree-networks

2.2.1 Broadcast in directed Tree.

We can notice that Algorithm 1 can be extended to the case where the tree T is directed from the source toward the leaves. The directed tree allows the source to call all nodes of the directed tree, and it allows an internal node v to call all nodes in T_v. Algorithm 1 must be modified because calls of type $\{v \rightarrow ?\}$ cannot be inserted in protocol $\mathcal{A}_v$ since such calls are directed from v to a node in $\overline{T_v}$, and therefore go upward the tree. Therefore, let us just consider the protocol $\mathcal{A}$ obtained using Algorithm 1 in which all calls of type $\{v \rightarrow ?\}$ are removed.

We claim that $\mathcal{A}$ is optimal in term of rounds. The proof is very similar to the proof of Theorem 1. Indeed, all broadcast schemes $\mathcal{A}_v$ can be described by a pair (t, L) as before. However, the list L never contains a negative value because a negative value corresponds to a call from T_v to $\overline{T_v}$. Furthermore, we can also define an order on the broadcast protocols as done in Section 2.1.2. This order allows Lemma 1 to be extended to directed trees. In the proof of Lemma 1 it is sufficient to not consider calls from T_v to $\overline{T_v}$.

2.2.2 Multicast problems.

Algorithm 1, and its extension, construct a broadcast scheme from node u in an undirected or directed tree. Both algorithms can be adapted in order to obtain a multicast scheme from a node u to a group of nodes D. This is true assuming the use of nodes not in D, these nodes will be only used to forward the piece of information. The adaptation to the multicast problem consists to modify the phase of the algorithms in which it is decided when v is informed. This phase is executed only if v is a node of D, or if there exists a round such that the difference between the number of calls of type $\{? \rightarrow v\}$, and the number of calls of type $\{v \rightarrow ?\}$ is strictly greater than 1, or strictly less than -1. It gives rise to Algorithm 2.

We do not prove that Algorithm 2 returns an optimal multicast protocol since the proof can be easily obtained from the proof described in Section 2.1.2.

3 Gossiping in undirected tree-networks

In this section, we prove a lower bound of the number of rounds required to gossip in any undirected tree. Afterwards, we describe an algorithm that returns a gossip protocol in T. This gossip protocol reaches the lower bound up to an additive constant factor of 1. First, let us give some notations.

- $b_{min}(T) = min_{v \text{vertex of } T} b(T, v)$.
- $B_{min}(T)$ is the set of vertices v of the tree T such that $b(T, v) = b_{min}(T)$.
- if a vertex v has d neighbors in T, then, by deleting v, we get d disjoint trees. We add vertex v to all theses trees, and we get d trees denoted by T_i^v. Moreover, these subtrees are ordered as follows: $T_i^v < T_j^v$ if and only if $b(T_i^v, v) \geq b(T_j^v, v)$. We denote by v_i the neighbor of vertex v such that it is in T_i^v.

Note that:

Remark 1 *For any vertex u in T, $b(T_1^u, u) = b(T, u)$.*

We get:

Lemma 3 *In any tree T of at least three vertices, there exists a vertex x of degree 2 in $B_{min}(T)$ such that $b(T_2^x, x) \geq b_{min}(T) - 1$.*

Proof. Assume, for a purpose of contradiction, that all vertices x in $B_{min}(T)$ satisfy $b(T_2^x, x) < b_{min}(T) - 1$. We will prove by induction on ℓ that there exists a path P of length ℓ such that

Algorithm 2 Multicast from u to a set D in a directed T

/ Initialization phase */*
Let h **be the height of tree** T **of root** u**.**
For each leaf v **of** T **do**
    / The character ? means that the sender of this call is unknown at this step. */*
    if $v \in D$ **then** $A_v[1] := \{? \to v\}$ **else** $A_v[1] := \emptyset$
End For
For $l := 1$ **to** $h - 1$ **do**
    For each node v **of level** l **do**
    / Construction of $\mathcal{A}_v$ */*
        Let t **be the maximum number of rounds of** $\mathcal{A}_y$ **where** $y \in \Gamma_{T_v}(v)$
        $i := t$ **and** $\tau_v := -1$
        $not\_end := true$
        While $(not\_end = true)$ **and** $(i > 0)$ **do**
            $\rho[i] = \left| \left\{ \{? \to x\} \ s.t. \ \{? \to x\} \in \mathcal{A}_v[i] \right\} \right|$
            case
                •$\rho[i] = 0$**: then** $\tau_v := i$
                •$\rho[i] > +1$**: then** $not\_end := false$
                / There exist at least two calls of type $\{? \to y\}$, that must cross the edge connecting the two subtrees T_v and $\overline{T_v}$. */*
            Endcase
            i:=i-1;
        EndWhile
        / Insertion of the call informing v */*
        if $(v \in D)$ **or** $(not\_end = false)$ **then**
            if $(\tau_v = -1)$ **then** $\tau_v := t + 1$
            $\mathcal{A}_v[\tau_v] := \mathcal{A}_v[\tau_v] \cup \{? \to v\}$
            For each round $i := \tau_v - 1$ **down to** 1 **do**
                every call $\{? \to x\}$ **is replaced by a call** $\{v \to x\}$
            End For
    End For
End For
/ case $v = u$ */*
For each round $i := \tau_v - 1$ **down to** 1 **do**
    $A_u[i] := \cup_{y \in \Gamma_T(u)} A_y[i]$
    every call $\{? \to x\}$ **is replaced by a call** $\{u \to x\}$
End For

1. all the vertices of P are in $B_{min}(T)$;
2. if $P = (p^1, \ldots, p^\ell)$, then, for all i, $i < \ell$, $p_1^i = p^{i+1}$ (p_1^i is the neighbor of p^i in $T_1^{p^i}$);
3. all the vertices of P are distinct.

Since $B_{min}(T) \neq \emptyset$, the property holds for $\ell = 0$. Assume that there exists a path P of length $\ell - 1$ satisfying the induction hypotheses. Let $P = (p^1, \ldots, p^{\ell-1})$, and $u = p^{\ell-1}$. By definition, vertex u is in $B_{min}(T)$. By Remark 1, we have $b(T_1^u, u) = b_{min}(T)$, and then we focus on the vertex u_1.

Assume that the vertex u_1 is a leaf. In this case, T_1^u is an edge (u, u_1). Therefore $b_{min}(T) = 1$ from the relation $b(T_1^u, u) = b_{min}(T)$. So, it is easy to see that tree T is an edge, and there is a contradiction with the hypothesis of the lemma. Therefore, vertex u_1 is not a leaf. We will prove that $u_1 \in B_{min}(T)$. Since $u \in B_{min}(T)$, there exist two broadcast protocols $\mathcal{A}_1$ and $\mathcal{A}_2$ where

- $\mathcal{A}_1$ is a broadcast protocol from u in T_1^u performed in at most $b_{min}(T)$ rounds;
- $\mathcal{A}_2$ is a broadcast protocol from u in $T \setminus T_1^u$ performed in at most $b_{min}(T) - 2$ rounds (because all vertices x in $B_{min}(T)$ satisfy $b(T_2^x, x) < b_{min}(T) - 1$).

Using $\mathcal{A}_1$ and $\mathcal{A}_2$, we construct a broadcast protocol $\mathcal{A}$ from u in T as follows. $\mathcal{A}_1$ and $\mathcal{A}_2$ are used simultaneously although $\mathcal{A}_1$ and $\mathcal{A}_2$ start at round 1 and round 2 of $\mathcal{A}$ respectively.

Let $\{u \to v\}$ be the call in T_1^u at the first round of $\mathcal{A}$. As u_1 is a vertex in a path connecting u with v in T, we can transform the protocol $\mathcal{A}$ into $\mathcal{A}'$ as follows. At the first round, the call $\{u \to v\}$ is replaced by a call $\{u \to u_1\}$, and a call $\{u_1 \to v\}$ is inserted at the first round. It is easy to see that the protocol $\mathcal{A}'$ performs the broadcast from u_1 in T in $b_{min}(T)$ rounds. So u_1 is a vertex of $B_{min}(T)$. Therefore, the path $P' = (p^1, \ldots, u, u_1)$ is a path of length ℓ which contains only nodes belonging to $B_{min}(T)$.

Now, to prove that all the vertices in P' are distinct, we will show that u_1 is not $p^{\ell-2}$. By hypothesis of induction, we have $u = p_1^{\ell-2}$. We also have $b(T', p^{\ell-2}) < b_{min}(T) - 1$ where T' is the subtree of T obtained by deletion of the edge $(p^{\ell-2}, u)$, and containing vertex $p^{\ell-2}$. Thus $b(T' \cup \{p^{\ell-2}, u\}, u) \leq b_{min}(T) - 1$. Thanks to remark 1, we get that u_1 is not $p^{\ell-2}$.

So, for any value of ℓ, there exists a path P of length ℓ such that all vertices in P are distinct. This is in contradiction with the fact that T has a bounded number of nodes. Thus there exists a vertex x in $B_{min}(T)$ such that $b(T_2^x, x) \geq b_{min}(T) - 1$. □

From Lemma 3, we derive the following lower bound.

Lemma 4 *Let T be a tree and x be a vertex of T such that $b(T_2^x, x) \geq b_{min}(T) - 1$. Then $g(T) \geq b_{min}(T) + b(T_2^x, x) - 1$.*

Proof. Thanks to lemma 3, there exists a vertex x of $B_{min}(T)$ such that $b(T_2^x, x) \geq b_{min}(T) - 1$ and $b(T_1^x, x) = b_{min}(T)$. At round $b(T_2^x, x) - 1$, there exists a piece of information of some vertex in T_2^x which is not known outside of T_2^x. The number of rounds to broadcast in T this piece of information from a node in T_2^x is at least the number of rounds to broadcast in T_1^x from x. Hence, gossiping in T needs at least $b(T_2^x, x) - 1 + b(T_1^x, x)$ rounds, and the lemma holds by application of remark 1 and lemma 3. □

Now, we will present a polynomial algorithm generating a near optimal gossip protocol. This algorithm is used to prove the upper bound stated in the following lemma.

Lemma 5 *let T be an tree of at least three vertices. There exists a vertex x in $B_{min}(T)$ such that $b_{min}(T) + b(T_2^x, x) - 1 \leq g(T) \leq b_{min}(T) + b(T_2^x, x)$*

Proof. Lemmas 3 and 4 give the lower bound. For the upper bound, we present a gossip protocol. First, one can select in polynomial time a vertex u in $B_{min}(T)$ such that $b(T_2^u, u) \geq b_{min}(T) - 1$. The gossip protocol performs as follows: all pieces of information of T are accumulated in u (Accumulation can be performed by just reversing a broadcast protocol). Then, u broadcasts all pieces of information in T. Let us count the number of rounds used by this protocol.

- If $b(T_2^u, u) = b_{min}(T)$, then this protocol performs the gossiping in $2b_{min}(T)$ rounds ($b_{min}(T) + b(T_2^u, u)$ rounds).
- If $b(T_2^u, u) = b_{min}(T) - 1$, then all pieces of information of T are collected in two vertices: node u and a node in T_1^u. During the round $b_{min}(T)$, they exchange their pieces of information. Indeed, this round corresponds to the last round of accumulation from T_1^u in u, and to the first round of broadcasting of the information of $T \setminus T_1^u$ from u in T_1^u. So, this protocol performs the gossiping in $2b_{min}(T) - 1$ rounds ($b_{min}(T) + b(T_2^u, u)$ rounds).

4 Further research

The broadcast and gossip problems are NP-complete in most of usual communication models considered in the literature. This is the case of the 1-port telephone model [28], and all-port line model [6]. Therefore, several approximation algorithms have been proposed (see for instance [2, 14, 20, 27]). The next step of this research is naturally to propose approximation algorithms for all-port (edge-disjoint) line model. $\lceil \log_2 n \rceil$ is an upper bound for broadcasting [13], but one can hope to do much faster, in particular for unbounded degree networks. Up to knowledge, this problem has not yet be investigated in arbitrary networks. Our result shows that, given a spanning tree of the networks, one can perform broadcast or gossip optimally on the tree. However, this complexity can be far from the best result (consider as a counter example an Hamiltonian graph with an Hamiltonian path as a spanning tree). A bread-first search spanning tree should provide better results but no proof of this fact has yet been derived.

References

[1] T. Ballardie, P. Francis, and J. Crowcroft. Core based trees. In *Proceedings of Communication Architecture and Protocols (SIGCOM)*, pages 85–96, 1993.

[2] D. Barth and P. Fraigniaud. Approximation algorithms for structured communication problem. In *10th ACM Symposium on Parallel Algorithms And Architectures (SPAA)*, 1997.

[3] B.D. Birchler, A. Esfahanian, and E. Torng. Information dissemination in restricted routing networks. In *Proceedings of the International Symposium on Combinatorics and Applications*, pages 33–44, 1996.

[4] B.D. Birchler, A. Esfahanian, and E. Torng. Sufficient conditions for optimal multicast communication. In *Proceedings of the International Conference on Parallel Processing*, pages 390–393, 1997.

[5] N.K. Cheung and al. Special issue on dense WDM networks. *IEEE Journal Selected Areas in Communications*, 8, 1990.

[6] J. Cohen. Broadcasting and gossiping under the all-port model. Technical report, LRI Université Paris sud, 91405 Orsay, France, 1998.

[7] J. Cohen, P. Fraigniaud, J.C. König, and A. Raspaud. Broadcasting and multicasting in cut-through routed networks. In *11th Int. Parallel Processing Symposium (IPPS)*, pages 734–738. IEEE, April 1997.

[8] S. Deering. Multicast routing in internetworks and extended LANs. In *Proceedings of Communication Architecture and Protocols (SIGCOM)*, volume 18, pages 88–100, 1988.

[9] O. Delmas and S. Perennes. Circuit-switched gossiping in 3-dimensionnal torus networks. In *Euro-Par'96 Parallel Processing*, number 1123 in Lecture Notes in Computer Science, pages 371–373. Springer, August 1997.

[10] R.F. DeMara and D.I. Moldovan. Performance indices for parallel marker-propagation. In *Proceedings of the International Conference on Parallel Processing*, pages 658–659, 1991.

[11] C. Diot, W. Dabbous, and J. Crowcroft. Multipoint communication: A survey of protocols, functions, and mechanisms. *IEEE journal on Selected Areas in Communications*, 15(3):277–290, April 1997.

[12] J.J. Dongarra and D.W. Walker. Software librairies for linear algebra computation onhigh performances computers. *SIAM review*, 37:151–180, 1995.

[13] A.M. Farley. Minimun-time broadcast networks. *Networks*, 10:59–70, 1980.

[14] P. Fraigniaud and S. Vial. Approximation algorithms for information dissemination problems. In *Proceedings of the IEEE second international conference on algorithms and architectures for parallel processing*, pages 155–162, 1996.

[15] S. Fujita, M. Yamashita, and T. Ae. Optimal group gossiping in hypercubes under wormhole routing model. In *Algorithms and computation*, volume 762 of *Lecture notes in Computer Science*, pages 276–286. Springer Verlag, December 1993.

[16] C-T. Ho and M-Y. Kao. Optimal broadcast in all-port wormhole-routed hypercubes. *IEEE Transactions on Parallel and Distributed Systems*, 6(2):200–204, 1995.

[17] J. Hromković, R. Klasing, W. Unger, and H. Wagener. Optimal algorithms for broadcasting and gossiping in the edge-disjoint path mode. In *4-th Scandinavian Workshop on Algorithm Theory (SWAT94)*, number 824 in Lecture Notes in Computer Science, pages 219–230. Springer-Verlag, 1994.

[18] C.C. Huang and P.K. McKinley. Communication issues in parallel computing across ATM networks. *IEEE Parallel and Distributed Technogy*, 1994.

[19] P. Kermani and L. Kleinrock. Virtual cut-through: a new computer communication switching technique. *Computers Networks*, 3:267–286, 1979.

[20] G. Kortsarz and D. Peleg. Approximation algorithms for minimum time broadcast. In *Proceedings of the 1992 Israel Symposium on Theory of Computer Science*, number 601 in Lecture Notes in Computer Science. Springer-Verlag, 1992.

[21] C. Laforest. Gossip in trees under line-communication mode. In *Euro-Par'96 Parallel Processing*, number 1123 in Lecture Notes in Computer Science, pages 371–373. Springer, August 1997.

[22] X. Lin, P. McKinley, and L. Ni. Deadlock-free multicast wormhole routing in 2D-mesh multicomputers. *IEEE Transactions on Parallel and Distributed Systems*, 5(8):793–804, 1994.

[23] X. Lin and L. M. Ni. Deadlock-free multicast wormhole routing in multicomputer networks. In *18th Annual International Symposium on Computer Architecture*, pages 116–125, Toronto, May 1991. ACM. (see also Technical Report MSU-CPS-ACS-29, 1990).

[24] L. Ni and P. McKinley. A survey of wormhole routing techniques in direct networks. *Computers*, 26(2):62–76, feb 1993.

[25] P. Pacheco. *Parallel programming with MPI.* Morgan Kaufmann, 1995.

[26] J.G. Peters and M. Syska. Circuit-switched broadcasting in torus networks. *IEEE Transactions on Parallel and Distributed Systems*, 7(3), March 1996.

[27] R. Ravi. Rapid rumor ramification: approximating the minimum broadcast time. In *35-th Symposium on Foundations of Computer Science*, pages 202–213. IEEE Computer Society Press, 1994.

[28] P. Slater, E Cockayne, and S. Hedetniemi. Information dissemination in trees. *SIAM Journal on Computing*, 10(4):692–701, 1981.

[29] A. Tanenbaum. *Computer Networks.* Prentice Hall, 1997.

[30] H. Xu, P. McKinley, and L. Ni. Efficient implementation of barrier synchronization in wormhole-routed hypercube multicomputers. *Journal of Parallel and Distributed Computing*, 16:172–184, 1992.