Networking Technologies and Applications

Rolland Vida
BME TMIT

October 18, 2017

Link-state protocols

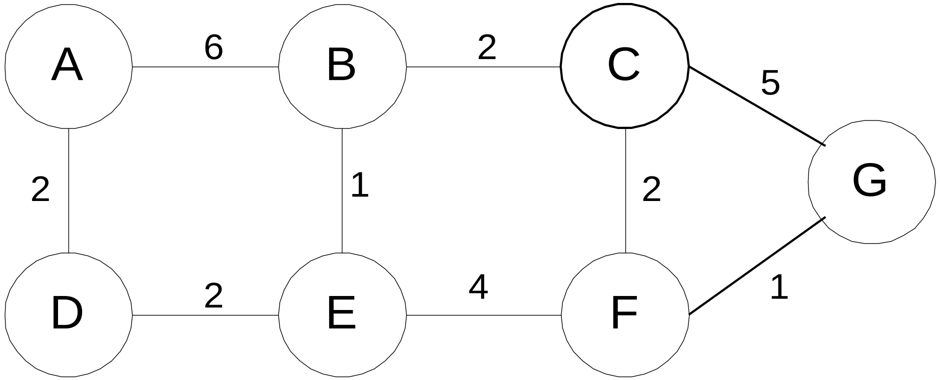
Operation of link-state protocols

- The operation of link-state protocols has two steps:

1. Each node discovers the network topology
   - Link state records advertised in the network
2. In the obtained graph it finds the shortest path and the next hop on the path

– Important!
   - The topology in each router should be the same
   - Finding the optimal path is done in the same way, in each node
     – If node A thinks the optimal route goes through B, and B thinks it goes through A – a loop is formed!

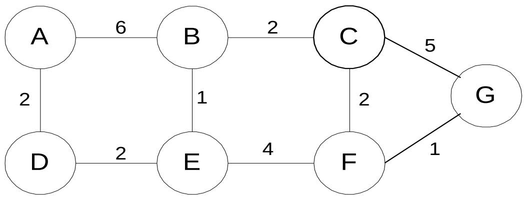
Link State Database

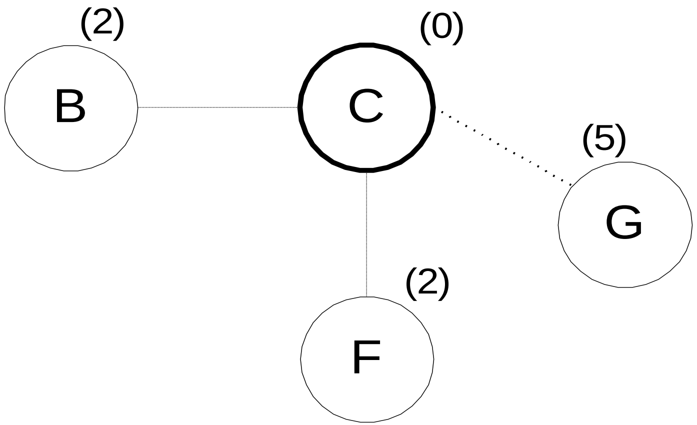


Link state Database						
A	B	C	D	E	F	G
B/6	A/6	B/2	A/2	B/1	C/2	C/5
D/2	C/2	F/2	E/2	D/2	E/4	F/1
	E/1	G/5		F/4		G/1

Dijkstra algorithm

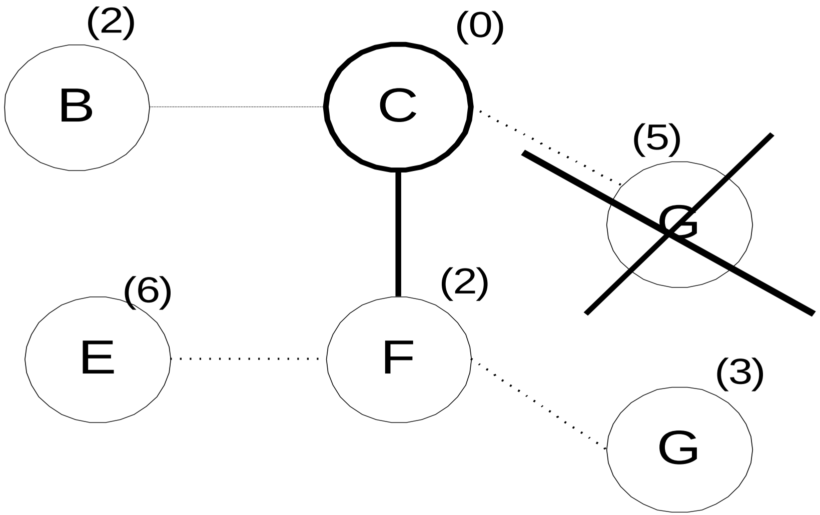
- Route selection based on the Dijkstra algorithm
  - Let C be the root
  - Let's calculate the cost of the paths to our neighbors

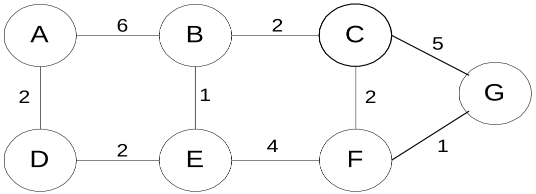




Dijkstra algorithm 2

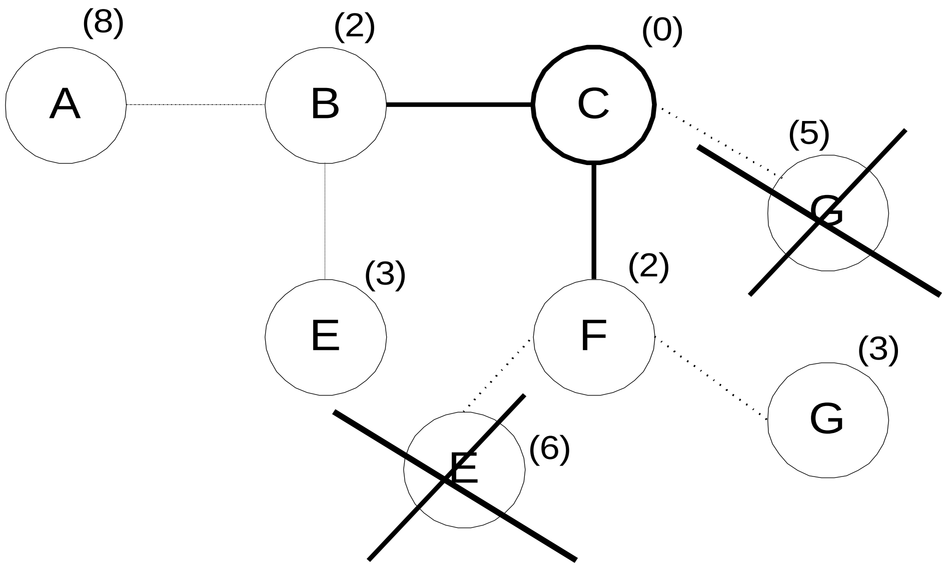
- Let's consider node F (the smallest cost, non-visited neighbor) and calculate the costs of the paths to the neighbors of F
- Shorter path to G through F. Node E gets in the picture

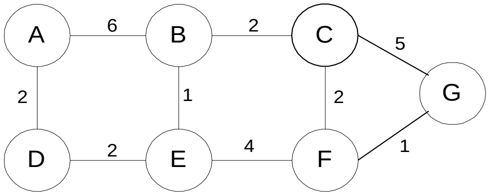




Dijkstra algorithm 3

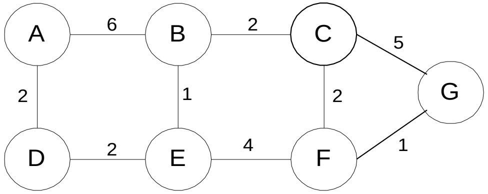
- Let's consider node B, and calculate the costs to its neighbors
- Shorter path to E through B. Node A gets in the picture

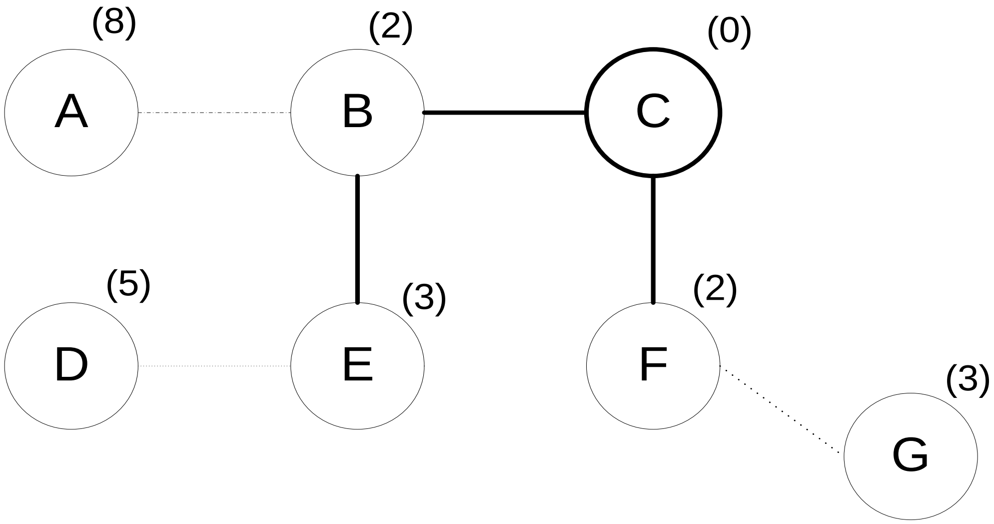




Dijkstra algorithm 4

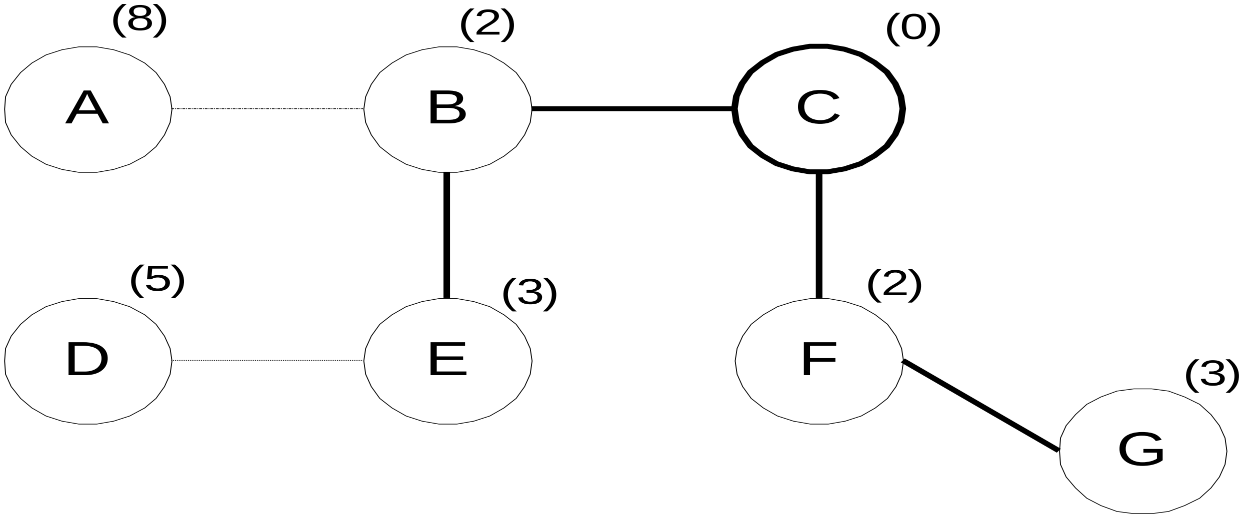
- Let's consider node E, and calculate the costs to its neighbors
- No changes, node D gets in the picture

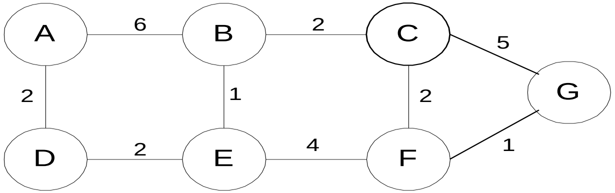




Dijkstra algorithm 5

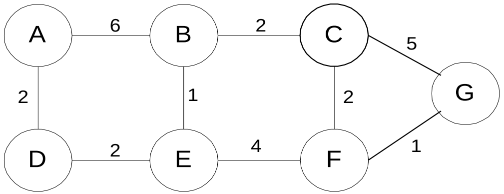
- Consider node G, and calculate the costs
- No changes

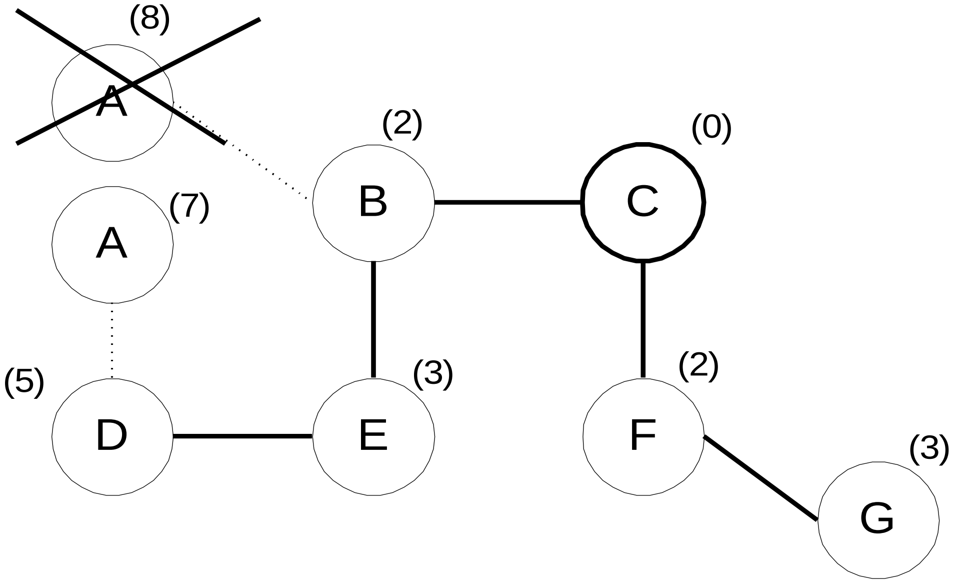




Dijkstra algorithm 6

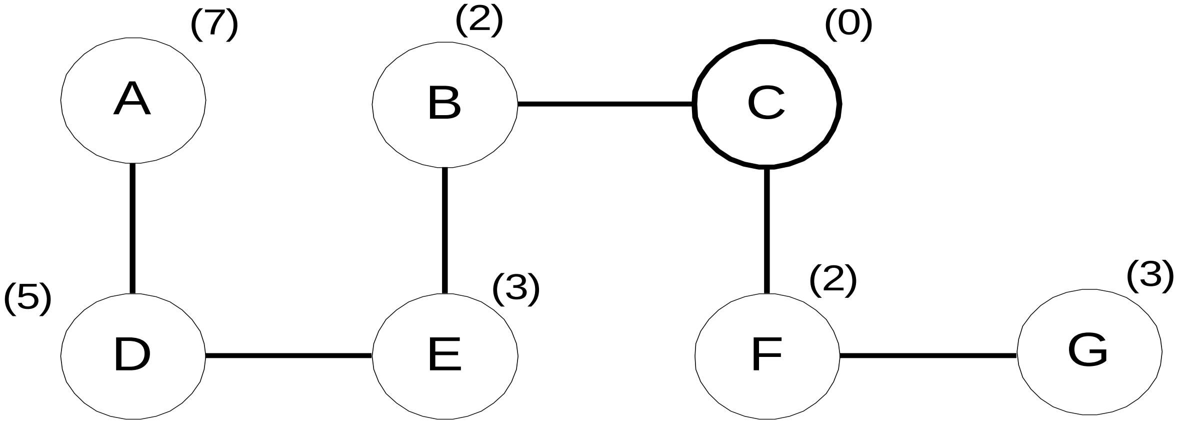
- Consider node D, and calculate the costs to its neighbors
- Shorter path to node A!

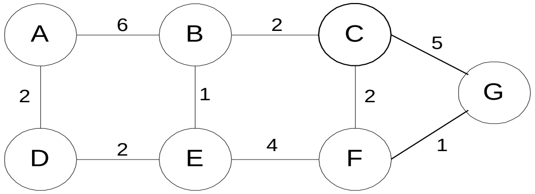




Dijkstra algorithm 7

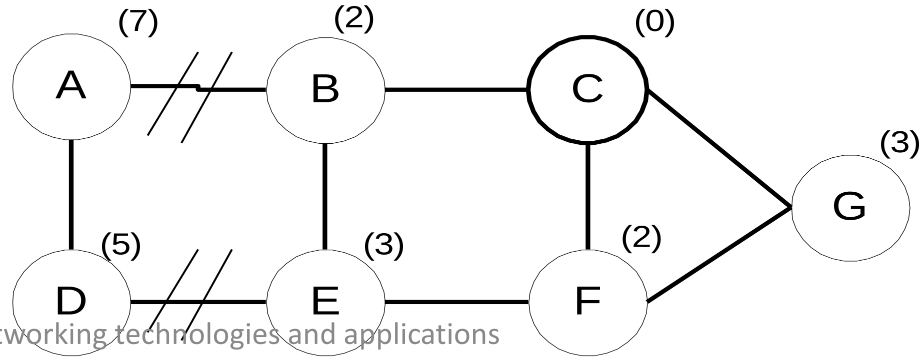
- Consider node A, and calcuate the costs
- No more neighbors
- End of story





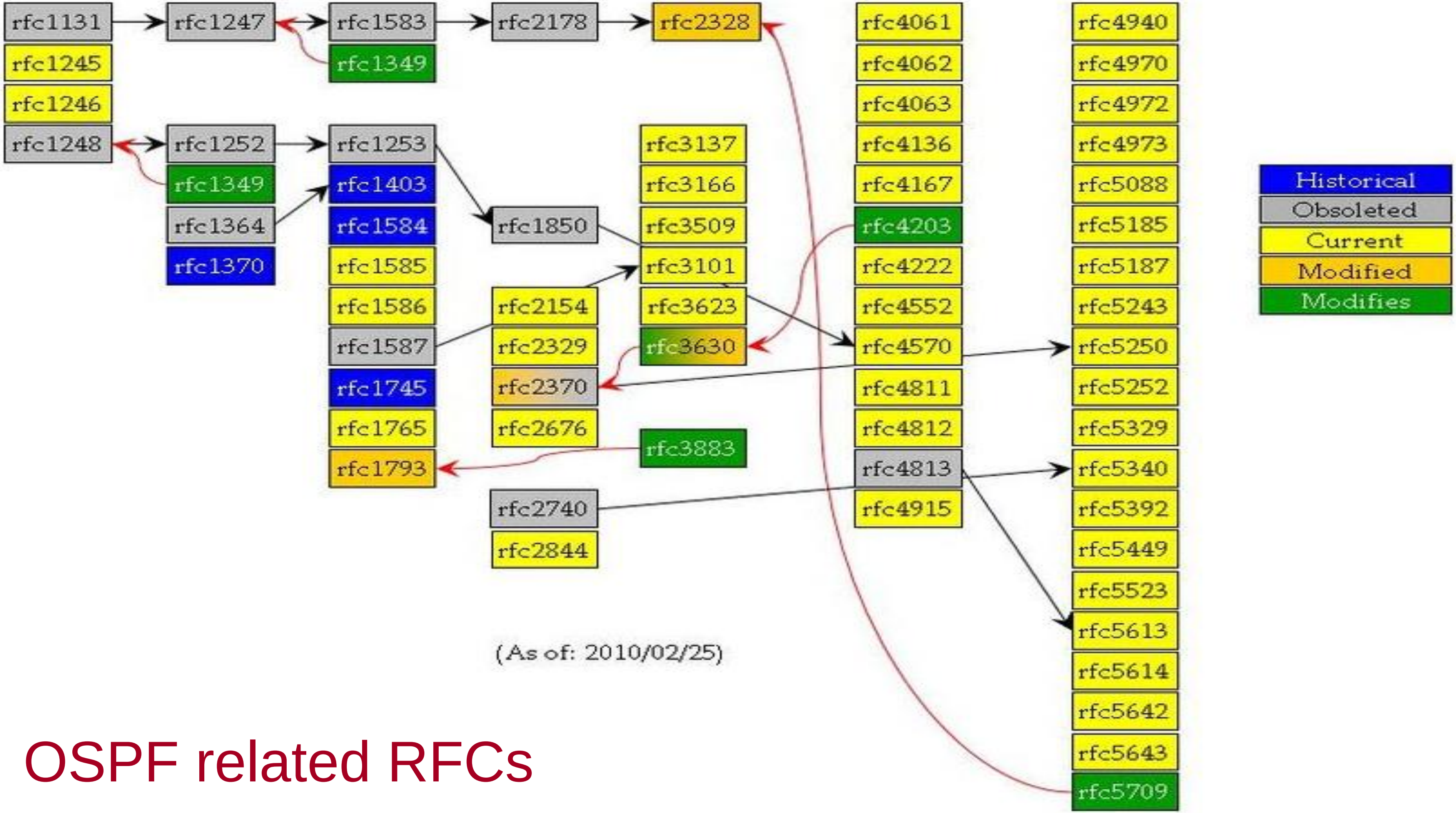
Consequences of a broken link

- Links A-B and D-E are broken
  - The network is partitioned
  - No update messages between the two partitions
- Nodes A and D consider the rest of the network unreachable
- After the link is re-established, the routers synchronize their databases
  - Topology update



Link-state protocols

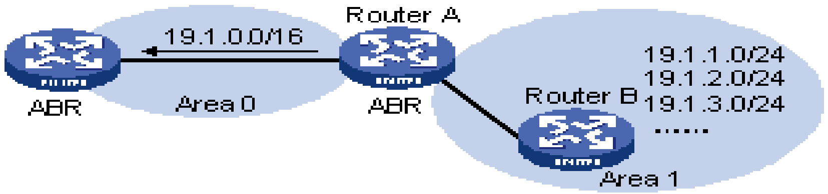
- OSPF – Open Shortest Path First
  - First standard – RFC 1131 ('89)
  - OSPFv2 – RFC 2178 ('97)
  - OSPFv3 – RFC 2740 ('99)
    - IPv6 version

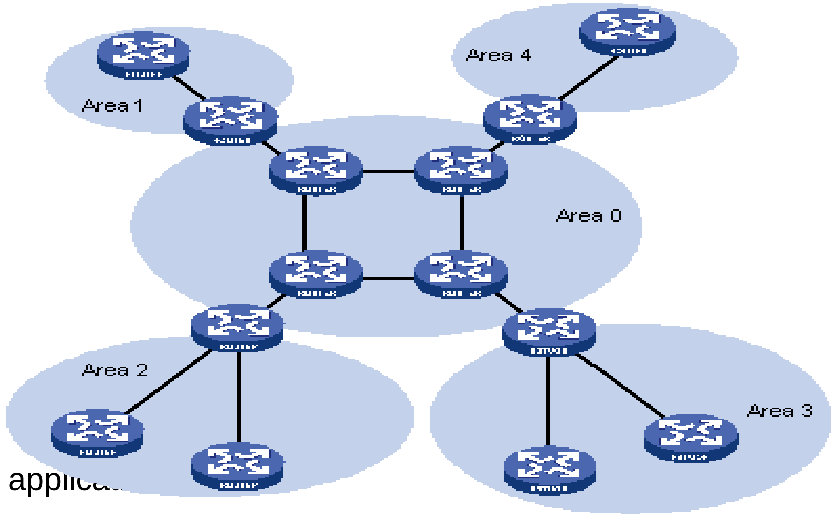


OSPF related RFCs

2 level hierarchy

- An OSPF domain split into areas
  - For scalability reasons
- LSA (Link State Advertisement) advertised inside the areas only
- Aggregation between the areas
  - The changes inside an area not visible from outside
  - Special area – Backbone area (AreaID=0)





OSPF protocol operation

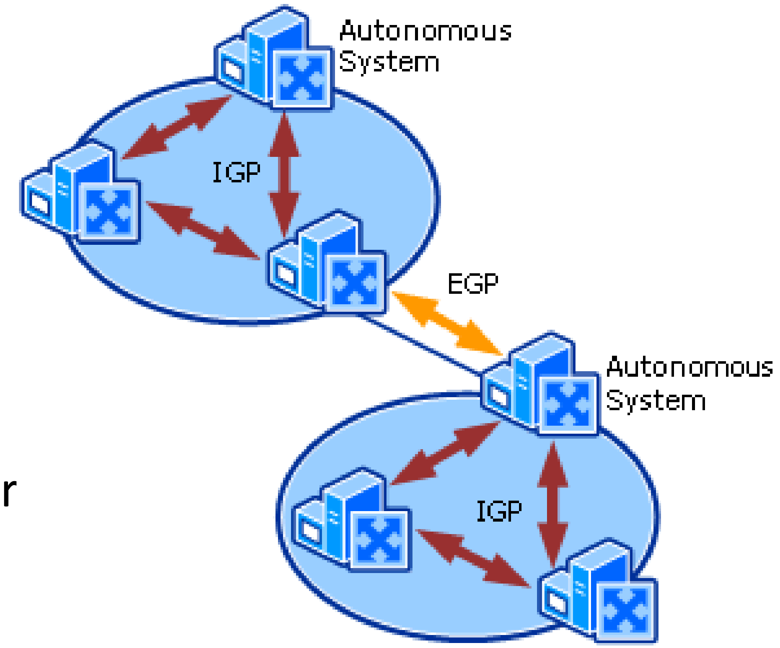
- Neighbor discovery
  - With the Hello protocol
- Chosing the Designated Router (DR) and Backup Designated Router (BDR)
  - Based on priorities
    - From 0 to 254
    - If priority set to 0, it will never be selected as DR or BDR
  - In case of equal priorities, the bigger Router ID wins
    - RID = the biggest configured loopback address on the router (127.x.x.x)
    - If no loopback address configured, RID = the biggest active interface address
  - If a higher priority router appears (is turned on) after the DR selection, it will not take over the DR role, until the DR and the BDR operate correctly
  - If the DR „dies”, the BDR takes over the role
    - New BDR is selected

OSPF protocol operation

- Forming adjacencies
  - Synchronizing the database and advertising the LSAs among the neighbors
  - The DR decreases the network traffic
    - The DR maintains a table about the entire network topology
    - Each router inside an area in a master-slave relation with the DR
    - Routers send updates to the 224.0.0.6 multicast address
      - All OSPF DR and BDR routers
    - The DR send the new table to the 224.0.0.5 multicast address
      - All OSPF routers

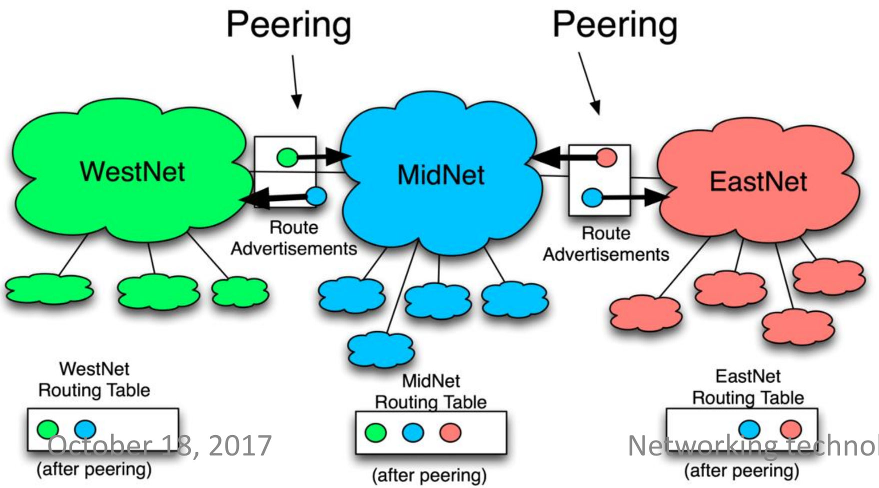
Autonomous sytsem

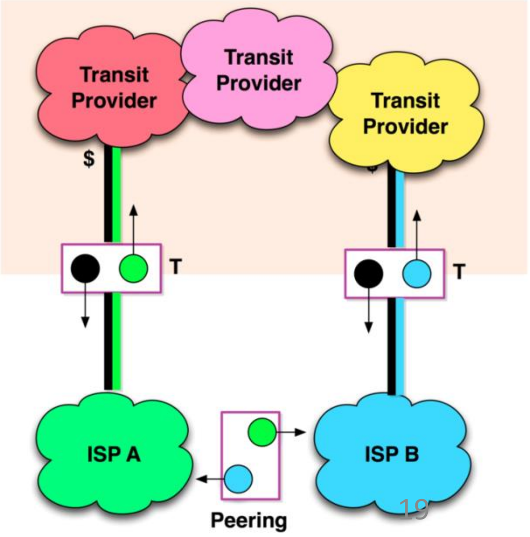
- **AS – autonomous system**
  - Set of routers inside a domain that is technically supervised by one entity
    - One ISP, one administration
  - Some IGP (Interior Gateway Protocol) protocol inside the AS
    - E.g., RIP, OSPF
  - Some Exterior Gateway Protocol (EGP) for inter-AS routing
    - E.g. BGP-4



Internet topology

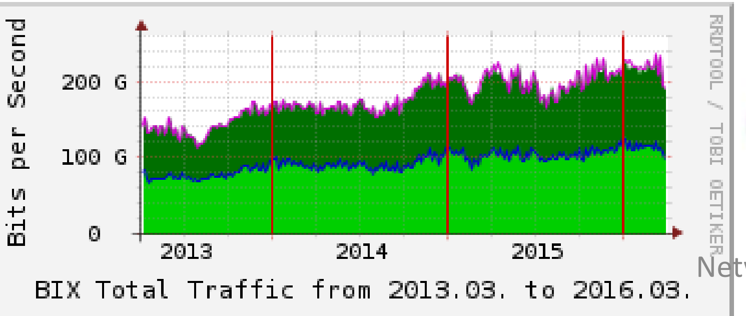
- Network of autonomous systems
  - Customer-provider relation
    - **Transit relation** – connecting to the global network
  - **Peering relation** - two equal rank ASs, between two equal rank providers
    - Not transitive

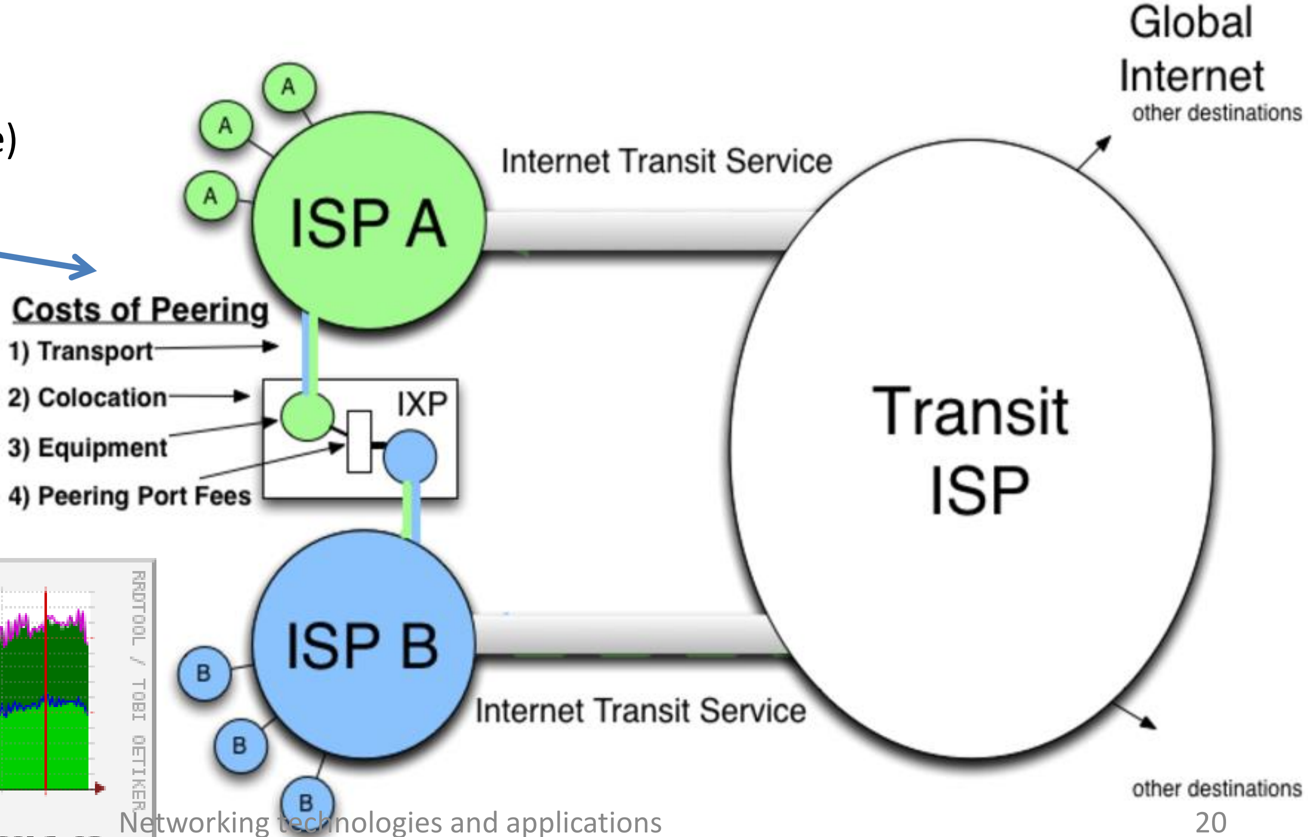




Tranzit vs. Peering

BIX2
(Budapest Internet Exchange)
210 Gbit/s (2014)





Internet topology

- Advantages of the IGP-EGP hierarchy
  - Scalability for large networks
    - Fewer prefixes to be sent
    - Faster convergence
  - Limits error propagation
  - Administrative autonomy
    - Inside each AS an IGP protocol of choice

IGP vs. EGP

- In IGP automatic neighbor discover
- In EGP specially configured peers

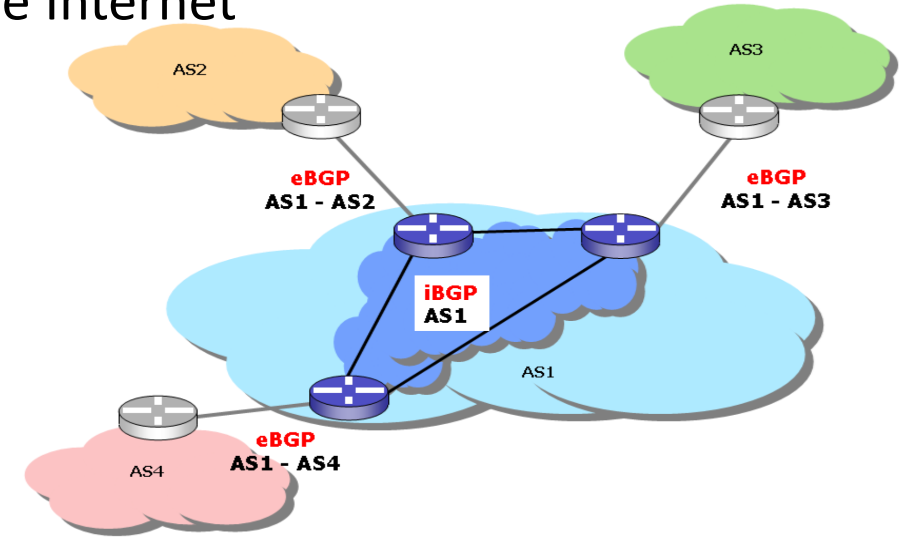
- In IGP you trust the routers
- In EGP you have limited trust in connections with other networks

- In IGP prefixes are distributed inside the entire network
- In EGP prefix distribution is administratively limited

- IGP connects routers of the same AS
- EGP connects the routers of different ASs

Border Gateway Protocol

- One of the main building blocks of the Internet
- BGP chronology
  - Initial standard – BGP – RFC 1105 ('89)
  - BGP-3 – RFC 1267 ('91)
  - BGP-4 – RFC 1771 ('95)
  - Last version – RFC 4271 ('06)



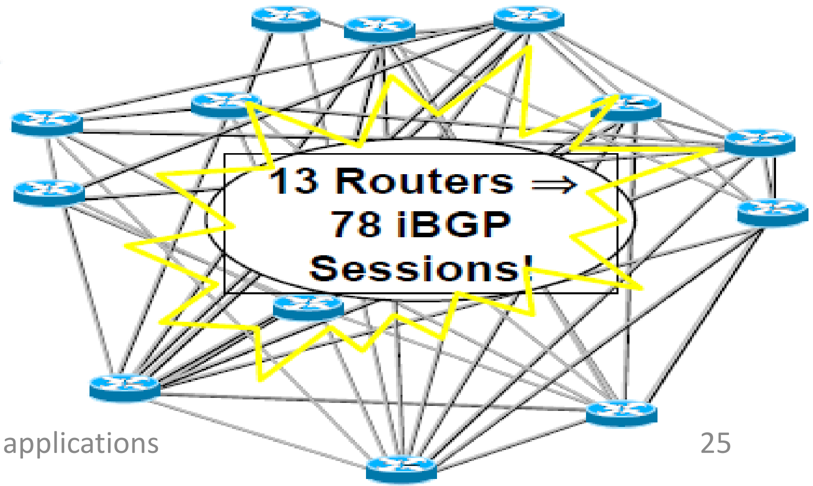
- External BGP (eBGP)
  - BGP connection with a neighbor router from a different AS
- Internal BGP (iBGP)
  - BGP connection with a neighbor router from the same AS

BGP properties

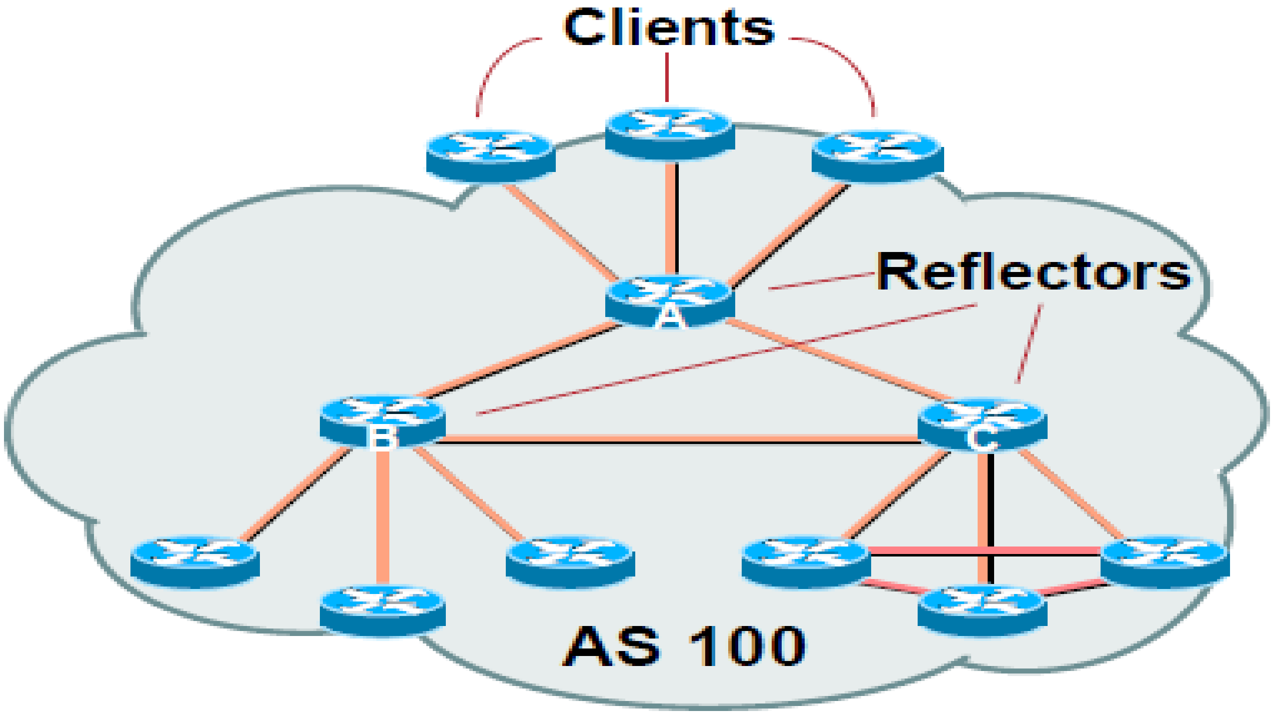
- **CIDR (Classless Inter-Domain Routing)** support
  - Variable length prefixes
  - Efficient address aggregation
- Manual neighbor configuration
  - No automatic discovery
- No periodic updates – hard state
  - Explicit UPDATE messages – NLRI records
    - Network Layer Reachability Information
      - (Destination prefix, AS path, next hop)
    - Loops can be avoide by listing the ASs
  - If a route becomes unavailable, it is also advertised explicitely

iBGP

- Distributes prefixes from eBGP neighbors
- iBGP nodes – full mesh
  - No iBGP routing
- Drawback – a full mesh is not scalable
  - If n=1000,
    n(n-1)/2 = 499.500
    iBGP sessions



Route reflector



Route reflector redundancy