Efficient Deadlock-free Multi-dimensional Interval Routing in Interconnection Networks*

Rastislav Kráľovič, Branislav Rovan, Peter Ružička, Daniel Štefankovič

Department of Computer Science
Faculty of Mathematics and Physics
Comenius University, Bratislava
Slovak Republic

Abstract: We present deadlock-free packet (wormhole) routing algorithms based on multi-dimensional interval schemes for certain multiprocessor interconnection networks and give their analysis in terms of the compactness and the size (the maximum number of buffers per node (per link)). The issue of a simultaneous reduction of the compactness and the size is fundamental, worth to investigate and of practical importance, as interval routing and wormhole routing have been realized in INMOS Transputer C104 Router chips.
In this paper we give an evidence that for some well-known interconnection networks there are efficient deadlock-free multidimensional interval routing schemes (DFMIRS) despite of a provable nonexistence of efficient deterministic shortest path interval routing schemes (IRS). For d-dimensional butterflies we give a d-dimensional DFMIRS with constant compactness and size, while each shortest path IRS is of the compactness at least $2^{d/2}$. For d-dimensional cube connected cycles we show a d-dimensional DFMIRS with compactness and size polynomial in d, while each shortest path IRS needs compactness at least $2^{d/2}$. For d-dimensional hypercubes (tori) we present a d-dimensional DFMIRS of compactness 1 and size 2 (4), while for shortest path IRS we can achieve the reduction to 2 (5) buffers with compactness 2^{d-1} ($O(n^{d-1})$).
We also present a nonconstant lower bound (in the form $\sqrt{d}$) on the size of deadlock-free packet routing (based on acyclic orientation covering) for a special set of routing paths on d-dimensional hypercubes.

1 Introduction

Interval routing is an attractive space-efficient routing method for communication networks which has found industrial applications in INMOS T9000 transputer design. Survey of principal theoretical results about interval routing can be found in [14, 6].

Interval routing is based on compact routing tables, where the set of nodes reachable via outgoing links is represented by intervals. The space efficiency can be measured by *compactness*, i.e. the maximum number of intervals per link.

* This research has been partially supported by the Slovak Research Grant VEGA 1/4315/97.

Previous work mostly concentrated on *shortest path* interval routing schemes (IRS). Shortest path IRS of compactness 1 are known to exist for a number of well-known interconnection networks including trees, rings, complete bipartite graphs, meshes, and tori. There are however networks that are known to have no shortest path IRS even for large compactness, which include shuffle-exchange, cube-connected cycles, butterfly, and star graphs. Several generalizations of IRS were therefore proposed.

Multidimensional interval routing schemes (MIRS) were introduced in [5] and were used to represent all the shortest paths information. MIRS with low memory requirements were proposed for hypercubes, grids, tori and certain types of chordal rings [5].

Another interesting aspect of the routing problem is related to deadlocks. A deadlock refers to a situation in which a set of messages is blocked forever because each message in the set occupies buffer in a node or on a link which is also required by another message. Deadlock-free routing is relevant in the framework of packet and wormhole routing protocols [2, 3, 8, 12, 13]. The first study dealing with deadlock-free IRS appeared in [11]. Further results were presented in [10, 15, 16]. We follow the model of buffered deadlock-free IRS introduced in [4] based on the notion of acyclic orientation covering. An s-buffered deadlock-free IRS with compactness k is denoted as (k, s)-DFIRS. Some results were already presented in [4]. For d-dimensional tori there exists a shortest path $(2, 2d+1)$-DFIRS; the reduction to 5 buffers can be achieved with compactness $O(n^{d-1})$. For d-dimensional hypercubes there is a shortest path $(1, d+1)$-DFIRS; the reduction to 2 buffers can be achieved with compactness 2^{d-1}.

We extend the model in [4] to buffered deadlock-free multi-dimensional interval routing (MIRS). We show that for some interconnection networks there are efficient deadlock-free MIRS even in the case when there does not exist efficient shortest path IRS. For butterflies of order d we give a deadlock-free d-dimensional MIRS with constant compactness and size, while each shortest path IRS needs compactness at least $2^{d/2}$. For cube connected cycles of order d we present a deadlock-free d-dimensional MIRS with compactness and size polynomial in d, while each shortest path IRS needs compactness at least $2^{d/2}$. For d-dimensional hypercubes we give a deadlock-free d-dimensional MIRS of compactnes 1 and size 2. And for d-dimensional tori we show a deadlock-free d-dimensional MIRS of compactness 1 and size 4.

There exist only few lower bounds on the size of deadlock-free packet routing, even for those based on specific strategies. The best lower bound is 3 (see [2]). We give the first nonconstant lower bound (in the form $\sqrt{d}$) on the size of deadlock-free packet routing (based on acyclic orientation covering) for a special set of routing paths on d-dimensional hypercubes. As a consequence, the set of routing paths induced by 1-IRS on the hypercube proposed in [1] is not suitable for the efficient deadlock-free packet routing based on acyclic orientation covering concept.

2 Definitions

An interconnection network is modeled by an undirected graph $G = (V, A)$, where V is a set of nodes and A is a set of links of the network. Assume $|V| = n$. Each node has a finite set of buffers for temporarily storing messages. The set of all buffers in the network G is denoted as B.

A *communication request* is a pair of nodes in G. A *communication pattern* $\mathcal{R}$ is a set of communication requests. We will consider certain significant communication patterns in G. A *static one-to-all* communication pattern is a set $\{(v, w) \mid w \in V\}$ for a given source node v. A *dynamic one-to-all* communication pattern is a set $\{(v, w) \mid w \in V\}$ for some (not given in advance) source node v. An *all-to-all* communication pattern is a set $\{(v, w) \mid v, w \in V\}$. A collection $\mathcal{P}$ of paths in G *satisfies* the communication pattern $\mathcal{R}$ if there is at least one path in G beginning in u and ending in v for each communication request $(u, v) \in \mathcal{R}$.

The *routing problem* for a network G and a communication pattern $\mathcal{R}$ is a problem of specifying a path collection $\mathcal{P}$ satisfying $\mathcal{R}$. A path collection is simple if no path contains the same link more than once, and it is a shortest path collection if for each $(u, v) \in \mathcal{R}$ only shortest paths from u to v in G are considered. Satisfying a communication request consists of routing a message along a corresponding path in $\mathcal{P}$. In this paper, the routing problem is solved by a path collection induced by interval routing schemes. In what follows we shall consider all-to-all communication patterns only unless otherwise specified.

An *Interval Labeling Scheme* (ILS) is given by labeling each node in a graph G by a unique integer from the set $\{1, 2, ..., n\}$ and each link by an interval $[a, b]$, where $a, b \in \{1, 2, ..., n\}$. We allow cyclic intervals $[a, b]$ such that $[a, b] = \{a, a+1, ..., n, 1, ..., b\}$ for $a > b$. The set of all intervals associated with the links incident with a node must form a partition of the set $\{1, 2, ..., n\}$. Messages to a destination node having a label l are routed via the link labeled by the interval $[a, b]$ such that $l \in [a, b]$. An ILS is valid if the set of paths specified by this ILS satisfies the all-to-all communication pattern. (Thus, if, for all nodes u and v in G, messages sent from u to v reach v correctly, not necessarily via shortest paths.) A valid ILS is also called an *Interval Routing Scheme* (IRS). An IRS thus specifies for each pair of distinct nodes u and v in G a (unique) path from u to v.

In a k-ILS each link is labeled with up to k intervals, always under the assumption that at every node, all intervals associated with links outgoing from the node form a partition of $\{1, ..., n\}$. At any given node a message with destination node labeled l is routed via the link labeled by the interval containing l. If k-ILS does not use cyclic intervals, the k-ILS is called *linear* or simply k-LILS. Valid k-ILS and k-LILS are called k-IRS and k-LIRS respectively. A k-IRS (k-LIRS) is said to be *optimal* if it represents a shortest path collection containing exactly one shortest path between any pair of nodes.

Multi-dimensional interval routing schemes (MIRS for short) are an extention of interval routing schemes. In (k,d)-MIRS every node is labeled by a unique d-tuple $(l_1, ..., l_d)$, where each l_i is from the set $\{1, ..., n_i\}$ $(1 \leq n_i \leq n)$. Each link is labeled by up to k d-tuples of cyclic intervals $(I_{1,1}, ..., I_{d,1}), ..., (I_{1,k}, ..., I_{d,k})$.

In any node a message with destination $(l_1, ..., l_d)$ is routed along any outgoing link containing a d-tuple of cyclic intervals $(I_1, ..., I_d)$ such that $l_i \in I_i$ for all i. In this case, multiple paths are represented by the scheme, so the intervals on the links of a given node may overlap, i.e. they do not form a partition of the nodes in V.

We intend to model the packet routing, i.e. the so called store-and-forward message passing in which the message from u to v passing via w has to be stored at the node w before it is sent further towards v. We shall assume each node contains a finite number of *buffers*. For a message to pass via a link (x, y) it means, that it has to be moved from a buffer at node x to a buffer at node y. This assumes the existence of an available (i.e., empty) buffer at y.

We follow the notions introduced in [4]. In packet routing, each message is represented by its source-destination pair. For a given message $m = (u, v)$ and a buffer b containing m, a controller $C : V \times V \times B \mapsto 2^B$ specifies the subset $C(u, v, b)$ of buffers which can contain m in the next step along the path to its destination v. We say that a controller C is deadlock-free if it does not yield any deadlock configuration. This property can be guaranteed if the resulting buffer dependencies graph is acyclic. In *buffer dependencies graph* [8] , each node represents a buffer and there is a directed edge between b_i and b_j if there is at least one message $m = (u, v)$ such that $b_j \in C(u, v, b_i)$.

Let us by s_u denote the number of buffers used by a controller C at the node u. For a network $G = (V, A)$ and a controller C for G, we define the *size* s of C as $s = max_{u \in V}(s_u)$.

Assume a path $\pi = v_1, ..., v_r$ connecting v_1 to v_r. We say that the controller C *covers* π if there exist r buffers $b_1, ..., b_r$ such that for each i, $1 \leq i \leq r$, b_i belongs to v_i and for each i, $1 \leq i \leq r-1$, $b_{i+1} \in C(v_1, v_r, b_i)$.

We need to extend the standard k-IRS to deadlock-free k-IRS. Notice that each k-IRS uniquely induces the set of simple paths, one for each pair of nodes in G. A (k, s)-DFIRS (*deadlock-free IRS*) for a graph G is a k-IRS for G together with a deadlock-free routing controller of size s for G which covers the set of paths represented by the k-IRS. The (k, s)-DFIRS is optimal if the k-IRS is optimal.

All controllers considered in this paper are based on the concept of an acyclic orientation covering. An *acyclic orientation* of a graph $G = (V, A)$ is an acyclic directed graph $DG = (V, DA)$ obtained by orienting all links in A. Let $\mathcal{G} = \langle DG_1, ..., DG_s \rangle$ be a sequence of (not necessarily distinct) acyclic orientations of a graph G and let $\pi = v_1, ..., v_r$ be a simple path in G. We say that $\mathcal{G}$ *covers* π if there exists a sequence of positive integers $j_1, ..., j_{r-1}$ such that $1 \leq j_1 \leq ... \leq j_{r-1} \leq s$ and for every i, $1 \leq i \leq r-1$, (v_i, v_{i+1}) belongs to DG_{j_i}.

Note that a path π need not be covered by $\mathcal{G}$ in a unique way. There could be different sequences $k_1, ..., k_{r-1}$ such that (v_i, v_{i+1}) belongs to DG_{k_i}. But there exists a unique sequence such that the corresponding $(r-1)$-tuple $(k_1, ..., k_{r-1})$ is minimal (w.r.t. the lexicographical ordering). We assume that the deadlock-free controller based on $\mathcal{G}$ works with minimal tuples. Such a controller is called *greedy*.

Let $\mathcal{P}$ be a set of simple paths connecting every pair of nodes in G. A sequence of orientations $\mathcal{G} = \langle DG_1, ..., DG_s \rangle$ is said to be an *acyclic orientation covering* for $\mathcal{P}$ of size s if $\mathcal{G}$ covers at least one path $\pi \in \mathcal{P}$ for each pair of nodes in G. A $((k,d),s) - DFMIRS$ (*deadlock-free MIRS*) for a graph G is a $(k,d) - MIRS$ for G together with a deadlock-free controller of size s for G which covers the set of paths induced by the $(k,d) - MIRS$.

The main problem covered is to design the deterministic packet routing protocol based on a possibly nondeterministic $(k,d) - MIRS$ with a deadlock-free routing controller (based on acyclic orientation covering $\mathcal{G} = \langle DG_1, ..., DG_s \rangle$) of size s for G. In this paper we solve this problem by applying the greedy mode. At the source node, the message destined for the node labeled l is routed via a link e having the interval containing l and satisfying $e \in DG_1$. Only if such a possibility does not exist it chooses the next orientation DG_2. Generally, at an arbitrary node, the protocol first chooses a link in the current orientation DG_j according to $(k,d) - MIRS$ and only if such a link does not exist, it switches to the next acyclic orientation DG_{j+1} in $\mathcal{G}$. We call this strategy a greedy one. All $((k,d),s) - DFMIRS$ in this paper are working with the greedy strategy.

The importance of acyclic orientation coverings is stated by the following classical result (see [11]) formulated for all-to-all communication patterns: given a network G and a set of simple paths $\mathcal{P}$ connecting all pairs of nodes in G, if an acyclic orientation covering of size s for $\mathcal{P}$ exists, then there exists a deadlock-free packet routing controller of size s for G which covers $\mathcal{P}$.

3 Results

The size of deadlock-free controllers for the optimal (shortest paths) packet routing on arbitrary networks strongly depends on the structure of communication patterns. The following fact for all-to-all communication patterns can be found e.g. in [11]: for any network G and a set of $n.(n-1)$ shortest paths connecting every pair of nodes in G, there is a deadlock-free controller (based on an acyclic orientation covering) of size $D+1$, where D is the diameter of G. The best lower bound on the size of deadlock-free controllers is 3 [2].

Considering all-to-all communication patterns on arbitrary networks, the problem is to determine nonconstant lower bound on the size of a deadlock-free controller (based on acyclic orientation covering concept) necessary for the optimal packet routing.

However, if we assume static one-to-all communication patterns, the requirements for the size of deadlock-free controllers are much lower. Namely, for any network G and a set of $n-1$ shortest paths connecting a node with all other nodes in G, there is a deadlock-free controller (based on acyclic orientation covering) of size 1.

For other types of communication patterns the problems are again unsolved. What is the number of buffers sufficient to realize dynamic one-to-all or permutation communication patterns? Can we do better than $D+1$ buffers per node?

We shall concentrate on specific networks. We shall study the relationship between the size and the compactness of deadlock-free packet routing, based on interval routing schemes, for certain interconnection networks including hypercubes, tori, butterflies and cube connecting cycles.

3.1 Hypercubes

A d-dimensional hypercube H_d is the cartesian product of d complete graphs K_2.

Lemma 1. *There exists a deadlock-free controller of size* 2 *for the optimal packet routing on a d-dimensional hypercube.*

Proof. A hypercube H_d is a node symmetric graph, so we can fix an arbitrary node as the initiator of H_d and assign it the string 0^d. Let the unique string of the nodes in H_d be from $\{0,1\}^d$ such that two nodes are neighbors if and only if their strings differ in exactly one bit. Define the acyclic orientation covering $\mathcal{G} = \langle DH_1, DH_2 \rangle$ of a hypercube such that in DH_1 all links are oriented from all the nodes towards the initiator and in DH_2 the orientation is opposite.

It is easy to verify that $\mathcal{G}$ forms a greedy deadlock-free controller of size 2 for H_d. There exists a collection of shortest paths between all pairs of vertices in H_d, covered by $\mathcal{G}$. Given any two nodes u and v in H_d with corresponding strings α and β, a shortest path from u to v follows

- in the first place links (in arbitrary order) changing bit 1 to 0 in all positions in which α has 1 and β has 0, and
- later on links (in arbitrary order) changing bit 0 to 1 in all positions in which α has 0 and β has 1.

□

When we consider dynamic one-to-all communication patterns instead of all-to-all communication patterns, we get the following consequence of the previous lemma.

Corollary 2. *There exists a deadlock-free controller of size* 2 *for the optimal packet routing on a d-dimensional hypercube with dynamic one-to-all communication patterns.*

The next two results are from [4]. When we consider linear interval routing schemes, the size $d+1$ can be obtained with compactness 1, and the reduction to the size 2 can be achieved with the compactness 2^{d-1}.

Lemma 3. *For every* i $(1 \leq i \leq d)$ *there exists a* $(2^{i-1}, \lceil d/i \rceil + 1) - DFLIRS$ *for a d-dimensional hypercube.*

Corollary 4. *There exists a* $(1, d+1) - DFLIRS$ *on a d-dimensional hypercube.*

We now show that using d-dimensional interval routing schemes (see [5]) the size 2 can be achieved with compactness just 1.

Theorem 5. *For every i $(1 \le i \le d)$ there exists a $((2^{i-1}, \lceil d/i \rceil), 2) - DFMIRS$ for a d-dimensional hypercube.*

Proof. Consider a d-dimensional hypercube $H_d = (V, A)$, given as the product of $\lfloor d/i \rfloor$ subcubes of dimension i and a subcube of dimension $d \ mod \ i$. For simplicity, assume $d \ mod \ i \ = \ 0$. Observe that each of these d/i subcubes $H_i^{(j)} = (V_j, A_j)$, $1 \le j \le d/i$, of dimensions i admits a $(\lceil 2^{i-1}/i \rceil, 1)$-MIRS.

We label each node in V by the d-tuple

$$(l_{1,1}, ..., l_{1,i}, l_{2,1}, ..., l_{2,i}, ..., l_{d/i,1}, ..., l_{d/i,i})$$

$(l_{p,q} \in \{0,1\}, 1 \le p \le d/i, 1 \le q \le i)$ where for each j, $(l_{j,1}, ..., l_{j,i})$ is the label of a node in V_j in the $(\lceil 2^{i-1}/i \rceil, 1)$-MIRS of $H_i^{(j)}$.

We label each link $e = ((l_1, ..., l_h, ..., l_d), (l_1, ..., \hat{l}_h, ..., l_d))$ in A, $\hat{l}_h = 1 - l_h$, by $\lceil 2^{i-1}/i \rceil$ d/i-tuples

$$(I_{1,1}, ..., I_{1,d/i}), ..., (I_{\lceil 2^{i-1}/i \rceil,1}, ..., I_{\lceil 2^{i-1}/i \rceil, d/i})$$

where $(k-1).i+1 \le h \le k.i$ (for some $k \in \{1, ..., d/i\}$), and for each m such that either $m < (k-1).i+1$ or $m > k.i$, $I_{1,\lceil m/i \rceil} = I_{2,\lceil m/i \rceil} = ... = I_{\lceil 2^{i-1}/i \rceil, \lceil m/i \rceil}$ is the interval containing the $\lceil m/i \rceil$-th dimensional component of all node labels, and $I_{1,\lceil h/i \rceil}, ..., I_{\lceil 2^{i-1}/i \rceil, \lceil h/i \rceil}$ are the $\lceil 2^{i-1}/i \rceil$ intervals associated at the node $(l_1, ..., l_h, ..., l_d)$ to the link $((l_1, ..., l_h, ..., l_d), (l_1, ..., \hat{l}_h, ..., l_d))$ in the $(\lceil 2^{i-1}/i \rceil, 1)$-MIRS for $H_i^{(j)}$, $1 \le j \le d/i$.

It is easy to verify that the described scheme correctly transmits messages via shortest paths. At each link the number of intervals is at most $\lceil 2^{i-1}/i \rceil$, hence it can be no worse than 2^{i-1} for each i. The dimension of the product cube H_d is clearly the sum of dimensions of all the subcubes, i.e. d/i. Following the proof of Lemma 1 we get a deadlock-free controller of size 2 working in the greedy mode for the optimal packet routing on H_d. □

Corollary 6. *There is a $((1, d), 2)$-DFMIRS on a d-dimensional hypercube.*

In Lemma 1 we proved that there exists a deadlock-free controller, for packet routing on a hypercube, which uses only two buffers in each node and allows messages to be routed via shortest paths. G. Tel [11] posed the question whether it is possible to obtain the set of the paths used by means of a (linear) interval routing scheme. We argue that there is no $(1, 2)$-DFLIRS (based on acyclic orientation controller) on a d-dimensional hypercube. (It is sufficient to show the nonexistence of $(1, 2) - DFLIRS$ on d-dimensional hypercubes for a small constant dimension.)

There exists an acyclic orientation covering of size $d + 1$ for the set of all shortest paths between all pairs of nodes in H_d. We show that the relevant lower bound is $\sqrt{d}$.

The d-dimensional hypercube has a node set consisting of all binary strings of length d with two nodes being connected if and only if they differ in exactly one bit. Thus every path in the hypercube corresponds to a sequence of changes of some bits. If the bits are changed in order from left to right then the path is called *monotone*.

Theorem 7. *Let $\mathcal{P}$ be a path system of a d-dimensional hypercube such that each path between any node v and its complement $\overline{v}$ in $\mathcal{P}$ is monotone. Every acyclic orientation covering for $\mathcal{P}$ has size at least $\sqrt{d}$.*

Proof. A movement of a message along the monotone path connecting a node v and its complement $\overline{v}$ can be simulated by a device consisting of a tape with d cells and a cursor which can be positioned either between any two neighboring cells or at the two ends of the tape. Initially the tape contains the string v and the cursor is on the left end of the tape. Moving a message along one link of the path corresponds to moving the cursor over one cell to the right and inverting the content of that cell. Reaching the destination is equivalent to reaching the right end of the tape. If we are given some acyclic orientation of the hypercube then we allow the cursor to advance only if the corresponding link is properly oriented in the current orientation.

If a sequence $\langle DG_1, ..., DG_s\rangle$ of acyclic orientations of the hypercube is an acyclic orientation covering for $\mathcal{P}$ then if we start the device on any node v and move the cursor according to $DG_1, ..., DG_s$ (in this order, using the greedy strategy) then the cursor reaches the right end of the tape.

Let us assume we shall start the device on all 2^d nodes simultaneously and consider the positions of cursors following the use of each acyclic orientation. An important observation is that for any acyclic orientation only few cursors can make long movements. For any positions of cursors $a, b \in \{0, ..., d\}$, $a < b$ and any acyclic orientation there are at most $2^d/(b-a+1)$ cursors that move between positions a and b in this orientation. For the sake of contradiction suppose that for some a, b there are more than $2^d/(b-a+1)$ cursors moving between positions a and b. From now on we consider only these cursors and their devices. For each device for each of the $b-a+1$ cursor positions between a and b the tape of the device has different contents. Therefore there must exist two devices that have the same tape content with both cursors between a and b. Let this content be $w_1w_2w_3$, the cursor of the first device being between w_1 and w_2 and the cursor of the second device being between w_2 and w_3. In this orientation the first device will move from $w_1|w_2w_3$ to $w_1\overline{w_2}|w_3$ and the second device moved from $w_1|\overline{w_2}w_3$ to $w_1w_2|w_3$. Therefore there is a cycle in the acyclic orientation between $w_1w_2w_3$ and $w_1\overline{w_2}w_3$ which is a contradiction.

Now we are ready to prove that after the i-th orientation at least $\left(1 - \frac{i}{\sqrt{d}}\right)2^d$ cursors are at most at position $i\sqrt{d}$. For $i = 0$ the claim holds since at the begining all cursors are at position 0. Let the claim holds after the i-th orientation. Based on the observation above at most $2^d/\sqrt{d}$ cursors can advance more than $\sqrt{d}$ positions to the right in the $i+1$-st orientation. Thus the claim holds also after the $i+1$-st orientation. Clearly the claim implies the theorem. □

In the 1-LIRS of the hypercube proposed in [1] every path between node and its complement is monotone. The consequence of the previous theorem is that this 1-LIRS is not suitable for the efficient deadlock-free packet routing (based on acyclic orientation covering).

One can observe that there exists a general deadlock-free controller of constant size covering the set of routing paths $\mathcal{P}$ from Theorem 7.

3.2 Tori

A d-dimensional torus $T_{n_1,\dots,n_d}$ is the cartesian product of d rings $R_1, \dots, R_d$, in which each R_i has n_i nodes.

Lemma 8. *There exists a deadlock-free controller of size 4 for the optimal packet routing on a d-dimensional torus.*

Proof. For simplicity, we will assume the case of 2 dimensions. The case of d dimensions is handled in a similar fashion. Fix an arbitrary node w of an $n \times m$ torus $T_{n,m}$. For simplicity, consider n, m even. Say $w = (n/2, m/2)$. Define the acyclic orientation covering $\mathcal{G} = \langle DT_1, DT_2, DT_1, DT_2 \rangle$ of a 2-dimensional tori $T_{n,m}$ such that in DT_1 the links are oriented from (i,j) to $(i+1,j)$ for $i = 1, 2, \dots, n/2-2, n/2, \dots, n$ and $1 \leq j \leq m$ and from (i,j) to $(i,j+1)$ for $1 \leq i \leq n$, $j = 1, 2, \dots, m/2-2, m/2, \dots, m$ and the links are oriented from $(n/2, j)$ to $(n/2-1, j)$ for $1 \leq j \leq m$ and from $(i, m/2)$ to $(i, m/2-1)$ for $1 \leq i \leq n$. In DT_2 all links are in opposite orientation. Edges $((n/2-1, j), (n/2, j))$ for $1 \leq j \leq m$ and $((i, m/2-1), (i, m/2))$ for $1 \leq i \leq n$ form row and column frontiers, respectively.

It is easy to verify that $\mathcal{G}$ forms a deadlock-free controller of size 4 for $T_{n,m}$. There is a collection of shortest paths between all pairs of nodes in $T_{n,m}$ that can be covered by $\mathcal{G}$. Given any two nodes u and v in $T_{n,m}$ with coordinates (i,j) and (k,l), respectively, there exists a shortest path from u to v that can be partitioned into four subpaths (where some of them may be empty) such that these subpaths are contained in coverings DT_1, DT_2, DT_1, DT_2, respectively. If the shortest path from u to v does not cross frontiers, the routing from u to v can be done using DT_1, DT_2. If the shortest path from u to v crosses one or two frontiers, the routing from u can reach frontiers using either DT_1 or DT_1, DT_2, then routing through frontiers can be performed with the next orientation in $\mathcal{G}$ and finally routing to v can be done with the next orientation in $\mathcal{G}$. □

The question remains whether it is possible to induce the set of paths achieved by deadlock-free controllers of size 4 by means of efficient interval routing schemes.

The next two results are from [4]. When we consider linear interval routing schemes, the size $2d+1$ can be obtained with the compactness 2, and the restriction to the size 5 can be achieved with the compactness $O(n^{d-1})$.

Lemma 9. *There exists a $(2, 2d+1)$-DFLIRS for a d-dimensional torus.*

Lemma 10. *For every n and i $(1 < i < d)$ there exists a $(\lceil n^i/2\rceil, 2.\lceil d/i\rceil + 1)$-DFLIRS on a d-dimensional torus.*

On the other hand, when using d-dimensional interval routing schemes (see [5]) the size 4 can be achieved with compactness of only 1.

Theorem 11. *For every n and i $(1 \le i \le d)$ there exists a $((n^{i-1}, \lceil d/i\rceil), 4) - DFMIRS$ on a d-dimensional torus.*

Proof. Consider a d-dimensional torus, given as the product of $\lfloor d/i \rfloor$ subtori of dimension i and a subtorus of dimension $d \ mod \ i$. For simplicity, assume $d \ mod \ i \ = \ 0$. Observe that each of these d/i subtori of dimension i admits $(n^{i-1}, 1)$-MIRS. Now, the proof follows in a similar way as the proof of Theorem 5 for hypercubes. Following the proof of Lemma 8 we get a deadlock-free controller of size 4 working in the greedy mode for the optimal packet routing on d-dimensional tori, based on $(n^{i-1}, \lceil d/i\rceil)$-MIRS. □

Corollary 12. *There exists a $((1, d), 4) - DFMIRS$ on a d-dimensional torus.*

3.3 Butterflies

The d-dimensional butterfly network (BF_d for short) has $(d+1).2^d$ nodes and $d.2^{d+1}$ links. The nodes correspond to pairs (α, p), where $p \in \{0, ..., d\}$ is the position of the node and α is a d-bit binary number. The two nodes (α, p) and (α', p') are connected by a link if and only if $p' = p+1$ and either α and α' are identical or α and α' differ only in the p'th bit.

Lemma 13. *There exists a deadlock-free controller of size 4 for the optimal packet routing on a d-dimensional butterfly.*

Proof. Let $u = (a_{d-1}...a_0, p)$ and $v = (b_{d-1}...b_0, q)$, $p \ge q$, be two nodes in BF_d. The distance $d(u, v)$ is $d(u, v) = p - q$ if $a_i = b_i$ for $i = 0, 1, ..., d-1$ and $d(u, v) = r_{max} - r_{min} + |p - r_{max}| + |q - r_{min}| + c$, where $c = 0$ for $p > r_{max}$, otherwise $c = 2$, and where $r_{max} = max\{i \mid a_i \ne b_i, \ 0 \le i \le d-1\}$ and $r_{min} = min\{i \mid a_i \ne b_i, \ 0 \le i \le d-1\}$. In order to reach the length $d(u, v)$, take the shortest path in BF_d from $(b_{d-1}...b_0, q)$ to $(b_{d-1}...b_0, r_{min})$ following $|r_{min} - q|$ links, then the shortest path from $(b_{d-1}...b_0, r_{min})$ to $(a_{d-1}...a_0, r_{max}+1)$ following $|r_{max} - r_{min}|$ links and finally the shortest path from $(a_{d-1}...a_0, r_{max}+1)$ to $(a_{d-1}...a_0, p)$ following $|r_{max} - p| - 1$ links when $p > r_{max}$, and $|r_{max} - p| + 1$ links otherwise.

Now, it is easy to verify that each shortest path between two nodes in BF_d can be partitioned into three subpaths such that each subpath either continuously increases or decreases the position parameter p. Hence, there is a deadlock-free controller of size 4 on BF_d. □

It was shown in [7] that there does not exist efficient IRS for d-dimensional butterflies.

Lemma 14. *Each optimal k-IRS for a d-dimensional butterfly needs* $k = \Omega(2^{d/2})$ *intervals.*

However, there are efficient MIRS on d-dimensional butterflies, with deadlock-free controllers of size only 4.

Lemma 15. *There is a* $((2,3),4)$*-DFMIRS on a d-dimensional butterfly.*

Proof. Consider the following machine. It has a working tape $[a_1a_2...a_i|a_{i+1}...a_d]$ with d cells containing bits "a_i" and a head "$|$" which can be positioned between cells or at any end of the tape. In one step the head moves to the left or to the right and writes 0 or 1 to the cell over which it has passed. The graph with vertices corresponding to states of this machine and links corresponding to steps is exactly the d-dimensional butterfly graph. This allows us to consider nodes of BF_d to be the states of the machine described.

Given BF_d and a node w of the form $[u\alpha|v]$, then there exist shortest paths from w to the nodes

- of the form (A): $[\beta_1|\beta_2]$, $|\beta_1| \geq |u\alpha|$, $u\alpha$ is not a prefix of β_1
- or of the form (B): $[\beta_1|\beta_2 0v]$

starting with the link e corresponding to moving the head to the left and writing 0. These are the only such nodes in BF_d.

We have to show that for BF_d there exists a $(2,3)$-MIRS. Let us label the nodes in the three dimensions as follows:

- 1st dimension: The number written on the tape.
- 2nd dimension: The number written on the tape read backwards.
- 3rd dimension: The position of the head.

For each node w and a link e described above it is possible to select the nodes of the form (A) and (B) using two triples of intervals. The first triple selects the nodes not starting with $u\alpha$ (these form a cyclic interval in the 1st dimension) and not having the head to the left of w's head (these form a cyclic interval in the 3rd dimension). The second triple selects the nodes ending with $0v$ (these form a cyclic interval in the 2nd dimension) and having the head to the left of w's head. The construction is similar for the other types of links.

Define $\mathcal{G} = \langle DBF_1, DBF_2, DBF_1, DBF_2\rangle$, where in DBF_1 the orientation of links is from $[a_1...a_i|a_{i+1}...a_d]$ to $[a_1...a_{i+1}|a_{i+2}...a_d]$ and from $[a_1...a_i|a_{i+1}...a_d]$ to $[a_1...\hat{a}_{i+1}|a_{i+2}...a_d]$ for $0 \leq i \leq d-1$ (where the head in position 0 means $[|a_1...a_d]$ and $\hat{a} = 1 - a$) and in DBF_2 the orientation is opposite. It is easy to verify that each shortest path induced by a $((2,3),4) - MIRS$ on BF_d can be covered by $\mathcal{G}$. □

3.4 Cube connected cycles

Let $u = (a_0...a_{d-1}, p)$ be a tuple consisting of a binary string and a cursor position from $\{0, ..., d-1\}$. The operations of shifting cursor cyclically to the left and to the right on u are denoted as $L(u)$ and $R(u)$, respectively, and the shuffle operation is defined as $S(u) = (a_0 ... \hat{a}_p ... a_{d-1}, p)$, where $\hat{a}_p = 1 - a_p$.

A d-dimensional cube connected cycles (denoted as CCC_d) is a network (V, A), where $V = \{u \mid u \in \{0,1\}^d \times \{0, ..., d-1\}\}$ and $A = \{(u, v) \mid R(u) = v \text{ or } L(u) = v \text{ or } S(u) = v\}$.

Lemma 16. *There exists an acyclic orientation covering of size $2d + 6$ for the system of all shortest paths between all pairs of nodes in CCC_d.*

Proof. Consider the following acyclic orientation DC_1: for each binary string $\alpha = a_0 ... a_{d-1}$ the cycle $(\alpha, 0), ..., (\alpha, d-1)$ is oriented $(\alpha, 0) \to ... \to (\alpha, d-1)$ and $(\alpha, 0) \to (\alpha, d-1)$; the remaining links are oriented arbitrarily provided that the resulting orientation is acyclic. The covering $\mathcal{G}$ consists of an alternating sequence of DC_1 and its opposite DC_2 of length $2d + 6$.

Consider an arbitrary shortest path $\pi = (\alpha_0, p_0), ..., (\alpha_k, p_k)$. It clearly contains at most d S-links (such that $p_i = p_{i+1}$). By cycle segment we mean maximal subpath of π that contains no S-link. If a cycle segment does not contain a link $(\alpha, 0), (\alpha, d-1)$ for some α then the entire segment is covered either by DC_1 or by DC_2. Call this segment as non-zero segment. Each zero segment consists of at most three paths such that each of them is covered either by DC_1 or by DC_2.

Because each shortest path contains at most two vertices $(\alpha_1, p), (\alpha_2, p)$ with the same cursor position p, there are at most two zero segments.

Thus π consists of at most $2d+5$ parts (i.e. d S-links, $d-1$ non-zero segments and two zero segments each of three paths) all of which are covered either by DC_1 or by DC_2. Hence π is covered by $\mathcal{G}$. □

Corollary 17. *There exists a deadlock-free controller of size $2d + 6$ for the optimal packet routing on a d-dimensional cube connected cycles network.*

It was shown in [7] that there does not exist an efficient shortest paths IRS for CCC_d (superpolynomial compactness in d is required !).

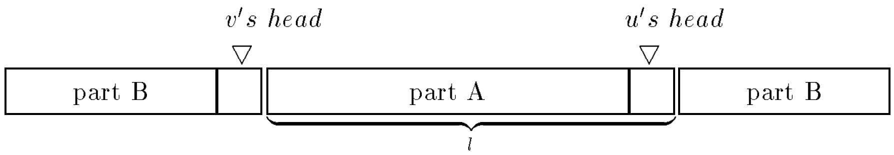
Lemma 18. *Each optimal k-IRS for a d-dimensional cube connected cycles network needs $k = \Omega(2^{d/2})$.*

Now we show that there are efficient d-dimensional IRS on CCC_d with compactness and size polynomial in d.

Theorem 19. *There exists a $((2d^3, d), 2d + 6) - DFMIRS$ on CCC_d.*

Proof. Let us define a machine whose state diagram is the d-dimensional cube-connected-cycles graph. Its working tape is a *circular* strip consisting of d cells. The head can be positioned above any cell. Each cell can contain one binary digit. In one step the head can change the content of the cell read or move one position to the left or to the right. Again we consider nodes being the states of the machine described.

Let u, v be two nodes of the CCC_d. Take $u\ XOR\ v$ (the tape is unwinded on the picture):



Denote a, b and a' the lengths of the longest runs of consecutive zeros in parts A, B and $A'(= A$ *without the rightmost cell*) respectively and b' the length of the run of consecutive zeros in part B starting immediately to the right of the position of u's head. There exists a shortest path from u to v starting with the left arc e if and only if either:

A: $a' = a$ and $2(l + b - a) \leq d$
or
B: $b' = b$ and $2(l + b - a) \geq d$ and u, v do not differ in the cell scanned by u's head.

The condition for the existence of a shortest path starting with the right arc is symmetric. There exists a shortest path from u to v starting with the shuffle arc if and only if u and v differ in the cell scanned by u's head. Now we briefly describe the $(2d^3, d) - MIRS$ of CCC_d.

The vertices in the i-th dimension ($i \in \{1, ..., d\}$) have numbers $1, ..., d$ according to the following lexicographic ordering:

- the first criterion is the position of the head
- the second criterion is the number written on the tape after the cyclic rotation by i bits to the left

In this labeling the vertices having the same position of the head form a block in each dimension. Another important property of the labeling is that selecting vertices having the head at any given position and containing (resp. not containing) any given binary substring at any given position of the tape can be done using at most two intervals in one block of *one* dimension. The dimension in which intervals are used is determined by the position of the substring.

Let u be any vertex of the CCC_d graph. Labeling the shuffle arc emanating from u is easy, as exactly messages to the vertices having a different symbol at the position of $u's$ head are to be routed along it. As there exists a dimension such that in each of its blocks such vertices form a cyclic interval, we need only d intervals per dimension.

Labeling the left arc is more complicated. We select vertices whose messages are to be routed along this arc for each position of their head independently. If for each given position we need at most q intervals per dimension to select such vertices then in total we need at most dq intervals per dimension.

Vertices satisfying the rule A and having the head at a given position are to be selected as follows:

- We choose the length a' of the longest run of consecutive zeros in the part A' of $u\ XOR\ v$ ($len(A')+1$ possibilities)
- We choose the position of this run ($len(A') - a' + 1$ possibilities)
- Given a' and the position of the run, vertices
  - having run of a' zeros at the choosen position
  - not having longer run of zeros in the part A
  - not having run of zeros in the part B longer than $a + \frac{d-2l}{2}$

  can be selected using two intervals per dimension, because we can fulfill these conditions by selecting the vertices having, or not having certain substrings at different positions.

Vertices satisfying the rule B and having the head at a given position are to be selected as follows:

- We choose the length b' of the run of consecutive zeros in the part B starting immediately to the right of the position of u's head. ($len(B)+1$ possibilities)
- Given b', vertices
  - having run of b' zeros in the part B starting immediately to the right of the position of u's head
  - not having longer run of zeros in the part B
  - not having run of zeros in the part A longer than $b + \frac{2l-d}{2}$
  - not differing from u in the cell scanned by u's head

  can be selected using two intervals per dimension, using the same reasoning as in the previous case.

It holds $(len(A')+1)(len(A')+1) + len(B) + 1 \leq d^2$, therefore we have used in total at most $2d^3$ intervals per dimension which gives us the $(2d^3, d) - MIRS$. □

4 Conclusions

We have presented efficient deadlock-free MIRSs on hypercubes, tori, butterflies and cube connected cycles. These results can be transformed also to an analogous wormhole routing model (as formulated in [4]). The main question remains whether there are efficient deadlock-free MIRS also for wider classes of graphs, e.g. vertex symmetric graphs, planar graphs etc.

We have also presented a nonconstant lower bound on the size of deadlock-free controllers (based on acyclic orientation covering) for a special set of routing paths in d-dimensional hypercubes. This is the first nontrivial lower bound on specific controllers. Moreover, this set of routing paths can be covered by general deadlock-free controllers of constant size, thus giving the first example of differences between sizes of general and specific controllers. The question is to determine nonconstant lower bounds on the size of deadlock-free controllers for general networks and to give size differences between general and specific deadlock-free controllers.

There are still many unresolved questions concerning DFMIRS (some of them are mentioned in Section 3). It would be nice to have a trade-off between compactness and size for deadlock-free MIRS on general graphs.

References

1. E. Bakker, J. van Leeuwen, R.B. Tan: *Linear interval routing schemes.* Algorithms Review 2, 1991, pp. 45–61.
2. R. Cypher, L. Gravano: *Requirements for Deadlock-Free, Adaptive Packet Routing.* In 14th Annual ACM Symposium on Principles of Distributed Computing (PODC), 1992, pp. 25–33.
3. W.J. Dally, C.L. Seitz: *Deadlock-free message routing in multiprocessor interconnection networks.* IEEE Transactions on Computers, C-36, 1987, pp. 547–553.
4. M. Flammini: *Deadlock-Free Interval Routing Schemes.* In 14th Annual Symposium on Theoretical Aspects of Computer Science (STACS), Lecture Notes in Computer Science 1200, Springer-Verlag, 1997, pp. 351–362.
5. M. Flammini, G. Gambosi, U. Nanni, R. Tan: *Multi-Dimensional Interval Routing Schemes.* In 9th International Workshop on Distributed Algorithms (WDAG), Lecture Notes in Computer Science, Springer-Verlag, 1995. To appear in Theoretical Computer Science.
6. C. Gavoille: *A Survey on Interval Routing Schemes.* Research Report RR-1182-97, LaBRI, Université Bordeaux I, October 1997. Submitted for publication.
7. R. Kráľovič, P. Ružička, D. Štefankovič: *The Complexity of Shortest Path and Dilation Bounded Interval Routing.* In 3rd International Euro-Par Conference, Lecture Notes in Computer Science 1300, Springer-Verlag, August 1997, pp. 258–265. Full version will appear in Theoretical Computer Science.
8. P.M. Merlin, P.J. Schweitzer: *Deadlock avoidance in store-and-forward networks.* IEEE Transactions of Communications, COM-27, 1980, pp. 345–360.
9. P. Ružička, D. Štefankovič: *On the Complexity of Multi-Dimensional Interval Routing Schemes.* Submitted for publication.
10. I. Sakko, L. Mugwaneza, Y. Langue: *Routing with compact tables.* Applications in Parallel and Distributed Computing (APDC), North-Holland, 1994.
11. G. Tel: *Introduction to Distributed Algorithms. (Chapter 5: Deadlock-free Packet Routing).* Cambridge University Press, Cambridge, U.K., 1994.
12. S. Toueg: *Deadlock-free and livelock-free packet switching networks.* In Proceedings of Symposium on Theory of Computing (STOCS), 1980, pp. 94–99.
13. S. Toueg, J. Ullman: *Deadlock-free packet switching networks.* SIAM Journal of Computing 10, 1981, pp. 594–611.
14. J. van Leeuwen, R.B. Tan: *Compact routing methods: A survey.* In 1st International Colloquium on Structural Information and Communication Complexity (SIROCCO), Carleton Press, 1994, pp. 99–110.
15. J. Vounckx, G. Deconinck, R. Lauwereins, J.A. Peperstraete: *Deadlock-free Fault Tolerant Wormhole Routing in Mesh-Based Massively Parallel Systems.* In Technical Committee on Computer Architecture (TCCA) Newsletter, IEEE Computer Society, Summer-Fall issue, 1994, pp. 49–54.
16. J. Vounckx, G. Deconinck, R. Cuyvers, R. Lauwereins: *Minimal Deadlock-free Compact Routing in Wormhole-Switching based Injured Meshes.* In Proceedings of the 2nd Reconfigurable Architectures Workshop, Santa Barbora, Ca, 1995.