# 10GbE 스위치간 링크 집합을 위한 프레임 분배방식

준회원 이 호 영*, 정회원 이 숭 희**, 이 종 협***

# Frame Distribution Methods for Link Aggregation between 10GbE Switches

**Ho-Young Lee*** *Associate Member*
**Soong-Hee Lee**, Jong-Hyup Lee***** *Regular Members*

요 약

10GbE 스위치간 링크 집합 기술은 프레임 분배 알고리즘의 설계에 따라 스위치에서 성능의 차이를 가져오므로 링크 집합의 장점을 살리려면 좋은 성능을 가지는 프레임 분배 알고리즘이 필요하다. 기존에 제시된 스위치와 스위치 사이에서의 프레임 분배 방식으로 정적 / 동적 프레임 분배 방식이 있으나 이 방식은 수신단말이 여러 개의 물리적인 링크 중 하나의 링크에 고정이 되어 프레임을 전송하기 때문에 집합된 링크를 모두 사용하지 못해 링크 집합의 이점을 충분히 활용하지 못한다. 이 문제점을 해결하기 위해 스위치간 링크 집합에 패딩을 이용한 분배 방식의 적용을 제안하고 성능 측면에서 정적 / 동적 분배 방식과 비교하였다. 그 결과 제안된 패딩 방식의 성능이 부하가 0.7이하이고 프레임 평균 길이가 954 바이트보다 더 긴 경우에 더 우수한 성능을 보였다.

Key Words : Link Aggregation, Frame Distribution Methods, 10GbE.

ABSTRACT

The link aggregation between 10GbE switches requires an advanced frame distribution method to be properly and efficiently applied. The fixed or dynamic frame distribution methods, formerly proposed, cannot fully utilize the aggregated links, where the receiving terminal only attaches to a pre-specified link among multiple physical links. A frame distribution method using tagging is proposed for the link aggregation between 10GbE switches to solve this problem. We compared the performance of the proposed method with those of the fixed and dynamic frame distribution methods. As a result, the proposed method shows a better performance when the applied load is below 0.7 and the average length of the frames is longer than 954 bytes.

* 인제대학교 광대역정보통신공학과 차세대통신망 연구실(2002b803@gurum.inje.ac.kr), ** 인제대학교 공과대학 전자정보통신공학부 (icshlee@inje.ac.kr), *** 한국전자통신연구원 액세스프로토콜팀 팀장(jhlee@etri.re.kr)
논문번호 : 030383-0903, 접수일자 : 2003년 9월 3일
※본 연구는 한국전자통신연구원 위탁과제로 수행되었습니다.

## Ⅰ. 서 론

인터넷 서비스의 급속한 확산으로 인해 초고속 통신망의 구축에 대한 사용자의 요구가 증대되고 있고 대도시의 주요 빌딩군으로 데이터 트래픽이 집중됨에 따라 MAN 구간의 병목 현상이 가중되고 있다. 또한, 초고속 인터넷 서비스 이용자들이 발생시키는 트래픽 양은 일반 인터넷 서비스 이용자들이 발생시키는 트래픽 양을 훨씬 초과하고 있어서 보다 큰 대역폭과 고속의 링크 제공이 요구되었다.

이를 위해 현재 망에서 기존의 링크를 사용하여 사용자가 원하는 대역폭을 얻기 위한 방법으로 802.3ad로 표준화된 링크 집합 기술을 적용할 수 있다. 이 기술은 여러 물리적인 링크들을 집합하여 하나의 논리적인 링크로 사용하는 기술이다. 이 기술은 10GbE으로 이루어진 망에서 대역폭의 증가가 필요할 때 새로운 링크의 추가 없이 기존의 링크를 결합함으로써 필요한 대역폭을 얻을 수 있다.

그림 1과 그림 2는 망을 구성하는 각각의 장치 사이에 적용된 집합된 링크와 링크 집합의 동작 처리 절차를 보여주고 있다[1][3][5]. 그림 1에서 볼 수 있듯이 링크 집합은 스위치와 스위치, 스위치와 서버, 서버와 서버 사이에서 하드웨어 업그레이드 없이 적용될 수 있다. 그림 2에서 물리적인 링크로의 프레임 분배는 분배 알고리즘에 의해 실행된다. 분배 알고리즘의 설계에 따라서 링크 집합 기능을 제공하는 스위치의 성능이 많이 달라질 수 있음을 알 수 있다. 따라서 효율적인 링크 집합을 위해서는 여러 개의 물리적인 링크에 프레임을 적절히 분배하는 분배 알고리즘을 설계하는 것이 중요하다. 분배를 위한 알고리즘은 표준안에 규정되어 있지 않다. 단, 아래의 사항들이 표준에 제시되어 있다[1].

- 해당 포트로 프레임이 전송되어질 때 프레임의 순서가 바뀌면 안 된다.
- 프레임의 중복이 발생하면 안 된다.

위 두 사항을 고려하여 분배 알고리즘을 설계하여야 한다. 이러한 분배 알고리즘으로는 기존에 제시된 스위치와 스위치 사이에서의 정적 / 동적 프레임 분배 방식이 있으나, 수신측에서 집합된 링크를 모두 사용하지 못하는 단점이 있다. 따라서 본 논문에서는 이 문제를 해결할 수 있는 프레임 분배방식을 제안하고자 한다.

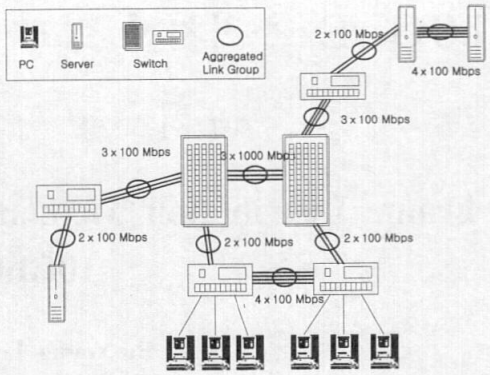



그림 1. 망 상에서의 링크 집합

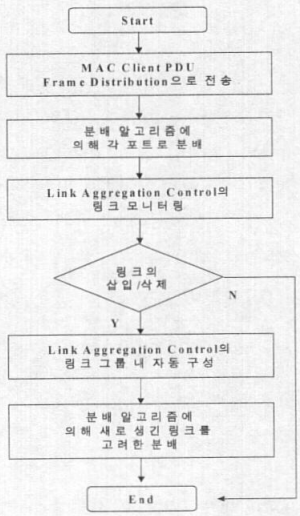



그림 2. 링크 집합의 동작 처리 절차

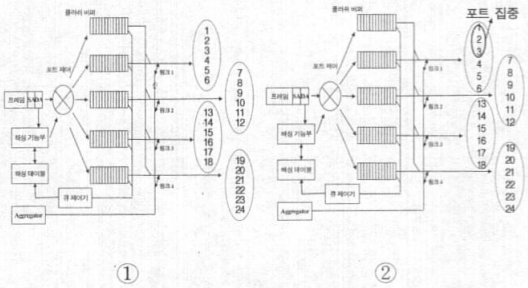
Ⅱ장에서 기존 프레임 분배 방식의 문제점을 제시하고 Ⅲ장에 이 문제점을 해결하기 위한 프레임 분배 방식을 제안하고 기존의 방식들과 비교한 후 Ⅳ장에서는 컴퓨터 시뮬레이션을 통해 기존의 분배방식과 패딩을 이용한 분배방식을 비교하여 성능을 분석하고 Ⅴ장에서 결론을 맺는다.

## Ⅱ. 기존 프레임 분배 방식의 문제점

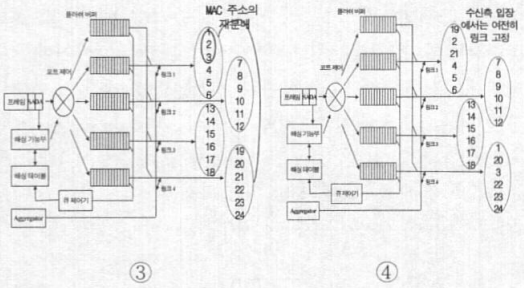
링크 집합은 여러 개의 물리적인 링크를 집합하였기 때문에 여러 단말들이 보내는 프레임들을 여러 개의 링크에 분배를 하여야 한다. 그런데 현재 표준안에는 특정한 프레임 분배 알고리즘에 과하여 규정을 하고 있지 않다. 이를 위해 이더넷 프레임들간에 순서의 불일치가 일어나지 않고 여러

개의 링크를 사용하는 장점을 활용하기 위해서 정적 프레임 분배 방식이 제안되었다[2]. 이 방식은 프레임들의 최종 수신 단말의 MAC주소에 따라 프레임을 각 링크로 분배하는 방식이다. 예를 들어 M개의 단말들이 N개의 링크로 집합된 스위치에 접속되어 있을 때, M / N개의 단말들을 하나의 호스트 그룹으로 묶여지고, 이 그룹에 속한 MAC 프레임들은 최종 수신 단말 측 주소에 따라 N개의 링크 중 고정된 하나의 링크를 통하여 프레임의 순서가 유지되면서 송신된다. 그러나 이 방식은 간단하기는 하나 수신 단의 MAC 주소에 따라 하나의 링크가 고정되기 때문에 특정 호스트로 프레임이 집중이 되면 링크의 효율이 떨어지고 집합된 링크를 모두 사용하지 못해 링크 집합의 장점을 살리지 못한다[2].

앞의 문제점을 해결하기 위해 동적 프레임 분배 방식이 제안되었다. 이 방식은 특정 링크로 집중되는 프레임들을 분산시키기 위해 기존의 링크를 동적으로 추가하는 것으로서, 기존의 링크 추가 시 프레임의 순서 불일치를 막기 위해 플러시 버퍼를 사용하고 있다. 동적 프레임 방식은 특정 링크에 MAC 프레임들이 집중되는 경우 해당 호스트 그룹에 대하여 N개의 출력 링크 중 추가로 출력 링크가 더 할당되도록 하는 방법이다. 이때, 추가되는 링크는 N-1개의 출력 링크 중 일정기간 동안 링크의 이용율이 가장 낮은 것이 선택되도록 한다. 하지만 이 방식은 시스템의 구현상의 복잡성이 있으며 부하가 증가함에 따라 집중되는 링크에 이용률이 낮은 링크를 추가하여 프레임을 재분배하는 동작이 계속해서 이루어지므로 오버헤드가 많이 발생하여 스위치의 성능 저하를 가져온다. 이 방식 또한 정적 프레임 분배 방식과 마찬가지로 특정 포트로 집중되는 프레임을 분산시킨다고 하더라도 수신측 입장에서는 집합된 링크를 모두 사용하지 못한다[2].




① ②




③ ④

그림 3. 동적 분배 방식의 문제점

그림 3은 동적 분배 방식을 사용하여도 수신측 입장에서는 여전히 링크 집합의 장점을 살리지 못 함을 보여주고 있다. 그림 3의 순서대로 이러한 동적 분배 방식의 문제점을 설명한다.

① 각 링크는 6개의 단말로 연결되어 있다
② 링크 1의 단말 중 MAC 주소 1,2,3번에 프레임의 집중이 생겨 링크 1에 부하가 많이 걸린다.
③ 링크의 이용율이 낮은 링크 4와 MAC 주소들을 재분배한다.
④ MAC 주소들을 재분배를 하여 링크 1의 부하를 줄일 수는 있지만 수신측 단말 입장에서는 여전히 4개의 링크 중 1개만 고정된다. 따라서 링크 집합의 이점을 충분히 활용하지 못한다.

## III. 정적 / 동적 분배 방식과 패딩을 이용한 방식의 비교

기존의 정적 / 동적 프레임 분배 방식은 수신측의 MAC 주소에 따라 링크를 고정하여 프레임의 순서 불일치를 해결한다. 그러나 이 분배 방식은 수신측 단말 입장에서는 링크가 고정되어 있기 때문에 링크 집합에 적용한 링크를 모두 사용하지 못해 링크기술의 장점을 살리지 못한다. 이를 해결하기 위해 링크 집합 기술의 장점을 살리면서 프레임의 순서 불일치도 일어나지 않도록 하기 위한 방법으로 패딩을 이용한 방식을 제안한다.

이 방식은 프레임을 여러 물리적인 링크로 보내기 전에 프레임을 최고 길이(1,500바이트)로 패딩을 하여 순차적으로 전송을 하므로 여러 개의 링크 중 어느 링크로 전송하더라도 프레임의 순서 불일치는 생기지 않는다. 또한, 수신측을 MAC 주소별로 링크를 고정 할 필요도 없어 링크를 모두 사용할 수

가 있고 구현 시 기존의 정적 / 동적 분배 방식보다 더 간단하다. 원래 이 방식은 단말과 단말 사이에 적용되던 기술로서[2] 스위치와 서버 또는 서버와 서버 사이에 소프트웨어에 의해서 구현해야 하는 경우에는 적합하지 않은 것으로 여겨졌다. 그러나 10GbE 스위치 등에서 네트워크 프로세서를 이용한 경우를 가정하면 이 방식의 적용이 가능해진다.

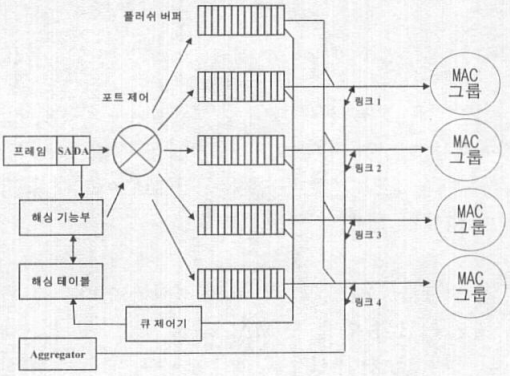



그림 4. 정적 / 동적 프레임 분배 방식의 예

그림 4는 4개의 링크에 각각 MAC 그룹이 정해져 있는 정적 / 동적 프레임 분배방식을 나타낸 것이다. 해싱 기능부는 해싱 테이블에 기록된 MAC 단말이 어느 링크로 전송되어야 되는지 판단하는 부분으로 동적 프레임 분배 방식에서는 한 개의 링크로만 분배되는 프레임들을 두 개의 링크 상에 재분배하여 해싱 테이블를 바꾸는 기능을 한다. 큐 제어기는 스위치의 출력버퍼에 임계값을 설정하고 집중되는 프레임에 의하여 이 임계값이 초과하면, 추가 링크 할당 동작이 개시 되도록 한다.

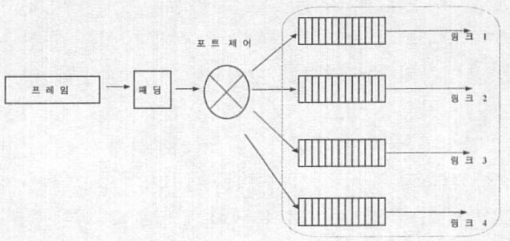



그림 5. 패딩을 이용한 분배방식의 예

그림 5는 패딩을 이용한 분배방식을 나타낸 것이다. 패딩을 이용한 분배방식은 각 링크로 프레임을 전송하기 전에 최고 길이(1,500바이트)로 패딩을 하고 있기 때문에 해싱 기능부, 해싱 테이블 및 큐 제어기가 필요가 없게 된다. 또한, 4개의 링크로 순차적으로 프레임을 분배하므로 프레임의 순서 불일치는 나타나지 않으며 기존 방식보다 훨씬 간단하다. 하지만 스위치에서 프레임을 최대 프레임 크기로 패딩 해야 하기 때문에 프레임의 크기가 작을 경우 링크의 효율이 떨어지는 단점이 있다.

## Ⅳ. 성능 분석

기존의 정적 / 동적 프레임 분배 방식과 제시한 패딩을 이용한 분배 방식을 컴퓨터 시뮬레이션을 통해 성능 분석을 하여 비교하였다.

시스템의 모델링은 송신측과 수신측 사이에 4개의 링크로 연결되어 있으며 각 링크는 프레임이 전송되는 동안 오류 없이 이상적으로 동작한다고 가정하였다. 각 프레임은 지수분포를 따르는 길이를 가지도록 발생시킨 다음 LAN 환경에서와 같이 프레임 형태를 얻기 위해 프레임의 최대 길이를 1,500 바이트, 프레임의 최소 길이를 64 바이트로 제한하였다. 입력 트래픽은 메트로 구간 등에서 다양한 유형의 서비스가 수용되는 경우를 고려하여 그림 6과 같은 2-state MMPP (Markov Modulated Poisson Process) 분포를 따른다고 가정하고, 시뮬레이션 시 전송되는 프레임 수는 800,000만개로 가정하였다. 스위치에서는 프레임들을 최대 프레임 길이로 패딩을 하고, 패딩된 프레임들은 4개의 링크에 1에서 4번 순으로 차례대로 분배된다고 가정하였다. 이때, 송신측 스위치의 각 링크별 출력 버퍼 크기는 4k 바이트와 8k 바이트 두 가지의 시스템을 가정하고, 시뮬레이션에서는 수신측이 각각의 링크에 무한버퍼를 갖는 시스템으로 가정하였다.

입력 트래픽으로 사용된 MMPP는 시간에 따라 도착률이 변하고, 도착 시간 간격 사이에 상관관계가 있는 특성을 이용해 보다 실제상황에 근접한 결과를 얻게 한다.

다음은 입력 트래픽으로 사용된 2-state MMPP에 가정된 조건이다[8].

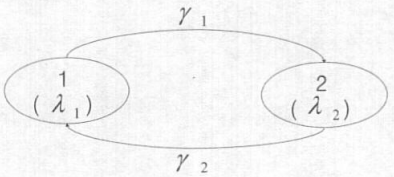



그림 6. 2-state MMPP

- 상태 1에 머무르는 시간 : 평균 1 / $\gamma_1$의 지수분포

- 상태 2에 머무르는 시간 : 평균 1 / $\gamma_2$의 지수분포
- 상태 1에서 고객 도착 분포 : 평균 비율 $\lambda_1$의 Poisson 과정
- 상태 2에서 고객 도착 분포 : 평균 비율 $\lambda_2$의 Poisson 과정

평균 프레임의 도착율은 다음의 식으로 구할 수 있다.

$$\lambda = \frac{\lambda_1/\gamma_1 + \lambda_2/\gamma_2}{1/\gamma_1 + 1/\gamma_2} = \frac{\lambda_1\gamma_2 + \lambda_2\gamma_1}{\gamma_1 + \gamma_2} \quad (1)$$

입력 트래픽에 사용된 2-state MMPP의 파라미터는 다음과 같다.

표 1. 2-state MMPP 파라미터

| $1/\lambda_1$ | $1/\lambda_2$ | $1/\gamma_1$ | $1/\gamma_2$ |
|---|---|---|---|
| 5 | 10 | 20 | 25 |

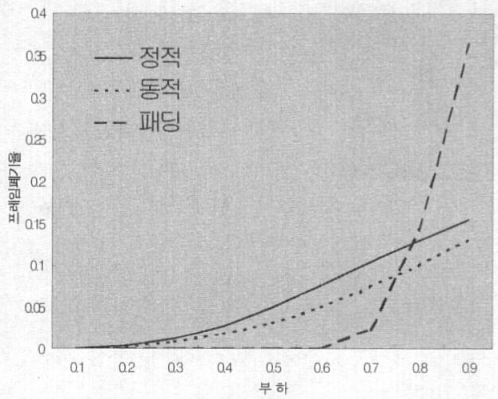



그림 7. 정적 / 동적 프레임 분배 방식과 패딩 방식의 프레임 폐기율 (버퍼크기 4 k바이트, 평균 프레임 길이 1,086바이트)

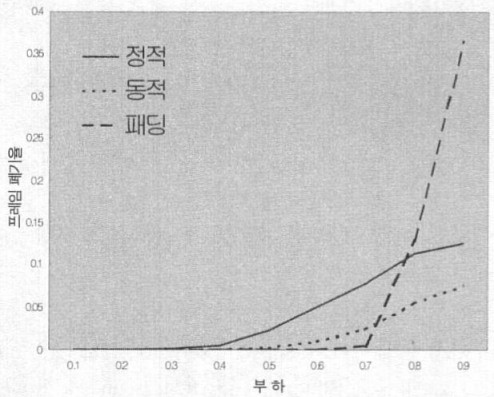



그림 8. 정적 / 동적 프레임 분배 방식과 패딩 방식의 프레임 폐기율 (버퍼크기 8 k바이트, 평균 프레임 길이 1,086바이트)

그림 7과 그림 8은 정적 / 동적 프레임 분배 방식과 패딩을 이용한 분배 방식의 프레임 폐기율을 나타내고 있다. 패딩을 이용한 분배 방식은 버퍼 크기 8 k이고 부하가 0.7일 때가 버퍼 크기 4 k일때보다 더 낮은 프레임 폐기율을 나타내고 다른 구간에서는 거의 같은 것을 알 수 있다. 버퍼 크기가 4 k일때는 부하가 0.75, 버퍼 크기가 8 k일때는 부하가 0.72일 때까지 정적 / 동적 분배 방식보다 프레임 폐기율이 작은 것을 볼 수 있으나 부하가 0.75, 0.72이상부터는 정적 / 동적 분배 방식보다 프레임 폐기율이 커짐을 볼 수 있다. 부하가 증가할수록 출력 버퍼내의 프레임들이 패딩에 의한 오버헤드가 급속도로 증가하기 때문에 제안한 패딩 방식의 폐기율이 정적 / 동적 분배 방식보다 커지게 된다.

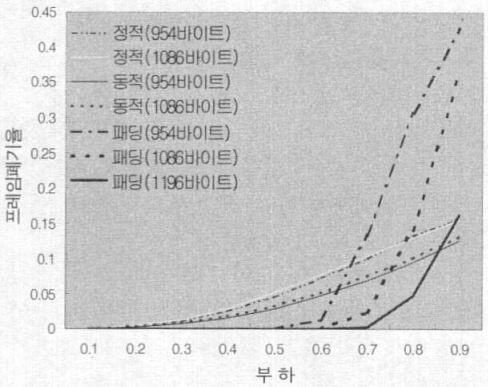



그림 9. 평균 프레임 크기에 따른 프레임 폐기율 (버퍼크기 4 k바이트)

그림 9는 정적 및 동적 분배 방식과 패딩 분배 방식을 평균 프레임 크기 954, 1,086, 1,196 바이트로 달리 하여 프레임 폐기율을 알아보았다. 정적 / 동적 분배 방식은 프레임의 크기에 따른 프레임 폐기율의 변화가 적다. 패딩을 이용한 분배 방식은 평균 프레임 크기가 954 바이트일 때 부하가 0.63까지는 프레임 폐기율이 작으나 그 이상일 때는 정적 / 동적 분배 방식보다 많아지는 것을 볼 수 있고 1,086 바이트에서는 부하가 0.75이상일 때 정적 / 동적 분배 방식보다 프레임 폐기율이 많아진다. 또한, 평균 프레임 길이가 1,196 바이트에서는 0.85이상일 때 정적 / 동적 분배 방식보다 프레임 폐기율이 많아지는 것을 알 수 있다. 이는 같은 전송시간에서 프레임 길이가 크면 출력 버퍼에서 출력되는 데이터 크기가 커지므로 출력 버퍼에서 받아들일 수 있는 데이터 양이 많아지기 때문이다. 이를 볼 때 평균 프레임 크기가 커질수록 정적 / 동적 프레임 분배 방식보다는 패딩을 이용한 분배 방식의 성능이 더 효율적임을 알 수 있다.

## V. 결 론

링크 집합의 구현에 중요한 역할을 하는 프레임 분배방식 중 기존의 정적 / 동적 프레임 분배 방식과 새로이 제안한 패딩을 이용한 분배 방식을 비교하여 성능을 분석하였다. 기존의 분배 방식들은 수신단의 MAC 주소에 따라 단말들이 1개의 링크로 고정되어 있어 집합된 링크를 모두 사용하지 못해 링크 집합의 장점을 살리지 못하는 단점이 있었다. 이 단점을 보완하고 프레임 순서 불일치도 일어나지 않는 패딩을 이용한 분배 방식은 링크 집합의 장점을 살릴 수 있고 구현 또한 간단해짐을 알 수 있었다. 또한, 컴퓨터 시뮬레이션을 통해 버퍼 크기에 따른 프레임 폐기율을 확인하였고 평균 프레임 크기에 따른 프레임 폐기율을 비교 분석하였다.

패딩을 이용한 분배 방식은 프레임의 크기가 작은 경우에는 기존의 방식에 비해 전송 효율이 낮은 문제점이 있지만 프레임의 크기가 클 경우에는 전송 효율이 훨씬 좋아지는 것을 알 수 있었다. 이는 프레임 길이가 길어지는 파일 전송 등의 경우에는 충분히 좋은 성능을 낼 수 있음을 말해준다.

앞으로 구현에 더 효율적인 분배 알고리즘이 되기 위해서는 링크의 추가 및 삭제 등 여러 경우에 대한 구체적 방안 등에 대한 연구가 추가로 진행되어야 한다.

## 참 고 문 헌


[1] IEEE, *IEEE Standard 802.3*, 2000 Edition

[2] 전우정, 윤종호, "통합링크기능을 가진 매체 접근제어기용 프레임 분배 방식의 성능 분석", *한국통신학회지*, 2000. 3.

[3] Link Aggregation according to IEEE 802.3ad White Paper, 2002

[4] Richard Foote, *"Link Aggregation 802.3ad"*, Corporate Systems Engineering, June 22, 20 01

[5] Walter Thirion, " *Link Aggregation Operations"*, jato Technologies, Inc., July 1998

[6] IEEE 802.3ad Link Aggregation Task Force Public archive area, http://grouper.ieee.org/groups/802/3/ad/public/

[7] Solving Server Bottlenecks Link Aggregation/FEC/GEC, http://www.pentium. co.kr/network/connectivity/solutions/server_bottlenecks/bot_sol2.htm

[8] Queueing systers Volume1:Theory, Leonard KIeinrock wiley



이 호 영(Ho-Young Lee) 준회원

2002년 2월 : 인제대학교 정보통신공학과 학사

2002년 3월~현재 : 인제대학교 광대역정보통신공학과 석사과정

<주관심분야> Ehternet 시스템, 차세대 인터넷 기술

이 숭 희(Soong-Hee Lee) 정회원

1987년 2월 : 경북대학교 전자공학과 학사

1990년 2월 : 경북대학교 전자공학과 대학원 석사

1995년 2월 : 경북대학교 전자공학과 대학원 박사

1987년~1996년 : 한국전자통신연구원 선임 연구원

1997년 3월~현재 : 인제대학교 전자정보통신공학부 조교수

<주관심분야> 초고속 통신망, NGcN, 통신 시스템

이 종 협(Jong-Hyup Lee) 정회원

1984년 : 고려대학교 산업공학과 학사

1986년 : 한국과학기술원(KAIST) 산업공학과 석사

1996년 : 한국과학기술원(KAIST) 산업공학과 박사

1986년~현재 : 한국전자통신연구원 책임연구원, 액세스프로토콜팀 팀장

<주관심분야> High-spped Network Design and Routing Switch and Router Technology, Network Protocols