

Beeping a Maximal Independent Set Fast*

Stephan Holzer¹ and Nancy Lynch²

- 1 MIT
holzer@mit.edu
- 2 MIT
lynch@csail.mit.edu

Abstract

We adapt a recent algorithm by Ghaffari [7] for computing a Maximal Independent Set in the LOCAL model, so that it works in the significantly weaker BEEP model. For networks with maximum degree Δ , our algorithm terminates locally within time $O((\log \Delta + \log(1/\varepsilon)) \cdot \log(1/\varepsilon))$, with probability at least $1 - \varepsilon$.¹ The key idea of the modification is to replace explicit messages about transmission probabilities with estimates based on the number of received messages.

After the successful introduction (and implicit use) of local analysis, e.g., in [2, 3, 7, 10], we study this concept in the BEEP model for the first time.

By doing so, we improve over local bounds that are implicitly derived from previous work (that uses traditional global analysis) on computing a Maximal Independent Set in the BEEP model for a large range of values of the parameter Δ . At the same time, we show that our algorithm in the BEEP model only needs to pay a $\log(1/\varepsilon)$ factor in the runtime compared to the best known MIS algorithm in the much more powerful LOCAL model. We demonstrate that this overhead is negligible, as communication via beeps can be implemented using significantly less resources than communication in the LOCAL model. In particular, when looking at implementing these models [14], one round of the LOCAL model needs at least $O(\Delta)$ time units, while one round in the BEEP model needs $O(\log \Delta)$ time units, an improvement that diminishes the loss of a $\log(1/\varepsilon)$ factor in most settings.

1998 ACM Subject Classification Dummy classification – please refer to <http://www.acm.org/about/class/ccs98-html>

Keywords and phrases Dummy keyword – please provide 1–5 keywords

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

1 Introduction

Computing a Maximal Independent Set (MIS) is a widely studied problem in distributed computing theory. One of the weakest models of communication in which this problem has been studied is the BEEP model, e.g., [1, 4, 22]. In this model, nodes can only send a beep

* Supported by: AFOSR Contract Number FA9550-13-1-0042, NSF Award 0939370-CCF, NSF Award CCF-1217506, and NSF Award CCF-AF-1461559.

¹ Errata note: In our brief announcement [13], we claimed as a side-effect of our local bound, the analysis of [7] can be used to show that this algorithm terminates globally within time $O(\log^2 \Delta) + 2^{O(\sqrt{\log \log n})}$ with high probability in n , the number of nodes in the network. While it is unknown whether this bound can be achieved, it is not clear that it can be derived via the graph-scattering technique used in [7] in combination with the deterministic algorithm of [19]. At least these techniques cannot be translated in the desired time in the beeping model as we thought of. The main reason is that in [19] nodes exchange more information than the BEEP model can handle in the time we hoped to achieve. Studying local complexity is of interest by itself as recently demonstrated by the papers cited in the abstract.



or listen in order to communicate, i.e., no sender-collision detection is available. In the version of this model with synchronized clocks, collision detection, and simultaneous wakeup, the authors of [1] showed how to compute an MIS w.h.p. in time $O(\log^2 n)$ by adapting a classical MIS algorithm by Luby [16]². Subsequently [22] showed how to improve this bound to $O(\log n)$ by adapting an improved version of Luby’s algorithm and tuning the probabilities of nodes joining the MIS over time.

Local analysis is a concept that was recently introduced or implicitly used, e.g., in [2, 3, 7, 10]. While traditional global complexity guarantees, that with high probability all nodes terminate after a certain runtime (that typically depends on the size n of the network), local complexity guarantees that each particular node v terminates with probability $1 - \varepsilon$ after a number of rounds, that typically depends only on the degree of the node (or network) and the probability $1 - \varepsilon$ that needs to be achieved. Local complexity is of interest in settings, where nodes could immediately continue with other computations once they found their own part of the solution to a problem. Constant probabilities often suffice, which results in speedups. This was demonstrated, e.g., in [10] that uses iterated computations of MIS in the SINR model, where only a very low local success probability is required that ultimately yielded an improvement of global broadcast in the SINR model, as previous approaches always worked with global w.h.p. guarantees. Further motivation on local analysis as an addition to global analysis can be found in [7].

We study the concept of local analysis in the BEEP model for the first time. This analysis improves over the canonical local complexity that can be derived from state-of-the-art algorithm and analysis [22] for a large range of values of the parameter Δ by reducing the runtime from $O(\log n)$ to $O((\log \Delta + \log(1/\varepsilon)) \cdot \log(1/\varepsilon))$ as we focus on local termination and local correctness:

► **Definition 1** (local $(1 - \varepsilon)$ -correctness of MIS algorithms). Let $\varepsilon > 0$. Any node v can produce output *IN* or *OUT* indicating v ’s membership to the MIS. This output cannot be revised once it is made.

- Local independence: If a node v outputs *IN*, then 1) with probability 1, none of v ’s neighbors previously decided to be *IN*, and 2) with probability at least $1 - \varepsilon$ none of v ’s neighbors decides *IN* at the same time, and 3) with probability 1 none of v ’s neighbors will output *IN* later.
- Maximality: If a node outputs *OUT*, then at that moment some neighbor is already *IN* the MIS. Each node will eventually make a decision with probability approaching 1.

► **Theorem 2** (Local $(1 - \varepsilon)$ -correctness (safety property)). *In our Algorithm of Section 3.2, when a node v terminates, it has made its (locally $(1 - \varepsilon)$ -correct) decision whether it is in the MIS or not.*

Note that this safety property is conditioned on the termination of the algorithm and the next Theorem states that the algorithm terminates in any specific node with probability $1 - \varepsilon$ within the stated time bounds.

► **Theorem 3** (Local termination complexity (liveness property)). Let $\varepsilon > 0$. *In our Algorithm of Section 3.2, for each node v , the probability that node v terminates within the first $O((\log \Delta + \log(1/\varepsilon)) \cdot \log(1/\varepsilon))$ slots and makes a locally $(1 - \varepsilon)$ -correct decision is at least $1 - \varepsilon$. This holds even if the outcome of the coin tosses outside $N_2^+(v) := V \setminus N_1(v)$ are determined adversarially.*

² They also consider other settings and argue that the assumption that all nodes wake up at the same time can be removed while keeping the same runtime.

We obtain this bound by adapting Ghaffari’s algorithm [7] for the LOCAL model to work in the BEEP model. The key idea in the proof and algorithm of Theorem 3 is that, instead of maintaining full information about its neighbors’ states, a node keeps a single binary estimate for the aggregate state of its entire neighborhood. In particular, the agents in the BEEP algorithm in this paper estimate probabilities by observing their neighbors’ probabilistically-generated transmission behavior. This results in improved runtimes and we show that the performance of the algorithms is close to the case in which the agents have exact information. If we shift our focus away from the pure notation of communication rounds or slots, it turns out that in many cases our new algorithms in the BEEP model are even more efficient than the original LOCAL model algorithm. The intuition for this is that it takes $\Omega(B\Delta)$ time units to emulate one round of the LOCAL model in the BEEP model in case the message size is B , whereas each time slot of the BEEP model can be emulated by one round in the LOCAL model.

Our approach is partly inspired by recent research on biological distributed algorithms such as [17], social networks, and other new forms of distributed algorithms. In such systems, agents do not obtain precise information about other agents’ preferences and tendencies, and estimate these from observing their behavior via sampling. It is interesting to understand how to improve efficiency by using stochastic information that arises from sampling distributions, rather than collecting exact information on these distributions.

► Remark. Note that this local bound is only a factor of $O(\log(1/\varepsilon))$ larger than the state-of-the-art $O(\log \Delta + \log(1/\varepsilon))$ bound in the LOCAL model [7].

1.1 Motivation, Related Work and Our Contribution

As pointed out above, communication in the LOCAL model and BEEP networks and their true implementation cost differ significantly. This requires us to be very careful when translating LOCAL algorithms into BEEP networks in order to not lose the strength of the techniques behind them and to obtain algorithms that are in many settings even more efficient due to the simpler nature of the BEEP model. The key difference between the models is, that within one communication round in the LOCAL model, a node can exchange arbitrarily large and different messages with all of its neighbors at the same time, while in BEEP networks a node can beep or not beep (this conveys less information than a conventional 1-bit message, where one has the option to send 0/1, or not send at all) and can only send or receive one message in the same slot. In the BEEP model, this message is restricted to contain one beep of information and therefore each round in the BEEP model needs much less resources than a round in the LOCAL model.

A straightforward and unfortunately inefficient way to emulate the behavior of one round in the LOCAL model in the BEEP model by performing a *local broadcast* of each node’s message that it would send in the LOCAL model. This takes $\Theta(B\Delta + \Delta \cdot \text{poly} \log(\Delta/\varepsilon_{ack}))$ time slots, where B denotes the size of the message in the LOCAL model, and the local broadcast is guaranteed to succeed with probability $1 - \varepsilon_{ack}$ (ack stands for acknowledged broadcast).

However, this simple technique of translating algorithms from the LOCAL model into the BEEP model is highly inefficient for algorithms that have small LOCAL runtimes, as Δ is a factor in the BEEP runtime. In particular, this is much higher than the (local termination) complexity of $O((\log \Delta + \log(1/\varepsilon)) \cdot \log(1/\varepsilon))$ that we achieve (note that ε is often a constant when using local complexity), which can be exponentially faster than the factor Δ lost by applying the simple transfer technique described above.

Finally, we remark that readers familiar with the $O(\log^* n)$ MIS algorithm of Schneider and Wattenhofer [21] in the LOCAL model for Bounded Growth Graphs might wonder why we did not translate their algorithm, as Bounded Growth Graphs capture most wireless network topologies in which the BEEP model is used. We show in Theorem 18 that their algorithm cannot be translated to the BEEP model without major modifications without paying a Δ factor in the runtime and would therefore be exponentially worse than our solution.

Local complexity:

model	time	probability	reference
BEEP	$O((\log \Delta + \log(1/\varepsilon)) \log(1/\varepsilon))$	$1 - \varepsilon$	Thm. 3
LOCAL	$O(\log \Delta + \log(1/\varepsilon))$	$1 - \varepsilon$	[7]

Global complexity:

model	time	probability	reference
BEEP	$O(\log n)$	w.h.p.	[22]
	$O(\log^2 n)$	w.h.p.	[1]
LOCAL	$O(\log \Delta) + 2^{O(\sqrt{\log \log n})}$	w.h.p.	[7]
	$O(\log n)$	w.h.p.	[16]
	$2^{O(\sqrt{\log n})}$	1	[19]

■ **Table 1** Overview of Results and Previous Work: Comparison of our results in the BEEP model with results in the LOCAL model and with previous work in the BEEP and LOCAL models. We include global bounds for completeness. While these global bounds yield w.h.p. guarantees, we want to stress that the key use of local analysis is when the success probability $1 - \varepsilon$ is small, e.g. constant. As demonstrated in [10], w.h.p. improvements on global broadcast can be achieved using local bounds on MIS computations that are locally successful with very low probability. Also our $\Omega(\Delta)$ lower bound in Theorem 18 on translating the Schneider Wattenhofer algorithm for Bounded Growth Graphs applies to global and local complexity, as this algorithm is deterministic, in which case these measures are the same.

2 Models and Definitions

Local and Beep Models: In both models, the network is abstracted as an undirected graph $G = (V, E)$ where $|V| = n$. All nodes wake up simultaneously. Communication occurs in synchronous rounds. In the LOCAL model (e.g., [7, 20]), each node knows its graph neighbors. Nodes communicate reliably, where in each round nodes can exchange an arbitrary amount of information with their immediate graph neighbors. On the other hand, in the BEEP model (e.g., [1, 4]), nodes do not know their neighbors. Nodes communicate reliably and a node can choose to either beep or listen. If a node v listens in slot³ t it can only distinguish between silence (no neighbor beeps in slot t) or the presence of one or more beeps (at least one neighbor beeps in in slot t).

Graph-related Definitions: We denote the set of h -hop neighbors of node v in G by $N_h(v) = \{u \in V \mid d(u, v) \leq h\}$, where $d(u, v)$ indicates the hop-distance between two nodes in a graph. By $\Delta := \max_{v \in V} |N_1(v)| - 1$ we denote the maximum degree of G . A set of

³ To disambiguate, we refer to the rounds of the BEEP model as *slots*.

vertices $I \subseteq V$ is an *independent set* of G if no two nodes in I are neighbors in G . An independent set $I \subseteq V$ is a *maximal independent set (MIS)* of G if, for all $v \in V \setminus I$, the set $I \cup \{v\}$ is not independent. An event occurs *with high probability* (w.h.p.), if it occurs with probability at least $1 - n^{-c}$ for some constant $c \geq 1$.

3 Algorithm

We first review the MIS Algorithm of [7] for the LOCAL model and then describe our modification for the BEEP model.

3.1 Algorithm of [7] in the Local Model

The MIS algorithm of [7] runs for

$$R := \beta(\log \Delta + \log(2/\varepsilon)) = O(\log \Delta + \log(1/\varepsilon))$$

rounds, where $\beta = 1300$. In each round t , each node v has a *desire-level* $p_t(v)$ for joining the MIS, which initially is set to $p_0(v) = 1/2$.

► **Definition 4** (Effective Degree, Ghaffari [7]). The sum of the desire-levels of neighbors of v is called its *effective-degree* $d_t(v)$, i.e., $d_t(v) = \sum_{u \in N(v)} p_t(u)$.

The desire-levels change over time:

$$p_{t+1}(v) = \begin{cases} p_t(v)/2, & \text{if } d_t(v) \geq 2, \\ \min\{2p_t(v), 1/2\}, & \text{if } d_t(v) < 2 \end{cases}.$$

The desire-levels are used as follows: In each round, node v gets *marked* with probability $p_t(v)$. If v is marked, and no neighbor of v is marked, v joins the MIS and gets removed along with its neighbors. Using the power of the LOCAL model, in each round t , nodes exchange exact values of $p_t(u)$ with all their neighbors.

3.2 Our Algorithm in the BEEP Model

In emulating the MIS algorithm of Section 3.1 in the BEEP model, we do not require that a node v learn the exact values of $p_t(u)$ for all neighbors u in order to compute $d_t(v)$. Instead, we allow node v to decide, based on how many beeps v receives within a certain number of rounds, whether $d_t(v)$ is more likely to be larger than $1/10$ or smaller than 22 . To estimate which of these two scenarios applies, node v beeps with probability $p_t(v)$ for a certain number of times and counts how often it received a beep when it is not sending. The number of received beeps serves as an indicator to estimate whether $d_t(v)$ might be smaller than 22 or larger than $1/10$. To perform this estimation, we define time intervals in the BEEP model. Eventually, a sequence of two intervals and one additional time slot is used to emulate each round of the LOCAL algorithm [7] in the BEEP model.

► **Definition 5** (Interval of slots). We define an *interval* to consist of

$$I := 2000(\ln(1500) + \ln(2/\varepsilon)) = O(\log(1/\varepsilon))$$

slots in the BEEP model.

During the first interval, the algorithm computes the ratio of the number of beeps received ($b_t(v)$) to the total number of slots in which v listened during the interval ($c_t(v)$) as follows: in each interval t , every node v maintains two counters $c_t(v)$ and $b_t(v)$. Counter $c_t(v)$ counts the number of slots that v is listening to the channel during interval t . Counter $b_t(v)$ counts the number of beeps v receives during interval t . Both counters $c_t(v)$ and $b_t(v)$ are initialized to 0 at the beginning of interval t . In each of the I slots of interval t , every node v decides randomly to beep with probability $p_t(v) \leq 1/2$. In each slot where v decides not to send, node v listens to the channel and increases $c_t(v)$ by one. If v receives a signal in this particular slot, node v increases its counter $b_t(v)$ by one. After all I time steps of interval t , node v compares $c_t(v)$ and $b_t(v)$. In case $c_t(v) \leq I/3$ we assume node v did not listen often enough to make an informed decision and let v randomly choose whether $b_t(v)/c_t(v) > \frac{1}{5}$ or not with probability $1/2$ for each choice – this is particularly important when $c_t(v) = 0$, as this avoids a division by 0. If $c_t(v) > I/3$, node v decides to update its desire-level:

$$p_{t+1}(v) = \begin{cases} p_t(v)/2, & \text{if } b_t(v)/c_t(v) > \frac{1}{5} \\ \min\{2p_t(v), 1/2\}, & \text{if } b_t(v)/c_t(v) \leq \frac{1}{5} \end{cases} . \quad (1)$$

Thus, we replace the condition $d_t(v) \geq 2$ in the algorithm of [7] by the condition $b_t(v)/c_t(v) > \frac{1}{5}$. The ratio is chosen to be $1/5$, as in the analysis it turns out that this is a good ratio in order to decide whether the effective degree $d_t(v)$ is larger than $1/10$ or smaller than 22 .

Notice that 1) these two ranges overlap, as we trade the uncertainty in making this decision for a shorter runtime while guaranteeing strong probabilities on correct decisions, and 2), that the overlap range $[1/10, 22]$ is chosen to capture a safety-distance around 2 that yields simple calculations in the proof.

For the sake of readability we replace β used in the definition of the number of rounds R in the LOCAL algorithm in Section 3.1 by γ , which is the analogous to constant β used in [7] and above when we analyze the BEEP algorithm. For the sake of simpler analysis, we set $\gamma := 80\beta = 104000$.

While in Ghaffari’s algorithm, in each round t , nodes exchange exact values of $p_t(u)$ with all their neighbors, we show how nodes estimate the value of $p_t(u)$.

During the second interval, a node decides whether to join a set M . Note that in Theorem 3 we state that M is locally an MIS with probability at least $1 - \varepsilon$. At the beginning of this second interval, a node v gets *marked* with probability $p_t(v)$ and does not change whether it is marked during the interval. If a node is marked in an interval, it selects half of the time slots in the interval uniformly at random and beeps in these time slots and listens in the others. If v is marked, and does not receive a beep in those time slots where v decides to listen, node v concludes that none of its neighbors is beeping and thus none of its neighbors is marked, and v joins M .

During the final time slot that completes the emulation of a round of the LOCAL algorithm, v beeps to indicate it joined M . In this time slot, each node that beeps or receives a beep gets removed, which corresponds to removing all nodes in M along with their neighbors in Ghaffari’s MIS algorithm.

4 Local Complexity of our MIS Algorithm

We demonstrate that for each node v , the accuracy of deciding whether v ’s effective degree is high or low is good enough for the translated algorithm of [7] to work correctly and fast in the BEEP model, i.e., our algorithm does not require v to learn exact desire-values of its neighbors. In Section 4.1 we define good nodes as those nodes that estimate the effective

degree (see Definition 4) accurately enough for our purposes, and bound the probability for a node being a good node in Lemma 7. In Section 4.2 we show that most of the time most nodes adjust their desire-values correctly in correspondence with the effective degree even they do not know its exact value. These Lemmas provide the tools for our modified analysis of [7] in Section 4.3.

4.1 For Most Nodes, Effective Degrees are Classified Correctly

We introduce the notion of good nodes in Definition 6, which are essentially nodes that correctly classify whether their effective degree is high or low. We show that if node v is good in interval r , node v (i) draws correct conclusions about whether its effective degree is high or low, and (ii) adjusts its desire-values in the same way as in the algorithm of [7]. The first statement follows directly from Definition 6 and the second materializes in the proof of Theorem 3. In Section 4.3, these insights will allow us to modify the analysis of [7] to obtain statements about good nodes. In order to apply this for a large part of the graph, we show that most nodes are good, which we use in Section 4.3 to obtain the desired statements about local complexity and correctness.

► **Definition 6** (Good node). A node v is a *good node* in an interval t , if at the end of the interval the following three conditions are satisfied:

1. $c_t(v) > I/3$, and
2. If $b_t(v)/c_t(v) > \frac{1}{5}$, then $d_t(v) \geq 1/10$, and
3. If $b_t(u)/c_t(v) \leq \frac{1}{5}$, then $d_t(v) \leq 22$.

The main result of this section is that many nodes are good, which is formalized as follows:

► **Lemma 7.** *For any node v and interval t , the probability that v is a good node is at least $1 - 2e^{-I/100}$.*

We prove Lemma 7 at the end of this Subsection. To prepare for the proof, we introduce two sub-Lemmas, Lemma 8 and 9, to bound the probability that $b_t(v)/c_t(v)$ reflects whether the effective degree is high or low based on the condition $b_t(v)/c_t(v) > \frac{1}{5}$ that we use rather than $d_t(v) \geq 2$ that is used by [7]. This differs from [7], as in the LOCAL model, full information on neighbor's effective degrees can be obtained within one round of communication, while we can only operate with beeps. These Lemmas use the Chernoff Bound stated in Lemma 20 in the Appendix for completeness.

The following Lemma states that for any node v , most of the time property 1. in Definition 6 is satisfied, such that v listens often enough to make an informed decision.

► **Lemma 8.** *For any node v and interval t , $\Pr(c_t(v) > I/3) > 1 - e^{-I/36}$.*

Proof. In each of the first I slots of interval t , the probability that v is not listening is $p_t(v)$, which is upper bounded by $1/2$ in the definition of our algorithm and we conclude $\mathbb{E}[c_t(v)] \geq I/2$. We apply Chernoff Bound 2. of Lemma 20 (see Appendix) for $X = c_t(v)$ with $\delta := 1/3$ and obtain that $\Pr(c_t(v) \leq I/3) \leq e^{-I/36}$. ◀

The next Lemma states that for any node v , most of the time property 2. or 3. in Definition 6 are satisfied.

► **Lemma 9.** *Assume $c_t(v) > I/3$ and let $b_t(v)/c_t(v)$ be the ratio computed by node v in an interval.*

1. If $b_t(v)/c_t(v) > \frac{1}{5}$, then $\Pr(d_t(v) \geq 1/10) \geq 1 - e^{-I/100}$, and
2. if $b_t(v)/c_t(v) \leq \frac{1}{5}$, then $\Pr(d_t(v) \leq 22) \geq 1 - e^{-I/100}$.

Notice that although $d_t(v)$ changes its value over time, we can bound $d_t(v)$ at time t with the probabilities stated in this Lemma independent of the history of $d_t(v)$.

Proof. For each time slot i of interval t , let $q_t(v)$ be the probability that at least one of the neighbors of node v beeps during slot i of interval t . Note that for each time slot i , the probability $q_t(v)$ is the same and choices are made for each i independently. In the following, for each t , we consider $c_t(v)$ to be fixed evaluations of the random variable describing them. Define independent random variables $X_1, \dots, X_{c_t(v)}$ for each of the $c_t(v)$ time slots in interval t during which v was listening, where $X_i = 0$ or 1 indicates whether v received a BEEP during the i 'th slot in which v listened. Define $X := \sum_{i=1}^{c_t(v)} X_i$, the random variable that indicates how many beeps v receives during interval t . We conclude that $\mathbb{E}[X] = q_t(v) \cdot c_t(v)$. Now observe that $b_t(v)$ is the sum of $c_t(v)$ evaluated random variables X_i , and thus an evaluation of X .

Proof of Statement 1: Assume $b_t(v)/c_t(v) > \frac{1}{5}$. We first show that in this case $\Pr[q_t(v) \geq 1/5] \geq 1 - e^{-I/66}$ and then derive the claimed statement on $d_t(v)$. Observe that the Bernoulli distribution X is defined using $q_t(v)$ and $c_t(v)$ only, and $\Pr[X/c_t(v) \leq 1/5]$ is a monotonically increasing function of $q_t(v)$ and vice versa. This allows us to turn the analysis around and analyze the probability that an evaluation of X is larger than $c_t(v)/5$ given $q_t(v)$, and draw conclusions on $q_t(v)$ from the event X is larger than $c_t(v)/5$. Therefore, we now assume for this part of the proof that $q_t(v) \leq 1/10$. In this case we know that $\Pr[X/c_t(v) \geq 1/5] = \Pr[X/c_t(v) \geq (1 + (\frac{1}{5q_t(v)} - 1))q_t(v)c_t(v)]$. We apply Chernoff Bound 1. of Lemma 20 (see Appendix) with $\delta = \frac{1}{5q_t(v)} - 1$, which we can do due to the assumption of $q_t(v) \leq 1/10$, and obtain that

$$\begin{aligned} & \Pr \left[X/c_t(v) \geq \left(1 + \left(\frac{1}{5q_t(v)} - 1 \right) \right) q_t(v)c_t(v) \right] = \Pr [X \leq (1 + \delta)\mathbb{E}[X]] \\ & \leq e^{-\frac{\delta^2}{2+\delta}\mathbb{E}[X]} = e^{-\frac{\left(\frac{1}{5q_t(v)} - 1\right)^2}{2 + \left(\frac{1}{5q_t(v)} - 1\right)} q_t(v)c_t(v)} = e^{-\frac{\left(\frac{1}{5} - 2q_t(v) + 5q_t(v)^2\right)}{1 + 5q_t(v)} c_t(v)} \leq e^{-\frac{c}{22}} \leq e^{-I/66} \end{aligned}$$

due to assuming $q_t(v) \leq 1/10$ and $c_t(v) > I/3$ (see statement of the Lemma). From this we derive that if $X \geq c_t(v)/5$, then $\Pr[q_t(v) \geq 1/10] \geq 1 - e^{-I/60}$. Now we lower bound $d_t(v)$ based on $q_t(v)$. We know that $q_t(v) = 1 - \prod_{u \in N(v)} (1 - p_t(u))$ and $d_t(v) = \sum_{u \in N(v)} p_t(u)$. As the $d_t(v)$ is minimized when the whole probability mass of $q_t(v)$ is aggregated in one node, we conclude that $\Pr[d_t(v) \geq 1/10] \geq 1 - e^{-I/66} \geq 1 - e^{-I/60}$.

Proof of Statement 2: Assume $b_t(v)/c_t(v) \leq \frac{1}{5}$. In the following proof, for each time slot of interval t , we lower bound the probability that the distribution-probability $q_t(v)$ is smaller than $1/2$. Then we upper bound $d_t(v)$ based on $q_t(v)$. To be more specific, we first show that in case an evaluation of X is smaller than $c_t(v)/5$ it is $\Pr[q_t(v) \leq 1/2] \geq 1 - e^{-I/100}$.

As argued above, we can turn the analysis around and assume $q_t(v) > 1/2$ and show in the text below, that under this assumption, $\Pr[X/c_t(v) > \frac{1}{5}] \geq 1 - e^{-I/90}$ and in turn draw conclusions on the probability of $q_t(v) > 1/2$ when $X/c_t(v) > \frac{1}{5}$ is given. Assuming $q_t(v) > 1/2$, we derive $\Pr[X/c_t(v) > \frac{1}{5}] = \Pr[X \leq c_t(v)/5] \leq \Pr[X \leq 2/5q_t(v)c_t(v)] =$ when $\Pr[X \leq 2/5\mathbb{E}[X]]$, as $\mathbb{E}[X] = q_t(v)c_t(v)$. We apply Chernoff Bound 1. of Lemma 20 (see Appendix) to upper bound the probability that an evaluation of the random variable X

is larger than 2 times its expectation, which we achieve with $\delta = 3/5$

$$\Pr[X \leq 2/5\mathbb{E}[X]] = \Pr[X \leq (1 - \delta)\mathbb{E}[X]] \leq e^{-\frac{\delta^2}{2+\delta}\mathbb{E}[X]} = e^{-\frac{9}{65}q_t(v)c_t(v)}$$

Using the assumption that $q_t(v) > 1/2$ and the assumption that $c_t(v) > I/3$ (see Lemma statement), we can bound this further by $e^{-3I/130} \leq e^{-I/100}$. Now we know that if $q_t(v) > 1/2$, then $\Pr[X/c_t(v) \leq 1/5] \leq e^{-I/100}$. We conclude that if $X/c_t(v) \leq 1/5$, then $\Pr[q_t(v) > 1/2] \leq e^{-I/100}$, as X is defined using $q_t(v)$ and $c_t(v)$ only, and defined in a way that the implied Bernoulli distribution monotonically increases when $q_t(x)$ increases. From this we derive that if $X/c_t(v) \leq 1/5$, then $\Pr[q_t(v) \leq 1/2] \geq 1 - e^{-I/100}$.

Next, we know that $q_t(v) = 1 - \prod_{u \in N(v)} (1 - p_t(u))$ and $d_t(u) = \sum_{u \in N(v)} p_t(u)$. As the whole probability mass of $q_t(v)$ could be distributed evenly among v 's neighbors, the worst case is $q_t(v) = 1 - (1 - p_t)^\Delta \geq 1 - e^{-\Delta/p_t}$ with $p_t(u) = p_t$ for any $u \in N(v)$. From this we conclude that $1/2 \geq 1 - e^{-\Delta/p_t}$, which in turn yields $p_t \leq -\Delta/\log(1/2) \leq 5\Delta < 22\Delta$.

Due to $q_t(v) \leq 1/2$ with probability $1 - e^{-I/100}$, we conclude $p_t/\Delta \leq 22$. The definition of $d_t(v)$ yields $d_t(v) \leq 22$. From this we conclude $\Pr[d_t(v) \leq 22] \geq 1 - e^{-I} \geq 1 - e^{-I/100}$. ◀

Now we are ready to prove Lemma 7, which follows from combining Lemma 8 and Lemma 9 to cover all properties of Definition 6 and multiplying the probabilities of the related events stated in the Lemmas we use. We state the full proof in the full version of the paper, see Appendix, Lemma 7.

Proof. (of Lemma 7). Due to Lemma 8, we know that $\Pr(c_t(u) > I/3) > 1 - e^{-I/36}$, such that property (1) of a good node (Definition 6) is satisfied. Now we can assume $c_t(u) > I/3$ with probability $1 - e^{-I/36}$, then the probability that properties (2) and (3) of a good node are satisfied is at least $1 - e^{-I/100}$ each due to Lemma 9. We conclude that all three conditions are satisfied for node u in an interval with probability larger than $(1 - e^{-I/36})(1 - e^{-I/100}) > 1 - 2e^{-I/100}$. Finally, notice that $b_t(v)$ is a realization of the random variable X . ◀

4.2 Changes of Effective Degrees Based on Neighbor's Behavior

We show in Lemma 10 that in any interval, the effective degree of a node v (see Definition 4), that is contributed by a set of neighbors with high effective degree in that interval, shrinks by almost a factor of 2 with significant probability. This is a key part in the modification of the analysis of [7] in Section 4.3. This is stated in a formal way in Lemma 10 and proven using Lemma 7.

The following Lemma is a key Lemma. It allows us to bound the amount of $d_{t+1}(v)$ that is contributed by neighbors of v with high effective degree based on the amount of $d_t(v)$ that is contributed by neighbors of v with high effective degree. A similar bound is used in [7], where it is obtained in a straight forward way thanks to the power of the LOCAL model. We need to (and already did) work a bit harder to obtain a similarly useful bound. This Lemma shows that the precise way of increasing/reducing the desire value of a node based on $d_t(v)$, which requires full knowledge of all these values of neighbors of v in [7], can be replaced by estimating $d_t(v)$ using beeps. Of course the bound is less strong as in [7], we obtain only a decrease of 51/100 vs. 1/2 in [7]. Also, as we operate probabilistically, we can only claim this bound with a certain probability. Fortunately this probability is sufficiently high and the ratio 51/100 strong enough to allow us to modify the analysis of [7] correspondingly, as we do in the rest of this section.

► **Lemma 10.** *For any interval t and node v , it is the case that $d_{t+1}(v|N_{1,t}^{\geq 22}(v)) \leq \frac{51}{100}d_t(v|N_{1,t}^{\geq 22}(v))$ with probability at least $1 - 300e^{-I/2000}$.*

The proof of this Lemma splits up the amount of $d_t(v)$ that is contributed by nodes with high effective degree into two parts. One part is contributed by good nodes, the other one by bad nodes. We mainly need to work to keep the contribution of bad nodes in check, as they may increase their desire values when they shouldn't (but they don't know). This can be done using Lemma 7 and a Chernoff Bound. However, problems arise when the set of high effective degree neighbors is small, smaller than $100I$ to be precise. In this case the probabilities that we obtain we Chernoff are not strong enough to modify the analysis of [7], i.e., not negatively exponential in I , and we treat this case of less than $100I$ nodes separately. The following notation helps us to formalizes some of the above:

► **Definition 11** ($N_{1,t}^{\geq 22}(v)$, $\bar{N}_{1,t}^{\geq 22}(v)$ and $d_t(v|S)$). Denote by $N_{1,t}^{\geq 22}(v) = \{u \in N_1(v) | d_t(u) \geq 22\}$ the neighbors of v with $d_t(u) \geq 22$. Denote by $\bar{N}_{1,t}^{\geq 22}(v) = \{u \in N_{1,t}^{\geq 22}(v) | u \text{ is not good}\}$ the set of nodes in $N_{1,t}^{\geq 22}(v)$ that are not good. Let $S \subseteq N_1(v)$ be a set of nodes, then we denote by $d_t(v|S) = \sum_{u \in S} p_t(u)$ the amount of $d_t(v)$ contributed by nodes in S .

Proof. (of Lemma 10). Based on how node v adjusts its value $p_t(v)$ when executing the algorithm (see Equation 1 that depends on the $b_t(v)/c_t(v)$ ratio), we can bound $d_{t+1}(v|N_{1,t}^{\geq 22}(v))$ to be smaller than $\frac{1}{2}d_t(v|N_{1,t}^{\geq 22}(v) \setminus \bar{N}_{1,t}^{\geq 22}(v)) + 2d_t(v|\bar{N}_{1,t}^{\geq 22}(v))$ and this is derived directly from the definition of the algorithm and in correspondence to how we modify the Algorithm of Ghaffari. Now we can write $\frac{1}{2}d_t(v|N_{1,t}^{\geq 22}(v) \setminus \bar{N}_{1,t}^{\geq 22}(v))$ as $\frac{1}{2}d_t(v|N_{1,t}^{\geq 22}(v)) - \frac{1}{2}d_t(\bar{N}_{1,t}^{\geq 22}(v))$ and when applied to the previous bound derive that $d_{t+1}(v|N_{1,t}^{\geq 22}(v)) \leq \frac{1}{2}d_t(v|N_{1,t}^{\geq 22}(v)) + \frac{3}{2}d_t(v|\bar{N}_{1,t}^{\geq 22}(v))$. To analyze the probability, that this is at most $\frac{51}{100}d_t(v|N_{1,t}^{\geq 22}(v))$, we distinguish two cases. In case 1, we consider $|N_{1,t}^{\geq 22}(v)| < 100I$ and in case 2 we consider $|N_{1,t}^{\geq 22}(v)| \geq 100I$.

Case 1, $|N_{1,t}^{\geq 22}(v)| < 100I$: The probability that no node in $N_{1,t}^{\geq 22}(v)$ is bad, i.e., $\bar{N}_{1,t}^{\geq 22}(v) = \emptyset$ is $(1 - 2e^{-I/100})^{|N_{1,t}^{\geq 22}(v)|}$, due to Lemma 7. By the assumption of case 1, that there are at most I neighbors u of v with $d_t(u) \geq 22$, this can be bounded by $\geq (1 - 2e^{-I/100})^{100I}$, which in turn is larger than $1 - 100I \cdot 2e^{-I/100} = 1 - 200e^{-I/100 + \log(I)} \geq 1 - 300e^{-I/100} \geq 1 - 300e^{-I/2000}$ due to the choice of I .

Case 2, $|N_{1,t}^{\geq 22}(v)| \geq 100I$: We bound the probability that at most a $1/150$ fraction of the nodes in $N_{1,t}^{\geq 22}(v)$ is bad, i.e., $|\bar{N}_{1,t}^{\geq 22}(v)| \leq |N_{1,t}^{\geq 22}(v)|/150$. To do so, we apply Chernoff Bound 2 of Lemma 20, see Appendix, for $X_i = \text{node } v_i \in \{v_1, \dots, v_{|N_{1,t}^{\geq 22}(v)|}\} = N_{1,t}^{\geq 22}(v)$ is good. Based on Lemma 7, we can conclude that $\mathbb{E}[X] \geq (1 - 2e^{-I/100}) \cdot |N_{1,t}^{\geq 22}(v)|$. Choosing $\delta = 1/300$, we can bound

$$\begin{aligned} Pr[X \leq (1 - 1/150) \cdot |N_{1,t}^{\geq 22}(v)|] &\leq Pr[X \leq (1 - 1/300)(1 - 2e^{-I/100}) \cdot |N_{1,t}^{\geq 22}(v)|] \\ &= Pr[X \leq (1 - \delta)\mathbb{E}[X]] = e^{-\frac{\delta^2}{2}\mathbb{E}[X]} = e^{-\frac{1}{180000}\mathbb{E}[X]} \leq e^{-\frac{1}{180000}(1 - 2e^{-I/100}) \cdot |N_{1,t}^{\geq 22}(v)|} \end{aligned}$$

Now we use the assumption $d_t(v|N_{1,t}^{\geq 22}(v))$ and the definition of I to derive that this is smaller than $e^{-\frac{1}{180000}(1 - 2e^{-I/100}) \cdot 100I} \leq e^{-I/2000} \leq 300e^{-I/2000}$. From this we conclude that $Pr[|\bar{N}_{1,t}^{\geq 22}(v)| \leq |N_{1,t}^{\geq 22}(v)|/150] \geq 1 - 300e^{-I/2000}$.

When combining both cases 1 and 2, we obtain that with probability at least $1 - 300e^{-I/2000}$, the value of $d_{t+1}(v|N_{1,t}^{\geq 22}(v))$ is smaller than

$$\frac{50}{100}d_t(v|N_{1,t}^{\geq 22}(v)) + \frac{3}{2} \frac{1}{150}d_t(v|N_{1,t}^{\geq 22}(v)) = \frac{51}{100}d_t(v|N_{1,t}^{\geq 22}(v)). \quad \blacktriangleleft$$

4.3 Proof of Theorem 3

Now we are prepared to follow the analysis of [7] and adapt it to our modifications of the algorithm. Using the notation used in the last two sections, Theorem 2 and 3 are derived from:

► **Theorem 12.** *For each node v , the probability that v makes a (locally $(1 - \varepsilon)$ -correct) decision within the first R intervals is at least $1 - \varepsilon$. Furthermore, this holds even if the outcome of the coin tosses outside $N_2^+(v)$ are determined adversarially.*

The rest of this Section is devoted to proving Theorem 12. First we define two kinds of *golden intervals* for a node v , by analogy with the definition of *golden rounds* in [7], then we show that it is likely that there are many golden intervals in case a node does not join M (Lemma 15). Then we argue that, if there are that many golden intervals, then it is likely that a node gets removed due to either joining M or having a neighbor that joins M (Lemma 16). To prove Lemma 15 we use Lemma 10; to prove Lemma 16 we use Lemma 7.

► **Definition 13.** A node v has *likely-low effective degree* if $b_t(v)/c_t(v) \leq \frac{1}{5}$, and has *likely-high effective degree* if $b_t(v)/c_t(v) > \frac{1}{5}$.

► **Definition 14 (Golden intervals).** Interval t is a golden interval of type 1, if $b_t(v)/c_t(v) \leq \frac{1}{5}$ and $p_t(v) = 1/2$. Interval t is a golden interval of type 2 if $b_t(v)/c_t(v) > \frac{1}{5}$ and at least $d_t(v)/11$ of $d_t(v)$ is contributed by neighbors u with $d_t(u) \leq 22$ (nodes of *low effective degree*).

These are called golden intervals because, as we will see, in the first type, v has a constant chance of joining M and in the second type, there is a constant chance that one of those neighbors of v with low effective degree joins M and thus v gets removed.

The following lemma and proof follow along the lines of a similar proof in [7], for Theorem 3.1, and is modified to our setting using the Lemmas proven so far.

► **Lemma 15.** *By the end of interval R , with probability at least $1 - 1500e^{-I/2000}$, either v has joined, or has a neighbor in M , or at least one of its golden interval counts reached $R/13$.*

Proof. (Note that we only prove claims about nodes joining M , not claims about nodes who join M being independent. An analysis of independence is performed as part of the proof of Theorem 12 at the end of this section). Let g_1 and g_2 respectively be the number of golden intervals of types 1 and 2 for v during this period. We assume, that by the end of interval R , node v is not removed and $g_1 \leq R/13$. Otherwise the statement of the Lemma would already be satisfied. Based on this assumption, we lower bound in the remaining part of this proof the number g_2 of golden intervals of type-2 while taking into account that any node u 's ratio $b_t(u)/c_t(u)$ might not always correctly represent whether $d_t(u) \geq 2/5$ or $d_t(u) \leq 22$.

Let h be the number of intervals where $b_t(v)/c_t(v) > \frac{1}{5}$. Notice that the changes in $p_t(v)$ are governed by the condition $b_t(v)/c_t(v) > \frac{1}{5}$ and intervals with $b_t(v)/c_t(v) > \frac{1}{5}$ are exactly the ones in which $p_t(v)$ decreases by a 2 factor. Since the number of 2 factor increases in $p_t(v)$ can be at most equal to the number of 2 factor decreases in it, we get that there are at least $R - 2h$ intervals in which $p_t(v) = 1/2$.

Now out of these $g_1 > R - 2h$ intervals, at most h of them can be when $b_t(v)/c_t(v) > \frac{1}{5}$. Hence, $g_1 \geq R - 3h$. As we have assumed $g_1 \leq R/13$, we get that $R - 3h \leq R/13$, and conclude that $h \geq R \cdot 4/13$.

Let us consider the changes in the effective-degree $d_t(v)$ of v over time. Note that $d_t(v)$ reflects all changes of each neighbor u 's value $p_t(u)$ based on whether $b_{t'}(v)/c_{t'}(v) > \frac{1}{5}$ in

previous intervals $t' < t$. This is independent of the actual value of $d_{t'}(v)$ at that time and thus $d_t(v)$ takes all previously made errors into account.

If $b_t(v)/c_t(v) > \frac{1}{5}$ and this is not a golden interval of type-2, then we know that at most $\frac{1}{11}d_t(v)$ of $d_t(v)$ is contributed by neighbors u with low effective degree $d_t(u) \leq 22$, such that the fraction of $d_t(v)$ contributed by those nodes doubles at most. On the other hand, at most all $d_t(v)$ of $d_t(v)$ is contributed by neighbors u with high effective degree $d_t(u) \geq 22$, i.e., $d_t(v|N_{1,t}^{\geq 22}(v)) \leq d_t(v)$. Due to Lemma 10 we know that $d_{t+1}(v|N_{1,t}^{\geq 22}(v)) \leq \frac{51}{100}d_t(v|N_{1,t}^{\geq 22}(v)) \leq \frac{51}{100}d_t(v)$ with probability at least $1 - 300e^{-I/2000}$. From this we conclude that with probability at least $1 - 300e^{-I/2000}$, $d_{t+1}(v) \leq 2\frac{1}{11}d_t(v) + d_{t+1}(v|N_{1,t}^{\geq 22}(v)) \leq \frac{765}{1100}d_t(v) < \frac{7}{10}d_t(v)$. There are g_2 golden intervals of type-2. We just showed that for all intervals with $b_t(v)/c_t(v) > \frac{1}{5}$, that are not among these g_2 golden intervals, the effective-degree $d_t(v)$ shrinks by at least a $7/10$ factor with probability at least $1 - 300e^{-I/2000}$ and this is independent of whether $b_t(v)/c_t(v) > \frac{1}{5}$ indicates the correct range of $d_t(v)$. Now let g be the number of intervals with $b_t(v)/c_t(v) > \frac{1}{5}$, that are not among these g_2 golden intervals. We show that with probability at least $1 - 1500e^{-I/2000}$ in at least $4g/5$ of these intervals the effective-degree $d_t(v)$ shrinks by at least a $7/10$ factor. Let k be the number of intervals in which the effective-degree $d_t(v)$ does not shrink by at least a $7/10$ factor, then $\mathbb{E}[k] \leq 300e^{-I/2000}g$. Using Markov's Inequality yields that $Pr(k \geq g/5) \leq \frac{\mathbb{E}[k]}{g/5} \leq \frac{300e^{-I/2000}g}{g/5} \leq 1500e^{-I/2000}$

In the g_2 golden intervals of type-2 and the $g/5$ intervals in which $b_t(v)/c_t(v) > \frac{1}{5}$ that are not golden intervals of type-2 and in which the effective-degree $d_t(v)$ does not shrink by at least a $7/10$ factor, the value of $d_t(v)$ increases by at most a 2 factor. Each of these $g_2 + g/5$ intervals cancels the effect of at most 2 shrinkage intervals, as $(7/10)^2 \times 2 < 1$. Thus, ignoring the total of at most $3(g_2 + g/26) \leq 3(g_2 + R/26)$ intervals lost due to type-2 golden intervals and their cancellation effects, every other interval with $b_t(v)/c_t(v) > \frac{1}{5}$ pushes the effective-degree of v down by a $2/3$ factor. This cannot (continue to) happen more than $\log_{3/2} \Delta$ times, as that would lead the effective degree to exit the $d_t(v) \geq 1/10$ region for any node. Hence, the number of intervals in which $b_t(v)/c_t(v) > \frac{1}{5}$ is at most $\log_{3/2} \Delta + 3(g_2 + g/5)$ with probability at least $1 - 1500e^{-I/2000}$. That is, $h \leq \log_{3/2} \Delta + 3(g_2 + g/5)$ with probability at least $1 - 1500e^{-I/2000}$. Since $h \geq R/13$, we get $g_2 > R/26$ with probability at least $1 - 1500e^{-I/2000}$ due to the definition of R . ◀

The following Lemma is adapted from a proof of Lemma 3.3 of [7] based on the new values and thresholds used in the modified algorithm and the corresponding definitions we introduced in this paper, and takes the error source and probabilistic behavior into account, which we generate due to not communicating effective degrees explicitly when (compared to what Ghaffari [7] does in the LOCAL model).

► **Lemma 16.** *In each type-1 (resp., type-2) golden interval, with probability at least $1/2000$, v joins M (resp., one of v 's neighbors joins M). If $R/13$ intervals are golden, then the probability that v has not decided whether it is in M during the first R intervals is at most $\varepsilon/2$. These statements hold even if the coin tosses in $N_2^+(v)$ are determined adversarially.*

Proof. In each type-1 golden interval, node v gets marked with probability $1/2$. In such an interval it is the case that $b_t(v)/c_t(v) \leq \frac{1}{5}$ and therefore $d_t(v) \leq 3$ with probability at least $1 - 2e^{-I/100}$ by Lemma 7. We conclude that the probability that no neighbor of v is marked

is

$$\begin{aligned} & \left(1 - 2e^{-I/100}\right) \cdot \prod_{u \in N(v)} (1 - p_t(u)) \geq \left(1 - 2e^{-I/100}\right) \cdot 4^{-\sum_{u \in N(v)} p_t(v)} \\ & = \left(1 - 2e^{-I/100}\right) \cdot 4^{-d_t(v)} \geq \left(1 - 2e^{-I/100}\right) \cdot 4^{-3} > 1/100. \end{aligned}$$

Hence, v joins M with probability at least $1/100 \cdot 1/2 > 1/2000$.

Now consider a type-2 golden interval. In such an interval it is $b_t(v)/c_t(v) > \frac{1}{5}$ and due to Lemma 7 we know that $d_t(v) \geq 1/10$ with probability at least $(1 - 2e^{-I/100})$. For the sake of analysis, suppose we walk over the set L of low effective degree neighbors of v one by one and expose their randomness until we reach a node that is marked. We will find a marked node with probability at least

$$\begin{aligned} & \left(1 - 2e^{-I/100}\right) \cdot \left(1 - \prod_{u \in L} (1 - p_u(t))\right) \geq \left(1 - 2e^{-I/100}\right) \cdot \left(1 - e^{-\sum_{u \in L} p_u(t)}\right) \\ & \geq \left(1 - 2e^{-I/100}\right) \cdot \left(1 - e^{-d_t(v)/11}\right) \geq \left(1 - 2e^{-I/100}\right) \cdot \left(1 - e^{-1/110}\right) \\ & > \left(1 - 2e^{-I/100}\right) \cdot 0.009 > 0.008, \end{aligned}$$

where the last bound is due to choice of $I \geq 2000 \cdot \log 1500$. When we reach the first low effective degree neighbor u that 1) satisfies the condition $d_t(u) \leq 22$ of v 's type-2 golden interval, and 2) that is marked, then the probability that no neighbor of u gets marked is at least

$$\prod_{w \in N(u)} (1 - p_t(w)) \geq 4^{-\sum_{w \in N(u)} p_t(w)} \geq 4^{-d_t(u)} \geq 1/64.$$

Hence, with probability at least $0.008/64 = 1/8000$, one of the neighbors of v joins M .

We now know that in each golden interval, v gets removed with probability at least $1/8000$, due to joining M or having a neighbor join M . Thus, using Lemma 15, we get that the probability that v does not get removed is at most

$$(1 - 1/8000)^{R/13} = (1 - 1/8000)^{8000(\log \Delta + \log(2/\varepsilon))} \leq \varepsilon/(2\Delta) \leq \varepsilon/2$$

due to the choice of $\gamma = 104000$ in the definition of R at the end of Section 3.2. \blacktriangleleft

Finally we are ready to prove Theorem 12.

Proof. (of Theorem 12). Due to Lemma 15, by the end of interval R , with probability at least $1 - 1500e^{-I/2000}$, either v has joined, or has a neighbor in M , or at least one of its golden interval counts reached $R/13$. In the latter case, we know due to Lemma 16, that with probability at least $1 - \varepsilon/2$, a node v terminates within R intervals and decides whether it is in M . Therefore, the probability that a node terminates within R intervals and decides whether it is in M is at least $1 - \varepsilon/2$, as we chose $I = 2000(\ln(1500) + \ln(2/\varepsilon))$.

Finally we need to prove local $(1 - \varepsilon)$ -correctness of the computation, which is handled in the second intervals of each emulated rounds of the LOCAL algorithm. The analysis above based on Lemma 15 argues that a node joins M with a certain probability or gets removed due to a neighbor joining M . This satisfies the maximality condition of an MIS - at any given point in time, no node that has already decided OUT could decide IN, i.e., be added to M , without violating independence among the nodes that have decided IN or OUT. Now we argue that independence is guaranteed locally with probability at least $1 - \varepsilon \frac{2^{-1000}}{\Delta}$, i.e., no

two neighboring nodes join M with probability at least $1 - \varepsilon \frac{2^{-1000}}{\Delta}$. As neighbors of a node v that joins the M are getting removed, we only analyze the probability that two neighbors join M in the same round (in which case the MIS condition is violated locally). This can only happen if: given a node v , a subset U of v 's neighbors marked themselves and all of them chose the same subset of time slots in the second interval to beep as v did, such that none of them recognizes that the other node marked itself as well. Furthermore, to ensure that there is a neighbor u of v that indeed joins M , node u must not have neighbors besides v that marked themselves, or all of u 's neighbors that marked themselves also chose a set of time slots to beep that is a subset of those slots that u (and thus as v) chose, and therefore u would not discover any of its neighbors that are marked - and thus u and v would both join M . The probability that this event happens is maximized if exactly one pair of neighbors v and u marked themselves and they have no further neighbors that marked themselves. Now the probability that both v and u choose the same subset of half of the time slots of the second interval that has I time slots in total is $1/\binom{I/2}{I/2}$. Due to the choice of $I = 2^{-2000(\log \Delta + \log(2/\varepsilon))}$, we conclude that this is smaller than $2^{-1000(\log \Delta + \log(2/\varepsilon))} = \varepsilon \frac{2^{-1000}}{\Delta}$. Combining this with the termination probability of the first paragraph of the proof yields that a node terminates and (locally) correctly decides whether it belongs to the MIS with probability $(1 - \varepsilon/2) \cdot \left(1 - \varepsilon \frac{2^{-1000}}{\Delta}\right) \geq 1 - \varepsilon$. ◀

Finally we are ready to prove Theorem 12 and Theorem 2.

Proof. (of Theorem 3 and Theorem 2). We use Theorem 12 and multiply the number of intervals R by the number I of time steps in each of the two intervals and add the final time slot that completes the emulation. We obtain that the total runtime is $R \cdot (2I + 1)O((\log \Delta + \log(1/\varepsilon)) \cdot \log(1/\varepsilon))$. Correctness follows from Theorem 12 as well. ◀

References

- 1 Yehuda Afek, Noga Alon, Ziv Bar-Joseph, Alejandro Cornejo, Bernhard Haeupler, and Fabian Kuhn. Beeping a Maximal Independent Set. *Distributed Computing*, 26(4):195–208, 2013.
- 2 Leonid Barenboim, Michael Elkin, Seth Pettie, and Johannes Schneider. The locality of distributed symmetry breaking. *Journal of the ACM (JACM)*, 63(3):20, 2016.
- 3 Kai-Min Chung, Seth Pettie, and Hsin-Hao Su. Distributed algorithms for the Lovász local lemma and graph coloring. In *Proceedings of the 2014 ACM symposium on Principles of distributed computing*, pages 134–143. ACM, 2014.
- 4 Alejandro Cornejo and Fabian Kuhn. Deploying wireless networks with beeps. In *International Symposium on Distributed Computing*, pages 148–162. Springer, 2010.
- 5 Alejandro Cornejo, Nancy Lynch, Saira Viqar, and Jennifer L Welch. Neighbor discovery in mobile ad hoc networks using an abstract mac layer. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 1460–1467. IEEE, 2009.
- 6 Alejandro Cornejo, Saira Viqar, and Jennifer L Welch. Reliable neighbor discovery for mobile ad hoc networks. *Ad Hoc Networks*, 12:259–277, 2014.
- 7 Mohsen Ghaffari. An improved distributed algorithm for Maximal Independent Set. In *Proceedings of the 2015 ACM-SIAM Symposium on Discrete Algorithms*, pages 270–277, 2016.
- 8 Mohsen Ghaffari, Erez Kantor, Nancy Lynch, and Calvin Newport. Multi-message broadcast with abstract MAC layers and unreliable links. In *Proceedings of the 33rd Annual ACM*

- Symposium on Principles of Distributed Computing, PODC 2014, Paris, France, July 15-18, 2014*, pages 56–65, extended version available at <http://arxiv.org/abs/1405.1671>, 2014.
- 9 Michel Goemans. Chernoff bounds, and some applications. In *MIT course 18.310 lecture notes*: <http://math.mit.edu/goemans/18310S15/chernoff-notes.pdf>.
 - 10 Magnus M Halldorsson, Stephan Holzer, and Nancy Lynch. A local broadcast layer for the sinr network model. In *Proceedings of the 34th Annual ACM Symposium on Principles of Distributed Computing, PODC 2014, Donostia-San Sebastián, Spain, July 21-23, 2015*, page tba, 2015.
 - 11 Magnus M Halldorsson, Stephan Holzer, and Nancy Lynch. A local broadcast layer for the sinr network model. *arXiv preprint arXiv:1505.04514*, 2015.
 - 12 Magnus M Halldorsson, Stephan Holzer, and Nancy Lynch. A local broadcast layer for the sinr network model. In *Proceedings of the 34th Annual ACM Symposium on Principles of Distributed Computing, PODC 2015, Donostia-San Sebastián, Spain, July 21-23, 2015*, pages 129–138, 2015.
 - 13 Stephan Holzer and Nancy Lynch. Brief announcement: Beeping a maximal independent set fast. In *Proceedings of the 30th International Symposium on Distributed Computing, DISC 2016, Austin, TX, USA, October 12-15, 2014*, page tba, 2016.
 - 14 Majid Khabbazi, Dariusz R. Kowalski, Fabian Kuhn, and Nancy A. Lynch. Decomposing broadcast algorithms using abstract MAC layers. *Ad Hoc Networks*, 12:219–242, 2014.
 - 15 Fabian Kuhn, Nancy A. Lynch, and Calvin C. Newport. The abstract MAC layer. *Distributed Computing*, 24(3-4):187–206, 2011.
 - 16 Michael Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing*, 15(4):1036–1053, 1986.
 - 17 Cameron Musco, Hsin-Hao Su, and Nancy Lynch. Ant-inspired density estimation via random walks. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing*, page tbd, 2016.
 - 18 Calvin C. Newport. Consensus with an abstract MAC layer. In Magnús M. Halldórsson and Shlomi Dolev, editors, *ACM Symposium on Principles of Distributed Computing, PODC '14, Paris, France, July 15-18, 2014*, pages 66–75, 2014.
 - 19 Alessandro Panconesi and Aravind Srinivasan. Improved distributed algorithms for coloring and network decomposition problems. In S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis, editors, *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 581–592, 1992.
 - 20 David Peleg. Distributed computing. *SIAM Monographs on discrete mathematics and applications*, 5, 2000.
 - 21 Johannes Schneider and Roger Wattenhofer. A log-star distributed maximal independent set algorithm for growth-bounded graphs. In Rida A. Bazzi and Boaz Patt-Shamir, editors, *Proceedings of the Twenty-Seventh Annual ACM Symposium on Principles of Distributed Computing, PODC 2008, Toronto, Canada, August 18-21, 2008*, pages 35–44, 2008.
 - 22 Alex Scott, Peter Jeavons, and Lei Xu. Feedback from nature: an optimal distributed algorithm for Maximal Independent Set selection. In *Proceedings of the 2013 ACM Symposium on Principles of Distributed Computing*, pages 147–156, 2013.

A Lower Bound on Translating the Schneider and Wattenhofer Algorithm for Bounded Growth Graphs

Schneider and Wattenhofer presented an algorithm running in time $O(\log^* N)$ on a restricted class of graphs, see Definition 17, that captures a variety of wireless network topologies. Here, $[1, N]$ indicates the ID space of the n nodes in the network. Theorem 18 claims that this algorithm cannot be translated to the BEEP model without losing its efficiency in the

following sense: direct translation of the algorithm fails. Major new techniques in the BEEP and LOCAL model would need to be developed and the algorithm be modified correspondingly.

► **Definition 17** (Bounded growth graphs). A graph $G = (V, E)$ is (*polynomial*) *growth-bounded* if there is a polynomial bounding function $f(r)$ such that for each node $v \in V$, the number of nodes in the r -neighborhood $N_{G,r}(v)$ of v in G that are in any independent set of G is at most $f(r)$ for all $r \geq 0$.

► **Theorem 18.** *A straightforward implementation of the MIS algorithm of [21] takes $\Omega(\Delta)$ time slots in the BEEP model.*

We provide some intuition behind Theorem 18 by reviewing how the algorithm of Schneider and Wattenhofer [21] works and argue that it cannot be translated into the BEEP model without losing a Δ factor, such that its efficiency can not be translated to the BEEP model without major modifications. To do so, we construct a network in the plane, in such a way, that the minimum value stored in the neighbors of a node v is different for most nodes v and argue that identifying this value for each node cannot be done fast in the BEEP model. This is a key ingredient of the Algorithm by Schneider and Wattenhofer, which therefore does not run fast in the BEEP model without major new insights.

Review of algorithm and result of [21]: In the LOCAL model, when applied on Bounded Growth Graphs, the algorithm runs in deterministic time $O(\log^* N)$, where nodes have IDs in the range of $[N]$, which is likely [poly n]. Recall that in the algorithm of [21], each node is in one of 5 states at any time and stores a value that changes over time. A node v changes its state and value based on the minimum value of v 's neighbors. The state change involves a bit by bit comparison between the node's current value and the minimum value around it. Therefore this minimum value needs to be known precisely and cannot be approximated.

Already Schneider and Wattenhofer [21] point out that the LOCAL model assumes perfect transmission of all messages in each round and therefore their algorithm is less appropriate for wireless networks. We show that this algorithm can not be emulated in the BEEP model using less than $\Omega(\Delta)$ slots.

Proof. (Proof sketch of Theorem 18) Consider a graph derived as follows: Δ nodes u_1, \dots, u_Δ located equidistantly on a line in ascending order with distance 1 to each. Let the transmission range be $\Delta/2$ and assume nodes are connected to each other when they are within transmission range. Assume each node u_i has value $2(i-1)+1$ or $2i$, which is determined by an adversary.

In this graph, the minimum value in the neighborhood of node $u_{\Delta/2+i}$ is $2(i-1)+1$ or $2i$. Node $u_{\Delta/2+i}$ can only determine this value by communicating with node u_i . This is true for all nodes $u_{\Delta/2+1}, \dots, u_\Delta$ and only one pair of nodes can communicate at the same time, such that at least $\Delta/2$ value need to be exchanged. This takes at least $\Omega(\Delta)$ slots. ◀

B Discussion and Implications for the Abstract MAC Layer

We describe a close connection between the BEEP model and abstract MAC layers (a.k.a. Local Broadcast Layers) that were introduced by Kuhn et al. [15] and recently got increased attention, e.g., in [5, 6, 8, 11, 12, 14, 15, 18]. We show how our MIS algorithm can be translated to this model. Abstract MAC layers were proposed as a model that provides an alternative approach to the various graph-based models with the goal of abstracting away low level issues with message contention. In this model one can express guarantees for local broadcast while hiding the complexities of managing message contention. These guarantees include message delivery latency bounds: an *acknowledgment bound* f_{ack} on the time for a

sender's message to be received by *all* neighbors, and a *progress bound* f_{prog} on the time for a receiver to receive *some* message when at least one neighbor is sending.

Of particular interest with respect to the BEEP model is the progress bound. More formally, the progress bound guarantee is as follows: fix some $(u, v) \in E$ and interval of length f_{prog} throughout which node u is broadcasting a message m ; during this interval node v must receive some message (though not necessarily m , but a message that some location is currently working on). We consider an enhanced definition of the abstract MAC layer [15], which provides nodes an abort interface that allows them to abort a broadcast in progress. This is useful, as we can stop a broadcast after time f_{prog} and know that each node that should receive a message has indeed received one.

We now provide a high-level idea of how to translate our result from the BEEP model to the abstract MAC layer. We emulate each slot of the BEEP model using f_{prog} time in the abstract MAC layer. For each round t and for any node u that wants to send a beep in slot t , we inject $bcst_v$ ("beep") into the MAC layer interface at time $(t - 1) \cdot f_{prog} + 1$ and an $abort_v$ command at time $t \cdot f_{prog}$ to stop the broadcast. Based on the definition of progress, each node v that has a neighbor that sends a message in slot t of the BEEP algorithm, received a message "beep" in the abstract MAC layer algorithm at time $t f_{prog}$ at the latest. Using Theorems 2 and 3 we conclude:

► **Theorem 19.** *Given an abstract MAC layer that supports aborts and executes the algorithm described above, when a node v terminates, it has made its (locally $(1 - \varepsilon)$ -correct) decision whether it is in the MIS or not, and the probability that node v terminates within the first $O((\log \Delta + \log(1/\varepsilon)) \cdot \log(1/\varepsilon) \cdot f_{prog})$ slots is at least $1 - \varepsilon$. This holds even if the outcome of the coin tosses outside $N_2^+(v)$ are determined adversarially.*

Thus the cost of our MIS algorithm over abstract MAC depends on the progress bound only, not the acknowledgment bound. Given that, for radio networks at least, acknowledgment bounds are much bigger than progress bounds, this produces an efficient MIS algorithm for the radio network model.

C Basic Chernoff Bounds Used in the Proofs

► **Lemma 20** (Chernoff Bounds [9]). *Let $X = \sum_{i=1}^n X_i$, where $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1 - p_i$, and all X_i are independent. Let $\mu = \mathbb{E}[X] = \sum_{i=1}^n p_i$. Then*

1. $Pr\left(X \geq (1 + \delta)\mathbb{E}[X]\right) \leq e^{-\frac{\delta^2}{2+\delta}\mathbb{E}[X]}$, for all $\delta > 0$, and
2. $Pr\left(X \leq (1 - \delta)\mathbb{E}[X]\right) \leq e^{-\frac{\delta^2}{2}\mathbb{E}[X]}$, for all $0 < \delta < 1$.