실시간 이더넷을 주도하는 VoIP

Tor Skeie - ABB Corporate Research에 소속되어 있는 선임 과학자

전통적인 이더넷의 실시간 특징은 빈약하다. thin 이더넷(coax - 10BASE-2) 또는 허브(10BASE 또는 100BASE)에 기반을 둔 이더넷 네트워크를 통한 지점간 대기시간(latency)은 네트워크의 로드에 따라 결정된다. CSMA/CD (Carrier Sense Multiple Access - Collision Detection) 구조의 비 결정적 특징은 이더넷을 실시간 요건을 갖는 어플리케이션을 위한 통신 솔루션으로 적합치 않다고 반대하는 주요 쟁점이었다.

그러나 이러한 상황은 새로운 우선순위 기능을 갖춘 이더넷 스위치가 소개되면서 변했다. 스위치 이더넷에 기반을 둔 실시간 어플리케이션은 급속히 성장하고 있는 VOIP (Voice Over IP) 비즈니스가 주도하고 있는 이 기술을 활용할 수 있다.

전통적 이더넷(Traditional Ethernet)

전통적인 이더넷은 실시간 중심이 아니다. CSMA/CD 구조는 본질적으로 비결정적인 방식으로 액세스한다. thin 이더넷(coax 10BASE-2)이나 허브(10BASE 또는 100BASE)에 연결되는 이더넷 컨트롤러는 매체가 다른 패킷을 전송하고 있으면 특정 패킷을 전송할 수 없다. 이더넷이 유휴 상태에 있게 되는 즉시 이더넷 컨트롤러는 패킷을 전송할 수 있다.

전송 중에, 스테이션은 wire를 거치는 데이터를 계속 감시하여 통신이 성공적으로 이루어지도록 한다. 두 스테이션이 동시에 정보 전송을 시도하는 경우, 전송이 겹쳐져서 충돌이 발생하게 된다. 충돌이 발생할 경우, 전송하는 스테이션은 네트워크에서 간섭을 인식하고 잼(jam)이라고 하는 비트 시퀀스(bit sequence)를 전송한다. 잼은 전송 중인 다른 스테이션이 충돌 발생 사실을 인지하는 데 도움을 준다. 무작위 지연(random delay) 후, 스테이션은 정보 재전송을 시도하며 프로세스는 반복된다.

충돌 확률은 충돌 도메인(이더넷 범위)과 네트워크의 로드에 따라 결정된다. 이용율이 20%인 전통적 CSMA/CD 이더넷의 충돌은 0.1% 미만이며, 네트워크 이용율이 40% 이상이면 5%의 패킷이 충돌을 경험한다. CSMA/CD 네트워크의 이용율이 40%가 되면 문제를 경험하게 되며, 로드가 계속 증가하더라도 순수 데이터 비율은 충돌로 인해 감소한다. 그러나 이러한 충돌은 오류가 아님을 명심해야 한다. 충돌은 이더넷 네트워크의 정상적인 부분이다. 다음 그림은 CSMA/CD의 원리를 보여 주고 있다.

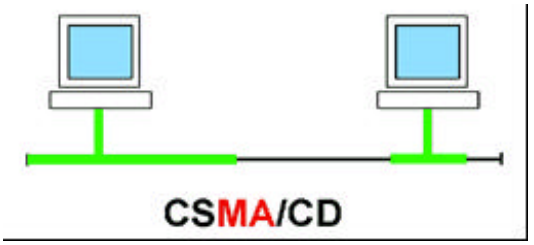


그림 2, 다중 액세스

그림 1, Carrier Sense

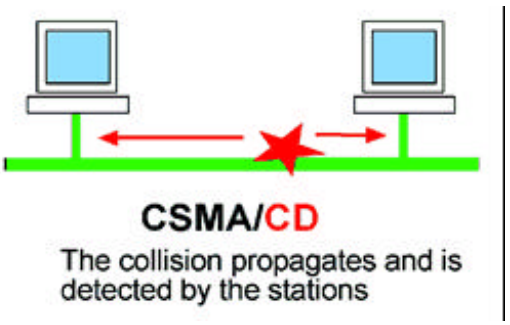


그림 3, 충돌 탐지

스위치 이더넷(Switched Ethernet)

기능적인 차원에서 볼 경우, 스위칭은 브리징과 완전히 같다. 그러나 스위치는 특수 설계된 하드웨어 ASIC(Application Specific Integrated Circuits)을 사용하여 브리지 기능과 패킷 전달(packet-forwarding) 기능을 수행한다. 이것은 중앙 CPU와 특수 소프트웨어를 사용한 구현과는 반대이다. 그 결과, 스위치는 브리지보다 훨씬 빠르다.

이더넷 스위치는 드롭 링크별로 10, 100, 1 Gbps 또는 10 Gbps를 제공하며, 10 Gbps는 현재 개발 중이다. 대역폭이 10 또는 100 Mbps이고 동일한 네트워크 세그먼트에 연결되어 있고 모든 사용자가 해당 대역폭을 공유하는 이더넷 허브와 비교해 볼 때, 이것은 확장성이 있는 거대한 대역폭의 증가를 의미한다.

또한 이더넷 스위치는 반이중(half duplex) 연결과 전이중(full duplex) 연결을 모두 제공한다. 이것은 전이중 연결이 사용될 경우 특정 이더넷 컨트롤러가 어떤 충돌도 경험하지 못한다는 사실을 의미한다.

- 전체 네트워크 로드는 스위치 엔진의 스위칭 기능을 초과한다. 즉 스위치는 각 드롭 링크의 전체 전선 속도를 처리할 수 없다.
- 출력 버퍼 용량이 충분하지 않다. 즉 출력 포트로 전송된 패킷량은 출력 버퍼가 처리할 수 있는 시간에 대한 해당 포트의 대역폭을 초과한다. 그러므로 여러 입력 포트의 패킷은 같은 출력 포트에 대해 경합을 벌이게 되고 비결정적인 버퍼링 지연이 발생한다. 스테이션의 보다 높은 수준의 프로토콜 레이어는 잃어버린 패킷을 처리해야 한다.

다음과 같은 이더넷 기법을 사용하여 이 두 시나리오를 방지할 수 있다.

- Back pressure : 해당 포트에서 수신한 패킷량이 스위치가 처리할 수 있는 양보다 많은 경우, 스위치는 반이중 모드로 가동되고 있는 포트에 트래픽을 시뮬레이트하는 잼 패턴을 전송할 수 있다.
- Flow control : 해당 포트에서 수신한 패킷량이 스위치가 처리할 수 있는 양보다 많은 경우, 스위치는 전이중 모드로 가동되고 있는 포트에 IEEE802.3x에 의거하여 PAUSE 패킷을 전송할 수 있다.
- 우선순위(Priority): 높은 우선순위 패킷으로 식별되는 이더넷 패킷은 높은 우선순위 대기열에 놓여진다. 높은 우선수위 대기열의 패킷은 낮은 우선순위 패킷 이전에 전송된다. 낮은 우선순위 패킷은 잃어버릴 수 있다. 대기시간에 민감한 실시간 데이터(latency sensitive real time data)의 최적 실시간 특징과 관련하여 이 기법은 가정 적절한 기법이다.

우선순위(Priority)

오늘날 이더넷 스위치는 포트별로 두 개 이상의 출력 대기열을 포함하는 우선순위를 지원할 수 있다. 여기서 높은 우선순위 대기열은 서비스 품질(QoS: Quality of Service)을 제공하는 실시간 중요 데이터(critical data)용으로 예약되어 있다. 우선순위 대기열 사이에서 스위치가 변경되는 방식은 공급업체마다 다르다. 두 개의 우선순위 대기열을 갖는 스위치의 변경 구조는 다음 중 한 구조를 취할 수 있다.

- 라운드 로빈 가중치(Round-robin weighting) : 하나의 패킷이 낮은 우선순위 대기열에서 전송되기전에 N 패킷이 높은 우선순위 대기열에서 전송된다.
- 엄격한 우선순위(Strict priority) : 모든 패킷이 높은 우선순위 대기열에서 전송된다. 낮은 우선순위 대기열의 패킷은 높은 우선순위 대기열이 비어 있는 경우에만 전송된다.

높은 우선순위 패킷이 출력 포트에 들어가기 전에 해당 패킷의 전송이 시작될 경우, 높은 우선순위 패킷은 낮은 우선순위 패킷 때문에 지연된다. 그럴 경우 높은 우선순위 패킷은 나머지 패킷을 지우는데 소요되는 시간만큼 지연된다. 최악의 경우는 최대 출력 길이(1518 바이트)를 갖는 이더넷 패킷의 전송이 막 시작된 경우이다. 100 Mbps일 경우의 추가 지연은 122 ms이며, 10 Mbps일 경우의 추가 지연은 1.22 ms이다.

또한 동일한 높은 우선순위 대기열에서 전송을 위해 이미 대기열에 놓여 있는 다른 실시간 패킷 때문에 높은 우선순위 패킷은 스위치를 통하여 지연된다. 그러나 종종 높은 우선순위 대기열을 사용하는 실시간 어플리케이션의 네트워크 로드와 트래픽 패턴이 알려져 있고 다른 모든 트래픽이 낮은 우선순위를 사용하는 경우, 그러한 패킷이 경험할 수 있는 최악의(worst-case) 스위치 대기시간은 쉽게 계산할 수 있다.

이러한 시스템에서 높은 우선순위 패킷의 최악의 스위치 대기시간은 일반적으로 드롭 링크별로 100 Mbps에 대해 수백 ms이다.

사례 :
- 전이중 연결의 100 Mbps는 모든 드롭 링크에 사용된다.
- 스위치는 저장 및 전달 스위치이며, 최소 스위치 대기시간은 10us이다.

스위치는 엄격 우선순위 일정(scheduling)을 사용한다.

실시간 패킷의 길이는 200 바이트이다. 여기에는 프리엠블(preamble), MAC, IP, UDP, 페이로드(payload), FCS (Frame Check Sequence) 및 최소 IPG(minimum Inter Packet-Gap)가 포함된다.
- 실시간 패킷은 높은 우선순위 패킷으로 취급된다. 다른 모든 패킷은 더 낮은 우선순위를 가진다.
- 최대 5개의 다른 스테이션은 해당 패킷이 대기열에 들어가기 전에 동일한 우선순위 대기열에 있을 수 있는 200 바이트의 비슷한 실시간 패킷을 발생시킬 수 있으며, 추가 스위치 지연을 야기시킨다.
- 모든 실시간 패킷은 순환 방식으로 발생된다. 실시간 패킷의 최악의 스위치 대기시간은 다음과 같다.

저장 및 전달(store-and-forwards): 16 us.

최소 스위치 대기시간(minimum switch latency): 10 us.

최대 패킷 길이를 갖는 패킷 비우기로 인한 최악의 대기시간: 122 us.

동일한 우선순위 대기열에 이미 놓여 있는 실시간 패킷 5개: 80 us,

합계: 228 us.

실시간 패킷의 최악의 대기시간은 더 낮은 우선순위를 갖는 기타 다른 네트워크 로드에 관계 없이 유효하다. 패킷을 높은 우선순위 패킷으로 식별하는 방식과 관련하여 여러 종류의 우선순위 구현이 존재한다. 우선순위 처리는 스위치 기능에 따라 다르다.

Layer 2 switch

레이어 2 스위치는 Ethernet MAC 대상 어드레스를 기반으로 하는 스위칭을 수행하며(그림 4 참조), 다음 사항을 기반으로 하는 우선순위 식별을 제공할 수 있다.

- MAC 어드레스(MAC address). 그림 4와 같이 MAC 원본 어드레스 및 대상 어드레스 모두 우선순위 식별에 사용할 수 있다. 사용자가 높은 우선순위 MAC 어드레스를 설정할 수 있으려면, 해당 스위치는 관리되는 스위치여야 한다. 이것은 그리 탄력적인 기능이 아니다.
- 이더넷 포트(Ethernet port). 스위치의 한 개 이상의 포트를 높은 우선순위를 위한 포트로 구성할 수 있다. 이것은 이들 포트에서 수신된 모든 패킷이 높은 우선순위 패킷으로 취급됨을 의미한다. 이 기능을 제공하는 스위치는 대부분 관리되는 스위치이다. 이 기능이 제공하는 이점은 제한적이지만, 이 기능은 지금까지 사용 가능한 유일한 우선순위 기능이었다.
- 우선순위 태깅(Priority tagging). IEEE 802.1p 및 IEEE 802.1Q 표준은 Ethernet MAC 헤더를 위한 추가 필드를 지정한다. 이 필드를 TCI(Tag Control Info - 태그 제어 정보) 필드라고 하며, 이 필드는 원본 MAC 어드레스와 이더넷 패킷의 MAC 유형/길이(MAC Type/Length) 필드 사이에 삽입된다(그림 5 참조). 이 필드에는 우선순위 처리에 사용되는 3비트 우선순위 필드가 포함되어 있다. 따라서 표준은 8개의 다른 우선순위 레벨을 정의한다. 그러나 우선순위 대기열 기능을 지원하는 이더넷 스위치 중에서 오늘날 사용 가능한 대부분의 이더넷 스위치는 2개 또는 4개의 대기열만을 가진다. 2개의 우선순위 대기열을 갖는 스위치는 우선순위 태그가 4개 이상으로 설정되어 있는 이더넷 패킷을 높은 우선순위 대기열에 놓으며, 다른 모든 패킷은 낮은 우선순위 대기열에 놓는다. 관리되지 않는 스위치와 관리되는 스위치 모두 이 기능을

그림 4, MAC헤더 (레이어 2)

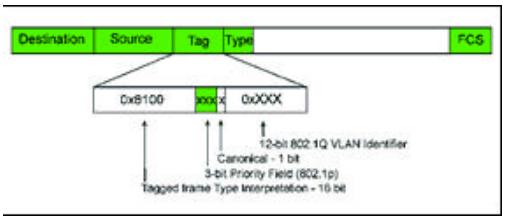


그림5, 태그가 있는 MAC헤더(레이어 2)

지원할 수 있다. 그러므로 스위치를 구성할 필요가 없다. 이 방법의 한 가지 단점은 지금까지의 거의 모든 스테이션이 우선순위 태깅을 지원하지 않는다는 점이다. 스위칭 후에 태그를 제거하도록 스위치를 구성하면 이 문제를 해결할 수 있다. 그러기 위해서는 관리되는 스위치를 가동해야 한다. 이 방법의 다른 단점은 네트워크에 우선순위 태깅을 지원하지 않는 다른 이더넷 스위치가 존재한다는

점이다. 즉 최대 패킷 크기는 태그 때문에 4 바이트씩 증가하여 최대 1522바이트로까지 증가할 것이며 일부 스위치는 패킷 길이가 1518바이트보다 큰 패킷을 전달하지 않을 것이다.

Layer 3 switch

레이어 3 스위치는 Ethernet MAC 대상 어드레스와 레이어 3 내용(라우터 기능)을 기반으로 하는 스위칭을 수행할 수 있다. 일례로 IP 패킷의 헤더 필드를 들 수 있다. 레이어 3 스위치는 레이어 2 스위치와 같은 범주를 기반으로 하는 우선순위 식별을 제공할 수 있다.

IP ToS(Type of Service-서비스 유형).

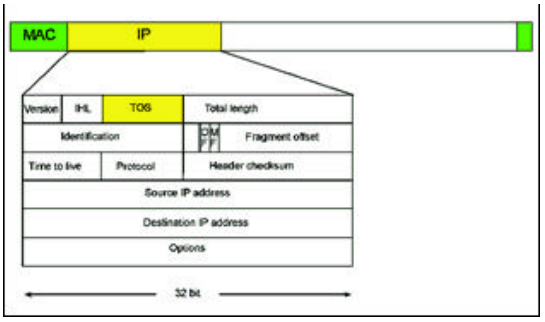


그림 6, IP 헤더 (레이어 3)

개별 IPv4 헤더에는 ToS 필드가 포함되어 있다(그림 6 참조). 차별화된 서비스(Differe-ntiated Service)로 알려진 최근의 표준은 ToS 필드를 두 개의 필드 즉 DSCP(6비트) 필드와 CU(2비트) 필드로 구분하고 있다. DSCP 필드는 필수 우선순위를 결정할 때 사용되는 필드이다. DSCP 필드의 6비트는 3개의 풀(pool)로 나누어져 있는 64개의 가능한 '코드 포인트(code point)'를 나타낸다.

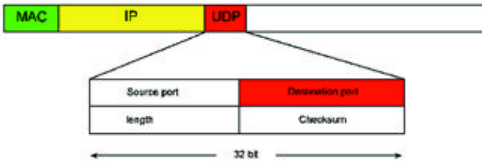
- Pool 1 DSCP = [0 .. 31] 표준 작업용(예: VOIP)
- Pool 2 DSCP = [32 .. 47] 실험용 또는 지역용. 그러나 향후 표준 작업용으로 할당 가능.
- Pool 3 DSCP = [48 .. 63] 실험용 또는 지역용.

가능한 64개 코드 포인트의 모든 하위집합은 스위치에서 높은 우선순위 식별 범주로 사용될 수 있다.

IP ToS 우선순위를 지원하는 스위치는 관리되지 않는 스위치이거나 관리되는 스위치이다. 관리되는 스위치일 경우, 높은 우선순위 코드 포인트는 대부분 사용자가 구성할 수 있다. 반면에 이와 상응하는 관리되지 않는 스위치를 위한 높은 우선순위 코드 포인트는 사전 구성되어야 한다. 사전 구성의 경우에는 스위치 구성이 필요 없다. 특정 비표준 IP를 기반으로 하는 실시간 어플리케이션을 위해 선호되고 있는 방법은 풀 3(Pool 3)의 코드 포인트이다.

실시간 중요 패킷의 IP ToS 필드를 위한 높은 우선순위 설정은 전송 스테이션의 IP 프로토콜에서 설정되어야 한다. TCP/UDP 소켓 레벨에서 setsockopt( ) 명령을 사용하여 대다수 운영체제의 클라이언트와 서버 소켓 사이드 모두에 대해 설정할 수 있다.

IPv6 헤더에는 트래픽 클래스(Traffic Class)라는 이름의 상응 필드가 포함되어 있다. 이 필드의 기능은 ToS 필드의 기능과 같다. 트래픽 클래스 8 진수는 ToS 필드가 IPv4 헤더에



가지고 있는 로케이션과 같은 로케이션을 IPv6 헤더에 가지고 있다.

Layer 4 switch

레이어 4 스위치는 Ethernet MAC 대상 어드레스(destination address)와 레이어 3과 레이어 4 내용 모두에 기반을 둔 스위칭을 수행할 수 있다. 일례로 TCP 패킷의 IP와 TCP 헤더를 들 수 있다. 레이어 4 스위치는 레이어 3 스위치와 같은 범주에 기반을 둔 우선순위 식별을 제공할 수 있다. 또한 다음 레이어 4 필드도 적절하다.

UDP 또는 TCP 대상 포트 번호.

UDP 또는 TCP 헤더의 대상 포트는 높은 우선순위 범주로 스위치에 또한 사용될 수 있다. 그림 7은 UDP 헤더를 위한 필드를 보여 주고 있다. 대부분의 레이어 4 스위치는 현재 관리되고 있으며 앞으로도 관리될 것이다. 그러므로 이 우선순위 기능은 사용자에 의한 스위치 구성이 필요하다.

특정의 실시간 애플리케이션에서 클라이언트와 서버 사이의 각 스위치는 해당 서버의 선택된 소켓 포트 번호에 대해 구성되어야 한다. 그러나 클라이언트의 소켓 포트 번호와 이에 상응하는 서버 소켓 포트 번호는 서로 달라야 함을 사용자는 명심해야 한다. 즉 UDP 패킷에서 대상 포트 번호와 원본 포트 번호는 달라야 한다. 실시간으로 중요 데이터를 서버에서 클라이언트로 전송하는 경우, 스위치는 두 포트 번호 모두에 대해 구성되어야 한다. 이것은 문제가 될 수도 있다. 왜냐하면 클라이언트 소켓 포트 번호는 대부분 동적으로 할당되기 때문이다.

결론

결정적 이더넷은 우선순위를 사용하여 달성된다. 스위치 이더넷 인프라를 통한 실시간 중요 데이터의 최악의 대기시간은 이 데이터가 우선순위로 보호될 경우 보장될 수 있다. 우선순위로 보호되는 패킷의 최악의 스위치 대기시간은 일반적으로 수백 분의 1초이다.
선호되는 방식은 IP ToS에 기반을 둔 스위치 우선순위 처리 방식이다. ToS 코드 포인트의 높은 우선순위 하위집합에 대해 미리 구성된 스위치는 사용자가 별도로 스위치를 구성할 필요가 없다. 이러한 스위치는 대부분 관리되지 않는 스위치이며, 비용 효율성이 가장 높은 방법이다. 그렇다면 IP ToS에 기반을 둔 우선순위 구성은 실시간 애플리케이션의 문제이다. 클라이언트와 서버 모두에 대해 소켓 레이어에서 쉽게 구성을 할 수 있다. 또한 IP ToS에 기반을 둔 우선순위는 이 기능을 지원하지 않는 다른 스위치나 스테이션에 대한 모든 충돌을 나타내지 않음에도 주의해야 한다. 일부 스위치와 스테이션이 태깅을 지원하지 않는 한, 이 사실은 태깅에 기반을 둔 우선순위(IEEE 802.1p)에 적용되지 않는다. 그러나 향후 대부분의 스위치와 NIC(Network Interface Cards) 그리고 OS가 이 기능을 지원할 경우, 우선순위 태깅은 적절한 대안이 될 수 있을 것이다.