울산대학교 공학연구논문집 제30권 2호 pp. 567~583, 울산대학교.
Journal of Engineering Research Vol. 30, No. 2, pp. 567~583, 1999
University of Ulsan.

Buffered 2×2 스위치들로 구성된 다층 연결 망의 Throughput 분석*

양명국
전기 전자 및 자동화 공학부

<요 약>

본 논문에서는, multiple buffered 2×2 크로스바 스위치들로 구성된 다층 연결 망 (Multistage Interconnection Network, MIN)의 성능 예측 모형을 제안하고, 스위치에 장착된 buffer의 개수 증가에 따른 성능 향상 추이를 분석하였다. Buffered 스위치 기법은 다층 연결 망 내부의 데이터 충돌 문제를 효과적으로 해결할 수 있는 방법으로 널리 알려져 있다. 제안한 성능 예측 모형은 먼저 네트워 내부 임의 스위치 입력 단에 유입되는 데이터 패킷이 buffered 스위치 내부에서 전송되는 유형을 확률적으로 분석하여 수립되었다. 확률 분석 과정의 수학적 복잡도 절감을 위하여 임의 싸이클 동안 buffer에 저장된 데이터 패킷 관련 확률식 유도 과정 등에 steady state probability 개념을 도입하였다. 제안한 모형은 스위치에 장착된 buffer의 개수와 무관하게 multiple buffered 2×2 크로스바 스위치의 성능 예측이 가능하고, 나아가서 이들로 구성된 모든 종류의 다층 연결 망 성능 분석에 적용이 용이하다. 제안한 수학적 성능 분석 연구의 실효성 검증을 위하여 병행된 시뮬레이션 결과는 상호 4% 이내의 미세한 오차 범위 내에서 모형의 예측 데이터와 일치하는 결과를 보여 분석 모형의 타당성을 입증하였다. 또한, 분석 결과 두 개 혹은 네 개의 데이터 패킷을 저장할 수 있는 buffer를 장착한 multiple buffered 2×2 크로스바 스위치들로 설계된 8×8 Baseline 네트워은 각각 83% 및 90%의 Throughput을 제공하는 것으로 관찰되었고, 여덟 개 이상 데이터 패킷을 저장할 수 있는 buffer를 장착한 경우 buffer 개수 증가에 따른 성능 향상 율이 둔화되는 것으로 나타났다. 따라서 buffered 2×2 크로스바 스위치들로 설계된 8×8 Baseline 네트워의 경우 두 개에서 네 개 가량의 데이터 패킷을 저장할 수 있는 buffer를 스위치에 장착 시키는 것이 효율적인 것으로 추론 되었다.

* 이 논문은 1998학년도 울산대학교 학술연구조성비에 의하여 연구되었음.

Throughput Evaluation of a Multistage Interconnection Network with Buffered 2×2 Switches

Myung K. Yang
School of Electrical, Electronic Engineering and Automation

<Abstract>

In this paper, a throughput evaluation model of the Multistage Interconnection Networks(MIN) with the multiple buffered *2×2* crossbar switches is proposed and examined. Buffered switch technique is well known to solve the data collision problems of the MIN. The proposed evaluation model is investigated by analyzing the transfer patterns of the data packets that are arrived at the input ports of a switch element in the network. Steady state probability concept is used to simplify the analysis processes. The model not only estimates the performance of the multiple buffered *2×2* crossbar switch with various size of buffers but also can be applied to evaluate MINs which are designed with buffered *2×2* crossbar switches. To validate the proposed evaluation model, the simulation is carried out on a Baseline network that uses the multiple buffered *2×2* crossbar switches. Less than 2% error between analysis and simulation results is observed. It is also shown that the *8×8* Baseline networks designed by buffered *2×2* crossbar switches provide 83%, and 90% of throughputs for buffer size of two, and four data packets, respectively. The throughput elevation is significantly reduced as the buffer size increases. This reveals that two to four buffers are optimal for the *8×8* Baseline network with multiple buffered *2×2* crossbar switches.

I. 서 론

WAN 으로부터 LAN에 이르기까지, 그리고 각종 병렬 컴퓨터 등의 상호 연결 기법으로 제안된 다층 연결 망(Multistage Interconnection Network, MIN)은 다양한 연결 망 기법 가운데 Crossbar 네트웍과 함께 넓은 Bandwidth, 네트웍 유연성 등의 장점을 보유한 효율적인 네트웍으로 평가되고있다. 다층 연결 망은 작은 스위치 소자를 단계(stage)별로 배열하고 이를 주어진 패턴으로 연결하여 스위치 소자간의 연결 루트를 형성하고, 시스템 크기에 따라 Stage 수를 조정하여 전체 상호 연결 망을 형성한다. 따라서 다층 연결 망을 통한 데이터 이동에는 각 Stage 에서 스위치 마다 제어가 요구되고, 데이터 이동 경로에 따라 특정 스위치에서 두개 이상의 데이터가 하나의 경로로 진행하고자 하는 데이터 충돌 현상이 초래하기도 한다. 데이터 충돌 현상은 네트웍 성능 저하를 유발함은 물론이고 전체 네트웍의 신뢰도에도 큰 영향을 미치게 된다. 다층 연결 망 내부의 데이터 충돌로 인한 문제를 해결하고 네트웍 Bandwidth 확장을 위하여 변형 스위치 소자 설계[1], 추가 Stage 삽입, 그리고 스위치 소자에

Buffer 장착[2-5] 등 다각도에서 연구가 진행되고있다. 본 논문에서는 앞서 나열한 방안 가운데 Buffered 스위치를 이용하여 데이터 충돌 문제를 해결한 다층 연결 망의 성능을 분석하였다.

Dias 와 Jump[2]는 한 개의 Buffer 를 장착한 스위치들로 구성된 단일 buffered(single buffered) Baseline 네트웍의 성능을 분석하였다. 수학적 분석과 시뮬레이션 결과를 통하여 Dias 와 Jump 는 buffered 다층 연결 망이 Crossbar 네트웍과 유사한 수준의 Bandwidth 를 제공하는 것으로 보고하였다. Jenq[3]는 단일 buffered Banyan 네트웍을 대상으로 분석 모형을 제시하고, 네트웍 throughput, delay, 및 internal blocking probability 등을 분석하였다. 또한, Krusal 과 Snir[4]는 unbuffered 및 무한 buffered (infinite buffered) Banyan 네트웍의 성능 분석 모형을 제시하고, 시뮬레이션과 수학적 모형 해석을 통하여 각 단(stage)별 대기시간 등 데이터 패킷 이동에 buffer 가 제공하는 영향을 연구하였다.

앞서 기술한 기존의 연구는 단일 buffer 혹은 무한 buffer 의 경우만을 대상으로 분석 모형을 제안하고 수학적 분석을 수행한 반면 복수 buffer(multiple buffers) 환경에 관한 부분에 대하여는 분석 모형 해석의 난이성으로 인하여 시뮬레이션을 통하여 성능 예측을 시도하였다. Yoon, Lee, 그리고 Liu[5]는 이와 같은 기존 연구의 문제점을 보완하여 임의 크기의 buffer 를 장착한 복수 buffered $N\times N$ Baseline 네트웍의 분석 모형을 제안하고, 모형의 실효성을 입증하였다. 단일 buffered 네트웍의 해석 모형을 확장하는 개념으로 제안된 Yoon 등의 복수 buffered $N\times N$ Baseline 네트웍 분석 모형은 실제 네트웍 상의 데이터 이동 패턴을 그대로 상태 변환도로 전환하고 이를 수식 화하여 설계되어, 연산 과정이 복잡하고 모형의 이해가 난해하다.

본 논문에서는 복수 Buffered 2×2 스위치들로 구성된 다층 연결 망의 Throughput 분석 모형을 제안하고, 성능을 평가 하였다. 먼저 Baseline 네트웍을 대상으로 네트웍 내부 스위치에서 발생하는 데이터 충돌에 대한 가능한 처리 기법들을 검토하고, 이를 기반으로 네트웍의 실질 성능에 영향을 미치지 않는 범위에서 네트웍 성능 분석이 용이하도록 수정된 데이터 충돌 처리 기법을 구상하였다. 이어서 복수 buffered 2×2 스위치의 분석 모형을 제시하고, 이를 다층 연결 망에 적용하여 buffer 개수 증가에 따른 네트웍 성능 변화를 평가하였다. 제안된 분석 모형은 Buffer 를 장착한 스위치 소자 내부의 데이터 이동 패턴을 확률적으로 해석하여 개발되었다. 본 논문에 제안된 복수 buffered 다층 연결 망의 성능 분석 모형은 스위치에 장착된 buffer 의 개수와 무관하게 적용 가능하고, 분석 과정에서 정상상태확률(Steady state probability) 개념을 도입하여 모형의 수식 이해가 용이하도록 하였다. 제안한 수학적 성능 분석 연구의 실효성 검증을 위하여 병행된 시뮬레이션 처리 결과는 상호 미세한 오차 범위 내에서 모형의 예측 데이터와 일치하는 결과를 보여 분석 모형의 타당성을 입증하였다. 또한 Baseline 네트웍을 대상으로 제안된 본 연구의 분석 모형은 모든 다층 연결 망의 성능 분석에 확대 적용 가능하다.

본 논문의 구성은 다음과 같다. 먼저, 서론에 이어 II 절에서는 다양한 다층 연결 망 구조와 함께 다층 연결 망 내부 스위치에서의 데이터 충돌 처리방식을 기술하였

다. III 절에서는 복수 Buffered 2×2 스위치로 구성된 다층 연결 망의 Buffer 크기에 따른 네트웍 Throughput 변화 추이를 예측할 수 있는 새로운 성능 분석 모형을 제시하고, 모형의 실효성을 검증하였다. 끝으로 본 연구의 성과와 결과를 마지막 절에 요약 기술하였다.

II. 다층 연결 망의 충돌 처리 방식

다층 연결 망(Multistage Interconnection Network, MIN)은 스위치 연결 방식에 따라 Data manipulator[6], Flip network[7], Omega network[8], Banyan network[9], Butterfly network[10], 그리고 Baseline network[11] 등으로 구분되어 각각의 특성에 따른 응용 분야와 함께 연구, 발표되었다. 이들 네트웍들은 비교적 넓은 Bandwidth, 네트웍 유연성 등의 장점을 기반으로 NYU Ultracomputer[12], IBM RP3[13], 및 BBN Butterfly GP1000[10], TC2000[14]등 다양한 병렬 컴퓨터의 상호 연결 망으로 활용되었다. 그러나 MIN은 데이터 전송을 위하여 구조상 여러 단(stage)의 스위치를 경유하여야 하고, 이 과정에서 여러 개의 데이터가 동시에 네트웍에 유입될 경우 전송 경로에 따라 특정 스위치에서 데이터 충돌이 불가피하게 된다. 앞서 열거된 병렬 컴퓨터들은 이와 같은 데이터 충돌 현상으로 인한 네트웍 성능 저하 효과를 줄이기 위하여 통신 패턴을 국지화 하여 네트웍 소통량을 줄이는 방안[10,14], 복수 MIN module 설정 방안[13], 그리고 스위치에 buffer를 장착하는 방안[12-14] 등을 채택하고 있다.

본 절에서는, 먼저 성능 변화 예측 모형 개발을 위하여 요구되는 다층 연결 망의 구조를 살펴보고, buffer를 장착한 스위치의 데이터 충돌 처리 방식을 기술하였다.

II.1. 다층 연결 망(MIN)의 구조

다층 연결 망은 시스템의 크기, 사용하는 스위치, 그리고 각 스위치 스테이지 간의 연결 패턴에 따라 구조가 결정된다. 예를 들어, N 개의 입력과 N 개의 출력을 갖는 $N\times N$ 다층 연결 망은, $a\times a$ crossbar 스위치들을 사용하여 설계할 경우, $\log_a N$ 스테이지로 구성된다. 각 스테이지는 N/a 개의 스위치들을 포함하고, 각 스테이지 스위치들을 일정한 연결 패턴에 따라 연결함으로써 네트웍을 구성하게 된다.

앞서 기술한 바와 같이 다층 연결 망은 각 스위치 스테이지간의 독특한 연결 방식에 따라 다양한 네트웍으로 구분된다. 그림 1은 기존에 연구 발표된 다층 연결 망들의 종류별 다양한 연결 패턴을 보여주고 있다. Wu와 Feng[18]은 이들 다층 연결 망들의 연결 구조와 특성을 분석하여, 각기 다른 형태로 설계된 기존의 네트웍들이 연결 패턴의 차이에도 불구하고, 구조적으로 동일함을 입증하고, 성능의 차이가 없음을 밝혔다. 따라서 본 논문에서는 복수 buffered 2×2 크로스바 스위치들을 사용한 8×8 Baseline 네트웍을 대상으로 성능을 분석하였고, 여기서 얻어진 결과는 다른 종류의 다층 연결 망에도 적용이 가능하다.

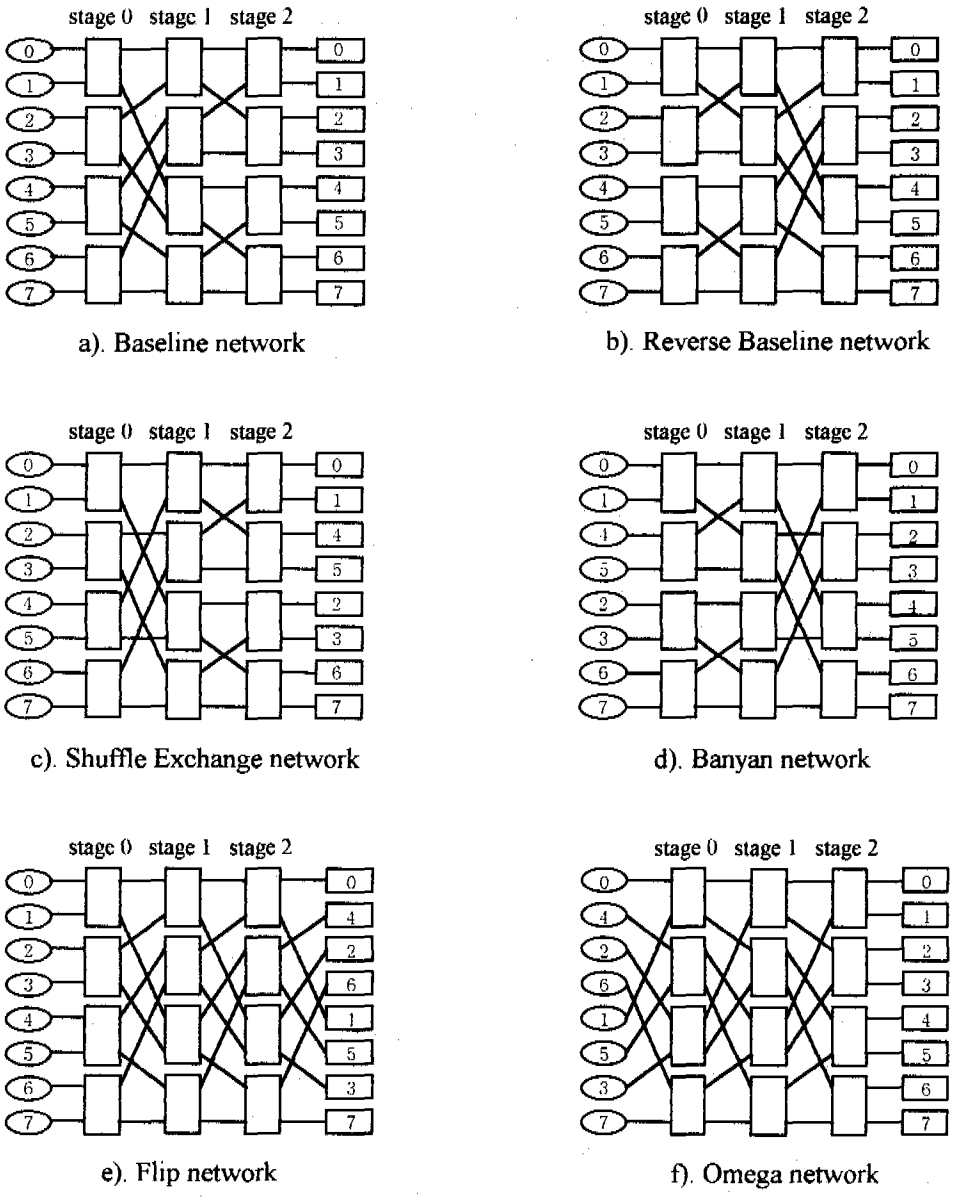


그림 1. 다중 연결 망

II.2. 충돌 처리 방식

다층 연결 망은 작은 스위치 소자를 스테이지별로 배열하고, 이를 주어진 패턴으로 연결하여 스위치 소자간의 연결 루트를 형성함으로 전체 상호 연결 망을 형성한다. 따라서 다층 연결 망에 유입된 데이터는 데이터 이동 경로에 따라 네트웍 내부 특정 스위치에서 두개 이상

의 데이터가 하나의 출력 단자(port)로 진행하고자 하는 데이터 충돌 현상이 초래하기도 한다. 스위치에 buffer를 장착하여 충돌에 관련된 데이터 패킷을 임시로 저장하여 데이터 손실을 줄이는 buffered 스위치 기법은 데이터 충돌 문제를 비교적 효율적으로 해결하는 방안으로 알려져 있다. 2x2 스위치 내부의 데이터 충돌 처리 시나리오를 buffer가 장착된 경우와 장착되지 않은 경우로 구분하여 기술하면 다음과 같다.

Case 1). 스위치에 buffer가 장착되지 않은 경우:
데이터 충돌 발생 시 무작위 선택 방식 혹은 특정 우선 순위에 의거 선택된 한 개의 데이터 패킷은 스위치를 통과하여 이동시키고, 나머지 데이터 패킷은 제거(reject)한다.(네트웍 프로토콜에 따라 필요 시 제거 대상 데이터 패킷이 유입된 네트웍 입력 단에 통보)

Case 2). 스위치에 buffer가 장착된 경우:

① buffer가 데이터 충돌에 관련된 데이터 패킷을 모두 수용할 수 있는 경우: 해당 싸이클에 스위치를 통과 하지 못한 데이터 패킷은 일단 buffer에 저장하고, 향후 순서에 따라 다음 스테이지로 전달한다. 따라서, 스위치에 buffer가 장착되지 않은 경우와 달리, 충돌로 인한 정보의 중도 유실을 방지할 수 있게 된다.

② buffer가 데이터 충돌에 관련된 데이터 패킷을 모두 수용할 수 없는 경우: 무작위 선택 방식 혹은 특정 우선 순위에 의거 가용 buffer 용량에 따라 선택된 데이터 패킷은 buffer에 저장하고, 나머지 데이터 패킷은 다음 두 가지 방법 가운데 한 방법으로 처리된다.

- ✓ 선택에서 제외된 데이터 패킷을 제거.(네트웍 프로토콜에 따라, 필요 시 제거 대상 데이터 패킷이 유입된 네트웍 입력 단에 통보)
- ✓ 해당 데이터 패킷을 이전 단계(stage)의 buffer에 저장.(이전 단계로부터 이동을 원천 봉쇄).

앞서 기술한 데이터 충돌에 대한 처리 방식 가운데 Case 1)과 Case 2)-①은 선택의 여지가 없는 반면, Case 2)-②의 경우 두 가지 처리 기법이 가용하고, 이들 가운데 두 번째 방법이 충돌로 인한 정보의 중도 유실을 방지할 수 있어 신뢰도 측면에서 좀더 효율적일 것으로 추론 된다. 그러나 해당 패킷을 이전 단계로부터 이동을 원천 봉쇄하기 위하여는 매 싸이클 마다 다음 스테이지의 buffer 상태 점검이 요구되고, 이로 인한 성능 저하를 초래하게 된다.

네트웍 성능 평가 시, 네트웍 각 입력 단들에 무작위 생성된 출력 단 주소를 가진 데이터 패킷을 유입하여 네트웍이 이들 데이터 패킷을 출력 단으로 전송하는 율(Throughput)을 계산하게 되는바, 이와 같은 환경에서 Case 2)-②에 제시된 두 가지 처리 방법이 제공하는 성능의 차이는 무시해도 무방한 정도일 것으로 예측되고, 이는 시뮬레이션을 통하여 입증되었다. 따라서 본 논문에서는 성능 분석 모형을 개발하는 단계에서 데이터 충돌 발생 시 처리 방법으로 Case 2)-②의 경우 분석이 용이한 첫번째 방법 '해당 출력 단자로 향하는 데이터 패킷을 제거하고, 해당 데이터 패킷이 유입된 네트웍 입력 단에 통보'을 채택하였다.

III. Buffered 다층 연결 망의 성능 분석

본 절에서는 다층 연결 망의 스위치에 장착된 buffer가 네트웍 성능에 미치는 영향을 분석 예측할 수 있는 수학적 분석 모형을 기술하였다. 먼저 분석 모형 개발에 적용된 네트웍 환경에 대한 일반적인 가정을 정리하고, 각 buffered 스위치 내부의 데이터 이동 패턴을 확률적으로 해석하였다. 끝으로 각 스위치 내부의 데이터 이동 확률을 토대로 buffered 스위치를 장착한 다층 연결 망 성능 평가의 주요 요소로 거론되어지는 네트웍 Throughput에 대한 새로운 성능 분석 기법을 기술하였다.

III.1. 네트웍 환경에 대한 일반적인 가정

복수 buffered 다층 연결 망의 분석 모형 개발과 시뮬레이션을 위해 본 논문에 적용된 일반적인 가정을 정리하면 다음과 같다.

- multiple-buffered 2×2 crossbar 스위치들을 사용한 8×8 Baseline 네트웍을 분석 대상으로 한다. Wu와 Feng[18]의 연구에서 밝혀진 바와 같이 기존의 모든 다층 연결 망은 기하학적으로 동일하여, 본 연구에서 baseline 네트웍을 대상으로 얻은 성능 분석 결과 및 분석 모형은 다른 모든 다층 연결 망의 해석에 활용이 가능하다.
- 네트웍은 스위치 클럭 싸이클에 따라 동기적으로 작동한다. 즉, 네트웍 내부 데이터 패킷은 스위치 클럭 Δt 동안 임의 스위치 출력 단을 출발, 다음 스테이지 스위치를 통과하여 해당 출력 단에 도달한다.
- 스위치에 장착된 buffer는 스위치 출력 단에 위치하고, buffer 공간 하나는 한 개의 데이터 패킷을 수용할 수 있다.
- 네트웍 각 입력 단으로 매 싸이클 마다 일정 크기의 데이터 패킷이 유입될 확률을 $\zeta_{stage\ 0}$라하고, 네트웍 내부 임의 스테이지 i에 위치한 스위치 입력 단으로 데이터 패킷이 유입될 확률은 $\zeta_{stage\ i}$라 한다. 따라서 매 싸이클 마다 네트웍 각 입력 단에 한 개씩의 데이터 패킷이 유입될 경우, $\zeta_{stage\ 0} = 1$ 이 된다.
- 네트웍 입력 단으로 유입되는 데이터 패킷의 네트웍 최종 출력 단 행선지는 무작위 선택 방식에 의거 결정된다.
- 데이터 충돌 발생 시 무작위 중재 방식에 의거 데이터 처리 우선 순위를 결정한다.

본 절에 기술한 가정은 기존의 네트웍 성능 평가 연구[3-5]에 보편적으로 적용되었다.

III.2. 데이터 이동 패턴

네트웍 내부 임의 2×2 crossbar 스위치 입력 단에 유입된 데이터 패킷은 데이터가 지향하는 행선지에 따라 스위치의 두 개 출력 단 중 어느 한 출력 단으로 향하게 된다. III. 1.의 가정에

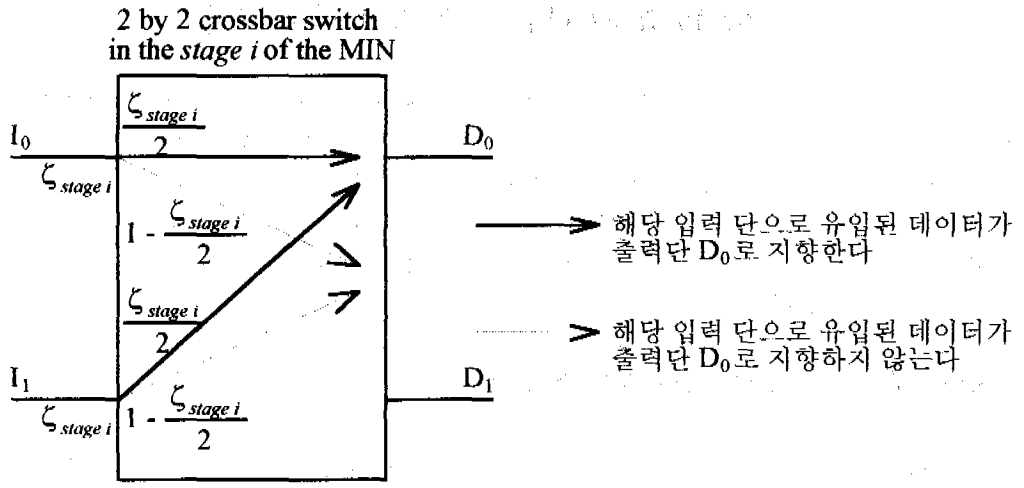


그림 2. 네트웍 내부 임의 스위치에서 데이터 패킷의 이동 확률

의거, 네트웍 입력 단에 처음 데이터 패킷이 유입될 때 최종 출력 단 행선지가 무작위 선택 방식에 의거 주어짐으로, 임의 스위치 입력 단에 도착한 데이터 패킷이 어느 한 출력 단으로 향하게 될 확률은 스위치 입력 단에 데이터 패킷이 유입될 확률의 반으로 계산된다. 즉, 네트웍 스테이지 i에 위치한 임의 스위치 입력 단에 데이터 패킷이 유입될 확률이 $\zeta_{stage\,i}$로 주어지면 해당 스위치 의 어느 한 출력 단으로 데이터 패킷이 향할 확률은 $\zeta_{stage\,i}/2$가 된다. 그림 2는 네트웍 내부 임의 스위치에서 데이터 패킷이 이동하는 패턴을 확률적으로 해석하여 도식화한 것이다.

네트웍 내부 스테이지 i에 위치한 임의 스위치 입력 단 I_0에 데이터 패킷이 유입될 확률이 $\zeta_{stage\,i}$로 주어지면, 그로 인하여 해당 스위치의 특정 출력 단 D_0로 데이터 패킷이 향할 확률은 $\zeta_{stage\,i}/2$가 되고, 특정 출력 단 D_0로 데이터 패킷이 향하지 않을 확률은 $(1-\zeta_{stage\,i}/2)$가 된다. 따라서, 스테이지 i에 위치한 임의 스위치 출력 단 D_0로 두 개의 데이터 패킷이 지향할 확률, $P(h=2)_i$은

$$P(h=2)_i = (\zeta_{stage\,i}/2)^2 \tag{1}$$

로 계산된다.

스위치 출력 단 D_0로 한 개의 데이터 패킷이 향할 경우는 먼저 스위치 입력 단 I_0에 유입된 데이터 패킷이 D_0로 향하고 스위치 입력 단 I_1에 데이터가 유입되지않거나, I_1에 유입된 데이터 패킷이 D_0로 향하지 않는 경우, 혹은 스위치 입력 단 I_1에 유입된 데이터 패킷이 D_0로 향하고 스위치 입력 단 I_0에 데이터가 유입되지않거나, I_0에 유입된 데이터 패킷이 D_0로 향하지 않는 경우 등 두 가지 경우로 구분된다. 따라서, 스테이지 i에 위치한 임의 스위치 출력 단 D_0로 한 개의 데이터 패킷이 지향할 확률, $P(h=1)_i$은

$$P(\hbar = 1)_i = {}_2C_1 \times (\zeta_{stage\,i} / 2) \times (1 - \zeta_{stage\,i} / 2) \quad (2)$$

이다.

스위치 출력 단 D_0로 한 개의 데이터 패킷도 향하지 않을 경우는 스위치 입력 단 I_0와 I_1으로 유입된 데이터 패킷이 없거나 혹은 스위치 입력 단 I_0와 I_1에 유입된 데이터 패킷이 D_0로 향하지 않는 경우이다. 따라서, 스테이지 i에 위치한 임의 스위치 출력 단 D_0로 한 개의 데이터 패킷도 지향하지 않을 확률, $P(\hbar = 0)_i$은

$$P(\hbar = 0)_i = (1 - \zeta_{stage\,i} / 2)^2 \quad (3)$$

와 같이 얻어진다.

네트웍 임의 스테이지 i에 위치한 2×2 crossbar 스위치 내부 데이터 이동 패턴은 확률적으로 그림 2와 식 (1), (2), 그리고 (3)과 같이 해석된다.

III.2. Buffered 다층 연결 망의 성능 분석

네트웍 내부 스테이지 i에 위치한 임의 2×2 crossbar 스위치 내부 데이터 이동 패턴의 확률적 분석을 토대로 Buffered 다층 연결 망의 성능 분석을 위하여 사용될 변수는 다음과 같다.

b	: 스위치에 장착된 buffer가 저장할 수 있는 데이터 패킷 수
ε	: buffer에 저장된 데이터 패킷 수
$P(\varepsilon = k)_i$	: buffer에 저장된 데이터 패킷 수가 k 개일 확률
$P(D_\gamma = 1)_i$	: 출력 단 D_γ로 데이터 패킷이 출력될 확률
$P(D_\gamma = 0)_i$	: 출력 단 D_γ로 데이터 패킷이 출력되지 않을 확률

스위치 출력 단 D_γ로 데이터 패킷이 출력되는 경우는 해당 출력 단 buffer에 데이터 패킷 저장된 경우, 혹은 스위치 입력 단에 새로이 유입된 데이터 패킷이 해당 출력 단으로 지향할 경우이다. 반대로 스위치 출력 단 D_γ로 데이터 패킷이 출력되지 않는 경우는 해당 출력 단 buffer에 데이터 패킷 저장되지 않은 상태에서, 스위치 입력 단에서 해당 출력 단으로 지향하는 데이터 패킷이 없는 경우이다. 확률 계산 측면에서 스위치 출력 단 D_γ로 데이터 패킷이 출력되는 경우의 확률, $P(D_\gamma = 1)_i$ 보다는 스위치 출력 단 D_γ로 데이터 패킷이 출력되지 않는 경우의 확률, $P(D_\gamma = 0)_i$ 계산이 보다 용이하다. 따라서 임의 싸이클 j에 스위치 출력 단 D_γ로 데이터 패킷이 출력되지 않을 확률, $P(D_\gamma = 0)_{i,\,cycle\,j}$을 구하면

$$P(D_\gamma = 0)_{i,\,cycle\,j} = P(\varepsilon = 0)_{i,\,cycle\,(j-1)} \times P(\hbar = 0)_{i,\,cycle\,j} \quad (4)$$

이 된다. 여기서 $j \geq b$ 이다. 식 (4)는 싸이클 $(j - k)$에 해당 출력 단 buffer가 k 개의 데이터 패킷을 저장하고 있다면, 이후 k번 싸이클 동안 연속 해당 출력 단 D_γ로 지향하는 데이터 패킷이

없고 또한 싸이클 j에서도 D_γ로 지향하는 데이터 패킷이 없어야 비로서 싸이클 j에 D_γ로 데이터 패킷이 출력되지 않는다는 것을 식으로 표현하였다. 따라서, 임의 싸이클 j에 스위치 출력 단 D_γ로 데이터 패킷이 출력 될 확률, $P(D_\gamma = 1)_{i,\ cycle\ j}$은

$$P(D_\gamma = 1)_{i,\ cycle\ j} = 1 - P(D_\gamma = 0)_{i,\ cycle\ j} = 1 - \{P(\varepsilon = 0)_{i,\ cycle\ (j-1)} \times P(\hbar = 0)_{i,\ cycle\ j}\} \tag{5}$$

로 계산된다. 식 (4)와 (5)에서 $P(\hbar = 0)$는 식 (3)에서 얻을 수 있고, $P(\varepsilon = 0)_{i,\ cycle\ (j-1)}$은 다음과 같이 계산된다. 먼저 싸이클 $(j-1)$ 종료 시점에 buffer가 저장하고있을 데이터 패킷의 수가 0일 경우는 다음과 같다.

- 싸이클 $(j-2)$ 종료 시 buffer에 저장된 데이터 패킷의 수가 하나이고, 싸이클 $(j-1)$에 해당 출력 단으로 향하는 데이터 패킷이 없는 경우: 이때 buffer에 저장 되었던 데이터 패킷은 싸이클 $(j-1)$에 해당 출력 단을 지나 다음 스테이지의 스위치 입력단을 향하고 buffer는 비게 된다.
- 싸이클 $(j-2)$ 종료 시 buffer에 저장된 데이터 패킷이 없고, 싸이클 $(j-1)$에 해당 출력 단으로 향하는 데이터 패킷이 하나인 경우: 해당 출력 단으로 향하는 데이터 패킷은 buffer에 저장되지 않고 그대로 출력 단으로 출력된다.
- 싸이클 $(j-2)$ 종료 시 buffer에 저장된 데이터 패킷이 없고, 싸이클 $(j-1)$에 해당 출력 단으로 향하는 데이터 패킷이 없는 경우

따라서, 임의 싸이클 $(j-1)$에 buffer에 저장된 데이터 패킷의 수가 0일 확률, $P(\varepsilon = 0)_{i,\ cycle\ (j-1)}$은

$$\begin{aligned} P(\varepsilon = 0)_{i,\ cycle\ (j-1)} &= P(\varepsilon = 1)_{i,\ cycle\ (j-2)} \times P(\hbar = 0)_{i,\ cycle\ (j-1)} \\ &+ P(\varepsilon = 0)_{i,\ cycle\ (j-2)} \times P(\hbar = 1)_{i,\ cycle\ (j-1)} \\ &+ P(\varepsilon = 0)_{i,\ cycle\ (j-2)} \times P(\hbar = 0)_{i,\ cycle\ (j-1)} \end{aligned} \tag{6}$$

로 계산된다. 식(6)의 $P(\varepsilon = 1)_{i,\ cycle\ (j-2)}$는 싸이클 $(j-2)$ 종료 시 buffer에 1개의 데이터 패킷이 저장될 확률이다. 이를 $P(\varepsilon = 0)_{i,\ cycle\ (j-1)}$ 분석과 유사한 과정을 거쳐 확률식으로 표현하면,

$$\begin{aligned} P(\varepsilon = 1)_{i,\ cycle\ (j-2)} &= P(\varepsilon = 2)_{i,\ cycle\ (j-3)} \times P(\hbar = 0)_{i,\ cycle\ (j-2)} \\ &+ P(\varepsilon = 1)_{i,\ cycle\ (j-3)} \times P(\hbar = 1)_{i,\ cycle\ (j-2)} \\ &+ P(\varepsilon = 0)_{i,\ cycle\ (j-3)} \times P(\hbar = 2)_{i,\ cycle\ (j-2)} \end{aligned} \tag{7}$$

이다. 같은 방법으로, 임의 싸이클 $(j-k-1)$에 buffer에 저장된 데이터 패킷의 수가 k 일 확률, $P(\varepsilon = k)_{i,\ cycle\ (j-k-1)}$은

$$\begin{aligned} P(\varepsilon = k)_{i,\ cycle\ (j-k-1)} &= P(\varepsilon = k+1)_{i,\ cycle\ (j-k-2)} \times P(\hbar = 0)_{i,\ cycle\ (j-k-1)} \\ &+ P(\varepsilon = k)_{i,\ cycle\ (j-k-1)} \times P(\hbar = 1)_{i,\ cycle\ (j-k-1)} \\ &+ P(\varepsilon = k-1)_{i,\ cycle\ (j-k-1)} \times P(\hbar = 2)_{i,\ cycle\ (j-k-1)} \end{aligned} \tag{8}$$

이다. 식 (8)은 싸이클 $(j-k-1)$에 buffer에 저장된 데이터 패킷의 수가 k 일 경우는

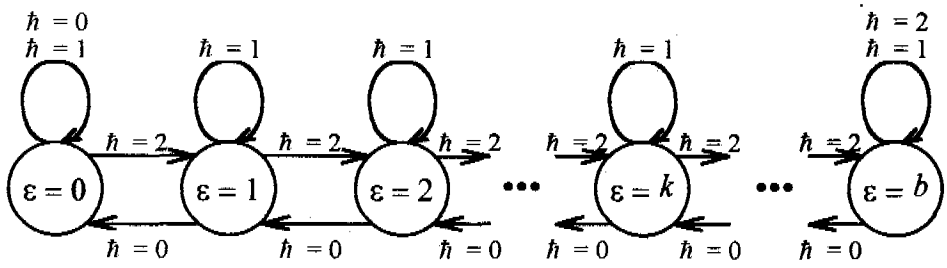


그림 3. buffer entry 상태 변환 도

싸이클 $(j-k-2)$에 buffer에 저장된 데이터 패킷의 수와 싸이클 $(j-k-1)$에 해당 출력 단으로 지향하는 데이터 패킷의 수의 합이 $(k+1)$임을 보여주고 있다. 이때 하나의 데이터 패킷은 싸이클 $(j-k-1)$ 동안 다음 스테이지로 이동하고 나머지 k 데이터 패킷은 buffer에 저장된다.

식 (6), (7), 그리고 (8)등의 확률 식에서 임의 buffer가 싸이클 j에 k 개의 데이터 패킷을 저장할 확률과 싸이클 $(j+1)$에 k 개의 데이터 패킷을 저장할 확률은 같다고 볼 수 있다. 즉, 이들 식에 정상 상태 확률(steady state probability) 개념 적용이 가능하고, 따라서 $P(\varepsilon=k)_{i,\ cycle\ j} = P(\varepsilon=k)_{i,\ cycle\ (j+1)}$ 그리고 $P(\hbar=x)_{i,\ cycle\ j} = P(\hbar=x)_{i,\ cycle\ (j+1)}$이다. 정상 상태 확률 환경 개념을 도입하여 buffer가 저장할 데이터 패킷의 수와 해당 buffer로 진행하고자 하는 데이터 패킷의 수의 관계를 상태 변환도로 표현하면 그림 3과 같고, 이에 따라 식 (6)은

$$\begin{aligned} P(\varepsilon=0)_i &= P(\varepsilon=1)_i \times P(\hbar=0)_i \\ &\quad + P(\varepsilon=0)_i \times P(\hbar=1)_i \\ &\quad + P(\varepsilon=0)_i \times P(\hbar=0)_i \end{aligned} \tag{9}$$

이 된다. 식 (9)를 정리하여 $P(\varepsilon=1)_i$를 $P(\varepsilon=0)_i$의 식으로 구하면

$$\begin{aligned} P(\varepsilon=1)_i &= P(\varepsilon=0)_i \times \frac{(1-P(\hbar=0)_i - P(\hbar=1)_i)}{P(\hbar=0)_i} \\ &= P(\varepsilon=0)_i \times \frac{P(\hbar=2)_i}{P(\hbar=0)_i} \end{aligned} \tag{10}$$

이다. 또한 식 (7)의 $P(\varepsilon=1)_i$은

$$\begin{aligned} P(\varepsilon=1)_i &= P(\varepsilon=2)_i \times P(\hbar=0)_i \\ &\quad + P(\varepsilon=1)_i \times P(\hbar=1)_i \\ &\quad + P(\varepsilon=0)_i \times P(\hbar=2)_i \end{aligned} \tag{11}$$

이 된다. 여기서 $P(\varepsilon=0)_i$를 식 (10)을 이용하여 $P(\varepsilon=1)_i \times \frac{P(\hbar=0)_i}{P(\hbar=2)_i}$로 치환하고, $P(\varepsilon=2)_i$를 $P(\varepsilon=1)_i$의 식으로 표현하면

$$P(\varepsilon=2)_i = P(\varepsilon=1)_i \times \frac{(1-P(\hbar=0)_i - P(\hbar=1)_i)}{P(\hbar=0)_i}$$
$$= P(\varepsilon=1)_i \times \frac{P(\hbar=2)_i}{P(\hbar=0)_i} \quad (12)$$

로 얻어진다. 같은 방법으로 buffer가 임의 싸이클 종료 시 $(k-1)$개의 데이터 패킷을 저장하고 있을 확률, $P(\varepsilon=k-1)_i$을 구하면

$$P(\varepsilon=k-1)_i = P(\varepsilon=k)_i \times P(\hbar=0)_i + P(\varepsilon=k-1)_i \times P(\hbar=1)_i + P(\varepsilon=k-2)_i \times P(\hbar=2)_i \quad (13)$$

이 되고, 여기서 $P(\varepsilon=k)_i$를 $P(\varepsilon=k-1)_i$의 식으로 표현하면

$$P(\varepsilon=k)_i = P(\varepsilon=k-1)_i \times \frac{P(\hbar=2)_i}{P(\hbar=0)_i} \quad (14)$$

로 계산된다. 식 (10), (12), 그리고 (14)의 $\frac{P(\hbar=2)_i}{P(\hbar=0)_i}$을 Ω로 놓고 회귀적 기법으로 $P(\varepsilon=k)_i$를 구하면

$$P(\varepsilon=k)_i = P(\varepsilon=0)_i \times \Omega^k \quad (15)$$

이 된다.

식 (15)의 Ω는 식 (1)과 (3)에서 계산되고, 식 (4)의 $P(D_i=0)_i$을 구하기 위한 $P(\varepsilon=0)_i$는 다음과 같은 연산 과정으로 얻을 수 있다. 스위치가 수용할 수 있는 데이터 패킷의 수가 b이면, 임의 싸이클 종료 시 buffer에 저장된 데이터 패킷의 수는 0개에서 b개 중 어느 하나일 것이다. 따라서

$$\sum_{k=0}^{b} P(\varepsilon=k)_i = \sum_{k=0}^{b} P(\varepsilon=0)_i \times \Omega^k = 1 \quad (16)$$

이 되어, 식 (15)의 $P(\varepsilon=0)_i$는

$$P(\varepsilon=0)_i = \frac{1}{\sum_{k=0}^{b} \Omega^k} \quad (17)$$

와 같이 계산된다.

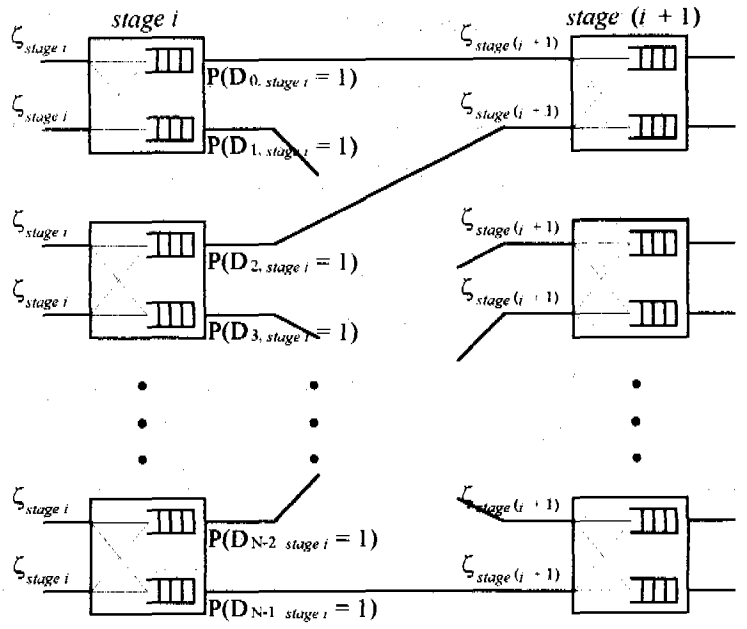


그림 4. 다층 연결 망 내부 스위치 입출력 확률의 관계

다층 연결 망 내부 스테이지 i에 위치한 임의 2×2 crossbar 스위치 출력 단 D_γ로 데이터 패킷이 출력될 확률, $P(D_\gamma = 1)_i$은 식(4), (15), 그리고 (17)을 이용하여 구할 수 있다. 네트웍 구조 상 스테이지 i에 위치한 스위치 출력 단은 스테이지 $(i + 1)$에 위치한 임의 스위치의 입력 단으로 연결됨으로, 스테이지 i의 임의 스위치 출력 단으로 데이터 패킷이 출력될 확률, $P(D_\gamma = 1)_i$은 그림 4와 같이 stage $(i+1)$에 위치한 해당 스위치 입력 단으로 데이터 패킷이 유입될 확률, $\zeta_{stage\,(i+1)}$이 된다. 따라서 네트웍 입력 단의 $\zeta_{stage\,0}$이 주어지면 이로부터 $P(D_\gamma = 1)_0$을 구하고, $P(D_\gamma = 1)_0$을 다시 $\zeta_{stage\,1}$로 놓고 $P(D_\gamma = 1)_1$을 구하는 과정을 반복하여 다층 연결 망 최종 스테이지의 스위치 출력 단으로 데이터 패킷이 출력될 확률, $P(D_\gamma = 1)_{last\,stage}$을 계산하게 된다. 일단 $P(D_\gamma = 1)_{last\,stage}$을 구하면, $N \times N$ 다층 연결 망의 경우 전체 네트웍 출력 단으로 출력되는 데이터 패킷의 수, OP는

$$OP = N \times P(D_\gamma = 1)_{last\,stage} \tag{18}$$

로 계산된다. 또한 다층 연결 망 입력 단으로 매 싸이클 마다 유입되는 총 데이터 패킷의 수를 $IP(= N \times \zeta_{stage\,0})$라 하면, 네트웍 정상 Throughput, NT(Normalized Throughput),은

$$NT = \frac{OP}{IP} = \frac{P(D_\gamma = 1)_{last\,stage}}{\zeta_{stage\,0}} \tag{19}$$

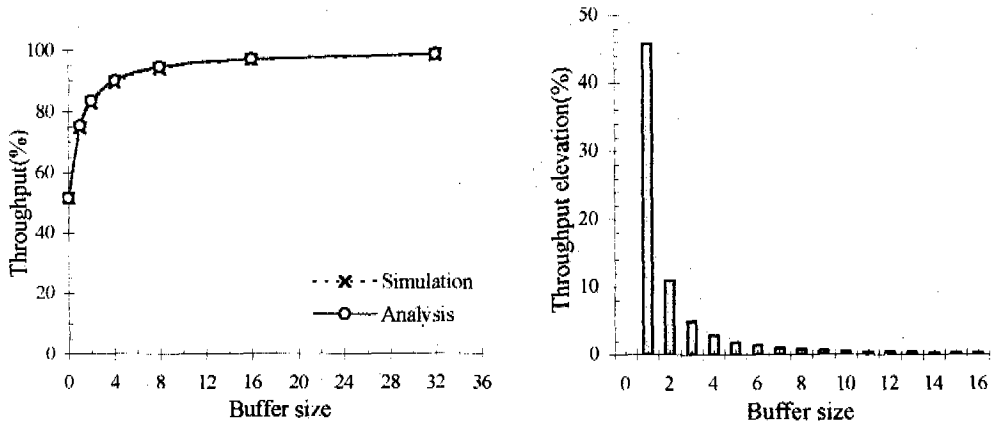
와 같이 얻어진다.

표 I. Multiple-buffered 2×2 crossbar switches 로 구성된 MIN 의 Throughput

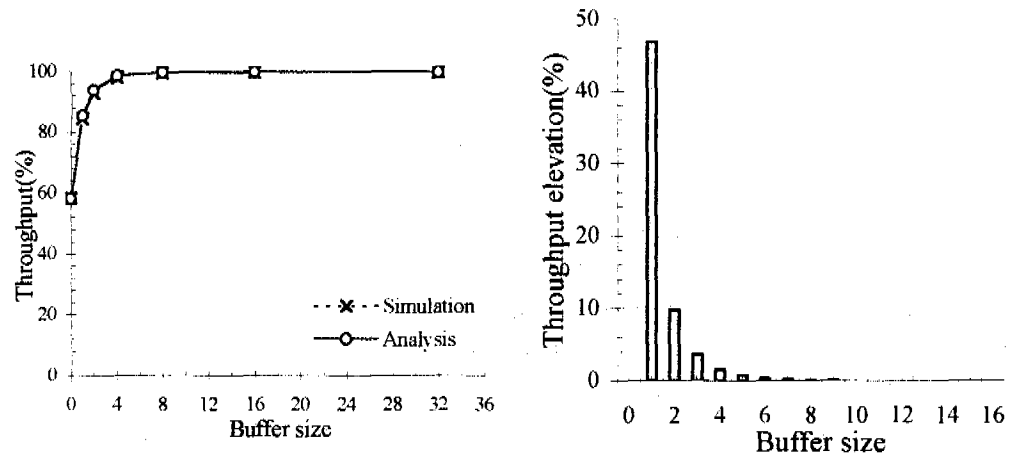
Number of buffer spaces	Throughput for $\zeta_{stage\,0} = 1.0$ (NT, %)		Throughput for $\zeta_{stage\,0} = 0.8$ (NT, %)	
	Analysis	Simulation	Analysis	Simulation
0	51.654	51.638	58.168	58.250
1	75.286	74.687	85.433	84.633
2	83.458	82.860	93.814	92.838
4	90.055	89.502	98.784	98.148
8	94.470	94.045	99.949	99.847
16	97.072	96.899	99.999	99.998
32	98.491	98.362	99.999	99.999
∞	99.999	99.999	99.999	99.999

표 I 과 그림 5 는 multiple-buffered 2×2 crossbar 스위치들을 사용한 8×8 Baseline 네트웍을 대상으로 본 논문에서 제안한 분석 모형을 적용하여 각 스위치에 장착된 buffer 크기 별 네트웍 정상 Throughput 의 예측 계산치를 표와 그래프로 정리하고, 이를 다시 시뮬레이션을 통하여 얻은 결과들과 비교하여 보여주고 있다. 표 1 과 그림 5 에 열거한 데이터는 8×8 Baseline 다층 연결 망 입력 단으로 매 싸이클 마다 한 개의 데이터 패킷이 유입되는 경우($\zeta_{stage\,0} = 1.0$), 그리고 매 싸이클 마다 데이터 패킷이 유입될 확률이 0.8 일 경우($\zeta_{stage\,0} = 0.8$) 네트웍 정상 Throughput 을 보여주고 있다. 여기서 $IP = OP = 8$. 시뮬레이션 데이터는 네트웍 운용 초기 buffer 가 비어 있는 상태 즉, cold start 환경에서 얻어지는 데이터 부분을 배제하고, 정상상태 동작 성능을 취하여 얻은 결과이다. 제안한 분석 모형의 타당성 검토를 위하여 수행된 시뮬레이션은 각각의 네트웍 성능 계산 치와 시뮬레이션 결과 치가 상호 미세한 오차 범위 내에서 일치하여 분석 모형의 우수성을 입증하였다.

표 I 과 그림 5 의 데이터들은, 스위치에 장착된 buffer 의 크기가 커지서 수용할 수 있는 데이터 패킷의 수가 증가함에 따라, 네트웍 Throughput 이 함께 증가하되, buffer size 가 일정 크기를 벗어나면 네트웍 Throughput 증가율이 둔화됨을 보여주고 있다. $\zeta_{stage\,0} = 1.0$ 일 경우 buffer size 증가에 따른 구체적인 성능 향상 추이를 살펴보면, buffer 가 장착 되지 않았을 경우 51.65%, 한 개의 데이터 패킷을 수용할 수 있는 buffer 가 각 스위치에 장착된 경우 75.29%, 두 개의 데이터 패킷을 수용할 수 있는 buffer 가 장착된 경우 83.46%의 Throughput 을 보여, buffer 가 장착 되지 않은 경우와 비교하여 각각 45.75%,와 61.59%의 증가율 보인다. 또한 네 개의 데이터 패킷을 수용할 수 있는 buffer 가 각 스위치에 장착된 경우 90.06%, 여덟 개의 데이터 패킷을 수용할 수 있는 buffer 가 장착된 경우 94.47%의 Throughput 을 보여, buffer 가 장착 되지 않은 경우와 비교하여 각각 74.37%,와 82.71%의 증가율을 기록하였다. 또한 네트웍 소통 빈도가 낮은 경우($\zeta_{stage\,0} = 0.8$) buffer 가 네트웍 Throughput 에 미치는 영향이 증가 후 극적으로



(a). $\zeta_{stage\,0} = 1.0$



(b). $\zeta_{stage\,0} = 0.8$

그림 5. Buffer size 증가에 따른 Throughput

감소하는 것으로 나타났다. 즉, 한 개 데이터 패킷을 수용할 수 있는 buffer가 장착된 경우 85.43%의 Throughput을 제공하여, buffer가 장착 되지 않은 경우와 비교할 때 각각 46.87%의 증가율 보인 반면, 두 개의 데이터 패킷을 수용할 수 있는 buffer 장착 시 93.81%의 Throughput과 61.28%의 Throughput 증가율을 보이고, 이후 buffer의 크기가 커지며 Throughput 증가율이 극적으로 둔화되어 열 개 이상의 buffer가 장착 될 경우 Throughput은 99.99%에 달하고, Throughput 증가율은 71.90%에 머무는 것으로 분석되었다. 그림 5의 Throughput elevation 그래프는 스위치 buffer 공간이 한 개씩 증가함에 따라 이전 상태와 비교하여 얻어지는 Throughput 증가율 보여주고 있다.

본 논문에서 제안한 성능 분석 모형은, 분석 결과의 신뢰도를 기반으로, 네트워 가격 대

성능 비, 혹은 요구되는 네트웍 Traffic 및 성능 등을 고려하여 스위치에 장착될 buffer의 적정 크기를 결정할 수 있게 하여 네트웍 설계 과정에 유용한 도구가 될 것으로 사료된다.

IV. 결 론

본 연구에서는 복수 buffered 2×2 스위치들을 사용하여 설계된 다층 연결 망의 분석 모형을 제시하였다. 제시된 분석 모형은 스위치에 장착된 buffer 의 개수와 무관하게 적용 가능하고, 분석 과정에서 간단한 데이터 충돌 처리 기법을 도입하여 모형의 수식 이해가 용이하다. 제안한 수학적 성능 분석 연구의 실효성을 검증하기 위한 시뮬레이션 처리 결과는 상호 미세한 오차 범위 내에서 모형의 예측 데이터와 일치하는 결과를 보여 분석 모형과 함께 수정된 데이터 충돌 방식의 타당 성을 입증하였다. 또한 복수 buffered 2×2 스위치들로 구성된 Baseline 네트웍을 대상으로 제안되고 입증된 본 연구의 분석 모형은 모든 다층 연결 망의 성능 분석에 확대 적용 가능한 한편, 임의 크기 $a \times a$ 스위치를 채용한 다층 연결 망 분석 모형 연구의 기반을 제공하였다.

참고문헌

[1] V. P. Kumar and S. M. Reddy, "Augmented Shuffle-Exchange Multistage Interconnection Networks", *IEEE Computer*, Jun. 1987.

[2] D. M. Dias and J. R. Jump, " Analysis and Simulation of Buffered Delta Networks", *IEEE Trans. on Computers*, Vol. C-30, No. 4. pp273-282, Apr. 1981.

[3] Y. C. Jenq, "Performance Analysis of a Packet Switch Based on Single Buffered Banyan Network", *IEEE J. Select. Areas Comm.*, Vol. SAC-3, No. 6, pp1014-1021, Dec. 1983.

[4] C. P. Krusal and M. Snir, "The Performance of Multistage Interconnection Networks for Multiprocessors", *IEEE Trans. on Computers*, Vol. C-32, No. 12. pp1091-1098, Dec. 1983.

[5] H. Yoon, K. Y. Lee, and M. T. Liu, "Performance Analysis of Multibuffered Packet-Switching Networks in Multiprocessor Systems", *IEEE Trans. on Computers*, Vol. C-39, No. 3. pp319-327, Mar. 1990.

[6] T. Feng, "Data Manipulating Functions in Parallel Processors and Their Implementations", *IEEE Trans. on Computers*, Vol. C-23, pp309-318, Mar. 1974.

[7] K. E. Batcher, "The Flip Network in STARAN", *Proc. Intl. Conf. on Parallel Processing*, pp65-71, Aug. 1976.

[8] D. K. Lawrie, "Access and Alignment of Data in an Array processor", *IEEE Trans. on Computers*, Vol. C-24, pp1145-1155, Dec. 1975.

[9] G. J. Lipovski and A. Tripathi, "A Reconfigurable Varistructure Array Processor", *Proc. Intl. Conf. on Parallel Processing*, pp165-174, Aug. 1977.

[10] -, "Butterfly GP1000 - Overview", *BBN Advanced Computer Inc.*, Nov. 1988.

[11] M. Lee, C. L. Wu, "Performance Analysis of Circuit Switching Baseline Interconnection Networks", Proc. 11th Computer Architecture Conf., pp82-90, 1984.

[12] Allan Gottlieb, Ralph Grishman, et al., "The NYU Ulyracomputer - Desiging an MIMD Shared Memory Parallel Computer", *IEEE Trans. on Computers*, Vol. C-32, No. 2, pp175-189, Feb. 1983.

[13] G. F. Pfister, W. C. Brantley, et al., "The IBM Research Parallel processor Prototype(RP3): Introduction and Architecture", *Proc. Intl. Conf. on Parallel Processing*, pp764-771, Aug. 1985.

[14] -, "TC2000 Technical Product Summary", *BBN Advanced Computer Inc.*, Jul. 1989.

[15] Chuan-Lin Wu and Tse-Yun Feng, "On a class of Multistage Interconnection Networks", *IEEE Trans. on Computers*, Vol. C-29, No. 8, pp108-116, Aug. 1980.