WDM-기반의 클러스터 구현을 위한 가상 토폴로지 재구성 알고리즘

Virtual Topology Reconfiguration Algorithm for Implementing the WDM-based Cluster

박병섭
인하공업전문대학 컴퓨터시스템과

Byoung-Seob Park(bspark@inhatc.ac.kr)

요약

본 논문에서는 새로운 가상 토폴로지 재구성 기법을 제안하여 WDM 기반의 클러스터 시스템 구현을 위한 이론 연구를 수행하였다. 제안 기법의 핵심 아이디어는 파장 할당이 요구되는 연결 요청을 최대한 이어서 집합을 형성하는 것으로, 각 집합은 서로 소이며 집합 내의 연결 요청들 중에는 링크의 중복이 일어나지 않도록 한다. 이러한 조건을 만족하도록 집합을 형성한 후, 각 집합마다 파장을 하나씩 할당한다. 이는 일괄적으로 재구성을 수행하는 방법이다. 제안된 기법은 OWns 시뮬레이션도구를 사용하여 측정한 결과, 블로킹 확률 및 ADM 이용률 측면에서 First-fit 기법에 비해 10% 정도의 블로킹 감소율과 30% 정도의 ADM 이용률이 향상되었음을 보였다.

■ 중심어 : | WDM | 토폴로지 | 재구성알고리즘 | ADM | 블로킹 |

Abstract

We completed a new cluster system based on WDM by proposing a virtual topology reconfiguration schemes. The key idea of the proposed scheme is to construct a set with the longest chains of requests of connecting nodes which need to be assigned a wavelength. All the sets have no common factor, that is, there is no duplicated link among the requests of connecting. After making the set satisfying this condition, we could assign a wavelength to per corresponding set. We could reconfigure a virtual topology with this scheme collectively. we compared our scheme to existing approaches by the OWns simulation tool. As the results, we gained improved performances, reducing 10% of blocking rate and improving 30% of ADM utilization in terms of the blocking probability and the ADM utilization.

■ keyword : | WDM | Topology | Reconfiguration Algorithm | ADM | Blocking |

I. 서 론

WDM 기술은 하나의 광섬유 내에 서로 다른 파장의 광 신호를 다중화하여 전송하고 수신단에서는 파장별로 광 신호를 분리함으로써 단일 광섬유의 전송 용량을 증대시키기 위한 것이다. WDM은 추가적인 광 섬유망의 구축과 고속 전송 장비를 사용하지 않고도 전송 용량을 배가시킬 수 있어, 인터넷 이용의 증가와 광대역 네트워

* 본 연구는 2005년도 인하공업전문대학 교내연구비 지원에 의해 수행되었습니다.

접수번호 : #060328-001
접수일자 : 2006년 03월 28일
심사완료일 : 2006년 05월 03일
교신저자 : 박병섭, e-mail : bspark@inhatc.ac.kr

킹의 진전 및 새로운 광대역폭 서비스의 출현에 따른 대량의 정보를 수용할 수 있는 기술로 각광받고 있다. WDM 기술의 주요 핵심 기술은 광주파수 안정화 기술, WDM 다중/역다중 기술, 광링크 전송기술, 시스템 운용 관리기술 및 전송망 관리 기술이 있다[1-3].

국내의 경우는 2.5Gbps 종속신호를 기본으로 한 20Gbps(8채널), 40Gbps(16채널) WDM 시스템, 32채널(80Gbps 급)에서 80채널(200Gbps 급)의 WDM 시스템이 개발이 되었다. 또한 10Gbps 종속 신호를 기본으로 한 WDM 시스템의 경우 16채널(160Gbps급)~40채널(400 Gbps급)이 개발되었고, 그 후 ETRI 주도로 개발을 추진하여, 2000년에 160Gbps ADM 링 WDM 시스템을 개발하였다. 이어 2001년에 80채널 시스템을 개발하였고, 2002년에는 1.6Tbps(160채널)급의 시스템 기술 개발에까지 이르렀다. 현재 본 논문관련 연구동향은 다음과 같다[4-6]; 모든 광 노드가 하나의 광 파이버(optical fiber)에 연결되어 임의의 노드간 연결이 하나의 파장으로 직접적으로 재구성될 수 있는 단일홉 네트워크와 임의의 노드간 연결이 반드시 전기적 라우팅을 요하는 다른 노드를 통해 전달되어야 하는 다중홉 네트워크 존재한다[7]. RWA(Routing and Wavelength Assignment) 문제에서 라우팅과 파장 할당 기법은 별개로 제안되며 둘 다 NP-hard에 속하는 문제로, 휴리스틱 기법이 연구되고 있으며[6], 가상 토폴로지에서 라우팅 문제는 가상 노드들 사이에 가상 링크가 설정되도록 최단 경로 알고리즘이 적용하여 연구하고 있다.

본 논문에서 제안한 기법의 핵심 아이디어는 파장 할당이 요구되는 연결 요청을 최대한 이어서 집합을 형성하는 것이다. 각 집합은 서로 소이며, 집합 내의 연결 요청들 중에는 링크의 중복이 일어나지 않는다. 이러한 조건을 만족하도록 집합을 형성한 후, 각 집합마다 파장을 하나씩 할당한다. PS_k는 $PS_k = \{t_{ij} \mid t_{ij} \in T_k, t_{ij} = 1$이고 $j-1 \le i < N$이다.$\}$ 와 같이 정의할 수 있다. 파장 할당이 요구되는 연결 요청을 최대한 이어서 집합 PS_k를 형성하는데, 사용 가능한 파장의 개수는 제한되어 있기 때문에 이로 인한 블로킹은 불가피하다. 최적의 해는 집합 PS_k의 수를 최소한으로 만들어 최소한의 파장을 할당하는 것이다. 본 논문은 서론에 이어 제2장에서는 입력 태스크 특성에 따른 토폴로지 결정정책에 대해 간략히 언급하고, 제3장에서는 본 논문에서 제안하는 가상 토폴로지 재구성 알고리즘을 설명한다. 제4장에서는 시뮬레이션의 성능비교 결과를 제시함으로서 제안된 기법의 타당성을 보이고, 마지막으로 제5장에서 결론을 맺는다.

II. 입력 태스크의 특성에 따른 토폴로지결정

1. 개요

하나의 큰 애플리케이션은 적당한 크기의 태스크로 분할되어 수행이 된다. 이런 태스크들은 그에 적합한 특성을 가진 토폴로지를 결정함으로써 태스크 수행 성능을 향상시킨다. 또는 DAG(Directed Acyclic Graph)라는 방향성을 가지면서 사이클이 존재하지 않는 그래프로 표현되어 진다. 이런 DAG 그래프의 노드는 태스크이며 태스크의 실행 시간을 알려준다. 또한 에지는 하나의 태스크에서 다른 태스크로 데이터를 전달하는 방향, 즉 데이터 상속성을 나타내며 데이터를 전달하는데 소요되는 통신 시간을 알려준다. 이 결과를 이용하여 애플리케이션에 적당한 토폴로지를 할당한다. 토폴로지를 결정할 때 태스크간 통신 시간을 최소화하여 전체 태스크 실행 시간을 최적화할 수 있도록 해야 한다. 태스크 이외에 애플리케이션에서 사용되는 데이터의 구조가 트리인지, 링크드 리스트인지를 고려해서 토폴로지를 결정할 수도 있다.

2. 시스템 모델

본 연구에서 가정하는 클러스터 시스템의 토폴로지는 선형 배열 구조를 따른다[그림 1]. 클러스터 시스템에는 로드 밸런싱을 관리한다든지 그 이외의 여러 컨트롤을 하는 노드가 있다. 그래서 클러스터 웹 서버의 경우에 이 노드는 디스패처로서 작용한다. WDM 기반 클러스터 시스템 구현을 위해 이 컨트롤 노드는 다른 클러스터 내부의 노드 상태와 네트워크의 상태 정보, 파장 할당 정보 등을 테이블의 형태로 저장하며 이를 이용하여 파장을 할당하고 가상 토폴로지를 재구성한다. 다시 말해서, 중앙 집중식 파장 할당 방식을 채택한다. 물론 단일

홉 네트워크 환경이며 파장은 양방향성을 가지므로, (1,2)와 (2,1)은 같은 연결로 간주한다. 수행할 태스크 그래프는 링이나 메쉬, 토러스, 하이퍼 큐브 등과 같이 병렬 태스크에서 많이 사용되는 토폴로지 일수도 있고, 임의의 불규칙한(irregular) 토폴로지를 형성하기도 한다. 네트워크를 구성하는 노드이자 클러스터 시스템을 이루는 노드가 N개 있어서 노드0에서부터 노드(N-1)까지 존재한다[그림 2]. 선형 배열 구조이므로 링크의 개수는 (N-1)개가 된다. 네트워크에서 사용 가능한 파장의 개수는 W개이며, 태스크 그래프는 G(V, E)로 표현되며, 트래픽은 행렬로 표현한다.

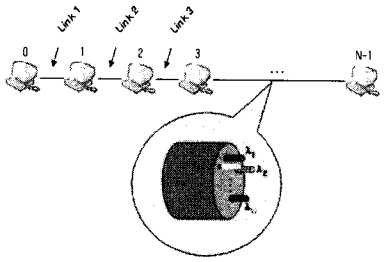


그림 1. 시스템 모델

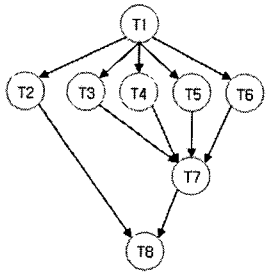


그림 2. DAG의 예

기존의 WDM 관련 연구에서는 무리한 가정으로 인하여 실제 네트워크와는 거리가 먼 모델링을 한 것이 사실이다. 예를 들어 블로킹 발생률을 최소로 하고자 하는 연구에서는 파장 전환 능력을 100%로 가정하고, ADM의 개수를 줄여 네트워크 비용을 줄이고자 하는 연구는 사용 가능한 파장의 개수가 무한하다고 가정을 한다. 무리한 가정은 발생 가능한 단점들을 원천적으로 고려하지 않기 때문에 신뢰성이 떨어진다고 볼 수 있다. 본 연구에서는 단일 홉 네트워크를 가정하는 만큼 각 노드에서의 파장 변환 능력은 0으로 고정시키고, 파장의 개수 역시 W개로 제한함으로써 보다 실제적인 네트워크 모델을 가정한다.

III. 제안된 가상 토폴로지 재구성 알고리즘

결정된 태스크 그래프의 각 에지(edge)에 적절한 파장을 할당함으로써 가상 토폴로지를 구성할 수 있는데, 이때 태스크 행렬을 미리 알고 있기 때문에 이는 정적인 가상 토폴로지 문제가 된다. 시간이 지남에 따라 수행하고자 하는 애플리케이션이 변화하여 다른 태스크 그래프가 요구되면 그에 적응적으로 파장을 할당해줌으로써 가상 토폴로지를 재구성한다. 최종적으로 제안하는 재구성 기법을 소개하겠다.

1. 파장 할당을 통한 가상 토폴로지 재구성 기법

파장은 양방향이므로 트래픽 행렬 T는 대칭(symmetric) 행렬을 이룬다. 출발지 노드와 목적지 노드가 같지 않기 때문에 아래의 트래픽 행렬 t_{00}, t_{11}, t_{22}, ... $t_{(N-1)(N-1)}$은 0이 되며 이 대각선을 기준으로 대칭을 이루어 symmetric matrix로 표현된다[그림 3].

$$T = \begin{pmatrix} t_{00} & t_{01} & t_{02} & \\ t_{10} & t_{11} & t_{12} & \cdots \\ t_{20} & t_{21} & t_{22} & \\ \vdots & & & \ddots \end{pmatrix}$$

노드1이 출발지인 트래픽

노드2가 목적지인 트래픽

$$T = \begin{pmatrix} 0 & t_{01} & t_{02} & \\ t_{10} & 0 & t_{12} & \cdots \\ t_{20} & t_{21} & 0 & \\ \vdots & & & \end{pmatrix}$$

$$t_{ij} = \begin{cases} 1 & traffic\ via\ i \rightarrow j \\ 0 & notraffic\ via\ i \rightarrow j \end{cases}$$

그림 3. Symmetric Matrix 구성

트래픽 행렬 T는 클러스터 시스템에서 수행하고자 하는 태스크 그래프로부터 얻어진다. 태스크는 여러 개의 sub-task로 나누어질 수 있으며 각 sub-task 간의 태스크 의존성을 토대로 그래프가 완성된다. sub-task는 클러스터 노드에 매핑이 되며, 의존성이 많은 태스크 간에는 자연히 통신량이 많음을 의미한다.

노드 i에서 노드 j로의 요청(request)이 있다면 t_{ij}는 1, 요청이 없다면 0으로 행렬에 표시할 수가 있으며, 파장은 양방향성 임을 가정하였기 때문에 노드 j에서 i로의 request 역시 t_{ji}로 나타낸다. 시스템 모델 부분에서 소개한 것처럼 노드의 개수는 N, 사용 가능한 파장의 개수는 W로 표시하며, 전체 트래픽의 양은 $T_{tot} = \sum_i\sum_j t_{ij}$이 된다.

클러스터를 구성하는 WDM 네트워크의 특성 상, 광경로를 설정하기 위해서는 라우팅을 통하여 물리적인 경로를 설정하고, 적절한 파장을 할당해 줌으로 인하여 가상적인/논리적인 경로를 설정해야한다. 가상 토폴로지 재구성 문제는 그래프 컬러링 문제로 해결할 수가 있는데, 이 때 파장의 개수를 최소로 할당하는 문제는 NP-complete이다[2]. 그러므로 그리디 알고리즘이나 휴리스틱 알고리즘 연구가 진행되었었다. 최소로 할당되는 파장의 개수는 노드를 어떤 순서로 배열하느냐에 따라 결과가 좌우된다.

제안 기법은 그래프 컬러링 문제와 유사하나 이를 적용할 수는 없다. 그래프 컬러링은 이웃한 노드나 이웃한 에지(edge)가 같은 색, WDM 환경에서는 같은 파장을 사용할 수 없다는 원칙을 기본으로 파장을 할당하는 기법이다. 여기서 이웃한다는 것은 물리적인 경로를 생각하거나, 라우팅 경로를 고려한 것이 아니다. 가상 토폴로지 내에서의 이웃함을 의미하는 것이다. 하지만 본 연구에서는 각 광경로의 물리적인 루트가 이미 결정되어 물리적인 경로의 고려가 불가피하므로 이를 고려하여 파장을 할당해야만 한다. 클러스터의 물리적인 토폴로지가 선형 배열이기에, 노드 1과 노드 3의 통신은 꼭 노드 2를 거쳐야만 한다.

제안 기법의 핵심 아이디어는 파장 할당이 요구되는 연결 요청을 최대한 이어서 집합을 형성하는 것이다. 각 집합은 서로 소이며, 집합 내의 연결 요청들 중에는 링크의 중복이 일어나지 않는다. 이러한 조건을 만족하도록 집합을 형성한 후, 각 집합마다 파장을 하나씩 할당한다. 이를 수식으로 표현하기 위해 중요한 변수를 먼저 소개하면 다음과 같다.

- $V=\{v_0, v_1, \ldots, v_{N-1}\}$ - 노드의 집합
- $E=\{e_0, e_1, \ldots, e_{Ttot-1}\}$ - 에지(edge)의 집합
- $T_{tot} = \sum_i\sum_j t_{ij}$ - 전체 트래픽의 양
- D_{in}^{v} - 노드 v의 in-degree 최대 연결의 수
- D_{out}^{v} - 노드 v의 out-degree 최대 연결의 수
- p_{ij} - 노드 i에서 j까지의 광경로
- PS_k - 트래픽 T의 부분집합, k는 인덱스

태스크 그래프 G는 G(V, E)로 나타내며, 여기에서 V는 태스크가 할당될 클러스터 노드의 집합이므로 v의 인덱스는 0부터 (N-1)까지로 V의 원소의 전체 개수는 N이다. 집합 E는 태스크 간의 연결을 나타내며, 전체 요구되는 태스크 간 연결의 전체 개수는 T_{tot}이므로 E의 인덱스는 0부터 T_{tot-1}이 된다.

다음으로 D는 클러스터 노드의 연결 가능한 최대 degree 수를 나타낸다. 시스템 모델 가정에서, 모든 노드의 파장 변환 능력은 없다고 했다. 하지만, 사용 가능한 파장이 W라고 가정했을 때, 모든 노드는 W개의 파장 송신기와 수신기를 갖고 있음이 묵시적으로 함께 가정되었다고 볼 수 있다. 앞의 그림에서 노드 T7의 in-degree는 4이고, out-degree는 1이다. 이때 in-degree 값이 최고인 노드의 in-degree 값도 W를 넘을 수가 없다. 모든 노드는 가정에 의하여 W개의 송신기와 수신기를 보유하고 있으므로 그 이상의 연결 요청은 블로킹된다. 그리고 태스크 간의 의존도는 통신 빈도와 연관이 있을 뿐이지, 방향성을 갖고 있음을 의미하지는 않는다. 또한 파장의 양방향성을 가정하였기 때문에 D_{in}^{v} 와 D_{out}^{v}의 구분이 없이 $D_{in}^{v} = D_{out}^{v}$이며, 가정에 의하여 최대 수락할 수 있는 송수신기의 개수는 W이므로, D_{in}^{v} 와 D_{out}^{v}의 값은 W이다.

p_{ij}는 광경로를 표시하는 변수로서, 출발지 노드는 i이고, 목적지 노드가 j인 광경로이다.

$$p_{ij}=\begin{cases}1 & \textit{lightpath establishment}\\ 0 & \textit{no establishment}\end{cases}$$

p_{ij} 연결 요청 (i, j)에 광경로가 할당되었다면 1을, 할당되지 않았다면 0의 값을 갖는다. $\sum_i\sum_j p_{ij} \le T_{tot}$ 과 $\sum_j p_{ij} \le W$을 만족한다. 설정된 광경로의 최대 값은 전체 트래픽의 양을 초과할 수 없으며, 또한 노드 i를 출발지로 하는 광경로의 최대 값이 사용 가능한 파장의 개수 W보다 많을 수 없다. 또한 $\sum_j p_{ij} \le D_{in}^{i}$이고, $\sum_j p_{ij} \le D_{out}^{i}$을 만족한다.

PS_k는 전체 트래픽의 부분 집합이고, k는 인덱스이므로 PS_0, PS_1, ... 과 같이 표현하며 각 PS_k에 파장이 하나씩 할당된다. 또한 모든 k에 대하여 PS_k는 서로 소이며, PS_k를 모두 합하면 전체 트래픽을 나타낼 수 있다. 파장 할당이 요구되는 연결 요청을 최대한 이어서 집합 PS_k를 형성하는데, 사용 가능한 파장의 개수는 제한되어 있기 때문에 이로 인한 블로킹은 불가피하다. 최적의 해는 집합 PS_k의 수를 최소한으로 만들어 최소한의 파장을 할당하는 것이다. 동시에, 토폴로지를 재구성 할 시에 이전 연결되었던 요청이 또 요청되는 경우에는 되도록이면 예전과 같은 파장이 할당되게 해줌으로써 파장 전환하는 오버헤드를 줄여야 한다. 최악의 경우는 (0, 1), (0, 2), ... (0, N-1)과 같은 트래픽이 행렬이 주어지는 경우이며, 이 때는 모든 집합이 하나의 원소만을 갖게 된다. $PS_k \le T_{tot}$ 을 만족하긴 하지만, $W<T_{tot}$라면 $(T_{tot}-W)$ 만큼의 블로킹이 발생함을 알 수 있다.

$PS_k = \{t_{ij} \mid t_{ij} \in T_k, t_{ij} = 1$이고 $j-1 \le i < N$이다.$\}$ 물론 트래픽 행렬이 symmetric 하므로 $i \neq j$를 만족해야 한다. 전체 트래픽 요청이 모두 집합에 포함될 때까지 반복적으로 집합을 계산해야 하는데, 이때 트래픽 전체 행렬 집합 T가 변경된다. 이미 PS_k에 포함된 원소들은 다음 PS_{k+1}을 선택하기 위해 고려되어야 하는 전체 행렬 T_{k+1}에는 제외되어야 한다. 즉, T_k는 PS_{k-1}의 여집합이 된다. T_0는 트래픽 행렬의 전체 집합임을 가정한다. 그러므로 $PS_0 \cap PS_1 \cap \ldots \cap PS_{N-1} = \varnothing$ 이다.

예를 들어, (1, 2), (2, 3), (1, 4), (2, 5), (3, 6), (4, 5), (5, 6), (4, 7), (5, 8), (6, 9), (7, 8), (8, 9)와 같은 태스크 그래프가 있다고 가정해 보겠다. 이 연결은 매쉬 구조를 이루며 다음의 [그림 4]와 같다.

그림 4. 물리적인 토폴로지와 수행 태스크 토폴로지의 예

위의 태스크 그래프를 트래픽 행렬 T로 표현하면 다음과 같다.

	1	2	3	4	5	6	7	8	9
1	0	1	0	1	0	0	0	0	0
2	1	0	1	0	1	0	0	0	0
3	0	1	0	0	0	1	0	0	0
4	1	0	0	0	1	0	1	0	0
5	0	1	0	1	0	1	0	1	0
6	0	0	1	0	1	0	0	0	1
7	0	0	0	1	0	0	0	1	0
8	0	0	0	0	1	0	1	0	1
9	0	0	0	0	0	1	0	1	0

==> symmetric matrix T_0

$PS_0 = \{(1, 2), (2, 3), (3, 6), (6, 9)\}$ 이 결정된다. 그 결과 변경된 트래픽 행렬 T_1은 다음과 같으며 이로 인해 얻어지는 $PS_1 = \{(1, 4), (4, 5), (5, 6), (7, 8), (8, 9)\}$ 가 된다.

	1	2	3	4	5	6	7	8	9
1	0	0	0	1	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	1	0	0	0	1	0	1	0	0
5	0	1	0	1	0	1	0	1	0
6	0	0	0	0	1	0	0	0	0
7	0	0	0	1	0	0	0	1	0
8	0	0	0	0	1	0	1	0	1
9	0	0	0	0	0	0	0	1	0

==> symmetric matrix T_1

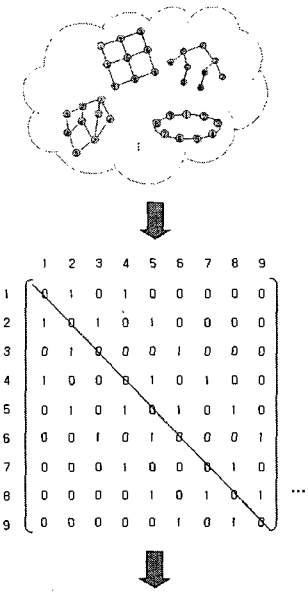
같은 방법으로 되풀이하여, PS_2 = {(2, 5), (5, 8)}, PS_3 = {(4, 7)}을 얻게 된다. 결과적으로 파장은 각 집합에 하나씩 할당되므로 총 4개의 파장이 요구되며, 인접한 광경로끼리 같은 파장이 할당되어 ADM(Add/Drop Multiplexer)를 공유할 수 있어 ADM의 이용률을 높일 수 있다는 장점이 있다. 또한 한번의 행렬 룩업을 통하여 PS_k 집합을 결정할 수 있으므로 계산 복잡도의 측면에서도 효율적인 방법이다.

T_0 행렬에서 PS_0 = {(1, 2), (2, 5), (5, 6), (6, 9)}로 구성할 수도 있다. 이럴 경우 PS_1 = {(1, 4), (4, 5), (5, 8), (8, 9)}, PS_2 = {(2, 3), (3, 6), (7, 8)}, PS_3 = {(4, 7)}로 집합을 구성할 수 있는데, 중요한 것은 집합의 개수에는 변함이 없다는 점이다. PS_k 집합을 형성하기 위해 연결 가능한 모든 경우의 수를 다 비교하여 가장 효율적인 연결이 어떤 것인지 결정하는 알고리즘을 개발할 수도 있을 것이다. 허나 어떤 순서로 연결을 하든 간에 결과적으로 얻어지는 집합의 개수에는 변함이 없기 때문에 휴리스틱한 알고리즘이 필요 없게 된다. 집합의 개수에 변함이 없다는 것은 어떤 방식으로 집합을 형성하든 간에 사용되는 파장의 개수는 같음을 의미하며 자원을 확보하는데 있어서 같은 효과를 보인다. 오히려 다른 휴리스틱 알고리즘 적용 시, 계산 복잡도가 증가하게 된다.

전체 트래픽 집합으로부터 위의 방법을 이용하여 가능한 모든 부분 집합을 구성하고 나면 원소의 개수가 가장 많은 부분 집합의 원소들부터 파장을 할당한다. 파장의 수에는 제한이 있기 때문에 이로 인한 링크 블로킹은 불가피하다. 하지만 원소의 개수가 많은 부분 집합에 먼저 파장을 할당해 줌으로써 블로킹 발생 확률을 감소시킬 수 있으며 시뮬레이션을 하여 실험적인 분석을 통해 확인하였다.

2. 제안 기법의 분석

제안 기법은 클러스터에서 수행할 태스크의 토폴로지가 정형화된 매시나 트리 구조이거나 불규칙한 토폴로지일지라도 영향 받지 않고 적용할 수 있다. 태스크 토폴로지를 행렬 T로 표현하는데, 이는 대칭 행렬로 표현되며 이러한 이유로, $O(N^2)$의 복잡도가 아닌 $O(N^2/2)$의 복잡도를 갖는다. 네트워크에서 사용 가능한 파장의 개수는 제한되어있다. 제한된 리소스를 효율적으로 활용하여 파장, ADM의 이용률을 증가시키고 동시에 블로킹 발생 확률을 감소시키는 제안 기법이 요구된다. 간략하게 제안 기법을 살펴보겠다[그림 5].



$PS_k = \{t_{ij} \mid t_{ij} \in T_k,\ t_{ij} = 1$이고 $j-1 \le i < N.\}$

그림 5. 제안기법의 매핑

위 그림에 따라 PS_0, PS_1, ... 집합을 완성하며 최대 PS_{W-1}개의 집합을 얻을 수 있다. 집합 마다 하나의 파장이 할당되므로 최대 할당 가능한 파장의 개수만큼의 집합만 유효하게 된다. 이런 제안 기법이 가능한 이유는 클러스터 시스템의 물리적인 토폴로지가 선형 배열이기 때문에 라우팅을 수행하기 않고도 물리적인 경로를 알고 있기 때문이다.

(1) First-Fit 방법과의 비교

파장을 할당하는 연구는 많이 진행되어 왔고, 앞에서도 소개한 바 있다. 하지만 WDM 기반의 클러스터 시스템에 관련된 연구는 전무한 상태이므로 제안 기법의 성능을 비교하기 위한 비교 대상 연구가 없다. 시뮬레이션을 위해 사용된 OWns 에 구현되어 있는 WDM 모듈은 First-Fit이다. 이는 프로세스 스케줄링 기법 중 선입 선

처리 방식과 흡사하다. 일반적인 파장할당 알고리즘은 다음 포와 같다.

표 1. Random 방식 및 First-Fit 알고리즘

기 법	특 징
Random	관련된 연구 분야에서의 첫 번째로 제안된 방식으로, 단순히 사용 가능한 파장 중의 하나를 무작위로 선택하는 방식이다.
First Fit	이 방식에서는 모든 파장들을 넘버링하고, 파장 할당이 요구될 때, 높은 번호보다는 낮은 호를 가진 파장을 선택하는 방법이다. 이 방식은 장점은 어떠한 전역 정보도 필요로 하지 는데 있다. 또한 랜덤 파장 할당에 비해서 계산 비용이 더 작다.

(2) 고찰

제안 기법은 정적인 가상 토폴로지 재구성 기법이며 일괄적으로 재구성을 처리하는 기법이다. 정적인 가상 토폴로지 재구성이 가능하기 위해서는 재구성 수행 이전에 처리해야 할 태스크 토폴로지, 다시 말해서 재구성해야 할 타겟 토폴로지가 미리 주어져 있어야 한다. 본 논문에서는 클러스터 시스템 환경을 구현하기 때문에 클러스터 노드의 수도 제한적이고, 태스크 토폴로지의 변화도 동적인 네트워크 상황에 비해 정적이라 할 수 있다. 또한 일반 네트워크 상황과 비교했을 때, 노드 폴트(fault)가 발생할 확률도 상대적으로 낮으며 이외의 네트워크에서 발생할 수 있는 다양한 상황들이 고려되지 않는다. 제안된 기법은 WDM 네트워크 분야에 적용이 가능하며, WDM 기반 클러스터 시스템 연구 자체를 보다 확장시킨다면 WDM 기반 그리드 시스템을 구현하는데 기여할 수 있을 것이다. WDM 네트워크에서 주된 연구 주제는 동적인 네트워크 환경에서 블로킹 발생률을 최소화하는 파장 할당 알고리즘, 가상 토폴로지 재구성 기법이나 혹은 토폴로지 재구성을 하는 시점을 결정하는 연구이다. 최근에는 네트워크 비용을 고려하여 ADM의 이용률을 최대화함으로써 비용을 최소화하고자 하는 연구가 주류이다. 연구 결과물로 발표된 논문들의 대다수가 위의 주제를 다루고 있으며, 모두 동적인 네트워크 환경을 가정하였다.

IV. 제안기법의 시뮬레이션

이 정에서는 제안하는 기법의 의사코드를 소개하며, 타 연구 분야에서의 적용 가능성을 고려하여 추가적으로 제안한 기법의 의사코드도 소개한다. 시뮬레이션은 OWns 기반 시뮬레이션을 수행했는데, OWns (Optical WDM network simulator)는 멀티 프로토콜 네트워크 시뮬레이터로써 네트워킹 리서치를 위해 널리 이용되고 있는 ns-2를 확장하는 VINT(Virtual Internet) 프로젝트의 한 부분으로써 개발되었다[8]. 네트워크의 스위칭과 라우팅 특성이 구현되어 있으며, WAssingLogic 모듈을 이용하여 물리적인 링크에서 여러 파장을 사용함으로써 개별적인 할당을 가능하게 하여 물리적인 토폴로지에 독립적으로 가상 토폴로지를 구성할 수 있게 한다. OWns 아키텍쳐의 핵심적인 부분은 광 스위칭 노드, 다수의 파장을 사용할 수 있는 링크, RWA 알고리즘 지원이다. OWns가 물리적인 토폴로지와 가상 토폴로지를 모두 지원한다고 하지만, 실제 애니메이터로 구현되는 것은 물리적인 토폴로지 뿐이다. 그러므로 사용자는 할당된 파장으로 구성된 가상 토폴로지를 시각적으로 볼 수가 없다. 이를 해결하고자 입력 트랙픽 파일을 수정하고, 이를 이용하여 우리가 제안한 기법을 OWns로 시뮬레이션하여 제안 기법의 정당성을 입증한다.

1. 제안 기법의 알고리즘

제안 기법은 물리적인 토폴로지에 적합한 가상 토폴로지를 형성함으로써 매핑에서의 오버헤드를 최소화하고 블로킹 발생을 최소화하는 것이 목표이다. 구현하고자 하는 클러스터는 선형 배열로 구성되어 있으므로 이에 최적화되도록 트래픽을 재구성하여 파장을 할당한다. 앞에서 설명한 방법으로 최대로 연결 가능한 트래픽들로 집합을 구성하며 집합마다 하나의 파장을 할당해준다.

<Algorithm>----------

Procedure IO_BLOCKING

IF source and destination nodes have available transceivers

next step

ELSE the request enters the wait-queue

Procedure WL_BLOCKING

IF the same wavelength can be assigned on all the links along the path

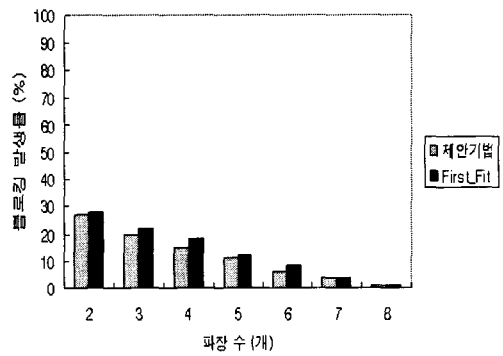
```
from the source to the destination
        wavelengths are candidate wavelengths
    ELSE the request enters the wait-queue

    Procedure RECONFIG_VT
    IF traffic is required and there is at least a wavelength in links between
source and destination node
        // parameter index stands for destination node
        call procedure GET_OPTIMAL(index)

    Procedure GET_OPTIMAL(index)
    IF index is smaller than maximum number of node
        IF there is a traffic whose source is index
            wavelength is assigned to that traffic
            call procedure GET_OPTIMAL(new_index)
```

2. 시뮬레이션 결과 분석

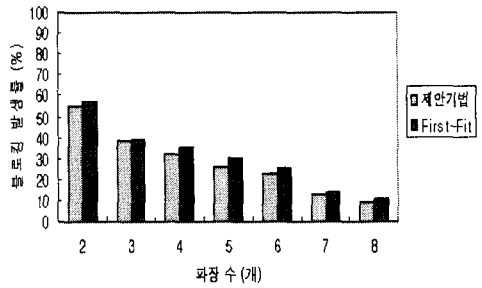
블로킹 발생률과 ADM(Add/Drop Multiplexer) 이용률에 대하여 시뮬레이션을 수행하였으며 결과는 다음과 같다.

(1) 블로킹 발생률

파장의 개수를 2개에서 8개까지 증가시키면서 제안하는 알고리즘과 기존 OWns에서 구현되어 있는 First Fit 방법을 블로킹 발생률을 기준으로 비교하였다. 노드의 개수를 10개에서 50개까지 증가시켜가면서 매 경우마다 시뮬레이션을 수행했으며, 그 중 노드의 개수가 20개, 30개일 때의 그래프는 다음과 같다.



a) 노드 20개인 경우



b) 노드 30개인 경우

그림 6. 파장 수 변화에 따른 블로킹 발생률

파장의 개수가 증가함에 따라 블로킹 발생률은 확실히 감소한다. First-Fit에 비해 최대 약10%의 블로킹 감소율을 보인다. 클러스터라는 제한적인 물리적 토폴로지를 기반으로 거의 정형화된 트래픽에 대한 시뮬레이션이므로 큰 성능의 향상을 보진 못했지만, 보다 동적인 네트워크 환경에 제안 기법을 적용한다면 First Fit과 비교할 시, 10% 이상의 성능 향상을 보일 거라 생각된다.

(2) ADM 이용률

약 10% 블로킹 발생을 감소시킨 제안 기법은 ADM 이용률 측면에서 보다 뚜렷한 성능 향상을 보였다. 제안 기법은 되도록 연속적으로 파장 할당이 가능하도록 하기 때문에 인접한 트래픽 요청이 ADM을 공유할 수 있도록 해준다. 기존의 방법들은 ADM의 이용률을 고려하지 않고 블로킹만을 고려해왔으며, 이는 평균 약 33%의

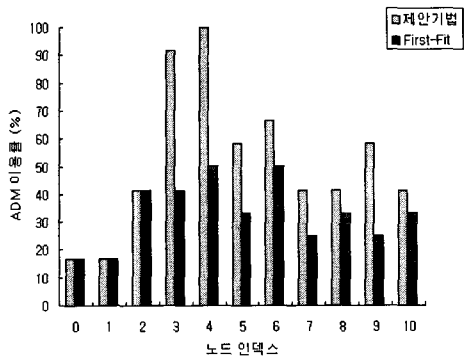


그림 7. 각 노드에서의 ADM 이용률

ADM 이용률을 기록한다. 아래의 그래프는 ADM 이용률의 측면에서 제안 기법과 First-Fit 간의 성능 평가 결과를 보여준다.

이는 임의의 상황에서 노드 0에서 노드10까지의 ADM 이용 현황을 비율로 표현한 것이다. 기존 기법은 최대 50%, 제안 기법은 인접한 광경로끼리 같은 파장이 할당되어 ADM를 공유할 수 있어 ADM의 이용률을 높일 수 있다는 장점이 있다. 따라서 최대 100%까지 ADM을 이용하므로 I/O 측면에서 비용 효율적인 기법이라 하겠다.

V. 결 론

제안 기법은 토폴로지를 재구성을 일괄적으로 처리한다. 병렬적으로 태스크를 처리할 때는 태스크 간의 의존도와 더불어 동기화 또한 중요한 요소이다. 일괄적으로 가상 토폴로지를 재구성 할 시에, 네트워크는 정지하게 된다. 동적인 네트워크 상황에서 정지 시간이 길어진다는 것은 커다란 문제가 되지만, 병렬 처리에서는 오히려 동기화가 자동적으로 이루어진다는 점에서 장점이 될 수 있다. 제안 기법을 OWns를 이용하여 시뮬레이션을 함으로써 OWns에서 구현되어 지원하는 First-Fit 기법과 비교 분석하였으며, 그 결과 약 10% 정도의 블로킹 감소율을 보임을 알 수 있었다. 또한 네트워크 비용의 대부분을 차지하는 ADM 이용률 관점에서 비교 분석한 결과, First-Fit 기법에 비해 약 30% 정도 ADM을 활용하고 있음을 알 수 있었다. ADM 이용률은 I/O 블로킹과 밀접한 관련이 있으며, 블로킹은 일반적으로 알려진 connection 블로킹을 의미한다. 클러스터 환경을 구현한 것이므로 노드의 수나 트래픽 발생에서도 제한이 많은 정적인 시뮬레이션이라고 할 수 있다. 보다 동적인 네트워크에 연구 결과를 적용한다면 보다 좋은 성능을 보일 것이라 예상된다.

참 고 문 헌

[1] J. P. Labourdette, G. W. Hart, and A. S. Acampora, "Branch-Exchange Sequences for Reconfiguration of Lightwave Networks," IEEE Transactions on Communications, Vol.42, No.10, pp.2822-2832, Oct. 1994.

[2] J. Rothe, "Heuristics versus Completeness for Graph Coloring," Chicago Journal of Theoretical Computer Science The MIT Press, Vol.2000, Article1, Feb. 2000.

[3] J. S. Choi and D. Su, "A Functional Classification of Routing and Wavelength Assignment schemes in DWDM networks: Static case," NRC 2000, New Jersey, USA, Apr. 2000.

[4] B. Kannan, S. Fotedar, and M. Gerla, "A two level optical star WDM metropolitan area network," Global Telecommunications Conference GLOBECOM '94 IEEE, NOV-DEC pp.563-566, 1994.

[5] H. Lee, J. Kim, S. J. Hong, and S. Lee, "Processor allocation and task scheduling of matrix chain products on parallel systems," Parallel and Distributed Systems, IEEE Transactions, Vol.14, No.4, pp.394-407, Apr. 2003.

[6] N. Baneriee and S. Sharan, "A evolutionary algorithm for solving the single objective static routing and wavelength assignment problem in WDM networks," Proceedings of International Conference, pp.13-18, 2004.

[7] D. Coudert and H. Rivano, "Lightpath assignment for multifibers WDM networks with wavelength translators," Global Telecommunications Conference, GLOBECOM '02. IEEE, Vol.3, pp.2686-2690, Nov. 2002.

[8] http://www.eecs.wsu.edu/~dawn/software/owns.html

저 자 소 개

박 병 섭(Byoung-Seob Park) 종신회원

- 1989년 2월 : 충북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : 서강대학교 대학원 전산학과(공학석사)
- 1997년 2월 : 서강대학교 대학원 전산학과(공학박사)
- 1997년 4월 : 국방과학연구소 선임연구원
- 2000년 3월 : 우석대학교 컴퓨터교육과 교수
- 2002년 9월 : 인하공업전문대학 컴퓨터시스템과 교수

<관심분야> : 컴퓨터네트워크, WDM, RFID/USN