Bounds on Oblivious, Conservative
Matrix Transposition Networks

Cynthia Dwork

TR 83-537

Computer Science Department
Upson Hall
Ithaca, N.Y. 14853

This work is partially supported by NSF grant MCS81-01220.

Abstract

A matrix transposition network of depth k is shown to require $\theta(kn^{1+1/k})$ edges.

Definition of the Model and Statement of the Problem

Let S be a fully connected distributed system consisting of a finite number of memoryless, minimally intelligent registers, and a set of tokens, initially distributed among certain distinguished registers called *inputs*. Communication between registers is restricted to passing nonempty packets of tokens, and occurs in a lockstep fashion. Packet delivery is instantaneous. A *step* of the system is the parallel execution by every register of:

Receive all packets sent at the previous step;
Unpack, permute, and repack all tokens into new packets,
    one for each intended receiver;
Send all new packets.

There is a unit cost associated with each packet sent, independent of the number of tokens it contains. No other messages are permitted. The tokens in the input registers are considered to have been sent at step -1.

In addition to the input registers we may designate a set of *output* registers, not necessarily disjoint from the inputs. A *protocol* for the system is a finite sequence of steps ending with every token in an output register.

Let S have m registers, n inputs and n outputs, the latter two sets numbered from 0 to $n-1$. With each input let us associate an ordered set of n tokens. This n-tuple may be viewed as a row of an $n \times n$ matrix. The *matrix transposition* problem is to route the i^{th} token of each input register to the i^{th} output register. We wish to minimize the total number of packets sent as well as the number of steps required. The problem is made nontrivial by assuming that registers have a maximum capacity of n tokens. A *transposition protocol* for the system S is an oblivious, conservative routing scheme solving the matrix transposition problem.

Let P be any k-step protocol for S. We construct the *message graph* $G(P) = (V,E)$, corresponding to P in a natural way. The vertex set, V, is a grid of $k+1$ columns and m rows. A vertex is denoted by specifying its grid coordinates. Column 0 represents the processors in their initial state, and in general column i represents the processors at time i ($i \leq k$). The edges of G represent the set of packets sent during execution of P, and are directed from sender to receiver. Specifically,

$$E = \{<(r,i),(q,i+1)>: \text{ r writes to q at time i } \}$$

A register may send a packet to itself. The *weight* of an edge is the number of tokens in the corresponding packet. The weight of a vertex v, denoted $w(v)$, is the sum of the weights of its incoming edges. If $v = (r,0)$ then $w(v)$ is the number of tokens initially contained in register r.

If G is the message graph of a k-step matrix transposition protocol then the subgraph of G induced by the paths of the tokens is an oblivious, conservative, n-matrix transposition network of depth k.

Lower Bound

For technical reasons we first study an extremely simplified version of the transposition problem. Let S be a system of $m \geq n$ registers with input set $\{a\}$ and outputs $\{b_0, \ldots, b_{n-1}\}$. Initially the input holds an ordered set of n tokens. The *distribution* problem is to send token i to output b_i, for $0 \leq i < n$.

Let D be the message graph for any protocol solving the distribution problem. We say D is an *n-distributor*. Let the tokens be $x_0, \ldots, x_{n-1}$, and let $p(a,b_j)$ denote the path taken by x_j in D. For each vertex v in D let $d(v)$ denote the outdegree of v, and let $\mu(j,v)$ be the characteristic function with value 1 if and only if x_j passes through v on its way from a to b_j. That is,

$$\mu(j,v) = 1 \text{ if } v \text{ is on the path } p(a,b_j)$$
$$\mu(j,v) = 0 \text{ otherwise.}$$

The number of tokens that pass through vertex v is given by $\sum_{0 \le j < n} \mu(j,v)$. This is precisely the weight of v, so we have,

$$w(v) = \sum_{0 \le j < n} \mu(j,v).$$

This quantity is bounded above by n, the maximum capacity of a register.

It is useful to measure how much of the distribution problem is solved by a given vertex, i.e., the progress made when tokens pass through the vertex. Intuitively, we see that a vertex requires large fanout in order to do a lot of distributing. It also must receive sufficiently many tokens to use its fanout in an interesting fashion. These considerations lead to a distribution measure which is the product of the outdegree of a vertex and its weight. Formally, for each vertex v in D the distribution measure of v, denoted $\delta(v)$, is defined as

$$\begin{aligned}\delta(v) &= d(v)\, w(v) \\ &= d(v) \sum_{0 \le j < n} \mu(j,v).\end{aligned}$$

The quantity $d(v)\, w(v)$ is defined for vertices of arbitrary message graphs, and is not restricted to n-distributors.

We define the related quantity, $\Delta(D)$, to be the sum of the distribution measures of the vertices of D. That is,

$$\begin{aligned}\Delta(D) &= \sum_{v \in D} d(v)\, w(v) \\ &= \sum_{v \in D} d(v) \sum_{0 \le j < n} \mu(j,v).\end{aligned}$$

Δ is well defined for any message graph.

Lemma 1 yields a lower bound for $\Delta(D)$ in terms of n and the depth of D.

Lemma 1: If D is an n-distributor of depth at most k then $\Delta(D) \ge kn^{1+1/k}$.

Proof:

Pippenger and Yao [1] have shown that if T is a *tree* with n leaves and depth at most k then $\Delta(T) \geq kn^{1+1/k}$. Let v be an arbitrary vertex in D with indegree $p \geq 2$. D may be modified by creating, for each edge e_i into v, a new vertex v_i, with unique inedge e_i, but with copies of the outedges of v. The modification clearly does not affect the depth of the dag, and

$$\begin{aligned}\sum_{1\leq i\leq p} \delta(v_i) &= \sum_i d(v_i)\, w(v_i) \\ &= \sum_i d(v)\, w(v_i) \\ &= d(v) \sum_i w(v_i) \\ &= d(v)\, w(v) = \delta(v).\end{aligned}$$

so $\Delta(D)$ is unchanged.

If the ancestors of a modified vertex all have indegree at most one then a newly created vertex will never be affected by modifications of its descendents. Thus, an easy induction on the depth of D shows that D can be unfolded to form a tree T of the same depth as D, with $\Delta(D) = \Delta(T)$, and such that T has at least as many outputs as D. It follows from the result of Pippenger and Yao that $\Delta(D) = \Delta(T) \geq kn^{1+1/k}$, as was to be shown. □

Theorem 1: The argument is a modification of the proof of an analagous result in [1] for shifting networks. Let $N = (V,E)$ be the message graph of an n-transposition protocol requiring k steps, and let $\#(E)$ denote the cardinality of E. Then $\#(E) \geq \Omega(kn^{1+1/k})$.

Proof:

For each i, $0 \leq i < n$, let D_i denote the subgraph of N rooted at input a_i. D_i is an n-distributor, as there must be a path from a_i to every output. Let x_{ij} be the j^{th} token initially at input a_i, and let the definition of μ be extended to

$\mu(i,j,v) = 1$ if v is on the path taken by x_{ij} from a_i to b_j
$\mu(i,j,v) = 0$ otherwise.

The weight of v is now given by

$$w(v) = \sum_i \sum_j \mu(i,j,v).$$

Accordingly, the definition of Δ may be extended to N (and to arbitrary message graphs) in the obvious way. Let

$$\begin{aligned}
\Delta(N) &= \sum_{v \in N} d(v)\, w(v) \\
&= \sum_{v \in N} d(v) \sum_{0 \le i < n} \sum_{0 \le j < n} \mu(i,j,v) \\
&= \sum_{0 \le i < n} \sum_{v \in N} d(v) \sum_{0 \le j < n} \mu(i,j,v) \\
&\ge \sum_{0 \le i < n} \sum_{v \in D_i} d(v) \sum_{0 \le j < n} \mu(i,j,v) \\
&= \sum_{0 \le i < n} \Delta(D_i) \\
&\ge kn^{2+1/k}, \text{ by lemma 1.} \quad (*)
\end{aligned}$$

By contrast, since $w(v) \le n$, we have

$$\begin{aligned}
\Delta(N) &= \sum_{v \in N} d(v)\, w(v) \\
&\le \sum_{v \in N} d(v)\, n \\
&= n \sum_{v \in N} d(v) \\
&= n\, \#(E).
\end{aligned}$$

Combining this with (*) yields

$$\begin{aligned}
& n\, \#(E) \ge \Delta(N) \ge kn^{2+1/k} \\
& \Rightarrow \#(E) \ge kn^{1+1/k}.
\end{aligned}$$

□

Note that even if the input and output vertices represent the same physical set of registers, and edges between copies of the same register are deleted, $\#(E)$ is still bounded from below by $kn^{1+1/k} - kn, = kn(n^{1/k} - 1)$.

Upper Bound

We now show that the bounds of the previous section are tight. Let $r \geq 2$ be an integer. An r-ary butterfly network of depth t, denoted B(r,t), is defined recursively for all $t > 0$ as follows:

$t = 1$: $B(r,1) = K_{r,r}$, the complete bipartite graph on r vertices, with all edges directed from inputs to outputs.

$t > 1$: B(r,t) is constructed from r copies of B(r,t-1). It has r^t inputs and outputs, each set numbered from 0 to $r^t - 1$. For each i in this range there is an edge directed from input i to input $i \ mod \ (r^{t-1})$ of each copy of $B(r,t-1)$. Thus, the inputs of $B(r,t)$ all have outdegree r, and each of the inputs of the r copies of $B(r,t-1)$ has indegree r.

Let $\#\ (G)$ denote the number of edges in an arbitrary graph G.

Claim: $\#\ (B(r,t)) = tr^{t+1}$.

Proof: The claim is immediate for $t = 1$. Assume, inductively, that $B(r,t-1)$ has $(t-1)r^t$ edges. Since each input of B(r,t) has outdegree r and there are r^t inputs, we have

$$\begin{aligned} \#(B(r,t)) &= (r^t)r + r(\#(B(r,t-1))) \\ &= r^{t+1} + r((t-1)r^t) \\ &= tr^{t+1}. \quad \square \end{aligned}$$

Let the vertices of B(r,t) have capacity r^t.

Lemma 2: B(r,t) is the graph of an r^t-transposition protocol.

Proof (sketch): For $t = 1$ the result is obvious. For $t \geq 2$ assume the claim inductively for $t - 1$. Let the r^t tokens initially at input i be numbered from 0 to $r^t - 1$, according to their destinations. Input i sends tokens jr^{t-1}, $jr^{t-1}+1,\ldots,(j+1)r^{t-1}-1$ to the j^{th} copy of B(r,t-1), for $0 \leq j < r$. By the inductive hypothesis the copies of B(r,t-1) are r^{t-1}-transposition networks, but with vertex capacity r^t. $\square$

Theorem 2: Let n and k be such that $r = n^{1/k}$ is an integer. Then there exists and n-transposition network of depth k and $kn^{1+1/k}$ edges.

Proof: Immediate from lemma 2 and the definition of $B(r,k)$. □

Finally, we note that if vertices at each level of this construction represent the same physical set of registers then kn edges may indeed be deleted from $B(r,k)$, and the lower and upper bounds are again identical.

Reference

[1] Pippenger, N. and Yao, A. C.-C., "Rearrangeable Networks with Limited Depth," IBM Research Report RJ3169, 6/23/81