IXP2400 네트워크 프로세서를 이용한 IPv6 멀티캐스트 포워딩 모듈의 설계 및 구현

송지수° 박우진 김대희 안순신
고려대학교 전자컴퓨터공학과
{jssong°, wjpark, dhkim, sunshin}@dsys.korea.ac.kr

Design and Implementation of an IPv6 Multicast Forwarding Module on the IXP2400 Network Processor

Jisoo Song, Woojin Park, Daehee Kim Sunshin An
Computer Network Lab. Dept. of Electronics and Computer Eng., Korea University

요 약

본 논문은 인텔사의 IXP2400 네트워크 프로세서를 이용하여 IPv6 multicast-enabled 라우터 개발의 예비단계로서 IPv6 멀티캐스트 모듈의 전체적인 설계 및 구현을 다룬다. 특히, 마이크로 엔진할당, IPv6 멀티캐스트 마이크로 블록 및 패킷 복사 마이크로 블록에 중점을 둔다. 우리의 IPv6 멀티캐스트 포워딩 모듈의 성능측정 결과는 이론적 한계치의 86%였다.

1. 서 론

오늘날 음성과 데이터 네트워크간의 통합을 하려는 움직임과 함께 통신 산업의 변화가 빨라지고 있으며 많은 새로운 서비스가 등장하고 있다. 이에 따라 새로운 서비스를 지원하는 네트워크 장비의 개발이 요구되고 있고 장비의 개발단계에서 시장으로 나오기까지의 기간을 단축하는 것이 점점 중요한 문제가 되고 있다. 결과적으로 네트워크 장비는 ASIC에서의 성능을 유지하면서 새로이 출현하는 표준에 신속히 적응하기 위해 프로그램가능성과 유연성이 필요하다. 이를 위해서는 근본적으로 새로운 접근이 필요한데, 이에 대한 하나의 해결책으로 네트워크 프로세서가 등장하게 되었다.

네트워크 프로세서는 하드웨어 레벨 성능을 소프트웨어적으로 프로그램 가능한 시스템에서 얻을 수 있게 하였다. 이를 통해 새로운 서비스에 대한 기능을 제공하는 네트워크 장비를 기존 ASIC장비의 개발보다 더욱 짧은 기간 내에 더욱 적은 비용으로 개발할 수 있다. 그리고 네트워크 프로세서는 소프트웨어적인 업데이트만으로 새로운 기술을 제공할 수 있기 때문에 실제 장비의 수명 또한 ASIC을 이용한 장비보다 길다할 수 있다.

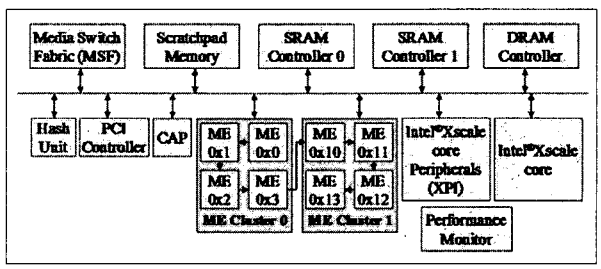
현재 사용되고 있는 IPv4의 32비트 주소체계에서는 앞으로 증가될 많은 수의 사용자와 인터넷의 확장에 의한 주소의 필요를 충족시키지 못한다. IPv6는 128비트 주소체계를 사용하여 지구상의 모든 네트워크 장치에 대해 고유의 IP 주소를 할당하고도 남을 만큼 충분한 주소공간을 지원한다.

IP 멀티캐스트는 실시간 멀티미디어 어플리케이션(예를 들어, 인터넷 라디오, 인터넷 텔레비전, 화상회의)과 같은 다수의 사용자에게 높은 대역폭을 요구하는 데이타 스트림을 낮은 지연시간으로 전송하기를 원하는 bandwidth-intensive 네트워크 어플리케이션에 대한 실질적인 해결책임에도 불구하고 효율적인 그룹 관리 방법의 개발이 어렵고 대부분의 상용 라우터들이 멀티캐스트 패킷에 대해 낮은 우선순위를 두고 여러 사용자에게 동일한 데이타를 전송할 때에 유니캐스트 패킷을 반복하여 전송하는 방법을 사용하고 있어 현재 인터넷에서 널리 사용되지 못하고 있다.

본 논문은 인텔사의 IXP2400 네트워크 프로세서를 이용하여 IPv6 multicast-enabled 라우터 개발의 예비단계로서 IPv6 멀티캐스트 모듈의 전체적인 설계 및 구현을 다룬다. 특히, 마이크로 엔진할당, IPv6 멀티캐스트 마이크로 블록 및 패킷 복사 마이크로 블록에 중점을 둔다. 우리의 IPv6 멀티캐스트 포워딩 모듈의 성능측정 결과는 이론적 한계치의 86%였다.

2 IXP2400 네트워크 프로세서의 개요

IXP2400 네트워크 프로세서는 인텔사의 2세대 네트워크 프로세서 패밀리 중 하나로 넓은 범위의 접근성과 edge 어플리케이션을 위해 설계되었다[1]-[3]. IXP2400은 단일 칩에서 복잡한 알고리즘을 수행하고 deep packet inspection을 수행하며 트래픽을 관리하고 wire speed로 포워딩하기위하여 고성능의 병렬 처리 구조를 가진 완전히 프로그램 가능한 네트워크 프로세서이다.



[그림 1] IXP2400 네트워크 프로세서의 블록 다이어그램

각각의 IXP2400은 8개의 multi-threaded packet-processing 마이크로엔진을 가지고 있고 저전력의 범용 Intel® XScale™ microarchitecture 코어, network media switch fabric 인터페이스, 메모리 컨트롤러, PCI 컨트롤러와 flash PROM과 주변 장치에 대한 인터페이스를 가지고 있다. 그림 1에서 IXP2400 네트워크 프로세서에 대한 블록 다이어그램을 보여준다[1]. 8개의 마이크로엔진은 프로그램 가능한 패킷 프로세서이고 각각은 8개의 thread까지 multi-threading을 지원한다.

* 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 육성·지원사업의 연구결과로 수행되었음

3. IPv6 멀티캐스트의 설계 및 구현

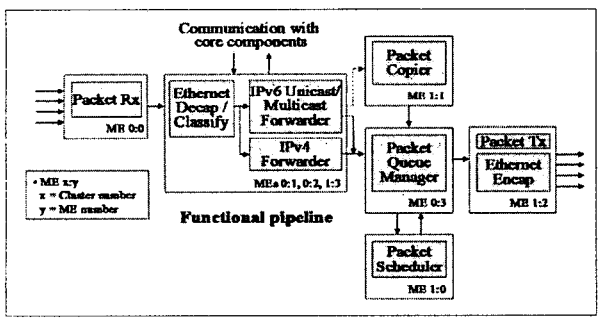
3.1 설계의 목적

본 논문의 최종목표는 멀티캐스트 패킷이 전체 네트워크 트래픽의 많은 부분을 차지하는 환경에서 사용하기 위한 초고속 IPv6 multicast-enabled 라우터를 설계 및 구현하는데 있다.

우리는 인텔사에서 제공하는 quad gigabit Ethernet forwarding pipeline application을 기반으로 하여 본 논문의 IPv6 멀티캐스트 모듈을 설계 및 구현하였다[4] [5]. 본래의 라우팅 툴은 IPv6 멀티캐스트 라우팅을 지원하지 않았기 때문에 IPv6 멀티캐스트 라우팅을 지원하도록 코드를 재작성하고 새로운 내용을 추가하였다. IPv6 멀티캐스트 포워더에 집중하기 위해 멀티캐스트 라우팅 프로토콜[8]이나 멀티캐스트 그룹 매니지먼트 프로토콜(MLD[9]와 같은)을 구현하지 않고 라우터와 호스트간에는 수동으로 설정하여 구성하였다. 우리의 구현은 다른 멀티캐스트 프로토콜에 대한 가정을 하고 구현하였으므로 다른 프로토콜에 적용 할 때에는 아주 작은 변화만을 가하여 쉽게 적용할 수 있다.

3.2 IPv6 Multicast and Packet Flow를 위한 Microblocks

그림 2는 IPv6 멀티캐스트를 위한 마이크로블록들과 패킷 흐름을 보여준다. IPv6 멀티캐스트 패킷 처리를 위해 필요한 수만큼의 패킷 복사를 수행하는 하나의 마이크로엔진(Packet Copier Microblock)을 할당하였다.



[그림 2] 마이크로 블록 할당 및 패킷 흐름

IPv6 패킷 포워딩의 종류는 IPv6 헤더의 destination field의 첫 번째 8비트에 의해 결정되는 세 가지의 경우로 나눌 수 있다. 각각에 대한 처리과정을 요약하면 다음과 같다:

- IPv6 유니캐스트 패킷 (Case I) - IPv6 유니캐스트 라우트 룩업후에 패킷 메타데이타(각 패킷마다 하나의 메타데이타를 지니고 있으며, DRAM의 패킷의 저장위치 및 출력될 포트번호등 패킷에 대한 추상적인 정보)의 모든 필드는 세팅된다. 이 패킷은 이후의 처리를 위해 IPv6 유니캐스트 마이크로 블록으로부터 패킷 큐 관리 마이크로 블록으로 전달된다.
- IPv6 멀티캐스트 패킷 (Case II) - IPv6 멀티캐스트 라우트 룩업의 결과로서 패킷 전송을 위한 출력 포트가 둘 이상으로 결정된 경우이다. 이 경우 패킷 복사(그림 2의 점선)를 위해 IPv6 멀티캐스트 포워더 마이크로 블록으로부터 패킷 복사 마이크로 블록으로 전달된다. 이 경우에 패킷 메타데이타의 모든 필드는 패킷 큐 관리 마이크로 블록에서의 적당한 처리를 위해 일부분 수정된다.
- IPv6 멀티캐스트 패킷 (Case III) - IPv6 멀티캐스트 라우트룩업의 결과로서 패킷 전송을 위한 출력 포트가 하나로 결정된 경우이다. 이 경우에 패킷은 복사될 필요가 없으며 패킷 전송 마이크로 블록에서의 멀티캐스트 L2 테이블 encapsulation을 제외하고 IPv6 유니캐스트 패킷과 동일하게 처리를 한다. Case II와 달리 이 패킷은 IPv6 멀티캐스트 포워더 마이크로 블록에서 패킷 큐 관리 마이크로 블록(그림 2의 실선)으로 전달된다. 이 패킷은 IPv6 유니캐스트 패킷과 구별되어야 하는데 이를 위해 패킷 메타데이타 포멧의 예약된 필드를 재정의하였다.

3.3 IPv6 멀티캐스트 포워더 마이크로 블록

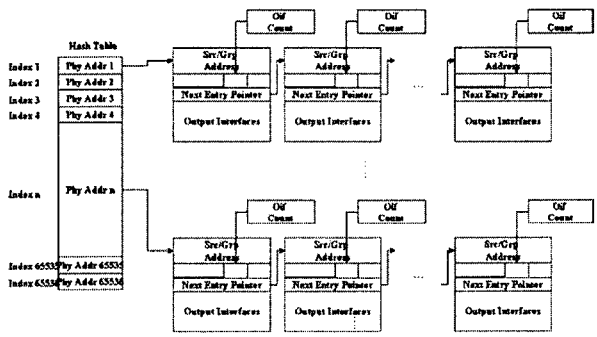
3.3.1 IPv6 멀티캐스트 라우팅 테이블

IPv6 멀티캐스트 라우팅 테이블은 DRAM에 저장이 되고 각 엔트리(크기: 4136 바이트)의 내용은 그림 3에서 보여지는 바와 같다. 우리의 IPv6 멀티캐스트 라우팅 테이블은 512개의 출력 인터페이스를 지원하도록 설계되었다. 각 출력 인터페이스는 멀티캐스트 L2 인덱스(패킷 전송 마이크로 블록에서 멀티캐스트 패킷의 L2 인캡슐 용도), nexthop type(포워딩 및 터널링 체크), 출력 포트 ID와 링크의 MTU(Maximum Transmission Unit)에 대한 정보를 지니고 있다.

Input Interface (Iif)	Olf Count	Reserved	D L
Next Entry Pointer			
Multicast L2 Index 0	Reserved	Next hop type	
Port ID	MTU		
Multicast L2 Index 1	Reserved	Next hop type	
Port ID	MTU		
⋮	⋮		
Multicast L2 Index n	Reserved	Next hop type	
Port ID	MTU		
Multicast L2 Index 510	Reserved	Next hop type	
Port ID	MTU		
Multicast L2 Index 511	Reserved	Next hop type	
Port ID	MTU		

[그림 3] IPv6 멀티캐스트 라우팅 테이블 엔트리 구조

3.3.2 IPv6 멀티캐스트 라우트 룩업



[그림 4] IPv6 멀티캐스트 라우트 룩업

IPv6 멀티캐스트 라우트 룩업의 효율을 증대시키기 위하여 FreeBSD에서와 같이 해쉬함수을 이용하였다. 우리의 해쉬함수의 동작은 다음과 같다:

- Source address와 Group address (총 256 비트)를 8개 (각각 32 비트)로 나눈다.
- 나뉘어진 8개를 순서대로 Exclusive-OR(XOR) 연산을 수행한다.
- 위의 연산의 결과는 32비트의 데이터이며, 이 값에서 하위 16비트를 해쉬키라 하며 해쉬 테이블의 인덱스로 사용한다.

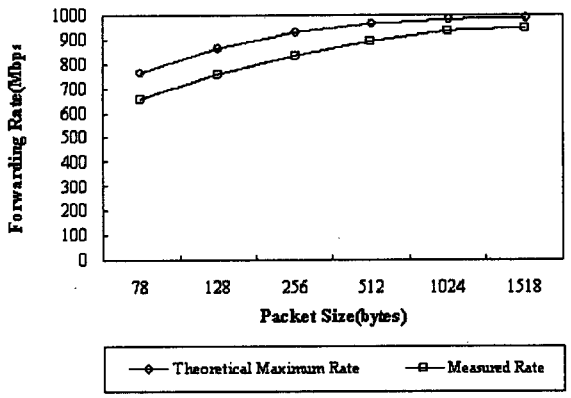
해쉬 테이블은 SRAM에 저장되며 256 kbytes(2^16*4 bytes)의 크기를 가진다. 그리고 해쉬키(16 비트)는 DRAM 메모리 접근에 의한 지연을 최소화하기 위하여 선택된다.

IPv6 멀티캐스트 라우트 룩업에서 해쉬 테이블은 그림 4에서 보여지는 바와 같이 해쉬키에 의해 접근이 되며 그 값은 첫 번째 IPv6 멀티캐스트 라우팅 테이블 엔트리로의 포인터를 가지며 동일한 해쉬키를 가지는 IPv6 멀티캐스트 그룹들은 "Next Entry Pointer"필드에 의해 링크드 리스트로 관리된다.

3.3 패킷 복사 마이크로 블록

패킷 복사 마이크로 블록은 두 개 이상의 출력 인터페이스를 가지는 IPv6 멀티캐스트 패킷의 복사를 한다. DRAM 버퍼에 저장되어 있는 패킷의 데이터를 복사하는 것은 많은 자원이 낭비되기 때문에 패킷 복사는 패킷의 실제 데이터를 복사하는 것은 아니다. 버퍼를 설계할 때 parent buffer handle과 child buffer handle을 고려하였다. 이 두 버퍼의 처리를 구분함으로 패킷 전송 마이크로 블록이 각 패킷에 적합한 유니캐스트/멀티캐스트 L2 테이블을 encapsulation하도록 하였다. 멀티캐스트를 위해 복사되어야 할 패킷 버퍼를 parent packet buffer라 하고 이 패킷의 버퍼 descriptor를 parent buffer descriptor라 한다. parent buffer descriptor의 여러 복사본을 child buffer descriptor라 한다. child buffer descriptor로부터 멀티캐스트 패킷이 전송될 인터페이스의 수가 결정되고 parent buffer descriptor의 reference count의 수만큼 DRAM에 저장되어 있는 original packet buffer가 재전송된다. 복사된 child metadata는 IPv6 멀티캐스트 포워더 마이크로 블록으로부터 들어오는 request message에 의해 결정된다. 패킷이 복사되는 경우, 버퍼에 얼마나 많은 children이 존재하는지를 나타내는 parent metadata의 Reference count는 수정된다.

4. 성능 분석 및 평가



[그림 5] 패킷 크기별 전송속도

하나의 IXP2400 네트워크를 사용하는 인텔사의 IXDP2401 advanced development platform의 네트워크 인터페이스에 패킷 생성기(Smartbits)를 연결하여 Gigabit Ethernet link에서 패킷을 생성하고 캡쳐하여 포워딩 속도를 테스트 하였다.

포워딩 속도에 대한 평가를 위한 테스트 시나리오는 다음과 같다:

- IPv6 멀티캐스트 라우트 테이블은 IPv6 Multicast Forwarding Plane Manager를 이용하여 수동으로 설정하였다. (입력 인터페이스: 포트0, 출력 인터페이스: 포트0, 포트1). 현재의 IXDP2401에서는 4개의 포트 중 2개만을 지원하고 있으므로 Reverse Path Forwarding은 고려하지 않았다.
- 패킷 생성기에 의해 생성된 IPv6 멀티캐스트 패킷(64 ~ 1518 바이트)은 IXDP2401의 0번 포트로 전송된다.
- 전송 속도는 패킷 생성기에 의해 측정되고 분석된다.

이론적인 최대 전송 속도는 full-duplex Ethernet을 기반으로 계산되어진다[10]. 그림 5는 패킷의 크기에 대한 이론적인 최대 전송 속도와 측정된 전송 속도를 그래프로 나타낸 것이다. 특히, worst cast로 IPv6의 minimum-sized Ethernet packet (78바이트)의 경우에는 다음과 같은 결과를 얻을 수 있었다 :

- Line Full Rate : 1,302,076 pps
- Transmission Rate : 1,124,034 pps
- 86% of Line Full Rate

측정된 전송 속도는 IPv6 멀티캐스트 포워더 마이크로 블록에서의 IPv6 멀티캐스트 라우트 룩업에 의한 지연과 패킷 복사 마이크로 블록에서의 패킷 복사에 의한 지연으로 인하여 나온 결과이다.

5. 결론 및 향후 연구

본 논문에서는 멀티캐스트 패킷이 전체 네트워크 트래픽의 많은 영역을 차지하고 있는 환경에서 필요한 초고속 IPv6 multicast-enabled 라우터 개발의 예비 단계로서 IXP2400 네트워크 프로세서상의 IPv6 멀티캐스트 모듈의 설계와 구현을 다루었다. 하드웨어 플렛폼을 보유하고 있었기 때문에 하드웨어 테스트를 통하여 구현에 대한 검증과 성능 분석을 할 수 있었다.

향후 초고속 IPv6 multicast-enabled 라우터를 개발하기 위해서는 본 논문에서 다루지 않은 멀티캐스트 라우팅 프로토콜(예를 들어, PIM-SM)들과 멀티캐스트 그룹 관리 프로토콜(예를 들어 MLD)들을 적용해야한다. 또한 현재 우리의 IPv6 멀티캐스트 라우트 룩업은 일치하는 엔트리를 찾기 위해 반복적으로 검색을 하도록 되어있는데 만일 일치하는 엔트리가 linked list의 마지막에 존재하는 경우에 많은 지연이 발생되는 문제가 있다. 보다 향상된 전송 속도를 얻기 위해서는 이러한 IPv6 멀티캐스트 라우트 룩업의 문제점들을 보완해 룩업시에 발생하는 지연을 줄여야 할 것이다.

6. 참고 문헌

[1]"Intel IXP2400 Network Processor Hardware Reference Manual", Intel Corporation, Nov 2003
[2]"Intel IXP2400 Network Processor Datasheet", Intel Corporation, Nov 2003
[3]"Intel IXP2400 Network Processor Programmer's Reference Manual", Intel Corporation, Nov 2003
[4]"Intel Internet Exchange Architecture Software Building Blocks", Intel Corporation, Mar 2004
[5]"Intel Internet Exchange Architecture Software Building Blocks Applications Design Guide", Intel Corporation, Nov 2003
[6]R. Hinden, S. Deering, "IPv6 Multicast Address Assignments", [RFC2375], July 1998
[7]R. Hinden, S. Deering, "Internet Protocol, Version 6 (IPv6) Specification", [RFC2460], Dec 1998
[8]D. Estrin, D. Farinacci et al, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification", [RFC2362], June 1998
[9]R. Vida, L. Costa, "Multicast Listener Discovery Version 2(MLDv2) for IPv6", [RFC3810], June 2004
[10]S. Karlin, L. Peterson, "Maximum Packet Rates for Full-Duplex Ethernet", Technical Report TR-645-02, Feb 2002