Netmanias.com
Analyze Trends, Technologies and Market
www.netmanias.com

NMC Consulting Group
www.nmcgroups.com

BGP(Border Gateway Protocol) 상세 동작 원리

2010년 7월 10일

NMC Consulting Group (tech@netmanias.com)

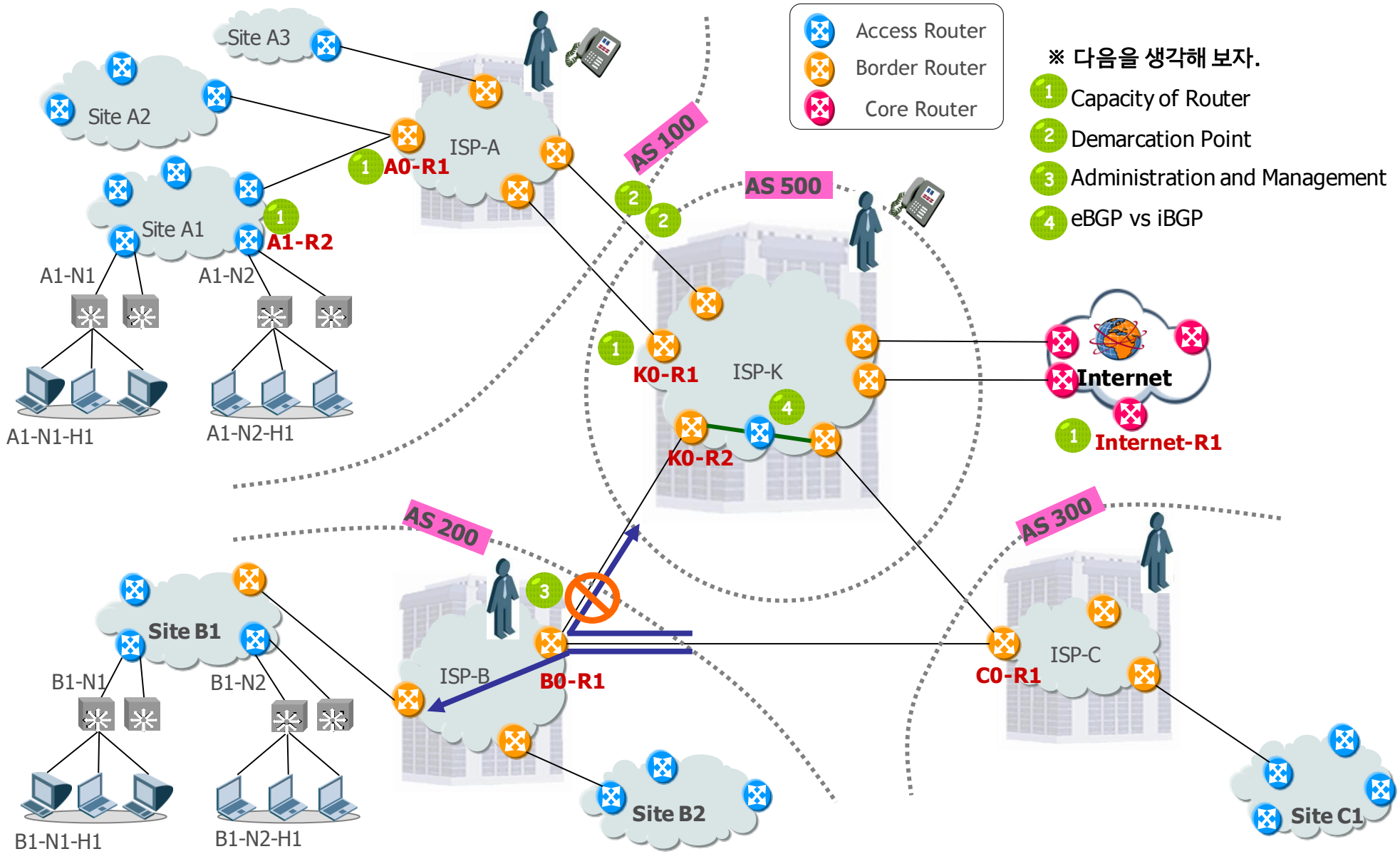
About NMC Consulting Group

NMC Consulting Group was founded on year 2002 and is advanced, professional network consulting company which is specialized for IP Network area like FTTH, Metro Ethernet and IP/MPLS, Service area like IPTV and IMS lastly, Wireless network area like Mobile WIMAX and LTE.

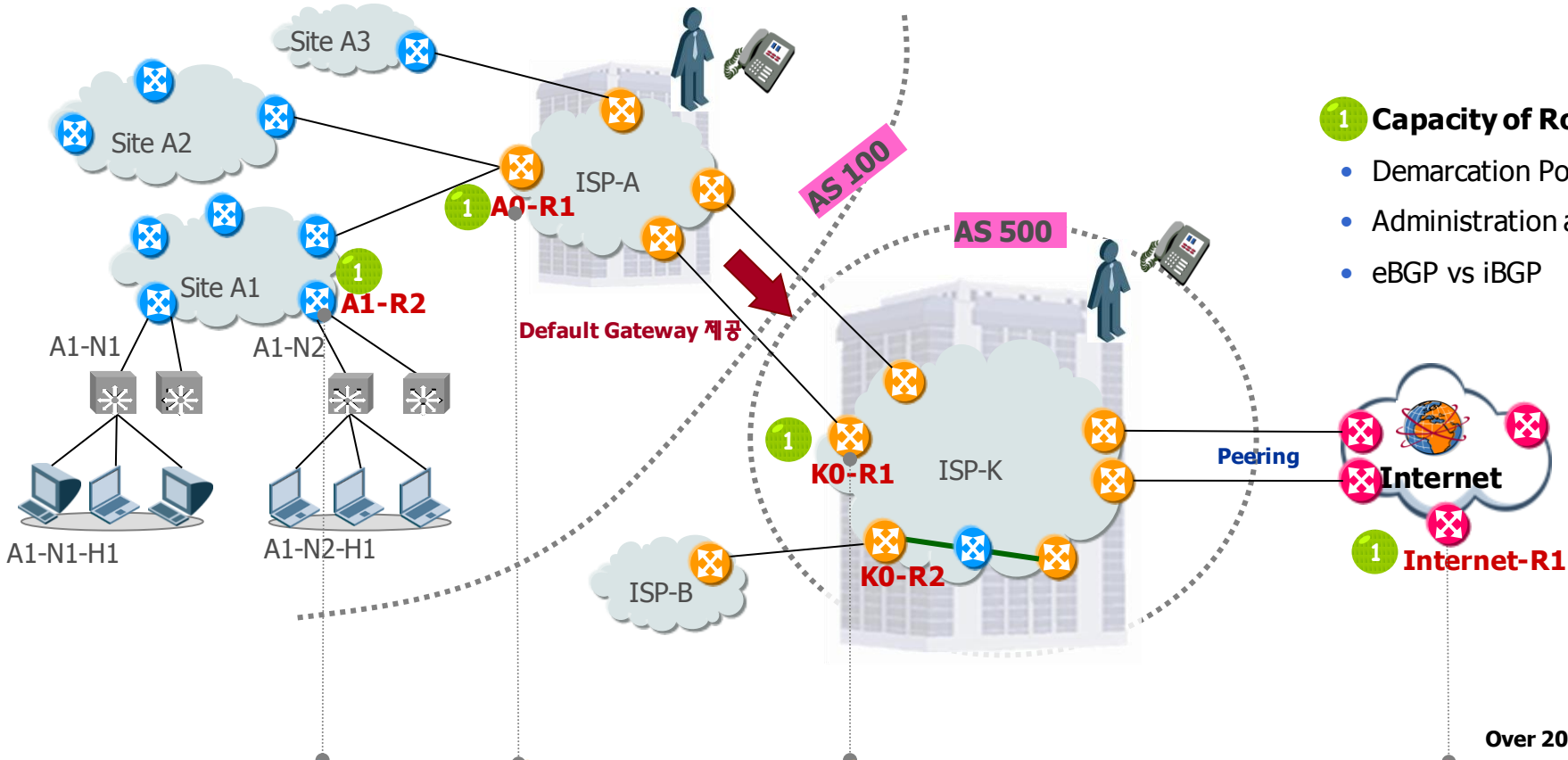
Contents

1. BGP Overview

1.1 BGP Overview (1/5)



1.1 BGP Overview – Capacity (2/5)



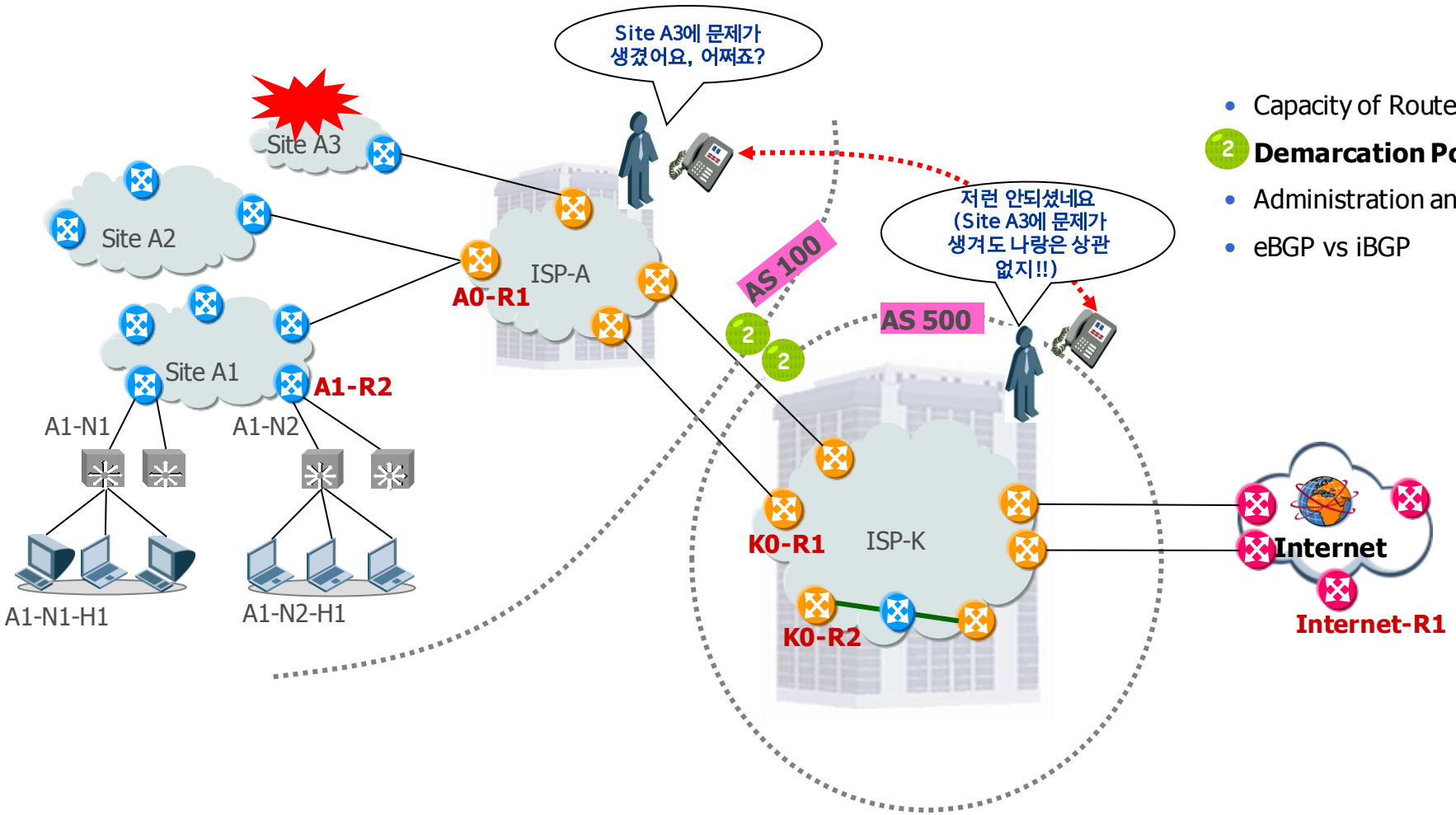
- **Capacity of Router**
  - Demarcation Point
  - Administration and Management
  - eBGP vs iBGP

Over 20K Routes

Routing Table Entry

A1-R2	A0-R1	K0-R1		Internet-R1		
A1-N1	A1-N1	A1-N1	B1-N1	A1-N1	B1-N1	Z1-N1
A1-N2	A1-N2	A1-N2	B1-N2	A1-N2	B1-N2	Z1-N2
	A2-N1	A2-N1	B2-N1	A2-N1	B2-N1.....	Z2-N1
	A2-N2	A2-N2	B2-N2	A2-N2	B2-N2	Z2-N2
	A3-N1	A3-N1	B3-N1	A3-N1	B3-N1	Z3-N1
	A3-N2	A3-N2	B3-N2	A3-N2	B3-N2	Z3-N2

1.1 BGP Overview – Demarcation Point (3/5)



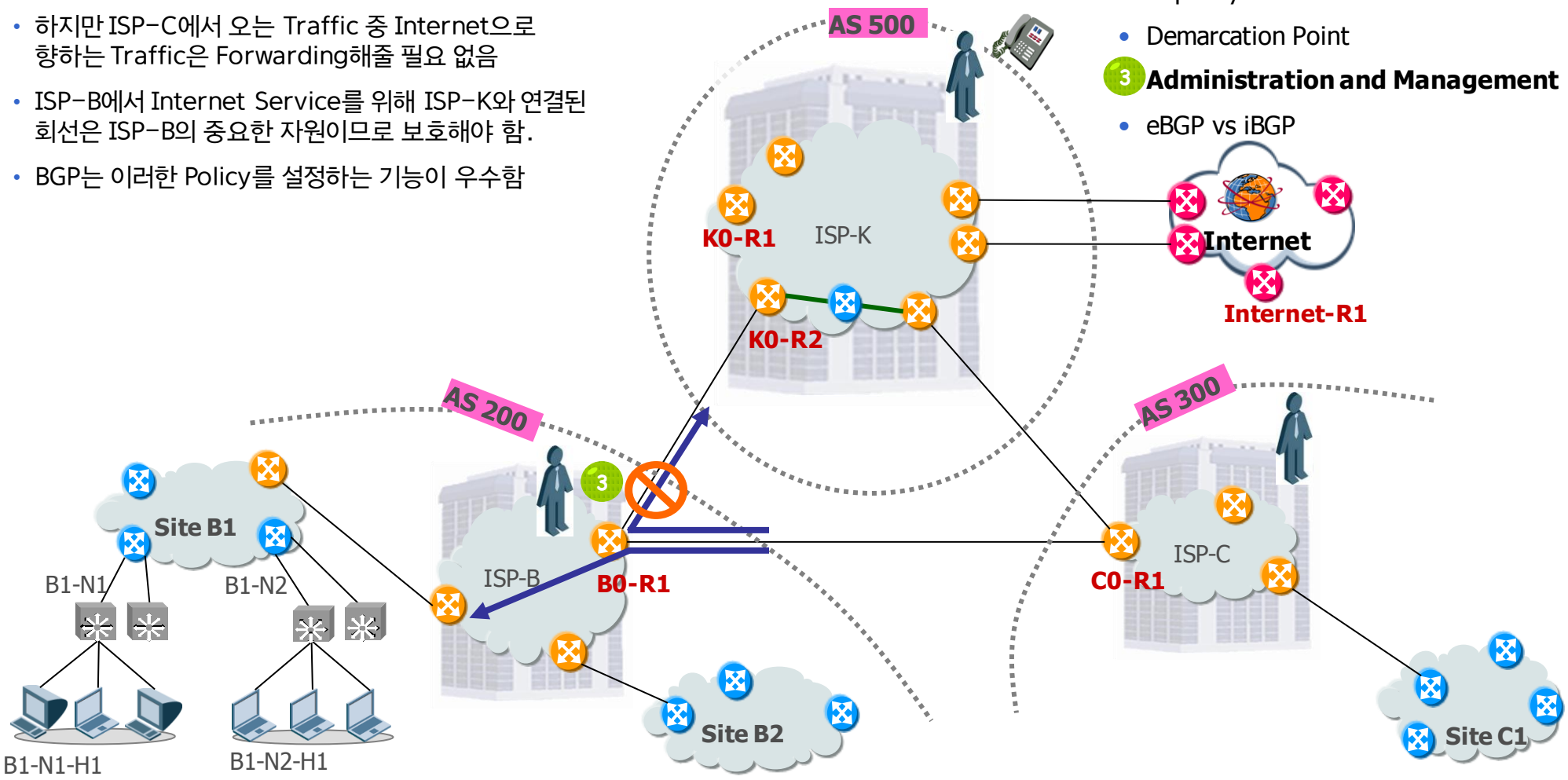
만약에 만약에, AS간에 OSPF로 연동했다면 → AS 내부의 Topology Change가 LSA Flooding을 통해 다른 AS로 전달 → 타 AS내의 SPF Calculation을 유발 → 서로 영향을 끼침
반면, BGP는? → Topology 정보(Link state info)없이 Network Prefix만 AS 간에 전달

1.1 BGP Overview – Administration (4/5)

※ ISP-B의 Site B1과 ISP-C의 Site C1 사이에 서비스를 제공하기 위해 B0-R1과 C0-R1 Border Router끼리 연결을 협약함

- 하지만 ISP-C에서 오는 Traffic 중 Internet으로 향하는 Traffic은 Forwarding해줄 필요 없음
- ISP-B에서 Internet Service를 위해 ISP-K와 연결된 회선은 ISP-B의 중요한 자원이므로 보호해야 함.
- BGP는 이러한 Policy를 설정하는 기능이 우수함

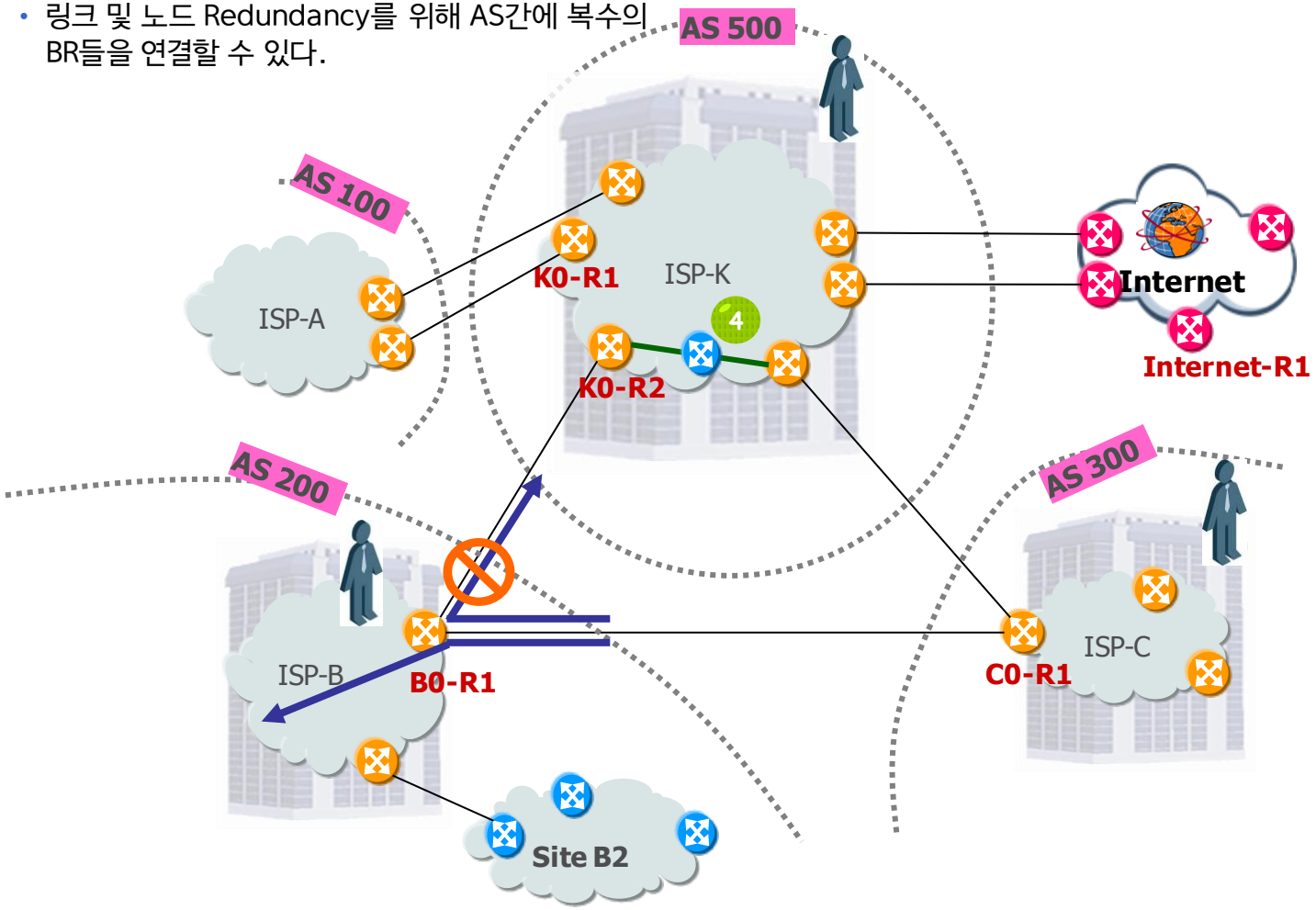
- Capacity of Router
- Demarcation Point
- 3 **Administration and Management**
- eBGP vs iBGP



1.1 BGP Overview – eBGP vs iBGP (5/5)

※ eBGP는 서로 다른 AS간의 연결을 위해 사용

- 서로 다른 AS간에는 일반적으로 IGP를 사용하지 않기 때문에 Connected Network을 사용하여 Neighbor를 설정한다.
- 링크 및 노드 Redundancy를 위해 AS간에 복수의 BR들을 연결할 수 있다.



- Capacity of Router
- Demarcation Point
- Administration and Management

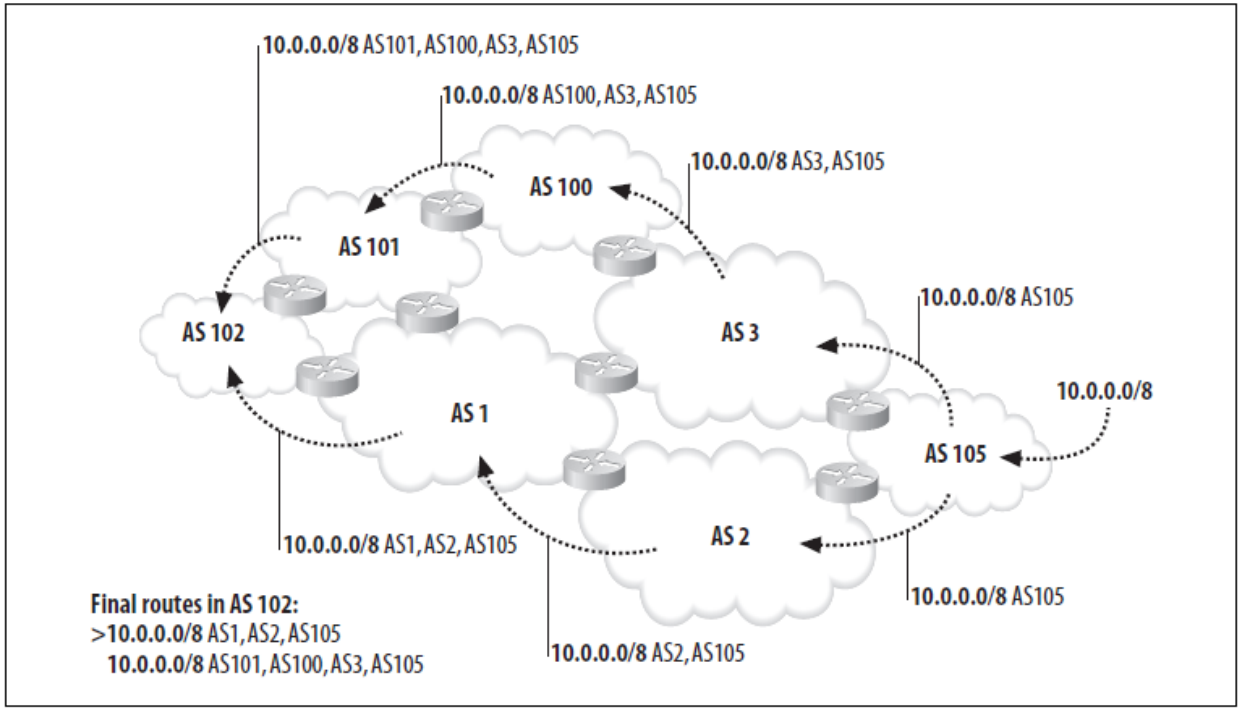
4 **eBGP vs iBGP**

※ iBGP는 동일한 AS내에 있는 Border Router간의 Routing Table의 동기를 맞추기 위해 사용

- 동일한 AS 안에서는 일반적으로 IGP를 사용하기 때문에 Loopback 주소등을 사용하여 Neighbor를 설정 (Redundancy)
- iBGP Neighbor간에 여러 Hop이 존재할 수 있음
- iBGP로 전달받은 Route들은 다른 iBGP Neighbor에게 광고하지 않는다.

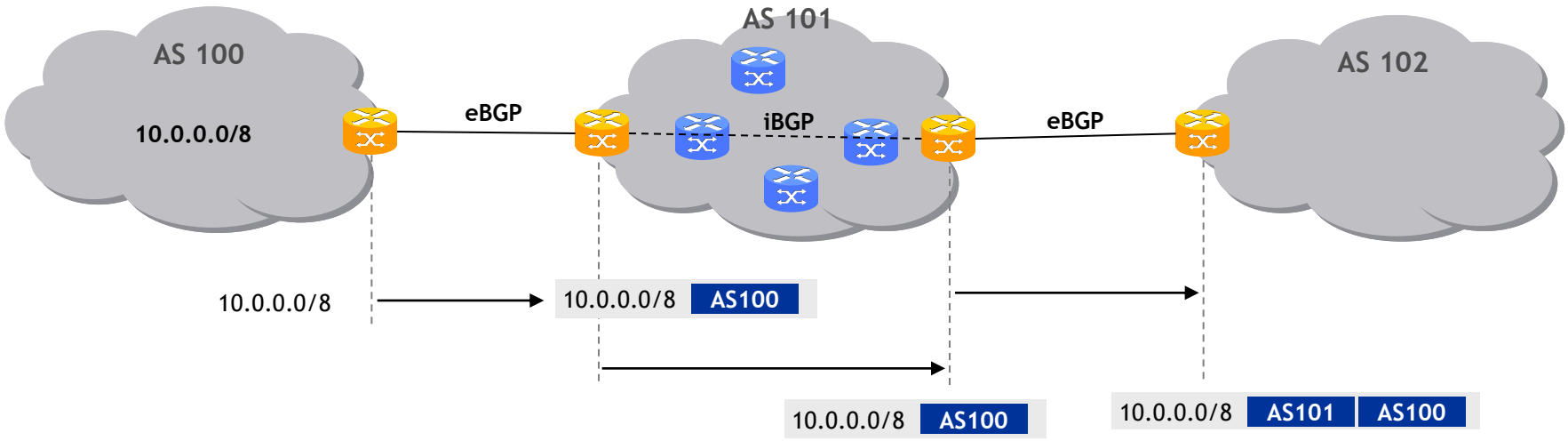
1.2 Routing Algorithm of BGP

- Route entry가 AS들 간에 교환됨
- 각 Route entry는 하나의 AS를 거칠 때 마다 AS number가 덧붙여짐
- 이 AS-Path의 길이가 짧은 경로를 Best Path로 선택하여 Routing Table에 install함
- 이상은 기본 Path-Vector algorithm에 대한 설명일 뿐, 실제로는 다양한 parameter가 정의되어 있어 AS-Path 이외의 많은 사항을 고려하여 Best Path를 판정함



1.3 eBGP vs. iBGP

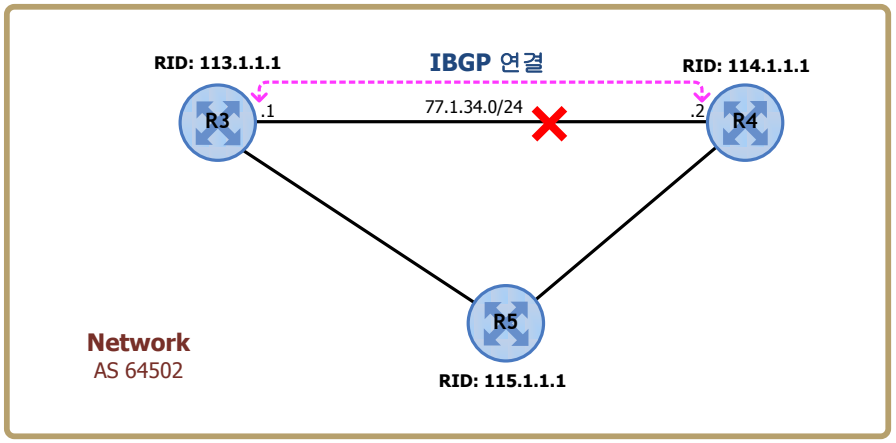
- 서로 다른 AS간의 BGP session → external BGP
- 동일 AS 내의 BGP router 간의 BGP session → internal BGP



2. BGP Parameter

2.1 BGP Session Type – IBGP 연결

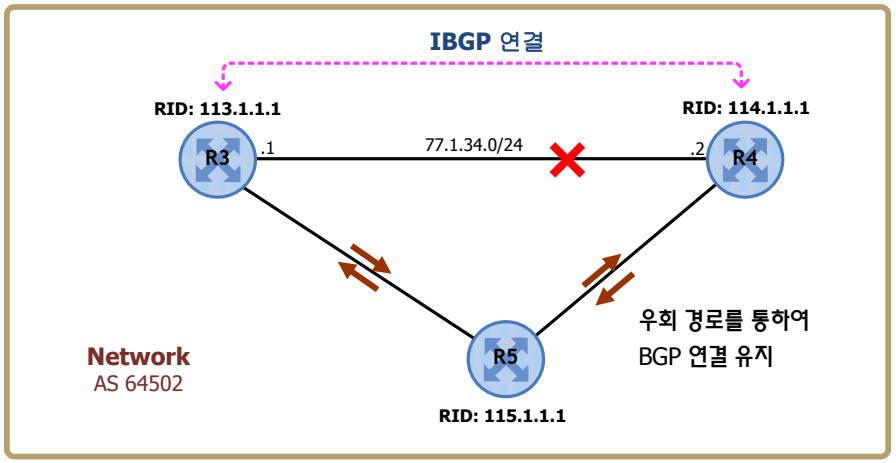
- **IBGP 연결**
  - 그림과 같이 동일한 AS 64502에 있는 R3와 R4 사이의 연결을 Internal BGP(IBGP) 연결이라고 한다.
  - BGP 연결에 사용하는 Neighbor 주소는 Physical Interface 주소와 Logical (Loop-back) Interface 주소를 사용할 수 있다.



Neighbor 주소로 Physical 주소를 사용한 IBGP 연결

- 다음은 Physical 주소를 이용해 Neighbor를 맺은 것이다.

```
R3(config-router)# neighbor 77.1.34.2 remote-as 64502
-------------------------------------------------
R4(config-router)# neighbor 77.1.34.1 remote-as 64502
```



Neighbor 주소로 Loopback 주소를 사용한 IBGP 연결

- 다음은 Loop-back 주소를 이용해 Neighbor를 맺은 것이다.

```
R3(config-router)# neighbor 114.1.1.1 remote-as 64502
R3(config-router)# neighbor update-source lo0
-------------------------------------------------
R4(config-router)# neighbor 113.1.1.1 remote-as 64502
R4(config-router)# neighbor update-source lo0
```

- Loop-back 주소를 사용하면, R3와 R4의 Link가 Down이 되어도 우회 경로가 있다면, IBGP 연결이 유지된다 (권장).

  (주: Loop-back Interface: Physical Interface가 아닌 Logical 개념의 Interface, Router가 Down 되지 않는 한 항상 UP 상태유지)

2.1 BGP Session Type – EBGP 연결

그림 A

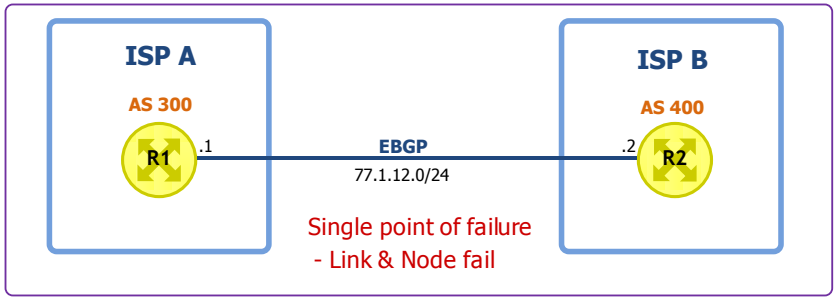
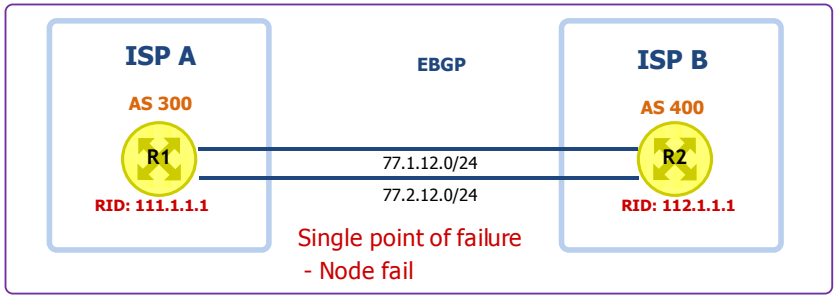


그림 B



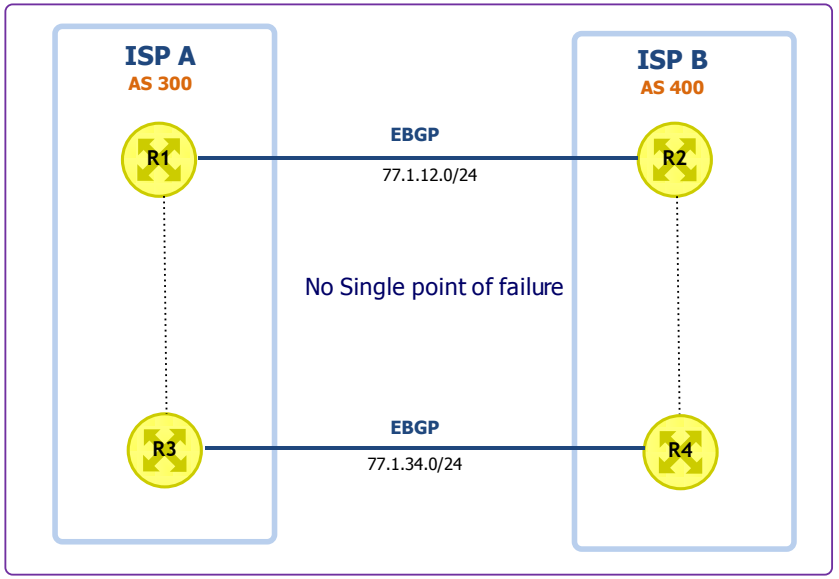


그림 C

EBGP 연결

- 그림과 같이 서로 다른 AS 300과 400 사이의 연결을 External BGP(EBGP) 연결이라고 한다.
- EBGP 연결은 주로 사업자 간 이루어지기 때문에 Peer-to-Peer 구성이 일반적이다 (그림 A).
- 우회 경로가 존재하는 것이 아니므로 보통 Loop-back Interface가 아닌 인접한 실제 Physical Interface 주소를 사용한다. 그러나, 이 경우는 Link failure/ Node failure의 문제가 있다.
- Redundancy를 제공하는 방법으로 그림 B와 그림 C 방법을 사용할 수 있다. 보통은 그림 C가 일반적이다.

〈 Physical 주소를 사용하는 경우(권장) 그림 A, C 〉

```
R1(config-router)# neighbor 77.1.12.2 remote-as 400
-------------------------------------------------
R2(config-router)# neighbor 77.1.12.1 remote-as 300
```

〈 Loop-back 주소를 사용하는 경우 그림 B 〉

```
R1(config-router)# neighbor 112.1.1.1 remote-as 400
R1(config-router)# neighbor 112.1.1.1 update-source lo0
R1(config-router)# neighbor 112.1.1.1 ebgp-multihop 2
-----------------------------------------------------
R2(config-router)# neighbor 111.1.1.1 remote-as 300
R2(config-router)# neighbor 111.1.1.1 update-source lo0
R2(config-router)# neighbor 111.1.1.1 ebgp-multihop 2
```

2.2 BGP 상태변화

[CLI 로 본 상태변화]

Idle 상태
```
R1# show ip bgp summary
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
112.1.1.1       4    400       4       4        1    0    0 00:00:02     Idle
```

Active 상태
```
R1# show ip bgp summary
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
112.1.1.1       4    400       4       4        1    0    0 00:00:02    Active
```

Connect 상태
```
R1# show ip bgp summary
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
112.1.1.1       4    400       4       4        1    0    0 00:00:02    Connect
```

Open Sent 상태
```
R1# show ip bgp summary
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
112.1.1.1       4    400       4       4        1    0    0 00:00:02  Open Sent
```

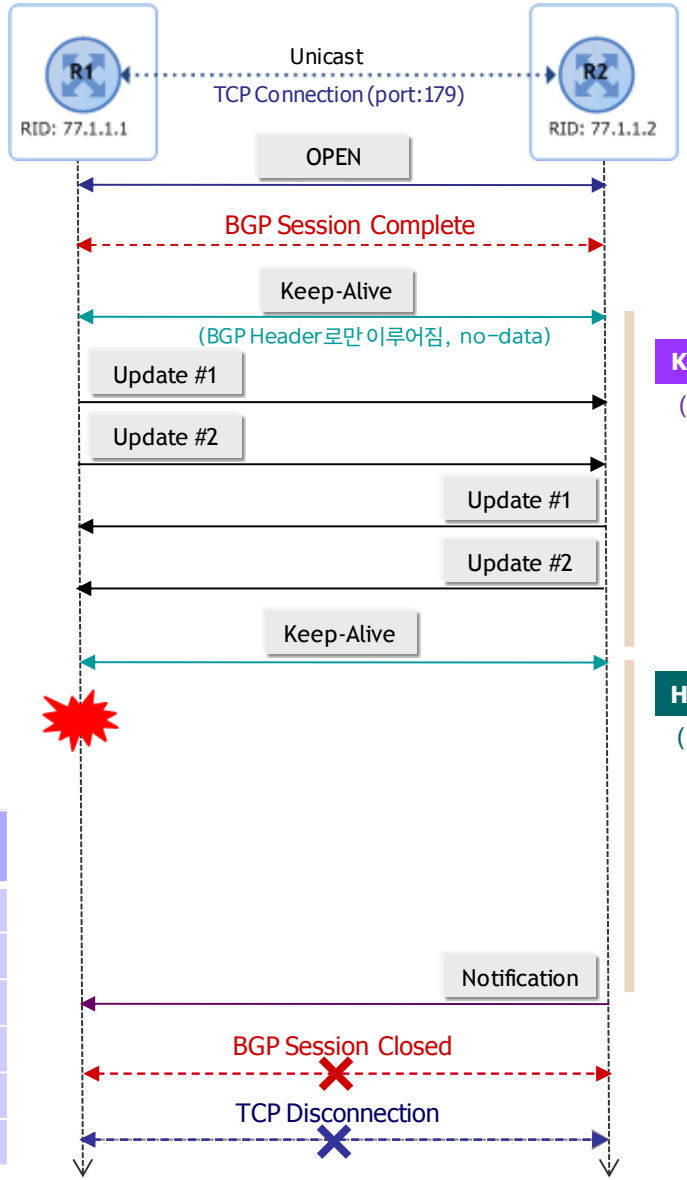
Open Confirm 상태
```
R1# show ip bgp summary
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
112.1.1.1       4    400       4       4        1    0    0 00:00:02 Open Confirm
```

Established 상태
```
R1# show ip bgp summary
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
112.1.1.1       4    400       4       4        1    0    0 00:00:02        0
```

실제로 상태 변화가 빠르게 진행되어 관찰하기 어려움

Cisco/ZebOS에서는 Established 상태가 숫자(0)으로 표시됨.

2.3 BGP Message 송수신 절차



Keep-Alive Time

(Default 60초)

자신이 살아있다는 Message를 Keep-Alive 주기로 BGP neighbor에게 보낸다.
보통, Hold-Time의 1/3로 설정된다.

```
R(config-router)#timers bgp 60 180
```

Hold-Down Timer

(Default 180초)

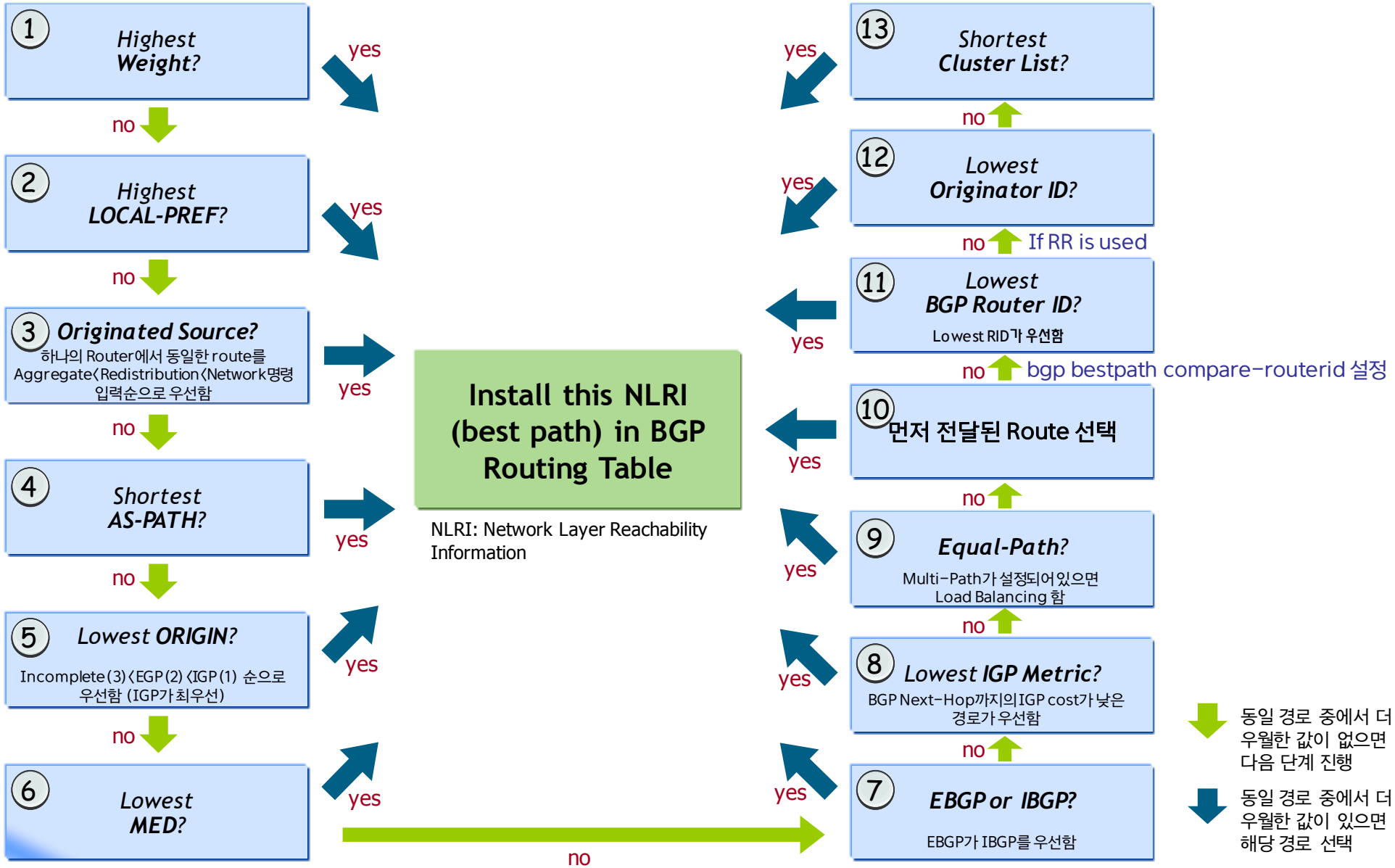
BGP neighbor에서 이 시간 동안 Keep-Alive나 Update Message가 오지 않으면, 상대편 Peer가 Down 되었다고 판단한다.
이 값은 Peer간 설정된 값이 같지 않아도 된다
→ Minimal 값으로 Adjust 됨

```
R(config-router)#timers bgp 60 180
```

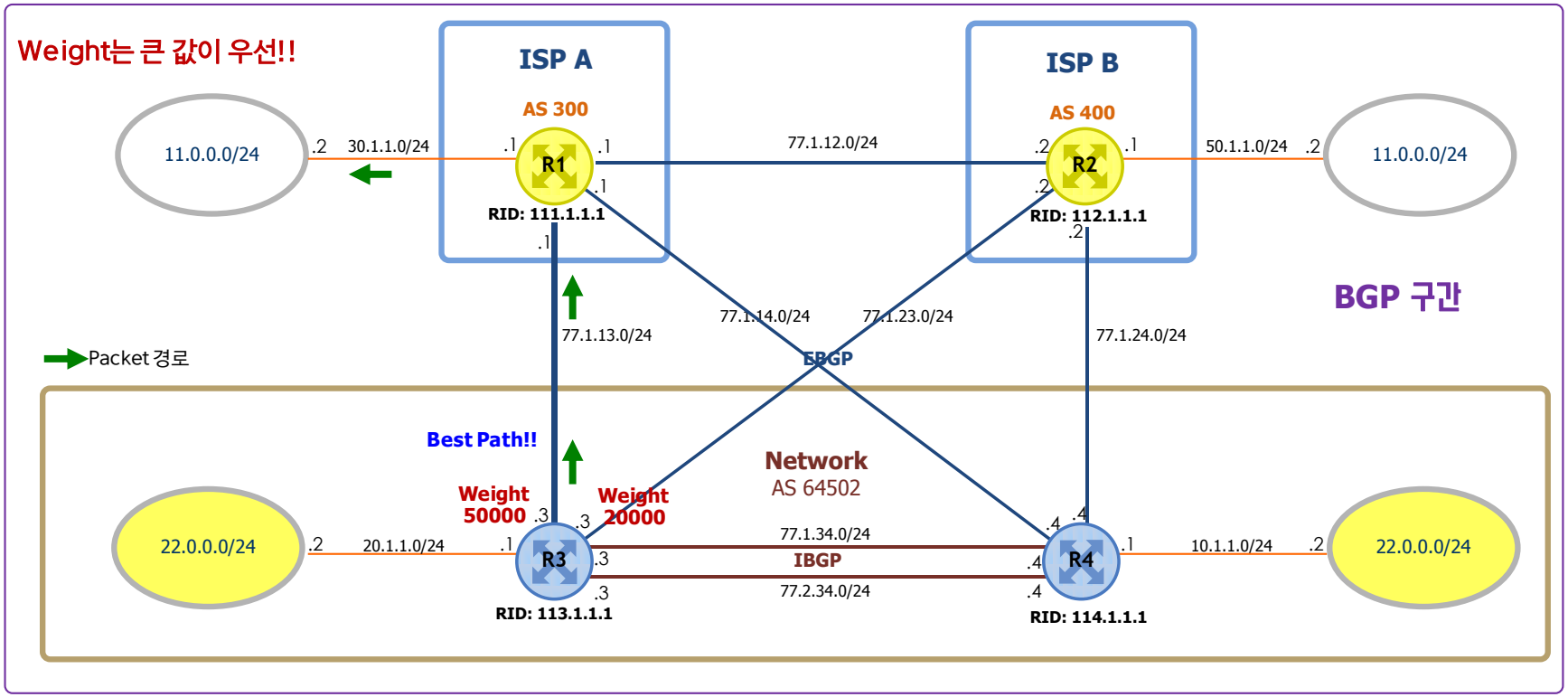
<Notification Message Error Code>

Error Code	Error
1	Message Header Error
2	Open Message Error
3	Update Message Error
4	Hold Timer Expired
5	Finite State Machine Error
6	Cease

2.4 BGP Attributes



2.4.1 BGP Attributes – Weight 값을 이용한 Packet 경로 변경



Administrative Weight Attribute

- Network AS 64502에 있는 R3는 목적지 11.0.0.0/24로 가는 경로가 두 개 존재한다(over ISP A and ISP B). 이 때, ISP A를 경유한 경로의 대역폭이 더 크다고 가정하자.
- R3에서 ISP A를 경유한 경로를 우선하도록 하기 위해, R3에서 Weight 값을 줄 수 있다. 즉, neighbor 111.1.1.1에 더 높은 Weight 값 50000을 준다.

Weight 값 설정

- Weight 값은 0~65,535 사이의 값을 가질 수 있으며, Default 값으로 neighbor에서 배운 경로는 weight=0을 가지며, 자신이 생성한 경로는 weight=32,768을 가진다.
- Weight 값은 Router 내부에서 특정 경로로 가는 out-going interface를 결정할 때만 의미가 있고 neighbor에게 전달되는 값은 아니다.

[R3 설정]

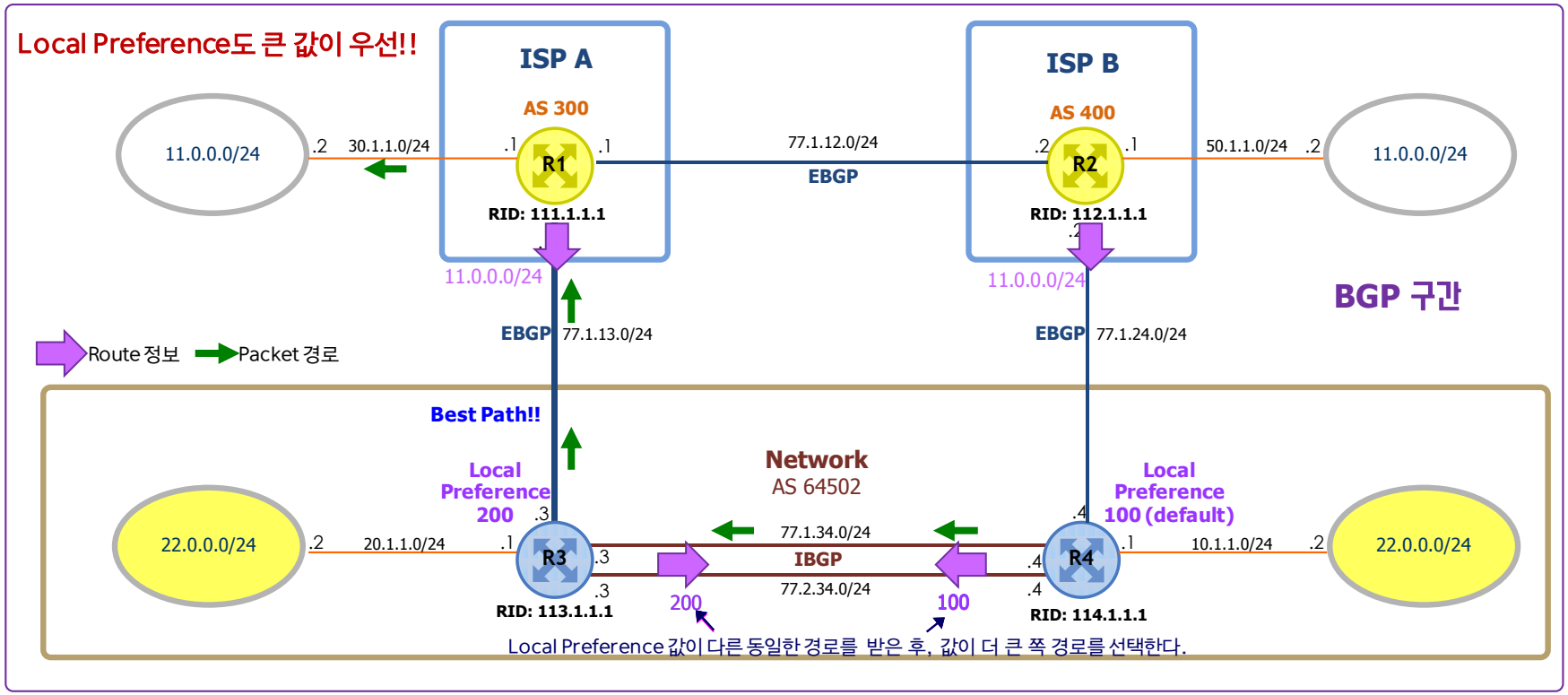
```
R3(config-router)# neighbor 77.1.13.1 weight 50000
R3(config-router)# neighbor 77.1.23.2 weight 20000
```

[R3 확인]

```
R3# show ip bgp
   Network          Next Hop    Metric LocPrf Weight Path
*> 11.0.0.0/24      77.1.13.1(R1)  0        50000 300 ?
*  11.0.0.0/24      77.1.23.2(R2)  0        20000 400 ?
```

Best

2.4.2 BGP Attributes – Local Preference 값을 이용한 Packet 경로 변경



Local Preference Attribute

- 상단 그림에서 **11.0.0.0/24**로 가는 경로가 두 개 존재한다.
- 이번에는 Local Preference 값을 이용한 **Best** 경로 선택을 보자.
- ISP A를 경유한 경로를 선택하도록 하기 위해 R3에서 Local Preference 값을 더 크게 설정한다 (Local Pref. =200).
- R4는 Default 값 100을 그대로 둔다.
- 이 값은 Weight 값과 달리 IBGP Peer간 값을 공유한다. 그러나, AS 외부로는 전달되지 않는다.

Local Preference 값 설정

- 이 값은 0~4, 294, 967, 295 사이의 값을 가질 수 있으며, Default 값은 100임

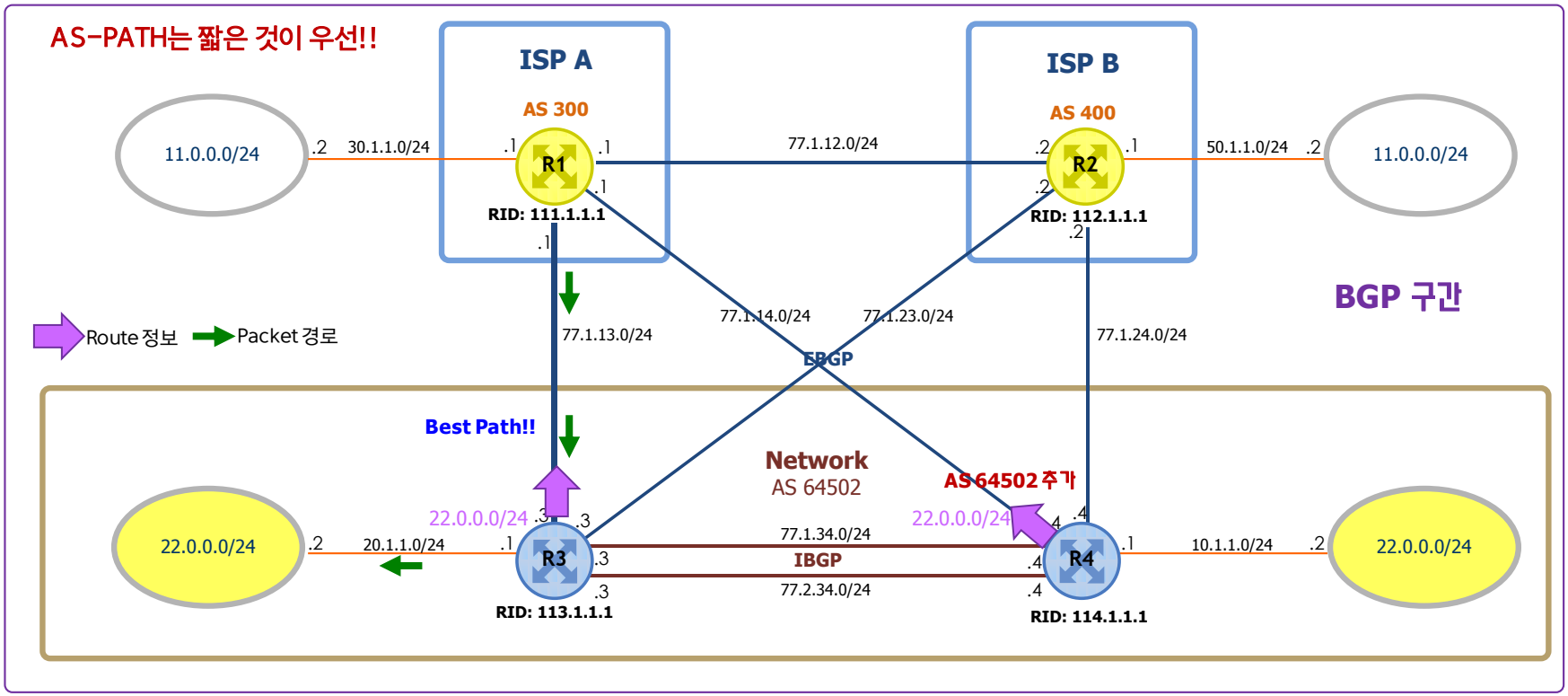
[R3 설정]

```
R3(config-router)# bgp default local-preference 200
```

[R3/4 확인]

```
R3# show ip bgp
   Network          Next Hop    Metric LocPrf Weight Path
*> 11.0.0.0/24      77.1.13.1(R1) 0    200      0    300 ?   <- Best
*  11.0.0.0/24      77.1.24.2(R2) 0    100      0    400 ?
------------------------------------------------------------
R4# show ip bgp
   Network          Next Hop    Metric LocPrf Weight Path
*> 11.0.0.0/24      77.1.13.1(R1) 0    200      0    300 ?   <- Best
*  11.0.0.0/24      77.1.24.2(R2) 0    100      0    400 ?
```

2.4.3 BGP Attributes – AS-PATH를 이용한 Packet 경로 변경



AS-PATH Attribute

- 이번에는 ISP A에서 AS 64502의 22.0.0.0/24 경로로 향하는 Incoming 경로를 생각해 보자.
- ISP A에서 22.0.0.0/24로 가는 경로는 두 개가 존재한다(Over R3 & Over R4)
- AS 64502 Network의 의 관리자는 ISP A가 대역폭이 더 큰 R3 경로를 선택하도록 하고 싶다.
- 이를 위해, R4에서 경로(22.0.0.0/24)를 advertise 할 때, 잉여의 AS를 추가해서 보낸다. 다른 조건이 동일하다면 AS-PATH의 길이가 짧은 경로가 우선하기 때문에 ISP A는 R3를 경유하는 경로를 선택한다.
  (이 때, 주의할 점은 추가할 AS 번호는 자신의 AS(64502)를 사용하는 것이다.)

[R4 설정]

```
R4# show run
router bgp 64502
 neighbor 77.1.14.1 route-map PATH out
route-map PATH permit 10
 match ip address 1
 set as-path prepend 64502
```

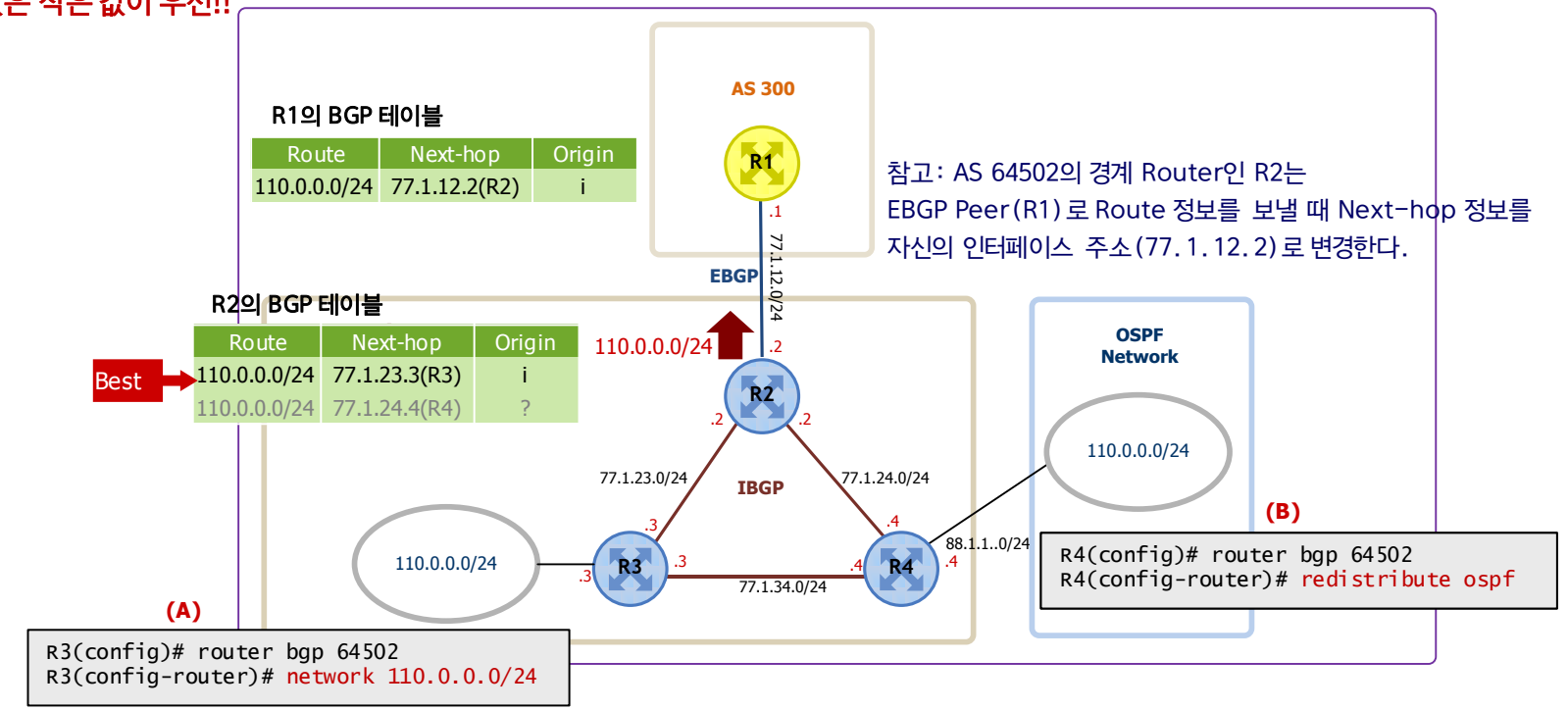
[R1 (ISP A) 확인]

```
R1# show ip bgp
   Network          Next Hop    Metric LocPrf Weight Path
*> 22.0.0.0/24      77.1.13.3(R3) 0         0      64502 ?
*  22.0.0.0/24      77.1.14.4(R4) 0                64502 64502
?
```

Best

2.4.4 BGP Attributes – Origin 값을 이용한 Packet 경로 변경

Origin 값은 작은 값이 우선!!



Origin Attribute

- BGP로 전달되는 Route는 갑자기 나타난 것이 아니라, 어디선가 배우거나 BGP로 유입된 경로이다.
- 즉, BGP **network** 명령으로 직접 BGP 경로를 발생할 수도 있고 **(A)**, **redistribute ospf (static) (B)** 등으로 유입되거나, 다른 **EGP Protocol**에서 배울 수도 있다.
- BGP 입장에서 이처럼 다양한 Origin에 대해 Priority를 줄 수 있다.
- BGP가 직접 **network** 명령으로 입력한 정보는 bgp table에서 I (IGP)로 표시되며, 가장 높은 Priority(1)을 가진다.
- BGP 이외의 다른 EGP Protocol로 배운 경로는 e (EGP)로 표시되며, 2의 값을 가진다. 실제로 BGP외에 다른 EGP Protocol은 현재는 존재하지 않으므로 이 값은 거의 찾아볼 수 없다.
- 마지막으로, OSPF나 Static 경로를 BGP에 유입(redistribute)한 경우이다.

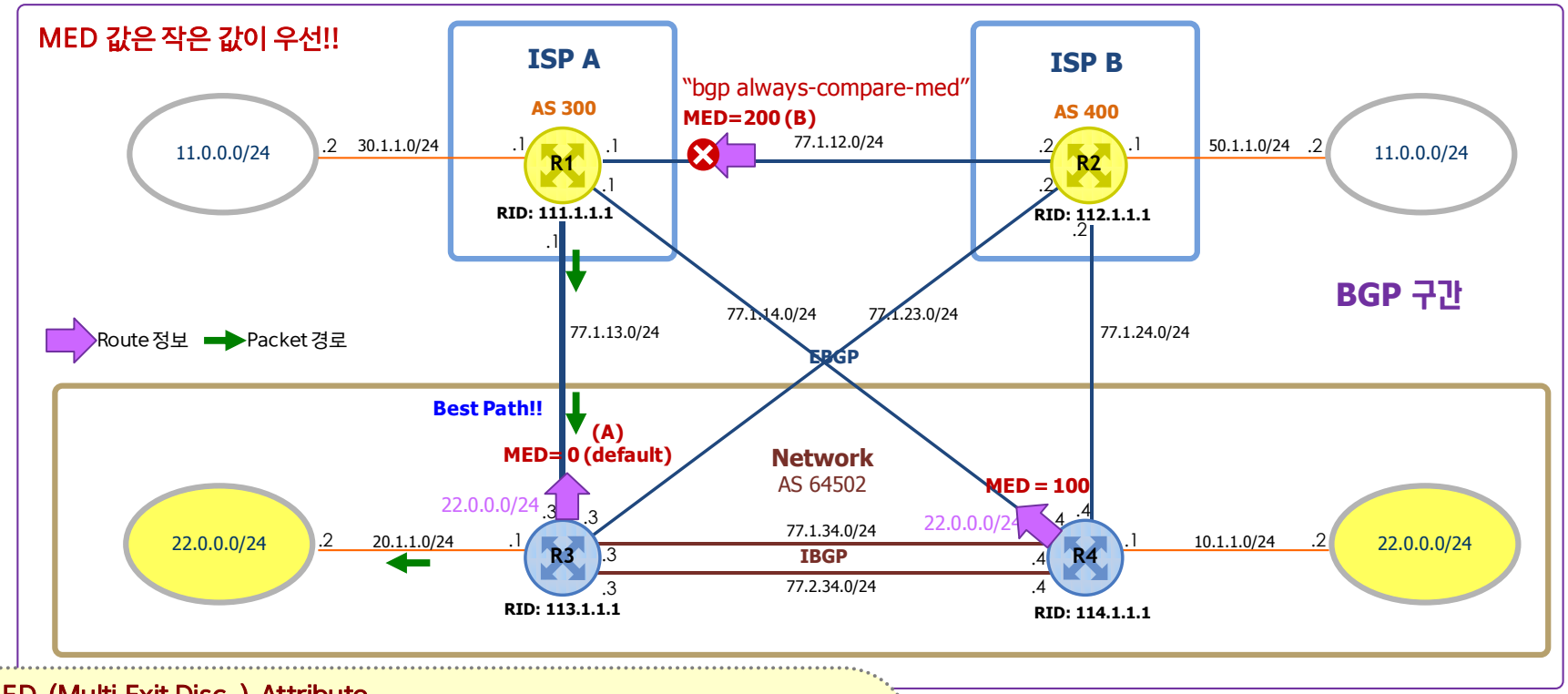
한 번 redistribution이 되면, 실제로 OSPF에서 배운 것인지 Static에서 배운 것인지 구별할 수 없어, ? (incomplete)로 표시한다 (값은 3).

IGP (1) 〉 EGP (2) 〉 Incomplete (3)

```
R2# show ip bgp

Origin Codes: i -IGP, e - EGP, ? - incomplete
   Network          Next Hop    Metric LocPrf Weight Path
*> 110.0.0.0/24     77.1.23.3(R3)  0    100         64502 i   <- Best
*  110.0.0.0/24     77.1.24.4(R4)  0    100         64502 ?
```

2.4.5 BGP Attributes – MED 값을 이용한 Packet 경로 변경



MED (Multi Exit Disc.) Attribute

- ISP A에서 AS 64502의 22.0.0.0/24로 오는 경로는 두 개가 존재한다 (Over R3 & R4).
- AS 64502의 의 관리자는 ISP A에서 22.0.0.0/24로 향하는 경로 선택 시, 대역폭이 더 큰 R3를 경유한 경로를 선택하도록 하고 싶다고 가정하자.
- 이를 위해, R3와 R4에서 ISP A로 해당 경로를 보낼 때, MED 값을 추가할 수 있다.

MED 값 설정

- MED 값은 0~4,294,967,295 사이의 값을 가질 수 있으며, Local Preference와는 달리 작은 값이 우선한다(보통 BGP Metric이라고 하면 MED 값을 의미함).
- 따라서, R3(MED=0)이 R4(100)보다 작으므로 R3를 경유한 경로를 선택한다 (A).

주) MED 값은 동일한 AS에서 온 것만 서로 비교 대상이 된다. 즉, AS 64502에서 온 MED 값과 AS 400에서 온 MED 값은 비교 대상이 되지 않는다(B).
→ 강제로 비교하게 하려면 "bgp always-compare-med"를 설정한다.

[R4 설정]

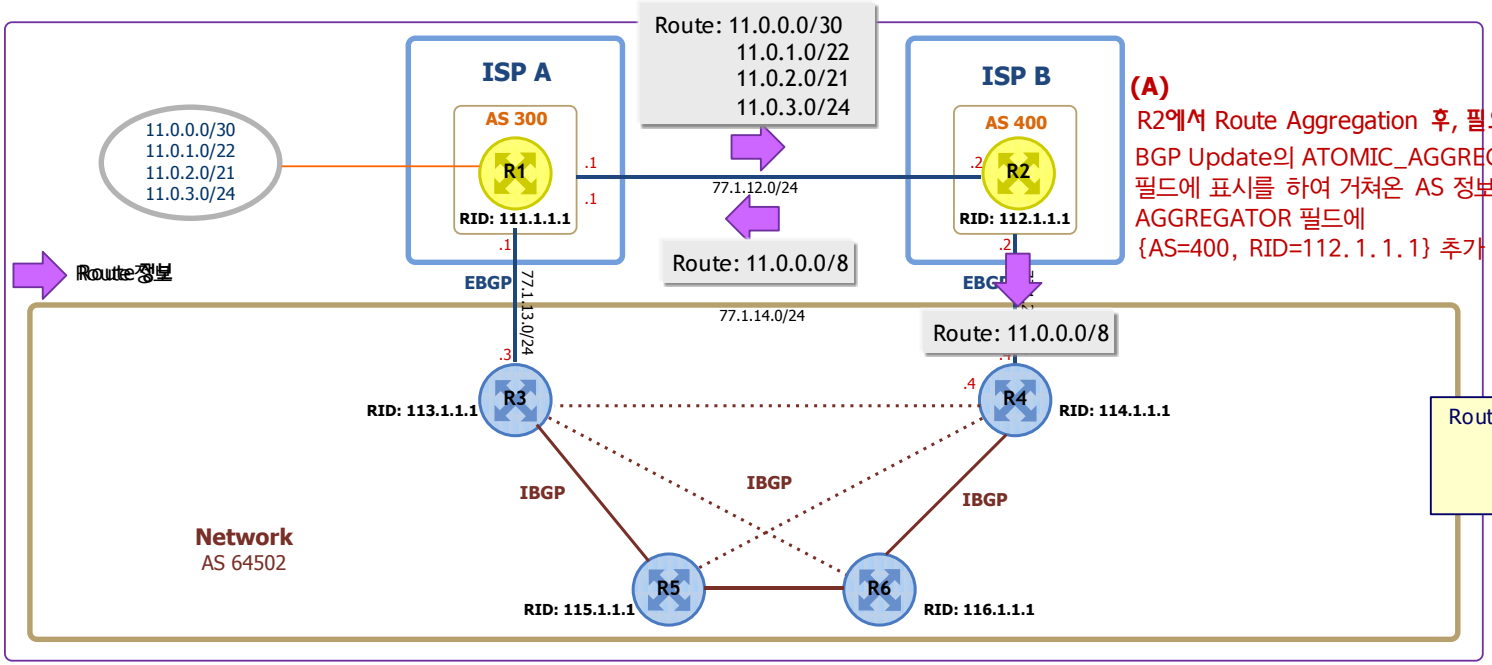
```
R4# show run
router bgp 64502
 neighbor 77.1.14.1(R1) route-map MED out
route-map MED permit 10
 match ip-address 1
 set metric 100
```

[R1 (ISP A) 확인]

```
R1# show ip bgp
   Network          Next Hop    Metric LocPrf Weight Path
*> 22.0.0.0/24 77.1.13.3(R3)    0        0     64502 ?
*  22.0.0.0/24 77.1.14.4(R4) 100         0     64502 ?
```

Best

2.4.6 기타 BGP Attributes – Atomic Aggregate & Aggregator (1)



(A)
R2에서 Route Aggregation 후, 필요한 정보 추가!!
BGP Update의 ATOMIC_AGGREGATE 필드에 표시를 하여 거쳐온 AS 정보가 사라졌음을 알림,
AGGREGATOR 필드에 {AS=400, RID=112.1.1.1} 추가

ATOMIC_AGGREGATE & AGGREGATOR Attribute

- R2에서 11.0.X.0/X 경로들을 전파할 때 Routing Entry를 줄이기 위해, 11.0.0.0/8로 Aggregation 할 수 있다 **(A)**.
- 이 때, R2는 축약된 경로를 전달할 때, 그 사실을 Update Packet에 표시하여야 한다.
  → Update Packet의 ATOMIC_AGGREGATE 필드에 표시를 하여 Aggregation 사실을 알려주고, Aggregation을 한 주체인 R2의 RID를 AGGREGATOR로 표시한다.

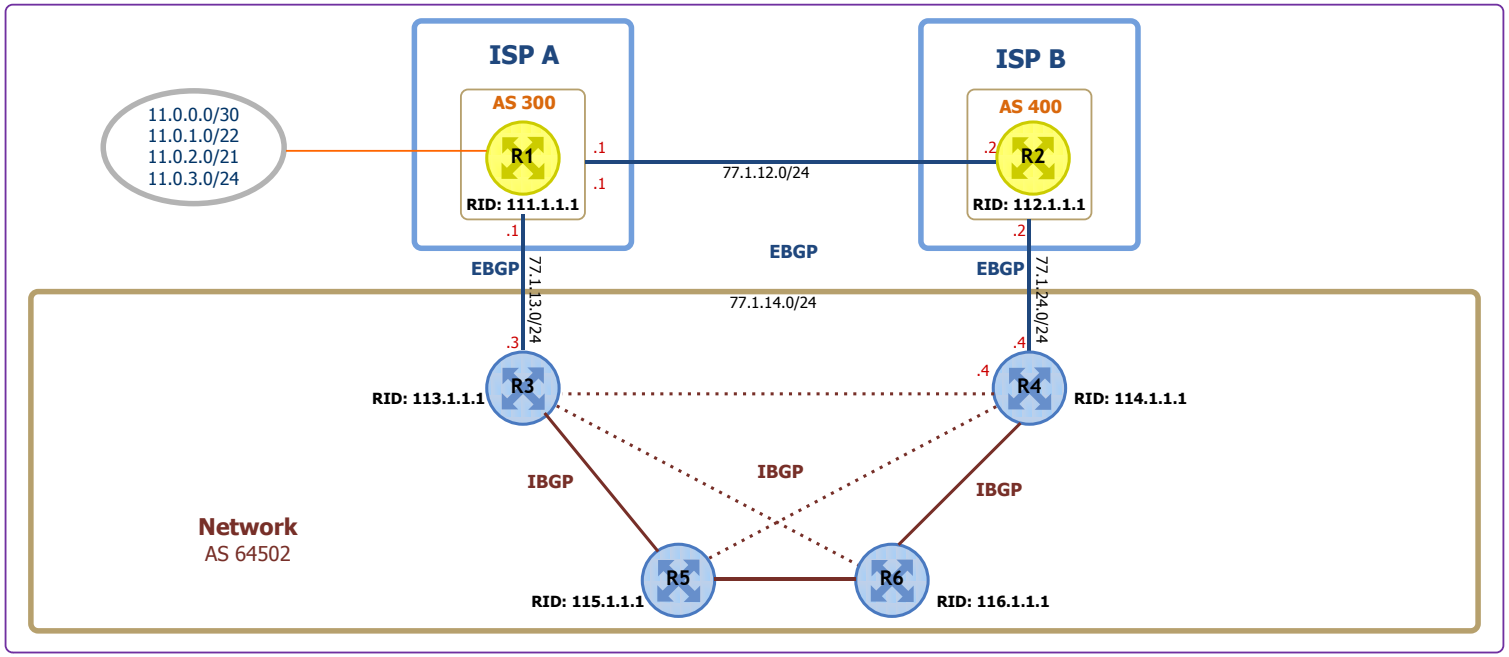
Route Aggregation 설정

```
R2#(config-router)# aggregate-address 11.0.0.0/8 summary-only
```

Route Aggregation 확인

```
R4#show ip bgp 11.0.0.0
BGP routing table entry for 11.0.0.0/8, version 98
Paths: (1 available, best #1, table Default-IP-Routing-Table)
  Advertised to update-groups:     1
   400, (aggregated by 400 77.1.24.2(R2))
    77.1.24.2 from 77.1.24.2 (77.1.24.2)
      Origin IGP, localpref 100, valid, external, atomic-aggregate, best
```

2.4.6 기타 BGP Attributes – Atomic Aggregate & Aggregator (2)



R1 BGP Table 확인

- R2에 의한 Route 축약의 결과, AS-PATH LIST에서 AS 300이라는 정보가 사라진다. 즉, R2에서 동일한 Route를 다시 R1으로 보내더라도 Loop를 감지할 수 없다.

```
R1# show ip bgp
    Network          Next Hop Metric LocPrf Weight Path
*> 11.0.0.0/8       77.1.12.2(R2)                  400 ?
*> 11.0.0.0/30       0.0.0.0 (R1)        32768     i
```

AS-SET 설정

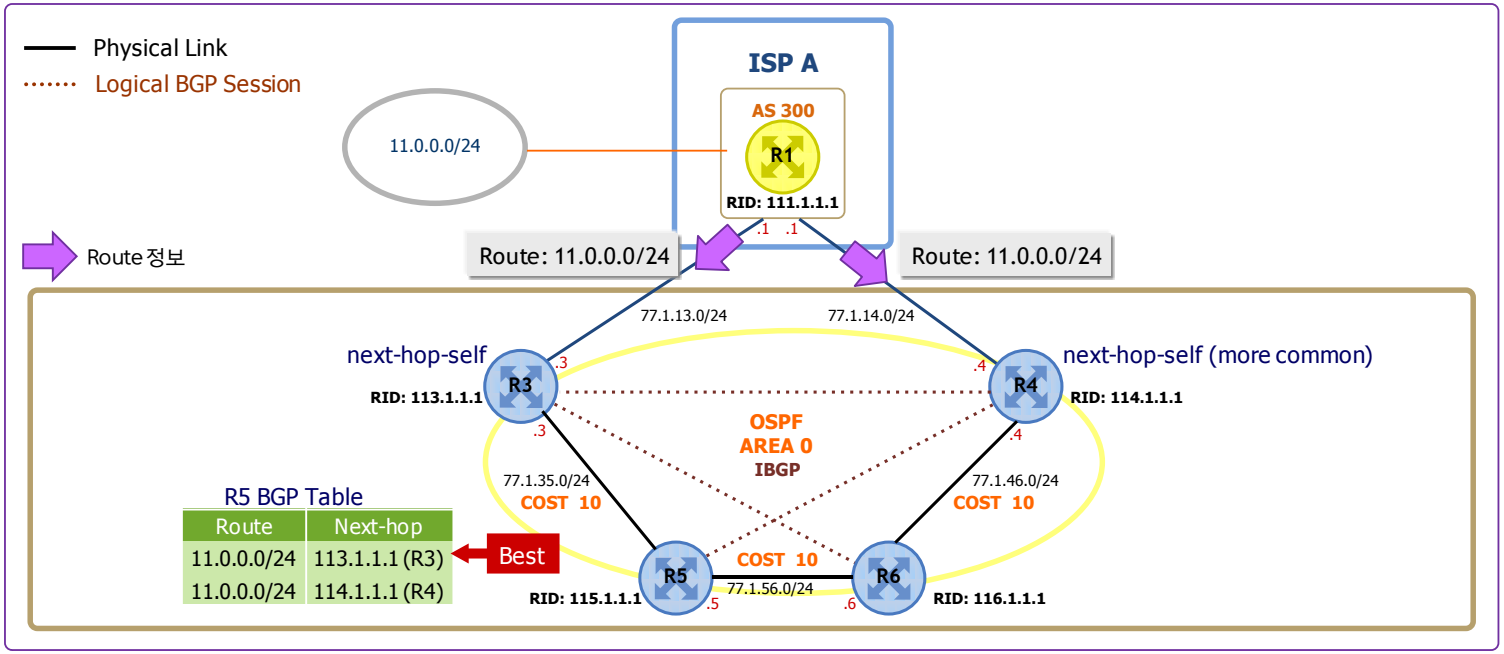
- 이를 방지하기 위해, 앞의 aggregate-address 명령에 as-set 명령을 추가한다.

```
R2#(config-router)# aggregate-address 11.0.0.0/22 as-set summary-only
```

- 결과, 해당 Route를 R1이 받더라도 AS-PATH List에 자신의 AS 300이 있으므로 BGP Table에 쓰지 않는다. R4의 BGP Table을 확인해 보면, 축약이 되기 전에 거쳐온 AS가 { }에 표시가 된다. { }안의 AS 번호는 Loop 방지 용도로만 사용하며, AS-PATH 길이에 의한 경로 선택에는 사용하지 않는다.

```
R4# show ip bgp
   Network          Next Hop Metric LocPrf Weight Path
*> 11.0.0.0/8       77.1.24.2                      400 {300} ?
```

2.4.7 기타 BGP Attributes – Next-Hop



Next-Hop Attribute

- 모든 Attribute값들이 같으면 해당 경로의 next-hop까지의 거리가 최단인 경로를 선택한다.
- 그림에서 R3, R4, R5, R6간에 Physical Link는 검은색 실선이라고 가정하자. BGP 연결은 fully 맺어져 있다.
- 다음의 경우를 고려해 보자.

1) R5는 R3, R4, R6와 BGP neighbor를 맺고 있다.

```
R5(config-router)# neighbor 113.1.1.1 remote-as 64502
R5(config-router)# neighbor 113.1.1.1 update-source lo0
R5(config-router)# neighbor 114.1.1.1 remote-as 64502
R5(config-router)# neighbor 114.1.1.1 update-source lo0
R5(config-router)# neighbor 116.1.1.1 remote-as 64502
R5(config-router)# neighbor 116.1.1.1 update-source lo0
```

2) 이제 R1에서 보낸 동일한 Route 정보 11.0.0.0/24를 R3와 R4에서 받는다 (R6는 IBGP Rule에 의거 해당 Route를 보내지 않는다.)

3) R5의 BGP Table에는 이래와 같이 11.0.0.0/24에 대한 경로 정보가 두 개가 존재한다.

```
R5# show ip bgp
   Network          Next Hop Metric LocPrf Weight Path
*> 11.0.0.0/24 113.1.1.1(R3) 0      0      300 ?
*  11.0.0.0/24 114.1.1.1(R4) 0      0      300 ?
```

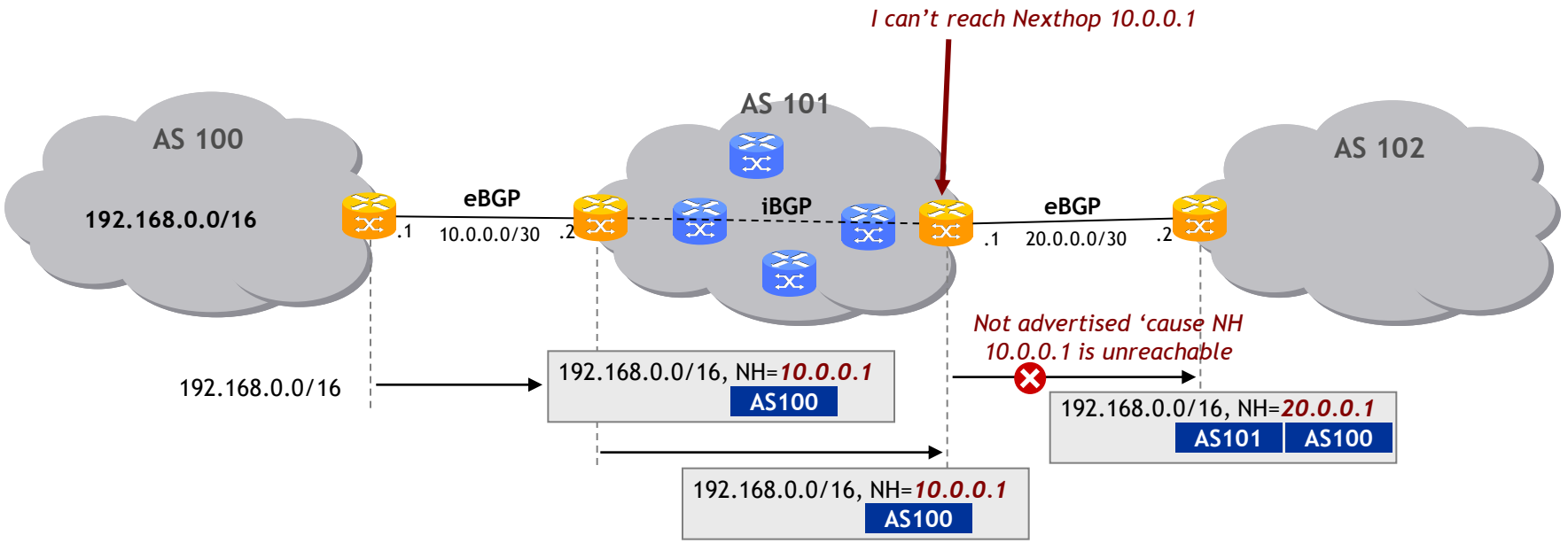
4) 여기서 Next-hop까지의 경로를 Routing Table로 확인한다. 다른 Attribute 값이 동일하다면 BGP는 next-hop까지의 거리가 가까운 경로 (Low IGP Cost) 를 Best 경로로 선택한다.

```
R5# show ip route

o 113.1.1.1 [110/10] via 77.1.35.3  00:10,45 gi3
o 114.1.1.1 [110/20] via 77.1.56.6  00:10,49 gi5
```

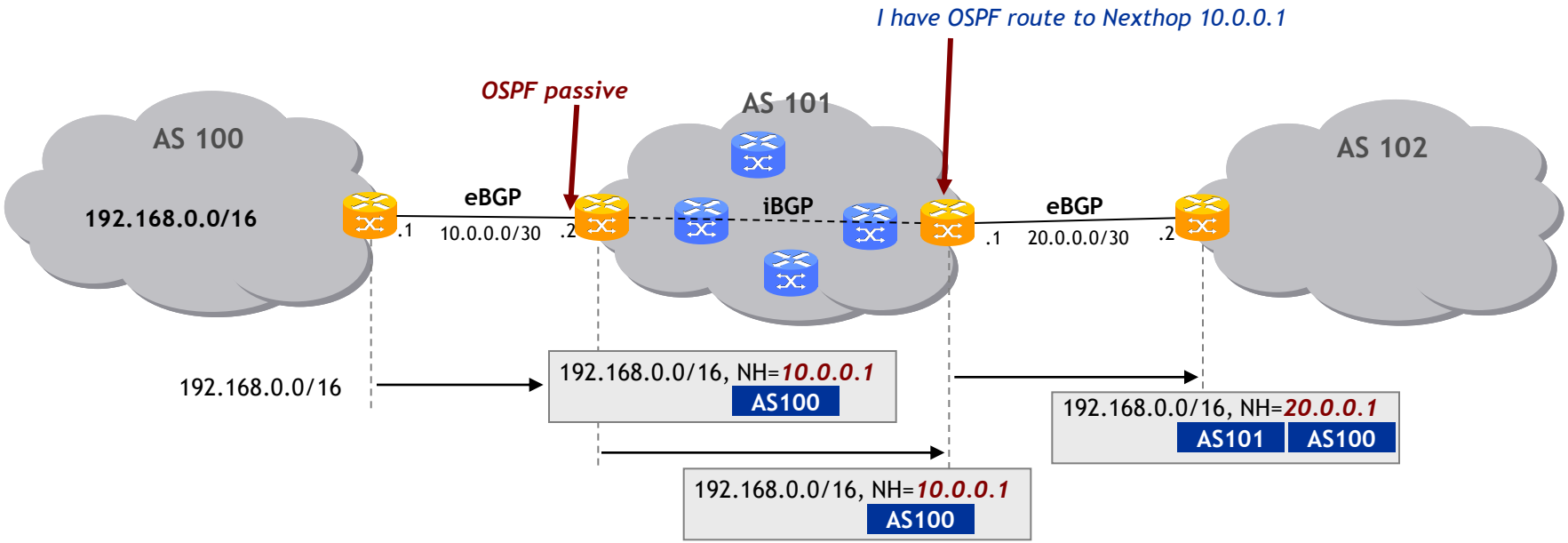
2.4.7 BGP Nexthop

- eBGP peer에게 NLRI를 보낼 때, Next-hop Attribute에 자신의 interface address를 씀
- iBGP peer에게 전달할 때에는 Next-hop을 변경하지 않고 그대로 보냄
- 수신한 BGP NLRI는, nexthop address가 IGP routing table에서 reachable하여야 만 유효하다.
- 그런데, 아래 예와 같이 Nexthop address가 unreachable하면 Routing Fail
  - AS100~AS101 간의 physical link의 address가 AS101 내에서 OSPF로 advertise되지 않은 경우임
- 해결책은? (Next page)



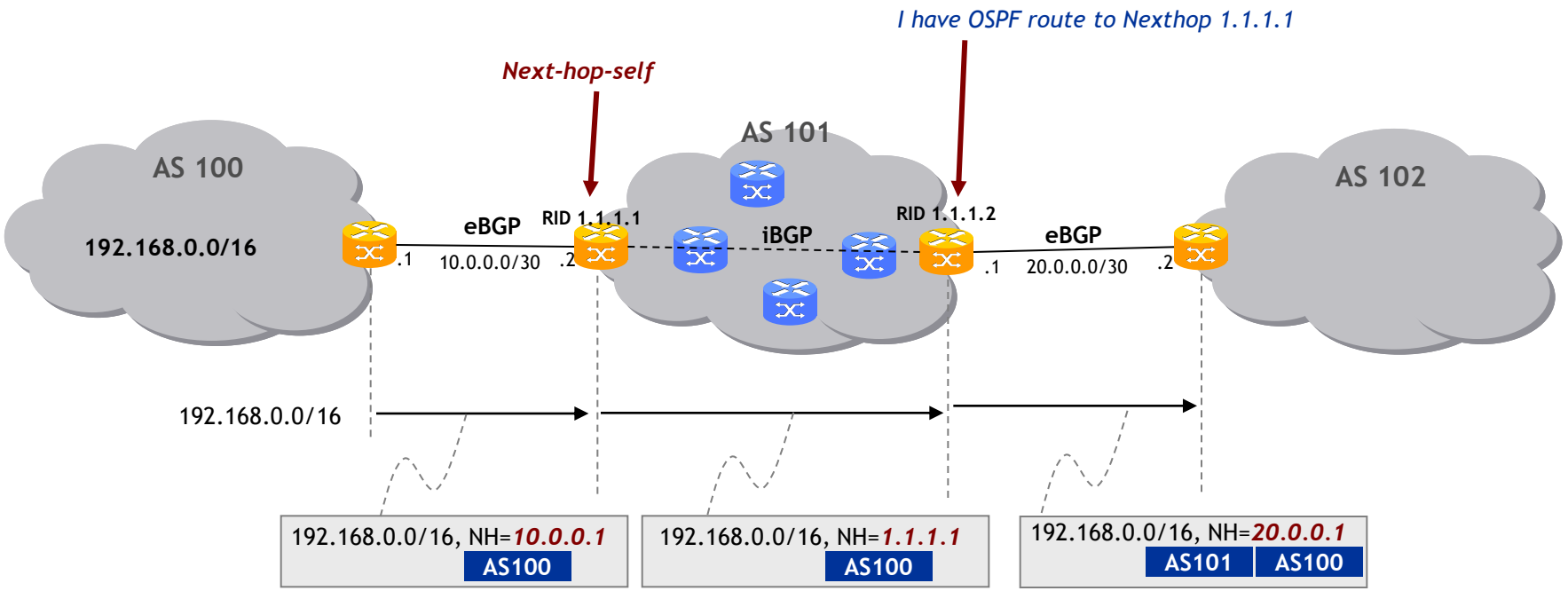
2.4.7 BGP Nexthop (cont)

1 Configure eBGP link as OSPF "passive interface" to advertise into OSPF domain

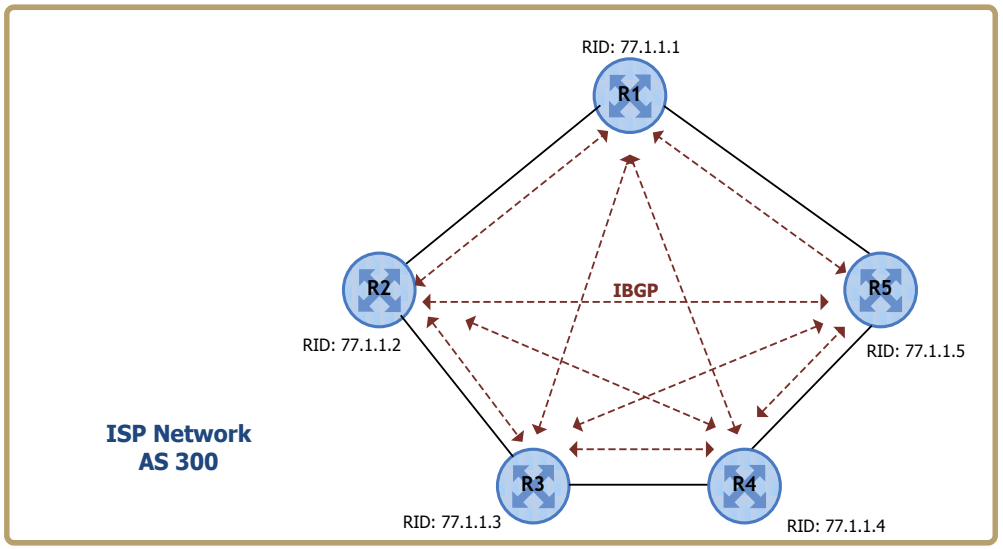


2.4.7 BGP Nexthop (cont)

2 Configure "next-hop-self" on ASBR to replace BGP-Nexthop attribute with its own RID



2.5.1 Route-Reflector

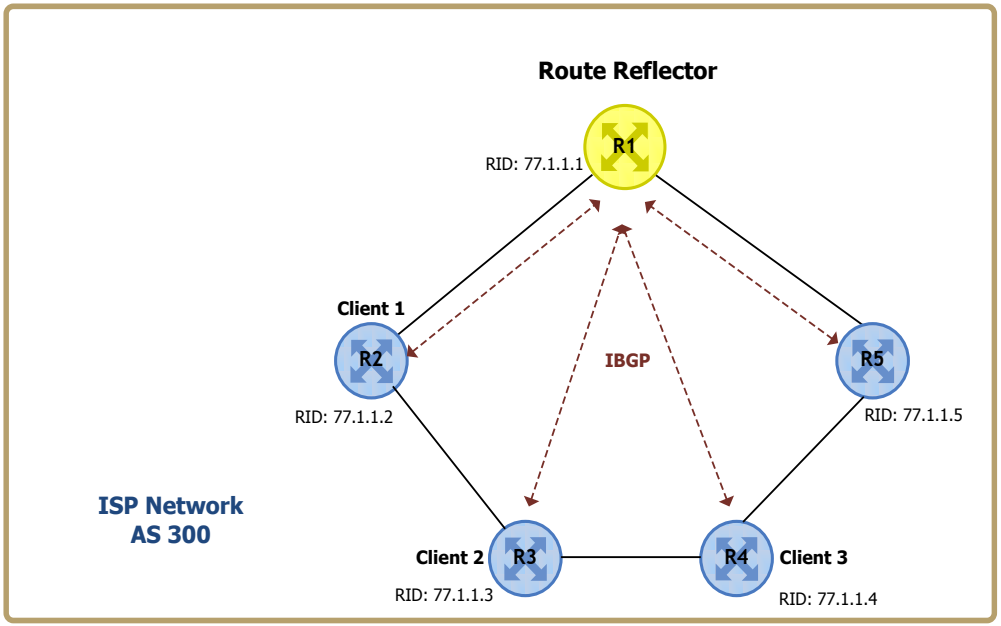


ISP Network에는 5대의 Router가 있다고 가정하자.
모든 Router에서 BGP 경로를 전달하려면, 10개의 IBGP 세션이 필요하다.
Router 개수가 증가함에 따라, IBGP 연결의 수는 기하급수적으로 증가한다.
(n(n-1)/2, n은 Router의 개수)

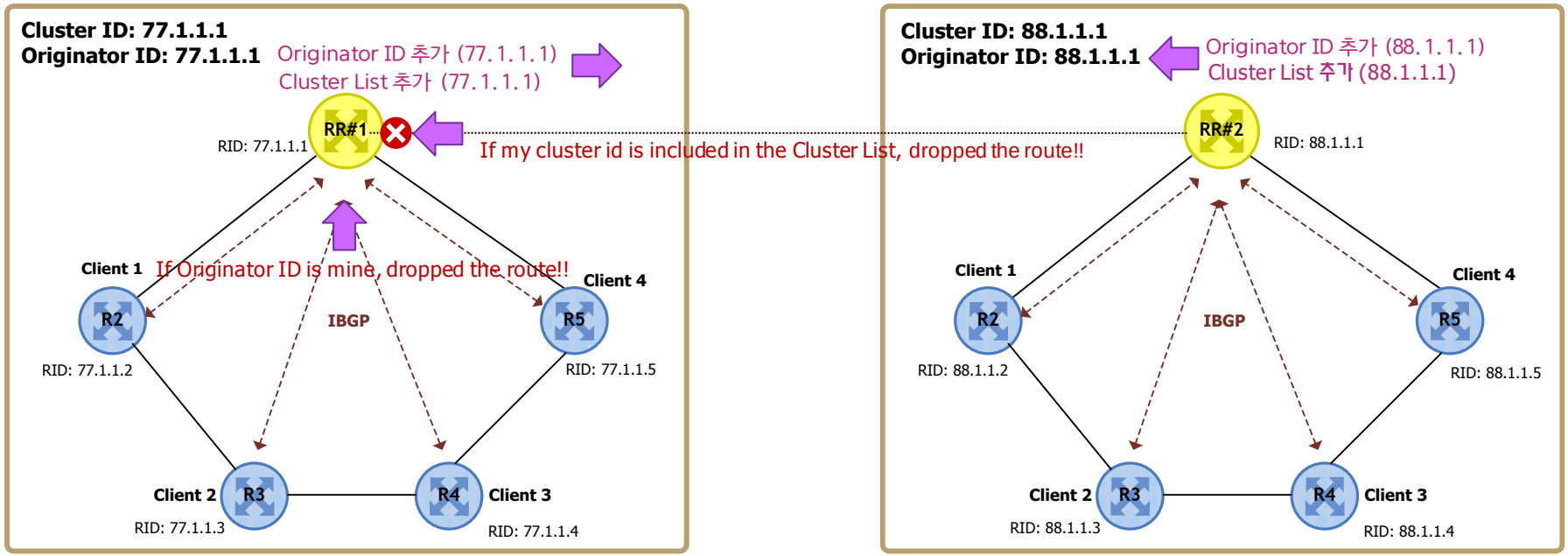
5개의 Router 중에서 가장 성능이 높은 R1을 Route Reflector (Sever)로 설정하고, 나머지 Router는 Client로 두면 오른쪽 그림과 같이 필요한 IBGP의 개수가 현저히 줄어든다.

```
R1(config-router)# neighbor 77.1.1.2 route-reflector-client
R1(config-router)# neighbor 77.1.1.3 route-reflector-client
R1(config-router)# neighbor 77.1.1.4 route-reflector-client
R1(config-router)# neighbor 77.1.1.5 route-reflector-client
```

- RR의 역할: Client에게 자신이 알고 있는 BGP 경로 정보 전달
- Client의 역할 : RR로부터 BGP 경로 정보 수신



2.5.2 Route-Reflector – Loop 방지기법



ORIGINATOR_ID를 이용한 Loop 방지 기법

- RR은 Route 정보에 자신의 Router ID를 ORIGINATOR_ID 속성에 추가한다(이 속성은 AS 내부에서만 사용).
- RR은 자신의 ORIGINATOR_ID와 같은 Route 정보를 받으면, Routing Loop으로 간주, 폐기한다.

CLUSTER_LIST를 이용한 Loop 방지 기법

- RR별로 RR과 그의 Client 그룹을 Cluster라 한다.
- 이 때, Cluster ID는 보통 RR의 Router ID이다.
  (여러 개의 RR을 묶어서 하나의 Cluster로 구성하려면 bgp cluster-id 명령 사용).
- RR은 Route정보의 Cluster List에 자신의 CLUSTER_ID를 추가한다.
- 자신의 CLUSTER_ID와 동일한 정보가 CLUSTER_LIST에 있을 경우, Loop으로 간주, 해당 경로를 폐기한다.
  참고로 ORIGINATOR_ID는 Cluster 내부의 Loop 방지용이고, CLUSTER_LIST는 AS 내부의 CLUSTER 간의 Loop 방지용도로 기억하자.

2.6 Route Selection among various Routing Protocol

Administrative Distance (Cisco)

Table 9-1. Routing protocols and their administrative distances

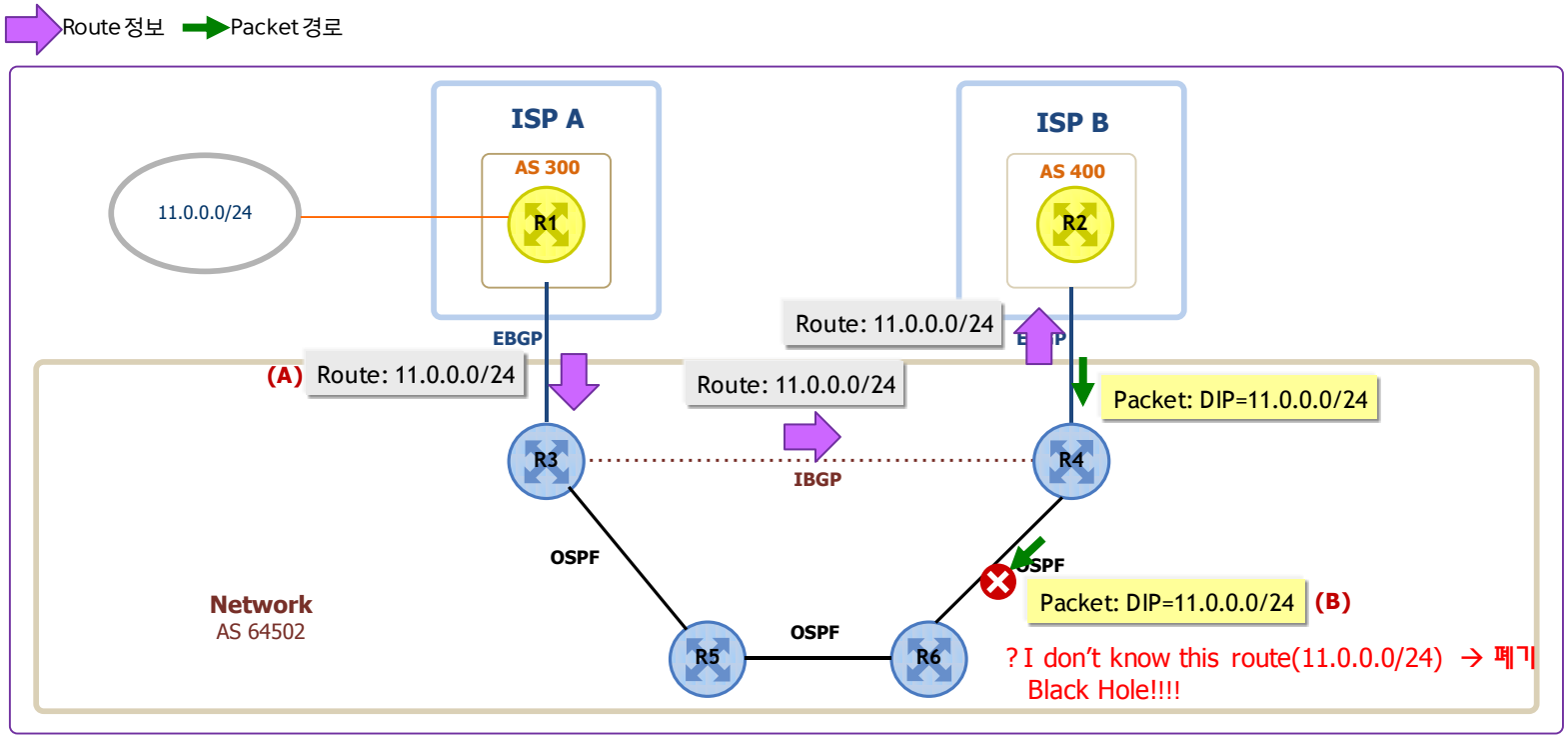
Route type	Administrative distance
Connected interface	0
Static route	1
Enhanced Interior Gateway Routing Protocol (EIGRP) summary route	5
External Border Gateway Protocol (BGP)	20
Internal EIGRP	90
Interior Gateway Routing Protocol (IGRP)	100
Open Shortest Path First (OSPF)	110
Intermediate System–Intermediate System (IS-IS)	115
Routing Information Protocol (RIP)	120
Exterior Gateway Protocol (EGP)	140
On Demand Routing (ODR)	160
External EIGRP	170
Internal BGP	200
Unknown	255

Routing Protocol Preference (Juniper)

How Route Is Learned	Default Preference
Directly connected network	0
System routes	4
Static	5
MPLS	7
LDF	8
LDP	9
OSPF internal route	10
IS-IS Level 1 internal route	15
IS-IS Level 2 internal route	18
Default	20
Redirects	30
Kernel	40
SNMP	50
Router Discovery	55
RIP	100
RIPng	100
PIM	105
DVMRP	110
Routes to interfaces that are down	120
Aggregate	130
OSPF AS external routes	150
IS-IS Level 1 external route	160
IS-IS Level 2 external route	165
BGP	170
MSDP	175

AD값, Preference값이 작은 것을 우선시 한다.

2.8 Synchronization 기능 (1)

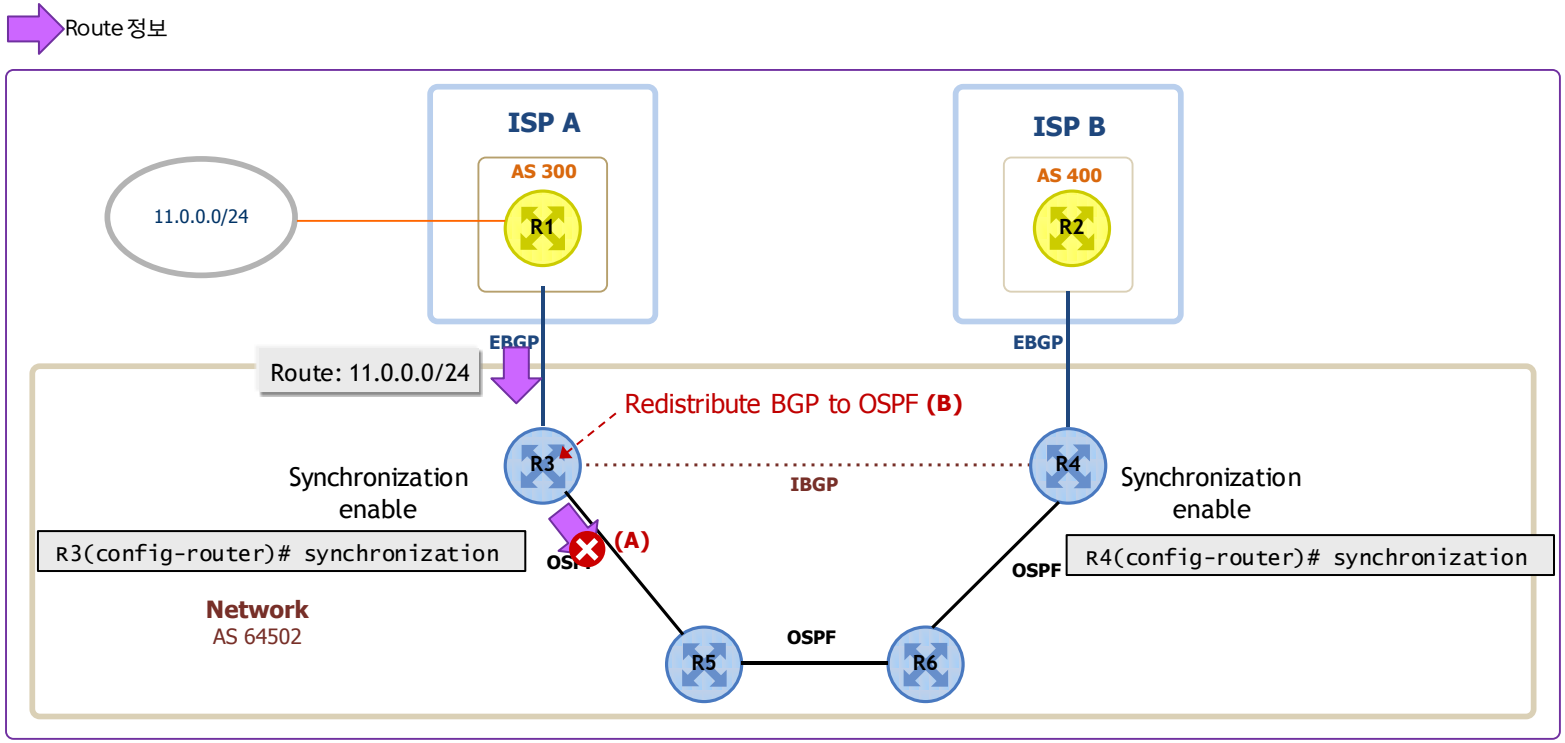


Non-BGP Router가 Network 내에 공존할 경우 발생 가능한 ISSUE

1. ISP A(AS 300)에서 ISP B(AS 400)으로 가기 위해, AS 64502 Network를 Transit Network로 사용한다고 가정하자.
2. AS 64502의 Border Router인 R3에서 **(A)** 11.0.0.0/24를 받았다.
3. AS 64502 Network는 그림과 같이, Non-BGP Router(R5/R6)가 두 개 포함돼 있다.
4. 이들 Router는 11.0.0.0/24 경로를 R4에게 bypass하고 자신의 RIB에는 쓰지 않는다. 즉, 11.0.0.0/24에 대한 정보가 없다.
5. 이 상태에서 11.0.0.0/24로 향하는 Traffic이 R4를 거쳐 R6로 도달한다. R6는 11.0.0.0/24에 대한 정보를 모르므로, 폐기한다. **(B)** → Black Holed

이러한 상황을 방지하기 위한 방법으로, Synchronization 기능이 있다 (다음 페이지).

2.8 Synchronization 기능 (2)

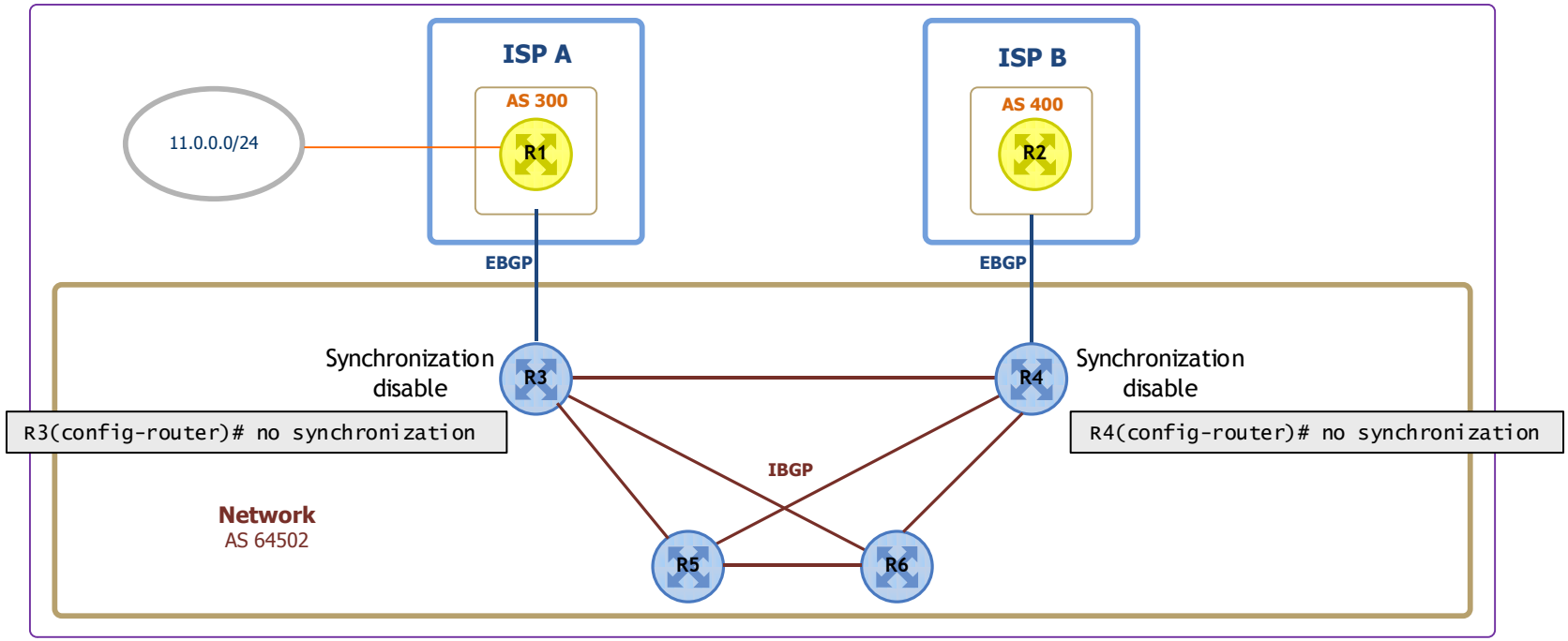


Synchronization 기능

1. Border Router R3는 ISP A에서 받은 경로정보(11.0.0.0/24)를 AS 64502 Network에 전파하기 전에, IGP Table을 검사한다 (여기서는 OSPF table).
2. IGP Table에 11.0.0.0/24 정보가 없다면, 해당 경로를 전파하지 않는다 **(A)**.
   → Black Hole은 예방되나, 해당 목적지로 향하는 Packet을 전달할 수 없다.

Solution 1: BGP 경로를 IGP로 Redistribution 한다 (B).
Solution 2: Non-BGP Router에도 BGP를 구동한다 (다음 페이지).

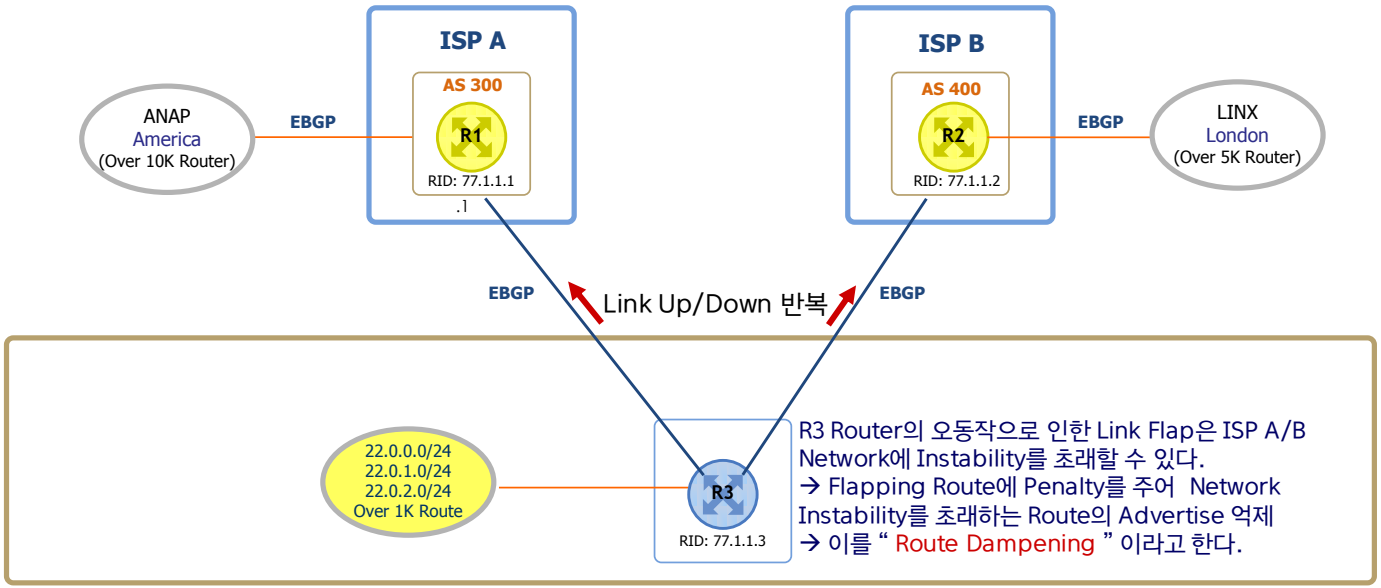
2.8 Synchronization 기능 (3)



Solution 2: Non-BGP Router에도 BGP를 구동한다.

1. AS 64502 Network에 있는 모든 Router(R3, R5, R6, R4)에 BGP를 모두 enable한다.
   → 이제, AS 64502에 있는 모든 Router들은 BGP 경로를 알게 된다.
2. 이 때는, Synchronization 기능 Disable 권장
   → IGP Table을 검색하거나, IGP가 Convergence 될 때까지 기다리면서 소요되는 추가적인 Delay 방지
   대부분의 장비에서는 no synchronization이 Default 값이다.

2.9 Route Flap Dampening



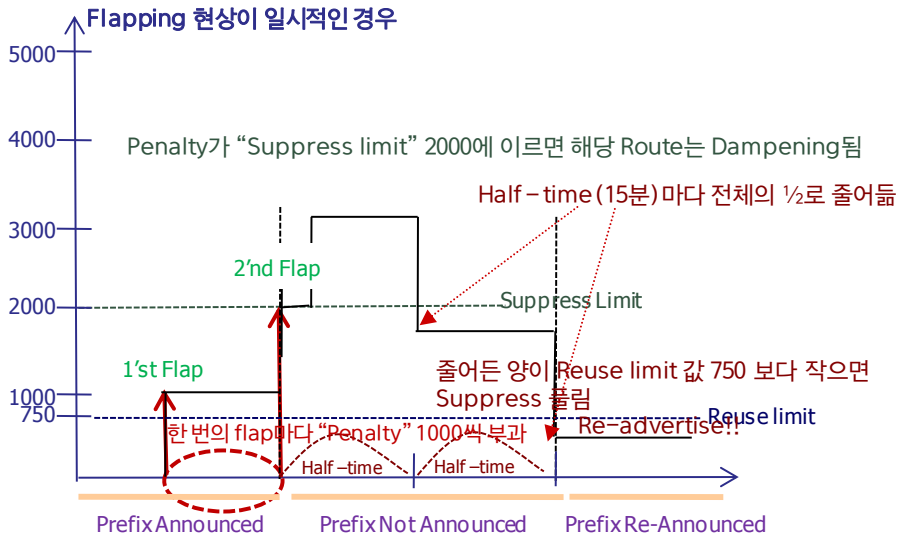
[Default Value – Cisco/ZebOS]
- Penalty : 1000 per flap
- Suppress Limit : 2000
- Half-Life-Time : 15 Min.
- Maximum Suppress Time : 60 Min. (4 x Half-Time)

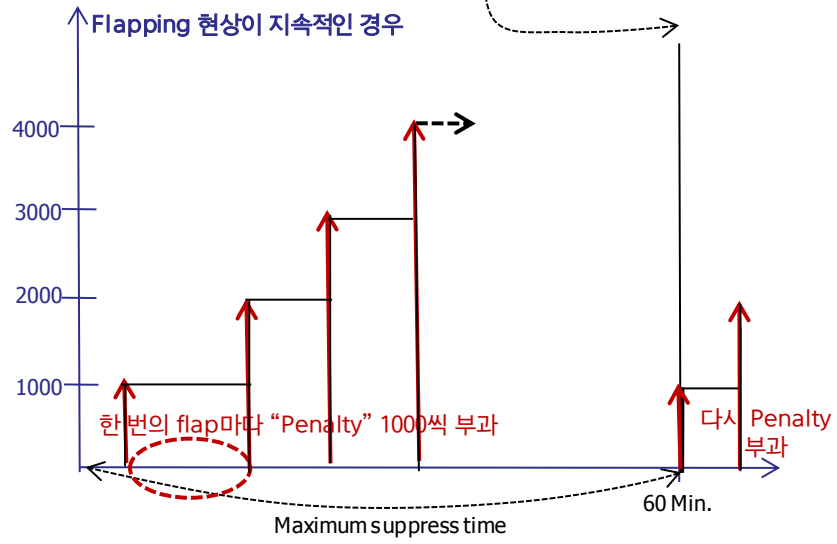
```
R(config-route-map)# set bgp dampening
{Half-time} {Reuse-limit} {Suppress-
limit} {Maximum-suppress-time}
```

```
R1(config-route-map)# set dampening 15
1000 2000 60
R2(config-route-map)# set dampening 15
1000 2000 60
```

아무리 Flap 현상이 지속적이어도 Maximum suppress time 이후에는 Suppress를 일단 풀어줌

Route Dampening 동작

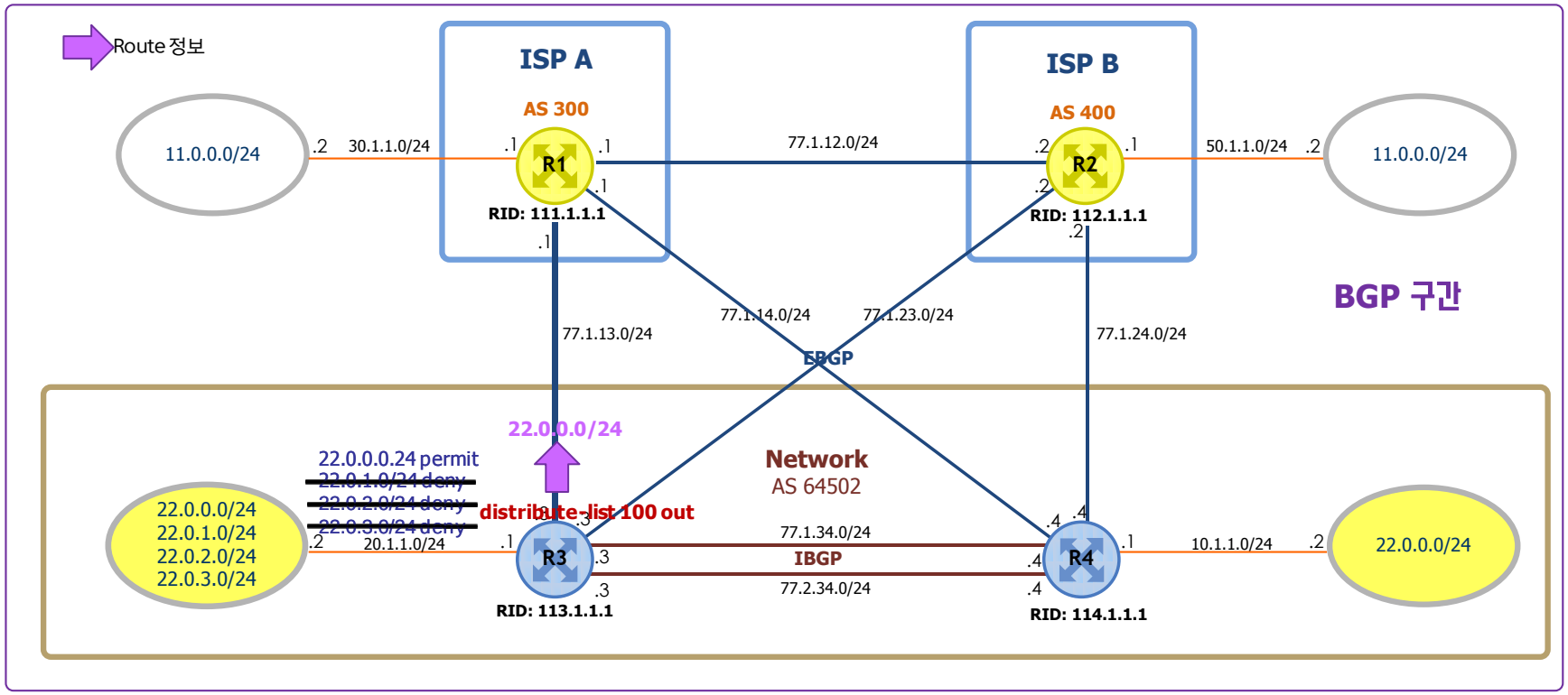




3. Filter

3.1 Distribute-List
3.2 Prefix-List
3.3 Filter-List

3.1 Distribute-List



ACL (Access List) 를 이용한 경로 차단

- AS 64502의 Network 중 22.0.0.0/24를 제외한 경로가 ISP A로 전달되는 것을 막고 싶다고 가정하자.
- 이 때, 가장 쉽게 사용할 수 있는 방법이 ACL을 이용하는 것이다.

ACL 기본구조

- **access-list** *Number* {**deny** | **permit**} prefix [prefix-wildcard] [log]

Neighbor에 적용하기

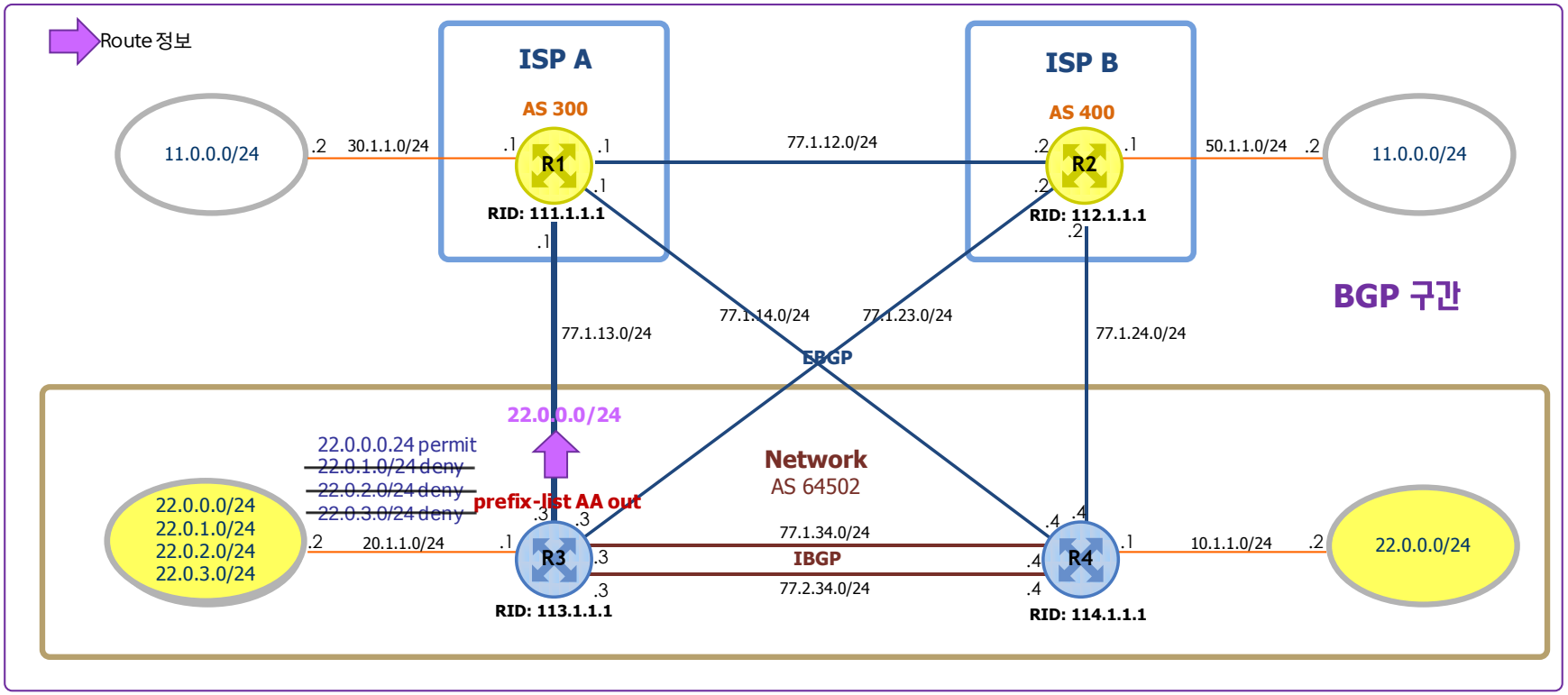
- **neighbor** x.x.x.x **distribute-list** *n* **out**

[R3 설정]

```
R3(config)# access-list 100 deny ip 22.0.1.0 0.0.0.0
255.255.255.0 0.0.0.0
R3(config)# access-list 100 deny ip 22.0.2.0 0.0.0.0
255.255.255.0 0.0.0.0
R3(config)# access-list 100 deny ip 22.0.3.0 0.0.0.0
255.255.255.0 0.0.0.0
R3(config)# access-list 100 permit ip any any
R3(config-router)# neighbor 77.1.13.1 distribute-list
                    100 out
```

```
R1# show ip bgp
   Network          Next Hop    Metric LocPrf  Weight Path
*> 22.0.0.0/24  77.1.13.3(R3) 0             0 64502 ?
```

3.2 Prefix-List



Prefix List를 이용한 경로 차단

- 앞 장에서 22.0.0.0/24를 제외한 경로를 차단하는 방법으로 ACL을 사용하였다.
- 이제, Route 차단 용도로 특화된 Prefix-List를 이용해 보자.
  (More flexible and high performance than ACL)

Prefix-List 기본구조

- **ip prefix-list** name [seq #] **deny** | **permit** prefix [ge/le]
  (Line마다 식별번호(sequence#)가 있어 해당 line만 delete/add 할 수 있다.)

Neighbor에 적용하기

- **neighbor** x.x.x.x **prefix-list** *n* **out**

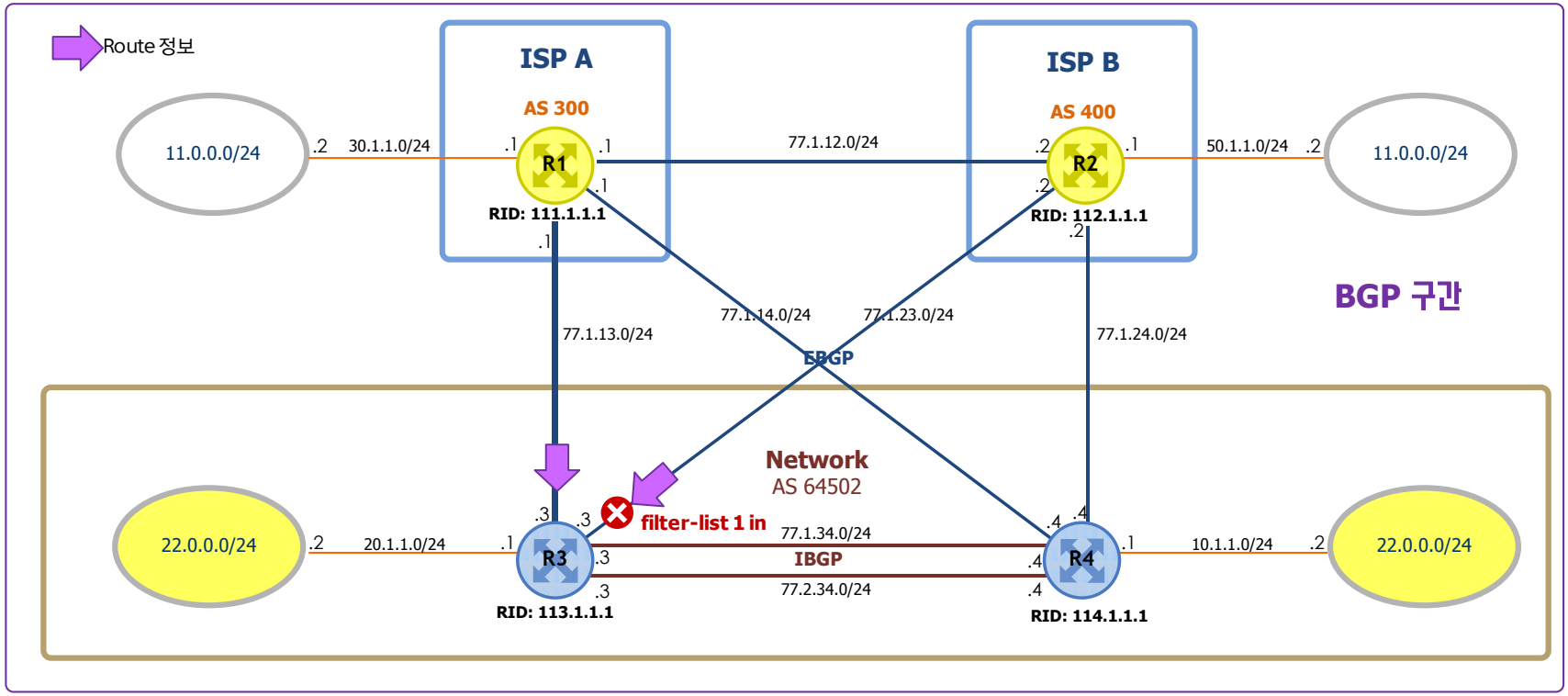
[R3 설정]

```
R3(config)# ip prefix-list AA seq 5 deny 22.0.1.0/24
R3(config)# ip prefix-list AA seq 10 deny 22.0.2.0/24
R3(config)# ip prefix-list AA seq 15 deny 22.0.3.0/24
R3(config)# ip prefix-list AA seq 20 permit 0.0.0.0/0 le 32
R3(config-router)# neighbor 77.1.13.1 prefix-list AA
                    out
```

[R1 확인]

```
R1# show ip bgp
   Network          Next Hop    Metric LocPrf Weight Path
*> 22.0.0.0/24 77.1.13.3(R3) 0             0 64502 ?
```

3.3 Filter-List



AS Path Access List를 이용한 경로 차단

- 이번에는 AS-Path를 이용한 경로차단을 보자.
- 예를 들어, AS 64502 Network의 R3는 11.0.0.0/24로 가는 경로 중 ISP B에서 오는 정보를 차단하고 싶다.
- 이때, AS-Path filter를 이용해 AS-Path 정보에 ISP B(AS 400)가 포함된 경로를 차단해 주면 된다.

AS-Path ACL 기본구조

- **ip as-path access-list** *n* **permit** | **deny** *regular expressions**

Neighbor에 적용하기

- **neighbor** x.x.x.x **filter-list** *n* **out**

[R3 설정]

```
R3(config)# ip as-path access-list 1 deny ^400$
R3(config)# ip as-path access-list 1 permit .*
R3(config-router)# neighbor 77.1.23.2 filter-list 1 in
```

[R3 확인]

```
R3# show ip bgp
   Network          Next Hop    Metric LocPrf Weight Path
*> 11.0.0.0/24      77.1.13.1(R1) 0             0 300 ?
```

4. Route-Map

4.1 Route-Map 설명

* Route-map은 다양한 match 조건과 Attribute setting 및 filter 기능을 하나로 모아놓은 것이다.

Route-map 설정 1 단계 – Match 조건 주기

- 사용 가능한 match 조건은 앞에서 설명한 acl, prefix-list, as-path list 외에도 다양한 조건을 줄 수 있다.

```
R1(config)# route-map AAA permit 10
R1(config-route-map)# match ?
 as-path      Match BGP AS path list
 interface    Match first hop interface of route
 ip           IP specific information
 metric       Match metric of route
 route-type   Match route-type of route
 tag          Match tag of route
```

Route-map 설정 2 단계 – Match 조건에 부합될 경우, Set 설정

- 앞에서 설명한 모든 Attribute 설정을 Route-map에서 설정할 수 있다.

```
R1(config-route-map) # set ?
 aggregator          BGP aggregator attribute
 as-path             Prepend string for a BGP AS-path attribute
 atomic-aggregate    BGP atomic aggregate attribute
 dampening           Enable route-flap dampening
 local-preference    BGP local preference path attribute
 metric              Metric value for destination routing protocol
 metric-type         Type of metric for destination routing protocol
 origin              BGP origin code
 originator-id       BGP originator ID attribute
 tag                 Tag value for destination routing protocol
 weight              BGP weight for routing tablee
```

Route-map 설정 3 단계 – Neighbor에 적용

```
R1(config-router)# neighbor router-id route-map
AAA {out|in}
```

Route-map 설정예제

- 아래와 같은 조건의 Route-map을 만들어 보자.

	Match 조건	Set Action
Weigh 값 변경	ip access-list 1 permit 11.0.0.0/24	set weight 50000
Local Pref. 변경	ip prefix-list AA permit 11.0.0.0/24	set local preference 200
AS-Path 추가	AS 400	set as-path prepend 400
MED값 변경	metric 0	set metric 100

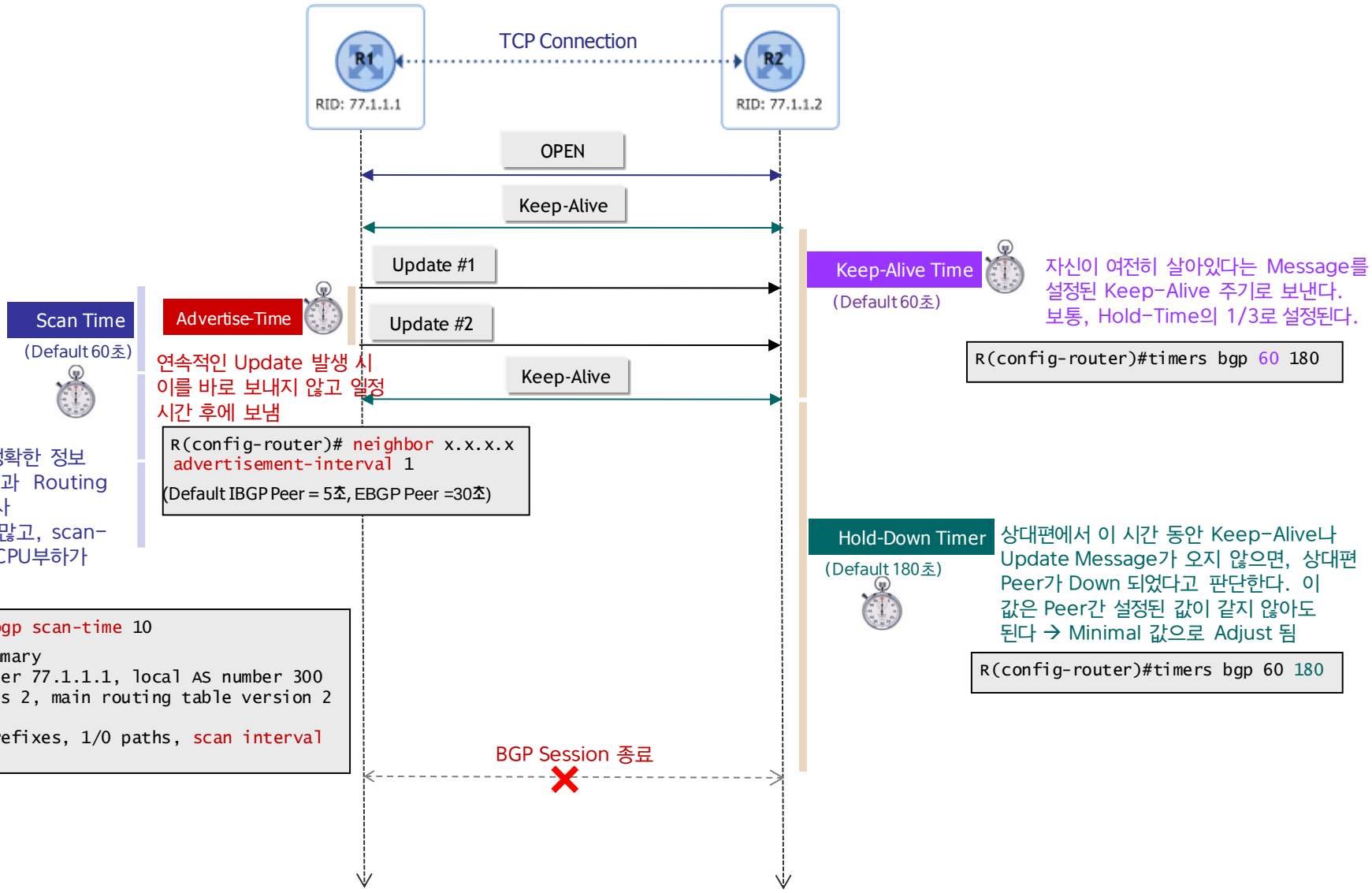
〈Router 설정〉

```
R1(config)# ip access-list 1 permit 11.0.0.0/24
R1(config)# ip prefix-list AA permit 11.0.0.0/24
-----------------------------------------------------
R1(config)# route-map IN permit 10
R1(config-route-map)# match ip address ?
 <1-199>       IP access-list number
 <1300-2699>   IP access-list number (expanded range)
 WORD          IP access-list name
 prefix-list   Match entries of prefix-lists
R1(config-route-map)# match ip address 1
R1(config-route-map)# set weight 50000
R1(config)# route-map IN permit 20
R1(config-route-map)# match as-path 400
R1(config-route-map)# set as-path prepend 400
R1(config)# route-map IN permit 30
R1(config)# route-map OUT permit 10
R1(config-route-map)# match ip address prefix-list AA
R1(config-route-map)# set local-preference 200
R1(config)# route-map OUT permit 20
R1(config-route-map)# match metric 0 metric 100
R1(config)#
R1(config)# route-map OUT permit 30
-----------------------------------------------------
R1(config-route)# neighbor 77.1.12.2 route-map IN in
R1(config-route)# neighbor 77.1.12.2 route-map OUT out
```

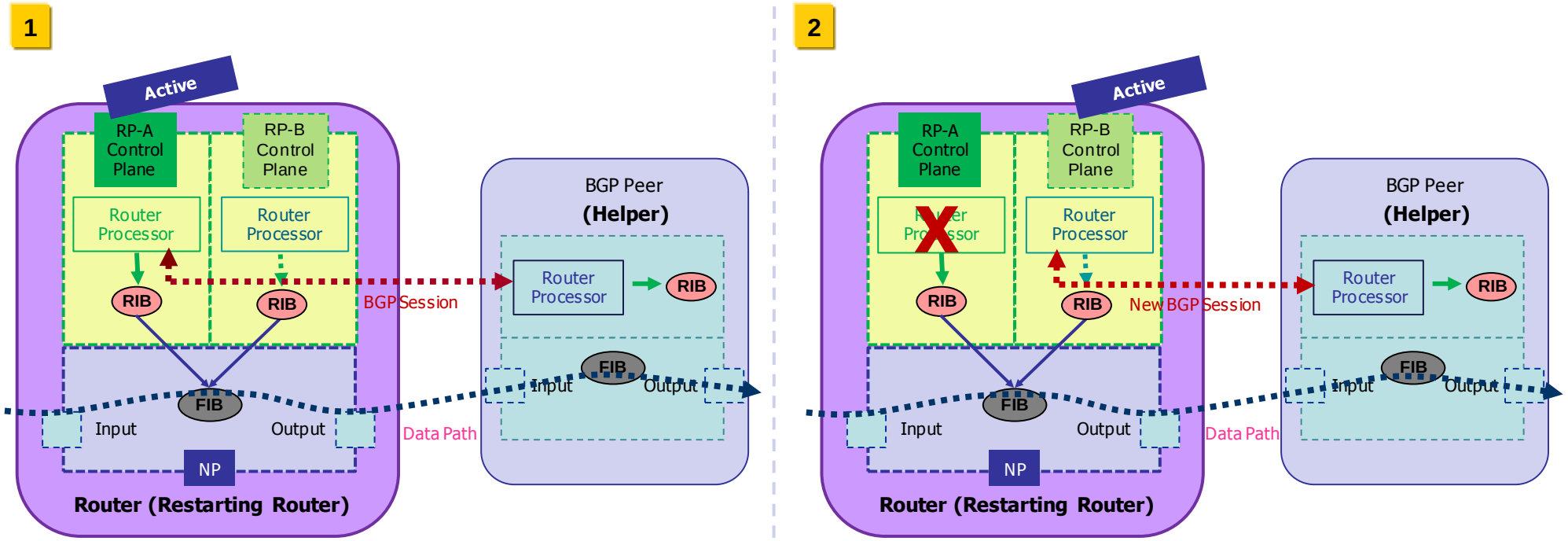
5. BGP Convergence

5.1 BGP Timer
5.2 Graceful Restart

5.1 BGP Timer



5.2 Graceful Restart (1)



Graceful Restart

- Data Plane과 Control Plane을 분리, Control Plane을 Restart 시키더라도 Data Plan은 영향을 받지 않고, 패킷 전송을 계속한다.
- Control Plane fail시, BGP Session은 새로 맺지만 Data Plane (Packet 전송)은 영향을 받지 않음.

그럼, RP-A가 Reset 되었다고 가정하자.

- 상단 그림에서 Data Plane은 변하지 않고, BGP Session을 RP-B로 새로 맺는다 (Graceful Restart Time (Default 100초) 동안 모든 과정을 완료해야 함).

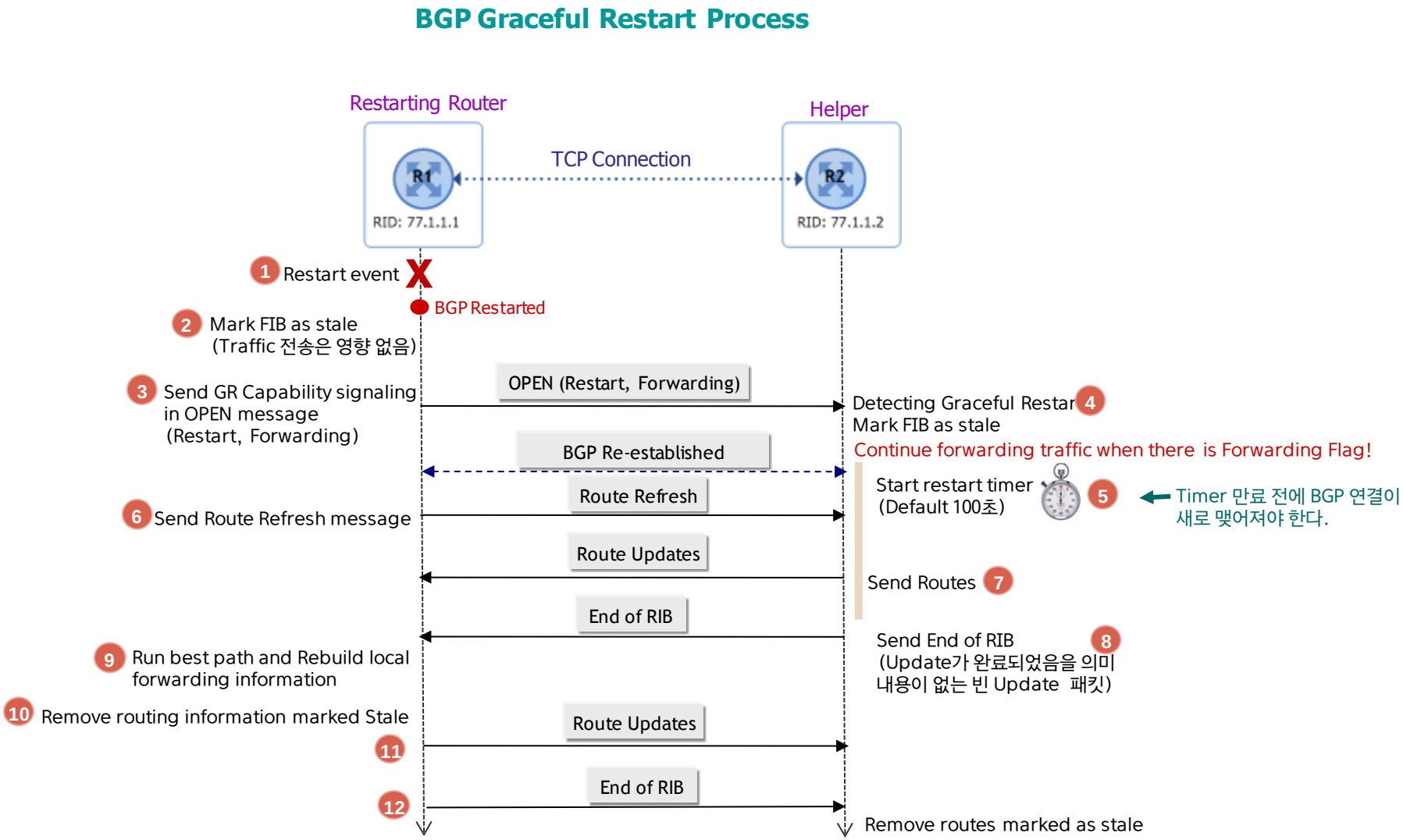
Graceful Restart 관련 명령

```
R(config-router)# enable graceful-restart bgp
(Default enable)

R(config-router)# bgp graceful-restart time 180
Restart 후에 세션을 새로 맺을 때까지 Helper에서 기다려 주는 시간
(Default 100초)

R(config-router)# neighbor 20.1.1.2 capability graceful-restart
Neighbor에게 Graceful Restart 기능을 지원한다는 것을 알려줌.
```

5.2 Graceful Restart (2)



References

- Russ White, Danny McPherson, Srihari Sangli. Practical BGP. Addison Wesley, 2005
- John W. Stewart III. BGP4(Inter-Domain Routing in the Internet). Addison Wesley, 1999
- Jeff Doyle, Jennifer DeHaven Carroll. Routing TCP/IP Volume II. Cisco Press, 2001
- Randy Zhang, Micah Bartell. BGP Design and Implementation. Cisco Press, 2004
- Iljitsch van Beijnum. BGP. O'REILLY, 2002
- Y. Rekhter, T. Li, S. Hares. A Border Gateway Protocol 4 (BGP-4). RFC 4271, 2006
- http://www.cisco.com
- http://www.juniper.net

End of Document