

다차원 링-델타 망: 고성능 고장감내 스위칭 망

정회원 박재현*

Multidimensional Ring-Delta Network: A High-Performance Fault-Tolerant Switching Networks

Jae-Hyun Park* *Regular Member*

요약

본 논문에서는 편향 자기 경로 제어 방법을 사용하는 새로운 고성능 고장 감내 스위칭 망을 제시한다. 기반이 되는 델타 망의 위상적 구조를 대수적 기법을 사용하여 해석한 결과로 부터, 편향 자기 경로 제어 알고리즘을 사용하는 다중 경로 스위칭 망인 다차원 링-델타 망을 유도한다. 제시된 스위칭 망에서는 델타 망의 링크들을 포함하는 모든 링크들이 결함 혹은 혼잡이 생긴 링크들을 우회하는 대체 경로들로써 사용된다. 본 연구에서 비균일 주소 분포를 가진 부하를 가했을 경우의 이 망의 성능을 시뮬레이션을 사용하여 분석하였다. 입력 셀의 도착률이 0.9 이고, Hot Ratio가 0.8인 경우, 1024×1024 크기의 제안된 스위칭 망의 통과율은 사이클릭 벤안 망에 비해서 36.5% 우수하고, 2차원 링-벤안 망에 비해서 13.3% 우수하다. 또한, 64×64 크기의 제안된 스위칭 망의 신뢰도는 Hui의 망에 비해 2.2배, 2차원 링-벤안 망에 비해 46.6% 높다.

Key Words : High-Performance Fault-Tolerant Switching Network, Deflection Self-routing, Augmented Delta Network, Topological Properties, Algebraic Formalism

ABSTRACT

In this paper, a high-performance fault-tolerant switching network using a deflection self-routing was proposed. From an abstract algebraic analysis of the topological properties of the Delta network, which is a baseline switching network, we derive the Multidimensional Ring-Delta network: a multipath switching network using a deflection self-routing algorithm. All of the links including already existing links of the Delta network are used to provide the alternate paths detouring faulty/congested links. We ran a simulation analysis under the traffic loads having the non-uniform address distributions that are usual in Internet. The throughput of 1024×1024 switching network proposed is better than that of the 2D ring-Banyan network by 13.3 %, when the input traffic load is 1.0 and the hot ratio is 0.9. The reliability of 64×64 switching network proposed is better than that of the 2D ring-Banyan network by 46.6%.

1. 서론

다단 스위칭 망은 센서 망을 구성하기 위해 처리 기들과 분산된 센서들의 배열 사이의 상호 접속 망으로서 사용되었다^[1]. 또한 최근에 이 스위칭 망은 저렴한 소형 라우터들 활용하여 대규모 다단 라우

터를 구성하기 위해서 사용되었다^[2]. 더욱이, 최근 주목 받고 있는 클라우드 컴퓨팅을 위한 데이터 센터 스위칭을 위해서는, 확장성이 있는 이러한 다단 스위칭 망이 필요하다^[3]. 다단 스위칭 망의 이러한 응용들에 있어, 성능과 고장 감내성을 개선하기 위해서, 입력 포트와 출력 포트 사이에 다중 경로를

* 중앙대학교 공과대학 컴퓨터공학부 (hyunie@cau.ac.kr)

논문번호:KICS2009-06-266, 접수일자: 2009년 6월 29일, 최종논문접수일자: 2010년 1월 7일

제공하는 많은 다단 스위칭 망들이 제안되어 왔다. Benes 망, Clos 망, 병렬 벤안 망, Itoh 망, Hui 망, 2차원 링-벤안 망이 그 예 이다^{[4],[6]}.

하지만, Benes 망은 중간 단계 이후에는 단일 경로만을 제공한다. 그래서 나머지 절반의 단계들에서의 장애가 발생하는 경우에 고장을 야기하고, 패킷들의 충돌이 발생하는 경우 역시 해결할 수가 없다^[4]. Clos망도 마찬가지로 중간 단계 이후에 단일 경로만을 제공한다. 결국, Clos 망은 첫 번째 단계에서만 다중 경로를 제공한다. 병렬 스위칭 망은 추가적인 스위칭 망들을 병렬로 사용함으로써 다중 경로를 제공한다. 그러나 이 방법 역시 첫 번째 단계에서만 다중 경로를 제공한다. Itoh의 망과 Hui의 망 역시 다중 경로를 제공한다^[5]. 그러나, 이러한 망들을 같은 단위 스위칭 장치를 사용하여 구성하는 것은 불가능하고, 또한 망의 구조도 상당히 복잡하다^[4]. 이러한 단점을 극복하며, 다중 경로를 제공하는 스위칭 망인 2차원 링-벤안 망도 제안되었다^[6]. 이 망은 모든 단계들의 모든 링크들을 다중 경로의 제공을 위해 사용한다^[6].

Clos 망을 제외한, 2차원 링-벤안 망을 비롯한 이러한 다중 경로를 제공하는 다단 스위칭 망은 벤안 망에 기반을 두고 있다. 반면에 전통적으로 다중 경로 망으로 널리 사용되어 온 Clos 망은 델타 망^[7]에 기반을 두고 있다. 그런데, 집적회로의 핀의 대역폭이 점차로 커지고 있기 때문에, 단위 스위치로 2x2 스위치를 사용하는 벤안 망보다, 더 큰 크기의 단위 스위치를 (예를 들어 4x4 스위치 혹은 8x8 스위치를) 구성요소로 사용하는 델타 망이 대용량의 스위칭 망으로써 적합하다고 할 수 있다^[8].

본 논문에서는, 델타 망을 기반 망으로 사용하여, 모든 단계의 모든 링크들을 다중 경로의 제공을 위해 사용하는 고성능 고장 감내 망인 다차원 링-델타 망을 제시한다. 그룹 이론 관점에서 분석한 델타 망의 위상적 특성으로 부터, 모든 단계들의 모든 링크들을 다중 경로의 제공을 위해 사용하는 델타 망 기반의 다중 경로 스위칭 망과 편향 자기 경로 제어 알고리즘을 제시한다.

본 논문에서는 모의 실험을 사용하여, 제안된 스위칭 망이 비균일 그래픽 아래 다른 스위칭 망 보다 더 좋은 성능을 가지고 있는 것을 보인다. 또한 제안된 스위칭 망의 신뢰도가 다른 스위칭 망들 보다 우수함을 제시한다.

2절에서 델타 망의 위상적 특성을 그룹 이론으로 분석한 결과로부터, 다차원 링-델타 망을 제시한다.

그리고 편향 자기 경로제어 알고리즘의 정당성을 증명한다. 3절에서는 제시된 스위칭 망의 성능 분석과 신뢰성 분석 결과를 제시한다. 끝으로 4절에서 결론을 맺는다.

II. 다차원 링-델타 망

2.1 델타 망의 위상적 특성

본 논문에서는 Wu^[9]의 명명법 (Naming Scheme)을 일반화 하여, 제안한 망의 형상과 경로제어 방법을 기술한다. 제시한 정의들과 따름정리들 그리고 정리들은 관련연구^[10]에 기술되어 있으나, 본 논문의 독립적인 가독성을 위해서 다시 기술한다. Wu가 사용한 유한체 (Finite Field) F_2 위에 정의된 $\log_2 N - 1$ 차원의 벡터 공간의 원소인 이진수 대신에, 유한체 (Finite Field) F_{2^m} 위에 정의된 $\log_{2^m} N - 1$ 차원의 벡터 공간의 원소인 2^m -진수를 사용하여 단위 스위치와 링크들을 명명한다^[10]. 스위칭 단계는 첫 단계를 1로 하며, 연속으로 $\log_{2^m} N$ 까지 번호를 붙인다. 각 스위칭 단계에 있는, N 개의 링크들은 $\log_{2^m} N$ 자리수의 2^m -진수로 표현하여 구분한다. 즉 $link(p_{\log_{2^m} N}, p_{(\log_{2^m} N)-1}, \dots, p_1)$ 와 같이 기술한다. 각 단계에서, 단위 스위치 (Switching Element, SE)는 $\log_{2^m} N - 1$ 자리수의 2^m -진수들로, $(p_{(\log_{2^m} N)-1}, p_{(\log_{2^m} N)-2}, \dots, p_1)$ 같이 표현되는데, 이는 위로부터 시작하여 스위칭 단계내에서의 위치를 2^m -진수로 표현한 것이다. 이 상대적인 위치를 수준 (level)이라고도 부른다. 입력 셀의 목적 주소는 $A = (a_1, a_2, \dots, a_{\log_{2^m} N})$ 로 표현 된다. 그러면, 델타 망은 다음과 같이 정의된다.

정 의 1 델타 망의 위상은 다음과 같이 기술할 수 있다. 스위칭 단계 i 에 있는 각 단위 스위치 $(p_{(\log_{2^m} N)-1}, p_{(\log_{2^m} N)-2}, \dots, p_1)$ 의 각 출력 링크는 스위칭 단계 $i + 1$ 에 있는 입력 링크와 연결되어 있다.

$$\beta_s [(p_l, p_{l-1}, \dots, p_1)_i] = (p_{l-1}, p_{l-2}, \dots, p_1, s)_{i+1} \text{ link}(p_l, p_{l-1}, \dots, p_1, s)_i \text{로 연결} \quad (1)$$

여기서 $1 \leq i \leq (\log_{2^m} N) - 1 = l$, $s \in F_{2^m}$ 이다. ■

두 SE 들 사이의 링크는 연산 β_s 로 표현된다. 즉, 식 (1)의 연산 β_s 는 단계 i 의 한 SE로 부터 단계 $i + 1$ 의 SE들로의 사상(mapping)으로써 링크

를 표현한다. (여기서, $1 \leq s \leq 2^m - 1$ 이다.) 연산 β_0 은 SE의 최상위 출력 링크를 통해서 도달 가능한 SE를 기술하기 위해서 사용되고, 연산 β_{2^m-1} 은 최하위 출력 링크를 기술하기 위해 사용한다. 그리고, 단계 i 의 수준 j 에 있는 단위 스위치를 간략히 표현하기 위해 기호 $n_{i,j}$ 를 사용한다.

이제 동일한 스위칭 단계 내에서의 단위 스위치들 간의 상대적 위치를 명확히 표현하기 위해, 함수 γ_k 를 다음과 같이 정의한다.

정의 2 델타 망의 각 스위칭 단계 내의 한 SE로부터 다른 SE의 상대적 위치를, 함수 γ_k 를 사용하여 다음과 같이 표현한다.

$$\gamma_k[n_{i,j}] = n_{i,(j+k) \bmod (N/2^m)} \quad (2)$$

여기서 $1 \leq k \leq N/2^m$ 이다. ■

Park의 연구^[11]의 정리들, 보조정리들, 정리들, 따름정리들의 모든 ‘2’를 ‘ 2^m ’로 대체하여, 다음의 정리들, 보조정리들을 얻을 수 있다^[10]. 그리고, 이들 역시 참임을 Park^[11] 방법과 유사한 방법을 사용하여 간단하게 증명할 수 있다^[10]. 다음에 제시된 정리들과 따름정리들은 편향 자기 경로제어 알고리즘을 유도하는데 유용한 속성들을 보여주며, 다음 절에서 제시되는 편향 자기 경로제어 알고리즘의 정당성 증명을 위해 사용된다.

정리 1 단계 i 의 하나의 동치 클래스 안에 있는 인접한 두 SE들 간의 위상적 거리는 $2^{m \cdot (i-1)}$ 이다. 즉

$$n_{i,j} \sim n_{i,k} \Leftrightarrow |j-k| = 2^{m \cdot (i-1)} \times l, \quad (3)$$

여기서 $l \in \{0, 1, \dots, N/2^{m \cdot i}\}$. ■

다음 따름 정리들은 모든 출력 링크들이 상호 대체하여 사용할 수 있음을 보인다.

따름정리 1 하나의 출력 링크로 시작되는 스위칭 경로는, 동치인 단위 스위치의 출력 링크로 시작되는 다른 스위칭 경로로 대체될 수 있다.

$$\beta_k[n_{i,j}] \sim \beta_k[\gamma_{2^{m \cdot (i-1)}}[n_{i,j}]] \quad (4)$$

$$\sim \beta_k[\gamma_{-2^{m \cdot (i-1)}}[n_{i,j}]],$$

여기서 $k \in \{0, 1, \dots, 2^m - 1\}$. ■

따름정리 2 만약 하나의 스위칭 단계 내에서 SE들 간에 연결들을 만들면, 한 출력 링크로 시작되는 경로는 동일한 SE의 다른 출력 링크로 시작되는 경로로 대체될 수 있다. 즉

$$\gamma_{d-s}[\beta_s[n_{i,j}]] = \beta_d[n_{i,j}] \quad (5)$$

그리고

$$\gamma_{2^{m \cdot (i-1)} + (d-s)}[\beta_s[n_{i,j}]] \sim \beta_d[n_{i,j}]$$

여기서 $1 \leq i < \log_{2^m} N$ 이고, $1 \leq s, d < 2^m - 1$ 이다. ■

정리 2 만약 함수 γ_k 에 해당하는 편향 라우팅 혹은 보정 라우팅을 할 수 있는 상호 연결을 제공하면, 단위 스위치로부터 추가된 링크로 시작되는 경로는 원래 단위 스위치의 출력 링크로 시작되는 경로로 대체될 수 있다. 즉,

$$\beta_j[\gamma_k[n_{i,j}]] \sim \gamma_{k \times 2^m}[\beta_j[n_{i,j}]], \quad (6)$$

여기서 $1 \leq i < \log_{2^m} N$ 이고, $j \in \{0, 1, \dots, 2^m - 1\}$ 이고 $k \in Z$. ■

2.2 다차원 링-델타 스위칭 망

$N \times N$ 다차원 링-델타 망은 $N \times N$ 델타 망에 다음과 같은 링크들을 추가해서 구성 할 수 있다.

정의 3 다차원 링-델타 망을 구성하기 위해 벤안 망에 추가되는 링크들은 다음과 같이 기술 할 수 있다. 아래는 $m = 2$ 인 경우의 다차원 링-델타 망을 기술 한 것이다.

$$\gamma_1[n_{i,j}] = n_{i,(j+1) \bmod (N/2^m)}, \quad (8)$$

$$\gamma_{-1}[n_{i,j}] = n_{i,(j-1+2^{2 \cdot (i-1)} \cdot (\log_{2^m} N - i + 1)) \bmod (N/2^m)}$$

$$\gamma_2[n_{i,j}] = n_{i,(j+2) \bmod (N/2^m)},$$

$$\gamma_{-2}[n_{i,j}] = n_{i,(j-2+2^{2 \cdot (i-1)} \cdot (\log_{2^m} N - i + 1)) \bmod (N/2^m)}$$

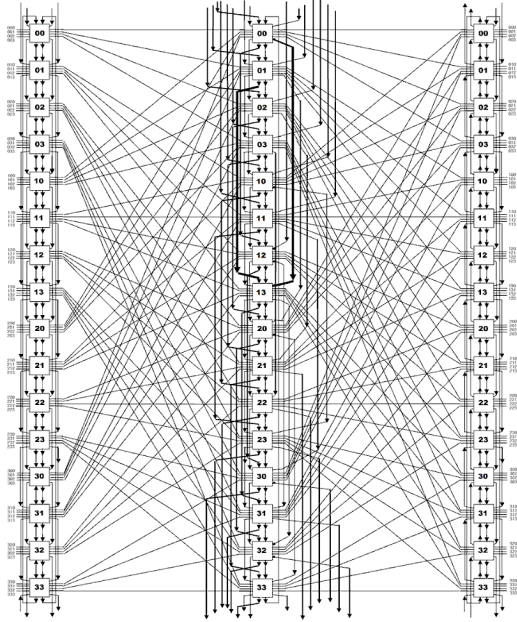


그림 1. 64 × 64 다차원 링-델타 망

여기서 $1 \leq i \leq \log_2 N$ 그리고 $0 \leq j < N/2^m$ 이다. 그리고 N 은 입력력 포트들의 수이다. ■

연산 γ_k 는 추가한 링크들을 하나 SE의 동일한 단계내의 다른 SE로 사상으로 표현한다. 일반적인 64 × 64 델타 망으로부터 다차원 링-델타 망 ($m=2$)을 구성한 예를 그림 1에 보였다.

2.3 다차원 링-델타 스위칭 망의 적응 자기 경로 제어

많은 다른 다단 스위칭 망들과 같이, 다차원 링-델타 망의 경로제어는 목적지 꼬리표(Tag)를 가지고 제어된다. 즉 목적지 꼬리표는 목적지 주소 $A=(a_1, a_2, \dots, a_{\log_2 N})$ 와 여기에 덧붙여진 고정된 크기의 이탈 꼬리표(Deviation Tag) k 가 사용된다. 즉, 행선지 주소는 (A, k) 쌍으로 표현된다.

이탈 꼬리표는 각 스위칭 단계에서 원래 도달하여야 하는 SE나 이것의 동치 SE로부터의 위상적 거리 차이를 나타내기 위해서 사용된다. k 값은 그것의 현재 값과 선택된 링크 번호 그리고 원래 목적된 링크 번호로부터, 다음에 나온 경로제어 알고리즘의 k 값을 수정하는 규칙을 사용하여 SE 별 분산적으로 계산할 수 있다.

단계 1의 입력 링크에 한 셀이 도착 했을 때, k 값은 0으로 설정된다. 즉 행선지 주소는 $(A, 0)$ 이

다. 제안된 편향 경로제어 알고리즘은 다음과 같다. 여기서 i 는 해당 SE의 단계 번호이다. 그리고 α_i 는 다음과 같은 델타 망의 자기 경로제어 함수이다.

$$\alpha_i[(p_1, p_{l-1}, \dots, p_l), (a_1, a_2, \dots, a_{l+1})] = (p_{l-1}, p_{l-2}, \dots, p_l, a_l)_{i+1}$$

여기서 $a_i \in F_{2^m}$ 이다.

정 의 4 링-델타 망의 SE $n_{i,j}$ 에서 사용되는 적응 자기 경로제어 알고리즘은 다음과 같은 링크 할당 절차 절차와 이탈 꼬리표 갱신 절차로 정의된다. (여기서 $s \in \{0, 1, \dots, 2^m - 1\}$ 이다.)

I. 링크 할당 절차

각 입력 셀에 대하여,

- 1) $k = 0$ 인 경우, 다음과 같은 순서로 셀을 링크에 할당하는 시도를 한다.

$$(\beta_{\varepsilon(0)}[n_{i,j}], \beta_{\varepsilon(1)}[n_{i,j}], \beta_{\varepsilon(2)}[n_{i,j}], \beta_{\varepsilon(3)}[n_{i,j}], \gamma_{\delta(j)}[n_{i,j}], \gamma_{\delta(j+1)}[n_{i,j}], \gamma_{\kappa(j)}[n_{i,j}], \gamma_{\kappa(j \times 2)}[n_{i,j}])$$

- 2) $K=1$ 또는 $(2^{m(i-2)} - K)=1$ 인 경우, 다음과 같은 순서로 셀을 링크에 할당하는 시도를 한다.

$$(\gamma_{\chi(1)}[n_{i,j}], \gamma_{\chi(2)}[n_{i,j}], \gamma_{\chi(-1)}[n_{i,j}], \gamma_{\chi(-2)}[n_{i,j}], \beta_{\varepsilon(0)}[n_{i,j}], \beta_{\varepsilon(1)}[n_{i,j}], \beta_{\varepsilon(2)}[n_{i,j}], \beta_{\varepsilon(3)}[n_{i,j}])$$

- 3) $K \geq 2$ 또는 $(2^{m(i-2)} - K) \geq 2$ 인 경우, 다음과 같은 순서로 셀을 링크에 할당하는 시도를 한다.

$$(\gamma_{\chi(2)}[n_{i,j}], \gamma_{\chi(1)}[n_{i,j}], \gamma_{\chi(-1)}[n_{i,j}], \gamma_{\chi(-2)}[n_{i,j}], \beta_{\varepsilon(0)}[n_{i,j}], \beta_{\varepsilon(1)}[n_{i,j}], \beta_{\varepsilon(2)}[n_{i,j}], \beta_{\varepsilon(3)}[n_{i,j}])$$

II. k 를 수정하는 규칙은 다음과 같다:

- 1) 만약 $\beta_d[n_{i,j}] = \alpha_i[n_{i,j}, A]$ 이고 $\beta_s[n_{i,j}]$ 가 선택된 경우, 만약 $K=0$ 이면,

$$K \leftarrow (2^{m_i} + (d - s)) \bmod 2^{m_i} \text{ 아니면}$$

$$K \leftarrow K \times 2^m + (d - s).$$

- 2) 만약 $\gamma_s[n_{i,j}]$ 가 선택되는 경우,

$$K \leftarrow (2^{m(i-1)} - s) \bmod 2^{m(i-1)}.$$

여기서 $\varepsilon(t) = s \quad \text{s.t.} \quad |s - a_i| \leq t \wedge s \in F_{2^m}$

$$\delta(t) = \begin{cases} 1 & \text{만약 } t \bmod 2 = 0 \text{ 이면} \\ -1 & \text{그밖의 경우} \end{cases},$$

$$\kappa(t) = \begin{cases} 2 & \text{만약 } (t/2) \bmod 2 = 0 \text{ 이면} \\ -2 & \text{그밖의 경우} \end{cases}, \quad \text{그리고}$$

$$\chi(t) = \begin{cases} t & \text{만약 } t < 2^{m-(i-2)} \text{ 이면} \\ -t & \text{그밖의 경우} \end{cases}.$$

다차원 링-델타 망의 편향 자기 경로제어 방법의 정당성 (Correctness)은 Park^[11]의 정리 6의 방법과 거의 같은 방법으로 다음과 같이 간단하게 증명할 수 있다.

정리 3 다차원 링-델타 망의 경로제어 알고리즘은 임의의 입력 패킷을 그의 행선지에 올바르게 보낸다.

증명: 먼저, 본 경로제어 알고리즘은 SE의 가능한 한 링크를 할당한다. 이 링크의 상호 대치적 할당의

정당성은 따름정리 1과 2에 의해서 증명된다.

두번째로, k 의 수정 규칙은 다음과 같이 증명된다: 규칙 1)은 정리 1과 정리 2 그리고 따름정리 2에 의해서 증명된다.

규칙 2)는 정의 3과 정리 1에 의해서 증명된다. 따라서, 본 수정 규칙들에 의해서 k 를 유지하는, 제시된 경로제어 알고리즘은 패킷을 마지막 단계의 의도된 출력단에 올바르게 내보낸다. 따라서, 이 경로제어 알고리즘은 정당하다. ■

III. 성능 및 신뢰성 분석

3.1 성능 분석

본 절에서는 $m = 2$ 인 다차원 링-델타 망의 성능을 비균일 (Nonuniform) 트래픽 하에서 분석한다. 성능분석을 위해서, 우리는 다음과 같은 일반적인 가정 하에 시뮬레이션을 행하였다.

- N 개의 근원들은 셀들을 비율 λ 의 포아송 확률 과정에 따라서 생성한다. (여기서 $0 \leq \lambda \leq 1$)
- 각 셀의 목적지는 0 부터 $N - 1$ 사이의 수를 비균일 분포로 생성하여 설정한다.
- 하나의 링크를 위해 셀들간에 충돌이 있으면, 하나의 셀이 임의로 선택되어 링크로 보내지고, 다른 셀들은 다른 링크들로 우회되어 진다.
- 통과율과 지연시간은 각 출력 포트에서 측정되

어지고, 네트워크의 크기로 평균되어진다.

비균일 트래픽은 $N \times N$ 부하 행렬 $D = \{\lambda_{ij}\}$ 로 나타낼 수 있는데, 여기서 λ_{ij} 는 출력 포트 j 가 목적지인 하나의 셀이 입력 포트 i 에 도달할 확률을 의미한다. 따라서, 행렬의 제 i 행의 합은 입력포트 i 에 부가되는 전체 부하를 나타내고, 제 j 열의 합은 출력 포트 j 에 부가되는 총 입력 부하를 나타낸다. 비균일 트래픽 패턴들은 그 가짓수가 무한대 이다.

이러한 패턴들 중 실제적인 트래픽 패턴들을 대표하는 Hot-group Model^[12] 이라고 하는 특정 비균일 트래픽 패턴을 사용한다. Hot-group Model의 입력 셀들은 모든 출력 포트들을 목적지로 하여, 다음과 같이 비균일적으로 분포 한다. 출력 포트는 “격렬히 출력하는 출력 집단”과 “대체적으로 쉬는 출력 집단”의 두개의 그룹들로 나뉜다. 부하 행렬 D 는 $D = [D_h(\lambda_h) D_c(\lambda_c)]$ 로 나뉜다. D_h 의 행의 합은 λ_h 이고 D_c 의 행의 합은 λ_c 이다. 여기서 $\lambda_h + \lambda_c = \lambda$ 이다. 다시 말해, 동일한 크기의 두 가지의 출력 포트 집단들이 있는데, 셀들이 목적지 집단 D_h 로 가는 확률은 λ_h 이고, 목적지 집단 D_c 로 갈 확률은 λ_c 이다. Hot Ratio를 r 이라고 정의하면, $\lambda_h = r\lambda$ 이고, $\lambda_c = (1-r)\lambda$ 이다. 여기서 r 은 $0.5 \leq r \leq 1.0$ 이다.

시뮬레이션 결과로 부터, 비균일 트래픽 하에서의 다차원 링-델타 망의 2-차원 링 벤안 망^[6]과 사

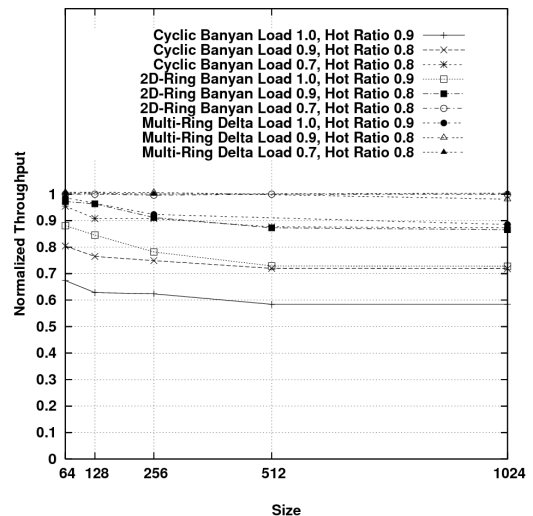


그림 2. 비균일 트래픽하에서의 다차원 링-델타 망과 2차원 링-벤안 망과 시클릭 벤안 망의 망 크기의 변화에 대한 정규화된 통과율

이클릭 벤안 망^[11] 대비 성능을 살펴보자. 그림 2에 보이는 바와 같이, 다차원 링-델타 망에 입력 셀의 도착률이 $\lambda = 0.9$ 이고, Hot Ratio r 을 0.8인 경우, 1024×1024 크기의 제안된 스위칭 망의 통과율은 사이클릭 벤안 망에 비해서 36.5% 우수하고, 2차원 링-벤안 망에 비해서 13.3% 우수하다. 256×256 망의 통과율은 사이클릭 벤안 망에 비해서 33.5% 우수하고, 2차원 링-벤안 망에 비해서 9.7% 우수하다. 이와 같이 제안된 스위칭 망은 셀의 도착률과 Hot Ratio가 높고 망이 큰 경우에 비교 대상이 되는 망들보다 더욱 우수하다. 이는 비균일 트래픽으로 모델링되는 인터넷 트래픽을 처리하는 클라우드 컴퓨팅을 위한 대규모의 스위칭 망을 구성하는데 적합하다고 말할 수 있다.

그림 3은 비균일 트래픽 하에서의 다차원 링-델타 망과 2-차원 링-벤안 망의 셀 지연 시간의 분포를 보여주고 있다. 다양한 크기의 스위칭 망에 대해, 부하 0.9와 Hot Ratio 0.8의 환경 하에서, 셀들이 제안된 스위칭 망을 통과하는 지연시간이 70 셀 이하임을 그림 3에서 알 수 있다. 그리고, 1024×1024 크기의 스위치의 지연시간의 최대치가 비교되는 스위칭 망에 비해 40% 정도 적음을 알 수 있었다. 결과적으로 실험된 스위칭 망의 순서의 무결성을 맞추기 위한 재 배열 윈도우의 크기, 즉 셀이 스위칭 망을 떠나기 전에 반드시 지연되어야 하는 크기가 주어진 입력 부하에 대해 70 셀 시간이라는 작은 크기로 유지할 수 있다. 따라서, 편향 자기 경

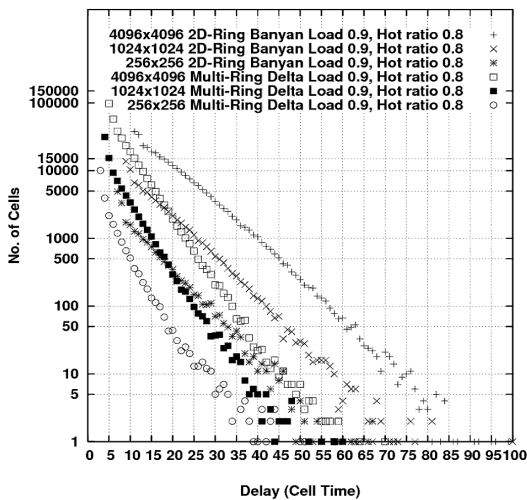


그림 3. 부하 0.9 비균일율 0.8인 비균일 트래픽 하에서의 다차원 링-델타 망과 2-차원 링-벤안 망의 망 크기의 변화에 대한 지연 분포

로제어 벤안 망^[11]과 유사한 방법으로 적은 하드웨어 비용으로 효과적인 재배열 버퍼를 구성할 수 있다.

3.2 신뢰성 분석

링-델타 망은 일반화된 사이클릭 벤안 망이다. 제안된 망은 사이클릭 벤안 망과 같이, 모든 링크들을 대체 링크 혹은 경로로 사용할 수 있다. 따라서, Park^{[10], [13]}의 방법과 거의 유사한 방법으로, $m=2$ 인 다차원 링-델타 망의 신뢰성을 분석할 수 있다. 표 1에 보여지는 바와 같이, 64×64 크기의 제안된 스위칭 망의 신뢰도가 벤안 망에 비해 85.5배, Tzeng의 망^[5]에 비해 3.14배가 높고, Hui의 망^[5]에 비해 2.2배가 높고, 2차원 링-벤안 망^[6]에 비해 46.6% 높음을 알 수 있었다.

표 1. 다차원 링-델타 망과 2-차원 링-벤안 망과 Tzeng 망과 Hui 망의 크기의 변화에 대한 고장을 초래하는 결합 요소들의 기대치들

N	Banyan	Tzeng	Hui	2D-Ring Banyan	MD Ring-Delta	MD - RingDelta	MD - RingDelta	MD - RingDelta
						2D - Ring	HuiNetwork	TzengNetwork
16	1.0	11.5	17.1	21.8	30.6	1.403	1.789	2.660
64	1.0	27.2	38.6	58.3	85.5	1.466	2.215	3.143

IV. 결 론

본 논문에서는 $2^m \times 2^m$ (여기서 $m \geq 2$) 스위치를 단위 스위치로 구성되는 델타 망에 링크들을 추가하여 모든 단계의 모든 링크들을 활용하여 다중 경로들을 제공하는 새로운 대용량 고성능 고장감내 망을 제시하였다. 그리고, 규칙적으로 연결된 델타 망의 기존의 링크들과 추가된 모든 링크들을, 고장난 링크들 혹은 혼잡이 발생한 링크들을 우회하는 대체 경로들로 활용하는 편향 자기 경로제어 알고리즘도 제시하였다. 대용량 스위칭 망에 기반이 되는 델타 망의 위상적 구조를 추상대수적 기법을 사용하여 분석한 결과를 활용하여, 제시한 편향 자기 경로제어 알고리즘의 정당성을 증명하였다.

모의 실험을 사용하여, 제안된 스위칭 망이 비균일 그래픽 아래 다른 스위칭 망 보다 더 좋은 성능을 가짐을 보였다. 입력 셀의 도착률이 0.9 이고 Hot Ratio가 0.8인 경우, 1024×1024 크기의 제안된 스위칭 망의 통과율은 사이클릭 벤안 망에 비해서 36.5% 우수하고, 2차원 링-벤안 망에 비해서 13.3% 우수하다.

또한 제안된 스위칭 망의 신뢰도가 다른 스위칭

망들 보다 우수하다는 결과를 제시하였다. 64×64 크기의 제안된 스위칭 망의 신뢰도는 벤안 망에 비해 85.5배 높고, Hui의 망에 비해 2.2배가 높으며, 2차원 링-벤안 망에 비해 46.6% 높다.

참 고 문 헌

[1] M. Choi, N.Park, and F. Lombardi, "Modeling and Analysis of Fault Tolerant Multistage Interconnection Network," *IEEE Trans. on Instrumentation and Measurement*, Vol.52, No.5, pp.1509-1519, 2003.

[2] A. Bianco, J.M. Finochietto, M. Mellia, F. Neri, and G. Galante, "Multistage Switching Architectures for Software Routers," *IEEE Network*, Vol.21, No.4, pp.15-21, July-August 2007.

[3] J. Turner, "A Dynamic Network Layer for Advanced Cloud Computing," Google Tech Talks, September 2, 2008. (Available at <http://www.youtube.com/>)

[4] M. Anan and M. Guizani, "A Fault Tolerant ATM Switching Architecture," *Proc. of the IEEE IPCCC '00*, pp.295-301, 2000.

[5] S.K. Hui, K. Seman, and J. Yunus, "An Augmented Chained Fault-Tolerant ATM Switch," *Proc. of 5th IEEE International Conference on High Speed Networks and Multimedia Communications*, pp.397-400, 2002.

[6] J.-H. Park, "Two-dimensional ring- Banyan network: a high-performance fault- tolerant switching network," *IEE Electronics Letters*, Vol.42, No.4, pp.249-251, 2006.

[7] H. Yoon, K.Y. Lee, and M.T.Liu, "Performance Analysis of Multibuffered Packet-switching Networks in Multiprocessor Systems," *IEEE Transactions on Computers*, Vol.39, No.3, pp.319-327, Mar 1990.

[8] J. Kim and W. J. Dally and D. Abts, "Flattened butterfly: a cost-efficient topology for high-radix networks," *Proc. of the 34th Annual international Symposium on Computer Architecture*, pp.126-137, 2007.

[9] C. Wu and T. Feng, "On a Class of Multistage

Interconnections Networks," *IEEE Trans. on Comp.*, Vol.C-29, pp.694-702, August, 1980.

[10] J.-H. Park, "The Deflection Self-routing Delta Network: a Dynamically Fault-tolerant High-radix Multistage Interconnection Network," *Journal of Supercomputing*, Springer, October 2009. *Published Online; DOI 10.1007/s11227-009-0337-y*

[11] 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.

[12] A. Pattavina and S. Gianatti, "Performance Analysis of ATM Banyan Networks with Shared Queueing part II: Correlated/unbalanced Offered Traffic," *IEEE/ACM Trans. on Networking*, Vol.2, No.4, pp.411-424, 1994.

[13] J.-H. Park and H. Lee, "Algebraic Analysis of the Topological Properties of a Banyan Network and its Application in Fault-tolerant Switching Networks," *Journal of Parallel and Distributed Computing*, Vol.66, No.6, pp.773-779, Academic Press, U.S.A., June 2006.

박 재 현 (Jae-Hyun Park)

정회원



1988년 2월 중앙대학교 전자계산학과

1991년 2월 한국과학기술원 전산학과 석사

1995년 8월 한국과학기술원 전산학과 박사

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

2000년3월~2002년8월 영남대학교 전자정보공학부 정보통신전공 교수

2002년9월~현재 중앙대학교 컴퓨터공학부 교수

<관심 분야> ATM Switch Arch., Multiprotocol Label Switching System, Routing Protocols, Ad Hoc Networking, Peer-to-Peer Networking, Grid Computing.