김해중
KBS 보도기술국

실무 네트워크 Design 2 : Layer 2 Protocol Basic

김해중 KBS 보도기술국

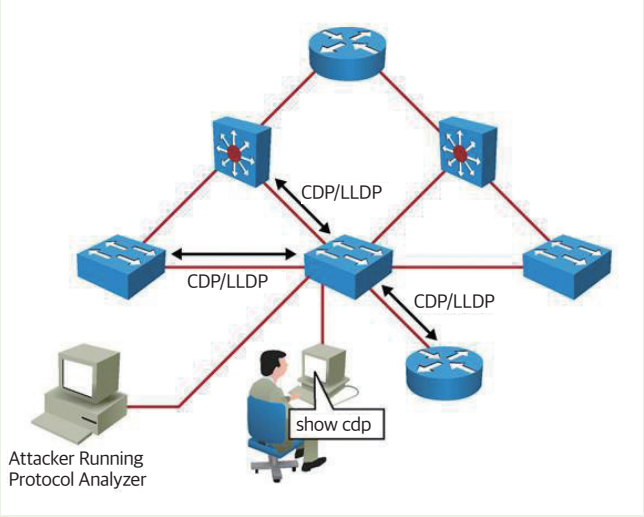
- 연재목록 -

이번 시간에는 OSI의 Layer 2(L2)에서 사용되는 protocol 중 NDP(neighbor discovery protocol)와 STP(spanning tree protocol), 이더 채널에 대해서 다루고자 합니다.

NDP(Neighbor discovery protocol)

개념

도면이 정확히 작성이 안 된 baseband 시스템을 대할 때, 처음 해보는 것은 해당 장비의 in, out이 어떤 장비랑 연결되어 있는지 파악하는 것이다. 그래서 일일이 케이블 연결을 해제하였다가, 다시 연결하는 방식을 택한다. 하지만 네트워크 기반 시스템은 자신이 어떤 네이버(이웃) 장비와 연결되어 있는지를 도면이 없이도 쉽게 파악을 할 수가 있으며, 그 역할을 하는 protocol을 NDP(neighbor discovery protocol)라고 한다. NDP의 종류로는 CDP(cisco discovery protocol), LLDP(link layer discovery protocol)가 존재한다. CDP는 cisco 사 고유의 protocol이고, LLDP는 IEEE 표준 기술이다. CDP는 60초마다 주기적으로 자신의 정보를 네이버에 전달한다.



[그림 1] 각 장비는 자신의 정보를 NDP를 통해 네이버 장비에게 전달한다


```
R1#sh cdp neighbor
Capability Codes: R - Router, T - Trans Bridge, B - Source Route Bridge
                  S - Switch, H - Host, I - IGMP, r - Repeater

Device ID        Local Intrfce     Holdtme    Capability  Platform  Port ID
R2               Fas 0/0           143           R        7206VXR   Fas 0/1
```

[그림 2] CDP를 통해 알게 된 정보

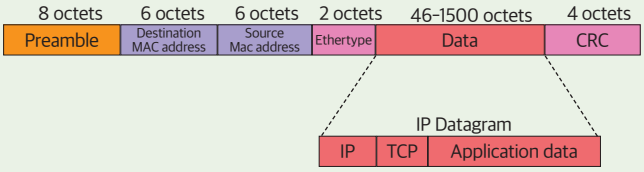
R1에서 (sh cdp neighbor) 명령을 입력하여, [그림 2]의 CDP 결과를 보면, R1의 f0/0(local interface)에 R2(device ID)의 f0/1 port(Port ID)가 연결되어 있으며, 네이버 장비의 기종은 (7206VXR)이며, 해당 장비는 R(Router) 역할로 동작함을 알 수 있다. 이 결과를 보면, 왼쪽 그림의 연결된 모습과 동일함을 알 수 있다.

고려 사항

NDP는 주기적으로 네이버에 자신의 정보를 전송하므로, 악의의 사용자가 네트워크망의 정보를 쉽게 알아버릴 수 있는 보안의 취약성이 존재한다. 그리고 실제 data가 전송되어야 하는 망에, 이런 background traffic이 존재해서 네트워크망의 효율이 떨어진다는 문제점이 발생한다. 최초 망 구축 시만 이용하고, 망이 구축된 후에는 되도록 NDP protocol을 사용하지 않는 것을 권장한다.

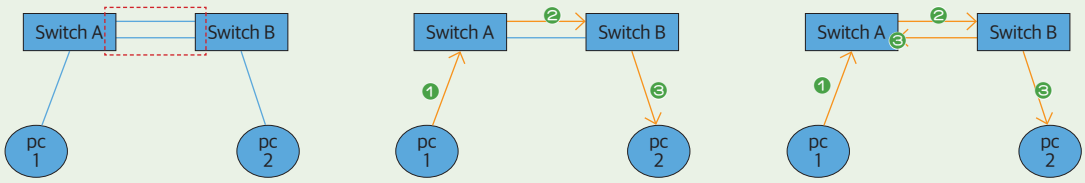
STP(Spanning Tree Protocol)

등장 배경 및 동작원리



[그림 3] Ethernet frame은 switch를 통과해도 frame이 변하지 않는다

Ethernet frame은 여러 대의 switch를 통과해도, frame 구조는 변하지 않는다. 즉 최초 출발지에서 전송한 frame과 목적지에 도착하게 된 frame이 동일한데, 이것을 transparent bridging이라고 한다. 마치 switch는 transparent 하게(없는 것처럼) 중간에서 동작을 하는데, 이런 동작 방식은 큰 문제점을 가지고 있다.



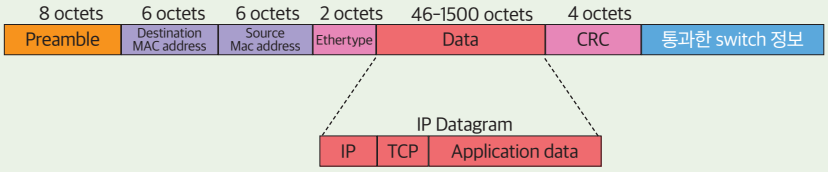
(a) Switch 간에 다중 경로 존재 시 (b) Unicast traffic 전송할 때 (C) Broadcast traffic 전송 시 문제점

[그림 4] switch 간에 다중경로 존재 시, broadcast traffic을 전송하면 loop가 발생한다

[그림 4(a)]처럼 switch A와 switch B 사이에 link 장애를 대비하여 다중 경로로 구현되어 있을 때, PC 1에서 PC 2로 보낸 unicast traffic은 원하는 traffic 경로인 [그림 4(b)]처럼 전달된다. 하지만 broadcast traffic을 전송하면 [그림 4(c)]처럼 목적지인 PC 2에도 패킷이 도착하지만, switch A와 switch B 사이로 프레임이 왔다 갔다 하는 loop 현상도 같이 발생한다. 왜냐하면 switch는 broadcast traffic을 받으면, 수신 port를 제외한 나머지 모든 port로 broadcast 프레임을 뿌리기 때문이다.

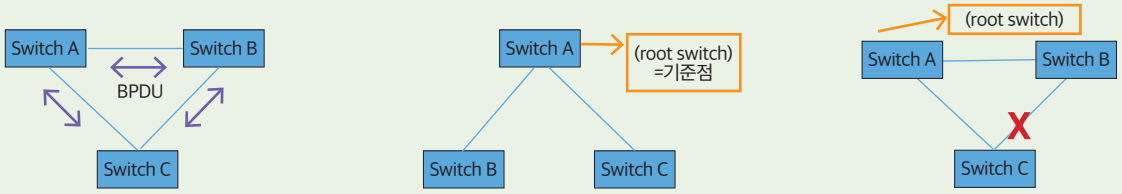
만약 switch가 transparent bridging으로 동작하지 않는다면, 즉 frame이 switch를 통과할 때마다, 통과한 switch 정보를 frame에 추가로 기록해 두는 공간이 frame header에 있다면, switch A가 전달한 frame이 다시 switch A로 돌아왔을 때, 해당 frame을 무시해 loop를 예방할 수 있을 것이다. 하지만 switch는 transparent 하게 동작하기에, 결국 loop를 방지하기 위해서 추가적인 protocol이 필요하게 되었는데, 바로 STP(spanning-tree protocol)이다. 즉, STP는 switch 간에 link 장애를 대비해서 만든 예비 경로가 있을 때, 오히려 link의 예비 경로로 인한 loop가 발생해, 이것을 방지하기 위해 출현하였다.

김해중
KBS 보도기술국



[그림 5] Ethernet frame이 위와 같은 구조였다면, STP는 필요 없을 것이다

STP은 어떻게 loop를 방지할까? STP는 loop를 방지하기 위해 tree 구조를 만든다. 마치 회사의 부장, 팀장, 팀원의 서열관계와 같이 계층 구조를 만들어서, 그 만들어진 tree 구조로만 traffic을 전달하고, tree에 해당하지 않는 경로는 사용하지 않는 원리이다.



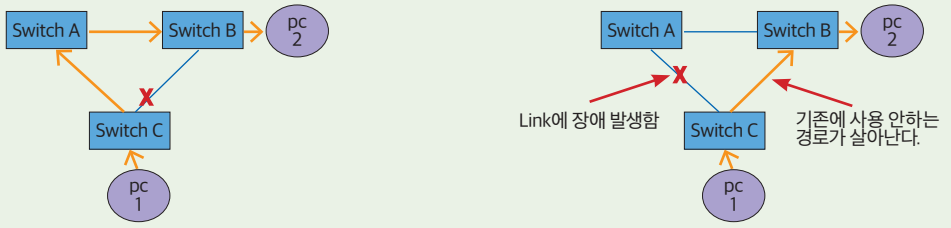
(a) Switch 간에 BPDU(명함)을 주고 받는다.　(b) root switch를 기준점으로 Tree를 그린다.　(c) STP로 인한 논리적인 결선

[그림 6] STP의 동작 원리

위의 [그림 6-(a)]을 보면 switch A에서 switch C로 가는 방법은 switch B를 경유하거나, switch C로 바로 가는 2가지 경로가 있기에 loop가 발생할 수 있다. 그래서 STP는 loop을 방지하기 위해서, [그림 6(a)]처럼 각 switch는 명함 역할을 하는 BPDU(Bridge protocol data unit)에 자신의 정보를 실어서, 다른 switch에게 전달한다. BPDU에는 누가 tree의 기준점인 root switch가 될 수 있는지에 대한 정보가 들어가 있다.

BPDU 교환 후, 1개 switch를 tree의 기준점인 root switch로 만든다. 관리자의 개입이 없을 때는, 가장 오래전에 만들어진 switch가 자동으로 root switch로 될 가능성이 있기에, 대부분의 회사는 관리자가 수동으로 root switch를 지정한다. 그래서 [그림 6(b)]처럼 switch A를 관리자가 root switch로 수동 세팅한 후, 나머지 switch B, switch C는 root switch인 switch A까지 최단 거리인 경로만 사용하는 tree 구조를 그린다.(단 이때 스위치간의 link는 동일한 속도라고 가정한다.)
대부분의 책에서는 복잡하게 tree를 만드는 과정을 설명하지만, 실무에서는 대부분 위의 그림과 같은 삼각형(triangle) 구조로 switch를 배치하기에, root switch가 아닌 각 switch에서 root switch로 직접 연결된 link만 active로 만들고, 나머지 link는 inactive로 만들면 빠르게 tree를 작성할 수 있다.
최종적으로 위 [그림 6(c)]처럼 실제 switch B와 switch C간의 경로는 link가 물리적으로 연결되어 있지만, tree 구조에 해당하지 않아서, 논리적으로는 끊어져 있는 방식으로 동작하게 된다.

그럼 실제로 data가 어떤 경로로 전달하는지 보다 자세히 살펴보기로 하자.



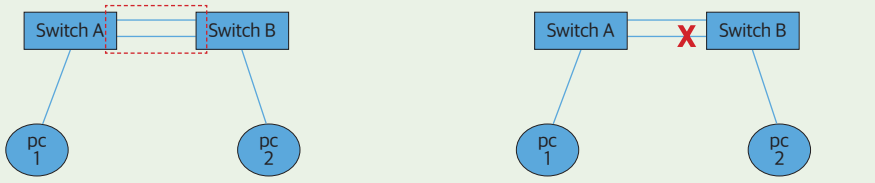
평상 시에 PC1에서 PC2로 전달되는 Data의 전달 경로　Switch A와 switch C 사이의 link 장애 시 data 전달 경로

[그림 7] STP Tree 완성 후의 data 전달 경로

switch C에 PC 1이 있고, switch B에 PC 2가 있을 때 실제로 data는 만들어진 tree 구조대로 전달되며, switch B와 switch C 간의 link는 평상시에는 사용되지 않다가, 장애 발생 시 사용하게 된다.

STP의 문제점(2가지)

STP는 특정 link를 논리적으로 끊어버려서 loop를 방지하는 기술이다 보니, 크게 2가지 문제점을 가지는데, 첫 번째는 loop를 방지하기 인해 발생한 차단된 link의 대역폭이 낭비된다는 문제점을 가지고 있다.



(a) 물리적인 결선 (b) STP 사용 후의 논리적인 결선(대역폭 낭비 발생)

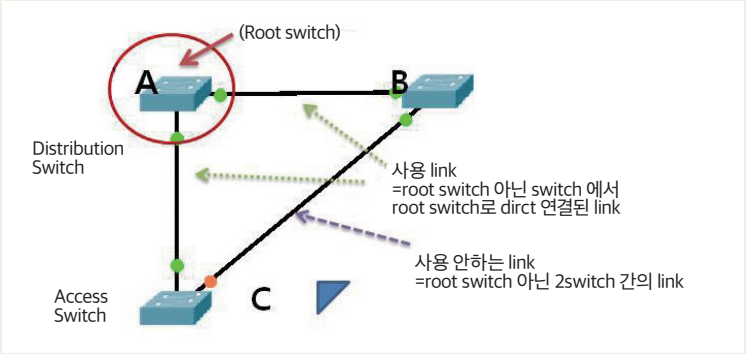
[그림 8] STP로 인한 대역폭 낭비 발생

두 번째는 STP는 tree 구조를 완성하는데 30 ~ 50초의 시간을 필요로 한다. 즉 link에 장애가 발생해서, 기존에 사용 안 하는 port가 살아나는 30~50초 동안은 네트워크 통신 두절이 발생한다는 치명적인 문제를 가지고 있다. 이러다 보니 실제 망에서 순수한 STP를 사용하는 곳은 거의 없다. 실제 망에서는 STP의 진보 기술인 RSTP(Rapid STP)가 가장 많이 사용되고 있다.

삼각형 구조의 Switching block 분석

1) triangle(삼각형) 1개 구조

실제 대부분의 회사는 일반적으로 삼각형(triangle) 구조로 switching block이 구성되어 있다. 그럼 누가 root switch가 되어야 할까? root switch는 distribution switch 중 1개를 tree의 기준점임 root switch로 하는 것을 권장한다. 그러면 이때 어떤 link가 사용되지 않을까? 당연히 기준점인 root switch로부터 가장 먼 link, 즉 switch B와 switch C사이가 사용되지 않을 것이다. 즉 root switch가 아닌 switch 간의 link가 사용되지 않는 직관적인 개념이다.



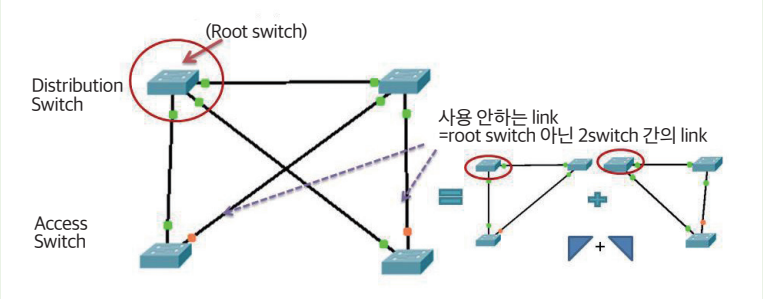
[그림 9] 삼각형 1개 구조의 switching block 분석

이것을 다르게 표현하면, access switch에서 2개의 distribution switch로 가는 link 중에서, root switch로 가는 link만을 사용하는 것이다. 참고로 [그림 9]의 주황색 동그라미로 표현된 port가 사용하지 않는 port이고, 녹색 동그라미로 표현된 port가 사용되는 port이다.

2) triangle(삼각형) 2개 구조

access switch가 1대 더 추가되어서, 삼각형 구조가 1개 더 추가되었다. 복잡하게 보이지만, 원리는 동일하다. 새로 만들어진 삼각형 구조에서 root switch에서 가장 먼 곳이 사용되지 않을 것이다. 실제로 대부분의 회사는 이러한 삼각형 구조의 switching block이 여러 개로 구성되어 있으며, 위의 원리를 이용하면 빨리 분석할 수 있을 것이다.

김해중
KBS 보도기술국

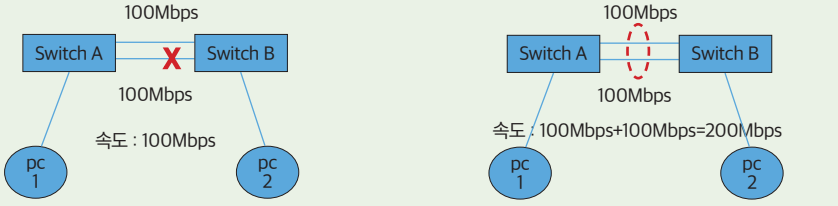


[그림 10] 삼각형 2개 구조의 switching block 분석

Etherchannel(이더 채널)

등장 배경

앞에서 다루었듯이, switch 간에 여러 개의 물리적인 link를 연결한다고 하더라도, STP로 인해 1개의 link만 사용이 가능하고, 나머지 link는 사용 못 하게 된다. 이런 link의 낭비를 해결하기 위해 여러 개의 물리적 link를 logical하게 1개로 보는 이더 채널 기술이 등장하였다. 이더 채널을 통해 대역폭 증가, load balancing의 효과를 얻을 수 있다.



(a) STP사용후 대역폭 낭비 발생　　(b) 이더 채널을 통해 논리적으로 1개 link로 간주

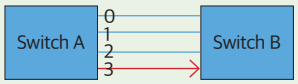
[그림 11] 이더 채널을 사용하면, 대역폭 증가 효과를 가져 온다

동작 방식

실제 이더 채널은 다양한 hash 방식을 통해 동작하는데, 가장 대표적인 방식인 src-dst-ip(source-destination-IP) 기반의 hash 방식의 동작을 설명해 보고자 한다. src-dst-ip 방식은 스위치 간의 연결 link가 2^n개 일 때, IP 주소 중에 last n bit를 XOR 연산하여, 사용하는 link를 선택하는 방식이다.

1) 10진수를 2진수로 변환
192.168. 1. 1(1을 2진수 표현:00000001)
172. 31 .67.46(46을 2진수 표현:00100110)
2) 2개 link 연결시: 끝에 1bit XOR연산
(1) XOR (0)= 1 ->10진수 변환(1)-> 1번 link사용

1) 10진수를 2진수로 변환
192.168. 1. 1 (1을 2진수 표현: 00000001)
172. 31 .67.46(46을 2진수 표현:00100110)
2) 4개 link 연결시: 끝에 2bi XOR연산
01) XOR (10)= 11 →10진수변환(3) -> 3번link사용

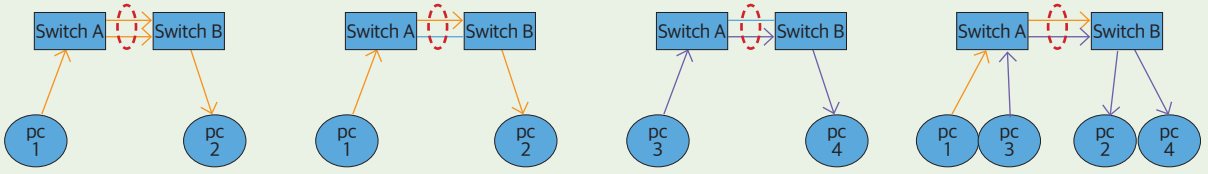


출발지 IP가 192.168.1.10이고, 목적지 IP가 172.31.67.46 패킷의 실제 전송경로

[그림 12] 이더 채널의 hash 방식(src-dst-ip)의 동작 원리

[그림 12]의 왼쪽 그림은 2개 link를 이더 채널로 사용하고 있으며, $2=2^1$ 이어서, IP 주소 중에 last 1 bit만 XOR 연산하면 된다. [그림 12]의 오른쪽 그림은 4개 link가 연결되어 있어 $4=2^2$이기에, IP 주소 중에 last 2 bit를 XOR 연산하면 된다. link 번호는 0번부터 시작하기에, 왼쪽 그림은 XOR 연산 결과가 1이 나와서, 1번 link를 사용하게 되고, 오른쪽 그림은 XOR 연산 결과가 3이 나와서, 제일 아래쪽 3번 link를 사용하게 된다.

많은 사람들이 이더 채널의 개념에 대해 오해를 하고 있다. 아래 그림을 보듯이 PC 1에서 보낸 패킷이 PC 2로 갈 때, 동시에 여러 개의 link를 같이 사용하는 것으로 오해를 한다. 하지만 PC 1에서 보낸 패킷은, 앞에서 설명한 hash 알고리즘을 통해 선택한 1가지 경로만을 사용할 뿐이다. 그리고 PC 3에서 보낸 패킷도 1가지 경로만을 사용할 뿐이다. 하지만 PC 1이 보낸 패킷과 PC 3가 보낸 패킷이 서로 다른 link를 사용하고 있어서, load-balancing 효과를 볼 수 있다.



(a) 이더 채널에 대한 일반적 오해(이렇게 동작 안 한다)

(b) PC 1에서 PC 2로 보낸 Data는 1경로만 사용한다.

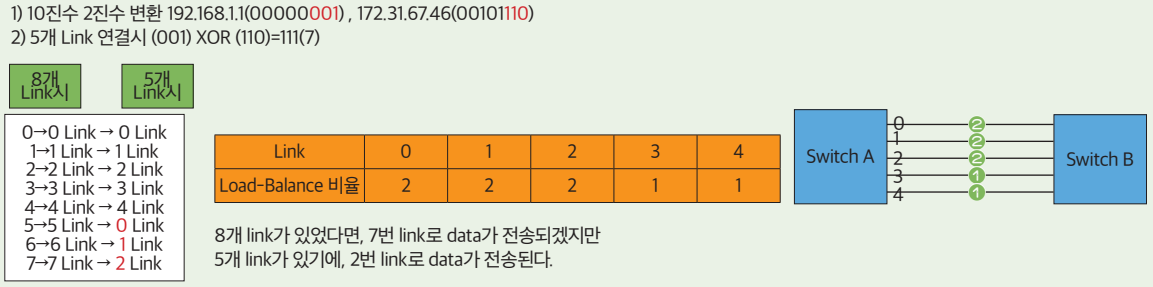
(c) PC3에서 PC4로 보낸 Data는 1경로만 사용한다.

(d) 최종 traffic 흐름

[그림 13] 이더 채널의 load-balancing 원리

설계 권장 사항

이더 채널을 설계할 때는 link수가 2^n이 되도록 설계하는 것을 권장한다. 아래 그림과 같이 5개 link일 경우는, 모든 link에 균등하게 load-balancing이 되지 않기 때문이다. 아래 그림에 보여 주듯이, 8개 link가 있다면 1:1:1:1:1:1:1:1로 분산되지만, 5개 link가 있어서 2:2:2:1:1로 부하가 분산되어서, 특정 link에 더 많은 traffic이 흐르게 된다.



(a) 5개 link일 때의 Ether-channel load-balancing 동작 원리

(b) 5개 link 사용할 때의 Load-balancing 비율

[그림 14] 이더 채널 사용시 2^n개의 link를 사용하는 것을 권장한다

정리

이번 시간에 학습한 내용을 아래 표에 정리해 보았다.

L2 Protocol	역할
NDP	어떤 장비(네이버)와 연결되어 있는지 알게 해준다.
STP	목적지까지 도달하는 여러 경로를 가질 때, loop를 방지해주는 기술이다.
이더 채널	switch 간에 다중 link로 연결되어 있을 때, 연결된 모든 link를 사용할 수 있게 해서, 대역폭 향상 및 load-balancing을 가능하게 한다.

L2 protocol 중에 핵심은 STP이다. 하지만 STP는 대역폭 낭비와 장애 발생 후 복구되는데 30~50초의 시간이 소요된다는 2가지 문제점을 가지고 있다. 그래서 이 문제점을 해결하기 위해서 다양한 기술들이 출시되었다.

다음 시간에는 STP의 대체 기술들과 그 기술들이 실제로 망에 어떻게 적용되어 있는지 다루도록 하겠습니다.