

Information Gathering in Ad-Hoc Radio Networks

Marek Chrobak

Department of Computer Science
University of California at Riverside

Kevin P. Costello

Department of Mathematics
University of California at Riverside

Leszek Gąsieniec

Department of Computer Science
University of Liverpool

Abstract

In the ad-hoc radio network model, nodes communicate with their neighbors via radio signals, without knowing the topology of the graph. We study the information gathering problem, where each node has a piece of information called a rumor, and the objective is to transmit all rumors to a designated target node. We provide an $\tilde{O}(n^{1.5})$ deterministic protocol for information gathering in ad-hoc radio networks, significantly improving the trivial bound of $O(n^2)$.

2012 ACM Subject Classification Discrete Mathematics → Combinatorics • Combinatorial Optimization • Theory of Computation → Design and Analysis of Algorithms → Distributed Algorithms • Networks → Ad-Hoc Networks

Keywords and phrases algorithms, radio networks, information dissemination

Digital Object Identifier 10.4230/LIPIcs...

Funding Research supported by NSF grant CCF-1536026.

1 Introduction

We address the problem of information gathering in ad-hoc radio networks. A *radio network* is represented by a directed graph, where nodes represent radio transmitters/receivers, and directed edges represent their transmission ranges; that is, an edge (u, v) is present in the graph if and only if node v is within the range of node u . When a node u transmits a message, this message is immediately sent out to all its out-neighbors. However, a message may be prevented from reach some out-neighbors of u if it collides with messages from other nodes. A collision occurs at a node v if two or more in-neighbors of v transmit at the same time, in which case v will not receive any of their messages, and it will not even know that they transmitted.

Radio networks, as defined above, constitute a useful abstract model for studying protocols for information dissemination in networks where communication is achieved via broadcast channels, as opposed to one-to-one links. Such networks do not need to necessarily utilize radio technology; for example, in local area networks based on the ethernet protocol all nodes communicate by broadcasting information through a shared carrier. Different variants of this model have been considered in the literature, depending on the assumptions about the node labels, on the knowledge of the underlying topology, and on allowed message size. In this work we assume that nodes are labelled $0, 1, \dots, n-1$, where n is the network size. We focus on the *ad-hoc model*, where the graph topology is unknown, and a protocol



© The copyright is retained by the authors;
licensed under Creative Commons License CC-BY

Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

needs to complete its task within a desired time bound, no matter what the topology is. We also do not make any assumptions about message size; thus, at any time a node can as well transmit all information it currently possesses.

Two most studied information dissemination primitives for this model are broadcasting and gossiping. In *broadcasting* (or *one-to-all communication*), a single source node s attempts to transmit its message to all nodes in the network. For broadcasting to be meaningful, we need to assume that all nodes in G are reachable from s . In *gossiping* (or *all-to-all communication*), each node has its own piece of information that we call a *rumor*. The objective is to distribute all rumors to all nodes in the network, under the assumption that the graph is strongly connected. Both these primitives can be solved in time $O(n^2)$ by a simple protocol called ROUNDROBIN where all nodes transmit cyclically one at a time (see Section 2). Past research on ad-hoc radio networks focussed on designing protocols that improve this trivial bound.

For broadcasting, gradual improvements in the running time have been reported since early 2000's [6, 17, 2, 3, 12, 11], culminating in the upper bound of $O(n \log D \log \log(D\Delta/n))$ in [10], where D denotes the diameter of G and Δ its maximum in-degree. This is already almost tight, as the lower bound of $\Omega(n \log D)$ is known [9]. For randomized algorithms, the gap between lower and upper bounds is also almost completely closed, see [1, 18, 11].

In case of gossiping, some major open problems remain. The upper bound of $O(n^2)$ was improved to $\tilde{O}(n^{3/2})$ in [6, 22] and then later to $\tilde{O}(n^{4/3})$ in [16], and no better bound is currently known¹. No lower bound better than the $\Omega(n \log n)$ bound that follows from [9] is known. In contrast, in the randomized case it is possible to achieve gossiping in time $\tilde{O}(n)$ [11, 19, 7].

The reader is referred to a survey paper [14] that contains more information about information dissemination protocols in different variants of radio networks.

In this paper we address the problem of *information gathering* (that is, *all-to-one communication*). In this problem, similar to gossiping, each node v has its own rumor, and the objective is to transmit these rumors to a designated target node t . (We assume that t is reachable from all nodes in G .) As indicated earlier, we are not assuming any bound on message size, so nodes are allowed to *aggregate* rumors; that is, they can combine all already received rumors, possibly adding other information, and transmit it all as one message.

The problem of information gathering for trees was introduced in [5], where an $O(n)$ -time algorithm was presented that uses message aggregation. Without aggregation (if each message can contain only one rumor), information gathering on trees can be solved in time $O(n \log \log n)$ [4]. Several other results in [5, 4] involve other variants of the problem, for example randomized algorithms or the model where message acknowledgements are provided to senders.

Our results. Our main result is a deterministic protocol that solves the information gathering problem in arbitrary ad-hoc networks in time $\tilde{O}(n^{1.5})$. To our knowledge this is the first such a protocol that achieves running time faster than the trivial $O(n^2)$ bound. The key new contribution is in solving this problem in time $\tilde{O}(n^{1.5})$ for acyclic graphs, where any protocols developed earlier for gossiping, that rely on feedback (see the discussion below), are not applicable. Our algorithm is based on careful application of combinatorial structures called strong selectors, combined with a novel amortization technique to measure progress of the algorithm. To extend this protocol to arbitrary graphs (with the target node t reachable

¹ We use notation $\tilde{O}(f(n))$ to conceal poly-logarithmic factors; that is, $g(n) = \tilde{O}(f(n))$ iff $g(n) = O(f(n) \log^c n)$ for some constant c .

from all nodes), we integrate it with a gossiping protocol. Roughly, the protocol for acyclic graphs is responsible for transferring information between strongly connected components, while a gossiping protocol disseminates it within each strongly connected component.

Additional context and motivations. If G is strongly connected then information gathering and gossiping are equivalent. Trivially, a gossiping algorithm gathers all rumors in t , solving the information gathering problem. On the other hand, one can solve the gossiping problem by running an information gathering protocol and then a broadcasting protocol with source node t . Thus, counter-intuitively, information gathering can be thought of as an *extension* of gossiping, since it applies to a broader class of graphs.

The crucial challenge in designing protocols for information gathering is *lack of feedback*, namely that the nodes in the network do not receive any information about the fate of their transmissions. This should be contrasted with the gossiping problem where, due to the assumption of strong connectivity, a node can eventually learn whether its earlier transmissions were successful. In fact, the existing protocols for gossiping critically rely on this feature, as they use it to identify nodes that have collected a large number of rumors, and then to broadcast these rumors (as one message) to the whole network, thus removing them from any further transmissions and reducing congestion.

Some evidence that feedback might help to speed up information gathering can be found in [4], where the authors developed an $O(n)$ -time protocol for trees if nodes receive (immediate) acknowledgements of successful transmissions, while the best known upper bound for this problem without feedback is $O(n \log \log n)$.

Various forms of feedback have been studied in the past in the problem of *contention resolution* for multiple-access channels (MAC), where a collection of nodes communicates via a single shared channel (ethernet is one example of such networks). Depending on more specific characteristics of this shared channel, one can model this problem as the information gathering problem either on a complete graph or a star graph, which is a collection of n nodes connected by directed edges to the target node t . (See [20, 21, 13] for information about contention resolution protocols.) For instance, in [5] a tight bound of $\Theta(n \log n)$ was given for randomized information gathering on star graphs (or MACs) even if the nodes have no labels (are indistinguishable) and receive no feedback.

Our algorithm relies critically on the rumor aggregation capability. This capability is needed to beat the $O(n^2)$ upper bound, as without rumor aggregation it is quite easy to show a lower bound of $\Omega(n^2)$, for both gossiping and information gathering, and even for randomized algorithms and with the topology known [15].

Interestingly, if we allow randomization, the randomized gossiping algorithms in [7, 19] can be adapted to information gathering without increasing the running time. Thus randomization can not only help to overcome collisions, but also lack of feedback.

2 Preliminaries

Graph terminology. Throughout the paper, we assume that the underlying radio network is represented by a directed acyclic graph G with a distinguished target node t that is reachable from all other nodes in G . By n we denote the number of nodes in G . We will treat G both as a set of vertices and edges, and write $u \in G$ if u is a node of G and $(u, v) \in G$ if (u, v) is an edge of G . If $(u, v) \in G$ then we refer to u as the *in-neighbor* of v and to v as the *out-neighbor* of u . For any node v , by $N^-(v) = \{u \in G : (u, v) \in G\}$ we denote the set of its in-neighbors.

For brevity, we will refer to strongly connected components of G as *sc-components*. For

each node v , the sc-component containing v will be denoted by $C(v)$. We partition the set of in-neighbors of v into those that belong to $C(v)$ and those that do not: $N_{\text{sc}}^-(v) = N^-(v) \cap C(v)$ and $N_{\text{acg}}^-(v) = N^-(v) - C(v)$. If $N_{\text{acg}}^-(v) = \emptyset$, we refer to v 's connected component $C(v)$ as a *source sc-component*.

The *extent* of v in G , denoted $G^-(v)$, is the set of all nodes of G from which v is reachable (via a directed path). We extend this definition in a natural way to sc-components of G ; if A is an sc-component then its extent is $G^-(v) = \cup_{v \in A} G^-(v)$.

Radio networks. As mentioned in the introduction we assume that each node of G has a unique label from the set $[n] = \{0, 1, \dots, n-1\}$. For convenience, we will identify nodes with their labels, so a “node u ” really means the node with label u .

The time is divided into discrete units that we refer to as *steps*, numbered with non-negative integers. We assume that all nodes start to execute the protocol simultaneously at time step 0. In the formal model of radio networks, at each step each node can be either in a *transmitting state*, when it can transmit a message (but it does not have to), or *receiving state*, when it can only listen to transmissions from other nodes. We will show below, however, that we can relax these restrictions and allow a node to simultaneously listen and transmit at each step. Only one message can be transmitted at each step. This is not an essential restriction because, as already mentioned in the introduction, we are not imposing any restrictions on the size or format of messages transmitted by nodes. However, a message transmitted at a given step cannot depend on messages received in the same step.

If a node u transmits a message at a time τ , this message reaches all out-neighbors of u in the same step. If v is one of these out-neighbors, and if u is the only in-neighbor of v that transmits at time τ , then v will receive this message. However, if there are two or more in-neighbors of v that transmit at time τ then a *collision occurs*, and v does not receive any information. In other words, for v , collisions are indistinguishable from absence of transmissions. There is no feedback mechanism available in this model, that is a sender of a message does not receive any information from the system as to whether the transmission was successful or not. (But it might learn this information later, indirectly, if there happens to be a path from the recipient to the sender in the graph.)

Selectors. A *strong (n, k) -selector* is a sequence of label sets $(S_0, S_1, \dots, S_{\ell-1})$ (that is, $S_i \subseteq [n]$ for each i) that “singles out” each label from each subset of at most k labels, in the following sense: for each $X \subseteq [n]$ with $|X| \leq k$ and each $x \in X$ there is an index i such that $S_i \cap X = \{x\}$. It is known [8] that there exist strong (n, k) -selectors of size $\ell = O(k^2 \log n)$.

Such selectors are often used for designing protocols for ad-hoc radio networks. The intuition is this: Consider a protocol that cyclically “runs” a strong (n, k) -selector; that is, each node u transmits in a step τ if and only if $u \in S_{\tau \bmod \ell}$. Suppose that u follows this protocol and starts transmitting its message at some time step. If v is an out-neighbor of u and v 's in-degree is at most k , then v will successfully receive u 's message in at most $O(k^2 \log n)$ steps, independently of the label assignment. Another basic protocol that is often used is called ROUNDROBIN. In this protocol all nodes transmit cyclically one by one; that is each node u transmits in a step τ if and only if $u = \tau \bmod n$. In ROUNDROBIN there are no collisions so, in the setting above, node u will successfully transmit its message to v in at most n time steps. Note that a protocol based on a strong (n, k) -selector can be faster than ROUNDROBIN only when $k = O(\sqrt{n/\log n})$.

Let $\theta = \frac{1}{2}(\log n - \log \log n) + 2$. For all $i = 0, 1, \dots, \theta - 2$, by 2^i -SELECT $= (S_0^i, S_1^i, \dots, S_{\ell_i-1}^i)$ we will denote a strong $(n, 2^i)$ -selector of size $\ell_i = O(4^i \log n)$. Without loss of generality we can assume that $\ell_{i+1} = 4\ell_i$ for all $i \leq \theta - 3$.

Note: To avoid clutter, in the paragraph above, as well as later throughout the paper,

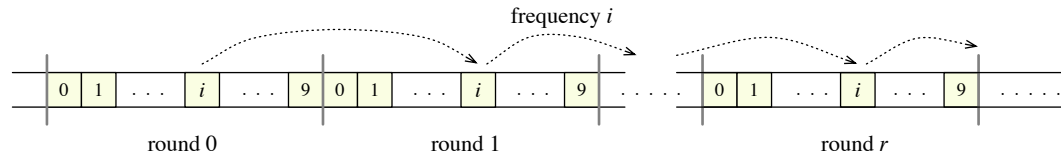
we omit the notation for rounding and assume that in all formulas representing integer quantities (the number of nodes, steps, etc.) their values are appropriately rounded. This will not affect asymptotic running time estimates.

Simplifying assumptions. To simplify presentation, in the paper we will assume a relaxed communication model with two additional features:

- (A) We assume that some number κ of radio *frequency channels* is available for communication. In a single step, a node can use all frequencies simultaneously. Further, for each frequency, it can also receive and transmit at this frequency in a single step. The restriction is that, for each step, the messages transmitted at all frequencies in a given step do not depend on the messages received in this step.

Below we explain how this relaxed model can be simulated using the standard radio network model.

Simulating multiple frequencies. We first explain how we can convert a protocol \mathcal{A} that uses κ frequencies and runs in time $O(T)$ into a protocol \mathcal{A}' that uses only one frequency and runs in time $O(\kappa T)$. This can be done by straightforward time multiplexing. In more detail: \mathcal{A}' organizes all time steps $0, 1, 2, \dots$ into *rounds*. Each round $r = 0, 1, 2, \dots$ consists of κ consecutive steps $r\kappa, r\kappa + 1, \dots, r\kappa + \kappa - 1$. Each step s of \mathcal{A} is simulated by round s of \mathcal{A}' . For each frequency i , the message transmitted at frequency i by \mathcal{A} is transmitted by \mathcal{A}' in step i of round s . At the end of round s , \mathcal{A}' will know all messages received in this round, so it will know what messages would \mathcal{A} receive in step s , and therefore it knows the state of \mathcal{A} and can determine the transmissions of \mathcal{A} in the next step.



■ **Figure 1** Partition of \mathcal{A}' 's time steps into rounds, for $\kappa = 10$ frequencies.

Simulating simultaneous receiving/transmitting. By the argument above, we can assume that we have only one frequency channel. We claim that we can disallow simultaneous receiving and transmitting at the cost of only adding a logarithmic factor to the running time. To see this, suppose that \mathcal{B} is some transmission protocol where nodes can transmit and listen at the same time. (Recall that the transmission of \mathcal{B} at any step does not depend on the information it receives in the same step.) We use a strong $(n, 2)$ -selector 2-SELECT of size $\ell_1 = \tilde{O}(1)$. We replace each step τ of \mathcal{B} by a time segment I_τ of length ℓ_1 . For any node u and any $j = 0, 1, \dots, |I_\tau| - 1$, if v is in the set of 2-SELECT with index j then at the j th step of segment I_τ node v transmits whatever message it would transmit in \mathcal{B} at time τ ; otherwise v is in the receiving state. Then, for any edge (u, v) , there will be a time step within I_τ at which u is in the transmitting state and v is in the receiving state, guaranteeing that u 's message will reach v . This way we convert \mathcal{B} into a protocol \mathcal{B}' where nodes do not transmit and receive at the same time.

In fact, for the type of protocols presented in the paper, allowing simultaneous reception and transmission does not affect the asymptotic running time at all. Our protocols are based on strong selectors and ROUNDROBIN. In case of ROUNDROBIN, the simultaneous reception and transmission capability is (trivially) not needed. For selector-based protocols,

the argument how this capability can be removed was given in [4]. Roughly, the idea is that whenever a protocol uses a strong (n, k) -selector, this selector can be replaced by a strong $(n, k + 1)$ -selector (whose size is asymptotically the same). This guarantees that for any node v with k in-neighbors and any v 's in-neighbor u there will be a step when v is in the receiving state and u is the only in-neighbor in the transmitting state.

3 Deterministic $\tilde{O}(n^{1.5})$ -Time Protocol for Acyclic Graphs

We first consider ad-hoc radio networks whose underlying graph G is acyclic and has one designated target node t that is reachable from all other nodes in G . We will give an information gathering protocol that will transmit all rumors to the target node t in time $\tilde{O}(n^{1.5})$ independently of the topology of G .

In the algorithm we will assume that each vertex knows the labels of its in-neighbors. This can be easily achieved in time $O(n)$ by pre-processing that consists of one cycle of ROUNDROBIN, where each node transmits only its own label. As explained in Section 2, we also make Assumption (A), namely that the protocol has multiple frequency channels available and on each frequency it can simultaneously receive and transmit messages at each step.

In the description of the algorithm below we use a sequence of $\theta + 1$ values $\beta_0, \beta_1, \dots, \beta_\theta$, defined as follows: $\beta_0 = 0$, $\beta_i = \sum_{g < i} \ell_g$ for $i = 1, \dots, \theta - 1$, and $\beta_\theta = \sum_{g < \theta} \ell_g + n$.

Protocol AccGather. The algorithm uses θ frequencies, where $\theta = \frac{1}{2}(\log n - \log \log n) + 2$, as defined in Section 2. The intuition is that each frequency $i < \theta - 1$ will be used to “run” selector 2^i -SELECT, while frequency $\theta - 1$ will be used to run ROUNDROBIN.

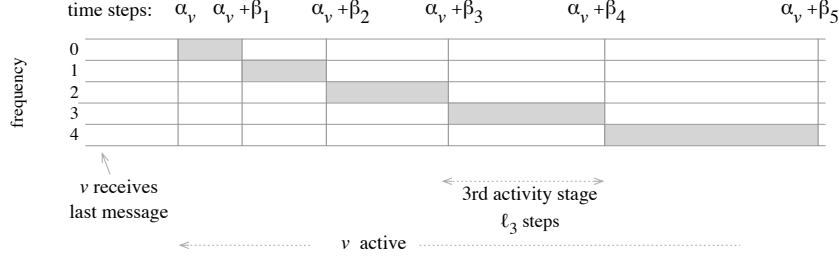
At each step, a node could be *dormant* or *active*. Dormant nodes do not transmit; active nodes may or may not transmit. A node v is active during its *activity period* $[\alpha(v), \alpha(v) + \beta_\theta]$, where $\alpha(v)$ is referred to as the *activation step* of v , and is defined below.

If v is a source node (that is, its in-degree is 0), then $\alpha(v) = 0$. Otherwise $\alpha(v)$ is determined by the messages received by v , as follows. Each message transmitted by a node u contains the following information: (i) all rumors collected by u , including its own, (ii) the label of u , and (iii) another value called *recommended wake-up step* and denoted rhs_u , to be defined shortly. For a non-source node v and its in-neighbor u , denote by $rhs_{u,v}^1$ the *first* rhs_u value received by v from u . (Note that this may not be the first rhs_u value *transmitted* by u , since these earlier transmissions might have failed.) Node v waits until it receives messages from all its in-neighbors, and, as soon as this happens, if u is the last in-neighbor of v that successfully transmitted to v , then v sets $\alpha(v) = rhs_{u,v}^1$.

The activity period $[\alpha(v), \alpha(v) + \beta_\theta]$ of v is divided into θ *activity stages*, where, for $i = 0, 1, \dots, \theta - 1$, the i th activity stage consists of the time interval $[\alpha(v) + \beta_i, \alpha(v) + \beta_{i+1}]$. (See Figure 2.) During its i th activity stage, for $i < \theta - 1$, node v transmits according to selector 2^i -SELECT using frequency i . During the $(\theta - 1)$ th activity stage, the protocol transmits using ROUNDROBIN on frequency $\theta - 1$. The recommended wake-up step value included in v 's message is $rhs_v = \alpha(v) + \beta_{i+1}$. At all other times v does not transmit.

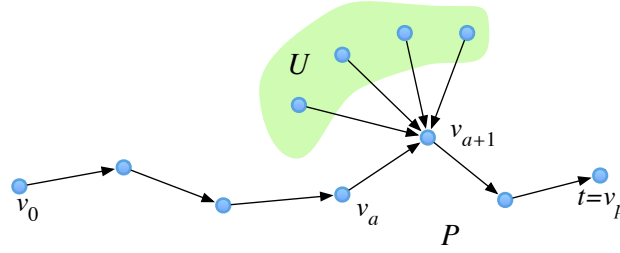
Correctness. We first note that the algorithm is correct, in the sense that each rumor will eventually reach the target node t . This is true because once a node becomes active, it is guaranteed to successfully transmit its message using the ROUNDROBIN protocol during its last activity stage.

Running time. Next, we derive an estimate of the completion time of ACCGATHER. We claim that this protocol completes information gathering in time $\tilde{O}(n^{3/2})$. To establish this bound, we choose in the graph G a *critical path* $P = (v_0, v_1, \dots, v_p = t)$, defined as follows:



■ **Figure 2** Illustration of activity stages. (The picture is not up to scale. In reality the length of activity stages increases at rate 4.) Shaded regions show frequencies used in different activity stages.

for each $a = p - 1, p - 2, \dots, 0$, v_a is the in-neighbor of v_{a+1} whose message was received last by v_{a+1} (formally, v_a is chosen so that $\alpha(v_{a+1}) = \text{rws}_{v_a, v_{a+1}}^1$), and v_0 is a source node. (Note that, since we define this path in the backwards order, the indexing of the nodes v_a can be determined only after we determine the whole path). Since v_{p-1} 's message is the last message received by t , it suffices to bound the time it takes for the algorithm to complete P .



■ **Figure 3** Illustration of the time analysis for acyclic graphs.

If at a current step a node v is in its i -th activity stage (that is, the current step is in the interval $[\alpha(v) + \beta_i, \alpha(v) + \beta_{i+1})$) then we refer to i as v 's *stage index*. We extend this (artificially) to dormant nodes as follows: if v has not yet started its activity period then its stage index is -1 , and if v has already completed its activity period then its stage index is θ . The stage index of each node is incremented $O(\log n)$ times, so the total number of these increments, over all nodes and over the whole computation, is $O(n \log n) = \tilde{O}(n)$.

Now consider some node v_a on P . (See Figure 4.) Our argument is based on the following key claim.

► **Claim 1.** There are $\tilde{\Omega}(n^{-1/2}(\alpha(v_{a+1}) - \alpha(v_a)))$ stage index increments in the time interval $[\alpha(v_a), \alpha(v_{a+1}))$.

Before we prove Claim 1, we argue that this claim is sufficient to establish our upper bound. Let T be the running time. Since $\alpha(v_0) = 0$ and $T \leq \alpha(v_p)$, we can bound the running time as $T \leq \sum_{a=0}^{p-1} (\alpha(v_{a+1}) - \alpha(v_a))$. Then Claim 1 implies that the total number of stage index increments during the computation is $\Omega(n^{-1/2}T)$. Since this number is also $\tilde{O}(n)$, it gives us that $T = \tilde{O}(n^{3/2})$.

Next, we prove Claim 1. Suppose that v_a succeeds first time in transmitting its message to v_{a+1} during its h -th activity stage.

► **Observation 1.** For $a < p$ and $h < \theta - 1$ we have $\alpha(v_{a+1}) - \alpha(v_a) = \tilde{O}(4^h)$.

This observation follows from the definition of P , as $\alpha(v_{a+1}) = rws_{v_a, v_{a+1}}^1 = \alpha(v_a) + \beta_{h+1}$, and $\beta_{h+1} = \sum_{g < h} \ell_g = O(4^h \log n)$.

We now consider three cases, depending on the value of h . First, if $h = 0$, then there is at least one stage increment in $[\alpha(v_a), \alpha(v_{a+1}))$ (namely the increment of the stage index of v_a from -1 to 0) and $\alpha(v_1) - \alpha(v_0) = \ell_0 = O(\log n)$, so the claim holds trivially.

Next, suppose that $0 < h < \theta - 1$. By the choice of h , v_a has not succeeded in its $(h-1)$ th activity stage $[\alpha(v_a) + \beta_{h-1}, \alpha(v_a) + \beta_h)$. Let U be the set of in-neighbors of v_{a+1} (including v_a) whose $(h-1)$ th activity stage overlapped that of v .

► **Observation 2.** $|U| > 2^{h-1}$.

To justify Observation 2, we argue by contradiction. Suppose that $|U| \leq 2^{h-1}$. During this activity stage v_a transmitted according to 2^{h-1} -SELECT using only frequency $h-1$. Further, by the definition of the protocol, at each step of this stage the in-neighbors of v_{a+1} with stage index other than $h-1$ did not use frequency $h-1$ for transmissions. So the transmissions from v_a to v_{a+1} can only conflict with transmissions from U to v_{a+1} . The definition of strong selectors and the assumption that $|U| \leq 2^{h-1}$ imply that then v_a would have successfully transmitted to v_{a+1} during its $(h-1)$ th activity stage, contradicting the definition of h . Thus Observation 2 is indeed true.

The $(h-1)$ th activity stage lasts ℓ_{h-1} rounds so all these $(h-1)$ th activity stages of the nodes in U end before time $\alpha(v_a) + \beta_h + \ell_{h-1} < \alpha(v_a) + \beta_{h+1} = \alpha(v_{a+1})$. This implies that in the interval $[\alpha(v_a), \alpha(v_{a+1}))$ the number of stage index increments is at least

$$|U| \geq 2^{h-1} = \frac{1}{2} \cdot 2^{-h} \cdot 4^h = \tilde{\Omega}(n^{-1/2}(\alpha(v_{a+1}) - \alpha(v_a))),$$

because $h \leq \frac{1}{2} \log n$ and $\alpha(v_{a+1}) - \alpha(v_a) = O(4^h)$.

Finally, consider the case when $h = \theta - 1$. Then $\alpha(v_{a+1}) - \alpha(v_a) = n$. But, by the choice of h , v_a has not succeeded in its $(h-1)$ th activity stage, where $h-1 = \frac{1}{2}(\log n - \log \log n)$. A similar argument as above gives us that the number of stage index increments during v_a 's $(h-1)$ th activity stage is $\Omega(n^{1/2})$, implying the Claim 1.

More precise time bound. We have established that Algorithm ACCGATHER runs in time $\tilde{O}(n^{1.5})$ on acyclic graphs. For a more precise bound, let us now determine the exponent of the logarithmic factor in this bound: one factor $O(\log n)$ is needed to simulate multiple frequencies with one, one factor $O(\log n)$ appears in the bound for the length of selectors, and we have another factor $O(\log n)$ that we ignored in the amortized analysis, since the number of stage index increments is $O(n \log n)$ (while we used the bound of $\tilde{O}(n)$). This gives us the main result of this section:

► **Theorem 1.** *Let G be an acyclic directed graph with n vertices and a designated target node reachable from all other nodes. Algorithm ACCGATHER completes information gathering on G in time $O(n^{1.5} \log^3 n)$.*

4 Deterministic $\tilde{O}(n^{1.5})$ -Time Protocol for Arbitrary Graphs

We now extend our information gathering protocol ACCGATHER from Section 3 to an arbitrary n -vertex directed graph G , achieving running time $\tilde{O}(n^{1.5})$. By t we denote the target node in G , and we assume that t is reachable from all other nodes in G .

The main obstacle we need to overcome is that protocol ACCGATHER depends critically on the graph being acyclic to coordinate the activity periods of different nodes. For instance,

in this protocol each node must wait until it receives messages from all its in-neighbors. If cycles are present, this leads to a deadlock, when each node in a cycle waits for its predecessor. On the other hand, as explained in the introduction, gossiping protocols from [6, 22, 16] do not work correctly if the graph is not strongly connected, because they rely on broadcasting to periodically flush out some rumors from the system and on leader election to synchronize computation. The idea behind our solution is to combine protocol ACCGATHER with the gossiping protocol from [16], using ACCGATHER to transmit information between different strongly connected components of G and using gossiping to disseminate information within strongly connected components. We also use gossiping to identify these strongly connected components.

Protocol GRXGossip for gossiping. We will refer to the gossiping algorithm from [16] as GRXGOSSIP. This protocol will be used as a black box, but we do exploit some properties of that algorithm, mainly the following two:

- (grx1) If the input graph is strongly connected and has k vertices, with the node labels from the set $[K] = \{0, 1, \dots, K - 1\}$, then algorithm GRXGOSSIP completes gossiping in time $O(k^{4/3} \log K \log^{10/3} k)$.
- (grx2) For arbitrary graphs (not necessarily strongly connected), after running algorithm GRXGOSSIP for time $O(k^{4/3} \log K \log^{10/3} k)$ the rumor from each node v will reach all nodes in the sc-component containing v .

Our algorithm will sometimes execute GRXGOSSIP on its sc-components, and we now give a more precise explanation of this process. Let A be an sc-component of G . To *execute* GRXGOSSIP on A means this: Starting at the same time, all nodes in A execute an instance of GRXGOSSIP in which the node labels are assumed to be from $[n]$ and the graph is assumed to be the subgraph of G induced by A . In particular, this means that this instance of GRXGOSSIP uses the value n_A as the graph size. Assuming that there is no interference from the nodes outside A , let $T_{\text{GRX}}(A)$ denote the running time of protocol GRXGOSSIP on A . By property (grx2), where $k = n_A$ and $K = n$, we have $T_{\text{GRX}}(A) = \tilde{O}(n_A^{4/3})$.

Executions of GRXGOSSIP for sc-components within predefined time intervals, that may be different for different sc-components. At time 0, each node v of an sc-component A partitions all time steps into disjoint *GRX-frames*, where the s th GRX-frame, for $s \geq 0$, is the time interval $[sT_{\text{GRX}}(A), (s+1)T_{\text{GRX}}(A))$. The “rumors” to be disseminated throughout A during this process will not be the original rumors from the instance, and will be specified in the algorithm’s description below.

Pre-processing. For convenience, we will assume that prior to the execution of the core algorithm, we will run a simple protocol, described below, that will identify sc-connected components in G . We start with one round of ROUNDROBIN, after which each node knows the labels of all its in-neighbors. This set of in-neighbors of a node v will be denoted $N^-(v)$. Next, we run the complete protocol GRXGOSSIP (on the whole graph), but instead of their original rumors, the nodes transmit their own labels. By property (grx2) above, after completing gossiping, each node will v receive a set Z_v of labels that includes the labels of all nodes in v ’s sc-connected component $C(v)$. We follow this with yet another execution of GRXGOSSIP, with each node u transmitting a pair consisting of its own label and set Z_u . When this is completed, each node v determines the labels of the nodes in $C(v)$ as follows: $u \in C(v)$ if and only if Z_u contains the label of v . Each node v now lets the identifier of $C(v)$ be the smallest label in $C(v)$. To complete the pre-processing we run ROUNDROBIN again, transmitting these identifiers. At this point each node v of G has the following information:

- The identifier and size $n_{C(v)}$ of its sc-connected component $C(v)$, as well as the list of all nodes in $C(v)$, and

- The list $N^-(v)$ of its in-neighbors, including their labels and the identifiers of their sc-connected components. We partition $N^-(v)$ into two types of nodes depending on whether they belong or not to $C(v)$: $N_{\text{sc}}^-(v) = N^-(v) \cap C(v)$ and $N_{\text{acg}}^-(v) = N^-(v) - C(v)$.

Algorithm ArbGather. Recall that protocol ACCGATHER used $\theta = O(\log n)$ frequencies for communication. We will use the same frequencies to simulate ACCGATHER and we will refer to them as ACG-frequencies. One additional frequency, called the SCC-frequency, will be used to simulate protocol GRXGOSSIP. The important thing to remember at this point is that, due to using different frequencies, these two protocols will never interfere with each other. As indicated earlier, GRXGOSSIP will be used to disseminate information within sc-components, so only messages from $N_{\text{sc}}^-(v)$ that are received by v on the SCC-frequency are used by the algorithm. Similarly, only messages from $N_{\text{acg}}^-(v)$ that are received by v on the ACG-frequencies are used. It may happen that v will receive messages on a “wrong” frequency, say a message from $N_{\text{sc}}^-(v)$ on some ACG-frequency or vice versa, but such messages are simply ignored by v .

It is convenient to think of the algorithm as running two subroutines, the SCC-subroutine and the ACG-subroutine, that interact with each other. Roughly, the completion of the ACG-subroutine by all nodes within an sc-component A triggers the SCC-subroutine for A , which runs GRXGOSSIP on A . If this SCC-subroutine completes, it triggers in turn the execution of ACG-subroutine for the nodes in A . In reality this is a bit more subtle, as we explain below.

The SCC-subroutine. Unlike in ACCGATHER, with each node v we now associate *two* activation times. The one relevant to the SCC-subroutine will be denoted $\alpha_{\text{sc}}(v)$ and called v 's *SCC-activation*. It is defined analogously to the activation time in ACCGATHER: If $N_{\text{acg}}^-(v) = \emptyset$ then $\alpha_{\text{sc}}(v) = 0$. Otherwise, $\alpha_{\text{sc}}(v)$ is the last-received value $\text{rws}^1(u, v)$ for $u \in N_{\text{acg}}^-(v)$, where $\text{rws}^1(u, v)$ denotes the first rws_u value received by v from u . (As explained earlier, these values will be received on the ACG-frequency.)

During the execution of this module, starting at time $\alpha_{\text{sc}}(v)$, v will be transmitting at the SCC-frequency. Let $A = C(v)$ and let s_v be the smallest s for which $s_v T_{\text{GRX}}(A) \geq \alpha_{\text{sc}}(v)$. Define the *A-rumor* of v to be a message that encapsulates all rumors received by v by time $\alpha_{\text{sc}}(v)$, plus the label of v itself. Node v repeatedly does the following: for $s = s_v, s_v + 1, \dots$, it executes GRXGOSSIP for A in the GRX-frame $[sT_{\text{GRX}}(A), (s+1)T_{\text{GRX}}(A))$, using its *A-rumor* as the rumor that needs to be disseminated by this execution of GRXGOSSIP. This process stops at time $\alpha_{\text{acg}}(A)$, which is equal to the beginning of the first GRX-frame after v receives *A-rumors* from all nodes in A .

The ACG-subroutine. We refer to the value $\alpha_{\text{acg}}(v)$ defined above as v 's *ACG-activation time*. This value now plays the role of v 's activation time in protocol ACCGATHER. In this subroutine v will transmit at the ACG-frequencies and v simply executes ACCGATHER, starting at time $\alpha_{\text{acg}}(v)$, in its activity period $[\alpha_{\text{acg}}(v), \alpha_{\text{acg}}(v) + \beta_\theta)$. The activity stages and the transmissions of each node are defined in exactly the same way as in protocol ACCGATHER (except that we use $\alpha_{\text{acg}}(v)$ instead of $\alpha(v)$).

Correctness. We justify correctness first. The following properties can be established by straightforward induction:

- First, a node v starts its SCC-subroutine, at time $\alpha_{\text{sc}}(v)$, only after it has received messages from all nodes in $N_{\text{acg}}^-(v)$, so when it starts, it will have all rumors from all nodes in $G^-(v)$. Thus at time $\alpha_{\text{sc}}(A)$, each rumor from $G^-(A)$ will already reach at least one node in A .

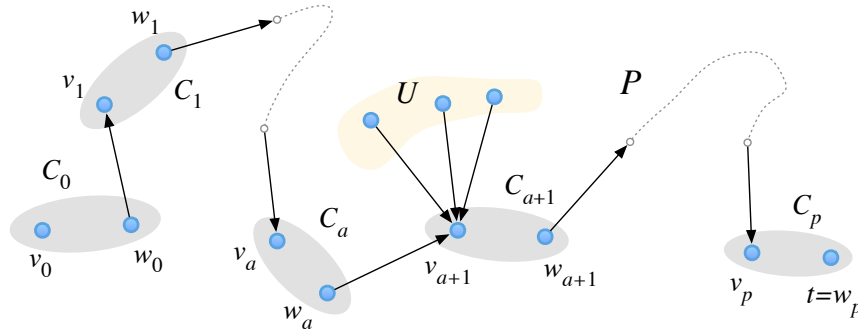
- Two, v will *not* receive all A -rumors from A before time $s_v T_{\text{GRX}}(A)$, because before time $s_v T_{\text{GRX}}(A)$ not all nodes in A yet started the execution of GRXGOSSIP in A .
- Three, as all nodes in A do a complete execution of GRXGOSSIP in A in the GRX-frame $[s T_{\text{GRX}}(A), (s+1) T_{\text{GRX}}(A))$, and no nodes interfere with this execution, GRXGOSSIP will successfully complete in time $T_{\text{GRX}}(A)$. This implies that $\alpha_{\text{acg}}(A) = \alpha_{\text{scc}}(A) + T_{\text{GRX}}(A)$, and that at time $\alpha_{\text{acg}}(A)$, v and all nodes in A know all rumors from $G^-(A)$.
- Four, v is guaranteed to successfully transmit during the ACG-subroutine because this subroutine involves a round of ROUNDROBIN.

Running time. Next, we estimate the running time. The main idea is this: When a node v starts its ACG-subroutine at time $\alpha_{\text{acg}}(v)$, the SCC-subroutine in $A = C(v)$ has already completed. By applying this property to the nodes in $N_{\text{acg}}^-(v)$, we obtain that when v starts its SCC-subroutine at time $\alpha_{\text{scc}}(v)$, the SCC-subroutines in all sc-components that are predecessors of v have already completed. Let $\alpha_{\text{scc}}(A) = \max_{u \in A} \alpha_{\text{scc}}(u)$. By the earlier observation, all nodes in A will already have all rumors from A and the preceding sc-components at time $\alpha_{\text{scc}}(A)$, and therefore the execution of GRXGOSSIP in A will be successful in the GRX-frame starting at $\alpha_{\text{scc}}(A)$. This implies that $\alpha_{\text{acg}}(A) \leq \alpha_{\text{scc}}(A) + 2$. (It might take two executions of ACCGATHER to complete the SCC-subroutine – one to distribute all gossips in A , and one more to distribute the success acknowledgements.)

The above paragraph implies that, for the nodes in A , the contribution per node of the SCC-subroutine to the overall running time is at most $2T_{\text{GRX}}(A)/n_A = \tilde{O}(n^{1/3})$. The analysis of the ACG-subroutine is the same as for protocol ACCGATHER.

To make this argument more precise, we extend the definition of a critical path from Section 3. In this section, the critical path is a sequence of nodes $v_0 w_0 v_1 w_1 \dots v_p w_p = t$ defined as follows:

- For each $a = p, p-1, \dots, 0$, suppose that w_a has already been defined, and let $C_a = C(w_a)$. If $\bigcup_{u \in C_a} N_{\text{acg}}^-(u) \neq \emptyset$, then let $v_a \in A$ be the node for which $\alpha_{\text{scc}}(v_a) = \alpha_{\text{scc}}(C_a)$. In other words, v_a is the node in C_a for which $\alpha_{\text{scc}}(v_a)$ is maximum. (It could happen that $v_a = w_a$.) On the other hand, if $\bigcup_{u \in C_a} N_{\text{acg}}^-(u) = \emptyset$ (that is, C_a is a source sc-component), then $a = 0$ and $v_0 \in C_a$ is arbitrary, for example we can take $v_0 = w_0$.
- For each $a = p-1, p-2, \dots, 0$, suppose that v_{a+1} has already been defined. Then w_a is the node in $N_{\text{acg}}^-(v_{a+1})$ whose message was received last by v_{a+1} (formally, w_a is chosen so that $\alpha_{\text{scc}}(v_{a+1}) = \text{rws}^1(w_a, v_{a+1})$).



■ **Figure 4** Illustration of the time analysis for arbitrary graphs.

Denote by T the running time of protocol ARBGATHER. We have $T = \alpha_{\text{acg}}(C_p)$ and $\alpha_{\text{scc}}(C_0) = 0$, so we can express T as

$$T = \sum_{a=0}^p [\alpha_{\text{acg}}(C_a) - \alpha_{\text{scc}}(C_a)] + \sum_{a=0}^{p-1} [\alpha_{\text{scc}}(C_{a+1}) - \alpha_{\text{acg}}(C_a)]$$

We estimate the two terms separately. As explained earlier, we have $\alpha_{\text{acg}}(C_a) = \alpha_{\text{scc}}(C_a) + T_{\text{GRX}}(C_a)$, so the first term is at most

$$\begin{aligned} \sum_{a=0}^p [\alpha_{\text{acg}}(C_a) - \alpha_{\text{scc}}(C_a)] &= \sum_{a=0}^p T_{\text{GRX}}(C_a) \\ &= \tilde{O}(n_{C_a}^{4/3}) = \tilde{O}(n^{4/3}), \end{aligned}$$

because the $\sum_{a=0}^p n_{C_a} \leq n$. To estimate the second term, note that the definition of v_{a+1} implies that $\alpha_{\text{scc}}(C_{a+1}) = \alpha_{\text{scc}}(v_a)$. Further, in the execution of ACCGATHER, node w_a gets activated at time $\alpha_{\text{acg}}(C_a)$. Then the analysis identical to that in Section 3 yields that we can estimate the second term by $\tilde{O}(n^{1.5})$. We thus obtain the main result of this paper:

► **Theorem 2.** *Let G be an arbitrary directed graph with n vertices and a designated target node reachable from all other nodes. Algorithm ARBGATHER completes information gathering in G in time $O(n^{1.5} \log^3 n)$.*

5 Final Comments

In this paper we provided an $\tilde{O}(n^{1.5})$ protocol for information gathering in ad-hoc radio networks, improving the trivial upper bound of $O(n^2)$.

Many open problems remain. The two most intriguing problems are about the time complexity of gossiping and of information gathering, as for both problems the best known lower bounds are only $\Omega(n \log n)$, the same as for broadcasting.

We hope that some ideas behind our algorithm will shed some light on these problems and help in further improvements. One idea that is particularly promising is the amortization technique in Section 3, where a failure of a node in transmitting its message is charged to the exponent in the size of the selectors used by the interfering nodes (that is, to stage index increments).

References

- 1 Noga Alon, Amotz Bar-Noy, Nathan Linial, and David Peleg. A lower bound for radio broadcast. *J. Comput. Syst. Sci.*, 43(2):290–298, 1991.
- 2 Danilo Bruschi and Massimiliano Del Pinto. Lower bounds for the broadcast problem in mobile radio networks. *Distributed Computing*, 10(3):129–135, 1997.
- 3 Bogdan S. Chlebus, Leszek Gasieniec, Alan Gibbons, Andrzej Pelc, and Wojciech Rytter. Deterministic broadcasting in ad hoc radio networks. *Distributed Computing*, 15(1):27–38, 2002.
- 4 Marek Chrobak and Kevin P. Costello. Faster information gathering in ad-hoc radio tree networks. *Algorithmica*, 80(3):1013–1040, 2018.
- 5 Marek Chrobak, Kevin P. Costello, Leszek Gasieniec, and Dariusz R. Kowalski. Information gathering in ad-hoc radio networks with tree topology. *Information and Computation*, 258:1–27, 2018.

- 6 Marek Chrobak, Leszek Gasieniec, and Wojciech Rytter. Fast broadcasting and gossiping in radio networks. *Journal of Algorithms*, 43(2):177–189, 2002.
- 7 Marek Chrobak, Leszek Gasieniec, and Wojciech Rytter. A randomized algorithm for gossiping in radio networks. *Networks*, 43(2):119–124, 2004.
- 8 Andrea E. F. Clementi, Angelo Monti, and Riccardo Silvestri. Selective families, superimposed codes, and broadcasting on unknown radio networks. In *Proc. 12th Annual Symposium on Discrete Algorithms (SODA'01)*, pages 709–718, 2001.
- 9 Andrea E. F. Clementi, Angelo Monti, and Riccardo Silvestri. Distributed broadcast in radio networks of unknown topology. *Theor. Comput. Sci.*, 302(1-3):337–364, 2003.
- 10 Artur Czumaj and Peter Davies. Faster deterministic communication in radio networks. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP'16)*, pages 139:1–139:14, 2016.
- 11 Artur Czumaj and Wojciech Rytter. Broadcasting algorithms in radio networks with unknown topology. *Journal of Algorithms*, 60(2):115 – 143, 2006.
- 12 Gianluca De Marco. Distributed broadcast in unknown radio networks. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'08)*, pages 208–217, 2008.
- 13 Antonio Fernández Anta, Miguel A. Mosteiro, and Jorge Ramón Muñoz. Unbounded contention resolution in multiple-access channels. *Algorithmica*, 67(3):295–314, 2013.
- 14 Leszek Gasieniec. On efficient gossiping in radio networks. In *16th Int. Colloquium on Structural Information and Communication Complexity (SIROCCO'09)*, pages 2–14, 2009.
- 15 Leszek Gasieniec and Igor Potapov. Gossiping with unit messages in known radio networks. In *Foundations of Information Technology in the Era of Networking and Mobile Computing, IFIP 17th World Computer Congress - TC1 Stream / 2nd IFIP International Conference on Theoretical Computer Science (TCS'02)*, pages 193–205, 2002.
- 16 Leszek Gasieniec, Tomasz Radzik, and Qin Xin. Faster deterministic gossiping in directed ad hoc radio networks. In *Scandinavian Workshop on Algorithm Theory (SWAT'04)*, pages 397–407, 2004.
- 17 Dariusz R. Kowalski and Andrzej Pelc. Faster deterministic broadcasting in ad hoc radio networks. *SIAM J. Discrete Math.*, 18(2):332–346, 2004.
- 18 Eyal Kushilevitz and Yishay Mansour. An $\Omega(D \log(N/D))$ lower bound for broadcast in radio networks. *SIAM J. Computg.*, 27(3):702–712, 1998.
- 19 Ding Liu and Manoj Prabhakaran. On randomized broadcasting and gossiping in radio networks. In *8th Annual Int. Conference on Computing and Combinatorics (COCOON'02)*, pages 340–349, 2002.
- 20 Gianluca De Marco and Dariusz R. Kowalski. Contention resolution in a non-synchronized multiple access channel. In *Proc. of the 27th Int. Symposium on Parallel Distributed Processing (IPDPS)*, pages 525–533, 2013.
- 21 Gianluca De Marco and Dariusz R. Kowalski. Fast nonadaptive deterministic algorithm for conflict resolution in a dynamic multiple-access channel. *SIAM Journal on Computing*, 44(3):868–888, 2015.
- 22 Ying Xu. An $O(n^{1.5})$ deterministic gossiping algorithm for radio networks. *Algorithmica*, 36(1):93–96, 2003.