Buffered-MIN의 성능 분석

신 태지 · 양 명국
울산대학교 전기전자 및 자동화 공학부

Evaluation of a Buffered Multistage Interconnection Network

Tai Z. Shin, Myung K. Yang
University of Ulsan

요 약

본 논문에서는, multiple-buffered crossbar 스위치를 이용한 다층 연결 망의 성능 분석 모형을 제안하고, 스위치에 장착된 buffer의 개수 증가에 따른 성능 향상 추이를 분석하였다. Buffered 스위치 기법은 다층 연결 망(Multistage Interconnection Network, MIN)의 내부의 데이터 충돌 문제를 효과적으로 해결할 수 있는 방법으로 알려져 있다. 제안된 성능 분석 모형은 먼저 네트웍 내부 임의 스위치 입력 단에 유입되는 데이터 패킷이 buffered 스위치 내부에서 전송되는 패턴을 확률적으로 분석하여 수립하였다. 분석 모형의 수학적 복잡도 절감을 위하여 확률식 유도 과정에 정상상태 확률 개념(steady state probability)을 도입하였다. 제안한 모형은 스위치의 크기 및 스위치에 장착된 buffer의 수와 무관하게 확대 적용이 가능하다. 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. 2×2 스위치로 구성된 8×8 MIN을 대상으로 분석을 시행한 결과 스위치에 2~4개의 buffer를 장착했을 경우 unbuffer 스위치 경우와 비교하여 네트웍 정상상태 Throughput의 증가율이 높고 네트웍 Delay 또한 낮아져 효율적인 것으로 나타났다. 따라서 2×2 crossbar 스위치로 구성된 MIN의 경우 스위치에 장착된 buffer의 개수가 네 개정도일 경우가 가격 대 성능비 면에서 가장 유리한 것으로 연구되었다.

1. 서 론

WAN으로부터 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 Network과 함께 넓은 Bandwidth, Network 유연성 등의 장점을 보유한 효율적인 네트웍으로 평가되고 있다. 다층 연결 망을 통한 데이터 이동에는 각 Stage에서 스위치마다 제어가 요구되고, 데이터 이동 경로에 따라 특정 스위치에서 두 개 이상의 데이터가 하나의 경로로 진행하고자 하는 데이터 충돌 현상이 초래되기고 한다. 데이터 충돌 현상은 네트웍 성능 저하를 유발함은 물론이고 전체 네트웍의 신뢰도에도 큰 영향을 미치게 된다.

본 논문에서는 이미 발표된 buffered MIN 관련 연구[1-4]의 취약점을 보완하고 Buffered $a \times a$ crossbar 스위치의 성능분석[5]을 바탕으로 하여 네트웍 성능 평가의 두 가지 주요 요소로 알려진 네트웍 Throughput과 Delay를 분석하였다. 본 논문에 제안된 복수 buffered 다층 연결 망의 성능 분석 모형은 스위치에 장착된 buffer의 개수와 무관하게 적용 가능하고, 또한 Baseline network을 대상으로 제안된 본 연구의 분석 모형은 모든 다층 연결 망의 성능 분석에 확대 적용 가능하다.

2. Buffered MIN의 성능 분석

2.1. 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 출력단으로 향하게 된다. 그림 1는 네트웍 내부 임의 스위치에서 데이터 패킷이 이동하는 패턴을 확률적으로 해석하여 도식화한 것이다. 스테이지 i에 위치한 임의 스위치 출력 단 D_0로 두 개의 데이터 패킷이 지향할 확률, $P(h=2)_i$,은

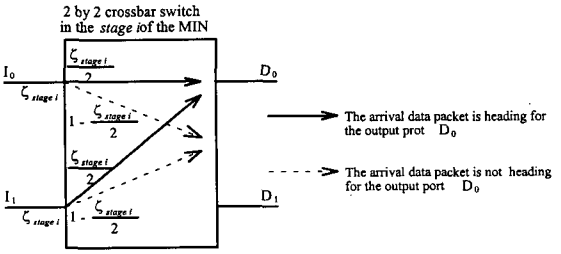

그림 1. 네트웍 내부 임의 스위치에서 데이터 패킷의 이동 확률

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

로 계산된다. 스테이지 i 에 위치한 임의 스위치 출력 단 D_0로 한 개의 데이터 패킷이 지향할 확률, $P(h=1)_i$,은

$$P(h=1)_i={}_2C_1\times(\frac{\zeta_{stage\ i}}{2})\times(1-\frac{\zeta_{stage\ i}}{2}) \tag{2}$$

이고, 스위치 출력 단 D_0로 한 개의 데이터 패킷도 지향하지 않을 확률, $P(h=0)_i$, 은

$$P(h=0)_i=(1-\zeta_{stage\ i}\ /\ 2)^2 \tag{3}$$

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

2.2. Buffered MIN의 성능 분석

네트웍 내부 스테이지 i에 위치한 임의 2×2 crossbar 스위치 내

부 데이터 이동 패턴의 확률적 분석을 토대로 Buffered 다층 연결 망의 성능 분석을 위하여 사용될 변수는 다음과 같다.

b : 스위치에 장착된 buffer 수

ε : buffer에 저장된 데이터 패킷 수

$P(\varepsilon=k)_i$: buffer에 저장된 데이터 패킷 수가 k개일 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

2.2.1 Throughput

스위치 출력 단 D_γ로 데이터 패킷이 출력되지 않는 경우는 해당 출력 단 buffer에 데이터 패킷이 저장되지 않은 상태에서, 스위치 입력 단에서 해당 출력 단으로 지향하는 데이터 패킷이 없을 경우이다. 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. 식 (4)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 저장하고 있을 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 되었든 데이터 패킷은 싸이클 $(j-1)$에 해당 출력 단을 지나 다음 스테이지의 스위치 입력 단을 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 출력단으로 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

$$\begin{aligned}P(\varepsilon=0)_{i,cycle(j-1)} &= P(\varepsilon=1)_{i,cycle(j-2)}\times P(\hbar=0)_{i,cycle(j-1)}\\ &+P(\varepsilon=0)_{i,cycle(j-2)}\times P(\hbar=1)_{i,cycle(j-1)}\\ &+P(\varepsilon=0)_{i,cycle(j-2)}\times P(\hbar=0)_{i,cycle(j-1)}\end{aligned} \quad (5)$$

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

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

이다. 식 (4), (5) 그리고 (6)등의 확률 식에서 임의 buffer가 싸이클 j에 k개의 데이터 패킷을 저장할 확률과 싸이클 $(j+1)$에 k개의 데이터 패킷을 저장할 확률은 같다고 볼 수 있다. 즉, 이들 식에 정상 상태 확률(steady state probability) 개념을 적용하여 $P(\varepsilon=k)_{i,cycle\ j}=P(\varepsilon=k)_{i,cycle(j+1)}$, $P(\hbar=x)_{i,cycle\ j}=P(\hbar=x)_{i,cycle(j+1)}$이 된다. 따라서 $P(\varepsilon=0)_i$을 다시 쓰면

$$\begin{aligned}P(\varepsilon=0)_i=&P(\varepsilon=1)_i\times P(\hbar=0)_i\\ &+P(\varepsilon=0)_i\times P(\hbar=1)_i\\ &+P(\varepsilon=0)_i\times P(\hbar=0)_i\end{aligned} \quad (7)$$

이 되며, 이를 정리하여 $P(\varepsilon=1)_i$를 $P(\varepsilon=0)_i$의 식으로 구하면

$$\begin{aligned}P(\varepsilon=1)_i&=P(\varepsilon=0)_i\times\frac{(1-P(\hbar=0)_i-P(\hbar=1)_i)}{P(\hbar=0)_i}\\ &=P(\varepsilon=0)_i\times\frac{P(\hbar=2)_i}{P(\hbar=0)_i}\end{aligned} \quad (8)$$

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

$$\begin{aligned}P(\varepsilon=k-1)_i=&P(\varepsilon=k)_i\times P(\hbar=0)_i\\ &+P(\varepsilon=k-1)_i\times P(\hbar=1)_i\\ &+P(\varepsilon=k-2)_i\times P(\hbar=2)_i\end{aligned} \quad (9)$$

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

$$P(\varepsilon=k)_i=P(\varepsilon=k-1)_i\times\frac{P(\hbar=2)_i}{P(\hbar=0)_i} \quad (10)$$

로 계산된다. 식 (8) 그리고, (10)의 $\frac{P(\hbar=2)_i}{P(\hbar=0)_i}$을 Ω로 놓고 회귀적 기법으로 $P(\varepsilon=k)_i$를 구하면

$$P(\varepsilon=k)_i=P(\varepsilon=0)_i\times\Omega^k \quad (11)$$

이 된다.

여기서 식 (4)의 $P(D_\gamma=0)_i$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

$$\sum_{k=0}^{b}P(\varepsilon=k)_i=\sum_{k=0}^{b}P(\varepsilon=0)_i\times\Omega^k=1 \quad (12)$$

이 되어, 식 (12)의 $P(\varepsilon=0)_i$s는

$$P(\varepsilon=0)_i=\frac{1}{\sum_{k=0}^{b}\Omega^k} \quad (13)$$

와 같이 계산된다.

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

$$P(D_\gamma=1)_i=1-P(D_\gamma=0)_i \quad (14)$$

이고, 이것은 식 (4), (11), 그리고 (13)을 이용하여 구할 수 있다. 네트웍 구조 상 스테이지 i의 임의 스위치 출력 단으로 데이터 패킷이 출력될 확률, $P(D_\gamma=1)_i$,은 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. $N\times N$ 다층 연결 망의 경우 전체 네트웍 출력 단으로 출력되는 데이터 패킷의 수, OP,는

$$OP=N\times P(D_\gamma=1)_{last\ stage} \quad (15)$$

로 계산된다. 또한 다층 연결 망 입력 단으로 매 싸이클마다 유입되는 총 데이터 패킷의 수를 $IP(=N\times\zeta_{stage_0})$라 하면, 네트웍 정상 Throughput, NT(Normalized Throughput),은

$$NT=\frac{OP}{IP}=\frac{P(D_\gamma=1)_{last\ stage}}{\zeta_{stage\ 0}} \quad (16)$$

와 같이 얻어진다.

2.2.2 Delay

임의 데이터 패킷이 다층 연결 망을 통과하는데 소요되는 시간의 분석과정은 먼저, 데이터 패킷이 망 내부의 각 스테이지 별 스위치를 통과하는데 필요한 시간을 분석하고, 이들 각 스테이지 별 체류 시간을 합하여 전체 네트웍 지연시간을 계산하게 된다.

다층 연결 망 내부 임의 2×2 복수 buffered crossbar 스위치에서 싸이클 j에 특정 출력 단으로 지향하여 buffer에 도착한 데이터 패킷이 k번째 buffer에 저장될 경우를 살펴보면 다음과 같다. 먼저, 싸이클 $(j-1)$종료 시 해당 출력 단 buffer에 k개 패킷이 저장된 상태에서, 싸이클 j에 새로운 한 개 이상의 패킷이 해당 출력 단을 지향하고, 이들 가운데 첫 번째로 선택되어 buffer에 저장될 경우를 들 수 있다. 또한 싸이클 $(j-1)$종료 시 해당 출력 단 buffer에 $(k-1)$개 패킷이 저장된 상태에서, 싸이클 j에 새로 2개 패킷이 도착되고 이들 가운데 두 번째 순위로 buffer에 저장될 경우 역시 해당 데이터 패킷이 k번째 buffer에 저장된다. 일단 데이터 패킷이

k번째 buffer에 저장되면 해당 데이터 패킷은 현 스위치에 $(k+1)\times\Delta t$ 시간만큼 머무르게 된다. 따라서 네트웍 내부 스위치 스테이지 i에 위치한 임의 스위치를 통과하여 다음 스테이지로 이동에 성공한 데이터 패킷이 해당 스위치에 체류한 시간, $\tau_{s,stage\ i}$, 을 확률식으로 구하면

$$\tau_{s,stage\ i}=\sum_{k=0}^{b}\Big[\sum_{\varphi=k-1}^{b}\Big\{P(\varepsilon=\varphi)_{i,cycle(j-1)}\times\sum_{y=k+1-\varphi}^{2}P(\hbar=y)_{i,cycle\ j}\Big\}\times(k+1)\Delta t\Big] \quad (17)$$

으로 얻어진다. 여기서 $P(\varepsilon=-1)=0$이다. 한편, 스위치 buffer 공간이 여의치 않아 데이터 충돌 시 경합에서 제외되는 데이터 패킷은 데이터 중도 유실을 감지하는 특정 과정을 거쳐, 네트웍 입력 단에서 재 전송된다. 따라서 데이터 충돌로 인하여 중도 소실된 데이터의 재전송 처리 시간을 $SDST$(Surplus Data Service Time)이라 정의하고, 이에 따른 중도 소실 지연시간, $\tau_{f,stage\ i}$, 을 데이터 소실 확률과 함께 계산하면

$$\tau_{f,stage\ i}=\{\zeta_{stage\ i}-P(D_r=1)_i\}\times SDST \quad (18)$$

여기서 $\{\zeta_{stage\ i}-P(D_r=1)_{stage\ i}\}$은 해당 데이터가 스테이지 i에 위치한 임의 스위치에서 충돌로 인하여 제거될 확률을 나타낸다. 네트웍 정상 Throughput 모형 개발에서와 같이 정상 상태 확률 개념을 도입하고 식 (11)를 이용하여 식 (17)을 정리하고 식 (18)과 함께 스테이지 i에 위치한 임의 스위치 통과에 소요되는 시간, $\Delta T_{stage\ i}$, 를 구하면

$$\Delta T_{stage\ i}=\tau_{s,stage\ i}+\tau_{f,stage\ i}=$$

$$\sum_{k=0}^{b}\Big[\sum_{\varphi=k-1}^{b}\Big\{\Omega^{\varphi}P(\varepsilon=0)_i\times\sum_{y=k+1-\varphi}^{2}P(\hbar=y)_i\Big\}\times(k+1)\Delta t\Big]+\{\zeta_{stage\ i}-P(D_r=1)_i\}\times SDST \quad (19)$$

이 된다. 여기서 $\Omega=\dfrac{P(\hbar=2)_i}{P(\hbar=0)_i}$이다. 식 (19)의 $P(\varepsilon=0)_i$와 $P(\hbar=y)_i$는 네트웍 스테이지별로 $\zeta_{stage\ i}$가 주어지면 식 (13)과 식 (1)~(3) 등을 이용하여 구할 수 있다. 일단 임의 스테이지 i를 통과하는데 소요되는 시간 $\Delta T_{stage\ i}$를 구하면, 전체 네트웍을 통과하는데 소요되는 시간, ΔT,은 각 스테이지별 지체 시간을 합하여

$$\Delta T=\sum_{i=0}^{n-1}(\tau_{s,stage\ i}+\tau_{f,stage\ i})$$

$$=\sum_{i=0}^{n-1}\Big[\sum_{k=0}^{b}\Big[\sum_{\varphi=k-1}^{b}\Big\{\Omega^{\varphi}P(\varepsilon=0)_i\times\sum_{y=k+1-\varphi}^{2}P(\hbar=y)_i\Big\}\times(k+1)\Delta t\Big]+\{\zeta_{stage\ i}-P(D_r=1)_i\}\times SDST\Big] \quad (20)$$

로 계산된다. 여기서 $n=\log_2 N$ 이다.

표 1은 2×2 multiple-buffered crossbar 스위치들을 사용한 8×8 Baseline 네트웍을 대상으로 매 싸이클 동안 하나씩의 입력이 유입되는 경우, 즉 $\zeta_{stage\ 0}=1$일 때 분석 모형으로부터 예측한 결과와 시뮬레이션 결과를 나타낸 것이다. 그림 2와 그림 3은 각각 Buffer 개수 증가에 따른 Throughput과 네트웍 Delay 변화를 그래프로 나타낸 것이다. ($SDST$=30 Δt)

3. 결론

본 연구에서는 복수 buffered 다층 연결 망의 분석모델을 제시하였다. 제시된 분석모델은 스위치에 장착된 buffer 의 개수와 무관하게 적용 가능하고, 분석과정에서 간단한 데이터 충돌 처리 기법을 도입하여 모델의 수식 이해가 용이하다. 제안한 수학적 성능 분석 연구의 실효성을 검증하기 위한 시뮬레이션 처리결과는 상호 미세한 오차 범위 내에서 모형의 예측 데이터와 일치하는 결과를 보여 분석 모형과 함께 수정된 데이터 충돌 방식의 타당성을 입증하였다. 또한 Baseline network 을 대상으로 제안된 본 연구의 분석모델은 모든 다층 연결 망의 성능 분석에 확대 적용 가능하다.

4. 참고문헌

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Kyung W. Park, Myung K. Yang, "Buffered $a\times a$ 스위치성능분석", 한국 정보 과학회 '98 가을 학술발표회 논문집, 제 25권 2호, pp630-632, 1998

표 1. 2×2 buffered crossbar switches를 이용한 8×8 MIN의 성능 (ζ=1)

buffer size	Normalized Throughput(NT,%)		Network Delay(Δt)			
			Delay for success packets(Δt)		Discarded 확률(%)	
	Analysis	simulation	Analysis	Simulation	Analysis	Simulation
0	51.65	51.64	3	3	48.35	48.36
1	75.29	74.69	4.45	4.45	24.71	25.31
2	83.46	82.86	5.78	5.79	16.54	17.14
4	90.06	89.50	8.40	8.43	9.95	10.50
8	94.47	94.04	13.58	13.67	5.53	5.95
16	97.07	96.90	23.92	24.17	2.93	3.10
32	98.49	98.36	44.59	44.95	1.51	1.64

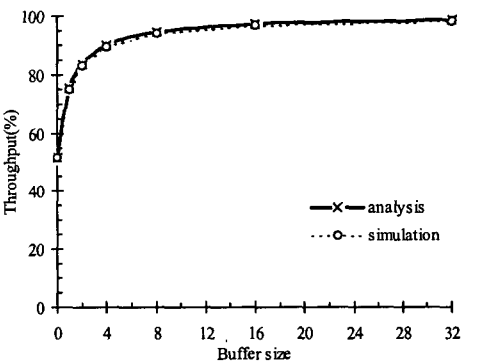

그림 2. Throughput vs. Buffer size

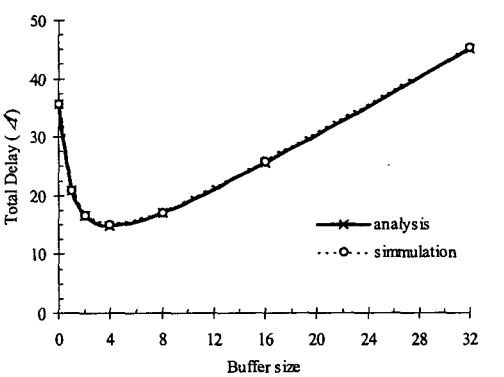

그림 3. Delay vs. Buffer size(ξ=1, $SDST$=30Δt)