논문 2012-49-11-4

동적 레인 제어방식을 적용한 에너지 절감형 광 이더넷 시스템의 성능분석

(Performance of Energy Efficient Optical Ethernet Systems with a Dynamic Lane Control Scheme)

서 인 수*, 양 충 열**, 윤 종 호*

(Insoo Seo, Choong-reol Yang, and Chongho Yoon)

요 약

본 논문에서는 광 이더넷 시스템에 대하여 상용 광 트랜시버 모듈의 사용이 가능하면서도 에너지 절감기능을 제공할 수 있도록 트래픽 예측모듈을 사용하는 동적 레인제어방식을 제안한다. 40/100Gbps급 상용 광 트랜시버는 4개 또는 10개의 광 트랜시버를 사용하는데 이들 각각은 트래픽 부하와 상관없이 항상 켜져 있어 많은 에너지를 소모한다. 이러한 에너지 소모를 감소시키기 위하여 제안된 동적 레인제어방식은 부하에 따라 일부 레인의 트랜시버를 끄고 나머지 활성화된 레인으로만 프레임을 처리하도록 한다. 이때 레인의 갯수가 변동될 때 발생할 수 있는 바이트 전송순서 어긋남을 보완하기 위하여 새로운 전송율 제어모듈을 xGMII 인터페이스 상위에 위치한 정합부계층에 설치하는 것을 제안하였다. 이것은 비활성화된 레인상으로 가상적인 바이트열을 삽입하는 기능을 수행하는 것으로써 이 바이트열들은 비활성화된 PMD에서 무시된다. 실제 이 모듈의 구현은 PHY모듈과 별개로 동작하므로 상용 PHY모듈의 사용이 가능한 장점을 제공한다. 이러한 시스템에서 변동되는 부하에 적응하여 활성화된 레인의 갯수를 결정하는 것이 중요하므로 구현관점에서 용이한 트래픽 예측기를 제시하였다. 이것은 주기적으로 샘플링된 현재의 송신버퍼크기와 지금까지 사용되었던 버퍼크기 예측값에 서로 다른 가중치를 부여하여 변화하는 트래픽에 적응하도록 한다. 이러한 시스템에 대하여 OMNET++기반의 시뮬레이터를 구현하여 적응정도와 에너지 절감효과를 분석하였다.

Abstract

In this paper, we propose a dynamic lane control scheme with a traffic predictor module and a rate controller for reconciling with commercial optical PHY modules in energy efficient optical Ethernet systems. The commercial high speed optical Ethernet system capable of 40/100Gbps employs 4 or 10 multiple optical transceivers over WDM or multiple optical links. Each of the transceivers is always turned on even if the link is idle. To save energy, we propose the dynamic lane control scheme. It allows that several links may be entirely turned off in a low traffic load and frames are handled on the remaining active links. To preserve the byte order even if the number of active links may be changed, we propose a rate controller to be sat on the reconciliation sublayer. The main role of the controller is to insert null byte streams into the xGMII of inactive lanes. For the PHY module, the null input streams corresponding to inactive lanes will be disregarded on inactive PMDs. It is very handy to implement the rate controller module with MAC in FPGA without any modification of commercial PHYs. It is very crucial to determine the number of active links based on the fluctuated traffic load, we provide a simple traffic predictor based on both the current transmission buffer size and the past one with different weighting factors for adapting to the traffic load fluctuation. Using the OMNET++ simulation framework, we provide several performance results in terms of the energy consumption.

Keywords: Ethernet, Energy, Traffic, PHY, Simulator

* 정회원, 한국항공대학교 항공전자정보통신공학부
(Department of Information and Telecommunication Engineering, Korea Aerospace University)
** 정회원, 한국전자통신연구원
(Electronics and Telecommunications Research Institute)
※ 본 연구는 MKE/KEIT의 ETRI연구개발지원사업의 일환으로 수행하였음.
[KI001911, 100Gbps급 이더넷 및 광전송 기술 개발
접수일자: 2012년1월27일, 수정완료일: 2012년10월29일

Ⅰ. 서 론

최근 10Gbps급의 초고속 이더넷 시스템이 활발히 도입되고 있으나, IPTV 등의 보급으로 인해 더욱 큰 대역폭을 제공할 수 있는 100Gbps급 이더넷 기술이 요구되고 있다. 이러한 요구에 따라 2010년 6월에 40Gbps 이더넷과 100Gbps 이더넷의 표준화를 완료하였다[1]. 이러한 40/100Gbps 이더넷은 4개 또는 10개의 트랜시버를 사용한 다중레인 기술을 사용한다[2].

한편 이더넷시스템의 소모전력을 절약하기 위하여 이더넷 기능부의 동작을 sleep과 normal 상태를 천이하면서 전력소모를 감소시키는 IEEE802.3az EEE(Energy Efficient Ethernet)표준화도 완료되었다[3~4]. 이러한 EEE기술은 모두 구리선 기반의 단일 링크를 사용하는 이더넷에만 적용된다.

최근에는 EEE기술을 4개 또는 10개의 광선로를 동시에 사용하는 40/100Gbps 이더넷 시스템에 적용하려는 연구가 진행 중이다. 이러한 광 이더넷 시스템에 대한 소모전력 절약 방법으로 고려할 수 있는 방법은 다음과 같다.

- 모든 레인에 대한 LPI 적용 방법
- 일부 레인을 OFF 시키는 방법

첫 번째 방법은 기존 단일 레인용 EEE기법을 각 레인마다 적용하는 방법이다. 이 방법은 Multi Lane Distribution(MLD)기능을 수행하는 Physical Coding Sublayer(PCS) 및 비트다중화(BitMux)기능을 수행하는 Physical Medium Access(PMA)부계층의 변경 없이 모든 Physical Medium Dependent(PMD) 레인으로 Sleep 및 Refresh신호를 보낼 수 있는 장점이 있다. 하지만 PMD에 장착된 레이저 다이오드(LD)의 큰 ON/OFF지연시간을 고려한다면 Off에서 On이 되어 정상상태가 되는 기간에도 여전히 전력이 소모되므로 이러한 잦은 on/off는 프레임 전송효율만 감소시키는 반면에 에너지 절약의 이점을 살릴 수 없는 문제점이 있다.

두 번째 방법은 특정 링크를 sleep하는 대신에 완전히 Off시키는 방법이다. 이 방법은 일부 레인이 Off 되기 때문에 Off된 레인에서는 Refresh메시지가 전송되지 않으므로 전력이 소모되지 않는다. 이렇게 함으로서 모든 레인에 LPI방식이 사용되는 것 보다 특정 레인을 ON 또는 OFF시키는 것이 에너지 절감 및 성능에 유리하다. 물론 이 경우에도 MLD를 수행하는 PCS에서는 sleep하는 레인에 대해서는 바이트열을 분배 시 할당하지 않아야 하며, PMA에서도 bitMux할 때 해당 레인에 대한 다중화기능을 수행하지 않아야 비트순서가 지켜질 수 있는 문제점은 여전히 존재한다.

본 논문에서는 상기 2가지 방법 중에서 2번째 방법을 선택하고 이 방법의 효율적인 에너지 절약 기능을 제공하기 위하여 새로운 Rate Controller모듈과 트래픽 예측모듈을 제안하였다.

먼저, 다중 레인을 사용하는 광 이더넷의 경우 레인 갯수의 변화에 따라 유효 전송율이 변동할 때 전송중인 바이트열의 순서가 유지될 수 있는 기능이 필요하다. 본 논문에서는 이를 위해 새로운 Rate Controller기능부를 정합부계층(Reconciliation sublayer)에 설치하는 방법을 제안하였다. 제안된 RateController는 MAC으로부터 전달되는 바이트열을 4개의 물리계층 레인으로 분배할 때 비활성 레인에 대하여 Null바이트열을 생성하여 물리계층으로 전달함으로써 PCS, PMA계층의 기능을 수정하지 않고도 유효전송률을 제공할 수 있는 장점을 제공한다. 또한 순간적인 데이터 유입에 대처가 가능하여 버퍼링 없이 데이터 송/수신이 가능하다.

또한 효율적인 멀티레인 운용을 위해서는 트래픽 부하를 예측하여 레인을 미리 ON/OFF하여 에너지를 절감하는 기능을 제공하는 트래픽 예측모듈을 제안하였다.

이것은 트래픽 예측을 위해 주기적으로 측정된 자신의 송신 버퍼크기 값과 직전까지 예측된 버퍼크기에 대하여 서로 상이한 가중치를 부여하여 이후의 버퍼크기를 예측하는 방법이다. 이어 만약 예측된 버퍼의 크기가 미리 설정된 임계값을 초과하면 상대방에게 레인추가메시지로 통보하여 트랜시버의 ON지연시간인 100msec 이후에는 새로운 레인이 추가 동작되도록 한다. 물론 임계값 이하로 떨어지면 레인 개수를 감소하도록 한다. 제안된 이 방식의 성능 평가를 위해 OMNeT++ 기반의 시뮬레이터를 구현하여 버퍼사이즈와 Threshold, 부하에 따른 트래픽 예측기의 동작여부와 패킷 손실율 및 전력절감효과를 분석하였다.

본 논문의 구성은 다음과 같다. 본 서론에 이어, 제Ⅱ장에서는 IEEE802.3az EEE 기술에서 사용되는 LPI 방법과 40/100Gbps급 이더넷 시스템의 구성을 분석한

다. 제 Ⅲ 장에서는 멀티레인 기반의 40/100Gbps급 광 이더넷에 대한 EEE 적용을 위한 방안을 제시하고 구현 가능성을 보장하는 Rate Controller기능을 제안한다. 제 Ⅳ 장에서는 동적으로 레인의 개수를 결정할 수 있는 트래픽 예측기를 제안하고 이의 동작 검증을 위한 OMNET++시뮬레이터를 구현하여 다양한 가중치를 적용하여 성능을 분석하였다. 마지막으로 제 Ⅵ장에서는 결론을 맺는다.

Ⅱ. EEE 시스템과 40/100Gbps 이더넷

1. Energy Efficient Ethernet 기술

IEEE802.3az의 EEE기술은 새로운 lower voltage mode를 사용하여 10BaseT 의 전력소모를 감소시킬 수 있는 10Base-Te 뿐만 아니라, 100Base-TX, 1000Base-T, XGXS(XAUI), 10Gbase-T, 10GBase-KX4, 10GBase-KR 등의 이더넷에 대하여 Low Power Idle(LPI)기법을 사용한 에너지 절감 기술에 대한 표준이다[5].

IEEE 802.3az 표준의 대부분은 LPI기법을 사용하여 이더넷 인터페이스 상에 송수신할 데이터가 없을 시에는 low power idle상태가 되어 MAC+PHY의 송신부 또는(및) 수신부를 Sleep모드로 진입시켜 전력을 절약하는 것이다. 이러한 절차는 모두 idle구간에서 수행된다. 그 절차는 다음과 같다.

- Initiator는 자신의 데이터 전송율이 일정수준 이하가 되면 마지막 데이터프레임 송신완료 즉시 특별한 SleepCode열을 전송하여 자신이 Quiet상태로 진입함을 peer에게 알린다. 이후 자신은 프레임 송신이 중지되고 자신의 MAC, 메모리, CPU 등 일부는 전원을 차단할 수 있다.
- 이에 Peer도 Quiet모드로 진입할 수 있다. 물론 Peer의 수신부도 전원절약모드로 진입할 수 있지만 RefreshCode나 WakeCode가 언제 도착할지 모르므로 Sleep모드로 진입하는 것은 추가적인 고려가 필요하다. 이를 해결하려면 LPI상태 시간이 어느 정도 지났을 때 자발적으로 데이터를 수신할 준비를 해야 한다.
- 이후 Initiator는 상대방 PHY레벨의 retiming을 위하여 Tq간격 (Quiet간격=20msec)마다 주기적으로 RefreshCode열을 전송한다.
- 만약 Initiator에 전송할 데이터가 일정 수준 버퍼링되면 상대방을 깨우기기 위해 WakeCode를 먼저 전송한 후 이어 실제 데이터를 송신한다. Peer는 WakeCode를 수신시 자신의 수신부를 Idle모드로 진입하여 이어 도착할 프레임의 수신을 대기한다.

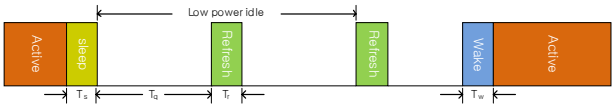


그림 1. LPI의 기본동작
Fig. 1. Operation of LPI.

2. 40/100Gbps급 이더넷 기술

IEEE 802.3ba규격은 40Gbps 및 100Gbps급 이더넷 기술의 표준이다. 이것은 4 또는 10개의 10Gbps 또는 25Gbps급 레인을 4개 또는 10개의 MMF를 사용하거나 1개의 SMF에 WDM하여 송신하는 것이다.

40GBase-R 및 100GBase-R이더넷의 RS는 MAC과 PCS간의 신호매핑을 수행하며, PCS는 xGMII로부터의 64비트폭의 데이터와 8비트폭의 제어신호를 기반으로 64B/66B인코딩을 수행하여 PMA와의 4개(40Gbps) 또는 20개(100Gbps)의 직렬인터페이스로 전달한다.

PMA는 PCS와의 4 또는 20개의 레인으로 구성된 전기적인 인터페이스를 4개 또는 10개의 PMD 광 레인으로 대응시키는 기능을 한다. 특징적으로 구리선 케이블 또는 백플레인용으로 FEC가 선택사양으로 추가되어 전달거리를 확장시킬 수 있도록 하였다.

40/100Gbps 이더넷에서는 여러 개의 물리적인 링크 4개 또는 10개를 동시에 사용하는 PMD를 지원하기 위해 PCS에 MLD(Multi Lane Distribution) 기술을 사용한다. MLD는 40/100Gbps 이더넷을 한 개의 광 선로로 이용할 수 없는 기술적 한계를 해결하는 기능이다. MLD의 가장 중요한 개념은 가상레인이다. 이것은 40/100G MAC프레임을 64B/66B 블록단위로 블록화 하여 각 블록을 n(4 또는 20)개의 가상레인에 Round Robin 방식으로 할당하여 송신하고, Rx MLD는 수신한 데이터를 정렬하여 원래의 MAC 프레임 형태로 복원하는 기능이다.

이를 정리하면 <그림 2>와 같다. Multi-Lane

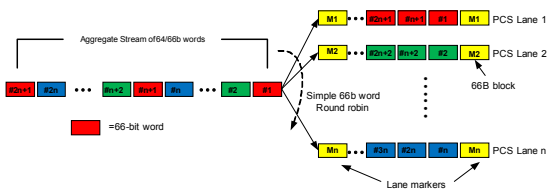


그림 2. PCS MLD의 기능
Fig. 2. Operation of PCS MLD.

Distribution기능에 의해 Data는 “n”개의 PCS 레인을 통하여 n개의 66 비트블록을 동시에 전송한다. 즉 일단 데이터가 인코딩된 후 스크램블되어 여러 개의 PCS레인에 분배되어 66비트블록으로 동시에 round robin방식으로 0번부터 높은 번호의 PCS레인으로 다음과 같이 분배 송신한다. 참고로 그림에는 주기적으로 삽입되는 deskew용 Alignment Marker블록도 함께 도시되었다.

- 40GBASE-R PCS의 경우에는 66-bit블록을 4개의 PCS레인으로 분배한다.
- 100GBASE-R PCS 는 이 블록을 20개의 PCS lane으로 분배한다.

수신시에는 PCS RX단에서 다시 정렬하여 원래의 MAC 프레임 형태로 복원한다. 하지만 각 레인은 각각 다른 경로를 갖기 때문에 스큐가 발생할 수 있다. 스큐 보상을 위해 A(alignment Marker)블록과 같은 특별한 블록을 사용하여 TX단에서는 주기적으로 삽입하여 보내주고, RX단에서는 이것을 삭제하여 스큐를 보상한다. 이러한 Alignment Marker는 제어블록 중 하나인 66비트블록이며, 모든 레인에 동시에 삽입되어 송신된다. 이 Marker는 IPG구간에 송신되며 이것은 스크램블링 되지는 않는다. 이렇게 스크램블링 하지 않고 송신하므로 수신측에서는 PCS레인별 skew를 조정한 후 재조립하여 데이터에 대한 디스크램블링을 수행해야 한다. 이러한 마커는 16383개의 66비트 블록을 송신한 즉시 이어 모든 레인에 동시에 삽입된다. 이러한 마커의 값은 각 레인별로 일정수의 변위를 보장하는 평형코드값을 가지는 값으로 미리 지정되어 있어 굳이 스크램블링을 하지 않도록 한다.

이러한 분배방식은 MAC프레임별로 활성화된 링크에 분배하는 802.3ad Link Aggregation기능과 달리, EEE에서는 활성화된 링크에 대하여 비트열이 분배되는 차이점이 있다.

참고로 CFP는 40/100Gbpsbps Form Factor Pluggable 패키지로서 산업체표준규격이다. CFP MSA 40G 광모듈은 10Gbps x 4채널로 동작하는데, 각 레인별 제어는 LD의 ON/OFF기능만 제공된다. 상용화된 40GBase-LR4용 PMD는 10.315Gbps x 4 lane CWDM PMD로써 XLAUI(4x10G) 전기인터페이스에 의해 PMA/PCS에 연결된다. 이것의 소모전력은 8W로써 각 레인별로는 2W이다.

Ⅲ. 멀티레인 광 이더넷 시스템의 구현을 위한 Rate Controller의 설계

1. 40/100Gbps급 광이더넷에 대한 EEE기술 적용 방법

(가) 모든 레인에 대한 LPI 적용 방법

이것은 기존 단일 레인용 EEE기법을 각 레인마다 적용하는 방법이다. 이 방법은 MLD를 수행하는 PCS 및 비트다중화기능을 수행하는 PMA와 상관없이 모든 PMD 레인으로 Sleep 및 Refresh신호를 보낼 수 있는 장점이 있다. 하지만 LD의 ON/OFF지연시간이 큰 경우 비 효율적이다. 즉 CFP규격에는 On/Off지연시간의 최대값은 100msec이다. Off에서 On이 되어 정상상태가 되는 이러한 100msec 동안에도 여전히 전력이 소모되므로 잦은 on/off는 프레임 전송효율만 감소시키는 반면에 에너지 절약의 이점을 살릴 수 없다.

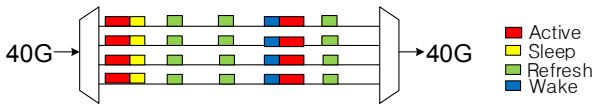


그림 3. 모든 레인에 대한 LPI적용방법
Fig. 3. LPI over all lanes scheme.

(나) 일부 레인을 OFF 시키는 방법

일부 레인의 동작을 sleep시키는 방법과 유사하지만 특정 링크를 sleep하는 대신에 off시키는 방법으로 본 논문에서 제안하는 방법이다. 이 방법은 일부 레인이 Off 되기 때문에 Off된 레인에서는 Refresh메시지가 전송되지 않는다. 즉 필요에 따라서 나머지 3개의 레인을 모두 off시키고, 필요시 활성화된 링크를 통하여 필요한

레인을 On시킴으로써 추가적으로 전력을 절약할 수 있다. 이렇게 함으로서 모든 레인에 LPI방식이 사용되는 것 보다 특정 레인을 On 또는 Off시키는 것이 에너지

그림 4. 일부 레인을 OFF시키는 방법
Fig. 4. Partial lane off scheme.

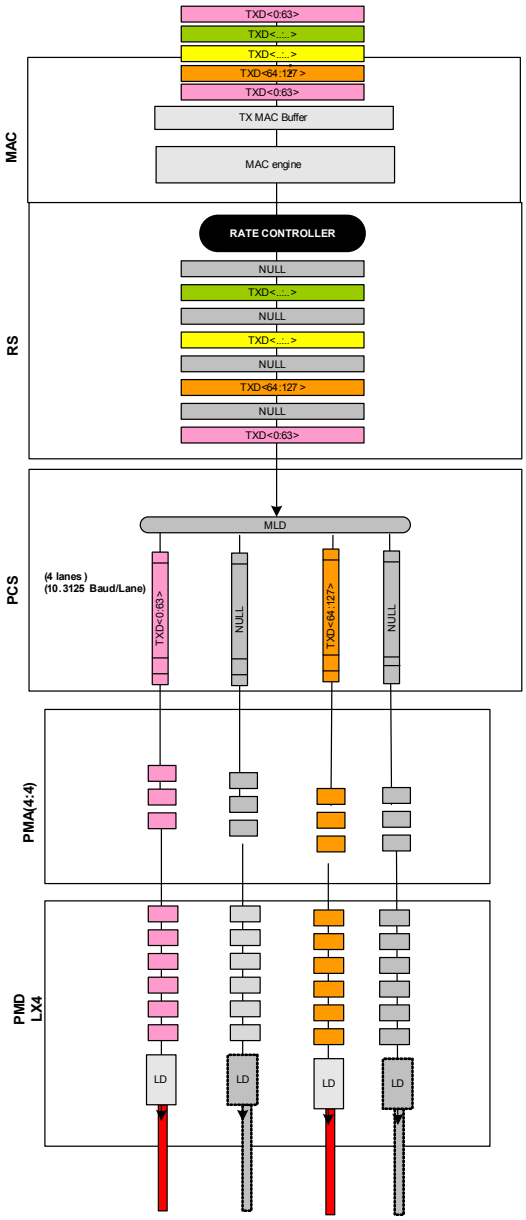


그림 5. Rate Controller가 추가된 멀티레인 이더넷 시스템의 동작
Fig. 5. Multilane Ethernet system with rate controller module.

절감 및 성능에 유리하다. 마찬가지로 이 경우 MLD를 수행하는 PCS에서도 Sleep하는 레인에 대해서는 분배열에서 제거해야 하며, PMA에서도 비트다중화시 해당 레인에 대한 다중화 기능을 수행하지 않아야 비트순서가 지켜질 수 있다.

2. Rate Controller

일부 PMD 레인의 동작을 sleep시키거나 off시키는 방법을 사용할 경우 MLD를 수행하는 PCS에서도 Sleep하는 레인에 대해서는 분배하지 않아야 한다. 또한 PMA에서도 비트다중화시 해당 레인에 대한 다중화를 수행하지 않아야 비트순서가 지켜질 수 있다.

이를 위하여 PCS나 PMA의 기능을 일부 수정하면 EEE방식을 적용할 수 있지만, PCS 및 PMA의 기능 수정시 xGMII로부터 유입되는 순간적인 데이터를 처리하기 위해서는 큰 용량의 버퍼링이 PHY계층에 필요하게 되고 이는 지연을 초래한다.

또한 이러한 물리계층에서의 버퍼링은 일반적인 이더넷 규격에 어긋난다. 본 논문에서는 이러한 문제점을 해결하기 위해 MAC/RS계층에 Rate Controller를 추가하는 것을 제안한다. 제안된 Rate Controller는 <그림

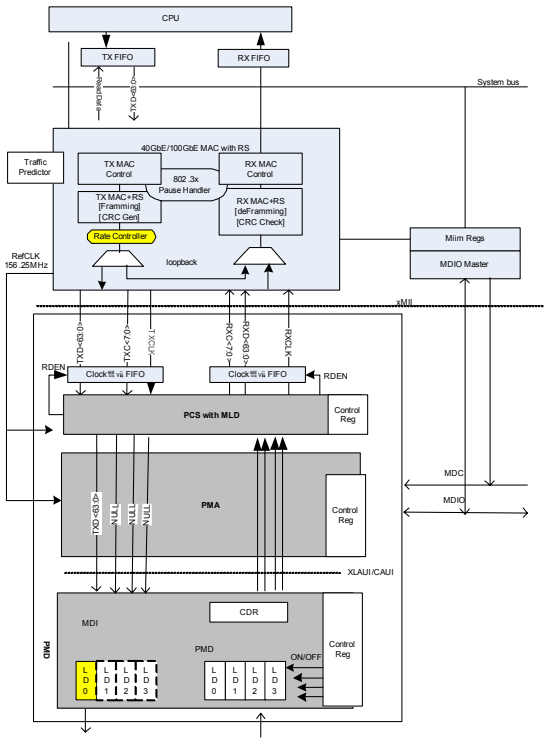


그림 6. Rate Controller의 하드웨어 구성
Fig. 6. Rate controller Module.

5>와 같이 비활성 레인의 개수에 부합하여 Null바이트를 삽입하여 전송함으로써 PCS, PMA계층의 기능 수정 없이 유효전송률을 제공할 수 있다. 또한, 순간적인 데이터 유입에 대처가 가능하여 버퍼링 없이 데이터 송/수신이 가능하다. 상용 PHY모듈을 그대로 사용할 수 있고, 구현상 용이한 장점이 있다.

<그림 6>은 제안하는 Rate Controller가 표함된 40Gbps급 이더넷 시스템의 구조를 도시한 것이다. 이것은 비활성화된 레인에 대하여 NULL바이트열을 삽입함으로써 용이하게 유효 bitrate를 조절하는 기능을 제공한다. 상용 CFP모듈을 사용할 경우 해당 LD의 ON/OFF설정은 MDIO에 의한 레지스터 설정에 의해 수행될 수 있다.

Ⅳ. 동적 멀티레인제어용 트래픽 예측기

1. 개요

본 장에서는 40/100Gbps급 광이더넷 시스템에서의 효율적인 멀티레인 운용을 위해 송신버퍼의 길이를 예측하여 활성화된 레인갯수를 동적으로 제어하는 방법을 제안하고, OMNet++기반의 시뮬레이터를 구현하여 성능을 분석한다.

2. 트래픽 예측 알고리듬

본 논문에서 사용한 트래픽 예측 알고리듬은 다음과 같다.

$$Q(i+1) = \alpha Q(i) + (1-\alpha) Q_M(i) \quad (1)$$

$$Q_p = \beta\, Q(i+1) \quad (2)$$

where

$Q(i)$: i번째 예측된 버퍼크기

$QM(i)$: i번째 실제 측정된 버퍼크기

$Q(i+1)$: i+1번째 예측 버퍼크기

Q_p : Margin이 곱해진 예측 버퍼 크기

α : 가중치

β : margin을 위한 가중치

이것은 주기적으로 시행하는 샘플링 시점에서 측정된 송신버퍼 크기($Q_M(i)$)에 대한 가중치를 (1-α)로 부여하는 반면에 지금까지 예측하여 사용하고 있던 버퍼크기($Q(i)$)에 가중치 α를 부여한 결과값 $Q(i+1)$을 다음에 사용할 예측값으로 결정한다. 이때 가중치 α가 0.5보다 크다면 지금까지 사용하고 있던 예측값에 가중치를 더 부여하는 반면 지금 측정된 버퍼크기값에는 작은 가중치가 부여된다. 이 경우 과거의 history에 큰 가중치를 두고 방금 측정된 값에는 작은 가중치를 두게 되는 일종의 low-pass filter역할을 수행한다.

그리고 실제 예측된 버퍼의 크기 Q_p 는 앞에서 예측된 버퍼크기에 일종의 margin factor인 β 를 곱하여 얻어진다. 여기서 β 는 1보다 큰 값을 사용하여 혹시나 있을 수 있는 급격한 유입량의 변화에 대처할 수 있도록 margin을 더 부여하는 값이다.

실제 예측된 버퍼의 크기 Q_p 가 미리 설정된 T_H 값을 초과하는 경우 상대방에게 레인추가를 통보한다. 반면에 미리 설정된 T_L 값 미만으로 감소하는 경우에는 레인감소를 통보하는 방법을 사용하여 트래픽의 증감 추세에 따라 레인의 활성화 유무를 결정한다. 물론 T_H 와 T_L 사이인 경우 현재 레인의 개수를 그대로 유지하도록 하였다.

참고로 제시된 예측식은 인터넷용 프로토콜인 TCP에서 사용하고 있는 RoundTripTime 예측방식을 참조하여 큐 크기 예측기법으로 활용한 것이다.

3. 시뮬레이터의 구성

본 논문에서 제시한 트래픽 예측 알고리즘의 성능 분석을 위하여 OMNet++기반의 시뮬레이터를 구현하였다. 본 시뮬레이터는 EEE가 적용된 멀티레인 환경의 광 이더넷 시스템을 모사한 것으로써 여기에는 트래픽 예측기가 포함된 것이다.

(가) 시스템 모델링

시뮬레이션에 적용된 네트워크 모델은 <그림 7>과 같이 2개의 광 이더넷 시스템(HOST)로 구성된다. 여기서 점선으로 표시되는 링크는 두 개의 energy efficient ethernet 기능을 지원하는 호스트를 연결하는 4개의 병렬 링크를 의미한다. 그리고 flatNetworkConfigurator는 각 host의 상위 IP계층 주소를 자동 설정해 주는 기능이다.

<그림 8>은 각 호스트의 내부구조이다. 각 호스트는 Poisson분포로 생성되는 125바이트 길이의 전달계층 패

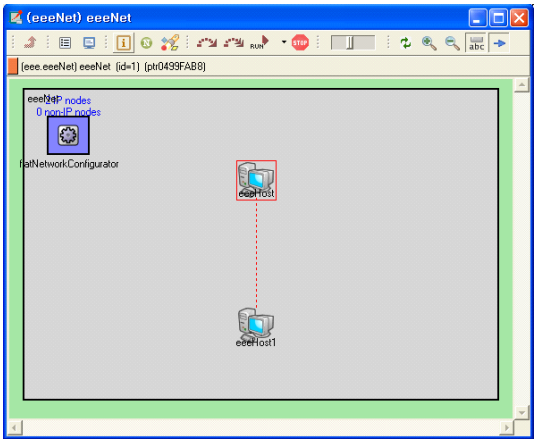


그림 7. 네트워크 모델
Fig. 7. Network Model.

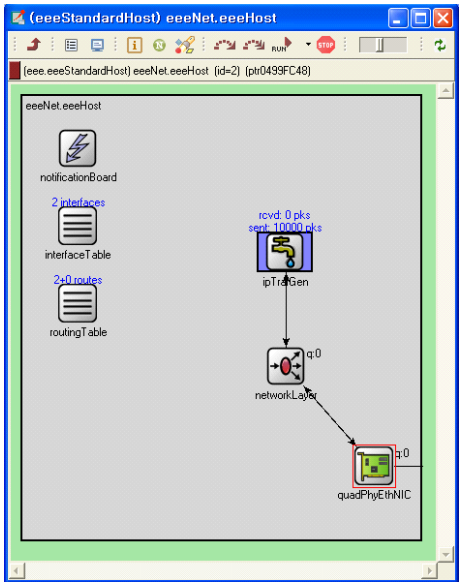


그림 8. 호스트 내부 구조
Fig. 8. Modules in a host.

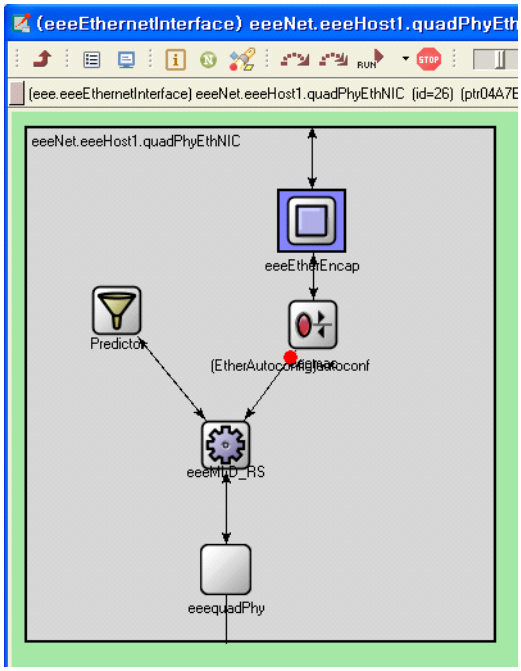


그림 9. QuadPhyEthNIC LAN카드의 내부 구조
Fig. 9. Modules for QuadPhyEthNIC LAN card.

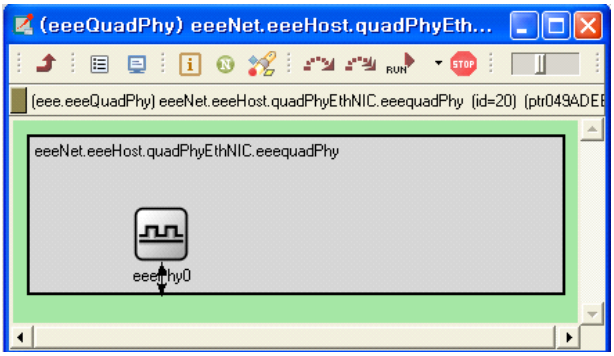


그림 10. eeeQuadPhy모듈의 내부 구조
Fig. 10. eeeQuadPhy Modules.

킷을 생성하는 ipTrafGen모듈과 이를 받아 IP헤더를 부착하여 quadPhyEthNIC로 전달하는 networkLayer 모듈, 그리고 networkLayer모듈로부터 전달된 IP패킷을 이더넷 MAC 프레임에 수납하여 케이블로 전송하는 quadPhyEthNIC모듈로 구성된다.

그림에서 routing table과 interfaceTable은 각각 IP입장에서 라우팅지원기능과 ARP를 지원하는 부가적인 모듈이다.

NetworkLayer모듈은 OMNET++에서 기본적으로 제공하는 모듈을 사용하였으며 ip모듈과 이를 지원하는 icmp, arp등의 서버모듈로 구성된다.

<그림 9>는 QuadPhyEtherNIC 모듈의 세부 구조이다. 이것은 NIC카드를 모델링한 것으로써 트래픽 예측기능을 수행하는 Predictor, eeeQuadPHY, 그리고 RS모듈로 구성하였다. 여기서 Predictor는 주기적으로 MAC의 송신버퍼인 TxQ의 크기를 샘플링하여 필요한 Queue 길이를 예측한 후, 미리 설정된 Threshold와 비교하여 레인의 추가 또는 삭제를 결정한다.

RS모듈은 Predictor에서 전달된 요구레인개수를 수납한 내부용 통보메시지를 전달받았을 때 MAC으로부터 전달된 MAC프레임의 헤더에 이 정보를 수납하여 상대방에게 전송한다.

이를 수신한 상대방은 자신의 레인갯수를 증가시켜야 할 경우에는 트랜시버의 ON지연시간인 100msec이

표 1. 모의실험 설정 값.
Table 1. Parameters for simulation.

파라미터	값
Bandwidth	10~40Gbps
Maximum Queue Size	10000
Average Length of message	1204byte
Traffic Load	0.0~1.0
Period of TxQ Polling	200us
High-threshold of Queue	3000~7000
Low-threshold of Queue	3~1000
Laser Diode On lagtime	100ms

후에 실제 활성화시키고, 감소시킬 경우에는 즉시 반영한다. 결과적으로 쌍방은 활성화 레인의 개수에 부합하여 10Gbps~40Gbps로 가변 동작한다.

<그림 10>은 NIC카드내부에 장착되는 4개의 PMD를 갖는 PHY모듈이다. 원래 각 PHY는 10Gbps로 동작하지만 용이한 시뮬레이터 구현을 위하여, 활성화된 레인개수에 따라 가변 전송율을 갖는 하나의 monolithic PHY로 설계하였다.

(나)시뮬레이션 파라미터

시뮬레이터에 사용된 파라미터는 <표 1>과 같다. Traffic Generator상위계층에서 생성하는 지수분포를 가지는 메시지의 평균길이는 1204바이트로 설정하였다. 이 메시지는 IP와 이더넷을 거치면서 IP헤더(20Byte), MAC헤더+FCS(=18Byte), Preamble(8byte)가 추가되어 평균길이 1250Byte로 송신된다. 또한 interarrival time도 지수분포를 가지도록 하였다.

V. 성능분석

<그림 11>은 유입되는 트래픽의 추세에 따라 실제 측정된 버퍼의 크기에 대하여 제안된 트래픽 예측기의 예측값이 LD의 ON lagtime에 따라 어느 정도 적응하는지를 비교한 것이다.

트래픽 예측기는 주기적으로 MAC의 TxQ를 샘플링하면서 다음 샘플링구간까지 사용할 버퍼크기를 예측하고, 또한 Threshold와 비교하여 레인의 활성화 유, 무를 판단한다. 현재 광케이블에 사용되는 레이저 다이오드(LD)가 ON되는 lagtime은 100ms이하이다. 따라서 예측된 버퍼크기가 Threshold를 초과한 경우 이를 상대방에 통보하더라도 실제로는 100ms이후에 레인이 활성화된다. 따라서 첫번째 그림에서 알 수 있듯이 LD lagtime을 고려한 경우에는 급격한 데이터유입 시에는 일시적으로 적응하지 못해 일부 프레임의 손실이 발생한다. 반면에 가상적이지만 LD lagtime=0인 경우라면 아랫쪽 그림과 같이 짧은 시간간격으로 적응을 시도하는 것을 알 수 있다.

<그림 12>는 레인의 활성화 개수의 변화를 도시한 것이다. 활성화된 레인의 감소는 Off 요청 즉시 바로 이루어지기 때문에 off 지연시간은 무시 할 수 있다. 하지만 활성화 레인이 증가해야 하는 경우는 레인증가를 통보받더라도 LD가 완전히 ON되어야 실제 송/수신이 가능하다. 즉 레인을 추가해야 한다고 판단된 경우 즉시 상대방에게 LD On 요청을 보낸다. 이에 수신측의 자신의 레인을 ON시키지만 100msec후 실제적으로 레인이 활성화 된다. 본 예에서는 0.006sec에 LD On 요청을 보

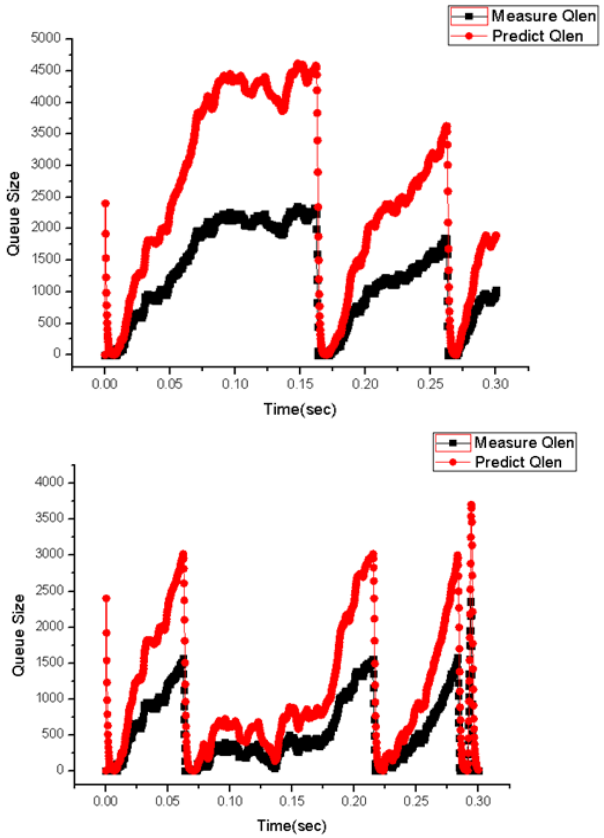


그림 11. 측정된 큐 길이와 예측된 큐 길이의 비교
Fig. 11. Measured and Esitmated queue lengths.

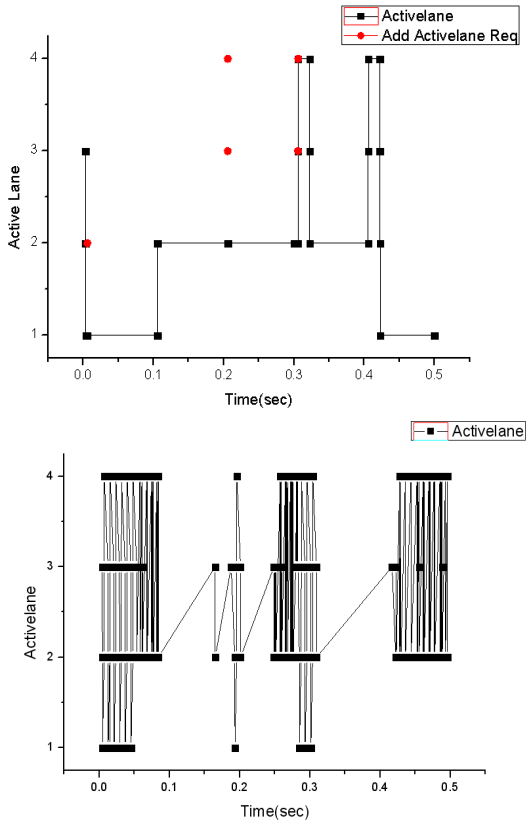


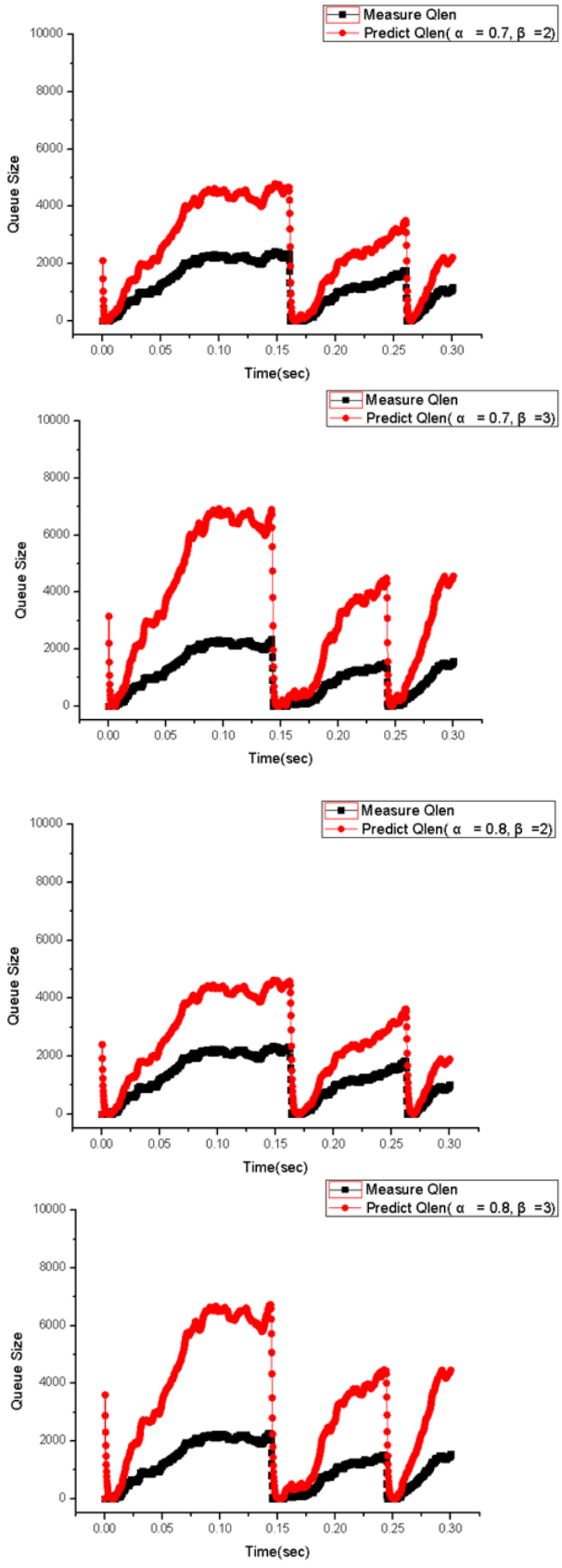
그림 12. 활성화된 레인 갯수의 변화
Fig. 12. Dynamics of active lane numbers.

내게 되고, 0.106sec에 레인을 활성화하는 것을 볼 수 있다. 참고로 아래쪽 그림은 LD lagtime을 고려하지 않은 가상적인 경우의 레인수 변화를 도시한 것이다. 이 경우 레인수의 변화가 아주 빈번함을 알 수 있다.

이러한 특성을 고려하여, 본 논문에서 제안하는 트래픽 예측기는 이전 예측값과 실제 측정된 값에 서로 다른 가중치를 두어 다음 샘플링 구간에서의 버퍼크기를 예측한다.

<그림 13>은 과거의 큐 길이 정보에 큰 가중치를 둘 것인지 방금 측정된 값에 가중치를 둘 것인지를 결정하는 가중치 α 값을 변경하면서 성능을 비교한 것이다. 여기서 현재시점 직전까지 사용된 예측 버퍼크기에 부가되는 가중치 α가 커질수록 완만하게 적응하는 것을 알 수 있다. 그리고 margin값 β 값이 증가할수록 실제 측정된 버퍼크기에 비해 여유 있는 값을 예측하는 것을 알 수 있다.

이를 통하여 실제 측정되는 버퍼크기에 비하여 큰 값을 예측하는 경우 미리 레인을 활성화시킬 수 있어 프레임의 손실을 줄일 수 있는 장점이 있지만, 활성화된



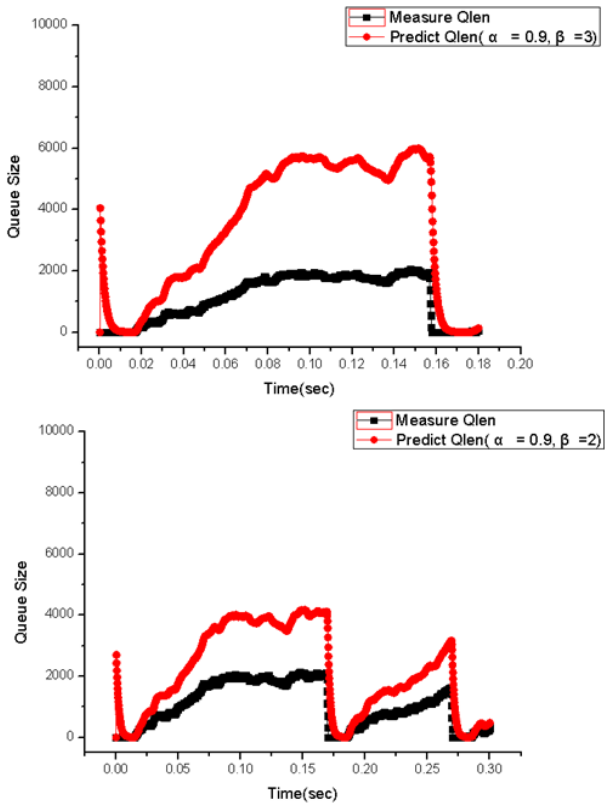


그림 13. 가중치에 따른 트래픽 예측기의 성능
Fig. 13. Dynamics of traffic estimator for various weighting factors.

표 2. Threshold 설정값
Table 2. Threshold values.

파라미터	값
High-threshold	3000~7000
Low-threshold	3~5000
load	0.5

레인의 개수를 증가시키므로 소모 전력량이 증가하는 단점이 있다. 패킷손실과 에너지 절감의 중요도에 따라 α, β 값을 조정하여 사용 가능하다.

또한 큐의 High-threshold와 Low-threshold의 적정 값을 도출하기 위하여 <표 2>과 같은 다양한 값을 를 사용하여 부하 0.5에서의 블록킹 확율과 소모전력을 비교 분석하였다.

<표 3>은 Queue의 High-threshold 에너지 소모율과 패킷 손실율을 나타낸다. High-threshold가 클수록 활성화되는 레인의 개수를 늘리기 위한 시도가 줄어들

표 3. Threshold값에 따른 패킷손실율과 소모전력 비교(부하=0.5)
Table 3. Packet losses and power consumptions for various threshold values(load=0.5).

High Threshold	Low Threshold	Packet loss Probability	Power consumtion
3000	3	0	5.84W
4000	3	0	4.46W
5000	3	0	3.81W
6000	3	5.376	3.92W
7000	3	6.58	2.85W

High Threshold	Low Threshold	Packet loss Probability	Power consumtion
5000	3	0	3.81W
5000	100	19.53	2.72W
5000	300	19.77	2.73W
5000	500	19.88	2.72W
5000	1000	20.00	2.71W

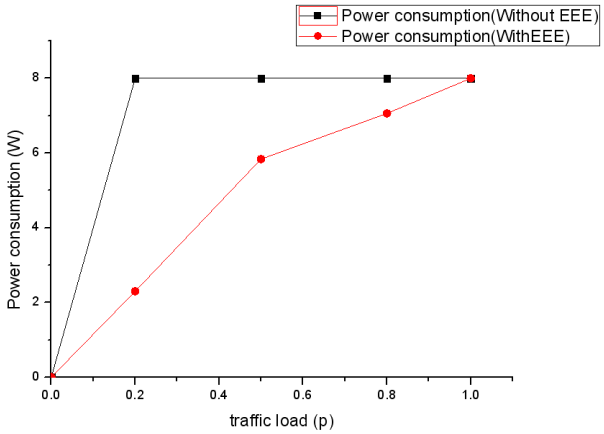


그림 14. 에너지 소모율 비교
Fig. 14. Comparison of energy consumptions.

게 되어 에너지 소모에서는 효율적이지만, 레인의 늦은 활성화로 인해 패킷손실이 증가하는 문제가 발생한다. 따라서 에너지 소모를 줄이기 위해서는 High-threshold 값을 크게 하지만 패킷의 손실을 줄이기 위해서는 High-threshold 값을 작게 해야하는 상호보완성이 있음을 알 수 있다.

<그림 14>는 본 논문에서 제안하는 EEE 방식과 EEE를 적용하지 않은 경우에 대하여 부하에 따른 전력 소모율을 비교한 것이다. 이것은 High-Threshold와

Low-Threshold값을 각각 3000과 3으로 설정한 경우이다. 부하 0.5의 경우 EEE를 적용하는 경우의 5.84W소모전력은 적용하지 않은 경우의 8W소모전력에 비하여 27% 전력 절감 효과를 가지는 것을 알 수 있다. 그리고 부하가 낮을수록 전력절감효과가 큰 것을 알 수 있다.

Ⅵ. 결 론

본 논문에서는 다중 레인을 사용하는 광 이더넷 시스템에 대한 EEE 적용 기법을 다루었다. 기존 방법의 장단점을 분석한 후 트래픽 부하에 따라 링크의 개수를 증감시키는 방법을 선택하였다. 이 방법의 구현 가능성을 제시하기 위해 MAC/RS 계층에 장착될 수 있는 Rate Controller 기능을 제안하여 기존의 PHY모듈을 그대로 사용하면서 에너지 절감과 레인 운용이 가능하게 하였다. RS계층에 추가된 Rate Controller는 활성화된 레인의 개수에 부합하여 비활성 레인에 Null 문자를 삽입하여 유효 전달율을 제공해 준다. 또한 PCS 및 PMA계층에 구현 시 문제가 되는 순간적인 데이터 유입에 대한 대처가 가능하게 된다.

또한 멀티레인을 효율적으로 운용할 때 필수적인 트래픽 예측기를 제안하고 시뮬레이션을 통해 기능을 검증하였다. 그 결과 패킷 손실과 에너지 절감 등 각각의 중요도에 따라 가중치를를 변경하여 사용자에 맞는 레인 운용 방법을 사용할 수 있음을 제시하였다.

본 논문에서 제안한 EEE 방법은 멀티레인을 사용하는 분야에서 새로운 전력 제어기법을 개발하는데 기여할 수 있을 것이다.

참 고 문 헌

[1] 신종윤 외, “40/100Gbps 이더넷 기술 및 표준화 동향”, *전자통신동향 분석* 제24권, 제1호, 2009.

[2] D’Ambrosia, Law, Nowell, “40 Gigabit Ethernet and 100 Gigabit Ethernet Technology Overview,” Ethernet Alliance White Paper, http://www.ethernetalliance.org/images/40G_100G_Tech_overview(2).pdf. , November 2008.

[3] “Energy-efficient Ethernet standard gains traction”. EETimes.com. http://www.eetimes.com/news/latest/showArticle.jhtml?articleID=207601205. Retrieved 2010-02-11.

[4] “Energy-Efficient Ethernet”. http://spectrum.ieee.org/computing/networks/energyefficient-ethernet. Retrieved 2010-02-11.

[5] “IEEE 802.3az: Energy Efficient Ethernet in the Works”. GoodCleanTech. 2008-09-04. http://www.goodcleantech.com/2008/09/ieee_8023az_energy_efficient_e.php. Retrieved 2010-02-11.

[6] S. Nedevschi, et al. “Reducing network energy consumption via sleeping and rate-adaptation,” *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, pp.323-336, 2008.

[7] P. Reviriego, et al., “Performance Evaluation of Energy Efficient Ethernet,” *IEEE Communications Letters*, Vol. 13, No. 9, pp. 697-699 , Sept. 2009.

[8] K. Christensen, P. Reviriego, B. Nordman, M. Bennett, M. Mostowfi, J. A. Maestro, “IEEE 802.3az: The Road to Energy Efficient Ethernet”, I*EEE Communications Magazine* , Vol. 48, No 11, pp. 50-56, November 2010.

[9] A. Maestro, P. Reviriego, “Energy Efficiency in Industrial Ethernet: the Case of Powerlink”, *IEEE Transactions on Industrial Electronics*, Vol. 57, No 8, pp. 2896-2903, August 2010.

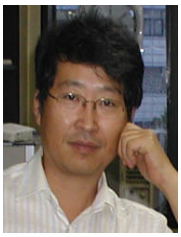
[10] P. Reviriego, J.A. Herndez, D. Larrabeiti, J.A. Maestro, “Burst Transmission in Energy Efficient Ethernet”, *IEEE Internet Computing* , Vol. 14, No 4, pp. 50-57, July/August 2010.

[11] Official page of OMNeT++, http://omnetpp.org.

[12] Andras Varga, “OMNeT++ User Manual Version 3.2”, March 2005.

저 자 소 개

서 인 수(학생회원)
2010년 한국항공대학교 정보통신공학부 학사
2012년 한국항공대학교 정보통신공학부 공학석사
2012년～현재 (주)현대오토에버 MES표준화팀
<주관심분야: 차량용 이더넷 시스템>



윤 종 호(정회원)
1984년 한양대학교 전자공학과 학사
1986년 한국과학기술원 전기및 전자공학과 공학석사
1990년 한국과학기술원 전기및 전자공학과 공학박사
1991년～현재 한국항공대학교 항공전자정보통신공학부 교수
<주관심분야: 실시간 및 항공용 이더넷 시스템>

양 충 열(정회원)
1983년 건국대학교 전자공학과 학사
1998년 충남대학교 전자공학과 석사
2007년 충남대학교 전자공학과 박사
1992년 6월～현재 한국전자통신연구원 광인터넷연구부 광전송기술연구팀 책임연구원
<주관심분야: 광통신, 광패킷스위칭, 광인터넷>