Lower Bounds on Broadcasting Time of de Bruijn Networks

Stéphane Perennes.

Laboratoire I3S U.R.A 1376
930 Route des Colles- B.P 145 F - 06903 Sophia Antipolis France

Abstract. Under the *telephone model*, the broadcasting time of most of the logarithmic networks (where the degree is fixed and the diameter logarithmic in the number of nodes) is not known, as the time of the known protocols is different from the known lower bounds. That is the case for de Bruijn. In this paper we present a technique enabling to derive better lower bounds on the broadcasting time of various networks, the technique is applied in the case of the de Bruijn graphs.

1 Introduction

The problem of dissemination of information has been investigated for most of the classical networks. We refer the reader to one of the recent surveys [8] [11] or the book [7]. Our aim is to describe a technique enabling to derive lower bounds on the broadcasting time of various networks. In this short version we apply the technique to obtain lower bounds on the broadcasting time of de Bruijn networks.

The *maximum degree* Δ of a graph is the maximum degree of the vertices. In the case of a digraph, the number of arcs going out (resp. entering) a vertex is called the out-degree (resp. in-degree) of this vertex. The out(in)-degree $d^+(d^-)$ of a digraph is the maximum out(in)-degree of the vertices.

Definition 1. The parameter d of a graph (resp. digraph) is its maximum degree minus one ($\Delta - 1$) (resp. its maximum out-degree d^+)

De Bruijn networks The **de Bruijn** digraph (resp. graph), denoted by $B(d, D)$ (resp. $UB(d, D)$), has d^D vertices, diameter D and in-degree or out-degree d (resp. degree $2d$). The vertices correspond to the words of length D over an alphabet of d symbols. The arcs (or edges) correspond to the shift operations: Given a word $x = x_1 \cdots x_D$ on an alphabet $\mathcal{A}$ of d letters, where $x_i \in \mathcal{A}$, $i = 1, 2, \cdots, D$, and given $\lambda \in \mathcal{A}$, the operations: $x_1 \cdots x_D \longrightarrow x_2 \cdots x_D\lambda$ and $x_1 \cdots x_D \longrightarrow \lambda x_1 \cdots x_{D-1}$ are called respectively left-shift and right-shift. In the de Bruijn digraph $B(d, D)$, the successors are obtained by left-shift operations, whereas in the de Bruijn graph $UB(d, D)$, the neighbors are obtained by either left or right-shift operations. The de Bruijn graph is obtained by removing the directions of the arcs. Note that, since the de Bruijn digraph contains loops and symmetric arcs, this leads to a multi-graph. Thus some authors define the undirected de Bruijn graph as obtained by removing loops and multiple edges. Here we still keep them for homogeneity purpose, and in order to have a regular graph. The symmetric de Bruijn digraph $B^*(d, D)$ is the digraph obtained by replacing in $UB(d, D)$ each edge by a pair of opposite arcs.

Broadcasting One of the main problems of information dissemination investigated in the current literature and used in parallel and distributed computing is **broadcasting**. Broadcasting (also called One to All) refers to the process of message dissemination in an interconnection network whereby a message, originated at one node x, is transmitted to all the nodes of the network. Broadcasting is accomplished by placing a series of calls over the communication lines of the network. Therefore the communication protocol is a sequence of rounds, each one being performed by a set of local calls.

Here we restrict ourselves to what is called the "telephone model" where during one round each vertex can communicate with at most one of its neighbors (out-neighbors in case of digraphs). The broadcasting time $b(x, G)$ is the minimum number of rounds to achieve in G a broadcasting protocol originated at x. The broadcast time $b(G)$ of a graph G, is the maximum of $b(x, G)$ on all vertices x. (In the notation of [8] it would be $b_{F_1}(G)$, F_1 standing for "full duplex one port model"). **In all the paper all the logarithms used are in base** 2.

2 Previous Results

Previous Upper bounds In the case of de Bruijn networks upper bounds have been given in many papers for example [4] [10] and [13] and have still be recently improved in [5, 6] . In particular this leads to:

$$b(B(2, D)) \leq 1.5(D+1) \quad b(B(3, D)) \leq 2(D+1)$$

Previous lower bounds Non trivial lower bounds have been derived using available results for graphs or digraphs of bounded degree [3, 12]. Here we recall the basic counting argument of these articles:

Let G be a graph (resp. digraph) of parameter d. The set of informed nodes cannot double at each round, as a node informed during some round can, in the optimal case, only forward the information during the d next rounds [1]. As example, when $d = 2$, for $t \geq 4$, the number $S(t)$ of nodes receiving an information during the round t is such that $S(t) \leq S(t-1) + S(t-2)$. We can then compute an upper bound $N(t)$ of $S(t)$. In fact $N(t)$ is the solution of $N(t) = N(t-1)+N(t-2)$ for $t \geq 4$ with $N(1) = S(1)$ and $N(2) = S(2)$, $N(3) = S(3)$. The difference between graphs and digraphs with the same parameter d appears in the computation of the initial values of $N(t)$. As example for graphs

[1] This is not true for the originator of the broadcasting in a graph, as such a node can possibly inform new vertices during $d + 1$ rounds

(digraphs) of parameter 2 we have : the values $N(2) = 2, N(3) = 4$ (resp.. . $N(2) = 2, N(3) = 3$). The lower bound is then deduced by:

$$\sum_{0 \le t \le b(G)} N(t) \ge \sum_{0 \le t \le b(G)} S(t) \ge |V|$$

The analysis of the recurrences leads to the bounds of [3, 12]:

$$d = 2, b(G) \ge 1.4404 \log(|V|)(1 + o(1)) \quad d = 3 b(G) \ge 1.1374 \log(|V|)(1 + o(1))$$

Note that $1.4404 = \frac{1}{\log(\tau)}$ where $\tau = \frac{1+\sqrt{5}}{2}$ is the **golden ratio**, that is the greatest positive root of polynomial $x^2 - x - 1$. This property can be generalized as in[3]: $b(G) \ge \frac{1}{\log(\tau_d)} \log(|V|)(1 + o(1))$, where τ_d is the largest positive root of $x^d - x^{d-1} - \ldots - 1 = 0$.

In the case of binary de Bruijn and Butterfly graphs, the lower bounds have been improved later in [11], the authors proved that $b(UB(2, D) \ge 1.31171D$, and that $b(\mathcal{WBF}(2, D)) \ge 1.7456D$.

In that paper we will introduce a technique enabling us to prove that broadcasting in the undirected de Bruijn graph is not significantly faster than in the digraph. We will prove that $b(B^*(2, D)) \ge 1.4404D, b(B^*(3, D)) \ge 1.8028D$. For a generalization of the technique to the case of *iterated line digraph* (Kautz and Butterflies networks) we refer to the extended version.

3 Counting informed vertices in a digraph (graph) of parameter d.

For this we consider a broadcast protocol in a graph (digraph) of parameter d as an infinite directed tree of out-degree d.

Definition 2. Given a digraph, the **broadcast tree** associated to the protocol is a directed infinite tree of out-degree d whose root is the originator. The arcs outgoing from a given vertex are labeled $0, 1, 2, \ldots, d - 1$ according to the following rule *the arc (x, y) is labeled i, if y is the $i+1-th$ vertex that x calls after having itself being informed.*

So the labeling reflects the strategy used locally by each node to forward the information in the tree, each label corresponding to the **delay** introduced by using the arc. In the case of a finite graph of out-degree d, the broadcast tree is finite and in fact obtained as a quotient graph of the infinite broadcast tree.

Definition 3. – Given a graph, the **broadcast tree** is defined in the same way, except that the out-degree of the originator is $d+1$ with delays $0, 1, 2, \ldots, d$.
– If w is a path from the root to x we will say that x is informed along w. If w if of length l we will say that x is at **depth** l

- The **delay** of a path is the sum of the delays on each arc. The **cost** of a path is the sum of its length and of the delays of its arcs, it corresponds to the time at which the end vertex of the path is informed along the path.
- We will denote $I(l,t)$ the ideal [2] number of nodes informed at time exactly t on a path of length l.
- In a broadcast protocol let $S(t)$ denotes the number of nodes informed at time exactly t, note that $S(t) \leq \sum_l I(l,t)$.

Proposition 4. *In a graph (digraph) of parameter d .*

$$I(l,t) = \binom{l}{t-l}_d$$

Where the multinomial coefficient $\binom{l}{i}_d$ is defined by the generating series identities:

$$(1+z+\ldots z^{d-1})^l = \sum_{i\geq 0} \binom{l}{i}_d z^i \text{ For digraphs}$$

$$(1+z+\ldots+z^{d-1}+z^d)(1+z+\ldots z^{d-1})^{l-1} = \sum_{i\geq 0} \binom{l}{i}_d z^i \text{ For graphs}$$

Proof. As our bound is valid for any graph (digraph), we study a broadcasting protocol in a directed infinite tree of out-degree d whose root is the originator. Under this assumption, the number of paths of length l and cost t (or equivalently of delay $t-l$) is exactly the number of words of length l and weight $t-l$ on the alphabet $\{0,1\ldots d-1\}$ that is $\binom{l}{t-l}_d$. In the case of a (di)graph of parameter d the value of $I(l,t)$ is always smaller or equal than this value. Indeed, a vertex can appear many times in the broadcast tree, and so might be counted as informed more than once. This only means that our counting does not take into account possible additional properties of the graph, and that the bound can be refined for some (di)graph. Note that in the case of graphs as the originator (.i.e the root of the tree) have degree $d+1$ instead of d with delays $0,1,\ldots d$, the generating serie differs a bit.

As one can check, the difference between graphs and digraph of parameter d is not significant when deriving asymptotical evaluations. Consequently we will just forget it, considering for graphs the serie $(1+z+\ldots z^{d-1})^l$ instead of $(1+z+\ldots+z^{d-1}+z^d)(1+z+\ldots z^{d-1})^{l-1}$.

Case $d=2$ In the case $d=2$, we can do a detailed analysis which gives not only the former bound $\frac{1}{\log(\tau)}\log(|V|)$ but much more information on how should be a protocol reaching the lower bound if any. As $d=2$, $I(l,t)$ the estimation of $I(l,t)$ is $\binom{l}{t-l}_2$, the number of combinaisons of $t-l$ elements of a set of cardinal l: $\binom{l}{t-l}$.

For a fixed time t, we shall estimate what is the length $l_0(t)$ which maximize $I(l,t)$ at time t. Indeed, we will compute at what depth are mainly located the

[2] That is a generic upper bound valid for a specific class of graph

nodes counted as informed at that time. The set of nodes informed at time t at depth $l_0(t)$ will be called the *main level* at time t. Stirling approximation or usual techniques of estimation for coefficients of generating series (see [9]) allow us to claim that: $\log(\binom{n}{xn}) \sim n\phi(x)$ where ϕ is the classical entropy function $\phi(x)) = -\log((1-x)^{1-x}x^x)$. Then setting $l = \alpha t$, we have to maximize for a fixed t: $\log(I(\alpha t, (1-\alpha)t) = (1+o(1))\alpha t\phi(\frac{1-\alpha}{\alpha})$ relatively to the variable α. Computations leads to: $l_0(t) = \alpha_0 t = \frac{1+\frac{1}{\sqrt{5}}}{2}t$. Thus the main level at time t is at depth $\alpha_0 t$ and its size is $\binom{\alpha_0 t}{(1-\alpha_0)t}$. Note that $\log(\binom{l_0(t)}{t-l_0(t)}_2) \sim t\log(\tau)$ [3] this means that the cardinality of the main level is potentially growing exponentially with the time. Due to the behavior of binomial coefficients one can shows that $\log(S(t))$ is bounded above by a quantity equivalent to $\log(I(l_0(t), t)$, that is $\log(S(t)) \leq (1+o(1))t\log(\tau)$. Furthermore due to exponential growing of the value bounding $S(t)$: $2^{(1+o(1))t\log(\tau)}$, one knows that $\log(\sum_{t\leq T} S(t)) \leq (1+o(1))T\alpha_0\phi(\frac{1-\alpha_0}{\alpha_0}) = T\log(\tau)(1+o(1))$. And the number of vertices informed at time t is bounded by $2^{t\log(\tau)(1+o(1))}$. This is a more precise way to derive the result of [3], as to achieve broadcasting at time T we must have $2^{T\log(\tau)(1+o(1))} \geq |V|$. Our refined calculus is useful, as it enables us to precise the depth at which the nodes informed at time t are mainly located. In fact one can easily check that at time t it is possible to consider that *all the vertices counted as informed are at depth* $\alpha_0 t$. As example, if a broadcast protocol finishes in a graph G of parameter 2 in $\frac{\log(|V|)}{\log(\tau)}$ rounds, then most of the informed vertices are at depth: $l = \alpha_0\frac{\log(|V|)}{\log(\tau)} = \gamma\log(|V|)$. Computation shows that $\gamma = \frac{\alpha_0}{\log(\tau)} \sim 1.0423 > 1$. So we can assert that if a graph (digraph) of parameter 2 has a nearly optimal broadcast protocol this one informs most of the nodes along paths of length around $\gamma\log(|V|)$. For example, in the case of degree 3 graphs, we need a graph with mean eccentricity less than $\gamma\log(|V|)$. As γ is very close to 1, such a graph is nearly a Moore graph (just note that the graph has $|V|$ vertices,[4] and most of these are supposed to lie at distance at most $\gamma\log(|V|)$). Our condition is clearly more restrictive than the straight one stating that the diameter of G is not greater than $1.4404\log(|V|)$, and gives also hints to tackle the following question:

Conjecture 1 $b(B(2,D)) \sim D\frac{1}{\log(\tau)} \sim 1.4404D$

Note that here $\log(|V|)$ is D, so a protocol proving this assumption should use paths of length γD instead of shortest paths of length of order D used in existing protocols.

[3] τ has been defined in section 2 as $\frac{1+\sqrt{5}}{2}$

[4] At the moment, no explicit procedure is known to construct such a graph, existence is proved by probabilistic methods

4 Counting in the de Bruijn graph $B^*(d, D)$

Any vertex in the digraph has $2d$ out going arcs. Among this $2d$ arcs, d correspond to a left-shift and lie in the de Bruijn digraph. These arcs will be called L (left) ones. The others corresponding to a right-shift are R (right) arcs. For purpose of simplicity we associate to each protocol P a derived protocol P_1 faster than P.

The derived protocol The protocol P is completely defined by a local labeling of the outgoing arcs. Suppose that at each vertex the outgoing arcs are sorted according to their labels, then the derived protocol P_1 is defined by the new labeling given below:

R arcs	$R_{a_0} < R_{a_1} \dots < R_{a_{d-1}}$
L arcs	$L_{a_0} < L_{a_1} \dots < L_{a_{d-1}}$

L arcs	L_{a_0}	L_{a_1}	$\dots$	$L_{a_{d-1}}$
labels	0	1	$\dots$	$d-1$
R arcs	R_{a_0}	R_{a_1}	$\dots$	$R_{a_{d-1}}$
labels	0	1	$\dots$	$d-1$

Counting in the derived protocol $\overline{I(l,t)}$ and $\overline{S(t)}$ will denote the values of the estimation in the derived protocol P_1. $I(l,t)$ and $S(t)$ still refer to a directed protocol. The counting problem is now simple; we consider once again the infinite tree associated to P_1; for which at each vertex there are $2d$ outgoing arcs. Using the labeling of P_1 they are:

$$\begin{cases} (R,0),\ (R,1),\ \dots,\ (R,d-1) \\ (L,0),\ (L,1),\ \dots,\ (L,d-1) \end{cases}$$

So, a path of length l and cost t, in the infinite broadcast tree is defined by the pair (a, w) where: $a = (a_1 \dots a_l)$ with $0 \le a_i \le d-1$ and $\sum a_i = t - l$, and w is a word of length l over the alphabet $\{L, R\}$. If the $i-th$ letter of the word w is L (resp. R) that means that the $i-th$ arc used in the path is (L, a_i) (resp. (R, a_i)). **The word w will be called the (L, R) word of the path**. A first bound on $\overline{I(l,t)}$ can be obtained by multiplying $\binom{l}{t-l}_d$ by the number of words of length l over the alphabet $\{L, R\}$. So the number of vertices informed at time t along a path of length l is at first glance: $\overline{I(l,t)} \le I(l,t)2^l$. Not surprisingly, this estimation is useless as it does not take into account properties of de Bruijn graphs. Hopefully we will prove that most of the paths are always redundant (i.e useless for the broadcasting protocol).

Useful (L, R) words

Definition 5.

- A path in the broadcast tree is redundant if, there exists either another path of smaller cost, or another path of equal cost with smaller length which leads to the same vertex. In other words, if this path informs a vertex which can be informed earlier by another path, or at the same time along a shorter path.
- A word w of $\{L, R\}^l$ is useless if each walk in the broadcast tree having w as $\{L, R\}$ word is redundant. A word is useful if it is not useless.

Lemma 6. *In the undirected de Bruijn graphs $UB(2,D)$ and $UB(3,D)$, useful (L,R) words have no sub-word of kind $L^iR^iL^i$ or $R^iL^iR^i$.*

Proof. We consider a path in the tree, leading to a node z, with as (L,R) word $w_1L^iR^iL^iw_2$. Let x (resp. y) be the vertex informed along the path corresponding to the word w_1 (resp. $w_1L^iR^iL^i$). Suppose that x is informed at time t. Then according to our protocol y will not be informed, thanks to the path, before time $t+3i$. Due to de Bruijn iterated line graph structure, there is a path with a (L,R) word L^i from x to y. So, once again, according to our protocol, y will be also informed at time at most $t+di$ (as $d \leq 3$, $t+di \leq t+3i$) along a path with (L,R) word w_1L^i . As z is informed from y, we can claim that any z informed along a path with (L,R) word $w_1L^iR^iL^iw_2$ can be informed sooner (or at the same time but along a shorter path) along a path with (L,R) word $w_1L^iw_2$. Hence $w_1L^iR^iL^iw_2$ is useless.

Lemma 7. *A partition of an integer n is an increasing sequence of integers $i_1 \ldots i_p$ such that $\sum_{k=1}^{k=p} i_k = n$. The number $P(n)$ of the partitions of n has the following asymptotic behavior: $P(n) \sim \frac{1}{4\sqrt{n}} exp(\pi\sqrt{\frac{2n}{3}})$*

Proof. A proof of this lemma is given for example in the book [1]

Proposition 8. *Let $U(l)$ denotes the cardinality of the set of useful (L,R) words of length l, then for $UB(2,D)$ and $UB(3,D)$; $\log(U(l)) = \Theta(\sqrt{l})$*

Proof. To a (L,R) word we associate a partition as follows: if the word is of the form $L^{i_1}R^{i_2}\ldots L^{i_p}$ we associate to it the sequence $i_1, i_2, \ldots, i_p$. As w is useful, the sequence $\{i_k\}$ is a *bitonic* one. (i.e. the sequence is increasing then decreasing). To check that, just notice that if $i_k+1 \leq i_k$ then $i_{k+2} < i_{k+1}$ (if not, one of the pattern $L^{i_{k+1}}R^{i_{k+1}}L^{i_{k+1}}$ $R^{i_{k+1}}L^{i_{k+1}}R^{i_{k+1}}$ would appear contradicting lemma 6). Then to each useful word we can associate an increasing sequence and a decreasing one with sums respectively n_1 and n_2 with $n_1+n_2=l$. So a crude estimation of $U(l)$ can be $U(l) \leq \sum_{n_1+n_2=l} P(n_1)P(n_2) \leq \sum_{0\leq i\leq l} P(i)P(n-i) \leq lP(\frac{l}{2})^2$. Lemma 7 allows us to conclude.

Proposition 9. $b(UB(2,D)) \geq \frac{1}{\log(\tau)}D + o(D), b(UB(3,D)) \geq \frac{\log(3)}{\log(\tau_3)}D + o(D)$

Proof. As $\overline{I(l,t)} \leq I(l,t)U(l)$, and as $\log(U(l)) = o(l)$, we have in fact $\log(\overline{I(l,t)}) = \log(I(l,t)) + o(l)$. As $l = \Theta(t)$ we can claim that $\log(\overline{S(t)}) = \log(\sum \overline{I(l,t)}) \leq \log(\sum I(l,t)U(l)) \sim \log(S(t)) + o(t)$. So, we get the same estimation than for directed graphs of out-degree d: $\log(\sum_{t\leq T} \overline{S(t)}) \leq \log(\tau_d)T + o(T)$. So, undirected de Bruijn graphs of small degree behave like the digraphs.

5 Conclusion

In this article we have used a general technique and the structural properties of the de Bruijn networks to derive new lower bounds on the broadcasting time.

The results obtained show that the protocol is not likely to be more efficient than in the de digraphs. We conjecture that bounds obtained give the right order for the broadcasting time in de Bruijn networks. So improvements have to been done in designing faster protocols. However our results show that the diffusion will not be done on shortest paths which implies that these protocols might be complicated or difficult to imagine. Extension of the technique to iterated line digraphs (butterflies and Kautz networks) and to n-dimensional grids will be addressed in the full paper.

Acknowledgment: This work has been supported by the European HCM project MAP, the french action RUMEUR and the Canadian/French project PICS.

I thanks very much J-C Bermond, B. Martin and O. Delmas for their advice.

References

1. George E. Andrews. *The Theory of partitions, Encyclopedia of Mathematics and its applications. Volume 2.* Addison-Wesley Publishing Company, 1976. Section: Number Theory.
2. C. Berge. *Graphes.* North Holland, 1981.
3. J-C. Bermond, P. Hell, A.L. Liestman, and J.G. Peters. Broadcasting in bounded degree graphs. *SIAM journal of Discrete Mathematics*, 5:10–24, 1992.
4. J-C. Bermond and C. Peyrat. Broadcasting in de Bruijn networks. Proceedings of the 19-th Southeastern conference on combinatorics, Graph theory and Computing, Congressus Numerantium, 66, 267-282, 1988.
5. J-C. Bermond, A. Marchetti, and X. Munoz. Improved broadcasting protocols for line digraphs. submitted to *EUROPAR 96*
6. J-C. Bermond and S. Perennes. Efficient broadcasting protocols on the de Bruijn and similar networks. In *Proceedings of SIRROCO'95*, 1995.
7. J. de Rumeur. *Communications dans les Réseaux de Processeurs.* Masson, 1994. To be translated.
8. P. Fraigniaud and E. Lazard. Methods and problems of communication in usual networks. *Discrete Applied Mathematics*, 53, Special issue on "Broadcasting and gossiping":79–133, 1995.
9. D. Gardy. The asymptotic behaviour of coefficients of large power of functions. In *Serie formelles et combinatoire algébrique*, pages 237–248, 1992.
10. M-C. Heydemann, J. Opatrny, and D. Sotteau. Broadcasting and spanning trees in de Bruijn and Kautz networks. *Discrete Applied Math.*, 27-28:297–317, 1992.
11. R. Klasing, B. Monien, R. Peine, and E. Stohr. Broadcasting and gossiping in the Butterfly and de Bruijn networks. In Spinger Verlag, editor, *Proc. STACS'92*, volume Lectures Notes in Computer Science 577, pages 351–362, 1992.
12. A.L. Liestman and J.G. Peters. Broadcast networks of bounded degree. *SIAM Journal of Disc. Math.*, 1(4), 1988.
13. S. Perennes. Broadcasting and gossiping on de Bruijn, shuffle exchange and similar networks. Technical report 93-53, I3S, 1993.