

다수의 결함이 있는 다단계 상호연결망에서 멀티캐스팅 알고리즘과 그 성능 평가

김 진 수*, 박 재 형**, 장 정 환***

*건국대학교 컴퓨터과학전공

**한국전자통신연구원 인터넷기술연구부

***부산외국어대학교 컴퓨터전자공학부

e-mail:jinsoo@kku.ac.kr

A Multicasting Algorithm in MINs with Multiple Faults and its Performance Evaluation

Jinsoo Kim*, Jaehyung Park**, Jung-Hwan Chang***

*Dept. of Computer Science, Konkuk University

**Network Technology Lab., ETRI

***School of Computer & Electronics Engineering, Pusan University of Foreign Studies

요약

본 논문은 다수의 결합 스위칭 소자를 갖는 다단계 상호연결망(MIN)에서 멀티캐스팅 알고리즘을 제시하고 그 성능에 대해 평가한다. 이 알고리즘은 MIN의 전체 스위칭 소자들을 같은 크기를 갖는 두 개의 집합으로 구분하고, 모든 결함이 동일한 집합에 속한 경우의 결함을 허용한다. 또한, 임의의 멀티캐스트 목적지들을 표현하기 위해 영역 부호화 방식을 사용하며, MIN을 통해 패킷을 두 번 순환시킴으로써 결합을 우회하면서 그 패킷의 목적지들로 멀티캐스팅을 한다. 그리고, 멀티캐스트 패킷이 사용하는 링크 수의 관점에서 이 알고리즘의 성능을 분석하고, 실험을 통해 그 성능을 기존의 알고리즘과 비교하여 평가한다.

1. 서론

다단계 상호연결망 (Multistage Interconnection Network, MIN)은 다중컴퓨터 시스템에서 프로세싱 노드들간의 연결을 위해 사용될 뿐만 아니라, 광대역 종합정보 통신망 (B-ISDN)을 위한 ATM (Asynchronous Transfer Mode) 스위칭 시스템의 내부 연결구조로도 보편적으로 사용되고 있다. 이것은 MIN에서 입력단에서 출력단으로 오직 하나의 라우팅 경로만 존재하며, 이 특성으로 고속의 셀프라우팅 기능을 제공하기 때문이다. 그러나, 이 특성 때문에 MIN은 다른 네트워크에 비해 결합 처리 능력이 떨어진다. MIN에서 결합 허용성(fault-tolerance)은 크게 두 가지 방법으로 구현된다. 첫 번째 방법은 원래 MIN에 하드웨어를 추가하여 다중의 경로를 제공함으로써 결함을 극복하는 것이다. 두 번째 방법은 부수적인 하드웨어 없이 결합 허용 라우팅 알고리즘을 제공하는 것이다. 그러나 MIN에서 결합 허용 라우팅에 대한 연구들은 주로 일대일 통신을 대상으로 진행되었다.

B-ISDN에서 케이블 TV, 다자간 영상회의, VOD (video-on-demand) 등과 같은 멀티미디어 서비스의 수요가 증가함에 따라, 일대다(one-to-many) 통신 형태인 멀티캐스트(multicast) 통신 서비스의 제공에 대한 중요성이 커지고 있다. MIN 기반의 ATM 스위치에서 멀티캐스트 기능을 제공하기 위해, 멀티캐스트 패킷을 복사하는 네트워크를 일대일 라우팅 네트워크 전단에 추가하는 방법과 라우팅 네트워크의 외부링크를 통해 출력단에서 입력단 쪽으로 패킷을 복사하면서 순환하도록 하는 방법 등이 사용된다 [1, 2, 3, 4]. 후자의 방법에서는 스위칭 소자(switching elements, SE)가 패킷의 유니캐스팅과 브로드캐스팅 기능을 모두 갖추어야 한다. 이를 대부분의 연구에서 결합 허용성은 고려되지 않았고, 순환 구조 MIN에서 한 개 SE의 결합을 허용하는 멀티캐스팅 알고리즘이 [3]에서 기 제안되었다. 본 논문에서는 [3]의 알고리즘을 확장하여 다수의 결합 SE를 허용하는 멀티캐스팅 알고리즘을 제안하고, 그 성능을 기존의 알고리즘과 비교하여 평가한다.

2. 시스템 구조와 기본 용어

2.1 시스템의 기본 구조

네트워크 크기가 N 인 MIN은 $n = \log_2 N$ 개의 단(stage)으로 구성되며 좌측부터 우측으로 $n-1$ 에서 0 으로 표현된다. 각 단은 위부터 아래로 0에서 $N/2-1$ 까지 $n-1$ 비트로 표현되는 $N/2$ 개의 2×2 SE들로 이루어지며, 각 단에서 N 개의 입출력 링크는 위부터 아래로 0에서 $N-1$ 까지 n 비트로 표현된다.

멀티캐스트 ATM 스위치는 패킷의 복사 능력을 지닌 SE들을 갖는 MIN으로 구성된다. 그림 1은 순환 구조 MIN을 기반으로 한 멀티캐스트 스위치를 예시하고 있다. 그림 1의 점선과 같이 오른쪽의 출력 제어기에서 왼쪽의 입력 제어기 쪽으로 역방향 링크들이 존재하며, 패킷의 순환을 위해 사용된다.

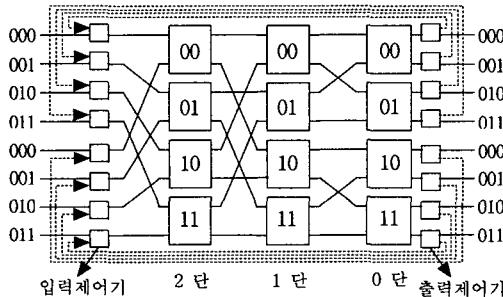


그림 1 : MIN 기반의 멀티캐스트 ATM 스위치.

2.2. 결합 모델과 용어

본 논문에서는 SE의 결합을 다루고 있으며, $(n-1)$ 단과 0 단에 있는 SE는 결합이 없다고 가정한다. 이것은 MIN에서 패킷의 출발지가 $(n-1)$ 단에서 결합이 있는 SE와 연결된 경우 또는 패킷의 목적지가 0 단에서 결합이 있는 SE와 연결된 경우에는 근본적으로 라우팅이 불가능하기 때문이다. 그리고, i ($n-2 \geq i \geq 1$) 단에 있는 하나의 SE를 $\alpha_{n-1} \cdots \alpha_{i+1} \beta_{i-1} \cdots \beta_0$ 으로 표시할 때, MIN상의 모든 SE를 다음과 같은 두 개의 집합 F_0 와 F_1 로 구분한다.

- $F_0 = \{ \alpha_{n-1} \cdots \alpha_{i+1} \beta_{i-1} \cdots \beta_0 \mid \alpha_{n-i} \oplus \beta_0 = 0 \}$
- $F_1 = \{ \alpha_{n-1} \cdots \alpha_{i+1} \beta_{i-1} \cdots \beta_0 \mid \alpha_{n-i} \oplus \beta_0 = 1 \}$

또한, 만일 모든 결합 SE들이 집합 F_0 에 속한 경우를 F_0 결합, 모든 결합 SE들이 집합 F_1 에 속한 경우를 F_1 결합이라고 한다. 본 논문에서는 F_0 결합 또는 F_1 결합인 경우를 처리한다.

일대다 통신 형태를 갖는 멀티캐스팅에서 목적지의 개수는 일정하지 않다. 이러한 멀티캐스트 메시지의 목적지를 표현하기 위해, MIN 기반의 ATM 스위치에서 널리 사용되는 방식 가운데 하나가 영역 부호화(region encoding) 방식이다. 이 방식은 연속된 주소를 갖는 다수 목적지들을 하나의 영역(region), 즉 최소와 최대 주소의 쌍인 [최소주소, 최대주소]로 표현하는 방식이다. 예를 들어 멀티캐스트 목적지가 001, 010, 011, 101 일 때, 목적지들을 두 개의 영역인 [001, 011], [101, 101]로 나타낼 수

있다. 즉 하나의 영역은 1개 이상의 연속된 목적지 주소로 구성되는 집합으로 볼 수 있다.

[정의-1] $D = \{R_1, R_2, \dots, R_m\}$ 이 멀티캐스트 목적지를 나타낼 경우, 이러한 집합 D 를 목적지 영역 집합이라고 정의하며. 임의의 두 영역 R_j 과 R_k ($1 \leq j < k \leq m$)에 속한 모든 d^j 와 d^k 에 대해 $d^j < d^k$ 을 만족하며 $R_j <_D R_k$ 로 표시한다. 그리고, R_j 에 속한 최대 주소를 d_j^{\max} , R_k 에 속한 최소 주소를 d_k^{\min} 으로 나타낼 때, R_2 와 R_1 의 차 $R_2 - R_1$ 은 $d_2^{\min} - d_1^{\max}$ 의 값으로 정의한다.

[정의-2] $M(0)$ 과 $M(1)$ 은 최상위비트가 각각 0과 1인 주소들의 집합으로, $L(0)$ 과 $L(1)$ 은 최하위비트가 각각 0과 1인 주소들의 집합으로 정의한다. 그리고, 목적지 영역 집합 D 에 대해, $R \in D$, $R \subseteq M(0)$, $R \not\subseteq M(1)$ 인 R 이 존재하면 R 을 두 개의 영역 R' $\sqsubseteq M(0)$ 과 $R'' \sqsubseteq M(1)$ 로 분리한 후, D 에서 R 을 빼고 R' 와 R'' 를 추가하여 D 를 수정한 후, D 를 다음과 같은 두 개의 분할 집합 $D0 = \{R | R \subseteq M(0)\}$ 과 $D1 = \{R | R \subseteq M(1)\}$ 로 구분하여 표기한다.

예를 들어, 멀티캐스트 목적지가 000, 001, 011, 100, 101일 때, $D0$ 은 {[000,010], [011,011]}이고, $D1$ 은 {[100,101]}이다.

[정의-3] $S = \{s^1, \dots, s^l\}$ 가 $s^1 < s^2 < \dots < s^l$ 인 출발지 주소들 집합이고, $D = \{R_1, R_2, \dots, R_k\}$ ($k \leq l$) 가 $R_1 <_D R_2 <_D \dots <_D R_k$ 인 목적지 영역 집합이라고 할 때, $S \Rightarrow^k D$ 는 출발지 s^i ($1 \leq i \leq k$)에서 목적지 영역 R_i 으로 k 개의 메시지들이 동시에 라우팅되는 것을 표현한다.

예를 들어, S 가 {1001, 1011, 1101}이고 D 가 {[1010,1011], [1110,1110]}일 때, $S \Rightarrow^2 D$ 는 출발지가 1001이고 목적지 영역이 [1010,1011]인 메시지와 출발지가 1011이고 목적지 영역이 [1110,1110]인 2 개의 메시지가 동시에 라우팅됨을 나타낸다.

3. 결합 허용 멀티캐스팅 알고리즘

본 알고리즘 **FTM-M**은 F_0 결합 또는 F_1 결합인 MIN에서 멀티캐스팅을 수행하며, 임의의 멀티캐스트 목적지를 목적지 영역 집합으로 표현한다.

3.1. 알고리즘 FTM-M

(1) 단계 1 : 출발지 $s = s_{n-1} \cdots s_1 s_0$ 에서 임의의 임시 출력 영역 IR 로 ($|IR| = 2 \times \max(|D0|, |D1|)$) 패킷을 복사한다. 이 때, F_0 결합이면 $IR \subseteq M(s_0)$ 을, F_1 결합이면 $IR \subseteq M(s_0)$ 을 만족하는 IR 을 선택한다.

(2) 단계 2 : IR 을 두 개의 집합 $S_0 \subseteq L(0)$ 과 $S_1 \subseteq L(1)$ 로 구분하여, S_0 과 S_1 에 속한 각 노드는 중간 출발지로서, 단계 1에서 복사된 메시지를 다음과 같이 실제 목적지 영역으로 라우팅한다. 이 때, 나머지 ($|IR| - |D0| - |D1|$) 개의 노드는 자신의 패킷을 버린다.

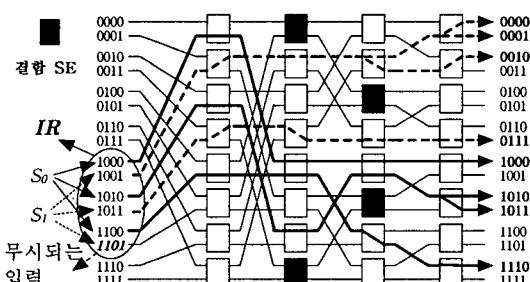
- F_0 결합이면, $S_0 \Rightarrow^{D1} D1$ 하고 $S_1 \Rightarrow^{D0} D0$ 함.
- F_1 결합이면, $S_0 \Rightarrow^{D0} D0$ 하고 $S_1 \Rightarrow^{D1} D1$ 함.

F_0 결합의 경우, $FTM-M$ 의 두 단계에서 모두 $L(0)$ 에 속한 출발지는 $M(1)$ 에 포함된 목적지로 라우팅하고, $L(1)$ 에 속한 출발지는 $M(0)$ 에 포함된 목적지로 라우팅한다. 따라서, 모든 패킷이 결합 SE를 통과하지 않는다. 그리고, 단계 2에서, $S_0 \subseteq L(0)$, $S_1 \subseteq L(1)$ 이므로, MIN의 특성상 $S_0 \Rightarrow^{DII} D1$ 의 패킷들과 $S_1 \Rightarrow^{DII} D0$ 의 패킷들 사이에는 충돌이 없다. 그리고, $S_0 \Rightarrow^{DII} D1$ 에 해당되는 패킷들의 실질적인 중간 출발지 집합을 $\{s^1, s^2, \dots, s^{|DII|}\} \subseteq S_0$, 목적지 영역 집합 $D1$ 을 $\{R_1, R_2, \dots, R_{|DII|}\}$ 라고 하면, $1 \leq j < |DII|$ 인 임의의 j 에 대해, $s^{j+1} - s^j = 2$ 이고 $R_{j+1} - R_j \geq 2$ 이며, $|R_j| \geq 1$ 이다. 따라서 $1 \leq j < k \leq |DII|$ 인 모든 j 와 k 에 대해, $s^k - s^j \leq R_k - R_j$ 이며, 기존의 연구 결과 [5]를 이용하여 $S_0 \Rightarrow^{DII} D1$ 인 패킷들 사이에 충돌이 없음을 보일 수 있다. 또한, $S_1 \Rightarrow^{DII} D0$ 의 패킷들 간에도 충돌이 없다.

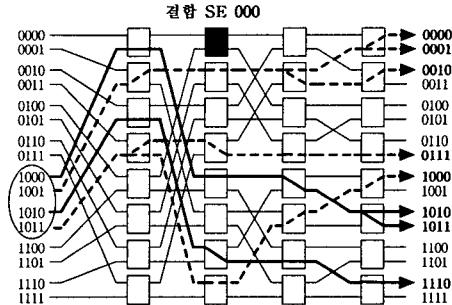
3.2. 알고리즘의 예와 기준 알고리즘과의 비교

그림 2와 3은 출발지 0에서 8개의 목적지 0, 1, 2, 7, 8, 10, 11, 14로 멀티캐스트 패킷을 전송할 때, $FTM-M$ 과 $FTMRA[3]$ 의 단계 2에 대한 라우팅 예를 각각 보이고 있다. $FTM-M$ 의 경우, MIN은 2 단에서 000, 111과 1 단에서 010, 101 등 4 개의 결합 SE가 있고, F_0 결합이다. $FTMRA$ 는 2 단의 SE 000이 결합이 있는 경우이다.

$FTM-M$ 의 경우, 멀티캐스트 목적지는 목적지 영역 집합 $D0 = \{[0000,0010], [0111,0111]\}$ 과 $D1 = \{[1000,1000], [1010,1011], [1110,1110]\}$ 으로 구분된다. $|D0| = 2$ 이고 $|D1| = 3$ 이므로, 단계 1에서 출발지 0에서 크기가 6인 임시 출력 리전 $IR = [1000,1101]$ 로 패킷을 복사한다. 단계 2에서, IC 는 $S_0 = \{1000, 1010, 1100\}$ 과 $S_1 = \{1001, 1011, 1101\}$ 로 구분된다. 그림 2의 굵은 실선 화살표와 같이, 1000, 1010, 1100은 복사된 패킷을 각각 $[1000,1000]$, $[1010,1011]$, $[1110,1110]$ 으로 전송한다. 또한, 그림 2의 굵은 점선 화살표와 같이, 1001과 1011은 복사된 패킷을 각각 $[0000,0010]$ 과 $[0111,0111]$ 로 전송한다. S_1 의 1101은 자신의 패킷을 버린다. 모든 패킷은 결합 SE들을 통과하지 않음을 알 수 있다.

그림 2 : $FTM-M$ 의 단계 2 예.

$FTMRA$ 의 경우, 멀티캐스트 목적지는 4개의 영역 $\{0000,0010\}$, $\{0111,1000\}$, $\{1010,1011\}$, $\{1110,1110\}$

그림 3 : $FTMRA[3]$ 의 단계 2 예.

으로 표현된다. 단계 2에서는 그림 3에서와 같이, 이전 단계의 임시 출력 노드인 1000, 1001, 1010, 1011에서 복사된 패킷을 각각 $[1010,1011]$, $[0000,0010]$, $[1110,1110]$, $[0111,1000]$ 으로 전송한다. 여기서 1011에서 $[0111,1000]$ 로 전송되는 패킷이 F_0 에 속한 2 단의 SE 111과 1 단의 SE 101을 통과하고 있다. 따라서, 이 SE 중 하나 이상이 추가로 결합이 발생하면 $FTMRA$ 는 멀티캐스팅을 수행할 수 없다.

4. 성능 평가

멀티캐스트 알고리즘의 성능은 그 알고리즘에 의해 표현되는 멀티캐스트 트리의 크기(volume)와 깊이(depth)에 의해 평가될 수 있다. 일반적으로, 결합을 허용하는 알고리즘은 결합을 고려하지 않은 알고리즘에 비해 성능이 저하되지만, 그 저하의 정도는 알고리즘의 효율성을 판단하는 척도중 하나이다.

본 알고리즘의 성능을 멀티캐스트 트리의 크기, 즉 사용되는 링크 수의 관점에서 기준에 제안된 알고리즘들 [3, 4]와 비교하여 분석한다. 알고리즘들은 모두 두 번의 순환을 통해 멀티캐스팅을 수행하므로, 트리의 깊이는 동일하다.

[4]에서 제안한 개선 알고리즘 I은 결합이 없는 MIN에서 제안된 것이며 EMA 으로 표기한다. 그리고, 멀티캐스트 목적지를 나타내는 영역의 집합은 EMA 와 $FTMRA$ 에서 $D = \{R_1, R_2, \dots, R_r\}$ 로 표시하고, $FTM-M$ 에서 $D0 = \{R_1^0, R_2^0, \dots, R_{|D0|}^0\}$ 과 $D1 = \{R_1^1, R_2^1, \dots, R_{|D1|}^1\}$ 로 표시한다. 여기서, $R \in D$, $R \not\subseteq M(0)$, $R \not\subseteq M(1)$ 인 R 이 없으면, $D=D0 \cup D1$ 이고 $r=|D0|+|D1|$ 이며, 그러한 R 이 존재하면 $r+1=|D0|+|D1|$ 이다. 또한, 알고리즘 A에 의해 사용되는 전체 링크 수는 $TNL(A)$ 로 표시한다.

EMA 는 첫 번째 순환에서 크기가 r 인 영역 FR_0 로 복사를 하고, 두 번째 순환에서 각각 크기가 $|R_j|$ ($1 \leq j \leq r$)인 r 개의 영역으로 라우팅하므로, $TNL(EMA)$ 의 근사치는 다음과 같다 [4].

$$TNL(EMA) \cong \sum_{i=0}^{r-1} \lceil \frac{|FR_0|}{2^i} \rceil + \sum_{j=1}^r \sum_{i=0}^{|R_j|-1} \lceil \frac{|R_j|}{2^i} \rceil$$

정확한 링크 수를 산출하기 위해, 한 패킷의 목적지 영역 R 을 $[a, b]$ 로 표시하고, a 와 b 의 n -비트 표현에서 일치하는 최상위 비트의 개수를 $s(a, b)$ 라고 표시하자. 이 패킷은 $(n-s(a, b)-1)$ 단에 있는 SE에서 첫 번째로 브로드캐스팅 기능을 수행한다. 그 SE에서 출력포트 0 번과 1 번을 통하여 도달하는 목적지의 수 $Up(a, b)$ 와 $Down(a, b)$ 는 각각 다음과 같이 표시된다.

$$Up(a, b) = 2^{n-s(a, b)-1} - a \bmod 2^{n-s(a, b)-1}$$

$$Down(a, b) = b \bmod 2^{n-s(a, b)-1} + 1$$

그리고, 이 패킷이 사용하는 전체 링크의 수 $NL(R)$ 는 다음과 정리할 수 있다.

$$NL(R) = s(a, b) + \sum_{i=0}^{n-s(a, b)-1} (\lceil \frac{Up(a, b)}{2^i} \rceil + \lceil \frac{Down(a, b)}{2^i} \rceil)$$

또한, $TNL(EMA)$ 는 다음과 같이 나타낼 수 있다.

$$TNL(EMA) = NL(FR_0) + \sum_{i=1}^k NL(R_i)$$

$FTMRA$ 에서 결합 SE 번호의 최상위비트를 α_{n-1} 라고 할 때, D'' 는 영역 R 드 $M(\alpha'_{n-1})$ 로 구성되는 집합이고, D'' 는 $D-D''$ 인 집합이다 [3]. $FTMRA$ 에서 사용되는 전체 링크 수는 첫 번째 순환에서만 EMA 와 차이가 생긴다. 첫 번째 순환의 중간 영역을 FR_1 ($|FR_1| = 2 \times \max(|D'|, |D''|)$)라고 하면, $TNL(FTMRA)$ 는 다음과 같다.

$$TNL(FTMRA) = NL(FR_1) + \sum_{i=1}^k NL(R_i)$$

$FTM-M$ 에서 단계 1의 중간 영역을 FR_m ($|FR_m| = 2 \times \max(|D0|, |D1|)$)이라고 하면 사용되는 전체 링크 수 $TNL(FTM-M)$ 은 다음과 같다.

$$TNL(FTM-M) = NL(FR_m) + \sum_{i=1}^m NL(R_i^0) + \sum_{i=1}^m NL(R_i^1)$$

그림 4는 네트워크 크기가 64 ($n=6$)인 경우, 알고리즘 $FTM-M$, $FTMRA$, EMA 에 대해 멀티캐스트 목적지의 수에 따라, 목적지 당 사용되는 전체 링크의 평균 개수를 보이고 있다. $FTM-M$ 은 $FTMRA$ 과 거의 일치하고, EMA 와 근소한 차이를 보인다. 모든 알고리즘에서, 목적지의 수가 증가할수록, 공유하는 링크가 증가하므로 목적지 당 링크 수는 감소하며 알고리즘 간의 성능차이가 더욱 좁혀진다.

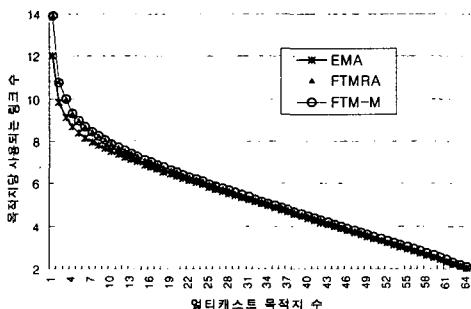


그림 4: 목적지 수에 대한 사용 링크 수 ($n=6$).

그림 5는 네트워크의 크기, 즉 단의 수 n 의 변화에 따라, 각 알고리즘에서 사용되는 링크 수에 대해 비교한 결과이다. 그림에서 p 는 출력 링크가 멀티캐스트 목적지에 포함될 확률이며, p 가 0.10, 0.25인 경우에 대한 결과를 보이고 있다. 네트워크 크기가 커질수록 근소한 성능의 차이가 더욱 좁혀지고 있다.

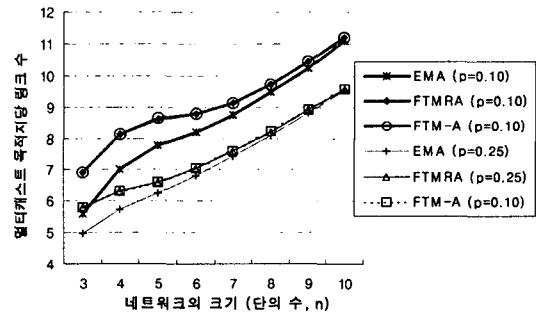


그림 5 : 네트워크 크기에 대한 사용 링크 수.

5. 결 론

본 논문은 순환 구조를 갖는 MIN에서 다수의 결합을 허용하는 멀티캐스팅 알고리즘을 제안하였다. MIN의 전체 스위칭 소자들을 두 개의 집합으로 구분하여, 모든 결합이 동일한 집합에 속한 경우에 패킷을 두 번 순환시킴으로써 멀티캐스팅을 수행한다. 또한, 임의의 멀티캐스트 패킷의 주소를 표시하기 위해 영역 부호화 방식을 사용한다. 그리고, 알고리즘에서 사용되는 전체 링크 수의 관점으로, 결합이 없거나 하나의 결합을 허용하는 기준의 알고리즘과 비교하여 시스템 성능을 평가하였다. 실험 결과는, 본 알고리즘이 기준의 알고리즘에 비해 성능의 저하가 미소하며, 멀티캐스트 목적지의 수와 네트워크 크기가 증가하면 성능의 차이가 거의 없어짐을 나타냈다.

6. 참고문헌

- [1] T. T. Lee, "Nonblocking Copy Networks for Multicast Packet Switching", IEEE Journal on Selected Areas in Communications, Vol. 6, pp.1455-1467, Dec. 1988.
- [2] X. Chen and V. Kumar, "Multicast Routing in Self-Routing Multistage Networks", Proc. of IEEE Infocom, pp.306-314, Apr. 1994.
- [3] J. Kim, J. Park, H. Yoon, and J. W. Cho, "Fault-Tolerant Multicasting in MIN's for ATM Switches", IEEE Communications Letters, Vol. 2, pp.331-333, Dec. 1998.
- [4] J. Park and H. Yoon, "Novel Recursive Multicast Algorithms Multistage Interconnection Networks for ATM Switches", High Speed Networks, Vol. 8, No. 4, pp.265-280, 1999.
- [5] J. Y. Hui, "Switching and Traffic Theory for Integrated Broadband Networks", Kluwer Academic Publishers, 1990.