Tight Bounds on the Ratio of Network Diameter to Average Internode Distance

Behrooz Parhami
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106-9560, USA
parhami@ece.ucsb.edu

***Abstract*—The average internode distance for a network is fairly difficult to derive. There is often no closed-form formula for this parameter, leading to the need for simulation-based derivation methods. Network diameter, by contrast, is relatively easier to determine and, for many networks of common interest, we have closed-form formulas for it. The bounds established in this paper show that the two parameters are usually not totally independent and that, from a practical standpoint, network diameter can be used in lieu of average internode distance for the evaluation of message-routing algorithms and assessment or comparison of communication performance, particularly for symmetric networks that are prevalent in many parallel systems.**

Keywords—Communication; Graph; Interconnection network; Parallel processing; Routing algorithm; Symmetric network

I. INTRODUCTION

A great many interconnection architectures have been proposed for linking bare-bone processing elements or full-fledged, independently-operating computers in parallel and distributed systems [4], [7], [11], [15]. Comparing such networks with respect to their suitability for a particular application domain is often difficult [8], given the multitude of static attributes (diameter, average distance, bisection width, VLSI layout area) and dynamic properties (routing algorithms, deadlock prevention, traffic balance, fault tolerance) that must be taken into consideration. Thus, introduction of new interconnection networks, while enriching the repertoire of parallel computer designers, also adds to the selection difficulty.

In this paper, we focus on two of the static attributes of a network, its diameter D and average internode distance Δ, and show that for practical networks that tend to be regular (node-symmetric), the two parameters are not totally independent of each other. We prove formally that for a regular, degree-d network, the two parameters D and Δ are within a factor of at most 2 from each other. Thus, worst-case communication latency, dictated by network diameter, and expected latency, which is a function of average distance, are also interdependent. Furthermore, we show that even for certain non-regular networks that are widely used in practice, namely meshes and binary trees, the average distance and diameter remain related, although not always satisfying the factor-of-2 relationship.

We take the average internode distance as the expected distance from a randomly chosen node to other nodes in the network, including the node itself. This inclusion of null paths, that is counting paths from each node to itself in computing the average distance, leads to cleaner results in most cases and has negligible effect on our conclusions.

A final note about terminology: We are interested in lower- and upper-bound on a parameter x, denoted by $\mathrm{lb}(x)$ and $\mathrm{ub}(x)$, when the actual value of x is unknown and/or difficult to derive. We have $x \le \mathrm{ub}(x)$ and $\mathrm{lb}(x) \le x$. The upper bound $x \le \mathrm{ub}(x)$ or lower bound $\mathrm{lb}(x) \le x$ is said to be tight when equality is possible in at least one instance of the cases under consideration.

II. AVERAGE DISTANCE IN MESHES AND BINARY TREES

Exact formulas for the average internode distance Δ have not been published for many useful networks, including the two most-widely used ones: q-dimensional meshes and complete binary trees. In this section, we aim to correct this deficiency [12] and will later use the results to draw general conclusions about the relationship between D and Δ.

The shortest-path length in a q-dimensional mesh can be found by adding the distances of the destination node from the source node along each of the q dimensions. Thus, all that is required for finding an exact formula for Δ is to have an exact formula for the average distance in a p-node linear array, also known as a p-path.

$$\Delta_{p\text{-path}} = (1/p^2)\Sigma_{0\le j\le p-1}[\Sigma_{0\le i\le j}(j-i) + \Sigma_{j\le i\le p-1}(i-j)] \tag{1}$$

Recall that we include the 0-length path from each node to itself in calculating the average, hence the normalizing term $1/p^2$ in the expression above, in lieu of $1/[p(p-1)]$. The two sums in the square brackets are sums of distances from node j to all to nodes to its left and right, respectively. Using the formulas for the sums of consecutive integers and their second powers, we obtain for $p \ge 3$:

$$\Delta_{p\text{-path}} = (1/p^2)\Sigma_{0\le j\le p-1}[j(j+1) - j(j+1)/2 + (p-j)(p-1+j)/2 - j(p-j)] = (1/3)(p-1/p) \tag{2}$$

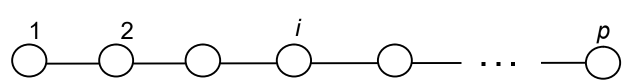

Fig. 1. Linear array with p nodes (a p-path).

The average internode distance in a q-dimensional $n_1 \times n_2 \times \ldots \times n_q$ mesh is thus:

$$\Delta_{q\text{D-mesh}} = (1/3)[\Sigma_{1\le i\le q}(n_i - 1/n_i)] \quad (3)$$

When the dimensions n_i are large, the average internode distance in (3) is roughly one-third of the diameter in (4):

$$D_{q\text{D-mesh}} = \Sigma_{1\le i\le q} n_i - q \quad (4)$$

The average internode distance and diameter for a p-ring are similarly derived in (5) and (6), again assuming $p \ge 3$, leading to the average internode distance and diameter for a q-dimensional $n_1 \times n_2 \times \ldots \times n_q$ torus in (7) and (8):

$$\Delta_{p\text{-ring}} = (1/4)[p - (p \bmod 2)/p] \quad (5)$$
$$D_{p\text{-ring}} = (1/2)[p - (p \bmod 2)/p] \quad (6)$$
$$\Delta_{q\text{D-torus}} = (1/4)\Sigma_{1\le i\le q}[n_i - (n_i \bmod 2)/n_i] \quad (7)$$
$$D_{q\text{D-torus}} = (1/2)\Sigma_{1\le i\le q}[n_i - (n_i \bmod 2)/n_i] \quad (8)$$

Deriving an exact formula for the average distance of a binary tree is much harder. We characterize an n-node complete binary tree with the parameter $m = n + 1 = 2^l$; we say that the tree has l levels, numbered from 1, for the root, up to l, for the leaves. So, T_m, $m = 2^l \ge 2$, refers to an $(m - 1)$-node complete binary tree.

First, some observations. The diameter of T_m is:

$$D_{\text{binary-tree}} = 2l - 2 = 2 \log_2 m - 2 \quad (9)$$

We begin by calculating the sum $\sigma(T_m)$ of the lengths of the paths from the root r to every node in T_m.

$$\sigma(T_m) = 1\times 2^1 + 2\times 2^2 + \ldots + (l - 1)\times 2^{l-1} = (l - 2)2^l + 2 \quad (10)$$
$$= m \log_2 m - 2m + 2$$

Now T_m can be viewed as consisting of three parts: The root node r, the left subtree L, which is a $T_{m/2}$ with root node r_L, and the right subtree R, also a $T_{m/2}$ with root r_R (see Fig. 2). To find the sum $S(T_m)$ of the lengths of all paths in T_m, we note that each such path must begin and end in one of the 3 parts, creating a total of 8 cases (ignoring the ninth case of a path from r to r), which entail symmetries.

$$S(L, L) = S(R, R) = S(T_{m/2}) \quad (11)$$
$$S(r, L) = S(r, R) = S(L, r) = S(R, r) = m/2 - 1 + \sigma(m/2) \quad (12)$$
$$S(L,R) = S(R,L) = (m/2-1)^2[2 + 2\sigma(m/2)/(m/2-1)] \quad (13)$$
$$= (m - 2)\sigma(m/2) + (m - 2)^2/2$$

In (12), each of the $m/2 - 1$ paths is one hop longer than a corresponding path beginning at a subtree root. In (13), each of the $(m/2 - 1)^2$ paths is 2 hops longer than the sum of two paths, one beginning at each of the subtree roots.

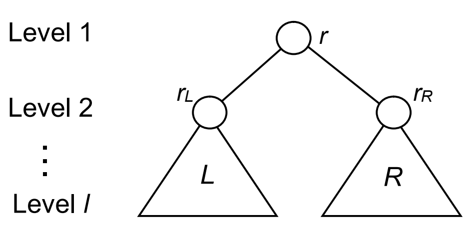

Fig. 2. Complete binary tree, with root r, left subtree L, and right subtree R.

Substituting $\sigma(m/2) = (m \log_2 m)/2 - 3m/2 + 2$ in $S(T_m) = 2S(L, L) + 4S(r, L) + 2S(L, R)$ and simplifying, we arrive at the recurrence:

$$S(T_m) = 2S(T_{m/2}) + m^2 \log_2 m - 2m^2 + 2m \quad (14)$$

The recurrence in (14) has a solution of the form:

$$S(T_m) = am^2 \log_2 m + bm^2 + cm \log_2 m + dm + e \quad (15)$$

Substituting in (14) and solving for the unknowns, we get $a = 2$, $b = -6$, $c = 2$, and $e = 0$, leading to:

$$S(T_m) = 2m^2 \log_2 m - 6m^2 + 2m \log_2 m + dm \quad (16)$$

Finally, from (16), $d = 6$ follows based on the initial condition $S(T_2) = 0$. Thus, we arrive at the solution (17) and average internode distance (18):

$$S(T_m) = 2m^2 \log_2 m - 6m^2 + 2m \log_2 m + 6m \quad (17)$$
$$\Delta(T_m) = (2m^2 \log_2 m - 6m^2 + 2m \log_2 m + 6m)/(m-1)^2 \quad (18)$$
$$= 2 \log_2 m - 6 + 2(3m \log_2 m - 3m - \log_2 m + 3)/(m-1)^2$$

Note that for very large m, the average internode distance for the $(m - 1)$-node complete binary tree T_m asymptotically approaches $2 \log_2 m - 6$, that is, 4 hops less than the diameter in (9); a rather counterintuitive result. Table 1 provides numerical values for Δ and D, as well as the ratio Δ/D, all numerically verified by direct calculation, for small complete binary trees.

So far, we have focused exclusively on complete binary trees. We have also studied balanced binary trees and obtained tight bounds for their diameters and average internode distances. Unbalanced binary trees need not be considered, as they offer no benefits.

Theorem 1: In an incomplete binary tree with more than one incomplete level, removing a node from an incomplete level k and adding a node to an incomplete level $k - j$ ($j > 0$) does not increase the diameter and always reduces the average internode distance. ■

Thus, we focus on incomplete balanced binary trees, in which all missing nodes are in the last level l. The diameter of such a tree is either $2l$ or $2l - 1$. As for average distance, Theorem 2 shows that it is best if the leaves are compressed on one side or the other, bunched together horizontally.

Theorem 2: In a balanced binary tree, with the final level l containing missing nodes in both subtrees, removing a node from a side with equal or fewer nodes and adding a node to the other side decreases the average internode distance, with no increase in diameter. ■

Based on these theorems, any incomplete binary tree can be transformed to a canonical balanced form that has the least average internode distance for its number of nodes.

Table 1. Average distance, diameter, and Δ/D for trees with ≤ 6 levels.

m	n	l	Δ	D	Δ/D
2	1	1	0	0	---
4	3	2	0.889 = 8/9	2	0.444
8	7	3	1.959 = 96/49	4	0.490
16	15	4	3.271 = 736/225	6	0.545
32	31	5	4.795 = 4608/961	8	0.599
64	63	6	6.482 = 25,728/3969	10	0.648

III. Relating Average Distance and Diameter

Clearly, the average distance in a graph is never greater than its diameter ($\Delta \le D$). Other than this obvious relationship, the two parameters are unrelated, in the sense that for any m, we can construct a graph whose diameter is about m times as large as its average internode distance. Referring to Fig. 3, assume that the n-node complete graph K_n is augmented with an m-node path (linear array) P_m emanating from its node v to form the overall graph G. The resulting graph G has diameter $D_G = m$ and the average internode distance.

$$\Delta_G = [n^2 + m(m^2–1)/3 + 2(n–1)(2+3+\ldots+m)]/(n+m–1)^2 \quad (19)$$

The first term inside the square brackets, n^2, is the sum of distances within K_n; the second term, $m(m^2 – 1)/3$, is the sum of distances within P_m, from (2). The third term is the sum of distances between a node in P_m and one in K_n, excluding v as a source or destination. It is readily verified that for a suitably large value of n, such that $m^2 = o(n)$, the average internode distance Δ_G will be $O(1)$, making the diameter $D_G = m$ a factor $O(m)$ larger than the average internode distance. In fact, Δ_G can be made as close to 1 as we wish, making the ratio D_G/Δ_G nearly equal to m.

Worst-case graphs such as G of Fig. 3 are not used in practice for interconnecting nodes in a parallel computer. So, a natural question that arises relates to the worst-case ratio of diameter to average distance in practical networks. A vast majority of practical networks of interest in parallel processing are node-symmetric. For such networks, we can prove the following general result, establishing the fact that the diameter and internode average distance are within a factor of at most 2 of each other.

Theorem 3 [13]: Given a node-symmetric network with node degree d, diameter D, and average internode distance Δ, we have $D/2 \le \Delta \le D$. ■

Both the lower and upper bounds of Theorem 3 are tight. The upper bound for Δ is attained by the complete graph K_n, which has $\Delta = D = 1$. The lower bound for Δ is attained by any ring or torus network, as discussed in Section II, in connection with equations (5) and (7). If we exclude the 0 distance of each node to itself in computing the average distance, then the inequalities in the statement of Theorem 3 become strict: $D/2 < \Delta < D$. However, even in this case we can get arbitrarily close to the bounds for sufficiently large networks. Binary n-dimensional hypercube [5] almost matches the lower bound, while n-dimensional radix-r generalized hypercubes [1], with non-constant n and r, have $\Delta = D – o(D)$.

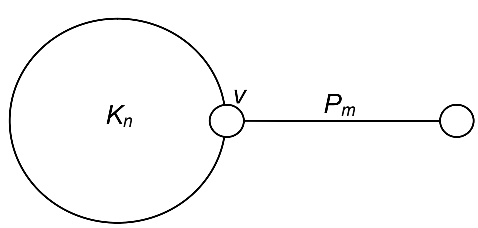

Fig. 3. Graph G composed of the n-node K_n and the m-node path P_m.

Theorem 3 [13] is new, and it is unrelated to previously reported work on the degree-diameter problem [10] or exploring the relationship between Δ and D (see, e.g., [14]). One other context in which the average distance has been studied is that of random, complex networks, but these results are both approximate (probabilistic, asymptotic) and pertain only to certain natural or human-made large-scale networks (see, e.g., [6]).

IV. Some Practical Implications

Let us take the number C of links or channels in an n-node network as a rough measure of its cost. This is a first-order approximation, because it ignores the actual connectivity pattern that affects the area cost of on-chip and on-board connections and the backplane and cabling accommodations for connecting modules, racks, and cabinets. Given a fixed cost C, we would prefer network topologies that minimize D (if worst-case performance is paramount) or Δ (when the average message latency is to be optimized). Because Δ and D are intimately related according to Theorem 3, and given that average performance is a more representative figure of merit, from now on we will use Δ as an indicator of message latency.

It is easy to see that average message latency is proportional to Δ when we use a store-and-forward protocol and the network is carrying light traffic, so that the effects of conflicts and waits are negligible. The same can be said about wormhole-routed messages that are fairly short, because the overall message latency is dominated by the routing time of the header flit. We now show that Δ, and hence D, is an important figure of merit even when we use wormhole switching with long messages, whose latencies are erroneously believed to be insensitive to internode distance, or when the network is heavily loaded.

We begin with a qualitative observation, which we then proceed to quantify. The number of links tied up by a long wormhole message is on average Δ, making the maximum number of in-transit messages no greater than C/Δ (it would generally be much less, given routing restrictions, path conflicts, and so on). So, a reduction in Δ would lead to greater aggregate network bandwidth or fewer conflicts with the same overall traffic.

We focus on oblivious routing, which is much more common in practice [3]. With oblivious routing, the path chosen to route a message from node i to node j is dependent only on the indices i and j and not on any other parameter or network state. Similar arguments can be applied to adaptive routing, but the quantification process is much more involved.

We want to compare two networks that have the same number C of links but different average internode distances Δ and Δ'. The number C of links is a very rough measure of cost and it makes sense to compare equal-cost networks. Even though, as stated earlier, other aspects of a network, such as VLSI layout area and the length of the longest wire affect its cost and per-hop latency, these cannot be taken into account in an architecture-independent discussion.

Table 2. Probability of being able to establish an *i*th routing path.

Δ↓ *i*→	2	3	5	7	10
1	0.990	0.980	0.960	0.940	0.910
2	0.960	0.921	0.846	0.773	0.671
3	0.912	0.829	0.679	0.548	0.385
4	0.847	0.713	0.492	0.327	0.162
5	0.770	0.584	0.319	0.161	0.046
6	0.683	0.455	0.183	0.063	0.008
7	0.592	0.336	0.092	0.019	0.001
8	0.500	0.234	0.040	0.004	0.000
9	0.395	0.154	0.014	0.001	0.000
10	0.310	0.095	0.004	0.000	0.000

We make one more simplifying assumption: That all routing paths have the average length Δ. Consider the probability p_i of being able to establish an *i*th routing path, given that $i - 1$ paths are already in place. This requires that all the required Δ links in the new path be available. Hence:

$$p_i = \binom{C-(i-1)\Delta}{\Delta} \Big/ \binom{C}{\Delta} \qquad (20)$$

For most values of Δ, the probability p_i is a sharply decreasing function of i. To get a feel for the numbers involved, let's take $C = 100$ and compute p_i for different values of i and Δ (Table 2). We see that the expected number of routing paths that can be established before conflicts make additional paths impossible is rather small.

The foregoing observations are basically the routing counterpart to the "birthdays" paradox [2]: That a relatively small number of people (23 to be exact) is needed in a room before there is a more than even chance for at least two of them to share a birthday. Here, as in the birthdays paradox, the probability of an event's occurrence, namely that of multiple paths requiring the use of the same link, is much larger than intuition would lead us to believe. Furthermore, D and Δ play important roles in increasing or decreasing the probability of conflicts.

V. Conclusion

We have shown that the two parameters Δ and D, that is, average distance and diameter, are closely related in regular networks, each being within a factor of 2 of the other. Networks do not have this property in general, but many recursive and semi-regular networks come close. We demonstrated that the relationship $D/3 < \Delta < D$ holds for mesh networks and that, given any arbitrarily small value ε, the relationship $(1 - \varepsilon)D < \Delta < D$ holds for all but a finite number $h(\varepsilon)$ of complete binary trees.

Research can continue in several directions. (1) Look for special classes of regular networks that lead to tighter bounds. Given that there are many Cayley graphs for which Δ can be as low as $D/2$, we suspect that the bounds of Theorem 3 are the best we can hope for, but we have no proof as yet. (2) Perform more detailed analytical and simulation studies to quantify the impact of Δ and D on communication performance more precisely. (3) Derive formulas for average distance in other nonregular and semi-regular networks to gain additional insights on how the two parameters are related. (4) Investigate the impact of routing algorithms, particularly dynamic or fault-tolerant variants, on our conclusions.

We close by reiterating that network diameter and other topological properties are not as unimportant as some researchers have claimed. The space of possibilities for network architectures and associated routing algorithms is vast; the choice is not limited to low- versus high-dimensional mesh/torus networks and variations on wormhole switching [9], although these choices have been dominant in the recent past. Furthermore, it is quite dangerous to generalize from a small number of high-level studies. It is even more dangerous to base the evaluation of research papers and proposals on industrial practices that may have been derived from nontechnical considerations. If a similar mentality prevailed in operating systems, for example, only research on Microsoft Windows, and perhaps Linux, would be deemed appropriate.

References

[1] L. N. Bhuyan and D. P. Agrawal, "Generalized Hypercube and Hyperbus Structures for a Computer Network," *IEEE Trans. Computers*, Vol. 33, pp. 323-333, April 1984.

[2] D. Bloom, "A Birthday Problem," *American Mathematical Monthly*, Vol. 80, pp. 1141-1142, 1973.

[3] J. Duato, "Why Commercial Multicomputers Do Not Use Adaptive Routing," *IEEE Computer Architecture Technical Committee Newsletter*, pp. 20-22, Summer/Fall 1994.

[4] J. Duato, S. Yalamanchili, and L. Ni., *Interconnection Networks: An Engineering Approach*, Morgan Kaufmann, 2002.

[5] K. Efe, "A Variation on the Hypercube with Lower Diameter," *IEEE Trans. Computers*, Vol. 40, No. 11, pp. 1312-1316, November 1991.

[6] A. Fronczak, P. Fronczak, and J. A. Holyst, "Average Path Length in Random Networks," *Physical Review E*, Vol. 70, #056110, 2004.

[7] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, R. Mortier, "Network Topology: Inference, Modeling, and Generation," *IEEE Communications Surveys & Tutorials*, Vol. 10, No. 2, pp. 48-69, 2008.

[8] K. J. Liszka, J. K. Antonio, and H. J. Siegel, "Problems with Comparing Interconnection Networks: Is an Alligator Better than an Armadillo?" *IEEE Concurrency*, Vol. 5, No. 4, pp. 18-28, October-December 1997.

[9] B. M. Maggs, "A Critical Look at Three of Parallel Computing's Maxims," *Proc. Int'l Symp. Parallel Architectures, Algorithms and Networks*, 1996, pp. 1-7.

[10] M. Miller and J. Siran, "Moore Graphs and Beyond: A Survey of the Degree/Diameter Problem," *Electronic J. Combinatorics*, Vol. 61, pp. 1-63, 2005. [A second edition, dated May 2013, is available on-line: http://www.combinatorics.org/ojs/index.php/eljc/article/viewFile/DS14/pdf/

[11] B. Parhami, *Introduction to Parallel Processing: Algorithms and Architectures*, Plenum, 1999.

[12] B. Parhami, "Exact Formulas for the Average Internode Distance in Mesh and Binary Tree Networks," *Computer Science and Information Technology*, Vol. 1, No. 2, pp. 165-168, 2013.

[13] B. Parhami and C.-H. Yeh, "Why Network Diameter Is Still Important," *Proc. International Conf. Communications in Computing*, June 2000, pp. 271-274.

[14] B. Wu, G. Liu, X. An, G. Yan, and X. Liu, "A Conjecture on Average Distance and Diameter of a Graph," *Discrete Mathematics, Algorithms and Applications*, Vol. 3, No. 3, pp. 337-342, September 2011.

[15] J. Xu, *Topological Structure and Analysis of Interconnection Networks*, Kluwer, 2001.