IP, OSPF and BGP in Action

Wolfgang Mühlbauer

wolfgang@net.in.tum.de

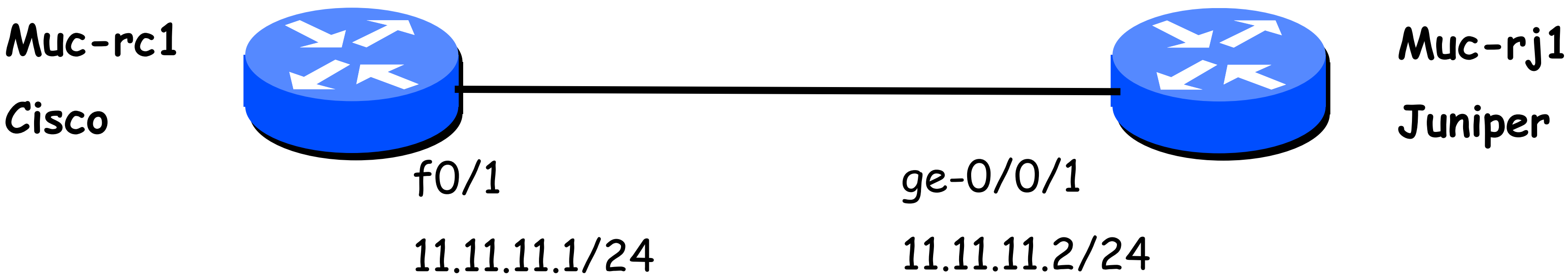
(Some of these these slides are taken from Prof. Anja Feldmann. Thanks!)

Content

- Routing protocols
  - Intra-domain: Open Shortest Path First (OSPF)
  - Inter-domain: Border Gateway Protocol (BGP)
- Configuration of real routers
  - Cisco IOS (C2691)
  - JunOS (M7I)
- Outline
  1. IP configuration
  2. OSPF configuration
  3. BGP configuration

IP Configuration

Demo: Reachability between Interfaces



Configure IP addresses:

- Cisco IOS

```
interface FastEthernet0/1
    ip address 11.11.11.1 255.255.255.0
```

- Juniper JunOS

```
set interfaces ge-0/0/1 family inet
address 11.11.11.2/24
```

Demo: Cisco/Juniper Configurations

Muc-rc1 (Cisco)

```
muc-rc1#show run
Building configuration...

Current configuration : 1560 bytes
!
version 12.3
hostname muc-rc1
!
interface FastEthernet0/1
 ip address 11.11.11.1 255.255.255.0
 duplex auto
 speed auto
!
end
```

Muc-rj1 (Juniper)

```
root@muc-rj1# show
## Last changed: 2008-11-12 00:16:34
    UTC
version 8.2R1.7;
interfaces {
   ge-0/0/1 {
      unit 0 {
         family inet {
            address 11.11.11.2/24;
         }
      }
   }
}

[edit]
```

Demo: Testing Connectivity (1)

- Ping
  - From muc-rc1 to muc-rj1:

```
muc-rc1#ping 11.11.11.2
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 11.11.11.2,
timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip
min/avg/max = 1/1/1 ms
```

  - From muc-rj1 to muc-rc1:

```
root@muc-rj1# run ping 11.11.11.1
PING 11.11.11.1 (11.11.11.1): 56 data bytes
64 bytes from 11.11.11.1: icmp_seq=0 ttl=255 time=3.427 ms
```

Demo: Testing Connectivity (2)

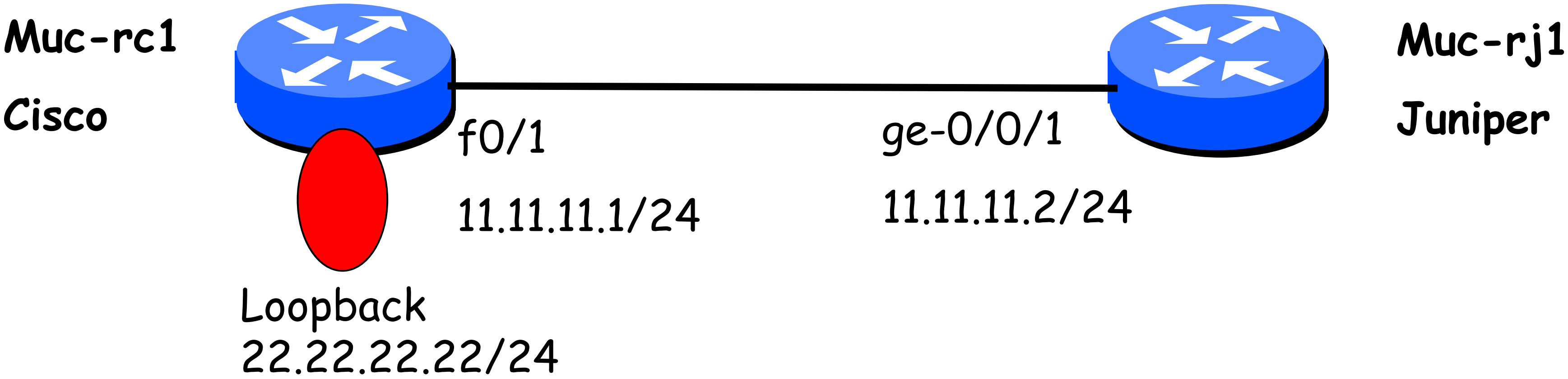
- Routing Table at muc-rc1

```
muc-rc1#show ip route
Codes: C - connected, S - static, R - RIP, M - mobile, B -BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 …
       ia - IS-IS inter area, * - candidate default,
       u - per-user static route o - ODR, P - periodic downloaded
       static route

Gateway of last resort is not set

     11.0.0.0/24 is subnetted, 1 subnets
C       11.11.11.0 is directly connected, FastEthernet0/1
```

Demo: Adding a Loopback Interface

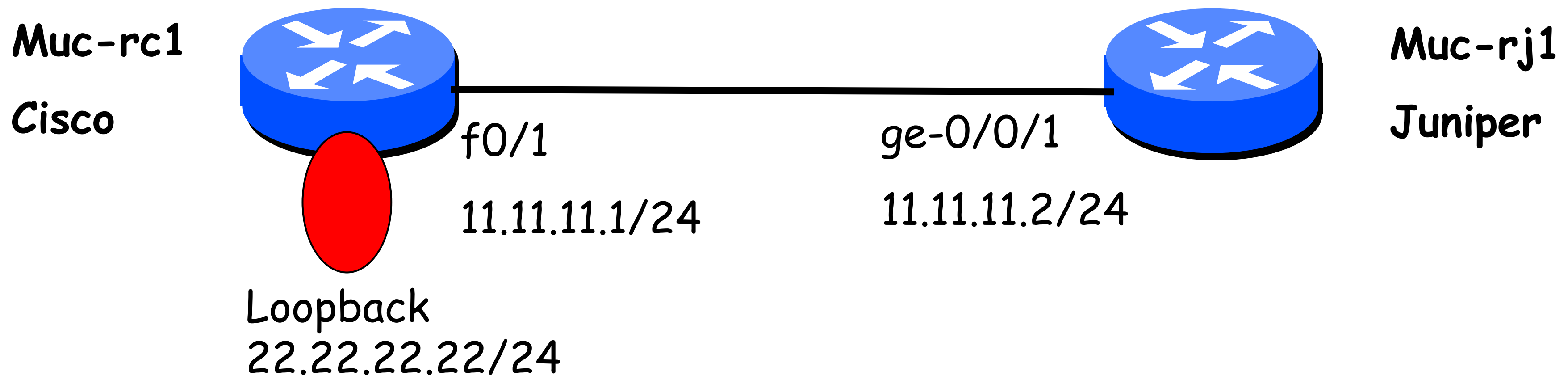


- Loopback interfaces
  - Required for many features
  - Describes routers as a whole (not "sum of interfaces")
  - Configuration:

```
interface Loopback0
  ip address 22.22.22.22 255.255.255.0
```

- Why can't muc-rj1 ping 22.22.22.22?

Demo: Adding a Loopback Interface (2)



- Why can't muc-rj1 ping 22.22.22.22?
  - Muc-rj1 does not have a route
- Add static route at muc-rj1 to Loopback
  - `set routing-options static route 22.22.22.0/24 next-hop 11.11.11.1`

OSPF Configuration

"Let's use OSPF rather than a static route"

OSPF (Open Shortest Path First)

- "Open": specification publicly available
  - RFC 1247, RFC 2328
  - Working group formed in 1988
  - Goals:
    - Large, heterogeneous internetworks
- Uses the Link State algorithm
  - Topology map at each node
  - Route computation using Dijkstra's algorithm
- Hierarchy
  - Multiple areas to keep routing scalable

OSPFv2: Tasks

- Neighbors
  - Discovery
  - Maintenance
- Database
  - Granularity
  - Maintenance
  - Synchronization
- Routing table
  - Metric
  - Calculation

Neighbors - Discovery and Maintenance

- Hello Protocol
  - Ensures that neighbors can send packets to and receive packets from the other side: bi-directional communication
  - Ensures that neighbors agree on parameters (HelloInterval and RouterDeadInterval)
- How
  - Hello packet to fixed well-known multicast address
  - Periodic Hellos
  - Broadcast network: electing designated router

Demo: Hello Protocol

- OSPF configuration:
  - Cisco

```
router ospf 1000
 log-adjacency-changes
 network 0.0.0.0 255.255.255.255 area 0
```

  - Juniper

```
protocols {
    ospf {
        area 0.0.0.0 {
            interface ge-0/0/1.0;}}}
```

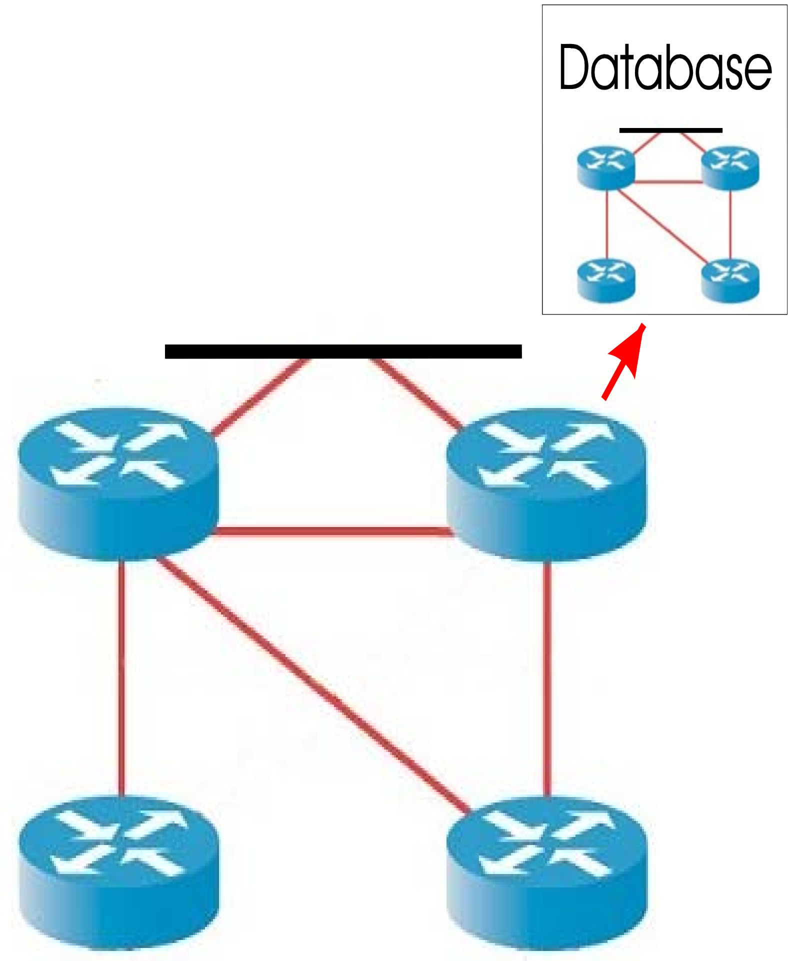
- Monitor "Hello" packets at Cisco interface

```
01:30:31.193995 Out IP 11.11.11.2 > 224.0.0.5:
          OSPFv2, Hello, length 48
```

- Observations
  - Hello packet sent every 10 seconds
  - Hello packets sent to multicast address (224.0.0.5)

Link State Database

- Based on link-state technology
  - Local view of topology in a database
- Database
  - Consists of Link State Advertisements (LSA)
  - LSA: data unit describing local state of a network/router)
  - Must kept synchronized to react to routing failures
- Currently empty



Link State Database: Example

LS-Type	*Link State ID*	*Adv. Router*	*Checksum*	*Seq. No.*	*Age*
Router-LSA	10.1.1.1	10.1.1.1	0x9b47	0x80000006	0
Router-LSA	10.1.1.2	10.1.1.2	0x219e	0x80000007	1618
Router-LSA	10.1.1.3	10.1.1.3	0x6b53	0x80000003	1712
Router-LSA	10.1.1.4	10.1.1.4	0xe39a	0x8000003a	20
Router-LSA	10.1.1.5	10.1.1.5	0xd2a6	0x80000038	18
Router-LSA	10.1.1.6	10.1.1.6	0x05c3	0x80000005	1680

LSAs

- Consists of a Header and a Body
- Header size is 20 Byte and consists of

<table>
<tr><td colspan="4">0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1</td></tr>
<tr><td colspan="2">LS Age</td><td>Options</td><td>LS Type</td></tr>
<tr><td colspan="4">Link State ID</td></tr>
<tr><td colspan="4">Advertising Router</td></tr>
<tr><td colspan="4">LS sequence number</td></tr>
<tr><td colspan="2">LS Checksum</td><td colspan="2">Length</td></tr>
</table>

LSAs (2.)

- Identifying LSAs
  - LS Type Field
    - Router LSA: Describes a router
    - Network LSA: Describes a network (e.g., Ethernet segment)
  - Link State ID Field
  - Advertising Router Field
- Verifying LSA Contents: LS Checksum Field
- Identifying LSA Instances
  (keep in mind that the topology changes)
  - LS Sequence Number Field
    - Linear sequence space

LSAs (3.)

- LS Age Field
  (to ensure consistency)
  - Goal: new sequence number every 30 minutes
  - Maximum value 1 hour
  - Age > 1 hour ⇨ invalid ⇨ removal
  - Enables premature aging
  - Ensures removal of outdated information

Demo: Link State Database

- Cisco

```
muc-rc1#show ip ospf database
         OSPF Router with ID (22.22.22.22) (Process ID 1000)
                Router Link States (Area 0)
Link ID         ADV Router      Age        Seq#      Checks.  #Links
11.11.11.2      11.11.11.2      75        0x80000002 0x00EFB6 1
22.22.22.22     22.22.22.22     75        0x80000002 0x00AD26 2
                Net Link States (Area 0)
Link ID         ADV Router      Age          Seq#    Checksum
11.11.11.1      22.22.22.22     75        0x80000001 0x000A3A
```

- Juniper

```
root@muc-rj1# run show ospf database
    OSPF link state database, Area 0.0.0.0
 Type        ID          Adv Rtr        Seq             Age  Opt  Cksum  Len
Router   *11.11.11.2   11.11.11.2    0x80000002    318  0x22 0xefb6   36
Router    22.22.22.22  22.22.22.22   0x80000002    321  0x22 0xad26   48
Network   11.11.11.1   22.22.22.22   0x80000001    320  0x22 0xa3a    32
```

Database Synchronization

- Central aspect: all routers need to have identical databases!
- 2 types of synchronization
  - Initial synchronization
    - After hello
  - Continuous synchronization
    - Flooding

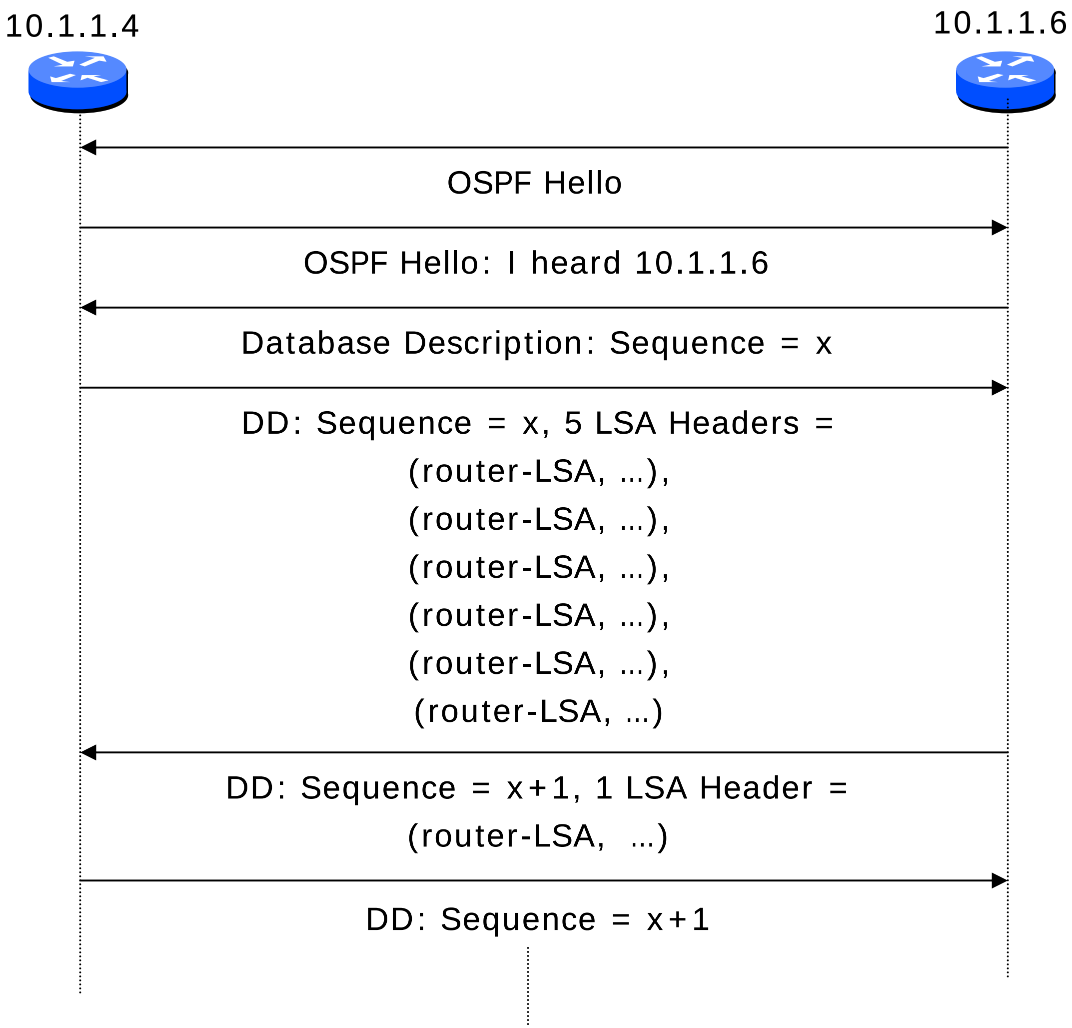
Initial Synchronization

- Explicit transfer of the database upon establishment of neighbor ship
- Once bi-directional communication exists
- Send all LS header from database to neighbor
  - OSPF database description packets (DD pkt)
  - Flood all future LSA’s

Initial Synchronization (2.)

- Database description (DD) exchange
  - Only one DD at a time
  - Wait for Ack
- Control of DD exchange
  - Determine Master/Slave for DD exchange
  - Determine which LSA's are missing in own DB
  - Request those via link state request packets
  - Neighbor sends these in link state update packets
- Result:
  - Fully adjacent OSPF neighbors

Example: Database Synchronization



Continuous Synchronization - Flooding

- Reliable flooding
  - LSAs must be acknowledged (implicit or explicit)
- Robustness
  - LSAs are aged
  - LSAs have checksums
  - LSAs cannot be sent at an arbitrary rate
    there are timers

Demo: Initial Database Synchronization

- Messages exchanged after re-establishing neighborhood between muc-rc1 and muc-rj1

```
01:30:31.193995 Out IP 11.11.11.2 > 224.0.0.5: OSPFv2, Hello, length 48
01:30:31.307993  In IP 11.11.11.1 > 11.11.11.2: OSPFv2, Database
   Description, length 44
01:30:31.308782 Out IP 11.11.11.2 > 11.11.11.1: OSPFv2, Database
   Description, length 32
01:30:31.308838  In IP 11.11.11.1 > 11.11.11.2: OSPFv2, LS-Request,
   length 48
01:30:31.358461 Out IP 11.11.11.2 > 11.11.11.1: OSPFv2, LS-Update,
   length 112
01:30:31.408428 Out IP 11.11.11.2 > 224.0.0.5: OSPFv2, LS-Update,
   length 60
01:30:31.513376  In IP 11.11.11.1 > 224.0.0.5: OSPFv2, LS-Update,
   length 76
01:30:31.513662 Out IP 11.11.11.2 > 11.11.11.1: OSPFv2, LS-Update,
   length 76
01:30:33.268557 Out IP 11.11.11.2 > 224.0.0.5: OSPFv2, LS-Update,
   length 64
01:30:33.278724  In IP 11.11.11.1 > 11.11.11.2: OSPFv2, LS-Ack, length
   44
01:30:33.938660  In IP 11.11.11.1 > 224.0.0.5: OSPFv2, LS-Ack, length
   104
```

Calculation of routing table

- Link state database is a directed graph with costs for each link
- Dijkstra‘s SPF algorithms
  - Add all routers to shortest-path-tree
  - Add all neighbors to candidate list
  - Add routers with the smallest cost to tree
  - Add neighbors of this router to candidate list
    - If not yet on it
    - If cost smaller
  - Continue until candidate list empty

OSPF: Summary

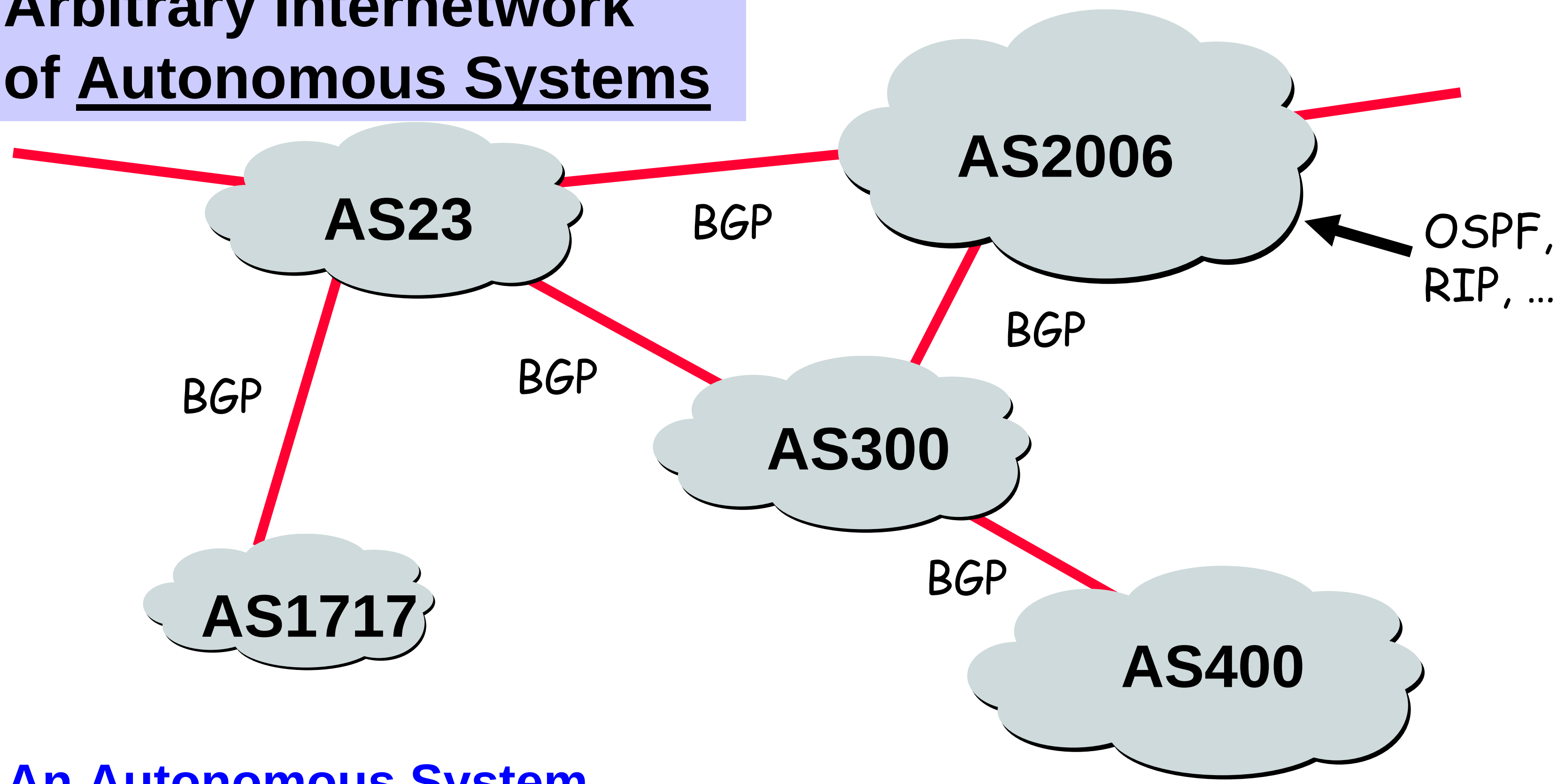
- Neighbors
  - Discovery — Multicast group
  - Maintenance — Hello protocol
- Database
  - Granularity — Link state advertisements (LSA)
  - Maintenance — LSA-updates flooding protocol
  - Synchronization — Synchronization protocol
- Routing table
  - Metric — Fixed values
  - Calculation — Local shortest path calculation

BGP Configuration

"Let's use BGP rather than OSPF"

Current Internet Architecture

Arbitrary Internetwork of Autonomous Systems



An Autonomous System is a unified administrative domain with a consistent routing policy

Currently there are approximately 30,000 ASs

Why Different Intra- and Inter-AS Routing?

Policy:

- Inter-AS: admin wants control over how its traffic routed, who routes through its net.
- Intra-AS: single admin, so no policy decisions needed

Scale:

- hierarchical routing saves table size, reduced update traffic

Performance:

- Intra-AS: can focus on performance
- Inter-AS: policy may dominate over performance

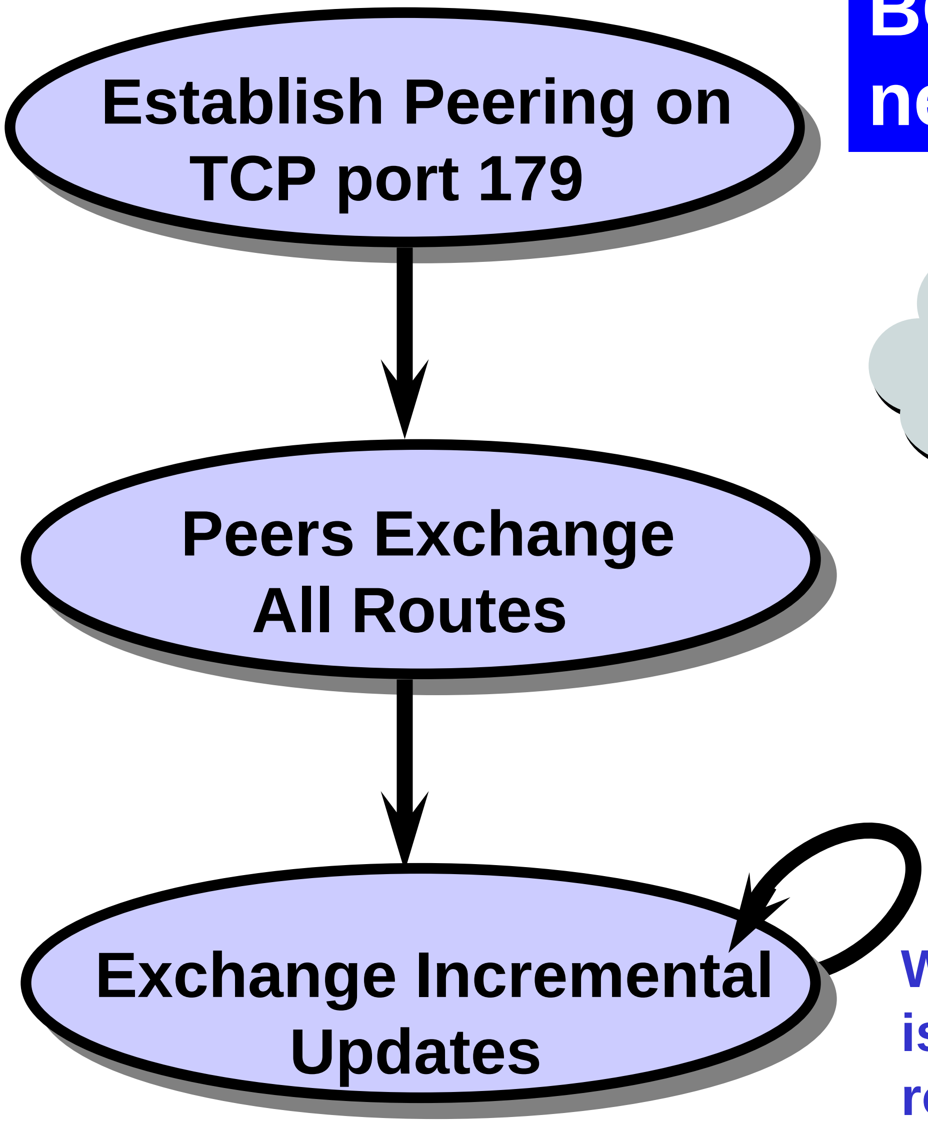
We need BOTH!

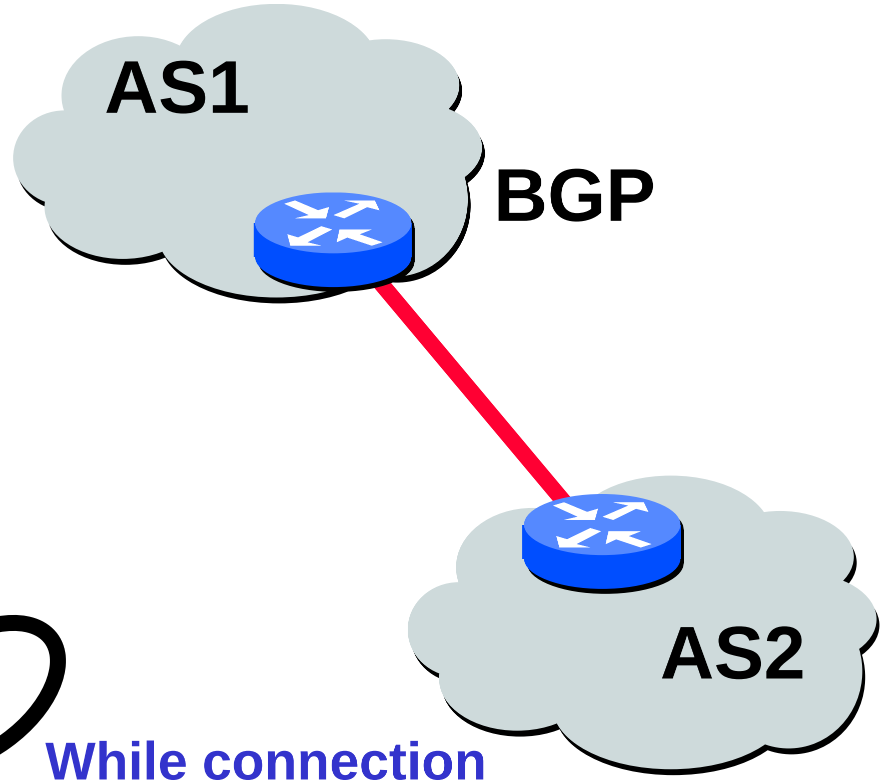
Internet Inter-AS Routing: BGP

- BGP (Border Gateway Protocol): *the* de facto standard
- BGP provides each AS a means to:
  1. Obtain subnet reachability information from neighboring ASs.
  2. Propagate the reachability information to all routers internal to the AS.
  3. Determine "good" routes to subnets based on reachability information and **policy**.
- Allows a subnet to advertise its existence to rest of the Internet: *"I am here"*

BGP Operations Simplified

BGP Route =
network prefix + attributes





While connection is ALIVE exchange route UPDATE messages

BGP Messages

Peers exchange BGP messages using TCP

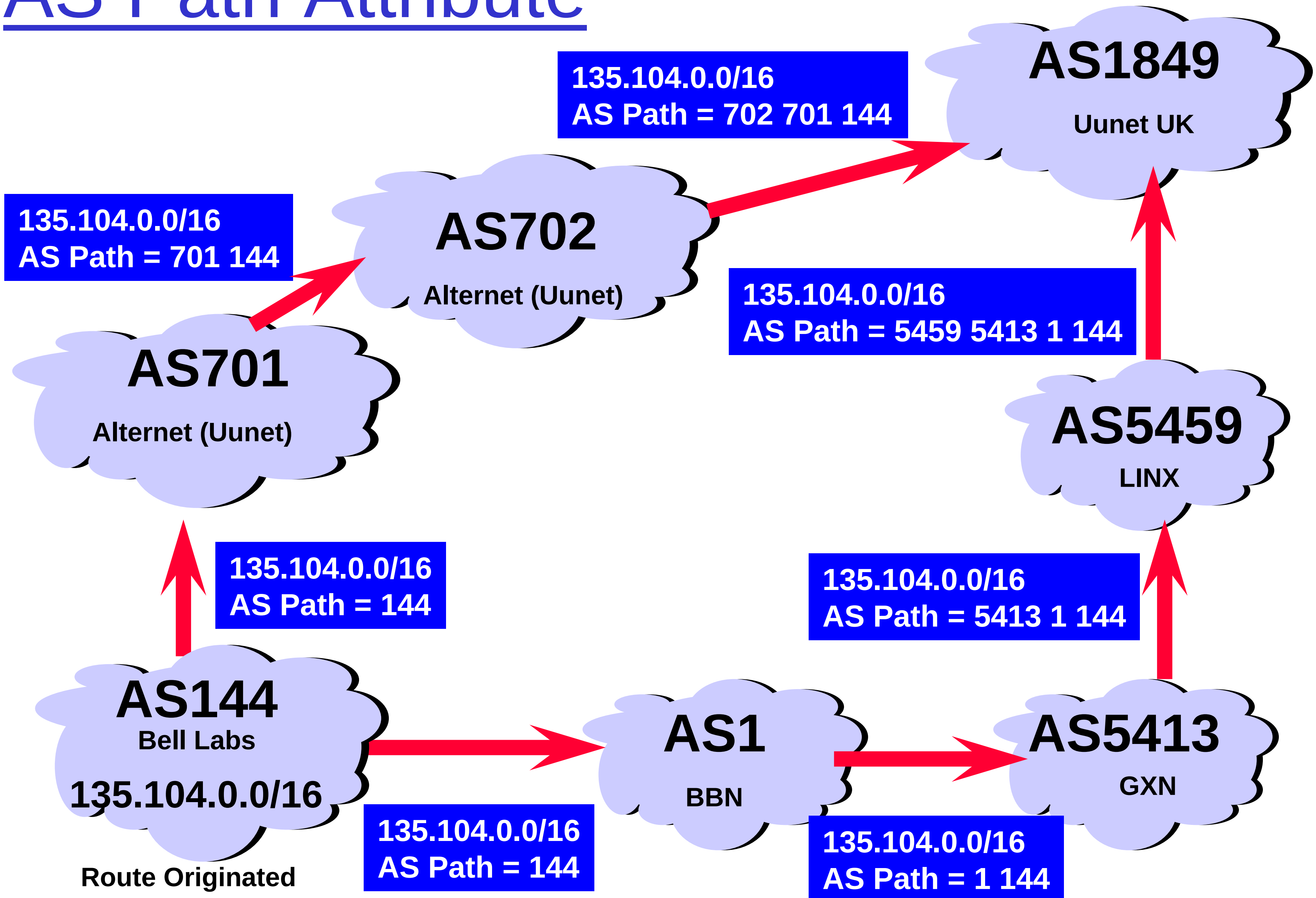
BGP messages:

- OPEN:
  - opens TCP conn. to peer
  - authenticates sender
- UPDATE:
  - advertises new path (or withdraws old)
- KEEPALIVE:
  - keeps conn alive in absence of UPDATES
  - serves as ACK to an OPEN request
- NOTIFICATION:
  - reports errors in previous msg;
  - closes a connection

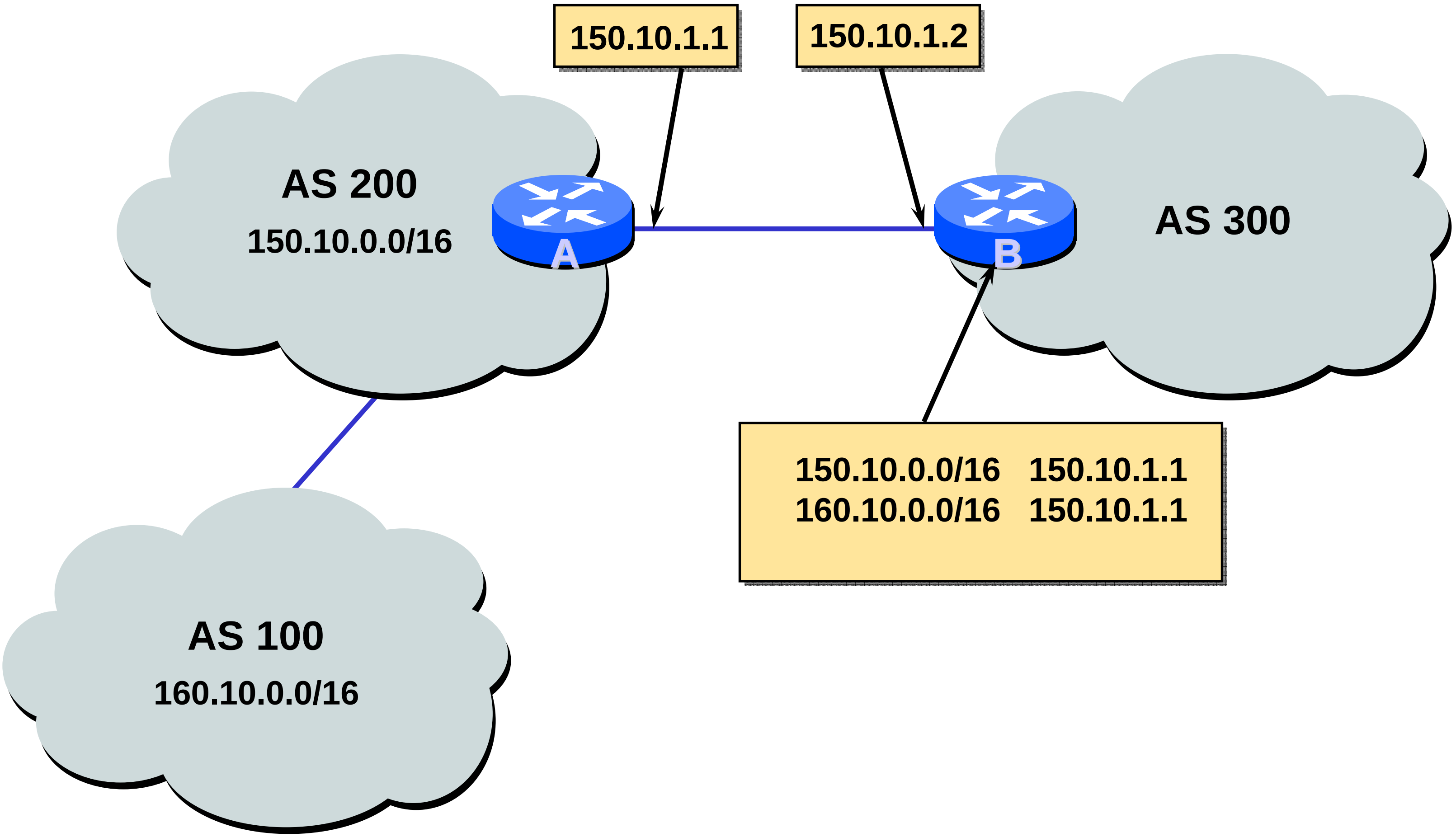
Path Attributes & BGP Routes

- When advertising a prefix, advertisement/update includes BGP attributes.
  - prefix + attributes = "route"
- Two important attributes:
  - AS-PATH: contains the ASs through which the advertisement for the prefix passed: AS 67 AS 17
    - used for loop detection / policies
  - NEXT-HOP: Indicates the specific internal-AS router to next-hop AS. (There may be multiple links from current AS to next-hop-AS.)
- When gateway router receives route advertisement, uses import policy to accept/decline.

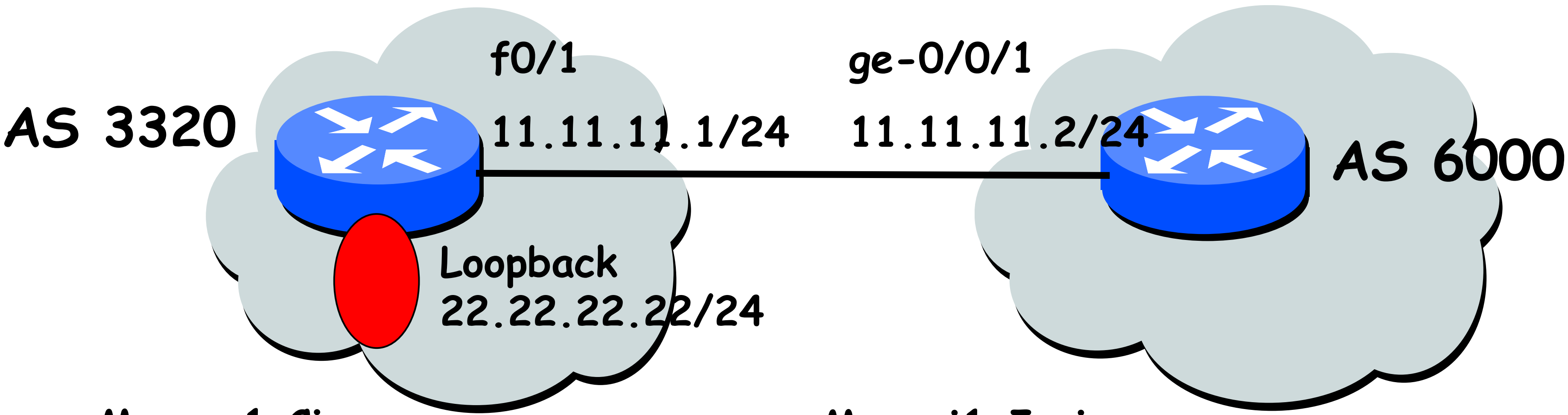
AS Path Attribute



Next Hop Attribute



Demo: BGP Configuration



Muc-rc1 Cisco

```
router bgp 3320
  no synchronization
  bgp log-neighbor-changes
  network 11.11.11.0 mask
        255.255.255.0
  network 22.22.22.0 mask
        255.255.255.0
  neighbor 11.11.11.2 remote-
        as 6000
  no auto-summary
```

Muc-rj1 Juniper

```
routing-options {
  autonomous-system 6000;
}
protocols {
  bgp {
    group gr-3320 {
      type external;
      peer-as 3320;
      neighbor 11.11.11.1;
    }
  }
}
```

Demo: BGP Routing Table

- Routing Table at muc-rj1

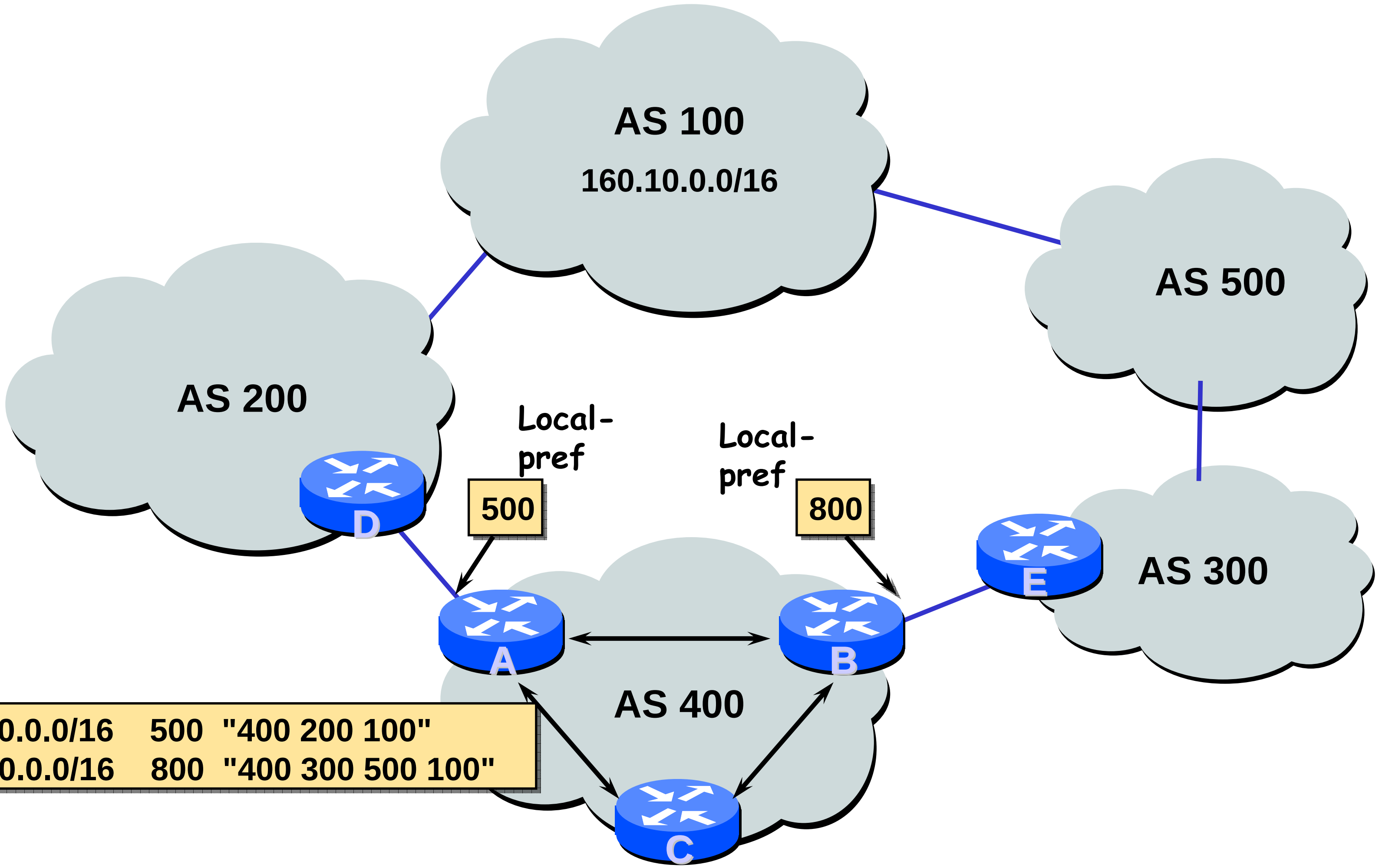
```
root@muc-rj1# run show route
inet.0: 3 destinations, 4 routes (3 active,
          0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

22.22.22.0/24 *[BGP/170] 00:00:08,
                MED 0, localpref 100
               AS path: 3320 I
              > to 11.11.11.1 via ge-0/0/1.0
```

BGP Route Selection

- Router may learn about more than one route to some prefix.
- Router must select route.
- Elimination rules:
  1. Highest local preference wins
  2. Shortest AS-PATH
  3. ... (many other criteria) ...
  4. Tie-breaking
     - E.g., pick route from router with lowest IP address

Local Preference Attribute



	160.10.0.0/16	500	"400 200 100"
>	160.10.0.0/16	800	"400 300 500 100"

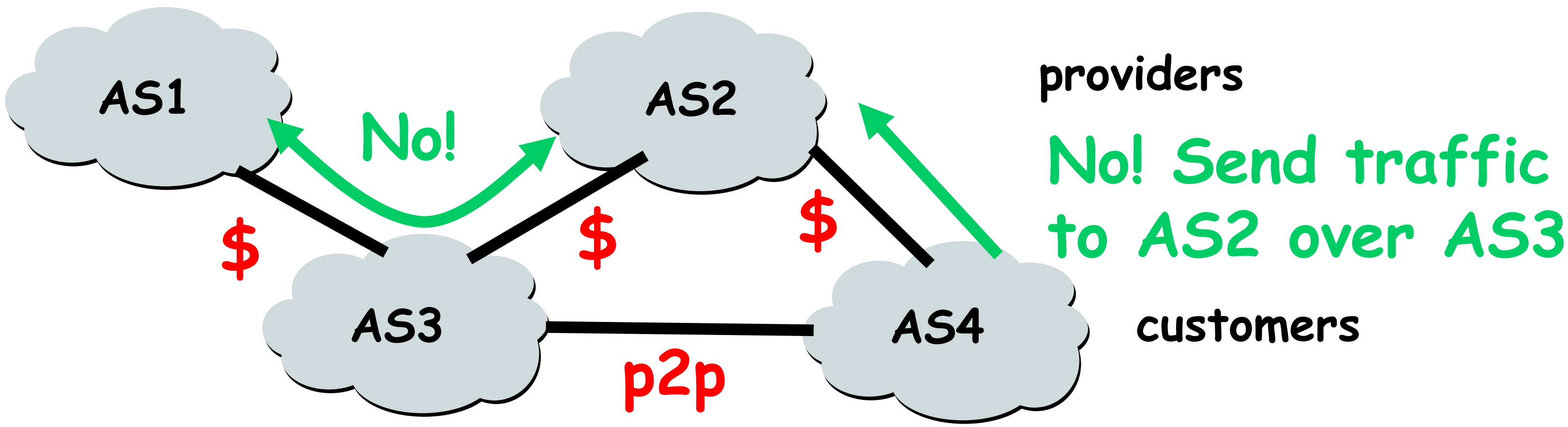
- Path with highest local preference wins although longer

Routing Policy

- Reflects goals of network provider
  - Which routes to accept from other ASes
  - How to manipulate the accepted routes
    - E.g., local preferences
  - How to propagate routes through network
  - How to manipulate routes before they leave the AS
  - which routes to send to another AS

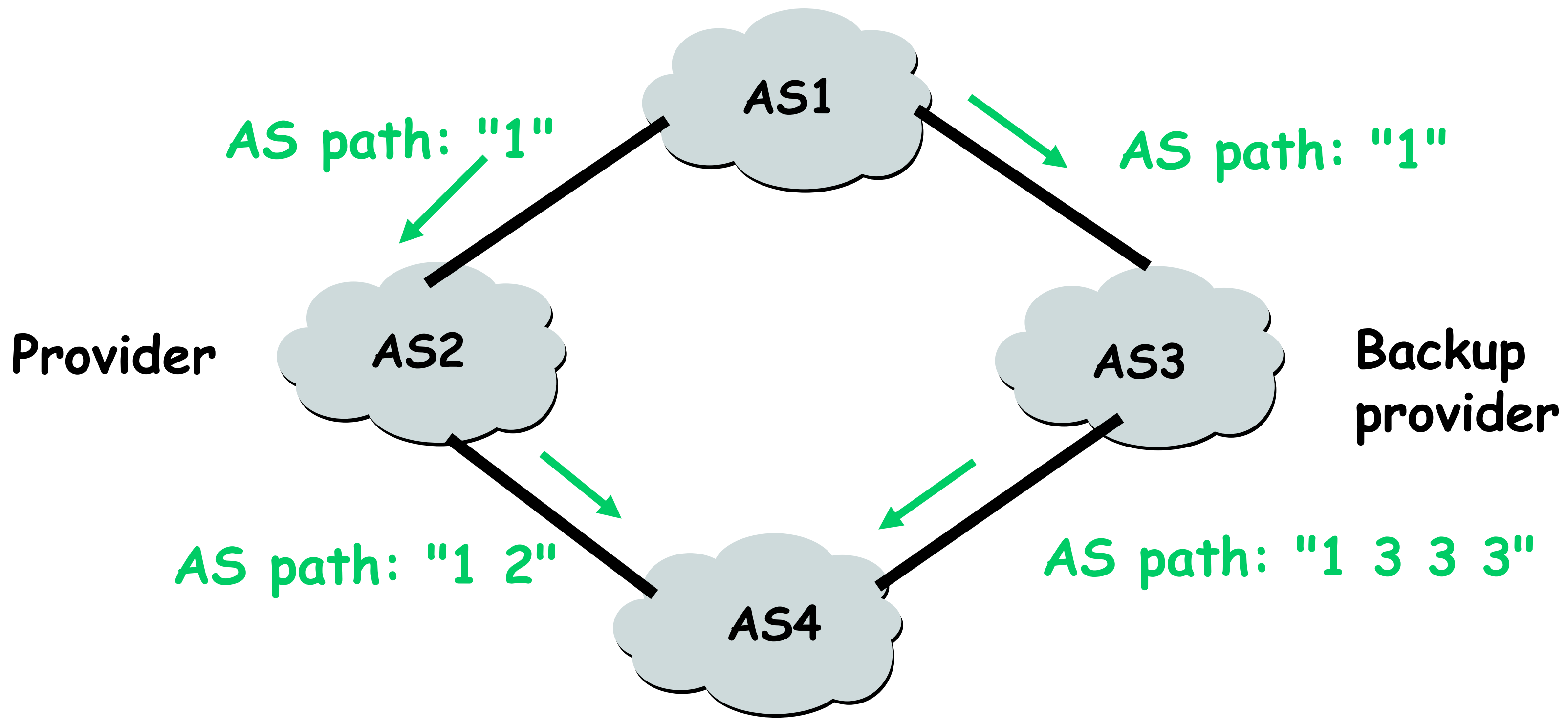
Policy Example – Business Relationships

- Types of relationships
  - Customer-provider: Customer pays for connectivity
  - Peer-to-peer (p2p): Costs are shared
- Consequences for routing
  - Prefer customer routes over p2p over provider routes
  - No export of provider routes to other providers

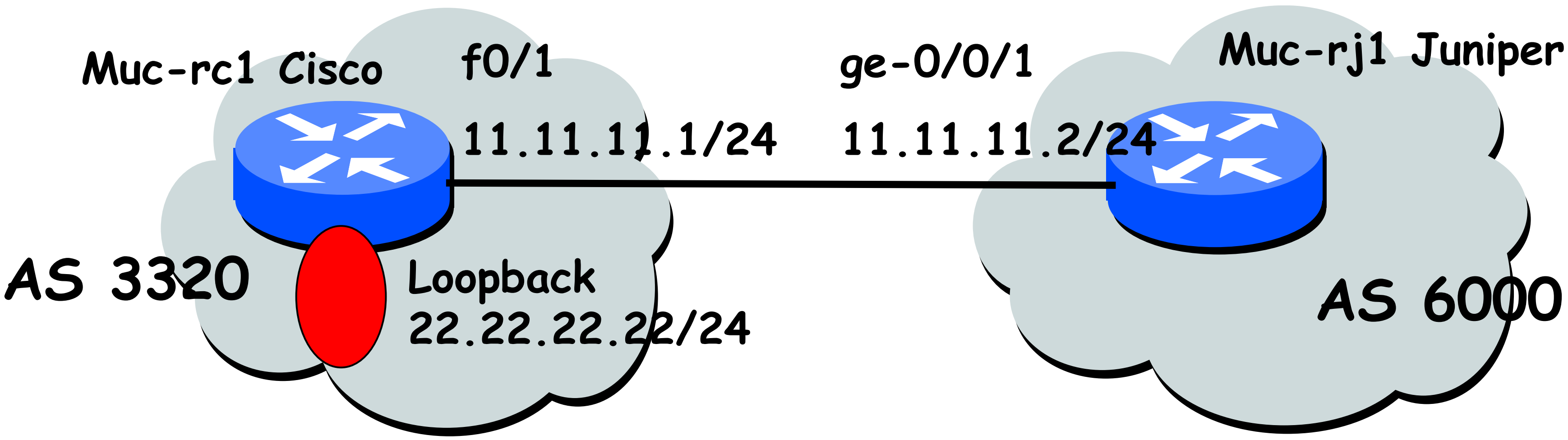


Policy Example – Backup Path

- AS 3 is backup provider (only used if AS 2 fails)
  - It prepends its AS number 3 times and thus makes AS path artificially longer



Demo: AS Path Prepending



- Make AS path artificially longer
  - AS 3320 prepends its AS number twice

```
Router bgp 3320
 neighbor 11.11.11.2 route-map prepend out
route-map prepend permit 10
  set as-path prepend 3320 3320
```

- Result at muc-rj1
  - `22.22.22.0/24   AS path: 3320 3320 3320`

Literature

- Andrew S. Tanenbaum, *Computer Networks*, 4th edition, 2003, Prentice Hall
- J.F. Kurose, K.W. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet*, 4th edition, 2007, Addison Wesley
- Further reading (if you want to know more)
  - John T. Moy, *OSPF - Anatomoy of an Internet Routing Protocol*, 1998, Addison-Wesley
  - John Stewart, *BGP – Inter-Domain Routing in the Internet*, 1998, Addison-Wesley

Summary

- Routing protocols
  - Open Shortest Path First (OSPF)
  - Border Gateway Protocol (BGP)
- Demos
  - IP configuration
  - OSPF configuration
  - BGP configuration
- **Questions?**

- **End title: … show tech …** ☺