

主 題

멀티미디어 광대역 ATM 스위치 개발기술 동향

건국대학교 전자공학과 김 영 범

차 례

- I. 서론
- II. 광대역 ATM 스위치 구현상의 문제점
- III. 연구 동향
- IV. 맺음말

I. 서론

고성능, 저가의 컴퓨터가 널리 보급되고 인터넷 및 WWW 사용자가 급속히 증가함에 따라 망 사용자 트래픽은 폭발적으로 증가하고 있으며 이러한 새로운 서비스의 증가는 비단 인터넷에만 국한되지 않고 전화망에서부터 케이블 TV, LAN, 셀룰라 및 위성통신에 이르는 전기통신의 모든 분야에 걸치고 있다. 이러한 망 사용자 트래픽의 증가는 단순히 양적인 면에서 뿐만이 아니라 다양한 특성을 갖는 트래픽 종류면에서의 증가를 의미하며, 이에 따른 기간 정보 통신망에 대한 속도, 전송 대역폭 및 서비스 품질에 대한 요구는 날로 증대되고 있는 상황이다.

하나의 통신망이 통신 링크와 망 노드에 해당하는 스위치로 구성되어 있다고 보았을 때, 광통신 기술의 지속적인 발달에 따라 통신 링크의 전송속도는 현재 상당한 수준에 도달해 있으므로 광대역 정보 전송을 위한 통신망 상의 bottleneck은 스위치

에 있다고 할 수 있다. 현재 컴퓨터 통신망에 부과되는 사용자 트래픽 및 향후 추가될 것으로 예상되는 트래픽의 양적 그리고 질적 다양성을 고려했을 때, 이를 효과적으로 수용하기 위한 광대역 ATM 스위치에 대한 필요 조건은 다음의 두 가지로 요약할 수 있다:

(R1) Wire speed (광통신 링크 상의 전송속도, 예: OC-12의 경우 포트당 556 Mbps)에 상응하는 스위치 셀 처리 속도.

(R2) 멀티미디어 트래픽을 지원하기 위한 서비스 품질 기능 (QoS functionality).

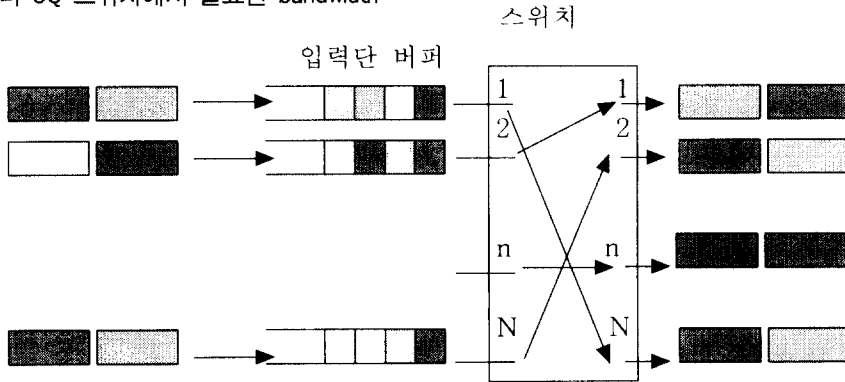
이 두 가지 필요조건은 스위치 설계에 있어서 서로 상충되는 요소를 지니고 있다. 즉, 필요조건 (R1)을 달성하기 위해서는 스위치 하드웨어 복잡도가 낮아야 하는 반면, 필요 조건 (R2)은 필연적으로 하드웨어 복잡도를 높이게 된다. 본고에서는 이러한 광대역 스위치 개발에 따른 문제점과 이를 해결하기 위한 연구 동향 등을 소개하기로 한다.

II. 광대역 ATM 스위치 구현상의 문제점

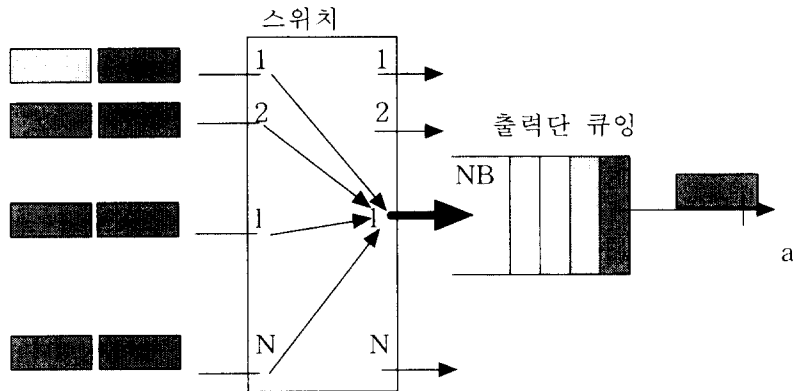
Optical fiber를 이용한 통신링크의 bandwidth는 현재의, 그리고 향후 예상되는 망 트래픽을 충분히 수용할 수준에 이르고 있으나, 통신 링크가 상호 연결되는 스위치의 경우 대부분의 셀 처리 과정은 전기적으로 수행되기 때문에, 광대역 트래픽을 처리하기 위한 통신망의 bottleneck은 망 노드에 해당하는 스위치라고 할 수 있다. 따라서 고속의 통신 링크 전송 속도에 상응하는 고속 ATM 스위치 구조의 개발은 아주 중요한 문제라 할 수 있다.

ATM에서는 망자원의 효율적 사용을 위하여 종
 그림 1. IQ와 OQ 스위치에서 필요한 bandwidth

래의 PSTN(Public Switched Telephone Network)에서 취하고 있는 Synchronous Transfer Mode와는 달리 사용자 트래픽을 통계적 다중화하는 방식을 취하고 있다. 따라서 셀 트래픽은 사전 조정 없이 (unscheduled) 스위치내로 유입되므로 서로 다른 VC에 속하는 셀간의 출력포트를 향한 충돌(output contention)은 필연적이다. 따라서 즉각적으로 처리할 수 없는 셀을 저장하기 위한 버퍼가 스위치내의 어느 위치에 있어야 하며 switch fabric을 기준으로 버퍼의 상대적인 위치에 따라 ATM 스위치 구조는 크게 입력 큐잉(IQ: Input Queueing)과 출력 큐잉(OQ: Output Queueing) 스위치로 나눌 수 있다 (그림 1).



(a) 입력 큐잉



(b) 출력 큐잉

입출력 링크 전송 bandwidth B 를 갖는 $N \times N$ ATM 스위치에 대해 요구되는 스위치내 셀 처리 bandwidth는 다음과 같다.

- IQ 스위치에 대해 $O(1)$: 입력 큐에 write하기 위한 memory bandwidth B , 입력 큐로부터 read하기 위한 memory bandwidth B , switch fabric의 전송속도 B .

- OQ 스위치에 대해 $O(N)$: 출력 큐에 write하기 위한 최대 NB 의 memory bandwidth, 출력 큐로부터 read하기 위한 memory bandwidth B , switch fabric의 전송속도 최대 NB .

OQ 구조에 대해 요구되는 bandwidth는 $O(N)$ 이므로 IQ 구조에 비해 스위치 내부의 동작 속도는 N 배 빨라야 한다. RAM의 access 속도는 유한하므로 $O(N)$ speedup은 OQ 스위치의 크기에 큰 제한을 주게 된다. 하나의 셀을 처리하기 위해서는 최소한 두 개의 memory operation (read, write)이 필요하므로 가장 빠른 50 ns cycle time을 갖는 DRAM을 사용했을 때의 OQ 스위치의 전체 처리율 (Aggregate throughput)은 4.24 Gbps로 제한된다.

좀더 빠른 OQ 스위치는 Banyan기반의 memory accessing network를 갖는 공유 메모리 구조를 사용하거나, customized memory control를 가진 memory design을 사용함으로써 구현이 가능하나 상당한 부가 비용이 필요하고, 여전히 공유 메모리 구조가 갖는 취약점인 switch fabric에 대한 N 배의 speedup이 요구된다. OQ 구조에서의 $O(N)$ 대역폭 요구는 Knockout switch에서와 같이 어느 정도의 cell loss를 허용함으로써 줄일 수 있으나 cell loss를 적정 수준(10^{-9})까지 줄이기 위해서는 역시 상당한 switch fabric에 대한 speedup(약 12배)이 요구된다.

현재 memory bandwidth는 network bandwidth에 대한 사용자 요구 추세를 만족할 수준으로 향상시키기 어렵기 때문에 통신링크의 전송속도의 N 배의

speedup을 요하는 공유 메모리 구조나 OQ 구조는 10Gbps급 이상의 대용량 ATM switch 설계에 적용하기 어렵다는 치명적인 단점을 가지며 보통속도의 ATM 스위치에서만 유용하다. 반면에 공유 메모리 구조와 OQ 구조에서는 도착 셀들의 출력 주소별로 큐가 구성되기 때문에 셀 처리율이 높고 (100%) 또한 QoS 기능 구현이 용이하다는 장점이 있다.

가. 광대역 ATM 스위치 구현을 위한 대안으로서의 IQ 구조

이 절에서는 OQ 스위치에 비해서 확장성 면에서 대용량 스위치 구현에 적합한 것으로 알려진 IQ 스위치 구조가 갖는 장점과 문제점들을 각각 살펴보기로 한다.

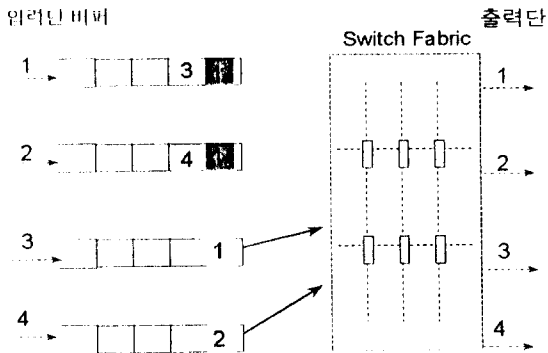
IO 스위치의 장점으로서의 첫째, OQ 스위치 구조에 비해 IQ 구조는 각 입력 큐 모듈에서 단일 통신 링크의 전송 속도로 도착 셀들을 버퍼에 저장하면 되기 때문에 상당히 우수한 확장성을 갖는다. Memory bandwidth 측면에서 살펴보면, 4Gb/s의 동작하는 link를 갖는 32-port ATM 스위치는 저비용, 고밀도의 DRAM device를 이용하면 128 Gb/s의 총 처리율이 실현 가능하다. 최소 bandwidth 측면외에, 둘째로 simulation을 통한 연구 결과[1]에 의하면 동일 buffer 크기에 대해 IQ 스위치는 상대적으로 bursty traffic을 보다 더 잘 처리하는 것으로 알려졌다. 셋째로, multicast traffic에 대한 측면을 살펴 보면, m 개의 output port로 multicast되는 n 개의 cell을 처리하기 위해 OQ 구조에서는 $m \cdot n$ cell buffer가 필요한 반면 IQ 구조에서는 단지 n 개의 cell buffer만을 필요로 한다. 이러한 사실들을 종합하면 IQ 구조는 광대역 ATM network에 대한 요구되는 필요 조건을 충족하기 위한 스위치 구조로서 적합하다.

IQ 구조는 앞에서 언급한바와 같이 확장성, bursty traffic 및 multicast traffic 처리면에서 우수

한 성능을 가짐에도 불구하고 광대역 ATM 스위치 구현을 위한 구조로써 적용되기 위해서는 선결되어야 할 중요한 문제점들을 가지고 있다. 이러한 문제점은 다음과 같이 요약할 수 있다.

첫째로, FIFO 원칙을 갖는 IQ 구조에서는 그림 2에서와 같은 HOL(head-of-the-line) blocking 현상이 존재함으로써 스위치의 셀 처리율을 상당히 제한한다는 단점이 있다. 그림 2에서 보듯이 입력큐 1,2의 두 번째 셀들은 원하는 출력 포트가 미사용 중임에도 불구하고 각기 큐 전단에 위치한 HOL 셀들에 의해 routing이 차단되고 있다. 이 HOL blocking 현상은 IQ 스위치의 최대 처리율을 0.586으로 제한하게 된다[2][3].

그림 2. 4x4 IQ 스위치에서의 Head-of-the-blocking 현상.



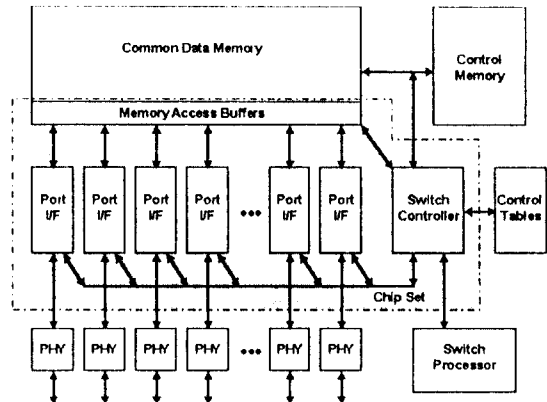
둘째로, IQ 구조에서는 다른 출력포트로 향하는 여러 VC에 속하는 셀들이 하나의 입력 포트에 혼합되어 있기 때문에, ATM이 갖는 최대의 장점인 QoS(서비스 품질) 기능을 실현시키기에 상당히 어려운 점이 있다.

나. QoS 기능 구현

ATM 망에서는, 하나의 호 설정 요구에 의해 할당되어 있는 스위치내의 자원 (스위치 내 셀 경로상의 대역폭, 버퍼, processing time, 스위치 출력 링크상의 대역폭 따위)은 여러개의 VC에 의해 통계적으로 공유되므로 각각의 VC에 대해 실제적으로 설정이 안되어 있을 수 있다. 자원의 공유는 여러개 VC간의 충돌을 피할 수 없기 때문에, 하나의 ATM 스위치내에 존재할 수 있는 수천에서 수만개의 VC간의 간섭에 의한 QoS 저하를 줄이기 위한 스위치 내 셀 흐름 제어 메카니즘의 개발은 매우 중요한 기술이다.

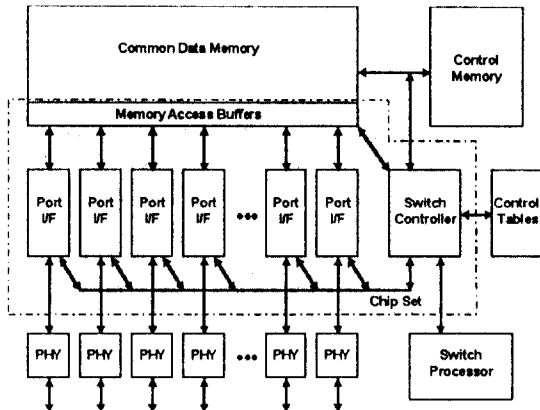
최근 미 MMC Networks, Inc.에서는 혼합된 VC간의 간섭 문제를 해결하기 위한 하나의 방법으로써, 공유 메모리와 pointer를 사용한 VC 별로의 큐를 구성하는 방식을 사용한 per-flow queueing(PFQ) 개념의 XStream chipset을 발표하였고(그림 3) CISCO Systems에서는 자사의

그림 3. MMC 스위치 구조



LightStream 1010 ATM 스위치에 이 chipset을 채택한바 있다. 그러나 MMC의 셀 처리 방법은 근원적으로 OQ 구조라 할 수 있는 공유 메모리 형태의 스위칭 구조를 갖기 때문에, 전에 언급된 바와 같이 확장성면에서 확연한 제한을 받는다. 따라서 이 스위치 구조는 campus나 Metropolitan Area Network (MAN) 용도의 수 Gbps 용량 스위치에 적합한 것으로 평가되고 있다.

그림 3. MMC 스위치 구조

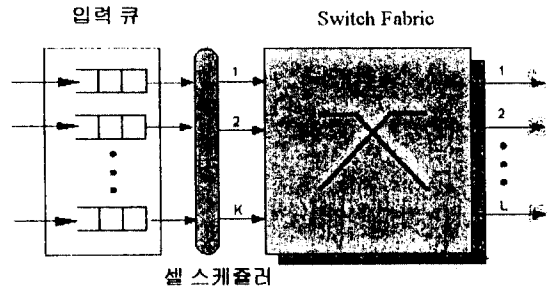


III. 연구 동향

이 장에서는 광대역 스위치 구현을 위한 방안으로써 IQ 구조가 갖는 단점을 해결하기 위한 연구 동향을 살펴보기로 한다.

IQ 스위치의 구성은 크게 switch fabric, 즉각적으로 스위칭이 될 수 없는 셀들을 저장하기 위한 입력 버퍼, 그리고 동일 출력주소를 갖는 경쟁상태의 셀들중에서 routing이 될 셀을 선택하기 위한 셀 스케줄러로 나눌 수 있다 (그림 4 참조).

그림 4. 입력 큐잉 스위치의 구성



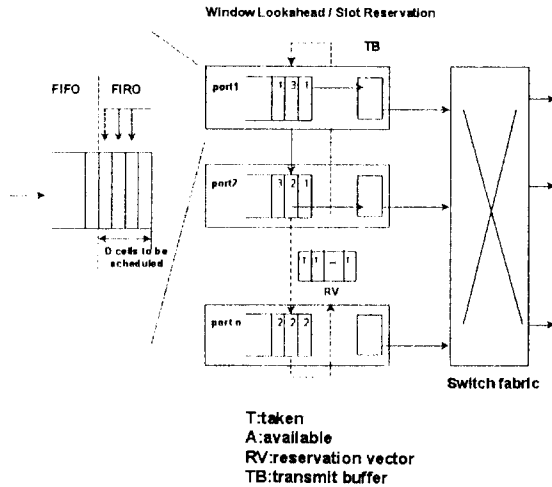
전에 언급한 입력 큐잉 스위치의 전송 대역폭면에서의 장점을 살리기 위해서는 FIFO 큐를 갖는 IQ 스위치에서 존재하는 HOL blocking을 제거하기 위한 방법이 모색되어야 하고 이는 효과적인 버퍼 구조와 셀 스케줄러의 설계에 달려 있다. 이러한 설계는 또한 수반되는 알고리즘이 하드웨어 구현 측면에서 간단해야 한다는 점도 고려되어야 한다.

가. Window lookahead 알고리즘

IQ 스위치의 처리율은 window lookahead 알고리즘을 사용함으로써 향상시킬 수 있다[4]. 이 알고리즘에서는 output contention이 생길 경우 각 큐의 HOL 셀만을 검색하여 출력 포트에 전송할 셀을 선정하는 대신 각 큐의 전단으로부터 d개 만큼의 셀들의 출력주소를 검색하여(즉 window 크기 = d) 전송할 셀을 결정함으로써 HOL blocking을 완화하는 방법을 취하고 있다(그림 5). 셀들은 도착 순서대로 전송이 되지 않을 수 있으므로 별도의

FIFO (First In Random Out) 큐와 복잡한 제어 로직이 필요하며 따라서 실제적인 d의 값은 상대적으로 작을 수밖에 없다.

그림 5. Window lookahead 알고리즘



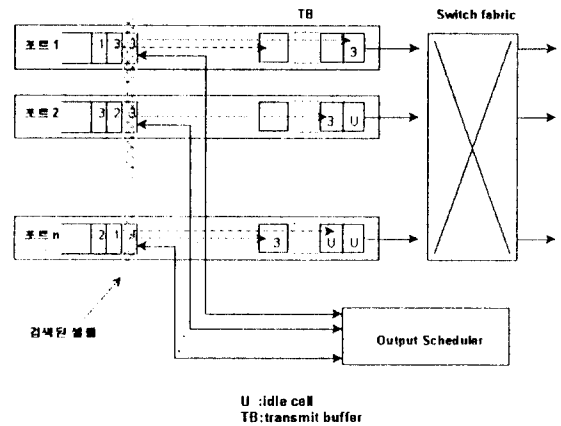
Simulation 결과에 의하면 Window lookahead 알고리즘은 균일 셀 트래픽에 대해 80% 정도의 처리율이 가능하나 실제적인 트래픽은 비균일하고 bursty한 성질을 가지므로 실제의 처리율은 이보다 떨어진다.

나. Output scheduling 알고리즘

FIFO 입력 큐는 output-scheduling 알고리즘[5]에 의해 스케줄링이 가능하다. Output-scheduling 알고리즘에서는 각 출력 포트에 대해 다음의 전송 가능한 시간 슬롯을 추적한다 (그림 6). 스케줄링 사이클동안 스케줄러는 각 입력 큐의 최전단 (HOL) 셀의 출력 주소를 일련적으로 검색하여 요구하는 출력 포트에 대해 전송가능한 슬롯 Ta를

지정해 준 다음, Ta를 1만큼 증가시키고 이 셀은 즉각적으로 상용하는 전송버퍼(TB: transmit buffer)로 이동이 된다. 지정된 전송 슬롯이 이미 다른 셀(출력 주소는 다름)에 의해 점유되어 있으면 이 셀은 현 사이클의 output contention에서 진 것으로 간주하고 FIFO 큐의 전단에 머물러 있게 된다. 모든 FIFO 큐의 전단 셀들은 동일 출력 포트에 향한다 하더라도 항상 차단(blocking)이 되지 않으므로 HOL blocking에 의한 처리율 열악화를 감소시키나 이 접근 방법에서 얻을 수 있는 최대 처리율은 70%로써 약간의 개선을 기대할 수 있다.

그림 6. Output scheduling



다. Non-FIFO 입력 큐잉

FIFO 원칙을 갖는 IQ 스위치에서의 HOL blocking 현상을 제거하고 QoS 기능을 구현하기 위한 방법으로는 각 입력단의 버퍼를 non-FIFO 큐로 구성하는 것이 있다. 그림 7(a)에 보이는 바와 같이 우선 순위별 큐의 구성은 QoS 기능 구현을 용이하게 한다. 그러나 단지 우선 순위에 의한 셀들

의 분류로서는 HOL blocking을 피할 수 없다.

다른 한편 VC별 (per-VC basis)로 큐를 구성하게 되면 각기 다른 VC에 속하는 셀들은 다른 출력 주소를 가질 확률이 높으므로 스위치의 처리율을 향상시킬 수 있다. 라운드 로빈 방식의 서비스 원칙을 사용하면 HOL blocking을 줄일 수 있고 VC 간에 공정하게 전송 대역폭을 할당시킬 수 있다. 또한 도착 셀들의 출력 주소별로 큐를 구성하면 HOL blocking을 완전하게 해소할 수 있다. 그러나 이 방법에서는 QoS 기능이 결여되어 있다.

그림 8. 3차원 큐잉.

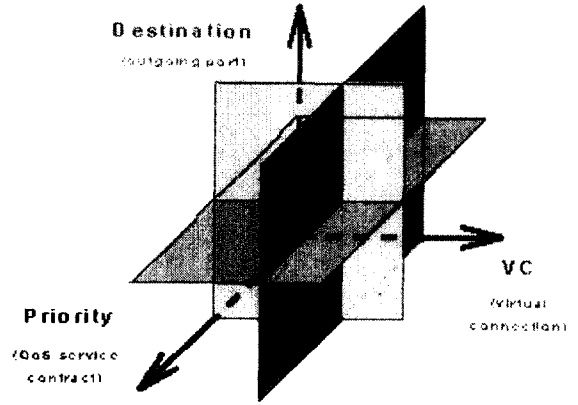
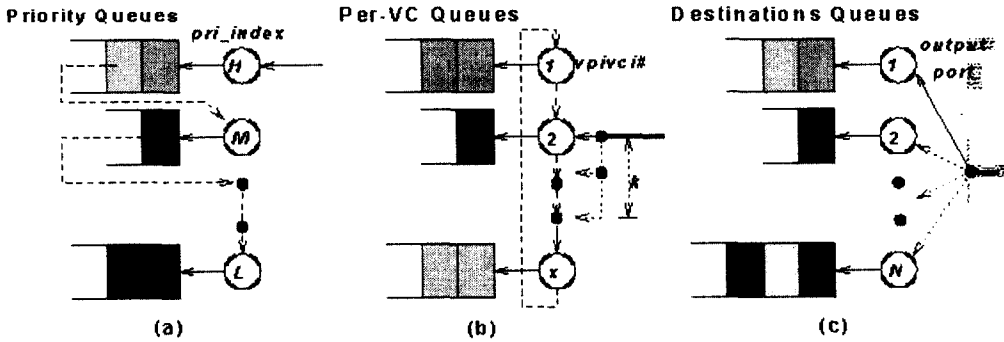


그림 7. Non-FIFO 입력 큐



이러한 점을 고려하여 참고문헌 [11]에서는 도착 셀들을 우선 순위, 출력 주소 및 VC에 의하여 큐를 구성하는 방안을 제시하고 있다 (그림 8).

라. 셀 스케줄링 알고리즘

IQ 스위치에서의 HOL blocking을 피하기 위한 하나의 방안으로서는 그림 9에서와 같이 각 입력 버퍼를 도착 셀들의 출력 주소에 따라 별도로 구성하는 방안을 생각할 수 있다. 출력 주소에 따른 각각의 큐를 가상 출력 큐라 부르기로 한다.

각 입력 포트의 가상 출력 큐중 어느 큐가 HOL 셀을 하나의 시간 슬롯동안 출력 포트에 전송할 수

있는지 결정하기 위해서는 셀 스케줄러가 필요하다. 포트당 OC-12(622 Mb/s)의 데이터 레이트를 가진 IQ 스위치에서 요구되는 반응 시간은 하나의 cell slot 길이에 해당하는 680ns 정도라야만 하며 data rate가 OC-48(2.5 Gb/s)로 증가할 때 요구되는 반응 시간은 170ns로 줄게 된다. 이 반응 시간에 대한 필요조건은 OC-192(10 Gb/s) traffic에 대해서는 42.5ns로 훨씬 줄어든다. 따라서 소프트웨어 기반의 셀 스케줄러는 아주 빠른 데이터 속도를 감당할수 없기 때문에 이러한 data rate에서는 적절하지 못하다.

그림 9. HOL blocking을 제거하기 위한 가상 출력큐의 구성

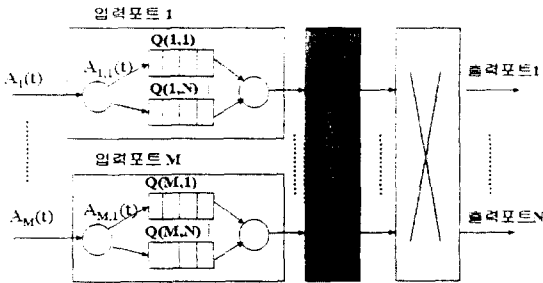


그림 9의 IQ 스위치를 고려해 보자. 이 스위치는 M개의 입력, N개의 출력, 비차단형 switch fabric, 그리고 스케줄러로 구성되어 있다. 각 입력 포트는 도착하는 셀을 저장하기 위한 N개의 가상 출력 큐를 갖는다. 출력 포트 j로 향하는 셀들의 입력 포트에서의 큐를 Q_{ij} 로 표시하기로 한다. 도착 셀들은 출력 주소에 따라 소팅이 되어 해당 큐에 저장되어 실제 전송이 되기까지 스케줄러의 결정에 따라 대기하게 된다.

대기중인 셀을 갖고 있는 모든 큐는 전송하기 위한 수락을 받기 위해 스케줄러에 스위칭에 관한 요청을 하게 되고 스케줄러는 스케줄링 알고리즘에 의하여 수락 여부를 결정한다. 이 스케줄링 문제에서의 제한 조건(constraints)은 다음과 같다:

첫째, 매 시간 슬롯동안 각 입력 포트로부터 최대 한 개의 셀이 원하는 출력으로 스위칭이 가능하고,

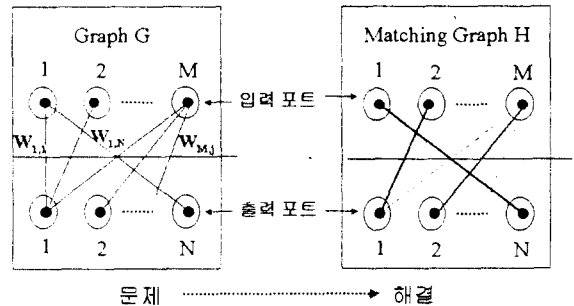
둘째, 매 시간 슬롯동안 하나의 출력 포트로는 최대 한 개의 셀만이 전송 가능하며,

셋째, M개의 입력 포트로부터 선택된 셀들의 출력 주소는 서로 다르다.

이러한 제한 조건하에서 스케줄링 알고리즘의

성능 목표는 최대 처리율을 얻는 것으로서, 하나의 시간 슬롯동안 출력 포트에 스위칭되는 셀의 개수가 최대가 되도록 셀들을 선택하는 것이다. 이러한 수락 여부 판정은 그림 10에 도시한 바와 같이 bipartite graph matching problem으로 귀결된다. bipartite graph에서 edge는 하나의 입력에서 출력으로의 스위칭 요구를 뜻하며 edge상의 숫자는 그 스위칭의 중요도를 나타내는 weight이다. Matching graph H는 주어진 스위칭 요구에 대한 하나의 해이며, H의 각 노드는 많아야 한 개의 edge에 연결되어 있음을 알 수 있다. 스케줄링이 완료되면 수락된 큐는 자신의 셀을 해당 출력포트로 전송할 수 있게 된다.

그림 10. Output scheduling problem의 예



앞에 언급된 제한 조건외에도 최적 스케줄러의 설계를 위해 고려해야 할 사항은:

첫째, 스위치에 유입되는 트래픽이 균일한 조건과 비균일한 조건하에서 셀 처리율이 최대가 되어야 하고,

둘째, 스위치내의 bandwidth는 동일한 priority(또는 QoS)를 갖는 VC간에 균등하게 할당되어야 한다.

이러한 필요조건을 만족하는 스케줄링 문제는 NP-complete로 알려져 있으며 스위치 용량이 커지는 경우 엄밀한 수학적 해를 구할 수 없다.

기존의 스케줄링 알고리즘으로서 신경 회로망을 사용하는 방법[6][8], Iterative searching을 이용한 iSLIP[9], LQF 알고리즘[10] 등을 들 수 있으나 신경회로망은 90% 이하의 처리율을 가지며 hardware 구현상의 complexity가 아주 높고, Iterative Searching의 경우 균일 트래픽에 대해서는 100%의 처리율을 보이나 비균일 트래픽에 대해서는 이에 미치지 못한다. 성능이 상대적으로 우월한 것으로 알려진 LQF 알고리즘의 경우 균일 및 비균일 트래픽에 대해서 100%의 처리율을 보이나 hardware complexity가 상당히 높고 (running time complexity $O(N^3 \log N)$), 한 개 이상의 큐에 대해 starvation 현상을 보인다는 단점이 있다.

IV. 맺음말

지금까지 날로 증대되고 있는 통신망에서의 전송 대역폭과 서비스 품질 기능에 대한 사용자 요구를 수용하기 위한 광대역 ATM 스위치 구현에 따르는 기술상의 문제점과 연구 동향을 살펴 보았다. 앞으로 지속적으로 연구되어야 할 부분은 특히 VC 단위의 서비스 품질 기능을 구현하기 위한 스위치내 버퍼 구성 및 버퍼 관리 기술 그리고 효율적인 셀 스케줄러의 개발 등을 들 수 있다.

참고 문헌

[1] S.C. Liew, "Performance of input buffered and output buffered ATM switches under bursty traffic: Simulation study," GLOBECOM, 90.
 [2] M.J. Karol, M. Hluchyj, and S.P. Morgan, "Input versus output queueing on a space-division packet switch," IEEE Transactions on

Communications COM-35 (1988).

[3] Y.-B. Kim and A.M. Makowski, "Large Size Asymptotics for Crossbar Switches with Input Queueing," In preparation for submission to Journal of Queueing Systems. Also available in Technical Report TR 95-66, Institute for Systems Research, Univ. of Maryland, College Park (MD).

[4] M. Mehmet-Ali, J.F. Hayes, A. Elhakeem, "Traffic analysis of a local area network with a star topology," IEEE Tr. on Comm., vol. 36, 1988.

[5] H. Obara, M. Sasagawa, I. Tokizawa, "An ATM cross-connection system for broadband transport networks based on virtual path concept," ICC, 1990.

[6] Ali, M., Nguyen, H. "A neural network implementation of an input access scheme in a high speed packet switch," GLOBECOM 1989.

[8] Brown, T., Liu, K. "Neural Network Design of a banyan network controller," JSAC, vol 8, 1990.

[9] McKeown, M. "Scheduling cells in an Input-Queued switch," IEE Electronic Letters, Dec 1993.

[10] McKeown, M. "Achieving 100% throughput in an Input-Queued switch," Ph.D. Thesis, Univ. of California, Berkeley, 1995.

[11] H. Duan, et. al., "A high-performance OC-12/OC-48 queue design prototype for input-buffered ATM switches", Proceedings of INFOCOM, 1997.



김 영 범

- 1984년 서울대학교 공과대학 전자공학과 졸업
- 1986년 서울대학교 전자공학과(석사)
- 1996년 미국 메릴랜드 주립대학 전자공학 박사
- 1986년 ~1988년 한국통신 품질보증단 전임연구원
- 1988년 ~1989년 과학기술대학 전기전자공학부 전임조교
- 1996년~ 1997년 TelPerion Network System, USA
- 1997년~ 현재 건국대학교 공과대학 전자공학과 조교수