

논문 2004-41TC-9-9

사이클릭 벤안 망의 셀 순서 무결성 보장을 위한 셀 재배열 버퍼

(The Cell Resequencing Buffer for the Cell Sequence Integrity Guarantee for the Cyclic Banyan Network)

박 재 현*

(Jae-Hyun Park)

요 약

*본 논문에서는, 고성능 결함 감내 셀 스위치인, 사이클릭 벤안 망의 셀 순서의 무결성 문제를 해결하기 위한 셀 재배열 버퍼를 제시한다. 사이클릭 벤안 스위치는, 편향 자기 경로제어를 사용하여, 입력 정합과 출력 정합 사이에 다중 경로들을 제공함으로써, 높은 신뢰성을 제공하고, 스위치의 내부 링크들의 혼잡 문제를 해결한다. 그런데, 이러한 다중 경로들은 길이가 서로 다를 수 있다. 따라서 셀들이 입력 정합에 도착한 순서와 다르게 출력 정합에 도달할 수 있다. 제안된 셀 재배열 버퍼는 이러한 셀 순서의 무결성 문제를 해결하는 일종의 하드웨어 슬라이딩 윈도우 매커니즘이다. 본 장치 구성의 주요 비용은 슬라이딩 윈도우를 구성하는 하드웨어 비용이다. 따라서 필요한 슬라이딩 윈도우의 크기를 계산하기 위해서, 비균일 주소 분포를 가진 트래픽 부하 하에서 스위치를 시뮬레이션하여, 셀들이 스위치를 통과할 때 발생하는 지연 분포를 분석을 하였다. 이 분석을 통하여, 적은 양의 범용 메모리와 제어 논리를 사용하여, 셀 순서의 무결성 문제를 해결하는 셀 재배열 버퍼를 만들 수 있다는 사실을 밝혔다. 본 논문에서 제시한 셀 재배열 버퍼는 다른 다중 경로 스위칭 망들을 위해서도 사용될 수 있다.

Abstract

In this paper, we present the cell resequencing buffer to solve the cell sequence integrity problem of the Cyclic banyan network that is a high-performance fault-tolerant cell switch. By offering multiple paths between input ports and output ports, using the deflection self-routing, the Cyclic banyan switch offer high reliability, and it also solves congestion problem for the internal links of the switch. By the way, these multiple paths can be different lengths for each other. Therefore, the cells departing from an identical source port and arriving at an identical destination port can reach to the output port as the order that is different from the order arriving at input port. The proposed cell resequencing buffer is a hardware sliding window mechanism. to solve such cell sequence integrity problem. To calculate the size of sliding window that cause the prime cost of the presented device, we analyzed the distribution of the cell delay through the simulation analyses under traffic load that have a nonuniform address distribution that express the property of traffic of the Internet. Through these analyses, we found out that we can make a cell resequencing buffer by which the cell sequence integrity is to be secured, by using a few of ordinary memory and control logic. The cell resequencing buffer presented in this paper can be used for other multiple paths switching networks.

Keywords : Cell Resequencing Buffer, Cell Sequence Integrity Guarantee, Cyclic Banyan Network

I. 서 론

일반 가입자들 및 작은 사무실들 그리고 기업들의 망들을 구축하기 위해, ADSL 및 VDSL, 그리고 Cable

Modem과 WLAN 같은 많은 대역폭을 요구하는 가입자 망 기술들이 사용되고 있다. 또한 이동 전화 같은 단말의 데이터 서비스 요구도 초고속 데이터 통신의 신규 수요를 창출하고 있다. 이러한 광대역 가입자 망들의 수용을 위해서는 대용량의 고속 셀 스위칭 라우터 장비가 필수적이다^[1].

대형 고속 셀 스위칭 시스템을 구성하기 위해서, 일반적으로 다단 스위칭 망으로 스위칭 시스템을 구성하게 된다^[2]. 다단 스위칭 망은 스위치 페브릭들로 구성된

* 정회원, 중앙대학교 공과대학 컴퓨터공학부
(School of Computer Science and Engineering,
Chung-Ang University)

※ 이 논문은 2004년도 중앙대학교 학술연구비 지원에 의한 것임.

접수일자: 2003년10월4일, 수정완료일: 2004년6월28일

다. 이러한 셀 스위칭 시스템의 주요 수요를 발생시키는 인터넷 트래픽은 폭발적이어서, 여러 입력 링크들을 통해 하나의 스위치 페브릭으로 동시에 도착하는 셀들이 동일한 출력 링크를 통해 전달될 가능성이 매우 크다. 더 나아가, 스위칭 시스템이 커질수록, 이러한 불균형의 트래픽 패턴은 더욱 심화되며, 결국 이러한 링크의 출력 경쟁의 확률은 더욱 높아진다^[2].

더불어 초고속의 데이터 통신이 기간 서비스가 되면서, 고품질의 서비스를 제공하기 위해서, 스위칭 라우터 시스템은 고신뢰성을 제공하여, 일반적으로 중단없이 운영된다. 특히 스위칭 망의 신뢰성은 이러한 스위칭 시스템의 신뢰성의 핵심 요소이다^[3]. 그런데 많은 스위칭 망들의 기반 망인 벤안 망은 각 입출력 쌍에 대해서 각각 유일한 경로만을 갖기 때문에 결합 감내성이 없는 문제점을 가지고 있다.

이러한 성능과 신뢰성 문제를 해결하기 위해, 많은 기법들이 개발되어 왔다. 이들 연구들은 라우팅 망 앞에 분산 망을 추가하거나, 여러 스위칭 망을 병렬로 사용하거나, 셀을 다시 입력버퍼에 넣거나, 스위칭 망에 하나 이상의 스위칭 단을 추가하거나, 셀을 편향(Deflection)시키거나, 내부에 추가적인 링크를 넣거나, 외부 속도에 비해 스위칭 망 내부의 속도를 증가시키거나, 내부 버퍼를 사용하거나, 다중 경로 스위칭 망을 사용하는 방법들이다^[2,3,4]. 그러나 이들 대부분은 지나치게 많은 하드웨어를 필요로 하거나, 복잡한 라우팅 방법을 필요로 하기도 하고, 심지어 둘 다 필요로 하기도 한다.

편향 자기 경로제어(Deflection Self-routing)를 사용하는 사이클릭 벤안 망(Cyclic Banyan Network)은 혼잡이 발생하거나, 결합이 내부 링크들에 생겼을 때, 스위치 내부의 모든 링크들을 사용하여, 다중 경로들을 제공한다^[2,5]. 이 스위치 구조는 많은 우회 경로들을 제공함으로써, 인터넷의 폭발적 트래픽 하에서 발생하는 내부 링크의 혼잡을 해결한다. 또한 결합이 발생하는 경우, 이를 우회하는 많은 경로들을 제공하여 높은 가용성을 가진다. 그러나 사이클릭 벤안 망이 제공하는, 동일한 원천지와 동일한 목적지 사이의 다중 경로들의 거리가 서로 다를 수 있기 때문에, 다른 모든 다중 경로 스위칭 망과 마찬가지로, 스위칭 되는 셀들의 순서의 무결성이 깨어질 수 있는 단점을 역시 가지고 있다.

다중 경로들을 제공하는 많은 스위치들의 셀 순서의 무결성 문제를 해결하기 위해, 다음과 같은 연구들이 있다. 각 스위치들은 각기 다른 방법으로 다중 경로들을 제공하기 때문에, 각각 다른 방법들로 셀 순서의 무

결성 문제를 해결하고 있다^[6,7,8,9,10]. 다중 경로를 제공하는 Clos 망의 경우, 중간 스위치 단계에서 버퍼를 사용하지 않도록 하여, 순서 일관성을 유지하는 방법이 제안되었다^[6]. 이 방법은 다중 경로들의 길이가 항상 일정할 때에 한하여 사용이 가능하다. 병렬 스위치들을 사용하여 셀들을 분산시키는 스위치 구조의 경우, 병렬 스위치들의 수가 입력 큐에 들어있는 셀의 갯수와 일치하지 않으면, 모조 셀을 추가하여, 셀들이 출력 단에 도착했을 때, 셀의 순서 무결성이 유지되도록 하는 방법이 제안되었다^[7]. 이 방법은 필요없는 셀들을 스위칭하여 스위칭 용량을 낭비하는 단점을 가지고 있다. 병렬 스위치 구조의 경우에 사용하는 또다른 방법으로는, 시간 정보를 셀에 추가하여, 출력 정합에서 이 시간 정보를 사용하여 순서의 무결성을 맞추는 하드웨어를 사용하는 방법이 있는데, 이 방법은 지나치게 하드웨어 비용이 많이 드는 것이 단점이다^[8,9]. Benes 망과 Clos 망의 경우는 같은 원천지와 같은 목적지를 가지는 셀들을 동일한 경로로 스위칭되도록, Per-VC 경로를 제공하는 경로 스케줄링을 해서, 셀 순서의 무결성을 유지하는 방법이 제안되었는데^[10], 이러한 방법 역시 주요 자원인 스위칭 용량을 낭비하는 결과를 가져온다.

본 논문에서는 사이클릭 벤안 스위치의 실제 활용도를 최대한으로 유지하면서, 스위칭 셀 순서의 무결성을 제공하기 위한, 셀 재배열 버퍼를 제안한다. 그리고, 셀 재배열 버퍼의 주요 하드웨어 비용의 대부분을 결정하는 슬라이딩 윈도우의 크기를 시뮬레이션을 사용하여 분석하여, 본 재배열 버퍼를 적은 하드웨어 비용으로 구성할 수 있음을 보인다.

II장에서 사이클릭 벤안 스위칭 망과 편향 자기 경로제어에 대해 간략히 설명한다. III장에서는 사이클릭 벤안 망의 셀 순서의 무결성 문제를 제기하고, 해결책으로 셀 재배열 버퍼를 제시한다. 끝으로 IV장에서 결론을 맺는다.

II. 사이클릭 벤안 망과 편향 자기 경로제어 셀 스위칭

1. 사이클릭 벤안 스위치

사이클릭 벤안 망은 벤안 망에 링크들을 추가하고, 이러한 링크들과 기존의 모든 링크들을 모두 경로 설정을 위해 사용하는 자기 경로제어 알고리즘을 채용한 다단 스위칭 망이다^[2,5]. 이 스위칭 망의 구조와 동작은 이전의 연구들^[2,5]에 잘 기술되어 있으나, 본 논문에서

기본적인 이해를 제공하기 위해 다시 기술 한다. 우리는 C. Wu^[11]의 명명법(naming scheme)을 확장하여, 사이클릭 벤안 망의 형상(configuration)과 경로제어 알고리즘을 기술하기 위해서 사용한다. 각 단계는 첫 스위칭 단계를 1로 하고, 마지막 스위칭 단계를 $\log_2 N$ 으로 하여, 번호를 연속으로 붙여 표현한다. N 개의 입출력 단들은 $\log_2 N$ 자리수의 이진수로 번호를 붙여 나타낸다. 즉 $link(p_{\log_2 N}, p_{(\log_2 N)-1}, \dots, p_1)$ 로 표현한다.

각 단계에서의, 하나의 단위 스위치(Switching Element: SE)는 $(\log_2 N) - 1$ 자리수의 이진수들로, 즉 $(p_{(\log_2 N)-1}, p_{(\log_2 N)-2}, \dots, p_1)$ 로, 표현한다. 이는 단계 내에서의 단위 스위치의 위치를, 위로부터 시작하여 아래로 내려가며, 이진수로 표현한 것이며, 뒤에서 레벨이라고도 부른다. 입력 셀의 목적지 주소는 $A = (a_1, a_2, \dots, a_{\log_2 N})$ 로 표현 한다. 이제 단위 스위치와 링크를 간략히 기술하기 위해, 약식 기호 $n_{i,j}$ 를 정의한다. 기호 $n_{i,j}$ 는 스위칭 단계 i , 레벨 j 의 SE를 나타낸다. 여기서 $1 \leq i \leq \log_2 N$ 이고 $0 \leq j < N/2$ 이다. 이제, 사이클릭 벤안 망을 다음과 같이 정의하자.

정 의 1: 사이클릭 벤안 망을 각 출력 링크에 대한 위상으로 기술하면 다음과 같다^[2,5]. 사이클릭 벤안 망의 스위칭 단계 i 의 각 SE $(p_i, p_{i-1}, \dots, p_1)_i$ 는, 그 SE의 각 출력 링크에 대해, 스위칭 단계 $i + 1$ 의 하나의 SE $(p_{i-1}, p_{i-2}, \dots, p_1, p_0)_{i+1}$ 로, 다음과 같이 연결된다.

$$\beta_0[(p_i, p_{i-1}, \dots, p_1)_i] = (p_{i-1}, p_{i-2}, \dots, p_1, 0)_{i+1}, \quad (1)$$

$link(p_i, p_{i-1}, \dots, p_1, 0)_i$ 로 연결되고,

$$\beta_1[(p_i, p_{i-1}, \dots, p_1)_i] = (p_{i-1}, p_{i-2}, \dots, p_1, 1)_{i+1}, \quad (2)$$

$link(p_i, p_{i-1}, \dots, p_1, 1)_i$ 로 연결된다.

여기서 $l = (\log_2 N) - 1, 1 \leq i \leq l$ 이다.

연쇄 링크들의 위상적 기술은 다음과 같다. 사이클릭 벤안 망의 스위칭 단계 i 의 각 SE $n_{i,j}$ 는, 그 SE의 각 추가된 링크에 대해, 같은 스위칭 단계의 하나의 SE로, 다음과 같이 연결된다.

$$\gamma[n_{i,j}] = n_{i,(j+1) \bmod (N/2)}, \quad (3)$$

$$\gamma_{-1}[n_{i,j}] = n_{i,(j+N/2-1) \bmod (N/2)} = \gamma_{N/2-1}[n_{i,j}]. \quad (4)$$

식 (1)과 식 (2)의 연산 β 는 스위칭 단계 i 의 한 SE로부터 스위칭 단계 $i+1$ 의 두 SE들로의 사상(Map-ping)으로 기존 링크들에 의한 상호 연결들을 표현한다. 연산 β 는 SE의 상위 출력 링크와 연결된 SE를 표

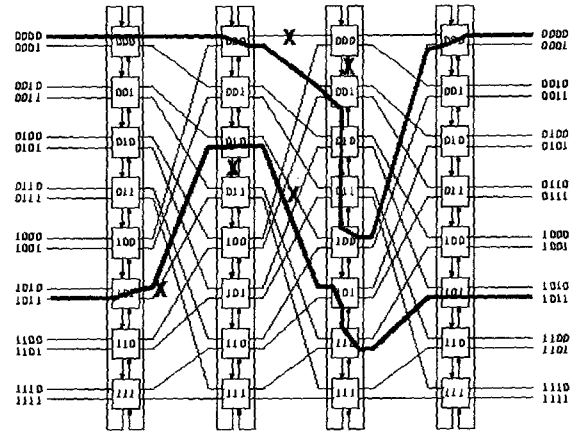


그림 1. 사이클릭 벤안 망과 적응 자기 경로제어 알고리즘

Fig. 1. The Cyclic Banyan network and the fully adaptive self-routing scheme thereof.

현하고, 연산 β 은 하위 출력 링크와 연결된 SE를 표현한다. 식 (3)과 식 (4)의 연산 γ 과 연산 γ_{-1} 는 추가된 링크들에 의한 연결들을 동일한 스위칭 단계 내의 다른 SE로의 사상들로 표현한다. 그림 1에서는, 일반적인 16×16 델타 망으로부터, 16×16 사이클릭 벤안 망을 구성하는 예를 도시하였다.

2. 편향 자기 경로제어 알고리즘

사이클릭 벤안 망의 경로 제어는, 많은 다른 다단 스위칭 망들과 동일하게, 목적지 주소를 가지고 이루어진다. 여기에 추가하여, $\log_2 N$ 비트들 크기의 이탈 꼬리표(Deviation Tag) k 가 사용된다. 이는 각 스위칭 단계에서, 원래 중간 기착지인 SE(혹은 이것의 동치 SE들)로부터 떨어진 위상적 거리를 나타낸다. 결과적으로 목적지 주소는 (A, k) 쌍으로 표현된다. 그리고 단계 1의 입력 링크에 한 셀이 도착 했을 때, k 값은 0으로 설정된다. 다시 말해, 입력 정함에 도달한 셀들의 행선지 주소는 $(A, 0)$ 이다. 이탈 꼬리표 k 의 값은 각 SE에서 갱신되는데, 이는 k 의 현재 값과 선택된 링크 그리고 원래 출력하려던 링크를 인수들로 하여, 간단한 하드웨어로 계산된다. 사용되는 편향(혹은 완전적용) 경로제어 알고리즘은 다음과 같다^[2]. (여기서 i 는 해당 SE의 스위칭 단계 번호이다.)

정 의 2: 사이클릭 벤안 망을 구성하는 각 SE의 편향 자기 경로제어 알고리즘은 다음과 같다^[2,5].

(여기서 $s \in \{0, 1\}$ 이고, $\alpha_s[(p_i, p_{i-1}, \dots, p_1)_i, (a_1, a_2, \dots, a_{i+1})] = (p_{i-1}, p_{i-2}, \dots, p_1, a)_i$ 이다. $(a_1, a_2, \dots, a_{i+1})$ 는 각

셀의 주소이다.)

각 입력 셀에 대하여,

1. $k = 0$ 인 경우,
 - (a) $\beta_s[\cdot]$ ($= \alpha_l[\cdot]$)로 보낸다.
 - (b) 만약 실패하면, $\beta_s[\cdot]$ ($\neq \alpha_l[\cdot]$)로 보낸다.
 - (c) 만약 실패하면, $\gamma_{\alpha_j}[\cdot]$ 로 보낸다,
 - (d) 만약 실패하면, $\gamma_{\alpha_{j+1}}[\cdot]$ 로 보낸다.

2. $k \neq 0$ 인 경우,
 - (1) $\gamma_{\alpha_{\theta(0)}}[\cdot]$ 로 보낸다,
 - (2) 만약 실패하면, $\gamma_{\alpha_{\theta(1)}}[\cdot]$ 로 보낸다 ,
 - (3) 만약 실패하면, $\beta_{\alpha_{\theta(0)}}[\cdot]$ 로 보낸다 ,
 - (4) 만약 실패하면, $\beta_{\alpha_{\theta(1)}}[\cdot]$ 로 보낸다 .

k 를 수정하는 규칙은 다음과 같다:

1. $k = 0$ 일때,
 - (1) 만약 $\beta[\cdot] = \alpha_l[\cdot]$ 그리고 $\beta[\cdot]$ 가 선택되었다면, $k \leftarrow 2^{i-1}$.
 - (2) 만약 $\beta[\cdot] = \alpha_l[\cdot]$ 그리고 $\beta[\cdot]$ 가 선택되었다면, $k \leftarrow 1$.
 - (3) 만약 $\gamma[\cdot]$ 가 선택되었다면, $k \leftarrow 2^{i-1} - 1$.
 - (4) 만약 $\gamma_{-1}[\cdot]$ 가 선택되었다면, $k \leftarrow 1$.
2. $k \neq 0$ 일때,
 - (1) $\beta[\cdot] = \alpha_l[\cdot]$ 그리고 $\beta[\cdot]$ 가 선택되었다면, $k \leftarrow k \times 2 - 1$.
 - (2) 만약 $\beta[\cdot] = \alpha_l[\cdot]$ 그리고 $\beta[\cdot]$ 가 선택되었다면, $k \leftarrow k \times 2 + 1$.
 - (3) 만약 $\beta[\cdot] = \alpha_l[\cdot]$ 그리고 $\beta[\cdot]$ 가 선택되었다면, $k \leftarrow k \times 2$, 여기서, $j \in \{0, 1\}$.
 - (4) 만약 $\gamma[\cdot]$ 가 선택되었다면, $k \leftarrow k - 1$.
 - (5) 만약 $\gamma_{-1}[\cdot]$ 가 선택되었다면, $k \leftarrow (k + 1) \bmod 2^{i-1}$.

여기서 $\alpha(j) = \begin{cases} 1, j \bmod 2 = 1 \text{ 경우} \\ -1, \text{ 나머지 경우} \end{cases}$ 이고

$\theta(m) = \begin{cases} (m+1) \bmod 2, k \leq 2^{i-2} \text{ 경우} \\ (m) \bmod 2, \text{ 나머지 경우} \end{cases}$ 이다.

사이클릭 벤안 망의 편향 자기 경로제어 방법의 정당성은 증명되어 있다^[2]. 그림 1에서는 본 편향 자기 경로제어 알고리즘의 두 가지 예들을 보여주고 있다. (여기서 혼잡 혹은 결함은 X로 표시되어 있다.) 먼저 셀이 입력 단 (0000)에서 출력 단 (0000)으로 전달되어야 하는 경우를 보자. 그림 1에서 보인 것처럼, 원래 할당된

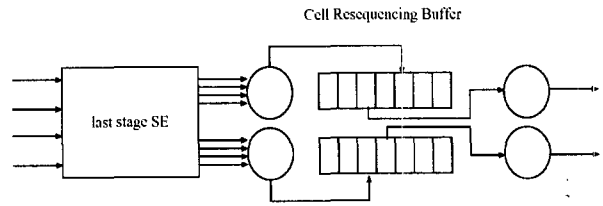


그림 2. 셀 재배열 버퍼의 논리적 구조도
Fig. 2. Logical structure of the cell resequencing buffer.

출력 링크에 결함이 발생한 SE는, 제시한 경로제어 알고리즘에 의해, 셀을 다른 출력 링크로 보낸다. 그 셀은 제시한 경로제어에 의해 추가된 링크들을 통과하여 나간 후, 원래 의도된 중간 기착지인 SE의 동치 SE 중의 하나로 보내지고, 이 SE로부터 최종 목적지가 되는 출력 정합으로 올바르게 보내진다. 또한, 그림 1에서 예시한 바와 같이, 셀들이 입력 단 (1011)로 부터 출력 단 (1011)로 가는 경우에, 정상적인 경로에 3개의 결함들이 발생하여도, 본 스위치와 경로제어 알고리즘을 사용하여, 셀들은 목적지인 출력 정합으로 스위칭 될 수 있다.

III. 사이클릭 벤안 망의 셀 순서 무결성 문제와 셀 재배열 버퍼

본 절에서는, 사이클릭 벤안 망이 거리가 서로 다른 대체 경로들을 제공하여, 셀 순서 무결성 문제를 가지고 있다는 사실을 제기하고, 이를 해결하기 위한 셀 재배열 버퍼를 제시한다. 셀 재배열 버퍼의 주요 하드웨어 비용 요소인, 필요한 슬라이딩 윈도우의 크기를 설계하기 위해, 인터넷 트래픽 하에서의 셀들의 스위칭 지연을 시뮬레이션을 사용해 분석한다.

1. 셀 순서 무결성

사이클릭 벤안 망은 스위칭 망 내부에 혼잡이 발생한 경우나, 링크 결함이 발생한 경우에, 이를 해결하기 위해, II장의 2절에서 제시한 경로제어 방법을 각 단위 스위치에서 사용하여, 적응적 셀 라우팅을 수행한다. 그런데 이 라우팅 방법을 사용하면, 셀들이 동일한 근원지를 출발하여, 동일한 목적지에 도달하는 경우에, 셀들은 근원지에 도착한 순서와 다르게 목적지에 도달할 수 있다. 다른 말로 하면, 셀 순서의 무결성이 깨어질 수 있다. 이는 사이클릭 벤안 망의 편향 경로제어 방법이 각기 다른 지연 시간을 갖는 경로들을 제공하기 때문이다. 이러한 거리가 다른 다중 경로들은 입력 포트의 수

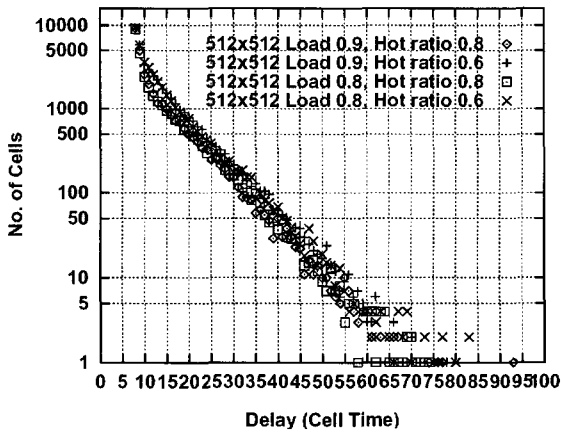


그림 3. 다양한 부하들과 Hot Ratio의 비균일 트래픽 하에서의 512 x 512 시클릭 벤안의 셀 지연 분포

Fig. 3. Cell delay distribution of the 512 x 512 Cyclic Banyan network under the nonuniform traffic of various input load and hot ratio.

를 N 으로 하였을 때, 총 $O(N \log_2 N - 1)$ 의 가지 수 만큼 많이 존재한다. (이는 총 N 개의 출력 링크들이 모두 각 스위칭 단계에서 우회 링크들로 사용될 수 있고, 이러한 스위칭 단계들이 스위치 내에 총 $\log_2 N - 1$ 개가 있기 때문이다.)

우리는 다음에서 셀들이 스위칭되며 발생하는 지연 시간을 버퍼안에서 균일하게 만들고, 이를 일정하게 출력함으로써 셀 순서의 무결성을 유지하는 셀 재배열 버퍼를 제시한다. 그림 2에 도시한 바와 같이, 제한하는 셀 재배열 버퍼의 논리적 구조는 정해진 지연 시간 후에 셀들을 전달하기 위해 셀들을 보관하는, 범용 메모리를 사용하는, 일종의 하드웨어 슬라이딩 윈도우 메커니즘이다.

2. 셀 버퍼의 윈도우 크기

본 절에서는, 스위칭된 모든 셀들의 셀 지연 시간을 균일하게 맞추기 위해 필요한, 슬라이딩 윈도우의 크기를 결정하기 위해서, 셀들이 스위칭 될 때 발생하는 지연 시간들의 분포를 시뮬레이션을 사용해 분석한다. 이를 위해 다음과 같이 일반적으로 사용되는 가정들을 도입한다.

가 정

- 새로운 셀들은 입력 정합들에 도착률 λ 를 갖는 포아송 확률 과정에 따라 도착한다. 여기서 λ 는 $0 \leq \lambda \leq 1$ 이다. 또한, 하나의 셀이 도착할 확률은 모든

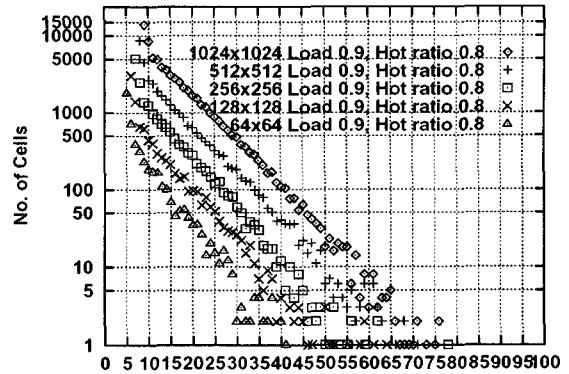


그림 4. 비균일 트래픽하에서의 시클릭 벤안 망의 망 크기의 변화에 대한 셀 지연 분포

Fig. 4. Cell delay distribution versus network size for the Cyclic Banyan network under the nonuniform traffic.

입력 정합에 대해 동일하다. 따라서 N 개의 원천지들은 셀들을 비율 λ 의 포아송 확률 과정에 따라서 생성한다.

- 입력 셀들은 모든 출력 정합들에 대해서 비균일하게 분산된다.
- 각 셀이 경쟁에서 이길 확률은 동일하다.

사용하는 비균일(Nonuniform) 트래픽은 인터넷 트래픽 하에서 스위칭 망으로 짧은 시간 범위 안에 인입되는 부하를 표현한다. 비균일 트래픽은 $N \times N$ 크기의 부하 행렬 $D = \{\lambda_{ij}\}$ 로 나타낼 수 있는데, 여기서 λ_{ij} 는 출력 정합 j 가 목적지인 하나의 셀이, 입력 정합 i 에, 도착할 확률을 의미한다. 따라서, 행렬의 제 i 행의 합은 입력 정합 i 에 추가되는 전체 부하를 나타내고, 제 j 열의 합은 출력 정합 j 에 추가되는 총 입력 부하를 나타낸다.

비균일 트래픽 패턴들은 그 가짓수가 무한대이다 [2,12]. 이러한 수많은 패턴들 중에서, 우리는 실제적인 트래픽 패턴들을 대표하는, Hot-group Model [12]이라고 하는, 비균일 트래픽 패턴을 사용한다. 모든 출력 정합들에 대하여, 입력 셀들은 다음과 같이 비균일적으로 분포한다. 즉 출력 정합은 "집중적으로 셀을 내보내는 출력 집단"과 "대체적으로 쉬는 출력 집단", 이들 두개의 그룹들로 나뉜다. 부하 행렬 D 는 $D = [D_h(\lambda_h) D_c(\lambda_c)]$ 로 정의된다. D_h 의 행의 합은 λ_h 이고 D_c 의 행의 합은 λ_c 이다. 여기서 $\lambda_h + \lambda_c = \lambda$ 이다. 다른 말로 하면, 동일한 크기의 두개의 출력 정합 집단이 있는데, 셀들이 목적지 집단 D_h 로 가는 확률은 λ_h 이고, 목적지 집단 D_c 로 가는 확률은 λ_c 이다. Hot Ratio를 r 이라고 정의하

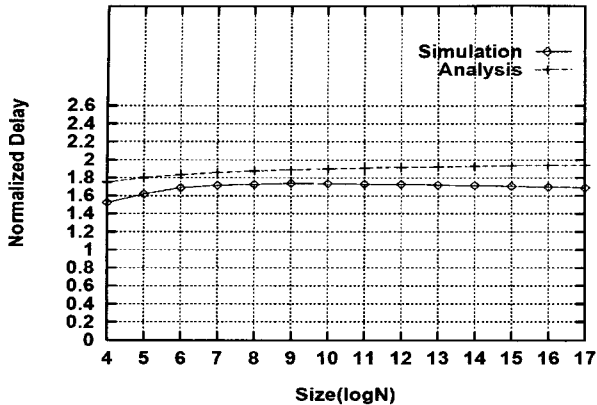


그림 5. 사이클릭 벤안 망의 망 크기의 변화에 대한 정규화된 지연^[2]

Fig. 5. Normalized delay versus network size for the cyclic Banyan network.

면, $\lambda_b = r\lambda$ 이고, $\lambda_c = (1-r)\lambda$ 이다. 여기서 r 은 $0.5 \leq r \leq 1.0$ 이다. 따라서 비균일 트래픽하에서의 모든 성능을 측정하기 위한 모형을 입력 부하 λ 와 Hot Ratio r 로 나타낼 수 있다^[2].

이제 비균일 트래픽 하에서의 사이클릭 벤안 망의 셀 지연 시간의 분포에 대해 알아보자. 그림 3에 보인 바와 같이, 스위칭 망에 과도한 입력 부하인 $\lambda = 0.8$ 과 $\lambda = 0.9$ 의 부하를 걸었을 때, 그리고 Hot Ratio r 을 0.6과 0.8로 했을 때, 512×512 스위칭 망을 통과하는 지연시간은 100 셀 시간 한도 내에 분포함을 볼 수 있었다. 이때 셀 수를 나타내는 Y 축이 로그 스케일 임에 유의하자. 심지어 다양한 크기의 스위칭 망에 대해, 부하 0.9와 Hot Ratio 0.8의 과도한 환경 하에서조차, 셀들이 스위칭 망을 통과하는 지연시간이 100 셀 이하임을 그림 4에서 알 수 있다.

시뮬레이션 결과들로부터 얻은 특기할 사항은, 그림 4에서 보는 바와 같이, 비균일 부하 하에서의 사이클릭 벤안 망을 통과하는 셀 지연이, 망의 크기를 최대 1024×1024 로 한정하였을 때, 100 셀 시간 이하라는 점이다. 또한 그림 5에서는 입력 부하가 1.0인 사이클릭 벤안 망의 크기 변화에 대한 정규화된 평균 지연을 보인다^[2]. 이 그림은 셀이 입력 부하가 1.0인 최대 부하를 걸었을 때에 정규화된 평균 지연이 망의 크기와 상관없이 2.0에 수렴함을 보이고 있다.

결과적으로 다중 경로 스위칭 망으로 사이클릭 벤안 망을 사용하는 경우, 순서의 무결성을 맞추기 위한 재배열 윈도우의 크기, 즉 셀이 스위칭 망을 떠나기 전에 반드시 지연되어야 하는 크기가 100 셀 시간이라는 작은 크기로 유지할 수 있으므로써, 적은 하드웨어 비용

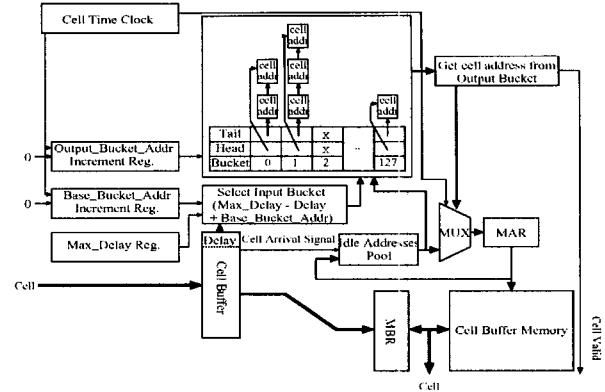


그림 6. 셀 재배열 버퍼의 상세 설계

Fig. 6. The detail design of the cell resequencing buffer.

으로 효과적인 재배열 버퍼를 구성할 수 있음을 알 수 있다.

3. 셀 재배열 버퍼의 상세 설계

본 절에서는 셀 재배열 버퍼의 상세 설계를 제시한다. 그림 6에 제시한 바와 같이, 셀 재배열 버퍼는 실제로 셀들이 저장되는 셀 버퍼 메모리와 셀 버퍼 메모리에 저장된 셀들의 주소들의 연결 리스트들을 유지하는 버킷들의 집합, 그리고 셀 버퍼 메모리의 셀 주소 중 유희한 주소들을 담은 가용 셀 주소 대기 풀, 그리고 셀 입력 기능과 셀 출력 기능으로 구성되어 있다.

셀 재배열 버퍼의 동작은, 다음과 같이, 크게 셀 입력 기능과 셀 출력 기능으로 나누어 설명할 수 있다. 셀 입력 기능은 입력 셀이 들어갈 입력 버킷을 선택하고, 가용한 셀 주소를 대기 풀로부터 할당 받아, 이 주소를 선택된 버킷의 연결 리스트에 추가하고, 셀을 셀 버퍼 메모리에 저장하는 역할을 한다. 셀 출력 기능은 출력 셀을 가져올 버킷을 선택하고, 셀의 버퍼 메모리 주소를 가져와서, 메모리로부터 셀이 배출되도록 하고, 사용한 셀 주소를 대기 풀에 넣는 역할을 한다. 여기서 셀 입력 기능은 출력 정합으로 배출되기 전에 셀이 기다려야 하는 셀 시간 스롯을 계산하여, 입력할 버킷을 선택하고, 셀들을 저장한다. 그러면 셀들은 여기에 머무르며 셀 지연 시간들이 균일화되어, 출력되는 셀들의 순서의 무결성이 보장되게 된다.

이제, 그림 7에 제시한, 셀 재배열 버퍼의 동작을 상세히 알아보자. 초기화 단계에서, 최대 지연 레지스터 (Max_Delay_Reg)는 시뮬레이션에 의해 얻어진 최대 지연시간에 설계 여유시간을 추가하여 만들어진 최대 지연 시간으로 초기화 된다. Base Bucket 주소 증가 레

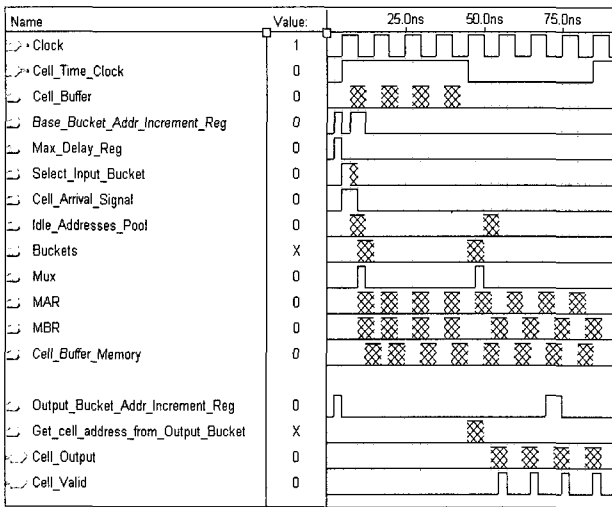


그림 7. 셀 재배열 버퍼의 상세 동작
Fig. 7. The detail behavior of the cell resequencing buffer.

지스터와 출력 버킷 주소 증가 레지스터도 역시 초기화 단계에서 0으로 각각 초기화 된다.

그림 7에 도시한 바와 같이, 셀 재배열 버퍼는 셀의 동기 정보인 셀 시간 클럭의 상승 모서리(Rising Edge)에 입력 셀을 셀 버퍼 메모리에 저장하고, 셀 시간 클럭의 하향 모서리(Falling Edge)에 출력 셀을 셀 버퍼 메모리로부터 가져와서 배출한다. 여기서, 셀은 버퍼 메모리에 16 비트들 병렬로 쓰고 읽혀진다.

먼저 셀들이 셀 재배열 버퍼에 입력되는 과정을 살펴보자. 셀이 스위칭 망의 마지막 단에서 나와서 셀 재배열 버퍼에 도착하면, Select_Input_Bucket은 입력 셀이 저장될 주소를 담은 버킷의 번호를 다음과 같이 계산한다. 즉 최대 지연 레지스터의 값에서, 도착한 셀의 실제 스위칭 지연 시간의 값을 빼고, 현재 슬라이딩 윈도우의 시작점을 가르키는 베이스 버킷 주소 증가 레지스터의 값을 더하여, 셀이 들어갈 버킷의 번호를 구한다. 결과적으로 계산한 값은 셀 재배열 버퍼에서 최대 지연 시간까지 얼마만큼의 시간이 남았는지 구한 것이며, 결국 셀을 버퍼에 잡고 있어야 하는 시간을 의미하게 된다.

셀들이 실제로 저장될 셀 버퍼 메모리의 유휴 주소는 Idle Addresses Pool에 유지되는데, 셀이 도착할 때, Cell Arrival Signal이 발생하여, 유휴 주소 중 하나를 할당한다. 이 유휴 주소는 멀티플렉서를 통해 메모리 주소 레지스터에 들어가고, 이 결과 셀 버퍼 메모리에 입력 셀을 저장하고, 해당 주소를 선택된 버킷의 연결 리스트에 유지한다. 여기서 메모리 주소 레지스터에 연결된 MUX는 셀 시간 클럭을 제어 정보로 해서, 입력

셀의 주소와 출력 셀의 주소 중 하나를 메모리 주소 레지스터로 보내는 역할을 한다.

앞에서 설명한 바와 같이, 입력 버킷을 주소를 구할 때, 각 셀들이 스위칭 망 통과할 때 발생한 각 셀의 실제 지연시간 정보가 필요하다. 실제 지연 시간 정보를 저장하기 위해, 우리는 셀의 헤더에 7 비트들 길이의 지연 정보 저장을 위한 필드를 추가한다. 이 지연 시간 정보 저장하기 위한 필드는 셀이 스위치에 도착할 때, 0으로 초기화되고, 셀들이 각 링크를 통해, 단위 스위치로부터 빠져 나올 때, 각 단위 스위치에서 1씩 증가시켜서, 셀 헤더에 갱신시킴으로써, 재배열 버퍼에 셀이 도달했을 때, 올바른 실제 지연 시간을 Select_Input_Bucket 회로에 제공하게 된다.

다음에서 셀들을 출력하는 기능에 대해 알아보자. 각 셀 시간에 셀은 다음과 같은 방법으로 출력된다. 셀 버퍼 메모리 내에서의 출력할 셀들의 주소를 연결 리스트로 유지하는 버킷의 번호는 출력 버킷 주소 증가 레지스터로부터 얻는다. 이 출력 버킷 주소 증가 레지스터는 버킷 슬라이딩 윈도우의 출력 버킷의 번호를 유지한다. 그리고 셀 시간 클럭의 하향 모서리에서, 해당 출력 버킷의 연결리스트로부터 가져온 주소를 메모리 주소 레지스터로 보내면, 셀 버퍼 메모리의 해당 주소에 저장된 셀이 메모리에서부터 출력 단으로 내보내진다. 결과적으로 본 셀 재배열 버퍼를 사용하여, 사이클릭 벤안 망을 통과하는 모든 셀들의 순서의 무결성을 보장할 수 있다.

IV. 결 론

본 논문에서는 완전 적응 자기 경로 제어 알고리즘을 사용하는 결합 감내 셀 스위치인 사이클릭 벤니안 망의 셀 순서의 일관성을 보장하기 위한 셀 재배열 버퍼를 제시하였다. 사이클릭 벤안 망은 인터넷 트래픽과 같은 폭발적인 트래픽을 스위칭 할 수 있고, 내부 링크들의 결합도 감내할 수 있지만, 거리가 다른 다중 스위칭 경로들을 제공하기 때문에, 동일한 입력 정합에서 출발하여 동일한 출력 정합에 도달한 셀들이, 입력 정합에 도착한 순서에서 벗어나서 출력 정합에 도달하게 만들 수 있다.

본 논문에서는 이를 바로 잡아 셀 순서의 무결성을 보장하기 위한, 일종의 하드웨어 슬라이딩 윈도우 메커니즘이라 할 수 있는, 셀 재배열 버퍼를 제시하였다. 제시된 셀 재배열 버퍼는 적은 제어 하드웨어와 범용 매

모리로 구성될 수 있었다. 이를 적은 하드웨어 비용으로 구성하기 위해서는, 제한된 작은 크기의 슬라이딩 윈도우를 사용하여야 한다.

본 연구에서는 필요한 윈도우 크기를 구하기 위해, 비균일 주소 분포를 가진 입력 부하에 대해, 셀들의 스위칭 지연을 시뮬레이션을 사용하여 분석하였다. 시뮬레이션의 결과에 의하면, 사이클릭 벤안 망의 셀 지연 변이는 100 셀 시간으로 제한되기 때문에, 슬라이딩 윈도우의 크기를 100 셀 시간에 공학적 여유를 약간 추가하여 선택하면 된다.

결과적으로, 본 셀 재배포 버퍼를 사이클릭 벤안 스위칭 망에 추가하면, 셀 스위칭 자원을 사용하지 않으면서, 셀 순서의 일관성을 보장하는, 대용량 고속 결합 감내 스위치를 구성할 수 있다. 본 논문에 제시한 셀 재배포 버퍼는 다른 다중 경로 스위칭 망들의 재배포 버퍼로도 사용될 수 있음을 밝혀둔다.

참 고 문 헌

- [1] H. Jonathan Chao, Kung-Li Deng, and Zhigang Jing, "A Petabit Photonic Packet Switch (P3S)," *Proc. INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*, IEEE, Vol. 1, pp. 775-785, 2003.
- [2] J.-H. Park, H. Yoon, and H. Lee, "The Deflection Self-routing Banyan Network: A Large-Scale ATM Switch using the Fully-Adaptive Self-routing and its Performance Analyses," *IEEE/ACM Trans. on Networking*, Vol.7, No.4, pp. 588-604, IEEE, U.S.A., August 1999.
- [3] R. Rojas-Cessa, E. Oki, and H.J. Chao, "Concurrent Fault Detection for a Multiple-plane Packet Switch," *IEEE/ACM Transactions on Networking*, Vol.11 No. 4, pp. 616-627, Aug. 2003
- [4] R. Rooholamini and M. Garver, "Finding the Right ATM Switch for the Market," *IEEE Comput.*, vol. 27, pp. 16-28, Apr. 1994.
- [5] J.-H. Park, H. Yoon, and H. Lee, "The Cyclic Banyan Network: A Fault Tolerant Multistage Interconnection Network with the Fully-Adaptive Self-routing," *Seventh IEEE Symposium on Parallel and Distributed Processing*, pp. 702-710, October 1995.
- [6] E. Oki, R. Rojas-Cessa, and H.J. Chao, "PCRRD: a Pipeline-based Concurrent Round-Robin Dispatching Scheme for Clos-Network Switches," *Proc. IEEE International Conference on Communications 2002*, IEEE, Vol. 4, pp. 2121-2125, May 2002.
- [7] N. Moriwaki, A. Makimoto, Y. Oguri, M. Wada, and T. Kozaki, "Large Scale ATM Switch Architecture for Tbit/s Systems," *Proc. Global Telecommunications Conference 1998*, IEEE, Vol. 1, pp. 334-338 Nov. 1998.
- [8] E. Oki and N. Yamanaka, "Tandem-Crosspoint ATM Switch with Input and Output Buffers," *IEEE Communications Letters*, IEEE, Vol. 2, No. 7, pp. 189-191, July 1998.
- [9] H.J. Chao, Byeong-Seog Choe, Jin-Soo Park, and N. Uzun, "Design and Implementation of Abacus Switch: a Scalable Multicast ATM Switch," *IEEE Journal on Selected Areas in Communications*, Vol. 15 No. 5, pp. 830-843, June 1997.
- [10] K.S. Chan, S. Chan, K.L. Yeung, "Design of Wide-Sense Nonblocking Multicast ATM Switches," *IEEE Communications Letters*, Vol. 2 No. 5, pp. 146-148, May 1998.
- [11] C. Wu and T. Feng, "On a Class of Multistage Interconnections Networks," *IEEE Trans. on Comp.*, vol.C-29, pp. 694-702, August, 1980.
- [12] S. Gianatti and A. Pattavina, "Performance Analysis of Shared-Buffered Banyan Networks under Arbitrary Traffic Patterns," *Proc. INFOCOM 93*, pp. 943-952, 1993.

저 자 소 개



박재현(정회원)

1988년 중앙대학교 전자계산학과 이학사

1991년 한국과학기술원 전산학과 공학석사

1995년 한국과학기술원 전산학과 공학박사

1995년 8월~2000년 2월 삼성전자 정보통신본부 데이터네트워크 개발팀 MPLS/ATM 개발

2000년 3월~2002년 8월 영남대학교 전자정보 공학부, 정보통신 전공 교수

2002년 9월~현재 중앙대학교 컴퓨터 공학부 교수
<주관심분야: ATM Switch Arch., 상호연결 망, Multiprotocol Label Switching System, Routing Protocols, Ad Hoc Networking, Peer-to-Peer Directory System>