Using Banyan Networks for Load-Balanced Switches with Incremental Update

Ching-Min Lien, Cheng-Shang Chang, Jay Cheng, Duan-Shin Lee and Jou-Ting Liao
Institute of Communications Engineering
National Tsing Hua University
Hsinchu 300, Taiwan, R.O.C.
E-mail: keiichi@gibbs.ee.nthu.edu.tw; cschang@ee.nthu.edu.tw;
jcheng@ee.nthu.edu.tw; lds@cs.nthu.edu.tw; jtliao@gibbs.ee.nthu.edu.tw

***Abstract*—Load-balanced switches have received a lot of attention lately as they are much more scalable than other existing switch architectures in the literature. One of the most salient features of load-balanced switches is its simplicity of implementing deterministic and periodic connection patterns for the switch fabrics. In particular, for an $N \times N$ load-balanced switch, its switch fabric only needs an $N \times N$ *rotator* that is capable of realizing all the powers of the circular shift permutation. In this paper, we consider the problem of incremental update of the number of linecards in load-balanced switches. For this, our idea is to consider a $2^M \times 2^M$ *degenerated* banyan network that only uses half of the 2^{M+1} inputs/outputs in the classical $2^{M+1} \times 2^{M+1}$ banyan network. We show how one can use the $2^M \times 2^M$ degenerated banyan network as a $p \times p$ rotator for any $2 \leq p \leq 2^M$. This is done by a specific rule of placing the p linecards in the 2^M input/output ports of the $2^M \times 2^M$ degenerated banyan network. In special, when $p = 2^M$, the $2^M \times 2^M$ degenerated banyan network can also be used as a *crosstalk-free* $2^M \times 2^M$ rotator, where all the routing paths do not share a common node. As such, one can use a $2^{M+1} \times 2^{M+1}$ banyan network as the switch fabric for a $2^M \times 2^M$ load-balanced switch that is capable of providing incremental update of the number of linecards.**

I. Introduction

Load-balanced switches (see e.g., [2], [8], [7], [6], [3]) have received a great deal of attention recently as they are much more scalable than other existing switch architectures in the literature. A typical load-balanced switch (see Fig. 1) consists of two stages: the first stage is for load-balancing that converts incoming traffic into the uniform traffic, and the second stage is for switching of the uniform traffic. By so doing, it was shown in [2] that such a switch architecture indeed provides 100% throughput (under a mild technical condition for the incoming traffic).

One of the most salient features of load-balanced switches is that the connection patterns in the switch fabrics of both stages are *deterministic* and *periodic*. Specifically, for an $N \times N$ load-balanced switch, its switch fabrics only need to realize in every period of N time slots any N $N \times N$ permutation matrices $P_1, P_2, \ldots, P_N$ that satisfy

$$P_1 + P_2 + \ldots + P_N = \mathbf{e}, \qquad (1)$$

where $\mathbf{e}$ is the $N \times N$ matrix with all its elements being 1. In the literature, there are two well-known conditionally nonblocking switches, the $N \times N$ rotator (that implements all the powers of the circular shift matrix) and the $N \times N$ symmetric TDM switch, that can be used for generating the needed connection patterns in (1).

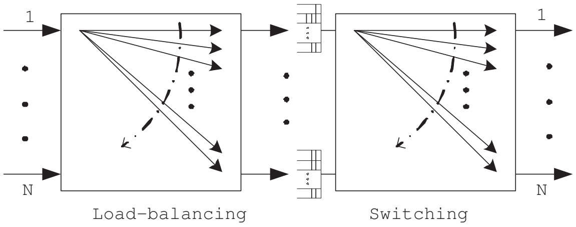

Fig. 1. The generic load-balanced switch architecture.

Keslassy, Chung, and McKeown [7] studied the problem of incremental update of the number of linecards in load-balanced switches. To solve the incremental update problem for an $N \times N$ load-balanced switch, it is required that the switch is capable of emulating any $p \times p$ load-balanced switch for $2 \leq p \leq N$, where p is the number of linecards placed in the $N \times N$ load-balanced switch. Instead of using the three-stage approach in [7], we are interested in finding a "universal" $N \times N$ switch (fabric) that solves the incremental update problem. Clearly, either an $N \times N$ crossbar switch or an $N \times N$ Benes network [1] can be used as the universal $N \times N$ switch because they both are (rearrangeable) nonblocking switches that can realize all the $N!$ permutations. However, the construction complexity of an $N \times N$ crossbar switch (in terms of the number of crosspoints) is $O(N^2)$ which does not scale well and thus prevents it from being used for large N. On the other hand, the lack of self-routing property in the $N \times N$ Benes network makes it difficult to find routing paths for realizing an $N \times N$ permutation. In our recent paper [11], we proposed a new class of multistage interconnection networks, called *twister networks*. We showed that twister networks possess the self-routing property and they can be used as the universal switch for the incremental update problem in load-balanced switches.

In the paper, we go one step further by showing that the classical banyan networks (see e.g., the books [14], [9], [4]) can also be used as the universal switch for the incremental update problem in load-balanced switches. Our idea is to consider a $2^M \times 2^M$ *degenerated* banyan network that is obtained from the classical $2^{M+1} \times 2^{M+1}$ banyan network by using only half of the 2^{M+1} inputs/outputs. We show how one can use the $2^M \times 2^M$ degenerated banyan network as a $p \times p$ rotator (and a $p \times p$ symmetric TDM switch) for any $2 \le p \le 2^M$. This is done by a specific rule of placing the p linecards in the 2^M input/output ports of the degenerated $2^M \times 2^M$ banyan network. In particular, when $p = 2^M$, the $2^M \times 2^M$ degenerated banyan network can also be used as a *crosstalk-free* $2^M \times 2^M$ rotator (and $p \times p$ symmetric TDM switch), where all the routing paths do not share a common node. As such, one can use the classical $2^{M+1} \times 2^{M+1}$ banyan network as the universal switch for a $2^M \times 2^M$ load-balanced switch that is capable of providing incremental update of the number of linecards.

This paper is organized as follows. In Section II, we introduce the degenerated banyan networks and prove their conditionally nonblocking properties. In Section III, we show how one can use degenerated banyan networks as rotators and symmetric TDM switches. The paper is then concluded in Section IV, where we address possible extensions of our work.

II. Degenerated Banyan Network

In this section, we introduce degenerated banyan networks. For $0 \le x \le 2^M - 1$, we denote the binary representation of x as the M-vector $(I_1(x), I_2(x), \ldots, I_M(x))$, where $I_k(x)$ is the k^{th} least significant bit of x. Note from the binary representation that

$$x = \sum_{k=1}^{M} I_k(x) 2^{k-1}.$$

Definition 1 ***(Degenerated Banyan Network)*** *Suppose that $N = 2^M$. An $N \times N$* degenerated *banyan network consists of $M+1$ stages with N nodes in each stage. Index the $M+1$ stages from 0 to M, and the N nodes at each stage from 0 to $N-1$. The N nodes at the 0^{th} (resp., $(M+1)^{th}$) stage are called the input (resp., output) nodes. For $k = 1, 2, \ldots, M$, $j = 0, 1, \ldots, N-1$, the j^{th} node at the $(k-1)^{th}$ stage is connected to the* two *nodes at the k^{th} stage whose M-bit binary representations can only differ from the M-bit binary representation of j in the k^{th} most significant bit.*

In Fig. 2, we show a 4×4 degenerated banyan network. In such a network, there are three stages with four nodes in each stage. Node 1 at the 0^{th} stage has the 2-bit binary representation $(1, 0)$. It is connected to node 1 at the 1^{st} stage with the 2-bit binary representation $(1, 0)$ and to node 3 at the 1^{st} stage with the 2-bit binary representation $(1, 1)$. We note that an $N \times N$ degenerated banyan network is in fact a classical $2N \times 2N$ banyan network with only half of the inputs and outputs. To illustrate this, we show a 8×8 classical banyan network in Fig. 3, where each node (switch) consists of two input links and two output links. By using only the first N input links at the 0^{th} stage and all the output links with even indices at the last stage, a classical $2N \times 2N$ banyan network reduces to the $N \times N$ degenerated banyan network (with all the output links are relabeled in the ascending order as shown in Fig. 2).

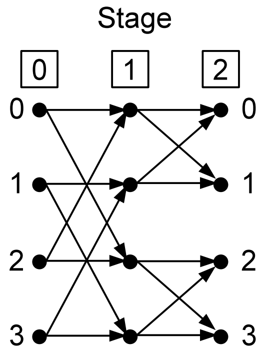

Fig. 2. A 4×4 degenerated banyan network.

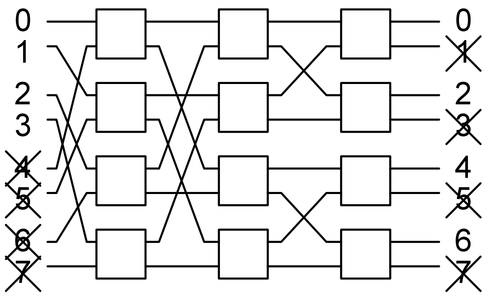

Fig. 3. An 8×8 banyan network.

A. Routing Path

In a $2^M \times 2^M$ degenerated banyan network, a routing path from an input node i to an output node o can be simply described by the $(M+1)$-vector $\mathbf{v} = (v_0, v_1, \cdots, v_M)$, where v_j is the index of the node traversed at the j^{th} stage for all $0 \le j \le M$ with $v_0 = i$ and $v_M = o$.

In the following definition, we specify a unique routing path in a $2^M \times 2^M$ degenerated banyan network for each pair of input/output nodes that will be used in this paper.

Definition 2 *(Unique Routing Path) Consider an $N \times N$ degenerated banyan network with $N = 2^M$ and an input/output pair (i, o). Let $(I_1(i), I_2(i), \ldots, I_M(i))$ and $(I_1(o), I_2(o), \ldots, I_M(o))$ be the binary representations for input node i and output node o, respectively. Then the routing path from input node i to output node o is represented by the $(M+1)$-vector $\mathbf{v} = (v_0, v_1, \ldots, v_M)$, where the binary representation of v_j is*

$$I_k(v_j) = \begin{cases} I_k(i), & \text{for } 1 \le k \le M - j, \\ I_k(o), & \text{for } M - j + 1 \le k \le M. \end{cases} \qquad (2)$$

Note that from (2), we have

$$\begin{aligned}&(I_1(v_{j-1}), I_2(v_{j-1}), \ldots, I_M(v_{j-1}))\\ &= (I_1(i), \ldots, I_{M-j}(i), I_{M-j+1}(i), I_{M-j+2}(o), \ldots, I_M(o)),\\ &(I_1(v_j), I_2(v_j), \ldots, I_M(v_j))\\ &= (I_1(i), \ldots, I_{M-j}(i), I_{M-j+1}(o), I_{M-j+2}(o), \ldots, I_M(o)),\end{aligned}$$

for $j = 1, 2, \ldots, M$. It is clear that the M-bit binary representations of nodes v_{j-1} and v_j can only differ in the j^{th} most significant bit for all $j = 1, 2, \ldots, M$, and it then follows from Definition 1 that the routing path specified in (2) is a feasible path in the degenerated banyan network. Furthermore, as $v_0 = \sum_{k=1}^{M} I_k(i)2^{k-1} = i$ and $v_M = \sum_{k=1}^{M} I_k(o)2^{k-1} = o$, it is a feasible path from input node i to output node o.

We note that the routing path defined in (2) is exactly the same as the routing path in the classical banyan network, where the k^{th} *most* significant bit is swaped from the binary presentation of the input node to that of the output node at the k^{th} stage.

B. Conditionally Nonblocking Properties

Like a switch or a switching network, a connection matrix for a $2^M \times 2^M$ degenerated banyan network is a $2^M \times 2^M$ sub-permutation matrix that specifies the connections from a subset of its 2^M input nodes to a subset of its 2^M output nodes. The routing paths specified by a connection matrix is said to be *link-disjoint* (resp., *node-disjoint*) if all the routing paths from the input nodes to the output nodes specified by the connection matrix do not share a common link (resp., node). In the switching context, one can view each node as a switch and a connection matrix with link-disjoint routing paths is said to be *realizable* (or *feasible*) as there are no conflicting links. On the other hand, a connection matrix with node-disjoint routing paths is sometimes said to be *crosstalk-free* as the crosstalk problem can be alleviated by allowing only one active input link in each switch (see e.g., [13], [15], [12], [5] and references therein). Clearly, a $2^M \times 2^M$ degenerated banyan network cannot realize all the $2^M \times 2^M$ sub-permutation matrices. As we will show later, it can realize a set of sub-permutation matrices that satisfy a certain condition. Such a property is called the *conditionally nonblocking property* in the literature [9], [10], and thus degenerated banyan networks can be used as conditionally nonblocking switches.

For the proof of the conditionally nonblocking property, we first introduce the *N-modulo distance* as defined in [11].

Definition 3 *The N-modulo distance $d_N(i,j)$ between two integers i and j is defined as*

$$d_N(i,j) = \min\left[(i-j) \textit{ mod } N, (j-i) \textit{ mod } N\right]. \tag{3}$$

The distance can be alternatively defined as

$$d_N(i,j) = \min\left[|i-j| \bmod N, -|i-j| \bmod N\right]. \tag{4}$$

One can easily see that the two definitions above are equivalent. In the special case that $0 \le i,j \le N-1$, (4) can be rewritten as

$$d_N(i,j) = \min[|i-j|, N-|i-j|]. \tag{5}$$

As discussed in [11], the *N-modulo distance* $d_N(i,j)$ is simply the length of the shorter arc between nodes i and j on the circle of circumference N when we place all the N input nodes (and output nodes) on a circle.

The following properties for the N-modulo distance are shown in [11]. To simplify the notation, we say that $i =_N j$ if i mod N equals j mod N.

Property 4 *Let i, j and k be all integers.*

(i) *(Nonnegativity)* $d_N(i,j) \ge 0$.
(ii) $d_N(i,j) = 0$ *if and only if* $i =_N j$.
(iii) *(Symmetry)* $d_N(i,j) = d_N(j,i)$.
(iv) *(Triangle Inequality)* $d_N(i,j) \le d_N(i,k) + d_N(k,j)$.
(v) *(Translation Invariance)* $d_N(i,j) = d_N(i+k, j+k)$.
(vi) $d_N(i,j) = d_N(-i,-j)$.
(vii) $d_N(i,j) = d_N(i,k)$ *if* $j =_N k$.
(viii) $d_N(i,j) = d_N(k,\ell)$ *if* $i+k =_N j+\ell$.

In the following theorem, we show conditionally nonblocking properties for a degenerated banyan network. The proof of Theorem 5 is given in the Appendix.

Theorem 5 *Consider an $N \times N$ degenerated banyan network with $N = 2^M$.*

(i) *If the connection matrix has the property that*

$$d_N(i_1,i_2) \le |o_1 - o_2| \tag{6}$$

for arbitrary two input/output pairs (i_1,o_1) and (i_2,o_2), then the routing paths are node-disjoint.

(ii) *If the connection matrix has the property that*

$$d_N(i_1,i_2) \le 2|o_1 - o_2| + 1 \tag{7}$$

for arbitrary two input/output pairs (i_1,o_1) and (i_2,o_2), then the routing paths are link-disjoint.

In comparison with twister networks in [11], we note that the condition in (7) is weaker than $d_N(i_1,i_2) \le 2d_N(o_1,o_2)$ in Theorem 11 (with $\gamma = 2$) of [11]. Thus, degenerated banyan networks can realize a richer class of connection matrices than those specified in Theorem 11 (with $\gamma = 2$) of [11]. This includes rotators and symmetric TDM switches discussed in the next section.

Banyan networks have been studied extensively in the literature (see e.g., the books [14], [9], [4]). There are also many conditionally nonblocking properties for banyan networks. In particular, the condition in (6) is known as a *circular expander* in [9], and all the routing paths in a banyan network are link-disjoint under (6). By using only half of the input/outputs in banyan networks, degenerated banyan networks go one step further by ensuring all the routing paths to be node-disjoint under (6). Moreover, the condition for link-disjoint routing paths in degenerated banyan networks is relaxed to (7).

III. Rotator and Symmetric TDM switch

As mentioned in the Section I, our objective is to show that a degenerated banyan network can be used as a universal switch (fabric) that can provide incremental update of the number of linecards in a two-stage load-balanced switch. For this, we formally define two kinds of switches that are commonly used as the switch fabrics for load-balanced switches. First, an $N \times N$ permutation matrix $P = (p_{ij})$ is called a circular shift matrix if

$$p_{ij} = \begin{cases} 1, & \text{if } j = (i+1) \bmod N, \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

where $0 \leq i, j \leq N-1$.

Definition 6 ***(Rotator)*** *Let P be the $N \times N$ circular shift matrix as defined in (8). An $N \times N$ switch (or switching network) is called a rotator if it can realize the N permutations, P^n, $n = 0, 1, 2, \ldots, N-1$.*

Note that each input/output pair (i, o) for the permutation matrix P^n satisfies $o =_N (i+n)$.

Definition 7 ***(Symmetric TDM Switch)*** *For $0 \leq n \leq N-1$, let $\tilde{P}_n$ be the permutation such that each input/output pair (i, o) satisfies $(i+o) =_N n$. An $N \times N$ switch (or switching network) is called a symmetric TDM switch if it can realize the N permutations, $\tilde{P}_n$, $n = 0, 1, 2, \ldots, N-1$.*

It is easy to see that the condition in (1) is satisfied for the N permutations in an $N \times N$ rotator and the N permutations in an $N \times N$ symmetric TDM switch. As such, they both can be used as the switch fabric for load-balanced switches.

Theorem 8 *Consider an $N \times N$ degenerated banyan network with $N = 2^M$. Then all the routing paths specified by any of the N permutation matrices in an $N \times N$ rotator (resp., symmetric TDM switch) are* node-disjoint. *As such, an $N \times N$ degenerated banyan network with $N = 2^M$ can be used as a* crosstalk-free *$N \times N$ rotator (resp., symmetric TDM switch)*

Proof: As each input/output pair (i, o) for the permutation matrix P^n in an $N \times N$ rotator satisfies $o =_N (i+n)$, we have for any two input/output pairs (i_1, o_1) and (i_2, o_2) that

$$o_1 - o_2 =_N i_1 - i_2.$$

Thus, we have

$$o_1 + i_2 =_N o_2 + i_1.$$

From Property 4 (viii) and (iii), it follows that

$$d_N(o_1, o_2) = d_N(i_2, i_1) = d_N(i_1, i_2).$$

As $d_N(o_1, o_2) \leq |o_1 - o_2|$ in (5), we then have

$$d_N(i_1, i_2) \leq |o_1 - o_2|,$$

and the result then follows from Theorem 5 (i).

For the N permutation matrices specified by an $N \times N$ symmetric TDM switch, we also have for any two input/output pairs (i_1, o_1) and (i_2, o_2) that

$$o_1 + i_1 =_N o_2 + i_2.$$

From Property 4 (viii), it follows that

$$d_N(o_1, o_2) = d_N(i_1, i_2).$$

The rest of the proof then follows the same argument for a crosstalk-free $N \times N$ rotator. ■

As commented immediately after Theorem 5, degenerated banyan networks can realize a richer class of connection matrices than those specified in Theorem 11 (with $\gamma = 2$) of [11] for twister networks. In [11], it was shown that an $2^M \times 2^M$ twister network can be used as a $p \times p$ rotator and a $p \times p$ symmetric TDM switch provided that the p linecards are placed as *evenly* as possible. Specifically, for any $2 \leq p \leq 2^M$, there exists $0 \leq m \leq M-1$ such that $2^m < p \leq 2^{m+1}$, and p can be written as $2^m + \ell$, where $1 \leq \ell \leq 2^m$. For $i = 0, 1, 2, \ldots, p-1$, place the i^{th} linecard in the $f(i)^{th}$ input/output port, where

$$f(i) = \begin{cases} i \cdot 2^{M-m-1}, & \text{for } 0 \leq i \leq 2\ell - 1, \\ (i - \ell) \cdot 2^{M-m}, & \text{for } 2\ell \leq i \leq p-1. \end{cases} \tag{9}$$

Here we adopt the same placement rule for the p linecards in a $2^M \times 2^M$ degenerate banyan network. Note that the gap between the placement of two consecutive linecards is either 2^{M-m-1} or 2^{M-m}. Thus, the maximum gap between two consecutive linecards is exactly twice of the minimum gap between two consecutive linecards. Following the same argument as in Theorem 14 of [11], one can show (the detailed proof is omitted here due to space limitation) that for any two input/output pairs (i_1, o_1) and (i_2, o_2) specified by any $p \times p$ permutation of a $p \times p$ rotator, the corresponding input/output pairs $(f(i_1), f(o_1))$ and $(f(i_2), f(o_2))$ in the $2^M \times 2^M$ degenerate banyan network satisfies

$$d_N(f(i_1), f(i_2)) \leq 2 d_N(f(o_1), f(o_2)).$$

As $d_N(f(o_1), f(o_2)) \leq |f(o_1) - f(o_2)|$, the condition in (7) is satisfied and the $2^M \times 2^M$ degenerate banyan network can be used as a $p \times p$ rotator. The argument for a $p \times p$ symmetric TDM switch is similar. The result is stated in the following theorem.

Theorem 9 *Consider a $2^M \times 2^M$ degenerated banyan network. Suppose that there are p linecards, indexed from $0, 1, \ldots, p-1$ and they are placed in a $2^M \times 2^M$ degenerated banyan network according to the placement rule in (9). Then, a $2^M \times 2^M$ degenerated banyan network can be used as a $p \times p$ rotator and a $p \times p$ symmetric TDM switch for these p linecards.*

As commented in [11], such a placement rule also allows one to incrementally update the number of linecards in a $2^M \times 2^M$ degenerated banyan network without repositioning the existing linecards. Specifically, suppose that there are already p linecards placed in a $2^M \times 2^M$ degenerated banyan network, and we would like to add a new linecard. As in the placement rule, write $p = 2^m + \ell$. If $\ell < 2^m$, then the new one is placed in $((2\ell+1)2^{M-m-1})^{th}$ input/output port of the $2^M \times 2^M$ degenerated banyan network. On the other hand, if $\ell = 2^m$, then the new one is placed in the $(2^{M-m-2})^{th}$ input/output port. For example, for an 8×8 degenerated banyan network, the order of placing new linecards in the input/output ports is 0,4,2,6,1,3,5,7.

For a $2^M \times 2^M$ degenerated banyan network, all the nodes can be made by 2×2 switches (with 2^M 1×2 switches for the 2^M input nodes and 2^M 2×1 switches for the 2^M output nodes). To use a $2^M \times 2^M$ degenerated banyan network as a $p \times p$ rotator, we first place the p linecards according to Theorem 9. For each (mapped) $p\times p$ connection matrix needed for a $p \times p$ rotator, we can find the routing paths according to the routing rule in (2). As such, all the connection patterns of the 2×2 switches in a degenerated banyan network can be determined accordingly. Specifically, consider the j^{th} node at the k^{th} stage, denoted by node (k, j), for some $1 \leq k \leq M-1$. Note that such a node is neither an input node nor an output node. If nodes $(k-1, j)$ and $(k+1, j)$ are connected through node (k, j), we say that the 2×2 switch for node (k, j) is in the "bar" state and in the "cross" state otherwise. Also, the j^{th} input (resp., output) node, denoted by node $(0, j)$ (resp., (M, j)), is said to be in the "bar" state if it connects to node $(1, j)$ (resp., node $(M-1, j)$) and in the "cross" state otherwise. A switch is said to be in the state "don't care" if the connection matrix of the degenerated banyan network is implemented no matter which state the switch is in.

We illustrate how one uses an 8×8 degenerated banyan network as a 5×5 rotator (see Fig. 4). According to Theorem 9, the five linecards are placed in the 0^{th}, 1^{st}, 2^{nd}, 4^{th} and 6^{th} input/output ports of the 8×8 degenerated banyan network. The five connection matrices that need to be implemented are P^n, $n = 0, 1, \cdots, 4$, where P is a 5×5 circular shift matrix. In Table I, we show all the states of nodes in this degenerated banyan network. The element in the m^{th} row and n^{th} column represents the states of switches with the same index m for the connection matrix P^n, where the states are represented as a sequence of "bar" (b), "cross" (x) and "don't care" (z), in the increasing order of their stages (from left to right). For example, the sequence xbzx in the 2^{th} row and the 1^{st} column of Table I indicates that the 2×2 switch for node (0,2) (resp., (1,2), (2,2) and (3,2)) should be set to the cross (resp., bar, don't care, cross) state for the 5×5 circular shift matrix P. Moreover, Table II shows all the states of nodes in this degenerated banyan network if they are used as a 5×5 symmetric TDM switch.

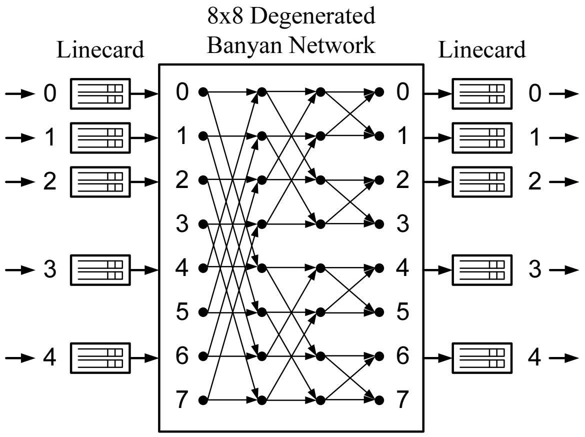

Fig. 4. Using an 8×8 degenerated banyan network as a 5×5 rotator.

	0	1	2	3	4
0	bbbb	bbxb	bxbb	xxxb	xbbx
1	bbbb	bxzx	xzzx	xzzx	bbxx
2	bbbb	xbzx	xbxb	bxbb	bxxb
3	zzzz	zzbz	zzzz	zzzz	zzzz
4	bbbb	bxxb	xzzx	xxbb	xbxb
5	zzzz	zzzz	zxxz	zbzz	zzzz
6	bbbb	xbxb	xxbb	xzzx	bxxb
7	zzzz	zzzz	zzzz	zzbz	zzzz

TABLE I
STATES OF SWITCHES IN AN 8×8 DEGENERATED BANYAN NETWORK FOR A 5×5 ROTATOR.

IV. CONCLUSION

In the paper, we studied the problem of incremental update of the number of linecards in load-balanced switches. We showed that a $2^M \times 2^M$ degenerated banyan network, obtained from the classical $2^{M+1} \times 2^{M+1}$ banyan network by using only half of the 2^{M+1} inputs/outputs, can be used as a $p \times p$ rotator (and a $p \times p$ symmetric TDM switch) for any $2 \leq p \leq 2^M$. This is done by a specific rule of placing the p linecards in the 2^M input/output ports of the degenerated $2^M \times 2^M$ banyan network. In particular, when $p = 2^M$, the $2^M \times 2^M$ degenerated banyan network can also be used as a *crosstalk-free* $2^M \times 2^M$ rotator (and $2^M \times 2^M$ symmetric TDM switch), where all the routing paths do not share a common node. As such, one can use the classical $2^{M+1} \times 2^{M+1}$ banyan network as the universal switch for a $2^M \times 2^M$ load-balanced switch that is capable of providing incremental update of the number of linecards.

Finally, we note that it is possible to extend our results to *generalized* banyan networks, where the interconnections between two consecutive stages are characterized by the

	0	1	2	3	4
0	bbbb	bbxx	bxxb	xxbb	xxxb
1	xzzx	bbxx	bbbb	bxzx	xzzx
2	xbxb	xxbb	bxxb	bxzx	bbbb
3	zzzz	zzzz	zzzz	zzbz	zzzz
4	xzxb	bbbb	bxxb	xxbb	xbzx
5	zbzz	zzzz	zzzz	zzzz	zxxz
6	xbzx	xxbb	bxxb	bbbb	xzxb
7	zzbz	zzzz	zzzz	zzzz	zzzz

TABLE II
STATES OF SWITCHES IN AN 8×8 DEGENERATED BANYAN NETWORK FOR A 5×5 SYMMETRIC TDM SWITCH.

generalized r-ary representation like twister networks in [11].

Appendix

In this appendix, we prove Theorem 5.

(i) We prove the theorem by contradiction. Suppose that the routing paths of the pair of input/output ports (i_1, o_1) and (i_2, o_2) share a node at stage j. Note that the shared node cannot be an input node or an output node as otherwise we would have either $i_1 = i_2$ or $o_1 = o_2$ that contradicts to the assumption that a connection matrix is a sub-permutation matrix. Thus, we know that $1 \leq j \leq M-1$. As these two routing paths traverse a common node at stage j, we have from the routing rule in (2) that

$$I_k(i_1) = I_k(i_2), \text{ for } 1 \leq k \leq M-j, \qquad (10)$$

$$I_k(o_1) = I_k(o_2), \text{ for } M-j+1 \leq k \leq M. \qquad (11)$$

From (10) and (11), we have that

$$\begin{aligned} |o_1 - o_2| &= \left| \sum_{k=1}^{M} I_k(o_1) 2^{k-1} - \sum_{k=1}^{M} I_k(o_2) 2^{k-1} \right| \\ &= \left| \sum_{k=1}^{M-j} I_k(o_1) 2^{k-1} - \sum_{k=1}^{M-j} I_k(o_2) 2^{k-1} \right| \\ &\leq \sum_{k=1}^{M-j} \left| I_k(o_1) - I_k(o_2) \right| 2^{k-1} \leq 2^{M-j} - 1. (12) \end{aligned}$$

Similarly, we also have from (10) and (11) that

$$\begin{aligned} |i_1 - i_2| &= \left| \sum_{k=1}^{M} I_k(i_1) 2^{k-1} - \sum_{k=1}^{M} I_k(i_2) 2^{k-1} \right| \\ &= \left| \sum_{k=M-j+1}^{M} I_k(i_1) 2^{k-1} - \sum_{k=M-j+1}^{M} I_k(i_2) 2^{k-1} \right| (13) \\ &\leq \sum_{k=M-j+1}^{M} \left| I_k(i_1) - I_k(i_2) \right| 2^{k-1} \\ &\leq N - 2^{M-j}. \qquad (14) \end{aligned}$$

On the other hand, we note from (13) that

$$\begin{aligned} |i_1 - i_2| &= \left| \sum_{k=M-j+1}^{M} I_k(i_1) 2^{k-1} - \sum_{k=M-j+1}^{M} I_k(i_2) 2^{k-1} \right| \\ &= 2^{M-j} \left| \sum_{k=1}^{j} (I_{k+M-j}(i_1) - I_{k+M-j}(i_2)) 2^{k-1} \right|. \qquad (15) \end{aligned}$$

Notice that $\sum_{k=1}^{j} (I_{k+M-j}(i_1) - I_{k+M-j}(i_2)) 2^{k-1}$ cannot be zero as otherwise we have $i_1 = i_2$ from (15). Thus, $| \sum_{k=1}^{j} (I_{k+M-j}(i_1) - I_{k+M-j}(i_2)) 2^{k-1}|$ is greater or equal to one, and hence

$$\begin{aligned} |i_1 - i_2| &= 2^{M-j} \left| \sum_{k=1}^{j} (I_{k+M-j}(i_1) - I_{k+M-j}(i_2)) 2^{k-1} \right| \\ &\geq 2^{M-j}. \qquad (16) \end{aligned}$$

From (5), (14) and (16), we see that

$$\begin{aligned} d_N(i_1, i_2) &= \min[|i_1 - i_2|, N - |i_1 - i_2|] \\ &\geq 2^{M-j} \qquad (17) \end{aligned}$$

From (17) and (12), it then follows that

$$d_N(i_1, i_2) \geq 2^{M-j} > |o_1 - o_2|,$$

which contradicts to the assumption in (6).

(ii) Suppose that the routing paths for two input/output pairs (i_1, o_1) and (i_2, o_2) share a common link between stages $j-1$ and j. Notice that $2 \leq j \leq M-1$ as otherwise they share the same input port or output port. Again, according to the routing rule (2), we have that

$$\begin{aligned} &I_k(i_1) = I_k(i_2), \quad \text{for } 1 \leq k \leq M-j+1 \\ &I_k(o_1) = I_k(o_2), \quad \text{for } M-j+1 \leq k \leq M. \end{aligned}$$

By using the same procedure as used in the proof of part (i), one can verify that $|o_1 - o_2| \leq 2^{M-j} - 1$ and $d_N(i_1, i_2) \geq 2^{M-j+1}$. It then follows that

$$d_N(i_1, i_2) \geq 2^{M-j+1} > 2|o_1 - o_2| + 1,$$

which contradicts to the assumption in (7).

References

[1] V. E. Benes. *Mathematical Theory of Connecting Networks and Telephone Traffic*. New York: Academic Press, 1965.

[2] C. -S. Chang, D. -S. Lee and Y. -S. Jou, "Load balanced Birkhoff-von Neumann switches, part I: one-stage buffering," *Computer Communicaitons,* vol. 25, pp. 611-622, 2002.

[3] C.-S. Chang, D.-S. Lee, Y.-J. Shih and C.-L. Yu, "Mailbox switch: a scalable two-stage switch architecture for conflict resolution of ordered packets," *IEEE Transactions on Communications*, vol. 56, pp. 136-149, 2008.

[4] H. J. Chao, C. H. Lam and E. Oki. *Broadband Packet Switching Technologies: A Practical Guide to ATM Switches and IP Routers*. John Wiley & Sons, Inc., 2001.

[5] Y. Deng and T. T. Lee, "Crosstalk-Free Conjugate Networks for Optical Multicast Switching," *Journal of Lightwave Technology* vol. 24, pp. 3635-3645, 2006.

[6] J.-J. Jaramillo, F. Milan and R. Srikant, "Padded frames: A novel algorithms for stable scheduling in load-balanced switches, " *IEEE/ACM Transactions on Networking,* vol. 16, no. 5, pp. 1212-1225, Oct. 2008.

[7] I. Keslassy, S. -T. Chung, N. McKeown, "A load-balanced switch with an arbitrary number of linecards," *Proc. IEEE INFOCOM*, 2004.

[8] I. Keslassy, S. -T. Chung, K. Yu, D. Miller, M. Horowitz, O. Sloggard, and N. McKeown, "Scaling internet routers using optics, " *ACM SIGCOMM 2003*, Karlsruhe, Germany, Sep. 2003.

[9] S.-Y. R. Li. *Algebraic Switching Theory and Broadband Applications*. Academic Press, 2001.

[10] S.-Y. R. Li and X. J. Tan, "Preservation of conditionally nonblocking switches under two-stage interconnection," *IEEE Transactions on Communications*, vol. 55, pp. 973-980. 2007.

[11] C.-M. Lien, C.-S. Chang, J. Cheng, D.-S. Lee and J.-T. Liao, "Twister networks and their applications to load-balanced switches," accepted by *Proc. IEEE INFOCOM*, 2010.

[12] G. Maier, A. Pattavina, "Design of photonic rearrangeable networks with zero first-order switching-element-crosstalk," *IEEE Transactions on Communications*, vol. 49, No. 7, pp. 1268-1279, Jul. 2001.

[13] K. Padmanabhan and A. Netravali, "Dilated networks for photonic switching," *IEEE Transaction on Communications* COM-35, 1357-1365, 1987.

[14] M. Schwartz. *Broadband Integrated Networks*. Prentice Hall, 1996.

[15] M. M. Vaez and C.-T. Lea, "Strictly nonblocking directional-coupler-based switching networks under crosstalk constraint," *IEEE Transaction on Communications*, vol.48, no.2, pp.316-323, February 2000.