

병렬처리 시스템에서의 Dual 네트워크의 설계 및 오류허용 라우팅 전략

正會員 최 창 훈* 正會員 金 聖 天*

Design and Fault Tolerant Routing Scheme of Dual Network in Parallel Processing System

Chang Hoon Choi*, Sung Chun Kim* *Regular Members*

※ 본 연구는 과학재단 연구비 지원에 의한 것임.

要 約

Gamma 네트워크는 많은 수의 중복경로를 포함하고 있기 때문에 오류발생시 그 오류를 허용할 수 있는 능력을 보유하고 있다. 그러나 근원지 번호와 목적지 번호가 동일할 경우는 경로가 한개 뿐이어서 오류가 발생한다면 오류를 허용할 수 있는 어떠한 방법도 존재하지 않는다. 더우기 그 네트워크가 동적 패킷 라우팅(dynamic packet routing) 전략을 사용할 경우에 오류가 발생한다면 다른 중복경로를 찾기위해 후진추적(backtracking)을 해야하는 문제점도 가지고 있다.

본 논문에서는 이러한 문제들을 새로운 스테이지 첨가 없이 해결할 수 있는 네트워크인 Dual 네트워크를 제안하고 있다. Dual 네트워크는 처음과 마지막 스테이지를 제외하면, Gamma 네트워크와 동일한 크기의 스위칭 소자를 사용하고 있으며, 특히 같은 크기의 Gamma 네트워크보다 스테이지(stage)수가 한개 줄어들어 매우 효율적인 네트워크이다. 또한, Dual 네트워크에서는 라우팅 방법으로 목적지 주소 라우팅(Destination Tag Routing) 방법을 사용하기 때문에, 보다 간편하고 빠른 라우팅을 수행할 수 있는 장점도 가지고 있다.

ABSTRACT

The Gamma Network contains the redundant path thereby it provides the ability to tolerate the faults occurred. However, in case of identical the source and destination number, only a single path exists, therefore there is no way of connecting for the fault situation. In addition, for the dynamic packet routing strategy, it should perform backtracking analysis to find the redundant path.

* 西江大學校 電子計算學科
Dept. of Computer Science, Sogang University
論文番號 : 9486
接受日字 : 1994年 3月 18日

In this paper we propose a new network, Dual Network, to resolve these drawbacks. The Dual Network uses switching elements about the same network size as the Gamma Network except first and last stage, and it is more efficient than the Gamma Network, for it has reduced the switching stages by one. And since it uses a destination tag routing scheme for the control algorithm, it has an advantage of becoming of simpler and faster routing control.

I. 서 론

다수 개의 프로세서와 다수 개의 기억장치 모듈을 갖고 있는 병렬처리 시스템에서는 프로세서와 프로세서 또는 프로세서와 기억장치 모듈간의 자료 전송 및 교환이 원활히 이루어지도록하여, 빠른 시간내에 주어진 작업을 처리할 수 있도록 하기 위한 상호 연결망(Interconnection Network)이 요구되어 진다. 이러한 환경에 있어서, 효율적인 상호 연결 네트워크의 구성은 전체적인 병렬 처리 시스템의 성능을 좌우하게 될뿐만아니라, 시스템의 신뢰도 향상에 중요한 역할을 하게 될것이다. 따라서 상호 연결 네트워크에서의 한요소(element)의 오류는 전체 시스템의 기능에 커다란 장애 요인이 될 수 있을 것이다.

다단계 상호연결망(Multistage Interconnection Network : MIN)에 관한 초기의 연구에는 Omega[1-6], Baseline[3,4], Generalized Cube Network[18]등이 있는데 이들 네트워크는 기본적으로 2×2 크기의 스위칭 소자, $\log_2 N$ 개의 스테이지와 각 스테이지마다 $N/2$ 개의 스위칭 소자로 구성되어 있다. 이러한 한개의 유일한 경로를 갖는 MIN에서는 임의의 근원지-목적지로 통하는 경로가 단지 한개 밖에 존재하지 않아서 한개의 스위칭 소자의 오류는 많은 수의 프로세서들 또는 메모리 모듈들을 사용 불가 또는 전체 시스템이 운영 불가하게되는 점이 초래하게 된다[20]. 따라서 오류허용(fault tolerance)을 위한 중복경로를 제공하기 위하여 2×2 스위칭 소자의 스테이지를 첨가시키는 연구[2]가 있었는데, 이러한 경우에 있어서는 근원지와 목적지간에 2개의 중복경로가 존재하여 한개의 오류는 허용할 수 있게 되었다. 또한 중복경로를 제공하는 또 다른 부류로서 plus-minus-2i 함수[11]연결형식을 기초로한 Data Manipulator[5, 17], Augmented Data Manipulator(ADM)[10, 12, 13], Inverse Augmented Data Manipulator(IADM)[9, 14-17]등이 있다. 특히 Parker[14]는 IADM을 변형시킨 Gamma Network(GN)을 제안하였다. GN에서는 다수 개의 경로를 표현하기 위하여 중복 수 체계

(redundant number system)[14]를 사용하였다.

따라서 근원지와 목적지 사이의 경로는 목적지 번호와 근원지 번호간의 차이 값의 여러 중복형식(redundant form)중에서 한개를 선택하여 사용하게 된다. 그러나 GN에서 많은 중복 경로를 가지고 있을지라도 근원지 번호와 목적지 번호가 같을 경우 경로가 한개 밖에 존재하지 않으므로 그 경로에서 오류가 발생한다면 원하는 연결이 이루어질 수 없게 된다. 또한, 패킷 스위칭 방법으로 동적 라우팅을 할 경우 라우팅 도중 스위칭 소자의 중간 링크(Middle Link)의 오류가 발생한다면 새로운 경로를 찾기위해 후진 추적하는 값이는 값어질 것이고 라우팅 태그의 수정 계산 시간 또한 문제시 되어질 것이다. 그러므로 GN은 중복경로를 많이 가지고 있어도 특별한 연결에서의 오류는 해결 불가능하거나 비록 해결 경로가 존재 하더라도 그 경로를 찾는데 많은 시간과 복잡한 제어를 필요로 한다.

이에 대한 최근 [21]과 같은 연구에서 이러한 Gamma 네트워크의 오류를 해결하기 위해 추가적인 스테이지를 사용하였다. 그러나 이러한 오류허용 방식은 추가된 스테이지로 인한 과다 비용과 더 많은 라우팅 시간을 초래하게 된다.

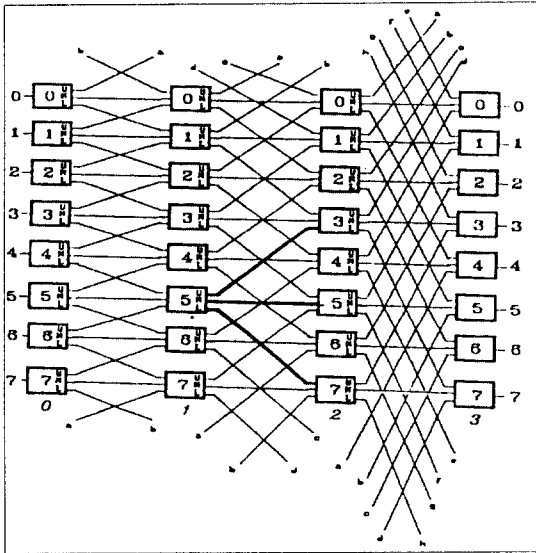
본 논문에서는 이러한 GN에서 발생하는 여러 문제들의 원인들을 분석하고 그원인이 되는 요소들을 재구성하여 새로운 네트워크인 Dual 네트워크를 설계한다. 그리고 Dual 네트워크의 라우팅 전략과 한개의 오류 및 다중 오류들을 허용할 수 있는 오류허용 라우팅을 제안하였다.

II. Gamma Network(GN)

2.1 GN의 구성

〈그림 1〉은 $N=8$ 인 GN을 나타낸 것이다. $N(=2^n)$ 개의 입/출력을 가진 GN은 $n+1$ 개의 스테이지와 각 스테이지당 N 개의 스위칭 소자로 이루어져 있다. 처음과 마지막 스테이지를 제외한 모든 스위칭 소자는 3개의 입력과 3개의 출력(3×3)으로 이루어져 있

고, 처음 스테이지(입력 스테이지)는 한개의 입력과 3개의 출력(1×3)으로 이루어져 있으며, 마지막 스테이지(출력 스테이지)는 3개의 입력과 한개의 출력(3×1)으로 이루어져 있다. 마지막 스테이지를 제외한 각 스테이지에 있는 스위칭 소자의 3개 출력중에서 상위출력을 Upper 링크(U), 중위 출력을 Middle 링크(M), 하위 출력을 Lower 링크(L)라고 한다. 이러한 소자들의 입출력을 이용하여 스테이지 $i(0 \leq i < n)$ 의 스위칭 소자를 스테이지 $i+1$ 의 스위칭 소자로 연결하는 형식은 다음과 같다. 스테이지 i 에 있는 스위칭 소자 $j(0 \leq j < N)$ 의 상위 출력 U는 스테이지 $i+1$ 의 스위칭 소자 $(j-2) \bmod N$ 으로 연결되고, 중위 출력 M은 스테이지 $i+1$ 의 스위칭 소자 j , 그리고 하위 출력 L은 스테이지 $i+1$ 의 스위칭 소자 $(j+2) \bmod N$ 으로 연결된다.



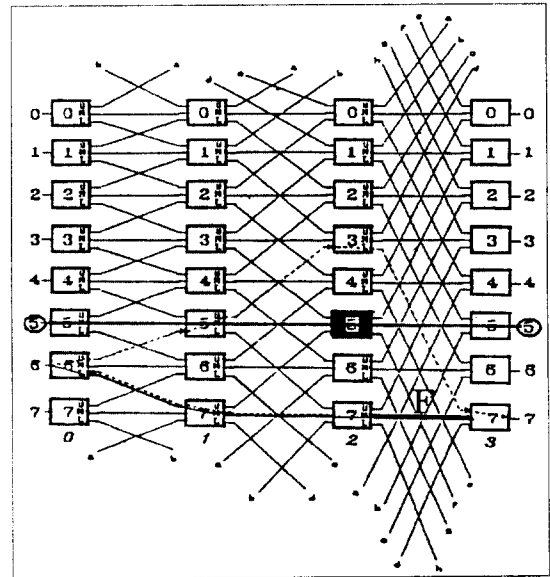
<그림 1> 8×8 GN(Gamma Network)

2.2 오류허용 라우팅 전략 및 문제점

<그림 2>에서 스테이지 2의 스위칭 소자 5번에서 오류가 발생하여서 사용할 수 없게 되었다고 하자. 이때 초기 태그(011)를 사용한다면, 오류가 발생한 5번 스위칭 소자 때문에 근원지와 목적지 사이의 연결은 불가능하게 될 것이다. 그러나 나머지 4개의 태그 값을 사용한다면 이들은 모두 스위칭 소자 5번을 통과하지 않게되므로 오류를 허용할 수 있게 된다.

이러한 오류허용 라우팅 과정을 보이기 위하여, 태그값 $111 (= t_2 t_1 t_0)$ 만을 예로써 보이겠다. 근원지로부터 직접연결된 스테이지 0에있는 스위칭 소자 2번으로 패킷이 전송되어 들어오면, t_0 가 1임을 검사하여 L링크로써 스테이지 1에 있는 스위칭 소자 3번으로 패킷을 전송한다. 이 패킷을 입력으로 받은 스위칭 소자 3번은 t_1 이 1이므로 U링크에 연결된 스테이지 2에 있는 스위칭 소자 1번으로 패킷을 전송하게 된다. 이 스위칭 소자에서는 t_2 가 1이므로 L링크를 사용해 마지막 스테이지에 있는 스위칭 소자 5번으로 전송하게 되어 원하는 목적지 5로 패킷을 전송하게 된다.

먼저 근원지 번호와 목적지 번호가 같은 경우에 있어서 오류허용에 대해 살펴도록 하자. 이같이 근원지와 목적지간의 번호가 동일할 경우에 있어서는 불행하게도 경로가 유일하게 한개 뿐이어서 그경로상에서의 오류를 허용할 수 있는 어떠한 방법은 전혀 존재하지 않는다.



<그림 2> GN에서의 오류허용 문제점

예를 들어, <그림 2>과 같이 근원지(5)와 목적지(5)를 연결하는 경로에서 스테이지 2에 있는 5번 스위칭 소자에서 오류가 발생했다면, 이의 연결은 불가능하게 된다. 이렇게 UP P MIN보다 다수의 중복경로를 제공하는 GN에서 단지 한개 요소의 오류 때문에 특정한 쌍들간의 정보교환을 이룰수 없게된다면,

이것은 커다란 자원의 손실일 뿐만 아니라, 시스템의 신뢰도를 떨어뜨리는 결과가 될 것이다.

다음은 스위칭 방법(switching methodology)에 관해서 살펴보자. 위에서 보았듯이 GN에서는 오류를 허용하기 위해서 중앙 제어로서 주기적인 오류 진단이 필요하다. 다시 말해서, 동적 패킷 스위칭(dynamic packet switching)이 어렵다는 것이다. 만약 동적 패킷 스위칭이 가능할지라도[15], 오류가 발생할 경우에 각 스위칭 소자에서 라우팅 태그를 다시 계산해야 하기 때문에, 병렬처리 시스템처럼 고속을 요구하는 시스템에서는 커다란 오버헤드가 될 것이다.

<그림 2>에서 'F'라고 표시된 굵은 링크에서 오류가 발생했다고 가정하자. 그리고 근원지 6에서 목적지 7로 패킷이 동적 패킷 스위칭으로 전송되고 있다고 하자. 초기 태그는 001($t_2t_1t_0$)이므로 스테이지 2의 7번 스위칭 소자까지 도착한 패킷은 $t_2=0$ 이므로 M링크를 사용하려고 하지만, 그 링크에서 오류가 발생했으므로 더 이상 원하는 목적지로 패킷을 전송할 수 없게 된다. 따라서 새로운 경로를 찾기위해 오류가 발생한 스테이지에서 입력 스테이지 방향쪽으로 진행해서 맨 처음으로 M링크를 사용하지 않은 스위칭 소자를 만날때까지 후진추적을 수행해야 한다. 일단 후진 추적이 끝나면, 오류가 발생한 링크를 피할 수 있는 새로운 태그를 다시 계산해야 한다.

이 예에서는 스테이지 0까지 후진 추적을 해야하고 태그 111를 다시 새로이 계산하여 원하는 목적지 7로 패킷을 무사히 도착시킨다. 이 같은 경우에 있어서, 네트워크의 크기가 커질수록 후진 추적을 해야하는 깊이는 깊어진다. 따라서 패킷이 통과해야하는 스테이지 수는 많아질 뿐만 아니라, 스위칭 소자에서 새로운 경로를 설정하기 위하여 라우팅 태그를 다시 계산해야 하기 때문에 시간적인 낭비를 초래하게 된다. 또한, 후진 추적을 수행하기 위해서는 양방향성(bi-direction)을 가진 스위칭 소자 및 링크를 사용해야 하므로 하드웨어적인 비용도 감수해야 할 것이다.

또한 GN에서의 목적지는 마지막 스테이지에 있는 한개의 스위칭 소자로만 연결되어 있으므로 이곳에서의 오류가 발생한다면 아주 치명적 오류로서, 그 목적지로의 향하는 어떠한 연결도 이루어질 수 없게 되어 시스템의 효율은 급격히 저하될 것이다.

결론적으로 GN에서 발생하는 문제점을 요약하면 다음과 같다.

첫째, 다수의 중복 경로를 포함한 GN일지라도 $S = D$ 인 쌍에서의 경로는 오직 한개 뿐이어서 오류 허

용이 불가능하다.

둘째, 동적 패킷 스위칭이 어렵기 때문에 네트워크 상의 오류를 찾기위해 주기적인 오류진단을 해야하고, 만약 동적 라우팅을 한다고 하여도 후진 추적을 해야하는 경우가 발생하기 때문에 시스템의 효율이 떨어진다.

셋째, 만약 마지막 스테이지에 있는 스위칭 소자의 오류가 발생한다면 그 스위칭을 통해 연결되는 목적지로의 어떠한 연결도 이루어질 수 없다.

다음상에는 중복성 방향 그래피(redundancy digraph)[8, 19]와 불력구조로 재구성된 GN을 통하여, GN에서 이러한 문제점들에 발생하게되는 원인을 찾아낸 다음, 본 논문의 주제가되는 Dual 네트워크를 설계하기로 하겠다.

III. Dual 네트워크

3.1 Dual 네트워크의 설계 배경

3.1.1 GN의 분석

(a) 경로의 수에 의한 GN의 분석

$2^n \times 2^n$ GN의 태그 t에 대한 경로의 수, $P_n(t)$ [11]는 다음과 같이 나타낼 수 있다($P_n(0) = 1$).

$$\begin{cases} \text{만약 } t \text{ 가 짝수이면, } P_n(t) = P_{n-1}(t/2) \\ \text{만약 } t \text{ 가 홀수이면,} \\ P_n(t) = P_{n-1}((t-1)/2) + P_{n-1}((t+1)/2) \end{cases}$$

<표 1> Gamma 네트워크에서 태그 값에 따른 경우의 수

t	N	2	4	8	16
0		1	1	1	1
1		2	3	4	5
2			2	3	4
3			3	5	7
4				2	3
5				5	8
6				3	5
7				4	7
8					2
9					7
10					5
11					8
12					3
13					7
14					4
15					5

〈표 1〉에서 태그값에 따른 경로의 수를 분석하여 보면, 아래와 같은 사실을 발견할수 있다.

- 1) $S = D$ 일때 한개의 유일한 경로만 존재한다.
- 2) 홀수인 태그값은 짝수인 태그값 보다 많은 수의 중복 경로가 존재한다.

이같은 원인은 중복 수 체제로서 확인되어 질 수 있다. 즉, 1)번의 $S = D$ 인 경우는 $D - S = 0$ 이므로 0 이외의 중복 수는 존재하지 않는다. 그리고 $S = D$ 인 경우에 있어서는 경로는 유일하게 한개만 존재하게 된다. 또한 2)번의 경우를 살펴보면, 태그값이 홀수인지, 또는 짝수인지는 t_0 값에 따라 결정되어 진다. 그리고 스테이지 0에서는 3가지 형태(-1, 0, +1)의 연결방법중 한가지를 선택하게 되어있다. 따라서, 만약 $(D - S)$ 가 짝수이면 스테이지 0에서 M링크가 선택되고 그외의 중복경로는 가질수가 없게되지만, $(D - S)$ 가 홀수이면 +1 또는 f중에서 한개를 선택할 수 있어 2개의 중복 경로를 가질 수가 있다. 이러한 현상은 다음 스테이지에서도 계속적으로 반복되므로, 근원지와 목적지에 따라서 경로의 수가 차이가 날 수 밖에 없는 것이다.

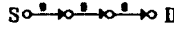
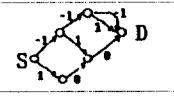
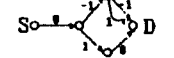
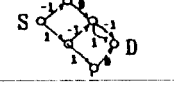
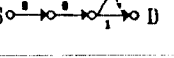


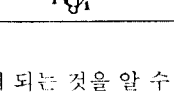
(b)중복성 방향 그래프를 이용한 GN의 분석

위에서 살펴 보았듯이 GN은 근원지와 목적지쌍에 따라 불규칙한 수의 경로가 존재하게 되는데 그 문제점들은 중복성 방향 그래프로써 분석될 수 있다.

$N = 8$ 인 GN의 중복성 방향 그래프는 〈표 2〉에 나타나 있다. 〈표 2〉에 나타난 중복성 방향 그래프의 각 노드에서는 태그가 0이면 간선의 수가 1개가 되고 그렇지 않으면 간선의 수가 2개가 되어 선택적 경로가 존재함을 볼 수 있다. 따라서 $N = 8$ 인 GN에서의 전체의 경로 수는 t 에 따라 1개에서 5개까지를 갖게 된다.

또한 〈표 2〉의 중복성 그래프를 살펴보면, $t = 0$ 뿐만 아니라 $t = 2, 4, 6$ 인 경우에도 문제가 발생하는 것을 찾아볼 수가 있다. $t = 2, 4, 6$ 인 경우에는 전체적인 중복 경로가 여러개 존재하지만, 첫 스테이지에서 부터 태그 비트가 0이므로 첫 스테이지에서 노드의 간선 수가 1개 밖에 존재하지 않는다. 그러므로 이곳에서 오류가 발생한다면 이들의 전체적인 다수 개의 중복경로들은 소용이 없게 된다. 따라서, t 값이 짝수인 모든 경로에 대해 이런 오류가 발생한다면, 다수 개의 중복경로는 유명무실하게 되어 오류의 허용이 불

〈표 2〉 t에 따른 중복경로의 표현

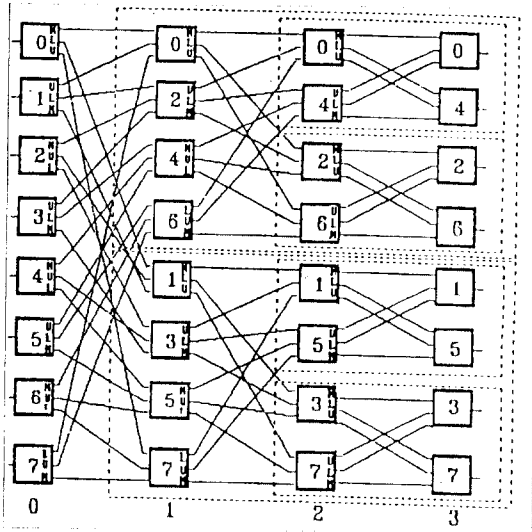
경로 t	중복성 방향 그래프	중복경로의 표현
0		000
1		001, 01-1, 1-1-1, -1-1-1
2		010, 1-10, -1-10
3		011, 10-1, 1-11, -10-1, -1-11
4		100, -100
5		101, 11-1, 0-1-1, -101, -11-1
6		110, 0-10, -110
7		111, 00-1, 0-11, -111

가능하게 되는 것을 알 수 있다.

3.1.2 GN의 블록 구조화

결론적으로 GN에 있는 경로의 불균형성은 〈그림 3〉로서 명백히 보여진다. 〈그림 3〉는 $N = 8$ 인 GN의 스위칭 소자 위치를 재구성시켜 블록 구조화 시킨 모습이다. 여기서 살펴볼 수 있듯이 각 스위칭 소자의 출력 U, M, L 링크중에서 U, L 링크들은 다음 스테이지의 같은 블록에 있는 스위칭 소자로 연결되어 있지만 오직 M 링크만이 홀로 다른 블록의 스위칭 소자로 연결되어 있다.

이렇게 GN의 연결형식이 균형을 이루지 못하므로, 특정 연결에 대해서는 적은 수의 경로, 심지어 $S = D$ 의 경우에는 하나의 경로만 존재하는 경우도 발생한다. 따라서 본 논문에서는 GN의 이러한 연결의 불균형성이 발생하는 요소들을 재구성을 하여 네트워크의 균형을 이루어서, 앞에서 살펴 보았던 여러 문제들을 해결하려 한다. 즉, GN에 있는 스위칭 소자들의 입출력 링크중에서 주축이되는 연결선만을 그대로 보존하며, 나머지 연결선들은 'Dual'이라는 개념을 도



<그림 3> 블럭구조로 재구성된 GN(N=8)

입해 연결시킨다. 이런 방법을 이용하면 근원지와 목적지간의 모든 쌍들이 동일한 갯수의 중복경로를 갖게 되어 오류 발생시의 연결이 불가능한 경우가 전혀 발생하지 않게 된다.

3.2 Dual 네트워크의 설계

3.2.1 링크의 제거

<그림 3>와 같은 특성을 갖고 있는 GN에 대해서 다음 블럭으로의 링크 연결이 균형을 이루기위해 GN을 변형시켜야 되는데, 이를 위하여 GN의 각 단계의 스위칭 소자들의 번호를 <그림 1>과 같이 위에서 부터 차례로 0, 1, ..., N-1로 붙여진다.

우선 네트워크의 균형성을 이루기 위해 각 스위칭 소자에 대해 U링크와 L링크중에서 한 개의 링크를 제거시켜야 한다. 먼저 스테이지 i에 있는 스위칭 소자 j의 2진 표현을 $j_{n-1} \dots j_1 j_0$ 라고 할때, 스테이지 i($0 \leq i \leq n-1$)에 있는 스위칭 소자 j의 i번째 비트 j_i 가 0이면 출력 링크중 U링크를 제거하고 j_i 가 1이면 출력 링크중 L 링크를 제거한다.

이러한 과정을 모든 스테이지를 통해 적용시키면 첫 스테이지와 마지막 스테이지의 스위칭 소자들은 각각 1x2, 2x1 크기가 되고, 중간 단계의 스위칭 소자들은 2x2크기로 형성된다.

3.2.2 스위칭 소자의 재배열

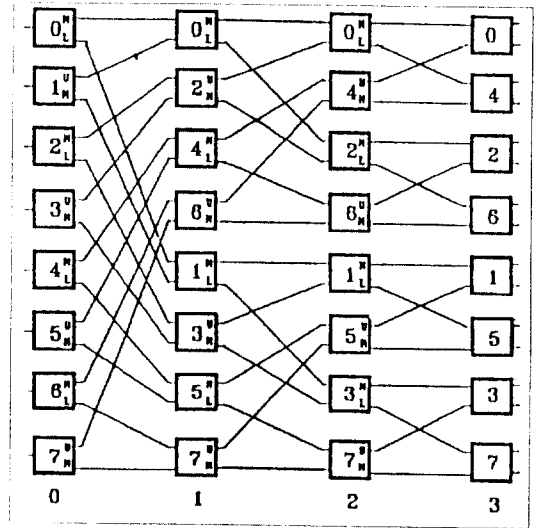
GN을 블럭구조로 변형시키기 위하여 스위칭 소자의 위치를 재정렬할 필요가 있다. 따라서 GN의 스위칭 소자를 재배열하기 위한 스위칭 소자의 매핑 함수(mapping function) Φ 를 정의한다.

[정의1] $\Phi(i, s_{n-1} \dots, s_{n-1} s_{n-2} \dots, s_1 s_0)$

$$= s_{n-1} \dots s_0 s_{n-1} \dots s_{n-2} s_{n-1}$$

즉, 스테이지 i에 있는 스위칭 소자 $s_{n-1} s_{n-2} \dots s_1 s_0$ 는 함수 Φ 에 의해 스위칭 소자 번호 $s_{n-1} \dots s_0 s_{n-1} \dots s_{n-2} s_{n-1}$ 로 매핑된다.

매핑 함수 Φ 를 적용시켜 재구성되어진 네트워크에서 마지막 스테이지의 스위칭 소자를 크기가 2x2인 스위칭 소자로 교체한다. 이렇게 형성된 네트워크를 MGN(Modified Gamma Network)이라고 하겠다. MGN은 이미 잘 알려진 Baseline 네트워크와 동일한 링크연결 형식을 가지고 있다. 예로써 N=8인 MGN이 <그림 4>에 나타나 있다.



<그림 4> MGN(Modified Gamma Network)(N=8)

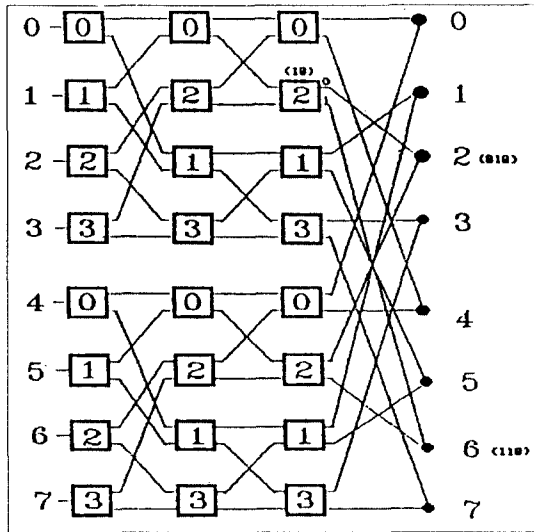
이렇게 형성된 MGN은 블럭구조를 형성하게 된다. 블럭구조를 기초로하여 구성된 MIN에서는 네트워크의 세어가 편리해질 뿐만 아니라 오류허용이 쉬어지는 장점이 있다. 또한 블럭구조는 단위(modular) 설계를 제공하기 때문에 VLSI로 구현하는데 장점이 될 수 있을 것이다[7].

3.2.3 MGN의 병합

우선, 두 MGN의 마지막 스테이지에 있는 스위칭 소자의 2N개 출력을 목적지 주소 N개에 연결시키는 작업이 필요하다. 예로써 네트워크 크기가 8인 Dual 네트워크를 설계하기 위해 크기가 4(=8/2)인 MGN 2개를 한개는 위쪽 다른 한개는 아래쪽으로 병합하여 목적지 주소로 연결한 형태가 <그림 5>에 나타나 있다. 먼저, <그림 5>의 마지막 스테이지의 스위칭 소자들의 번호를 2진수 $j_{n-2} \dots j_{1j_0}$ 로 표현할때, 각 MGN의 마지막 스테이지에 있는 스위칭 소자의 상, 하위 출력은 다음과 같은 방법으로 목적지 주소로 연결시킨다.

즉,

- 상위 출력은 $0j_{n-2} \dots j_{1j_0}$ 인 목적지 주소로 연결시키고,
- 하위 출력은 $1j_{n-2} \dots j_{1j_0}$ 인 목적지 주소로 연결시킨다.



<그림 5> 4x4 MGN 2개를 목적지 주소로 연결한 형태

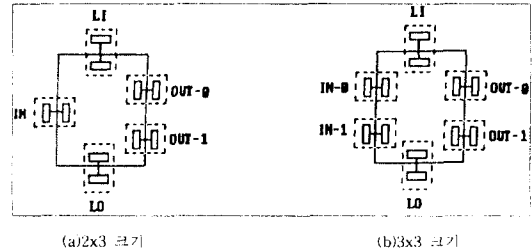
[정의2] $N \times N$ MGN에서 스테이지 i 의 k 번째($0 \leq i \leq n-1, 0 \leq k < 2$)분할을 $P_{i,k}$ 로 나타낸다.

[정의3] $N \times N$ MGN에서 스테이지 i 의 j 번째($0 \leq j \leq N$) 스위칭 소자를 $SW_{i,j}$ 로 나타낸다.

MGN-0와 MGN-1은 서로 동일한 크기의 MGN이므로 각각의 MGN에서는 모두 $P_{i,k}$ 와 $SW_{i,j}$ 를 가지고 있어 다음과 같이 정의한다.

[정의4] 두 MGN에서 서로 동일한 $P_{i,k}$ 를 듀엣 분할(duet partition)이라고 하고, 또한 동일한 $SW_{i,j}$ 를 듀엣 스위치(duet switch)라고 한다.

Dual 네트워크를 설계하기 위하여 <그림 5>과 같이 두개의 MGN을 연결한 상태에서 마지막 스테이지를 제외한 각 스테이지의 듀엣 스위치를 연결시켜야 된다. 이렇게 스위칭 소자를 연결시키기 위해서는 스위칭 소자들을 교체하여야 한다.



<그림 6> 스위칭 소자

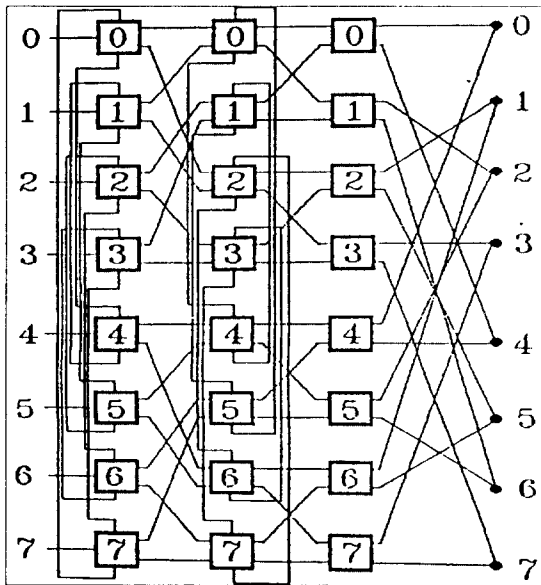
<그림 6>은 교체하게될 새로운 스위칭 소자의 모형을 나타낸 것이다. <그림 6(a)>는 첫번째 스테이지에서 사용되어질 2x3크기의 스위칭 소자의 형태이다. 여기서 2개의 입력중 한개의 바로 전 스테이지의 스위칭 소자에서 들어오는 입력(IN)이고, 다른 한개의 입력은 같은 스테이지의 다른 스위칭 소자의 LO와 연결된 입력(LI)이다. 그리고 3개 출력중 2개의 출력은 다음 스테이지의 스위칭 소자로 연결된 출력(OUT-0, OUT-1)이며 1개의 출력은 같은 스테이지의 다른 스위칭 소자의 LI와 연결된 출력(LO)이다. <그림 6(b)>는 중간 스테이지에서 사용하게될 3x3 크기의 스위칭 소자의 형태이다. 여기서 3개의 입력중 2개의 입력(IN-0, IN-1)은 전 스테이지의 스위칭 소자와 연결되어 있고, 다른 한개의 입력(LI)은 같은 스테이지의 다른 스위칭 소자의 LO와 연결된 입력이다. 또한 3개의 출력은 <그림 6(a)>의 3개 출력과 동일하다.

먼저, 듀엣 스위치 중에서 MGN-0에 속한 $SW_{i,j}$ 의 LO와 MGN-1에 속한 $SW_{i,j}$ 의 LI를 연결한다. 이과

정을 마지막 스테이지를 제외한 모든 스테이지에서 수행한다.

MGN-0에 속한 $SW_{i,j}$ 는 LI, MGN-1에 속한 $SW_{i,j}$ 는 LO가 연결되지 않은 상태로 놓여있다. 이러한 LI, LO는 듀엣 분할을 연결시키는데 이용된다. 듀엣 분할을 연결시킨다는 의미는 듀엣 분할내의 스위칭 소자들을 연결시켜 루프(loop)를 형성하는 것이다. 그래서 듀엣 분할을 연결시키기 위하여 MGN-1과 MGN-2에 듀엣쌍의 스위치를 연결한다.

이렇게 마지막 스테이지를 제외한 모든 스테이지의 듀엣 분할을 연결한다면, <그림 7>와 같은 네트워크가 형성되는데 이것이 Dual 네트워크(N)이다. 'Dual'의 의미는 앞에서 언급했듯이 MGN-0와 MGN-1의 동형 네트워크에 있는 듀엣 스위칭 소자들을 루프를 이용해 연결하면, MGN-0와 MGN-1을 이중으로 사용할 수 있다는 것이다. 이러한 특성은 네트워크상에서 스위칭 소자나 링크의 오류를 허용하는데 큰 장점 [8]이 된다.



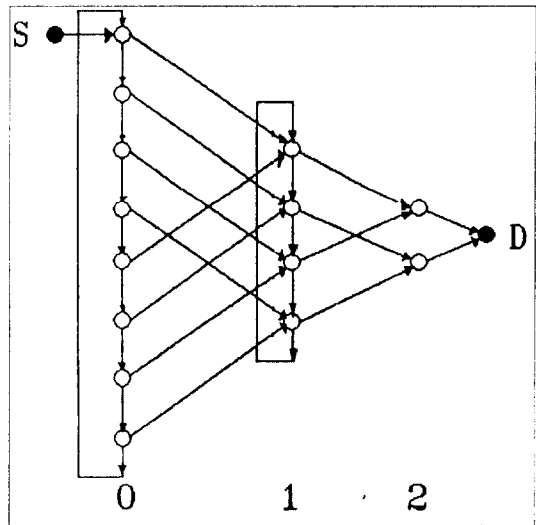
<그림 7> Dual 네트워크(N=8)

또한, 한개의 루프내에서 서로 인접되어 있는 스위칭 소자들은 다음 스테이지에 있는 서로 다른 스위칭 소자와 연결되어 있기 때문에, 다음 스테이지에 있는 스위칭 소자의 오류가 발생했을 경우 출력 LO를 이

용하여 바로 인접해 있는 스위칭 소자를 통하여 다음 스테이지로 연결시킬 수가 있어 오류의 허용을 가능하게 해준다. 다시 말해서, 스위칭 소자의 출력 LO는 MGN-0와 MGN-1을 서로 교환시켜 주는 역할을 하여 오류의 허용을 가능하게 해주는 것이다.

결론적으로 Dual 네트워크는 $\log_2 N$ 개의 스테이지로 이루어져 있으며, 각 스테이지는 N개의 스위칭 소자로 이루어져 있고, 입력과 출력 스테이지를 제외한 모든 스테이지에서는 3×3 크기의 스위칭 소자로 구성되어 있으며, 입력 및 출력 스테이지는 각각 2×3 , 2×2 크기의 스위칭 소자들로 구성되어 있다.

단계수가 줄어들었다고 하여 같은 크기의 GN보다 중복경로가 줄어들어 경로선택의 폭이 좁아진 것은 아니다. 오히려 경로의 수는 증가되었다. 이것은 중복성 방향 그래프로써 입증될 수 있다.



<그림 8> Dual 네트워크에 대한 중복성 방향 그래프

<그림 8>는 $N=8$ 인 Dual 네트워크의 중복성 방향 그래프를 나타낸 것이다. <그림 8>와 <표 2>에 나타난 중복성 방향 그래프를 비교해 보자. $N=8$ 인 GN에서 가장 많은 수의 중복 경로를 가지고 있는 것은 5개를 가진 $t=3, 5$ 이다. 반면 <그림 8>에서와 같이 동일한 크기의 Dual 네트워크에서는 모든 근원지와 목적지쌍에 대한 중복 경로수가 32개로 일정함을 볼 수 있다. 일반적으로 네트워크의 크기가 N일때의 경로의 수는 다음 정리로써 보여진다.

[정리1] $N \times N$ Dual 네트워크에서 한쌍의 근원지와 목적지간을 연결할 수 있는 경로는 총 $\prod_{i=1}^n 2^i$ 개가 존재한다($N=2^n, n \geq 2$).

(증명)

i) $n=2$ 일때 4개가 존재하므로 성립한다.

ii) $n=k$ 일때, $\prod_{i=1}^k 2^i$ 개가 존재한다고 가정하자.

iii) $n=k+1$ 일 $i=2$ 경우, 스테이지 0의 한 노드에서 다음 스테이지로 연결되는 경로가 한개뿐이므로 스테이지 0에 있는 임의의 한 노드에서 목적지로 연결되는 경로의 수는 $\prod 2^i$ 가 존재한다. 또한 스테이지 0에서의 노드의 갯수는 $2k+1$ 개이므로 총경로의 수는 $\prod 2^i \times 2^{k+1} = \prod 2^i$ 이다. 그러므로 네트워크 크기 N 인 Dual 네트워크에서 한쌍의 근원지와 목적지간의 연결해 줄 수 있는 전체 경로의 수는 $\prod 2^i$ 개가 존재한다.

IV. 라우팅 전략

4.1 목적지 주소 방식의 라우팅 전략

- Dual 네트워크의 라우팅 전략 -

- 스테이지 0에 있는 스위칭 소자들은 다음 단계를 따른다.

단계 1) 근원지로부터 직접 연결된 스테이지 0의 스위칭 소자로 패킷이 도착되면, 패킷의 목적지 주소 부분에 있는 d_0 를 검사한다.

단계 2) 만약 $d_0=0$ 이면, 패킷을 상위 출력인 OUT-0로 전송하고, 만약 $d_0=1$ 이면, 패킷을 하위 출력인 OUT-1으로 전송한다.

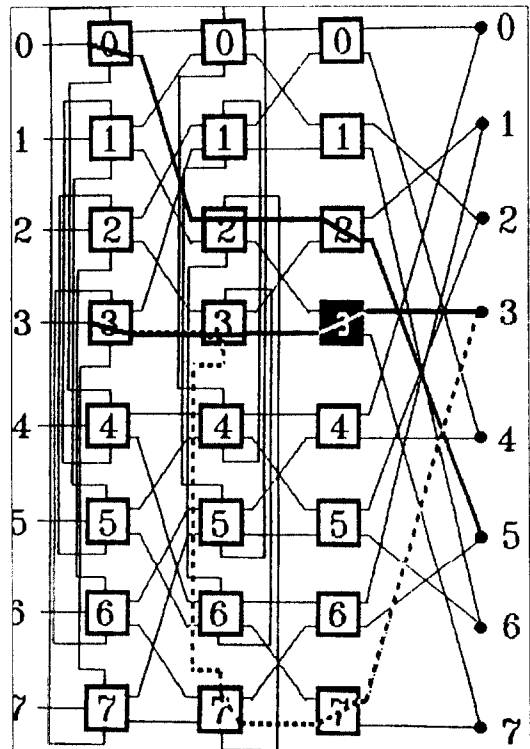
- 스테이지 $i(0 < i \leq n-1)$ 에 있는 스위칭 소자들은 다음 단계를 따른다.

단계 1) 스테이지 $i-1$ 에 있는 스위칭 소자로 부터 패킷이 전송되어 오면, 그 패킷의 목적지 주소 부분에 있는 d_i 를 검사한다.

단계 2) 만약 $d_i=0$ 이면, OUT-0로 패킷을 전송하고, 만약 $d_i=1$ 이면, OUT-1로 패킷을 전송한다.

<그림 9>과 같은 $N=8$ 인 Dual 네트워크에서 근원

지 0에서 목적지 5로 패킷을 전송하려한다. 목적지 주소 5는 2진수로 101이다. 먼저 스테이지 0에 있는 스위칭 소자 0에서 d_0 를 검사한다. $d_0=1$ 이므로 하위 출력인 OUT-1로 패킷을 전송한다. OUT-1으로 연결된 스테이지 1의 스위칭 소자 2번으로 도착한 패킷은 $d_1=0$ 임을 검사한후 상위 출력인 OUT-0로 전송하게 된다. 이렇게 해서 마지막 스테이지의 2번 스위칭 소자에 도착한 패킷은 d_2 를 검사하고 d_2 가 1이므로 OUT-1로 패킷을 출력하여 원하는 목적지 5로 도착시키게 된다.



<그림 9> Dual 네트워크에서의 라우팅 전략

4.2 Dual 네트워크의 오류허용 라우팅 전략

모든 네트워크상에서는 스위칭 소자나 출력 링크의 오류가 발생할 수 있다. 다중 프로세서를 사용하는 시스템에서 이같은 오류를 허용할 수 있는 능력을 부여함은 시스템의 신뢰도 및 성능의 향상에 크게 기여할 것이다.

오류허용이 포함된 Dual 네트워크의 라우팅 전략

- 모든 근원지에 대하여:
모든 근원지는 원하는 목적지와 연결을 하기 위하여 직접 연결된 입력 스테이지에 패킷을 전송한다.
- 입력 스테이지에 있는 스위칭 소자에 대하여:
(2개의 입력 링크에서 각각 패킷이 입력된다.)
입력 들어온 패킷에서 목적지 주소 부분에 있는 d_0 를 검사한다. 그리고 그 값에 따라 OUT-0 또는 OUT-1으로 패킷을 출력시킨다. 만약에 요구된 출력 링크가 바쁘거나(busy) 오류가 발생하면, 루프 출력인 LO를 통하여 루프에 연결된 다음 스위칭 소자로 패킷을 출력한다.
- 스테이지 $i(0 < i < n-2)$ 에 있는 스위칭 소자에 대하여:(3개의 입력 링크에서 각각 패킷이 입력된다.)
입력 링크에서 들어온 패킷에서 목적지 주소부분의 d_i 를 검사한다. 그리고 그 값에 따라 OUT-0 또는 OUT-1으로 패킷을 출력한다. 만약에 요구된 출력 링크가 바쁘거나 오류이면, 루프 출력인 LO를 통하여 루프로 연결된 다음 스위칭 소자로 패킷을 출력한다.
- 출력 스테이지에 있는 스위칭 소자에 대하여
(2개의 입력 링크에서 각각 패킷이 입력된다.)
각각의 입력 링크로 들어온 패킷은 목적지로 향하는 1개의 출력 링크로 출력한다.

특히, 이 방법에서는 오류 허용을 위한 후진 추적이 전혀 발생하지 않는다. 이같은 사실은 오류를 허용하기 위해 라우팅 태그의 변형이나 후진 추적이라는 오버헤드와 비교해 볼때 두드러진 시스템 효율의 향상을 가져온 것이다. 그리고 특히 주목해야 될것은 근원지와 목적지번호가 동일한 경우에 경로상의 오류가 발생했을때, GN에서는 그 오류는 해결이 불가능 하였지만, Dual 네트워크에서는 완전하게 오류를 허용할 수 있게 되었다는 것이다. 한가지 예로써 근원지 번호와 목적지 번호가 같을 경우를 살펴보자.

<그림 9>에서 스테이지 2의 스위칭 소자 3번에서 오류가 발생했기 때문에 그 스위칭 소자를 통해서 원하는 목적지로 전송할 수 없게 되었다고 하자. 이때 근원지 3에서 목적지 3으로 패킷을 전송하려 한다. 목적지 3의 2진 표현이 011이기 때문에, 스테이지 0

에 있는 스위칭 소자 3번에서는 d_0 가 1이므로 패킷을 OUT-1로 출력한다. 그래서 그 패킷은 스테이지 1의 스위칭 소자 3번에 도착한다. 스위칭 소자 3번에서는 $d_1=1$ 이므로 다음 스테이지에 있는 스위칭 소자 3번으로 패킷을 전송해야 하지만 그 스위칭 소자가 오류이므로 스테이지 1의 스위칭 소자 3번의 LO를 통해 같은 스테이지에 있는 스위칭 소자 7번(스위칭 소자 3번과는 듀엣 스위치 임)으로 패킷을 전송한다.

스위칭 소자 7번에서는 $d_1=1$ 이므로 OUT-1로 연결된 스테이지 2의 7번 스위칭 소자로 패킷을 전송한다. 스테이지 2의 스위칭 소자 7번에서는 d_0 가 0이므로 OUT-0로 원하는 목적지 3번으로 전송하여 근원지 3에서 목적지 3으로의 라우팅은 무사히 끝나게된다.

[정리2] Dual 네트워크의 스테이지 $i(0 < i < n)$ 에 있어서 $(N/2^i - 1)$ 개 스위칭 소자의 오류는 허용되어질 수 있다.

(증명) 이를 증명하기 위해 Dual 네트워크의 중복성 방향 그래프를 살펴보자. 마지막 스테이지를 제외한 모든 스테이지에서의 각 노드는 루프로 연결되어 있고, 각 스테이지에 있는 노드들은 다음 스테이지의 노드들과 간선으로써 연결되어 있다. 그러므로 스테이지 $i-1$ 에 있는 임의의 한 노드는 그 스테이지의 루프를 통해 스테이지 i 에 있는 어떠한 노드와도 연결시킬 수가 있다. 따라서 스테이지 i 에서 $N/2^i - 1$ 개의 노드가 오류가 발생한다고 해도 1개의 노드만 있으면, 다음 스테이지로의 연결이 가능하다. 그러므로 스테이지 i 에서 $N/2^i - 1$ 개의 스위칭 소자의 오류는 허용이 가능하다.

[정리2]으로써 2장에서 지적되었던 GN에서의 문제점중 세번째인 마지막 스테이지에 있는 스위칭 소자의 오류도 Dual 네트워크에서는 해결될 수 있음을 알 수 있다. 즉, 마지막 스테이지에서 최소한 1개의 스위칭 소자 오류는 허용될 수 있다.

따라서 GN에서 해결 불능이었던 모든 오류들은 Dual 네트워크에서 모두 해결되었다. 예로써 <그림 9>에서 마지막 스테이지에 있는 스위칭 소자 3번의 오류가 발생했을 경우에 이를 해결할 수 있는 다른 경로가 존재함을 볼 수가 있다.

V. 기존의 MIN들과의 비교

기존의 다른 MIN들과 Dual 네트워크를 비교하기에 앞서, GN과 Dual 네트워크를 비교하여 Dual 네트워크의 특징을 살펴졌다. <표 3>에서 나타나 있듯이 Dual 네트워크는 $\log_2 N$ 개의 스테이지만으로도 오류허용이 가능하며, 목적지 주소방식을 사용하기 때문에 간편하고 빠른 라우팅을 할 수 있다. 또한 $S=D$ 일때의 오류뿐만 아니라 마지막 스테이지에 있는 스위칭소자의 오류도 허용할 수 있어 GN보다 효율적인 네트워크이다.

<표 3> GN과 Dual 네트워크와의 비교

항목	MIN	GN	Dual 네트워크
스테이지수		$\log N + 1$	$\log N$
스위칭소자 크기 (입출력스테이지)		3×3 ($1 \times 3, 3 \times 1$)	3×3 ($2 \times 3, 2 \times 2$)
태그비트수		$2n$	n
라우팅전략		D-S값의 중복 형식 중 한개를 선택	목적지 주소방식
후진추적		오류허용 경로를 찾기위해 필요함	필요없음
마지막스테이지의 입출력 오류허용 여부		오류허용불가함	최소한1개의 오류는 허용됨
$S=D$ 의 경로의 갯수		경로가 1개뿐이어서 오류발생때는 오류허용이 불가능하다	$\prod_{i=1}^n 2^i$ 개의 경로가 존재함

($\log_2 N \Rightarrow \log N$, D는 목적지 번호, S는 근원지 번호)

Dual 네트워크와 기존의 MIN들은 <표 4>와 같이 비교될 수 있다. 중복경로를 제공하는 다른 MIN과는 달리, Dual 네트워크에서는 MIN에서 사용되는 최소

한의 스테이지 수로써 오류를 허용할 수 있다. Dual 네트워크는 목적지 주소로써 스위칭 소자를 개별 제어하므로 중앙 제어에서 발생할 수 있는 잦은 오류검진과 경로 설정을 위한 계산 시간이 필요없게 되어, 라우팅 제어가 보다 간편해지고 빨라진다. 또한, 오류가 발생하되 태그의 수정이 필요없기 때문에 기존의 어느 네트워크보다 효율적인 네트워크이다.

VI. 결 론

많은 수의 중복경로를 제공하는 GN에서 한개의 오류로 인하여 특정 근원지와 목적지간의 연결이 불가능하게 된다는 것은 UPP MIN과 다를바가 없다. 또한 GN에서는 오류가 발생했을 경우 중복 경로를 찾기위해 후진추적을 감수해야 하기 때문에, 고속을 요하는 병렬처리 시스템에서는 커다란 오버헤드가 아닐 수 없다.

따라서 본 논문에서는 이러한 GN의 문제점을 분석한 후, 그 원인들을 발견하고 또한 그 원인이 되는 요인들을 'Dual'개념을 이용해 위의 문제들을 해결할 수 있는 새로운 Dual 네트워크를 설계하였다. Dual 네트워크에서는 라우팅 전략으로 목적지 주소 라우팅 전략을 사용하므로 라우팅 제어가 간편하고 오류허용시에도 태그의 수정이 필요 없어 보다 효율적인 네트워크라고 말할 수 있다. 특히, Dual 네트워크는 스테이지 수가 $\log_2 N$ 개이므로 MIN에서 사용되는 최소한의 스테이지 수로써 오류를 허용할 수 있기 때문에 성능면에서도 우수하다고 할 수 있을 뿐만 아니라, 하드웨어적인 복잡도(hardware complexity)면에 있어서 기존의 많은 중복 경로를 가지고 있는 GN 또는 Extra Stage GN보다 적은 수의 스테이지로써 비용의 절감도 피하였다.

<표 4> 기존의 MIN과 Dual 네트워크와의 비교

항목	MIN	Benes Network	Augment Shuffle Exchange Network	GN	Dual
스테이지 수		$2\log N - 1$	$\log N$	$\log N + 1$	$\log N$
스위칭 소자 크기		2×2	3×3	3×3	스테이지 1,2에서는 2×2 , 중간스테이지에서는 3×3
라우팅 전략		looping Algo.	목적지 주소방식	근원지, 목적지	목적지주소의 2진역 표현
중복 경로 여부		있 음	있 음	존재(단 근원지와 목적지가 같을 경우 없음)	있 음
오류허용을위한라우팅태그의 수정 여부		필 요	필요없음	필 요	필요없음
스위칭 제어		중앙제어	개별제어	중앙제어	개별제어

참 고 문 헌

1. 김성천, "상호 연결 네트워크와 MIN(Multistage Interconnection Network)," 정보과학회지, 제6권, 제6호, pp.39-48, Dec. 1988.
2. G.B.Adams and H.J.Siegel, "The Extra stage Cube: A Fault-Tolerant Interconnection Network for Supersystems," IEEE Trans. Compt., Vol.C-31, pp.443-454, May 1982.
3. D.P.Agrawal, "Graph Theoretical Analysis and Design of Multistage Interconnection Networks," IEEE Trans. Compt., pp.637-648, July, 1983.
4. T.Feng, "A Survey of Interconnection Networks," IEEE Compt., pp.12-27, Dec. 1981.
5. T.Feng, "Data Manipulating Functions in Parallel Processors and Their Implementations," IEEE Trans. Compt., pp.309-318, Mar. 1974.
6. K.Hwang and F.Briggs, Computer Architecture and Parallel Processing, McGraw Hill, 1984.
7. S.C.Kothari, "Multistage Interconnection Networks for Multiprocessor Systems," Advances in Compt. Vol.26, pp.155-199, 1987.
8. V.P.Kumar and S.M.Reddy, "Augmented Shuffle-Exchange Multistage Interconnection Networks," IEEE Compt., pp.30-40, June, 1987.
9. K.Y.Lee and W.Hegazy, "The Extra Stage Gamma Network," IEEE, Trans. Compt., Vol. 37, No.11, pp.1445-1450, Nov. 1988.
10. K.Y.Lee and D.Lee, "On The Augmented Data Manipulator Network in SIMD Environments," IEEE Trans. Compt., pp.574-584, May, 1988.
11. K.Y.Lee and H.Yoon, "The PM221 Interconnection Network," IEEE Trans. Compt., pp. 302-307, Feb. 1989.
12. R.J.Mcmillen and H.J.Siegel, "Routing Schemes for the Augmented Data Manipulator Network in an MIMD System," IEEE Trans. Compt., pp.184-196, Dec. 1982.
13. M.D.Palmer Leland, "On the Power of Augmented Data Manipulator Network," in Proc. 1985, Int'l Conf. Parallel Processing, pp.74-78, Aug. 1985.
14. D.S.Parker and C.S.Raghavendra, "The Gamma Network," IEEE Trans. Compt., pp.367-373, Apr. 1984.
15. D.Rau and H.J.Siegel, "Destination Tag Routing Techniques Based on State Model for the IADM Network," the 15th Annu. Int'l. Symp. on Compt. Arch., pp.318-324, 1988.
16. H.J.Siegel, "The Theory Underlying the Partitioning of Permutation Networks," IEEE Trans. Compt., Vol.C-29, No.9, pp.791-801, Sept. 1980.
17. Interconnection Networks for Large-Scale Parallel Processing, D.C.Heath and Company, 1985.
18. H.J.Siegel and R.J.McMillen, "The Mutistage Cube: A Versatile Interconnection Network," IEEE Compt., pp.65-76, Dec. 1981.
19. N.F.Tzeng, P.C. Yew, and C.Q.ZHU, "Realizing Fault-Tolerant Interconnection Networks Via Chaning," IEEE Trans. Compt., Vol.37, No.4, pp.458-462, Apr. 1988.
20. A.Varma and C.S.Raghavendra, "Fault-Tolerant Routing in Multistage Interconnection Networks," IEEE Trans. Compt., pp.385-393, Mar. 1989.
21. Jon-Seok Han and Tack-Don Han, "A Fault Tolerant Design of the Gamma Network," Proc. of the Fouth ISMM int'n Conf. on Parallel and Distributed Computing and Systems, 1992 pp.30-34.



최 창 훈(Chang Hoon Choi) 정회원

1988년 2월 : 명지대학교 전자계산학과 졸업

1990년 2월 : 서강대학교 대학원 전자계산학과(공학석사) 졸업

1990년 1월 ~ 9월 : 대우통신 기술개발부 근무

1992년 ~ 현재 : 서강대학교 대학원 전자계산학과 박사과정 재학중

※주관심분야 : Computer Architecture, parallel processing system



金 聖 天(Sung Chun Kim)정회원

1975년 : 서울대학교 공과대학 공업교육학(전기전공)학사

1976년 ~ 1977년 : 동아컴퓨터(주) Sys. Eng.

1977년 ~ 1978년 : 스펀리유니맥 Sales Rep.

1979년 : Wayne State Univ. 컴퓨터공학 석사

1982년 : Wayne State Univ. 컴퓨터공학 박사

1982년 ~ 1984년 : 캘리포니아주립대 조교수

1984년 ~ 1985년 : 금성반도체(주) 책임연구원

1986년 ~ 1989년 : 서강대학교 공과대학 전자계산소 부소장

1989년 ~ 1991년 : 서강대학교 공과대학 전자계산학과 학과장

1985년 ~ 현재 : 서강대학교 공과대학 전자계산학과 조교수(1985. 8 ~ 1987. 8), 부교수(1987. 9 ~ 1992. 8), 교수(1992. 9 ~ 현재)

1989년 ~ 현재 : 한국정보과학회 병렬처리시스템 연구회 부위원장, 대한전자공학회 및 한국통신학회 논문지 편집위원(1991, 1993), 한국정보과학회 학회지 부위원장(1993)

※주관심분야 : 병렬처리시스템(Parallel Computer Architecture, Interconnection Network), Neural Network, Computer Network