

MINIMAL-DISTANCE ROUTING FOR KYKLOS II

D. Z. Du
MIT
Cambridge, Mass.

F. K. Hwang & A. M. Odlyzko
AT&T Bell Laboratories
Murray Hill, NJ

Y. J. Zhang
University of California
Berkeley, CA

ABSTRACT

We propose a new routing strategy for the KYKLOS II multiprocessor interconnection network which achieves minimum distance for the path between any two processors. For KYKLOS II with 2^n processors, the average distance is shorter than those of previous routing strategies by approximately $2 \log_2 n$. The traffic density, a measure of traffic concentration, is comparable or better than previous strategies for up to two thousand processors, but is soundly beaten by a strategy proposed by Jenevein and Menezes for larger n .

MINIMAL-DISTANCE ROUTING FOR KYKLOS II

D. Z. Du
MIT
Cambridge, Mass.

F. K. Hwang & A. M. Odlyzko
AT&T Bell Laboratories
Murray Hill, NJ

Y. J. Zhang
University of California
Berkeley, CA

1. INTRODUCTION

A *double binary tree* (DBT) of order n is a graph consisting of two copies, T_1 and T_2 , of an n -level complete binary tree T with the $N = 2^n$ terminal nodes of T_1 one-to-one mapped and identified to the 2^n terminal nodes of T_2 . Different mappings result in different DBTs. The DBTs have been proposed for multiprocessor interconnection structures where processors are located at those nodes which are obtained by merging the terminal nodes of T_1 and T_2 . We will label the 2^n processors by the 2^n binary sequences of length n . The identity mapping results in the DBT as studied by Bentley and Kung [1], and also by Imai, Tateizumi, Yoshida and Fukumura [3] (called KYKLOS I by Jenevein and Menezes). The inverse mapping, i.e., mapping the node (a_1, \dots, a_n) of T_1 to the node (a_n, \dots, a_1) of T_2 , results in KYKLOS II, a multiprocessor interconnection structure which Jenevein and Menezes have shown to have many nice properties in several articles [4-7]. In [5] they proposed two routing strategies, the $M2$ -routing and the $H2$ -routing, and suggested that $H2$ has shorter average distance between processors, though it can have larger distances for certain pairs of processors. More importantly, they showed that the traffic density for $H2$ grows on the order of $O(N^{1.5})$ versus $O(N^2)$ for $M2$ and most other trees.

In this paper we propose a new routing strategy $D2$ for KYKLOS II and prove that it has the shortest distance for every pair of processors. We show that its average distance is shorter than that of $M2$ and $H2$ by approximately $2 \log_2 n$ and its traffic density is comparable to that of $H2$ for up to one thousand processors but worse off for larger N .

2. THE $D2$ ROUTING STRATEGY

Let $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ be two processors in KYKLOS II. We will let $d_S(a, b)$ denote the distance between a and b in T under the routing strategy S and let $d_S^1(a, b)$ ($d_S^2(a, b)$) denote the part of $d_S(a, b)$ in T_1 (T_2). Let $d^1(a, b)$ ($d^2(a, b)$) denote the distance in T_1 (T_2) alone. By a *run* we

mean a sequence I of consecutive indices such that $a_i = b_i$ if $i \in I$ but $a_i \neq b_i$ if i is the index (if any) immediately preceding or succeeding the sequence. Let \Rightarrow_i denote the length of the run containing index i ; if $a_i \neq b_i$, then we define \Rightarrow_i to be zero. The following lemma is well known [5] and easily verifiable.

Lemma 1. $d^1(a, b) = 2(n - \Rightarrow_1)$, $d^2(a, b) = 2(n - \Rightarrow_n)$.

We now describe $D2$. Let $L(a, b) = (i + 1, \dots, i + k)$ be a longest run with length $\Rightarrow(a, b) = k$. Then $D2$ chooses one of the two following paths randomly. The first path routes a to the processor $s_1 = (a_1, \dots, a_{i+k}, b_{i+k+1}, \dots, b_n)$ through T_1 and then routes s_1 to b through T_2 . The second path routes a to the processor $s_2 = (b_1, \dots, b_{i+k}, a_{i+k+1}, \dots, a_n)$ through T_2 and then routes s_2 to b through T_1 . It is easily verified that these two paths are node-disjoint, except when $i = 0$ or $n - k$, then the two paths overlap. By Lemma 1 both paths have length $2(n - k)$. We now prove that there is no shorter path. Define $d(a, b) = \min_s d_S(a, b)$.

Theorem 1. $d(a, b) = d_{D2}(a, b) = 2(n - \Rightarrow(a, b))$.

Proof: An equivalent statement of Theorem 1 is that if $d(a, b) = 2d$, then $\Rightarrow(a, b) = n - d$. We prove this equivalent statement by induction on d . It is trivially true for $d = 0$. For general d let p denote a shortest path between a and b . If p contains no other processor, then p uses only one of the two trees T_1 and T_2 . Without loss of generality assume that p lies completely in T_1 . Then the length of p , which by assumption is $2d$, equals $d^1(a, b)$ by Lemma 1. Therefore $\Rightarrow(a, b)$ is at least $n - d$, the length of the first run. On the other hand, $\Rightarrow(a, b)$ is at most $n - d$ for otherwise the $D2$ routing will find a path of length shorter than $2(n - d)$, contrary to our assumption. Hence $\Rightarrow(a, b) = n - d$.

Next we consider the case that p contains a processor $s = (c_1, \dots, c_n)$. Note that

$$d(a, s) + d(b, s) = d(a, b) = 2d .$$

Therefore

$$\max \{d(a, s), d(b, s)\} < 2d .$$

By the inductive assumption

$$\Rightarrow(a, s) = n - d(a, s) ,$$

$$\Rightarrow(b, s) = n - d(b, s) .$$

Now

$$\begin{aligned} \Rightarrow(a, b) &\geq |L(a, s) \cap L(b, s)| \\ &\geq \Rightarrow(a, s) + \Rightarrow(b, s) - n \\ &= n - d(a, s) + n - d(b, s) - n = n - d(a, b) . \end{aligned}$$

As before, we know $\Rightarrow(a, b) \leq n - d(a, b)$; hence equality holds.

Corollary. For each given processor there exists only one other processor which is away from it by the maximum distance $2n$.

3. A COMPARISON OF ROUTING STRATEGIES

The $M2$ routing connects a to b by choosing the shorter of the two paths, one lies completely in T_1 and the other lies completely in T_2 . By Lemma 1

$$\begin{aligned} d_{M2}(a, b) &= \min \{d^1(a, b), d^2(a, b)\} \\ &= 2(n - \max \{\Rightarrow_1, \Rightarrow_n\}) . \end{aligned}$$

Since KYKLOS is processor-symmetric, we need only to study the distances of all processors to processor 0. Let $P_r(x)$ denote the number of processors with distance x to processor 0 under the routing strategy r . Menezes and Jenevein [4] showed that

$$\begin{aligned} P_{M2}(2d) &= 2^d && \text{aif } 1 \leq d \leq \left\lfloor \frac{n}{2} \right\rfloor , \\ &= 2^d - \left\lfloor 3 \cdot 2^{2d-n-2} \right\rfloor && \text{if } 2 \leq d \leq n . \end{aligned}$$

In particular, 2^{n-2} processors have the maximum distance $2n$ to processor 0. The average distance for even n is

$$\begin{aligned}
& \frac{1}{2^n} \left[\sum_{d=1}^n 2d \cdot 2^d - 2^d - \sum_{n/2+1}^n 2d (3 \cdot 2^{2d-n-2}) \right] \\
&= \frac{1}{2^n} \left[4(n2^n - 2^n + 1) - \left[2n2^n - \frac{10}{3} 2^n + n + \frac{13}{3} \right] \right] \\
&= 2^n - \frac{10}{3} + \frac{n + 13/3}{2^n} .
\end{aligned}$$

For odd n the last term is replaced by $(2n + 5/6)/2^n$.

The H_2 routing connects a to b using only the subgraph of T induced by the nodes from level i of T_1 to level $n - i$ of T_2 (the processors are at level 0 of both trees). Furthermore, the path can contain at most one other processor. Therefore

$$d_{H_2}(a, b) = 2(n - \overline{\Rightarrow}_i) \quad \text{where } \overline{\Rightarrow}_i = \Rightarrow_i \text{ if } \Rightarrow_i \geq 1 \text{ and } \overline{\Rightarrow}_i = \Rightarrow_{i+1} \text{ if } \Rightarrow_i = 0 .$$

Since $P_{H_2}(2d)$ and the average distance of processors have not been given explicitly, we derive them in the following. Without loss of generality, assume that $i \leq n - i$. Note that for $\overline{\Rightarrow}_i < i$ there are $\overline{\Rightarrow}_i + 1$ ways of obtaining $\overline{\Rightarrow}_i$, for $i \leq \overline{\Rightarrow}_i < n - i$ there are $i + 1$ way and for $n - i \leq \overline{\Rightarrow}_i$ there are $n - \overline{\Rightarrow}_i + 1$ ways. Furthermore, if the run contains neither index 1 nor index n , then $\overline{\Rightarrow}_i + 2$ bits are fixed; if the run contains either one, then $\Rightarrow_i + 1$ bits are fixed. Thus we have

$$\begin{aligned}
P_{H_2}(2d) &= (d + 3) 2^{d-2} && \text{aif } 1 \leq d \leq i \\
&= (i + 2) 2^{d-2} && \text{if } i < d \leq n - i \\
&= (n - d + 1) 2^{d-2} && \text{if } n - i < d \leq n .
\end{aligned}$$

Again, 2^{n-2} processors have the maximum distance $2n$ from processor 0. The average distance can be more easily computed by computing $d_{H_2}^1$ and $d_{H_2}^2$. Let \Rightarrow'_i denote the length of L_i truncated at index i . Then $d_{H_2}^1 = i - \Rightarrow'_i$. Since there are $2^{i - \Rightarrow'_{i-1}}$ paths of length exactly \Rightarrow'_i for $\Rightarrow'_i < i$ and one path of length i , the average length is

$$\begin{aligned}
 Ed_{H2}^2 \equiv E_i &= 2^{-i} \sum_{\Rightarrow'_i = p}^{i-1} 2(i - \Rightarrow'_i) 2^{i - \Rightarrow'_i - 1} = i \sum_{\Rightarrow'_i = 0}^{i-1} \left[\frac{1}{2} \right]^{\Rightarrow'_i} - \frac{1}{2} \sum_{\Rightarrow'_i = 1}^{i-1} \left[\frac{1}{2} \right]^{\Rightarrow'_i} - 1 \\
 &= i \cdot \frac{1 - \left[\frac{1}{2} \right]^i}{1 - \frac{1}{2}} - \frac{1}{2} \cdot \frac{1 - i \left[\frac{1}{2} \right]^{i-2} + (i-1) \left[\frac{1}{2} \right]^i}{\left[1 - \frac{1}{2} \right]^2} \\
 &= 2i - i \left[\frac{1}{2} \right]^{i-1} - 2 + i \left[\frac{1}{2} \right]^{i-2} - (i-1) \left[\frac{1}{2} \right]^{i-1} \\
 &= 2i - 2 + \left[\frac{1}{2} \right]^{i-1} \quad \text{for } 1 \leq i \leq n - i .
 \end{aligned}$$

Therefore, the average distance is

$$Ed_{H2}^1 + Ed_{H2}^2 = E_i + E_{n-i} = 2n - 4 + \left[\frac{1}{2} \right]^{i-1} + \left[\frac{1}{2} \right]^{n-i-1} \quad \text{for } 1 \leq i \leq n - i .$$

The minimum average distance occurs when $i = \left\lfloor \frac{n}{2} \right\rfloor$ and equals

$$2n - 4 + \left[\frac{1}{2} \right]^{\left\lfloor \frac{n}{2} \right\rfloor - 1} + \left[\frac{1}{2} \right]^{\left\lceil \frac{n}{2} \right\rceil - 1} .$$

Note that this distance is shorter than the average distance of $M2$ only by a constant $\frac{2}{3}$.

Since the length of a longest run is certainly no less than that of the run L_i , we have

$$d_{D2}(a, b) \leq \min \{ d_{M2}(a, b), d_{H2}(a, b) \} .$$

Note that $\Rightarrow(0, b)$ is simply the length of the longest zero-run in the n -sequence b . If we average $\Rightarrow(0, b)$ over all b including 0, then the average equals the expected length of the longest zero-run which was given by Guibas and Odlyzko [2] to be $\log_2 n - 3/2 + \gamma/\log 2 + v(\log_2 n) + O((\log n)^3/n)$ where γ is the Euler constant ($\approx .577 \dots$) and v is a nonconstant continuous periodic function with period 1 and mean 0.

By eliminating $\Rightarrow(0, 0)$ and using Theorem 1 we obtain

Theorem 2. The average distance of two processors in KYKLOS II under $D2$ routing is

$$\frac{2^n}{2^n - 1} \cdot 2 \left[n - \log_2 n + \frac{3}{2} - \frac{\gamma}{\log 2} + v(\log_2 n) + O \left[\frac{(\log n)^3}{n} \right] \right].$$

Hence the average distance of $D2$ is shorter than that of $M2$ and $H2$ by approximately $2 \log_2 n$.

4. TRAFFIC DENSITY

Let $F_{D2}(k)$ denote the frequency that the level- k links (links connecting level- k nodes and level $(k + 1)$ nodes) of T_1 are traveled considering all $\binom{n}{2}$ paths under the $D2$ routing. By symmetry, the corresponding frequency for any given level- k link $T_{D2}(k)$ is $F_{D2}(k)/2^{n-k}$. Define the traffic density of $D2$ as

$$\max_{1 \leq k \leq n} T_{D2}(k) .$$

Then the traffic density is a measure of the concentration of traffic under a routing strategy. Jenevein and Menezes [5] showed that the traffic density for $M2$ is $N^2/64$ and occurs at level $n - 1$, while the traffic density for $H2$ is $N^{1.5}/2$ for even n and $N^{1.5}/\sqrt{2}$ for odd n and occurs at level $\left\lceil \frac{n}{2} \right\rceil$. Thus a major attraction of the $H2$ routing is the reduction of traffic density from order N^2 to order $N^{1.5}$.

Though we are unable to solve for the $D2$ traffic density since we do not have the probability distribution $P_{D2}(2d)$, we can still compute the $D2$ traffic density directly for small n by the following method. By symmetry we need only study the path from processor zero to all other processors and also the traffic density in T_1 . For a fixed n we enumerate the set S_n of all the binary sequences of length n except the one containing all zeros. For each sequence we find its longest run of zeros (if there are more than one, select one randomly) and counts the number of remaining bits after the run. For the sequence of all ones we define the count to be zero or n with equal probability. A count k signifies that the highest node in T_1 that the path reaches is at level k . Let $g_{D2}(k)$ denote the expected number of sequences in S_n with count k , ‘‘expected’’, since k is not fixed when there are more than one longest run. Since any path with count k must go through two links at each level i below k , we have

$$F_{D_2}(k) = 2 \sum_{i=k+1}^n g_{D_2}(i) \quad \text{for } k = 0, \dots, n-1,$$

from which $T_{D_2}(k)$ and the traffic density can be easily computed.

In the following table we show the traffic density and the level it occurs of D_2 for $3 \leq n \leq 10$ and compare it with H_2 .

n	3	4	5	6	7	8	9	10
D_2 -density	8	24	71	213	648	2106	6879	23402
level	1 or 2	2	3	4	5	6	6	7
H_2 -density	32	32	256	256	2048	2048	16384	16384
level	2	2	3	3	4	4	5	5

Thus we see that for $n \leq 10$, the traffic density of D_2 is comparable to H_2 for even n and quite a bit better for odd n . For $n \geq 12$ discrepancy of D_2 and H_2 widens as is clear from the following table.

n	12	14	16	18	20
D_2 -density	2.8×10^5	3.4×10^6	4.3×10^7	5.4×10^8	7.1×10^9
level	9	11	13	15	17
H_2 -density	1.3×10^5	1.0×10^6	8.0×10^6	6.5×10^7	5.3×10^8
level	6	7	8	9	10

For n large the probability of $\Rightarrow (0, b) = i$ can be approximated by $(e^{-n/2^{i+1}} - e^{-n/2^i})$ by using an asymptotic estimate of the probability that a random binary n -sequence contains no run of length i or more [2]. For given i each position in the sequence from $n-i$ to $n-1$ then has approximately equal probability of ending a longest zero-run while position n has about twice as much probability. Ignoring this difference, then we can approximate $g_{D_2}(i)$, $1 \leq i \leq n$, by

$$\sum_{j=0}^{n-i} (e^{-n/2^{j+1}} - e^{-n/2^j}) \frac{1}{n-j}.$$

However, it is not easy to solve for $\max T_{D_2}(k)$.

5. CONCLUSION

We propose a routing strategy $D2$ for KYKLOS II which assures a shortest path between any two processors and is shorter by approximately $2 \ln n$ than $H2$. The traffic density for $D2$ is comparable to $H2$ for up to two thousand processors, in fact, significantly better when $n \leq 9$ is odd. However, $H2$ has much better traffic density characteristic for large n .

REFERENCES

- [1] J. L. Bentley and H. T. Kung, A tree machine for searching problems, Proc. 1979 Inter. Conf. Paral. Proc., pp. 257-266.
- [2] L. J. Guibas and A. M. Odlyzko, Long repetitive patterns in random sequences, Z. Wahrscheinlichkeitstheorie verw. Gebiete 53 (1980), 241-262.
- [3] M. Imai, Y. Tateizumi, Y. Yoshida and T. Fukumura, The architecture and efficiency of a DON: a combinatorial problem oriented multicomputer system, 4th Inter. Conf. Dist. Comput. Systems, pp. 174-182.
- [4] R. M. Jenevein and B. L. Menezes, Mathematical foundations for KYKLOS, Kyklos Tech. Report #1, Mar. 1985.
- [5] R. M. Jenevein and B. L. Menezes, KYKLOS: low tide high flow, Proc. 6th Inter. Conf. Dist. Comput. Systems, pp. 8-15.
- [6] B. L. Menezes and R. M. Jenevein, KYKLOS: a linear growth fault-tolerant interconnection network, Proc. 1985 Inter. Conf. Paral. Proc., pp. 498-502.
- [7] B. L. Menezes, R. M. Jenevein and M. Malek, Reliability analysis of the KYKLOS interconnection networks, Proc. 6th Intern. Conf. Dist. Comput. Systems, pp. 46-53.