

핫스팟 상황 하에서 출력 버퍼형 $a \times a$ 스위치로 구성된 다단 연결망의 성능분석

김정윤 신태지 양명국

울산대학교 전기전자 정보시스템공학부

basseam@hanmail.net shintaezi@bcline.com mkyang@mail.ulsan.ac.kr

Performance Evaluation of a Multistage Interconnection

Network with Buffered $a \times a$ Switches under Hot-spot Environment

Jung-Yoon Kim⁰ Tae-Zi Shin Myung-Kook Yang

Dept. of Electrical & Computer Engineering University of Ulsan

요약

본 논문에서는, $a \times a$ 출력 버퍼 스위치로 구성된 핫스팟이 발생된 상황 하에서 다단 연결망(Multistage Interconnection Network, MIN)의 성능 예측 모형을 제안하였다. 제안한 성능 예측 모형은 먼저 네트워크 내부 임의 스위치 입력 단에 유입되는 데이터 패킷이 스위치 내부에서 전송되는 유형을 확률적으로 분석하여 설계하였다. 성능분석 모형은 스위치에 장착된 버퍼의 개수와 무관하게 버퍼를 장착한 $a \times a$ 스위치의 성능, 네트워크 정상상태 처리율(Normalized Throughput, N)과 네트워크 지연시간(Network Delay)의 예측이 가능하고, 나아가서 이를로 구성된 모든 종류의 다단 연결망 성능 분석에 적용이 용이하다. 제안한 수학적 성능 분석 연구의 실효성 검증을 위하여 병행된 시뮬레이션 결과는 상호 미세한 오차 범위 내에서 모형의 예측 데이터와 일치하는 결과를 보여 분석 모형의 타당성을 입증하였다.

1. 서 론

넓은 대역폭과 네트워크 유연성 등의 장점을 가진 다단 연결망(Multistage Interconnection Network : MIN)은 각종 병렬 컴퓨터의 상호 연결망으로 시작하여, 최근 고성능 컴퓨터 네트워크의 스위칭 소자 및 B-ISDN의 기본이 되는 ATM 스위칭 소자등으로 다양하게 사용되어지고 있다.[1]

다단 연결망을 통한 데이터 이동에는 다단 연결망의 구조상 각 스테이지에서 스위치마다 제어가 요구되고, 데이터 이동 경로에 따라 특정 스위치에서 두 개 이상의 데이터가 하나의 경로로 진행하고자 하는 데이터 충돌 현상이 초래된다. 이러한 현상은 핫스팟이 발생할 때 더욱 빈번해진다. 데이터 충돌 현상은 네트워크 성능저하를 유발할 뿐 아니라 네트워크 신뢰도에도 큰 영향을 미치게 된다. 이러한 네트워크 내부의 데이터 충돌로 인한 데이터 손실을 최소화하고, 네트워크 성능 향상을 위한 다양한 연구가 진행되고 있다.[2] 이를 기운데 스위치 소자에 버퍼를 장착하는 기법은 데이터 충돌로 인하여 소실될 데이터 패킷을 버퍼의 여유 공간에 저장함으로, 데이터 손실을 막고, 나아가서 네트워크의 성능을 증가시키는 방법으로 널리 알려져 있다.

Mahmoud Saleh 와 Mohammed Atiquzzaman[3]는 핫스팟 상황 하에서 Shared Buffer Delta Network의 성능을 분석하였다. 그러나 스위치에 데이터 소통양이 높아지면서 분석과 시뮬레이션 데이터 간의 오차가 커지는 문제를 보이고 있다. 본 논문에서는 네트워크 성능 평가의 두 가지 주요 요소로 알려진 네트워크의 성능 평가의 두 가지 주요 요소로 알려진 네트워크의 정상상태 처리율(Normalized Throughput, N)과 네트워크 지연시간(Network Delay)을 분석하였다. 본 논문에 제안된 성능 분석 모형은 스위치 크기 및 스위치에 장착된 버퍼의 개수와 무관하게 적용 가능하다. 제안한 수학적 성능 분석 연구의 실효성 검증을 위하여 병행된 핫스팟 상황 하에서의 시뮬레이션 결과는 상호 미세한 오차 범위 내에서 모형의 예측 데이터와 일치하는 결과를 보여 분석 모형의 타당성을 입증하였다.

2. 성능 분석 모형

2.1 네트워크 환경에 대한 일반적인 가정

복수 buffered 다단 연결망의 분석 모형 개발과 시뮬레이션을 위해 본 논문에 적용된 일반적인 가정을 정리하면 다음과 같다.

- $a \times a$ multiple-buffered crossbar 스위치들로 구성된 $N \times N$ Baseline network를 분석대상으로 한다. Wu와 Feng[4]의 연구에서 밝혀진 바와 같이 기존의 모든 다층 연결망은 기능적으로 동일하여, 본 연구에서 baseline network를 대상으로 얻은 성능 분석 결과 및 분석 모형은 다른 모든 다층 연결망의 해석에 활용이 가능하다.

- 네트워크는 스위치 Clock Cycle, Δt 에 따라 동기 적으로 작동한다. 즉, 네트워크 내부 데이터 패킷은 스위치 클럭 동안 임의 스위치 출력 단을 출발, 다음 스테이지 스위치를 통하여 해당 출력 단에 도달한다.

- 스위치에 장착된 버퍼는 스위치 출력 단에 위치하고, 버퍼 공간 하나는 한 개의 데이터 패킷을 수용할 수 있다.

- 데이터 패킷은 네트워크 입력 단의 각 소스 노드에서 같은 확률로 발생하며 네트워크 내부 임의 스테이지에 위치한 스위치 입력 단으로 데이터 유입될 확률은 ζ_{stage} 라 한다. 따라서 매 사이클마다 네트워크 각 입력 단에 한 개씩의 데이터 패킷이 유입될 경우, ζ_{stage} 는 1이 된다.

- 네트워크 입력 단으로 유입되는 데이터 패킷의 네트워크 최종 출력 단 행선지는 핫스팟 출력 단 지향율(ζ_H)과 일반 출력 단 지향율(ζ_C)로 다음과 같이 나누어진다. [3]

$$\zeta_H = \zeta_{stage} \left(H_0 + \frac{1-H_0}{N} \right), \quad \zeta_C = \zeta_{stage} \left(\frac{1-H_0}{N} \right)$$

$$\zeta_H + (N-1)\zeta_C = \zeta_{stage}$$

여기서, H_0 은 네트워크 한 핫스팟율이다.

- 데이터 충돌 발생 시 무작위 중재 방식에 의거 데이터 처리 우선 순위를 결정한다.
- 핫스팟은 최종 출력 단에 오직 하나만이 존재한다.

핫스팟 상황 하에서 네트워크 내부 스테이지에 위치한 임의 스위치 내부의 데이터 이동 패턴을 확률적으로 분석하기 위하여 사용된 변수는 다음과 같다.

H_i	: i 번째 스테이지로 유입되는 데이터의 핫스팟 지향율
R_i	: 최종 핫스팟 출력 단과 관련된 네트워크 초기 입력 단의 수
D_j	: 임의의 SE (Switch Element)의 j 번째 출력 단 ($0 \leq j \leq a-1$)
$R(h=a)$: 스테이지 i 에 위치한 임의 SE의 한 출력 단으로 r 개의 패킷이 지향할 확률

핫스팟 상황 하에서 네트워크를 구성하는 스위치는 크게 3가지 상태로 구분된다. 먼저, 스위치 내에 핫스팟을 지향하는 경로가 포함된 상태(Type 1), 그리고 해당 스위치가 핫스팟 지향 경로를 포함하지 않지만 이전 스테이지에 연결된 스위치가 Type 1인 경우(Type 2), 마지막으로 해당 스위치가 핫스팟 지향 경로를 포함하지 않고 이전 스테이지에 연결된 스위치가 Type 1이 아닌 상태(Type 3) 등으로 구분 할 수 있다.

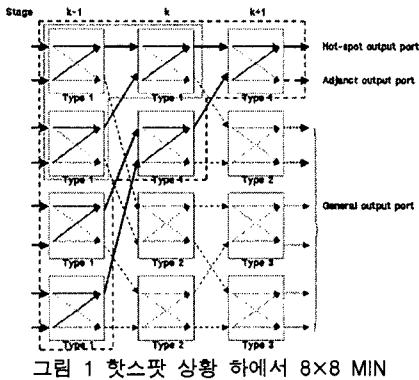


그림 1 핫스팟 상황 하에서 8x8 MIN

2.2 스위치 내부에서의 데이터 이동 패턴

2.2.1 핫스팟 영향이 없는 경우

네트워크 내부 임의의 크로스바 스위치 입력 단에 유입된 데이터 패킷은 데이터가 지향하는 행선지에 따라 스위치의 a^2 개 출력 단 중 어느 한 출력 단으로 향하게 된다. 네트워크 환경의 일반적인 정의에 의거, 네트워크 입력 단에 처음 데이터 패킷이 유입될 때 최종 출력 단 행선지가 핫스팟 출력 단 지향율 (ζ_H)과 일반 출력 단 지향율 (ζ_G)로 주어지며, 핫스팟 영향이 없는 경우이므로, 핫스팟 출력 단을 지향하는 비율, H_0 ,이 0%가 된다. 그러므로 네트워크 스테이지 i 에 위치한 스위치 중, 핫스팟 영향이 없는 임의의 스위치 입력 단에 데이터 패킷이 유입될 확률이 $\zeta_{stage,i}$ 로 주어지면 해당 스위치의 어느 한 출력 단으로

데이터 패킷이 향할 확률은 $\frac{\zeta_{stage,i}}{a}$ 가 된다.

네트워크 내부 스테이지 i 에 위치한 임의의 스위치 입력 단 D_i 에 데이터 패킷이 유입될 확률이 $\zeta_{stage,i}$ 로 주어지면, 그로 인하여 해당 스위치의 특정 출력 단 D_0 로 데이터 패킷이 향할 확률은 $\frac{\zeta_{stage,i}}{a}$ 가 되고, 특정 출력 단 D_0 로 데이터 패킷이 향

하지 않을 확률은 $(1 - \frac{\zeta_{stage,i}}{a})$ 가 된다. 따라서 스테이지 i 에 위치한 임의의 스위치 출력 단 D_0 로 a 개의 데이터 패킷이 지향 할 확률, $R(h=r)_{i,D_0}$,은

$$R(h=r)_{i,D_0} = {}_a C_r \times \left(\frac{\zeta_{stage,i}}{a} \right)^r \times \left(1 - \left(\frac{\zeta_{stage,i}}{a} \right) \right)^{a-r} \quad (1)$$

로 계산된다. 여기서, $(0 \leq r \leq a)$ 이다.

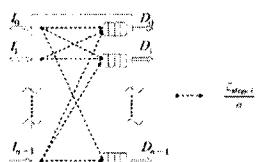


그림 2 핫스팟 영향이 없는 경우 (Type 2 or 3)

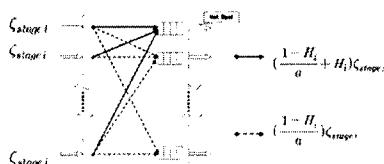


그림 3 핫스팟 영향권에 포함된 경우 (Type 1)

2.2.2 핫스팟 영향권에 포함된 경우

스위치 내에 핫스팟을 지향하는 경로가 포함된 crossbar 스위치(Type 1) 입력 단에 유입된 데이터 패킷은 데이터가 지향하는 행선지에 따라 스위치의 a^2 개 출력 단 중 어느

한 출력 단으로 향하게 된다. 네트워크 환경의 일반적인 정의에 의거, 네트워크 입력 단에 처음 데이터 패킷이 유입될 때 최종 출력 단 행선지가 핫스팟 출력 단 지향율 (ζ_H)과 일반 출력 단 지향율 (ζ_G)로 주어지므로, 스위치에서 데이터 패킷이 지향할 확률도 두 가지로 나누어진다.

첫째로, 핫스팟 출력에 인접한 출력 단으로 데이터 패킷이 지향할 확률은 식 (1)의 유도과정에서 데이터 패킷이 지향할 확률, $\frac{\zeta_{stage,i}}{a}$, 을 핫스팟 영향권에 포함된 임의의 스위치에서 핫스팟 출력 단에 인접한 출력 단으로 r 개의 데이터 패킷이 지향할 확률, $R(h=r)_{i,A}$, 은

$$R(h=r)_{i,A} = {}_a C_r \times \left(\frac{1-H_i}{a} \right) \zeta_{stage,i}^r \times \left\{ 1 - \left(\frac{1-H_i}{a} \right) \zeta_{stage,i} \right\}^{a-r} \quad (2)$$

로 계산된다. 여기서, $(0 \leq r \leq a)$ 이다.

둘째로, 핫스팟 출력 단으로 데이터 패킷이 지향할 확률은 식 (1)의 유도과정에서 데이터 패킷이 지향할 확률, $\frac{\zeta_{stage,i}}{a}$, 을 핫스팟 영향권에 포함된 스위치에서 핫스팟 출력 단으로 데이터 패킷이 지향할 확률, $(\frac{1-H_i}{a} + H_i) \zeta_{stage,i}$, 를 변경하면, 네트워크 스테이지 i 에 위치한 스위치 중, 핫스팟 영향권에 포함된 임의의 스위치에서 핫스팟 출력 단으로 r 개의 데이터 패킷이 지향할 확률, $R(h=r)_{i,H}$, 은

$$R(h=r)_{i,H} = {}_a C_r \times \left(\frac{1-H_i}{a} + H_i \right) \zeta_{stage,i}^r \times \left\{ 1 - \left(\frac{1-H_i}{a} + H_i \right) \zeta_{stage,i} \right\}^{a-r} \quad (3)$$

로 계산된다. 여기서, $(0 \leq r \leq a)$ 이다.

2.3 정상상태 처리율 분석

네트워크 내부 스테이지에 위치한 임의의 $a \times a$ 출력 버퍼 스위치 내부 데이터 이동 패턴의 확률적 분석을 토대로 buffered 다단 연결망의 성능 분석을 위하여 사용될 변수는 다음과 같다.

b	: 스위치에 장착된 버퍼가 저장할 수 있는 데이터 패킷 수
	: 버퍼에 저장된 데이터 패킷 수
$P(h=k)_i$: 버퍼에 저장된 데이터 패킷 수가 k 개일 확률
$P(D_d=1)_i$: 출력 단 D_d 로 데이터 패킷이 출력될 확률
$P(D_d=0)_i$: 출력 단 D_d 로 데이터 패킷이 출력되지 않을 확률
T_H	: 네트워크 입력 단에 유입되는 데이터 패킷 중 Hot 출력 지향율들의 합
T_C	: 네트워크 입력 단에 유입되는 데이터 패킷 중 Cold 출력 지향율들의 합

네트워크 성능 분석의 두 가지 주요 요소는 네트워크 정상상태 처리율과 네트워크 지연시간이다. 이 요소들은 핫스팟 영향이 없는 최종 출력 단의 경우, 네트워크 정상상태 처리율 (NT)은 네트워크 최종 스테이지의 스위치 출력 단으로 데이터 패킷이 출력될 확률, $P(D_d=1)_{laststage}$, 을 네트워크 초기 스테이지의 스위치 입력 단으로 데이터 패킷이 유입될 확률, $\zeta_{stage,0}$, 로 나누어서 식 (4)와 같이 계산된다. 여기서 버퍼 크기에 따른 네트워크 마지막 스테이지의 스위치 출력 단으로 데이터 패킷이 출력될 확률은 식 (5,6,7,8) 등을 식(4)에 대입하면 구할 수 있다.[5]

$$NT = \frac{P(D_d=1)_{laststage}}{\zeta_{stage,0}} \quad (4)$$

$$P(D_d=1)_{i,cycle,j} = 1 - P(\epsilon=0)_{cycle(j-1)} \times P(h=0)_{i,cycle(j-1)} \quad (5)$$

$$P(\epsilon=0)_{i,cycle,j} = \frac{1}{\sum_{k=0}^b \phi_x} \quad (6)$$

$$\phi_i = \frac{1}{P(h=0)} \times \sum_{y=1}^a P(h=y) \quad (7)$$

$$\phi_x = \phi_{x-1} + \sum_{k=0}^{x-2} \phi_{x-k-1} \times \phi_k \quad (8)$$

핫스팟 영향권에 포함된 최종 출력 단의 경우, 핫스팟 출력에서의 네트워크 정상상태 처리율(NT)은 $P(D_d=1)_{last stage}$ 을 네트워크 입력 단에 유입되는 데이터 패킷 중 핫스팟 출력 단으로 지향하는 확률들의 합, T_{H_i} 으로 나눔으로서 식 (9)와 같이 계산되고. 핫스팟에 인접한 출력 단으로 지향하는 확률은 식(9)에서 T_{H_i} 대신에 T_{C_i} 으로 나눔으로서 식 (10)과 같이 계산된다. [6]

$$N.T_{HotSpotOutlet} = \frac{P(D_d=1)_{last stage}}{T_{H_i}} \quad (9)$$

$$N.T_{AdjacentOutlets} = \frac{P(D_d=1)_{last stage}}{T_{C_i}} \quad (10)$$

$$T_{H_i} = \left(\frac{1-H_1}{R_i} + H_1 \right) R_i \times \xi_{stage 0} \quad (11)$$

$$T_{C_i} = \left(\frac{1-H_1}{R_i} \right) R_i \times \xi_{stage 0} \quad (12)$$

식 (9) 와 (10)에서 N/T (Normalized Throughput)을 구하기 위해서는 각 스테이지에서의 유입률, $\xi_{stage i}$ 를 구하여야 한다. 값은 식 (2) 와 (3)에 유입률 $\xi_{stage i-1}$ 과 핫스팟 출력 단 지향율, H_i 를 대입하여 구할 수 있다. i 번째 스테이지로 유입되는 데이터 중에 핫스팟 출력 단 지향율은 식(13)에 초기 핫스팟 출력 단 지향율, H_0 과 최종 핫스팟 출력 단과 관련된 네트워크 초기 입력 단의 수, R_i 를 대입하면 구할 수 있다.

$$H_i = \frac{H_1}{\frac{1-H_1}{R_i} + H_1} \quad (13)$$

예를 들어 8×8 다단 연결망의 경우, Single 핫스팟을 가지는 다단 연결망의 데이터 패킷의 흐름은 그림 1과 같이 나타낼 수 있다. 그림 (1)에서 스테이지 $k-1$ 의 각 Type 1 SE (Switch Element)들의 핫스팟 출력을 구하기 위하여 네트워크 입력 단의 초기 유입률, $\xi_{stage k-1}$, 과 초기 핫스팟 출력 단 지향율, H_{k-1} , 을 식 (3)에 대입하면 스테이지 $k-1$ 로부터의 핫스팟 출력 단 지향율(스테이지 $k-1$ 의 Type 1 으로 유입되는 확률)을 구할 수 있다. 이렇게 계산된 스테이지 $k-1$ 의 Type 1 으로 유입되는 확률, $\xi_{stage k-1}$, 과 스테이지 k 에서 핫스팟 출력 단 지향율을 (3)에 대입하면 스테이지 $k+1$ 에 있는 하나의 Type 1 SE의 유입률 $\xi_{stage k+1}$ 을 계산할 수 있다. 계산된 $\xi_{stage k+1}$ 과 (9)식에서 계산된 스테이지 $k+1$ 에서의 핫스팟 출력 단 지향율, H_{k+1} , 을 식 (3)에 대입하면 $P(D_d=1)_{k+1}$ 의 값을 계산 할 수 있다. 마지막으로 식 (9)와 (10)에 $P(D_d=1)_{last stage}$ 을 대입하면 핫스팟 출력과 핫스팟의 영향권에 포함된 SE내의 핫스팟에 인접한 출력들의 NT를 구할 수 있다.

2.3 네트워크 지연시간 분석

네트워크 성능 평가에 있어서 정상상태 처리율과 함께, 다른 주요 평가 지표로 네트워크 지연시간(Network Delay, τ_s)을 들 수 있다. 네트워크를 성공적으로 통과한 데이터 패킷의 전체 네트워크 지연시간, τ_s 는 각 스테이지 지체시간을 합하여 식(14)와 같이 계산되는 반면, 종도 소실 데이터 패킷은 종도 유실 감지 시간(Lost Data Detection Time, LDDT)만큼의 오류 검사 과정을 거쳐 재전송 된다. [6]

$$\tau_s = \sum_{i=0}^{n-1} \tau_{s, stage i} \quad (14)$$

여기서 $n = \lceil \log_2 N \rceil$ 이고, N 은 다단 연결망의 입력 단 수를 나타낸다.

$$\tau_{s, stage i} = \left\{ P(\epsilon = \rho)_{i, cycle(j-1)} \times \sum_{y=0}^{a-1} \frac{1}{y+1} \times P(h=y)_{cycle j} \right\} \times \Delta t \quad (15)$$

$$\tau_s = \tau_s + \frac{(1-NT)}{NT} \times LDDT \quad (16)$$

따라서 임의 데이터 패킷이 전체 네트워크를 통과하는데 걸리는 평균 시간, τ_s 는 식(16)과 같이 계산된다. 여기서, N/T 는 식(1,2,3)에서 계산 될 수 있고, τ_s 는 식(14,15)로부터 계산할

수 있으며, $LDDT$ 는 네트워크 특성에 따라 상수로 주어진다.

3 결론

본 논문에서는, $a \times a$ 출력 버퍼 스위치로 구성되는 핫스팟 상황 하에 다단 연결망의 성능을 확률 식으로 분석하고, 실효성을 입증하였다. 제안된 분석 기법은 네트워크 스위치 내부에서 핫스팟이 발생하는 데이터 패킷과 일반 데이터 패킷의 이동 상태를 관찰하여 확률 식으로 정리하고, 이를 토대로 네트워크 전체의 정상상태 처리율 및 네트워크 지연시간을 예측한다. 분석 모형의 수립 단계에서 정상상태 확률 개념을 도입하여 간단한 근사화(approximation)를 시도하여 모형의 해석과 확률 식 전개를 용이하게 하였다. 또한 본 논문에서는 모형의 이해를 둘기 위하여 네트워크 트래픽 제어 및 종도 소실 패킷에 대한 처리 기능 등 최근 개발되는 스위치 네트워크의 부기기능을 배재하고 수식을 정리하였다. 그러나 제안된 분석모형은 이를 다양한 성능 향상 기술이 적용된 네트워크, 그리고 다양한 크기의 네트워크 성능분석에도 쉽게 적용이 가능하다. 모형의 실효성 검토를 위하여 병행된 시뮬레이션 결과는 분석 모형에 의하여 얻은 결과와 상호 미세한 오차 범위 내에서 일치하여, 제안된 분석 기법의 우수성을 입증하였다.

표 1. 8×8 다단 연결망의 성능

Buffer size	데이터 패킷 입력률(≤ 1.0)					
	정상상태 처리율(NT, %)		네트워크 통과한 데이터 패킷 탈락 확률(%)			
	해석	시뮬레이션	해석	시뮬레이션	해석	시뮬레이션
0	42.33	42.36	3.000	3.000	51.67	57.64
1	55.55	55.59	4.768	5.000	44.45	44.41
2	58.05	58.15	7.089	7.300	41.95	44.85
4	58.80	58.81	12.37	12.50	41.20	41.19
8	58.71	58.81	23.75	23.91	41.29	41.19
16	58.89	58.83	47.31	47.62	41.11	41.17
32	58.70	58.81	95.52	95.69	41.30	41.19

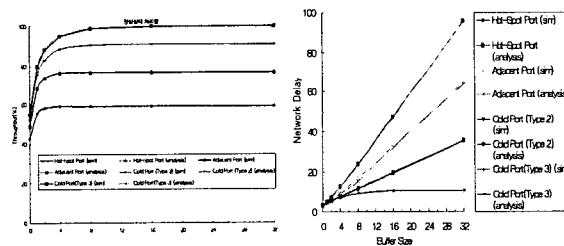


그림 4 8×8 MIN 최종 출력 단들의 성능 분석

4 참고 문헌

- [1] S. H. Byun, D. K. Sung, "The UniMIN Switch Architecture for Large-Scale ATM Switches", IEEE Trans. on Networking, Vol. 8, NO. 1, pp109-120, Feb. 2000.
- [2] 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.
- [3] Mahmoud Saleh and Mohammed Atiquzzaman, "Analysis of Shared Buffer Multistage Networks with Hot Spot", IEEE First International Conference on Algorithms and Architectures for Parallel Processing, 1995. ICAPP 95
- [4] Chuan-Lin Wu and Tse-Yun Feng, "On a class of Multistage Interconnection Networks", IEEE Trans. on Computers, Vol. C-29, No. 8, pp 108-116, Aug. 1980.
- [5] Myung K Yang and Tae Z Shin, "Performance Evaluation of the Buffered MIN with $a \times a$ Switches", KISS Conf. on Parallel Processing, pp244-246, Nov. 2000.
- [6] G.F. Pfister and V.A. Norton, "Hot spot contention and combining in multistage interconnection networks," IEEE Transactions on Computers, Vol. C34, no.10, pp.943-948, October 1985.