Circuit-Switched Gossiping in 3-Dimensional Torus Networks

Olivier Delmas[1] and Stéphane Perennes[1]

SLOOP Project (CNRS-INRIA-UNSA)
Laboratoire I3S - CNRS URA 1376
930 Route des Colles, B.P. 145 - 06903 Sophia Antipolis Cedex
E-mail: {delmas, sp}@unice.fr

Abstract. In this paper we describe an efficient gossiping algorithm for short messages into the 3−dimensional torus networks (wrap-around or toroidal meshes) that uses synchronous circuit-switched routing. The algorithm is based on a recursive decomposition of a torus. It requires an optimal number of rounds and a quasi-optimal number of intermediate switch settings to gossip in a $7^i \times 7^i \times 7^i$ torus network.

1 Introduction

Distributed memory multicomputer architectures in which the processors communicate by exchanging messages over an interconnection network are very useful techniques for achieving cost-effective high-performance computing. At present the "circuit-switched" like routing (wormhole, direct connect, virtual cut-through, ...) is used in many recent multicomputer systems such as the Intel Paragon, IBM SP2, Cray-T3D or more recently the new Cray-T3E.

In this paper, we study the circuit-switched gossiping in 3−dimensional torus networks. After the description of the model of communications, we recall some classical definitions used in this study. In section 4 we establish a new non-trivial lower bound for the number of rounds from a circuit-switched gossiping protocol. In the last section we present a new circuit-switched gossiping algorithm for 3−dimensional torus networks which uses linear cost model. We prove that our algorithm is optimal in terms of number of rounds and quasi-optimal in terms of number of intermediate switch settings. This algorithm gives an efficient protocol when the messages are short or when the time to initiate a message transmission is much larger than the unit propagation time of a message along a link. This is the situation in many current multiprocessor networks.

2 Models of communication

In this paper we will consider the circuit-switched routing model. We will use the linear cost model in which the transmission time for a message of length L to be sent at distance d is $\alpha + d\delta + L\tau$, where α is the time to initiate a new message transmission, δ is the time to switch an intermediate node, and $1/\tau$

is the bandwidth of the communication links. We will use the all-port model of communication in which a processor can use all of its communications links simultaneously. We also assume that the communication links are full-duplex so that messages can travel in both directions simultaneously. Finally, we assume that each node has an initial distinct message, but all these messages have the same length L and we allow messages to be concatenated with negligible cost.

3 Definitions

In this article, $\mathbb{Z}_q$ will denote the set of integers modulo q. G will denote a **digraph** of order N with vertex set $V(G)$ and arc set $A(G)$. The **distance** $d_G(x,y)$ will denote the length of the shortest dipath from a vertex x to a vertex y. $D(G)$ will denote the **diameter** of a digraph G (i.e. $D(G) = max_{(x,y)\in V^2(G)} d_G(x,y)$). In a symmetric digraph G, $\Delta(G)$ (or shortly Δ) will denote **maximum in-degree** of G, that is the maximum over the in-degrees of all vertices $V(G)$.

Definition 1 [5]. The $k-$dimensional torus is the cartesian sum of k symmetric circuits of orders $p_1, p_2, \ldots, p_k$ and is denoted by $TM(p_1, p_2, \ldots, p_k) = C_{p_1} \Box C_{p_2} \Box \cdots \Box C_{p_k}$, where C_{p_i} denote the symmetric circuit of order p_i.

Remark. When $p_1 = p_2 = \cdots = p_k$, we will use the abbreviated notation $TM(p)^k$, We will assume that $p \geq 3$.

Definition 2. The total time necessary to achieve a gossiping protocol in a digraph G will be denoted by $g(G) = g_\alpha(G)\alpha + g_\delta(G)\delta + g_\tau(G)\tau$ where $g_\alpha(G)$ is the number of rounds, $g_\delta(G)$ the sum of the maximum communication distances of each couple of processors implicated in each round of the gossiping protocol and $g_\tau(G)$ measure the flow of information.

As said before, here we consider only short messages (or equivalently suppose $\tau << \alpha$ and $\tau << \delta$). So we are mainly interested in determining the optimal $g_\alpha(G)$ and $g_\delta(G)$. For $g_\delta(G)$ a trivial lower-bound is the diameter $D(G)$. In the next section we give a new non-trivial lower-bound for $g_\alpha(G)$.

4 Lower bounds

First let $\pi(G)$ be the **arc-forwarding index of the digraph** G (see [2]). For any digraph we have establised the following theorem.

Theorem 3. *Let G be a digraph with maximum degree Δ and order N and $t_0 = \lceil \log_{\Delta+1}(N) \rceil$. If $g_\alpha(G) \leq 2t_0$ then $g_\alpha(G) \geq t_0 + \log_{\Delta+1}(\frac{\pi(G)}{N}) - O(\log_{\Delta+1} \log_{\Delta+1}(N))$.*

The idea of the proof is based on a precise enumeration of the load of dipaths which can be used in a gossiping protocol. This notion is similar to the arc-forwarding index which uses the load of the arcs. Now, with this theorem we are able to state the following corollary.

Corollary 4. *Given a gossip protocol in the digraph $TM(n)^k$, the number of rounds necessary to achieve this protocol is* $g_\alpha(G) \geq (k+1)\log_{2k+1}(n) - O(\log_{2k+1}\log_{2k+1}(n))$.

Proof. This corollary is correct as in [2] it has been proved that $\pi(TM(n)^k) = \frac{n^{k-1}}{2}\lfloor\frac{n^2}{4}\rfloor$ and in [1] it has been shown that the total number of rounds to achieve a broadcasting protocol in the $TM(n)^k$ digraph is t_0. Then a trivial gossiping protocol will be the concatenation of 2 broadcasting protocols and $g_\alpha(G) \leq 2t_0$. Therefore the result follows immediatly.

5 Gossiping in the 3−dimensional torus $TM(7^i)^3$

5.1 Case of $TM(7)^3$

The idea of this section come from the original study of J.G. Peters and M. Syska [4] for the circuit-switched broadcast in the 2-dimensional torus. Here G denote the $TM(7)^3$ symmetric digraph for which we have established the following proposition.

Proposition 5. *There exists a gossiping protocol on the symmetric digraph $G = TM(7)^3$ with time $g(G) = 4\alpha + 12\delta + (1+7+7^2+7^3)L\tau$.*

The proof is based on the description of the gossiping protocol. But before describing it, we need some additional notations and definitions.

We will consider the vertices of G as elements of the 3−dimensional vector-space $\mathbb{Z}_7^3$, with canonical base $\{e_1, e_2, e_3\}$ [1]. If M is a 3−dimensional matrix and U is a set of vectors, MU will denote the image of U by $M : \{Mx \mid x \in U\}$. The sum of two sets of vectors U_1 and U_2 will be $U_1 + U_2 = \{x \mid x = u_1 + u_2, u_1 \in U_1, u_2 \in U_2\}$. We will denote by B_1 the set $\{e_1, e_2, e_3, 0, -e_1, -e_2, -e_3\}$ of vectors whose norm is less than or equal to 1. Note that B_1 is the sphere of radius 1 centered at zero of $\mathbb{Z}_7^3$ for the Lee distance (see [3]). $x + B_1$ is the sphere of radius 1 centered at x and it also contains the neighbours of x in G union x.

Definition 6 [3]. The **code** $\mathcal{C}$ is the set of vertices such that $\mathcal{C} = \{(x_1, x_2, x_3) \in \mathbb{Z}_7^3 | x_1 + 2x_2 + 3x_3 = 0\}$.

Remark [3]. $\mathcal{C}$ is a linear code of length 3 defined over $\mathbb{Z}_7$. As $\mathbb{Z}_7^3 = \mathcal{C} + B_1$, then $\mathcal{C}$ is a perfect Lee code. Note that $\mathcal{C}$ has 49 elements like $(0,0,0), (2,-3,-1), \cdots$

To describe the gossiping protocol of proposition 5 we introduce an additional notation.

Let x be a vertex of a digraph G and $\mathcal{A} \subset V(G)$. The notation $x \to \mathcal{A}$ *(resp.* $x \leftarrow \mathcal{A}$*)* is used when the vertex x sends its message towards all the vertices of $\mathcal{A}$ *(resp. all the vertices of* $\mathcal{A}$ *send their own message towards the vertex* x*)*.

---

[1] Vertex $(x_1e_1 + x_2e_2 + x_3e_3)$ will be denoted as the vector $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$.

We have found (see [1]) a 3–dimensional matrix M_0 which performs the following algorithm as a gossiping protocol.

Begin ——— *Gossiping Algorithm* ——— *in* $TM(7)^3$.

- **Round 1 - Concentration:** $\forall x \in \mathcal{C}$, $x \leftarrow \{x + B_1\}$ ..... [*Cost:* $\alpha + \delta + L\tau$].
- **Step 2 - Gossiping between the vertices of** $\mathcal{C}$
  - **Round 2-a:** $\forall x \in \mathcal{C}$, $x \rightarrow \{x + M_0 B_1\}$ ..... [*Cost:* $\alpha + 5\delta + 7L\tau$].
  - **Round 2-b:** $\forall x \in \mathcal{C}$, $x \rightarrow \{x + M_0^2 B_1\}$ ..... [*Cost:* $\alpha + 5\delta + 7^2 L\tau$].
- **Round 3 - Final broadcasting:** $\forall x \in \mathcal{C}$, $x \rightarrow \{x + B_1\}$. [*Cost:* $\alpha + \delta + 7^3 L\tau$].

End ——— *Gossiping Algorithm* ——— *in* $TM(7)^3$.

With a precise analyse of the algorithm we have been able to exhibit a set of dipaths in G realizing each round of communication of the algorithm.

5.2 Generalization for torus $TM(7^i)^3$

Here G_i denotes the symmetric digraph $TM(7^i)^3$. We have generalized the previous result to the $TM(7^i)^3$ torus digraph.

Proposition 7. *There exists a gossiping protocol on the symmetric digraph* $G_i = TM(7^i)^3$ *with time* $g(G_i) = 4i\alpha + \frac{12}{9} D(G_i)\delta + [\frac{57}{49}(7^{i-1} - 1) + \frac{7^3}{7^{2i}} - \frac{1}{7^{3i}}]\frac{NL}{6}\tau$.

The main idea of this section is to apply recursively the gossiping protocol designed for the torus $TM(7)^3$ to the torus $TM(7^i)^3$. For this we use the **code** $\mathcal{C}_i$ which is the subset of the vertices of G_i defined as $\mathcal{C}_i = \{(x_1, x_2, x_3) \in \mathbb{Z}_{7^i}^3 | x_1 + 2x_2 + 3x_3 \equiv 0 \pmod 7\}$. The recursion is possible because this code is once again a perfect code for the Lee distance. Indeed, spheres of radius 1 centered at each vertex of the code $\mathcal{C}_i$ cover completely the digraph G_i. That is $V(G_i) = \mathbb{Z}_{7^i}^3 = B_1 + \mathcal{C}_i$. The code $\mathcal{C}_i$ has 7^{3i-1} elements.

Acknowledgments

The authors are grateful to J-C. Bermond, M. Syska and J. Yu for helpful discussions and remarks.

References

1. O. Delmas and S. Perennes. Diffusion en mode commutation de circuits. In *RenPar'8*, pages 53–56, Bordeaux, France, 20-24 May 1996. Edition spéciale, présentation des activités du GDR-PRC Parallélisme, Réseaux et Systèmes, Richard Castanet and Jean Roman.
2. M.C. Heydemann, J.C. Meyer, and D. Sotteau. On forwarding indices of networks. *Discrete Applied Mathematics*, 23:103–123, 1989.
3. F.J. MacWilliams and N.J.A. Sloane. *The theory of Error-Correcting Codes.* North-Holland, 1977.
4. J.G. Peters and M. Syska. Circuit-switched broadcasting in torus networks. *IEEE Transactions on Parallel and Distributed Systems*, 7(3):246–255, March 1996.
5. Jean de Rumeur. *Communication dans les réseaux de processeurs.* Collection Etudes et Recherches en Informatique. Masson, 1994. (English version to appear).