La Sapienza
Università degli Studi di Roma
Dipartimento di Informatica e Sistemistica

Computer Networks II

OSPF - **O**pen **S**hortest **P**ath **F**irst

Luca Becchetti

Luca.Becchetti@dis.uniroma1.it

A.A. 2009/2010

Link State Routing

- Link state protocols suitable for large networks I
- Underlying principles:
  - Routers are responsible for contacting neighbouring routers and acquire their identities (Hello packets)
  - Routers send Link State Packets (LSP) containing list of subnets directly connected to router and costs to traverse each of them
  - LSPs transmitted to all other routers
  - At equilibrium, all routers have same data set and can therefore reconstruct the same **"network map"**, i.e., the same view of the network
  - Network map used to construct shortest path tree to all destination networks

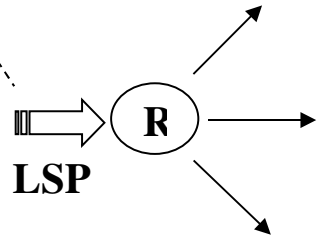
Link State Routing

- In LS algorithms it is necessary to model every link as a point-to-point one
- Sub-optimal procedure in LANs, since this implies mapping each router - end system route with a point-to-point link
- OSPF defines two link types:
  - Router Link: point-to-point connection between 2 routers (e.g., 2 routers connected over a serial link)
  - Network Link: point-to-point collection between a router and a network
  - Each (directed) link and its cost maintained in a Link-State Data Base (**LSDB**)

Link State Routing

- LSP sent when:
  - A router contacts a new neighbour router
  - A link fails
  - The cost of a link changes
  - Periodically every fixed time interval
  - Networks delivers LSPs using **flooding**
  - An LSP is forwarded by a router over all interfaces, *with the exception of the one over which it was received*
  - LSPs have **timestamps** and **sequence numbers** so that:
    - forwarded packets are not forwarded again
    - allow correct ack by receiver

Contains state of links leaving a node



Goal: reduce amount of signalling traffic overhead

Flooding

Since, in general, the same router receives the same information multiple times, probability of complete information increases

- Ensures that all routers in a network:
  - Reconstruct a consistent database of the network state
  - Possess the same information on link states
- When LSP received:
  - Router examines fields of the LSP: **link identifier**, **metric**, **timestamp** or **sequence number**
  - If LSP data not in database, data is stored and LSP is forwarded over all interfaces but the one on which it was received
  - Same happens if LSP data is in database, **but** it is more recent than data about this link contained in database
  - If LSP data is less recent than data stored in database, new LSP with data stored in database is sent back *only* over the link from which the first LSP was received
  - If LSP data is the same as database data, no operation is performed

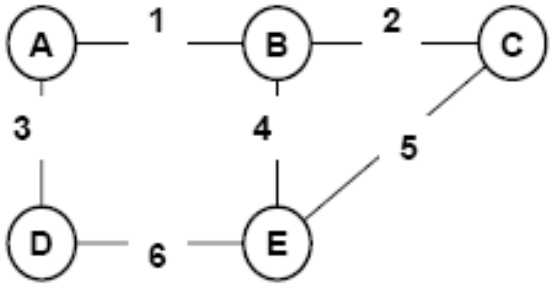
Flooding

- Flooding has following advantages:
  - It explores all possible source-destination routes
  - It is extremely reliable and robust
  - At least *one copy of every LSP* fill follow shortes path to each other router in the network
- Overhead traffic depends on network size and it can be very high

Possible fixes:

- Do not send LSPs back over receiving network interface
- Avoid retransmitting already transmitted packets
- Split an AS into areas in each of which “pure” link-state protocol is adopted and among which “light” version is used, implying exchange of summary, aggregate data

Link State Protocol: example 1/3



Topology database

Da	A	link	Dist
A	B	1	1
A	D	3	1
B	C	2	1
B	E	4	1
D	E	6	1
E	C	5	1

Routing Table

	Destinazione	A	B	C	D	E
A	Distanza	0	1	2	1	2
	Link	local	1	1	3	1

	Destinazione	A	B	C	D	E
B	Distanza	1	0	1	2	1
	Link	1	local	2	1	4

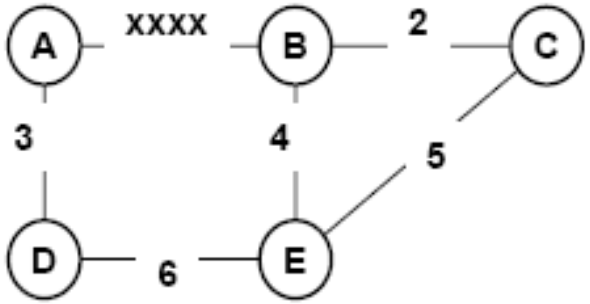
	Destinazione	A	B	C	D	E
C	Distanza	2	1	0	2	1
	Link	2	2	local	5	5

	Destinazione	A	B	C	D	E
D	Distanza	1	2	2	0	1
	Link	3	3	6	local	6

	Destinazione	A	B	C	D	E
E	Distanza	2	1	1	1	0
	Link	4	4	5	6	local

Link State Protocol: example 2/3

Fault at link AB



Routing Table

A	Destinazione	A	B	C	D	E
	Distanza	0	inf	inf	1	inf
	Link	local	1	1	3	1

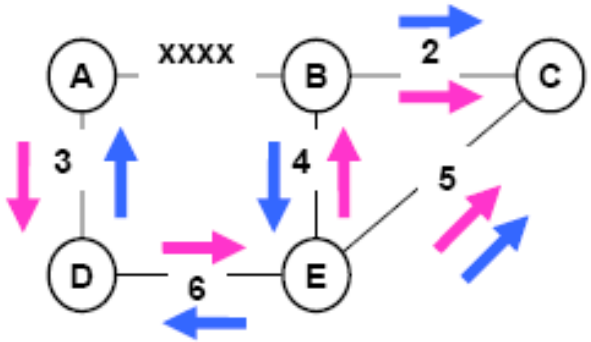
B	Destinazione	A	B	C	D	E
	Distanza	inf	0	1	inf	1
	Link	1	local	2	1	4

C	Destinazione	A	B	C	D	E
	Distanza	2	1	0	2	1
	Link	2	2	local	5	5

D	Destinazione	A	B	C	D	E
	Distanza	1	2	2	0	1
	Link	3	3	6	local	6

E	Destinazione	A	B	C	D	E
	Distanza	2	1	1	1	0
	Link	4	4	5	6	local

Link State Protocol: example 3/3



A		
	Link	1
	Metric	Inf

B		
	Link	1
	Metric	Inf

Topology database

Da	A	link	Dist
A	B	1	Inf
A	D	3	1
B	C	2	1
B	E	4	1
D	E	6	1
E	C	5	1

Routing Table

A	Destinazione	A	B	C	D	E
	Distanza	0	3	3	1	2
	Link	local	3	3	3	3

B	Destinazione	A	B	C	D	E
	Distanza	3	0	1	2	1
	Link	4	local	2	4	4

C	Destinazione	A	B	C	D	E
	Distanza	3	1	0	2	1
	Link	5	2	local	5	5

D	Destinazione	A	B	C	D	E
	Distanza	1	2	2	0	1
	Link	3	6	6	local	6

E	Destinazione	A	B	C	D	E
	Distanza	2	1	1	1	0
	Link	6	4	5	6	local

External routing

- AS connected to other Ases by one ore more **AS Border Routers (ASBR)**
- If one single AS Border Router --> it is enough to update every internal router with default route to exterior of the AS
- If more than one AS Border Router --> every ASBR will send to internal routers
- Cost of route to exterior of the AS
  - External record

OSPF

- OSPF is a Link State Routing Protocol
- First version: October 1989 (RFC 1131), Current: RFC 2328
  - Only protocol whose implementation is mandatory in IPv4 routers (RFC 1812)
  - "Open": available to public
- Consists of three protocols [common header]:
  - Hello
  - Exchange
  - Flooding
- Main ingredients
  - Topology map (LSDB)
  - Topology information (contained in LSPs) flooded via Link State Advertisements (**LSA**s)
  - Dijkstra's algorithm
  - Hierarchical structure

OSPF

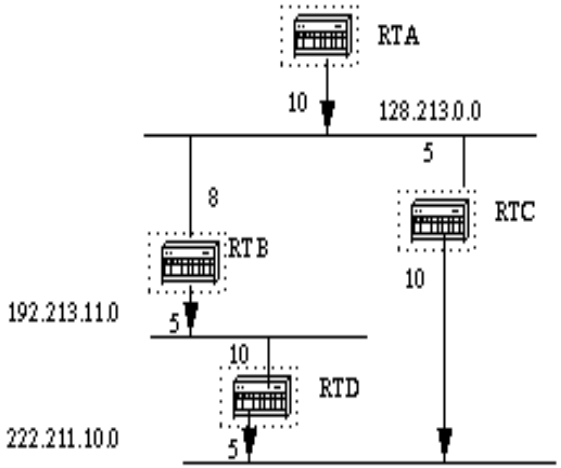
- **OSPF notification message contains one record for every adjacent neighbour**
- Notification sent to whole Autonomous System (using flooding)
- Advanced characteristics:
  - security: all OSPF messages are authenticated
  - More than one shortest path possible (only 1 in RIP)
    - Load balancing
  - For every link, different metrics for different values of TOS field
  - Example: satellite link cost “low” for best effort traffic but “high” for real time traffic
  - Support for multicast Supporto per unicast e multicast:
    - Multicast OSPF (MOSPF) uses same database (routing tables) as OSPF
  - Hierarchical OSPF for large ASes

OSPF

- **Analysis of OSPF:**
  - Basic topology and LSDB
  - Shortest path tree
  - Scalability:
    - Hierarchical routing - Areas
    - Routing tables and **route summarization** [used to reduce RT sizes]
  - State info collection
    - Link State Advertisements
    - Neighbours

Basic topology and LSDB

- We assume simple case of 1 area
- For areas, see further

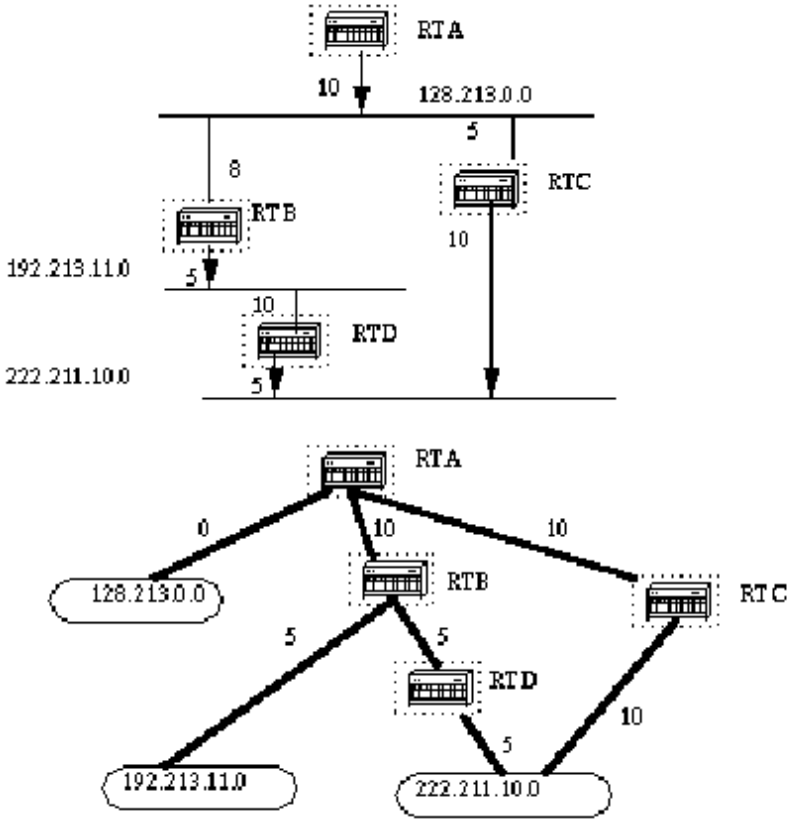


FROM/TO	RTA	RTB	RTC	RTD	128.213.0.0	192.213.11.0	222.211.10.0
RTA		10	10		10		
RTB	8		8	5	8	5	
RTC	5	5		10	5		10
RTD		10	5			10	5
128.213.0.0	0	0	0				
192.213.11.0		0		0			
222.211.10.0			0	0			

- **LINK STATE DATA BASE (Topology database)**
- **At every router**

OSPF

- ***Shortest Path Tree –SPT*:**
  - *For every router*
  - Example:
    - SPT computed at RTA
    - Alternative paths possible (same length)
    - Directed links
    - Costs may depend on direction
      - E.g.: different congestion levels in different directions
      - In the picture: cost(RTA, RTB) ≠ cost(RTB, RTA)



Splitting large networks into areas

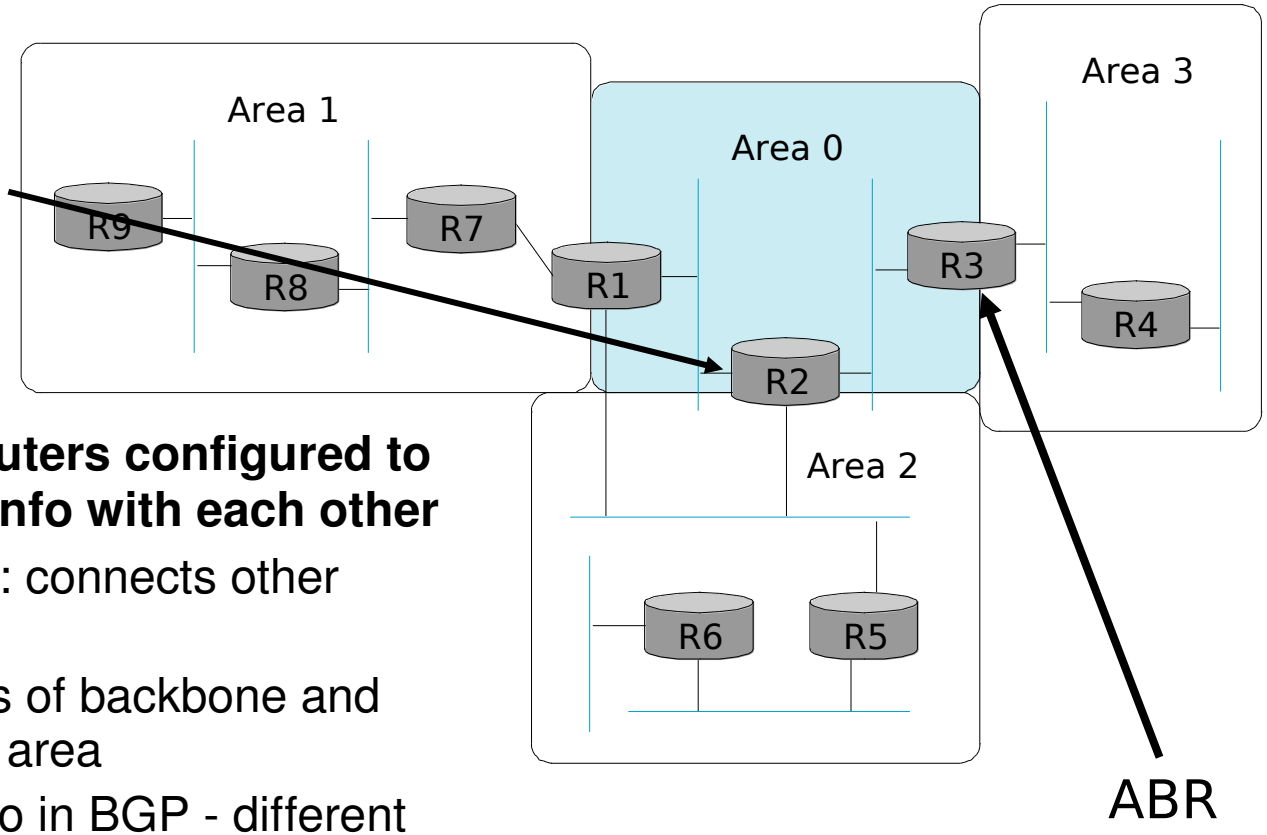
- If network has large size:
  - Number of database records grow and so does memory necessary at every router
  - Shortest path computational time increases
  - Signallign traffic overhead for LSP flooding increases
- OSPF solves problem using hierarchical routing:
  - Network organized in areas
    - Areas correspond to independent portions of the network
    - Disjoint databases
    - Independent flooding mechanisms
  - Areas connected by backbone

Splitting large networks into areas

- Some routers (**Area Border Router - ABR**) belong to different areas
  - Every area contains at least one ABR
  - Every area connected to backbone
- One ABR:
  - Contains databases of areas it belong to
  - Sends suitable messages (summary records) containing list of subnets reachable in the areas it belongs to

Areas

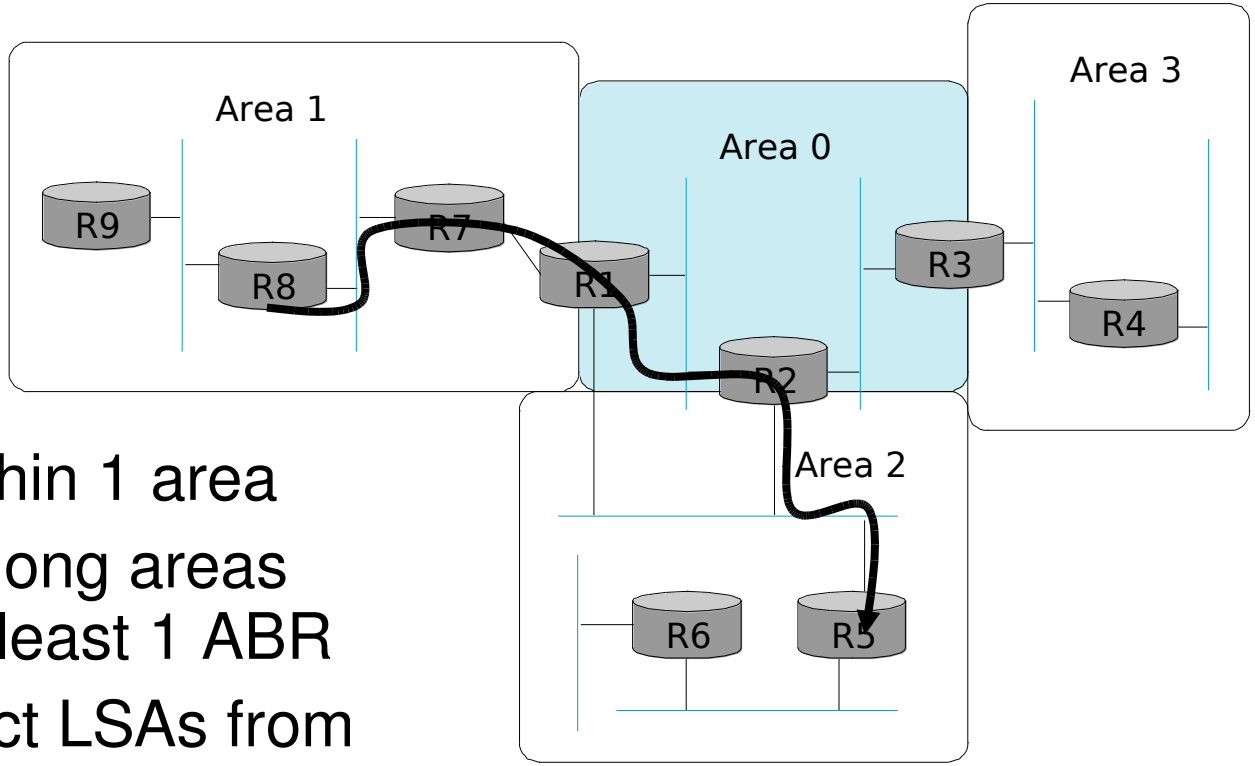
Note: to go to area 2
traverse R2
Why important?



- Area: **set of routers configured to exchange LS info with each other**
- Backbone area: connects other areas
- ABRs members of backbone and non- backbone area
- Note: areas also in BGP - different meaning
- Internal Routers have all interfaces within same area
- Allows for scalability
- Every area: unique 32-bit ID

Routing table at R7
Assume unitary link costs

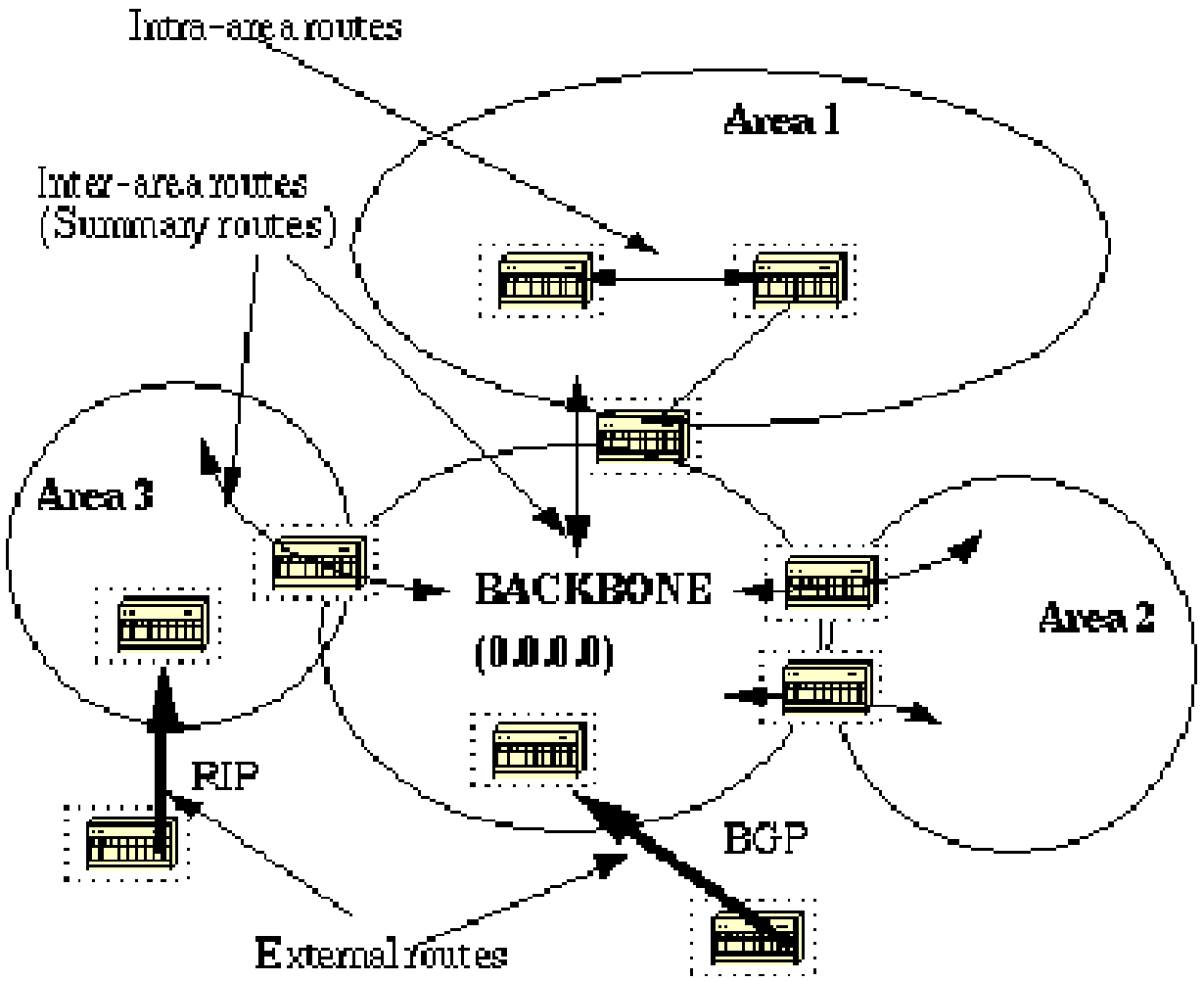
Routing



- Optimal within 1 area
- Routing among areas involves at least 1 ABR
- ABRs collect LSAs from areas to which they are connected and flood them in backbone and so in other areas (route summarization)

Area 0 - backbone

- One area must be **Area 0**
- **Area 0** interconnects all others
- Backbone redistributes routing info among different areas
- **Summary route**: route between different areas
- **External route**: routes to networks using different routing protocols



OSPF router hierarchy

- IR (Internal Router)
  - Router internal to an area: they exchange link state info with routers in the same area by flooding
- Backbone Router
- ABR (Area Border Router)
  - Interface routers toward area 0
  - Every ABR is also a backbone router
  - Send link state info about area to which they are connected to other ABRs in the backbone
  - They relay link state info about other areas into the area they are responsible for
- ASBR (Autonomous System Boundary Router)
  - Border routers towards other ASes
  - Import/export routing information to/from other Autonomous Systems
  - Any router can an ASBR

Routing tables

- In principle, every router might know optimal route to any other destination in the network
- In practice, routes are optimal only within each area
- Route summarization
  - Only aggregate info about subnets belonging to different areas
  - Info collected, aggregated and relayed by ABRs

Route summarization

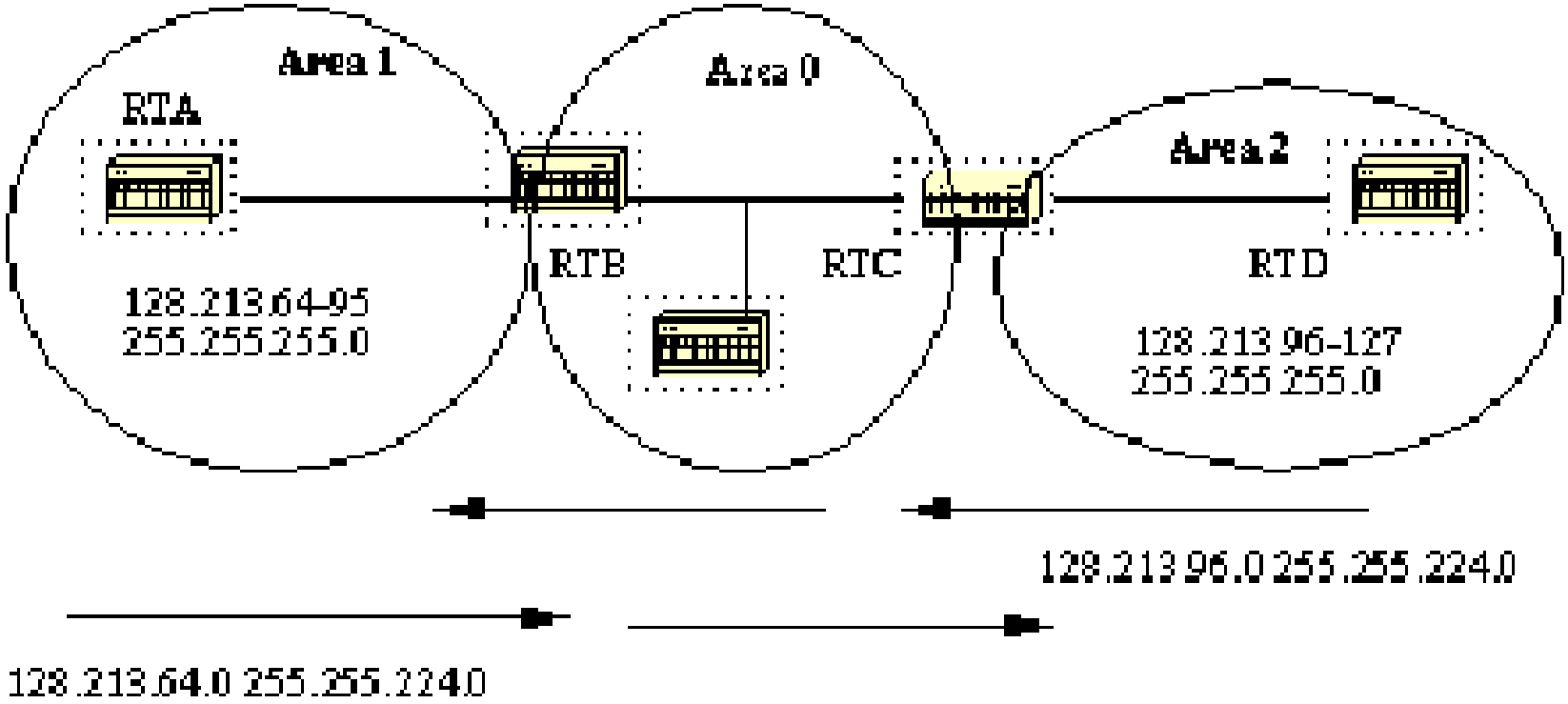
- Aggregation of information about multiple routes in a single LSA (Link State Advertisement)
- Done by ABRs
- Normally: from areas towards backbone
- **Inter-area routes**: routes between areas of the same As
- **External routes**: typically routes to other ASes, communicated to an ASBR bordering a different AS

Route summarization/cont.

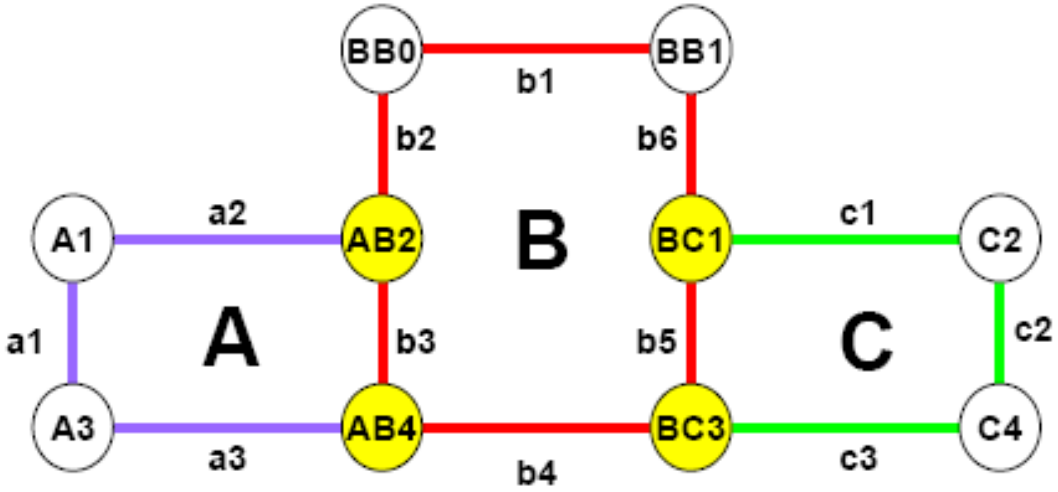
- Uses network masks
- Example: represent block 128.213.64.0 - 128.213.95.255 compactly
  - 2^{13} = 8192 IP addresses
- Binary network prefix common to all addresses in block: 10000000 11010101 010 (19 bits) -> Net mask = 255.255.224.0 (first 19 bits)
- Compact notation: 128.213.64/19
- Works is we refer adjacent blocks
- Follows CIDR rules

Route summarization/cont.

- 3 more significant bits in 3rd byte masked -> 32 class C networks in every area
- How many addresses in Area 1?
- Eg.: RTB sends summary info about networks 128.213.64-95
- Mind overlaps!!



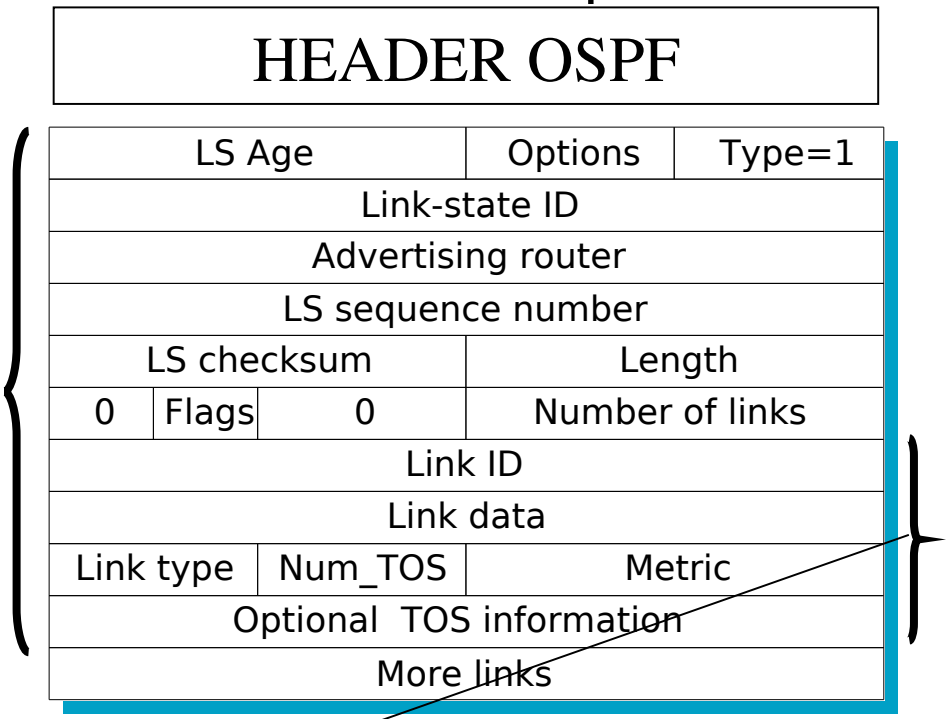
Splitting large networks into areas - recap



- The database of a router in area A contains:
  - Records for links a1, a2, a3, communicated by routers A1, A3, AB2, AB4
  - Summary records for subnets belonging to backbone and area C, communicated by routers AB2 and AB4
    - Associated to every subnet: cost to reach the subnet
    - Analogy with distance vector protocols
  - External records sent by routers BB0 and BB1 and forwarded by routers AB2 and AB4
    - Every destination has an associated cost to reach it

LSA - structure

- Es.: LSA di tipo 1 - costo di link tra router
- LS ID = advertising router
- Sequence number: identifica LSA vecchi o duplicati



- Link ID/Link Data
  - identificano link
- metric: costo
- link type: ex. ETH
- num_TOS: descrive il significato del costo

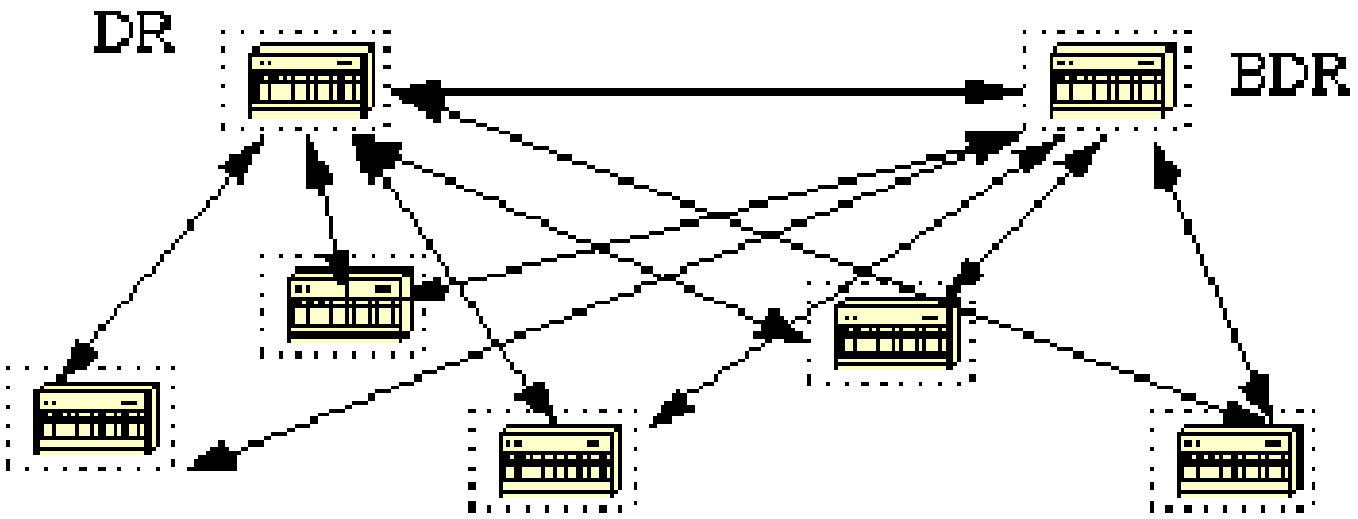
Corrisponde a un LSP

Link

LSA

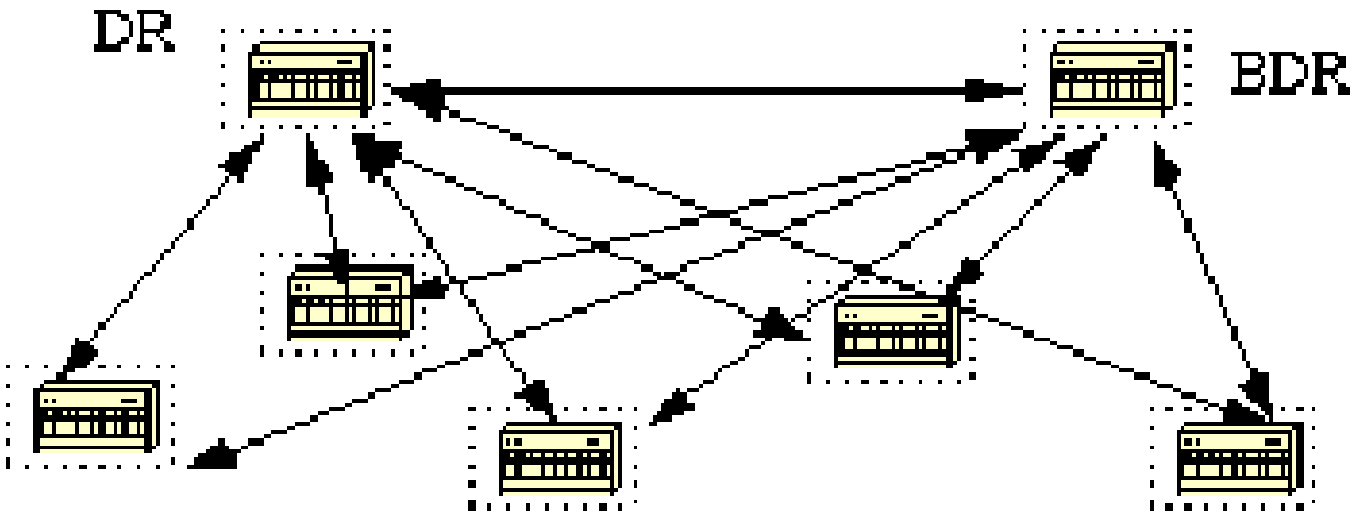
Neighbours

- Router on same LAN segment
  - On multiple acces networks: all pairs
  - On point-topoint networks: physically adjacent routers



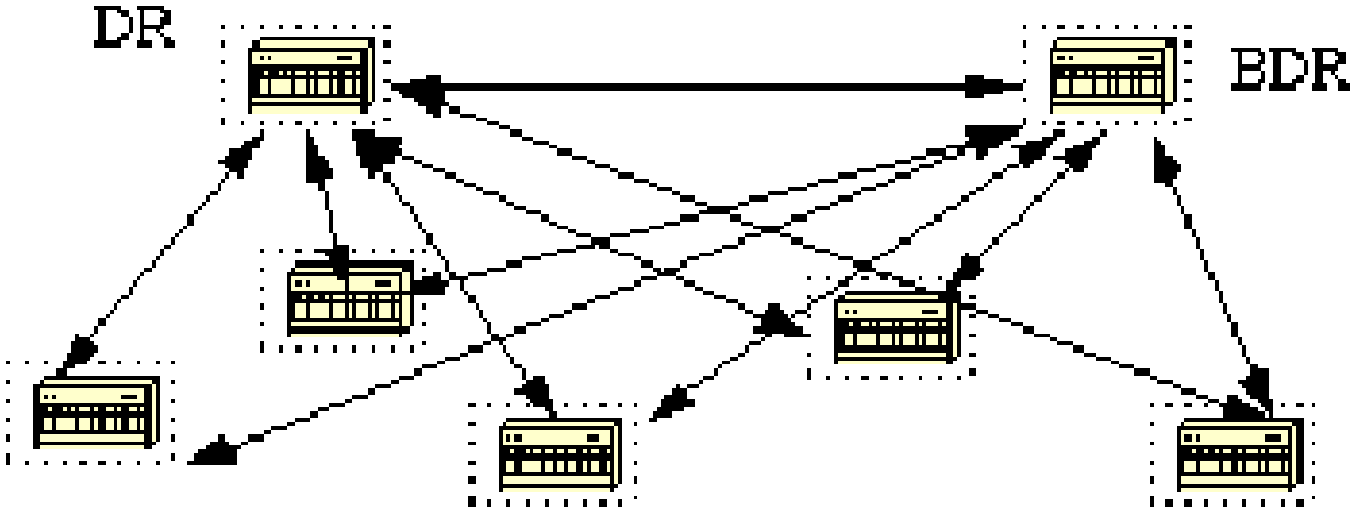
Adjacency

- Nodes become **adjacent** using **Hello protocol**
- On broadcast networks (e.g. Ethernet): two neighbour nodes become adjacent if one of the two is a DR (Designated Router) or a BDR (Back-up Designated Router)



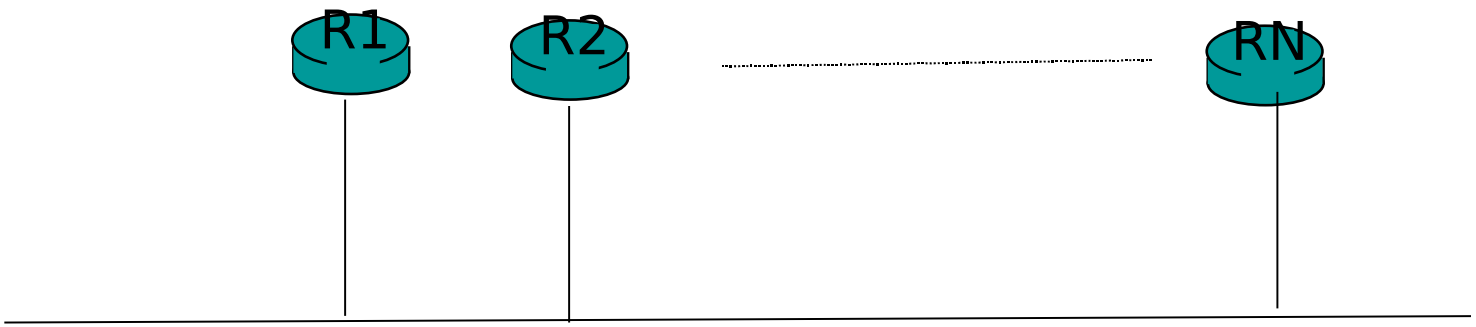
Adjacency/cont.

- DRs reduce signalling LAN
- I vicini si scambiano l'informazione di base che serve a ricostruire l'informazione di instradamento all'interno di un'area



Adjacency/cont.

- LAN segment internal to an OSPF area
- N routers
- How many messages for every router to have updated info on LAN segment without a DR?
- How many when DR?



OSPF

- Open Shortest Path First (OSPF) follos following principles
  - Separation between hosts and routers
  - Support for Broadcast networks (e.g. LANs)
  - Support for non broadcast networks
  - Splitting of large networks into areas

Separation between hosts and routers

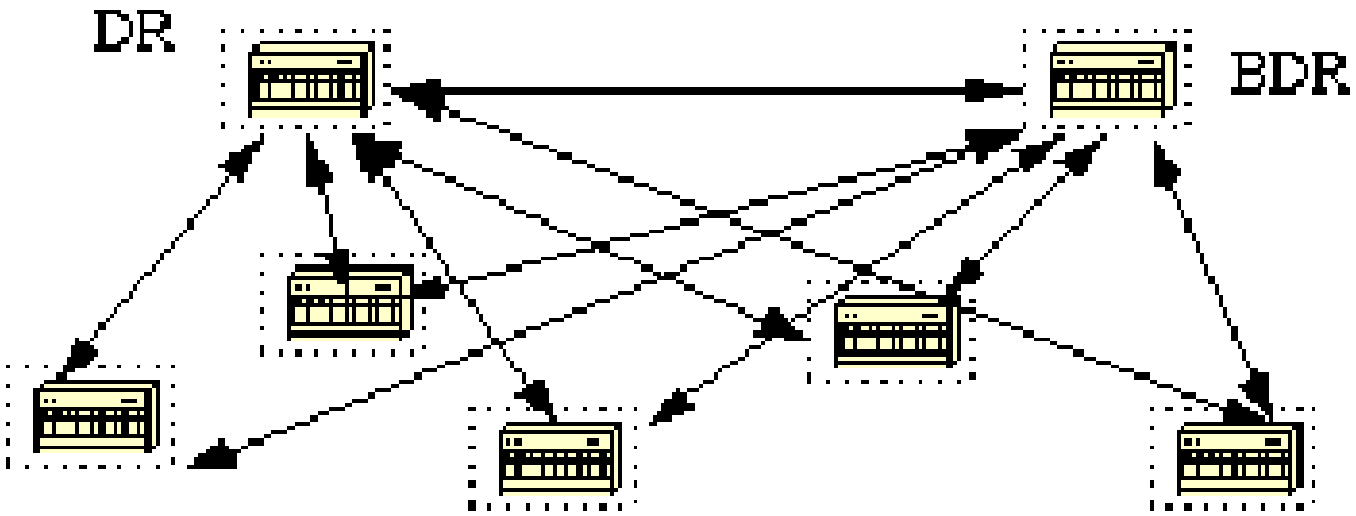
- Router typically connected to many hosts over a local network (LAN)
- OSPF uses concept of **Subnet** as defined by IP addressing scheme
  - Whole subnet represented as a single link in database --> link identified by subnet address
  - Link to a stub network
- This greatly limits number of records in network database

Support for broadcast networks

- Local network (LAN) characterized by:
  - Full connectivity
  - Diffusive transmission
- If LAN contains N routers:
  - Every router must manage N-1 adjacency relationships with other routers and a “stub-network link” with network hosts
  - N^2 records overall in network database
- Problems fixed by assigning a **Designated Router** (**DR**) in every LAN
- In a LAN: every router establishes adjacency relationships only with designated router
  - Number of records reduces to N
  - Link state info flooding only managed by DRs

Adjacency

- On broadcast networks: LSAs exchanged through a Designated Router
- Backup DR present for fault tolerance
- Signalling traffic overhead reduced



Support for non-broadcast networks

- A packet network in geographycal area offers :
  - Full connectivity among network nodes
  - Non-diffusive transmission
- If a geographical network contains N routers:
  - Need to define a static configuration of Virtual Circuits
    - complexity N^2 for a full mesh
  - Applying the concept of Designated Router (DR)
    - Difference with broadcast networks: DR must send a copy of every LSP over every virtual link towards all other routers in the network

OSPF

- OSPF is widespread in medium- and large-sized networks:
  - Based on the exchange of LSPs said **Link State Advertisements** (**LSA**) is OSPF terminology
  - Supports different metrics depending on value of TOS field
  - Supports use of variable length subnet mask
  - Supports authentication service between routers
  - Supports indication of **specific routes**
  - Reduction in routing table sizes through the concept of Designated Router
  - Support of the concept of **virtual link** for interconnection of non-contiguous areas

Supporto password

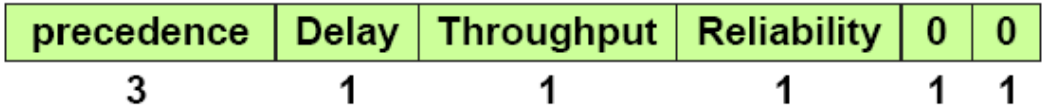
Idea: diverisfy topology (link cost) according to type of service. Example: same link may have low cost for reliable service and high cost for low-delay service

OSPF

- Metrics:
  - Best path can be, according to specific needs:
    - Shortest path
    - Largest capacity path
    - Lowest-delay path
    - …
  - OSPF support different metrics on the same link at the same time
- Different metrics:
  - Important to use same metric for routing the same packet in order to avoid loops
  - Important to provide indication of specific metric to use
    - **Choice of metrics determined by value of TOS field in IP packet header**

OSPF

- TOS metrics:
  - OSPF allows to define metrics according to TOS field of IP header :
    - 64 possible Types Of Service in theory
    - In pratice, **precedence** bits ignored
    - Modifications are currently under consideration
  - TOS field currently “under review” also by other network standardization groups [e.g. DiffServ]



OSPF terminology

- **Area**
  - Logical set of networks and routers (geographically and administratively,...)
  - Goal: limit size of topology database within routers
  - Within an area: routers must have *identical* network topology databases
  - Information on portion of network external to an area are contained in Area Border Routers
  - An Area Border Router relays LSAs containing information on networks external to the are under consideration
  - All OSPF networks have at least one area, the **backbone**

Aggregated info, in the sense of “already processed”

OSPF terminologyh/cont.

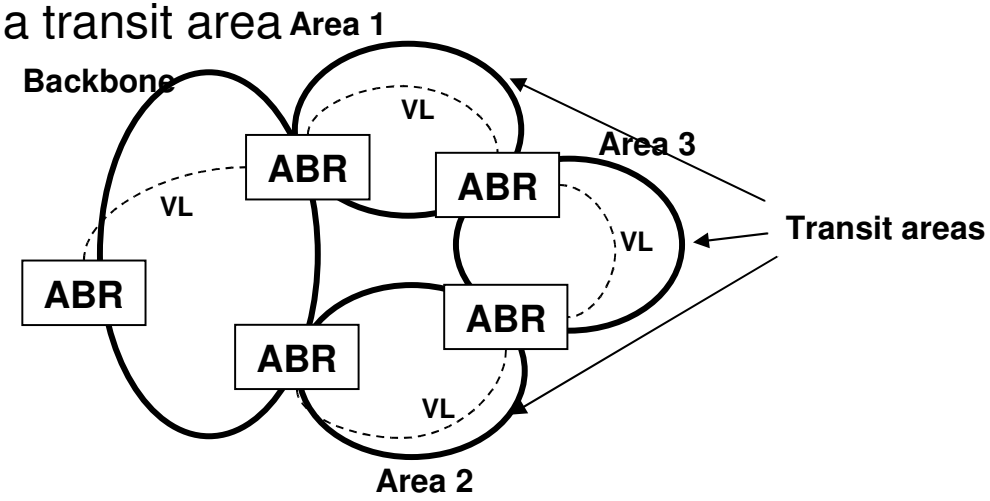
- **Intra-Area Router (IAR)**
  - Routers internal to an OSPF area
  - They exchange LSAs with all other routers inside the area
  - Each manages a copy of area's topology database
- **Area Border Router (ABR)**

  Operates in 2 modes: pure ABR + IAR for each area to which ABR is connected
  - Connected to 2 or more areas
  - Manage topology databases for *all* areas to which they are connected
  - Relay insided every area LSAs for networks contained in every other area
- **AS Boundary Router (ASBR)**

  Needs to implement 2 protocols: IGP for every area to which ASBR is connected + EGP for each AS to which it is connected
  - Routers on the border of an OSPF domain
  - They exchange LSAs containing info about reachability of routers an networks belonging to different ASes
  - They relay LSAs inside own domain, containing info about external routes

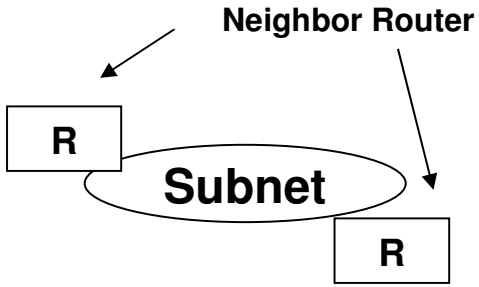
OSPF terminology/cont.

- **Virtual Link (VL):**
  - 2 possibilities:
    - Part of backbone
    - Endpoints are 2 ABRs sharing a *non-backbone* area
- **Transit Area**
  - Area physically spanned by a VL
  - Of course, backbone is a transit area

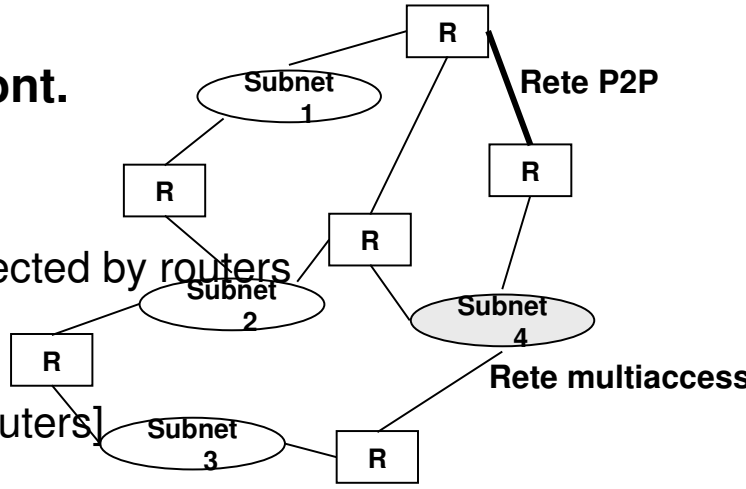


OSPF terminology/cont.

- **Neighbor Router (NR):**
  - "Physical" adjacency: routers interface on the *same* subnet
- **Adjacent Router (AR)**
  - "Logical adjacency": NRs that exchange LSAs to synchronize respective topology databases for the *common* area they belong to
  - **Not every pair of NRs becomes a pair of ARs**

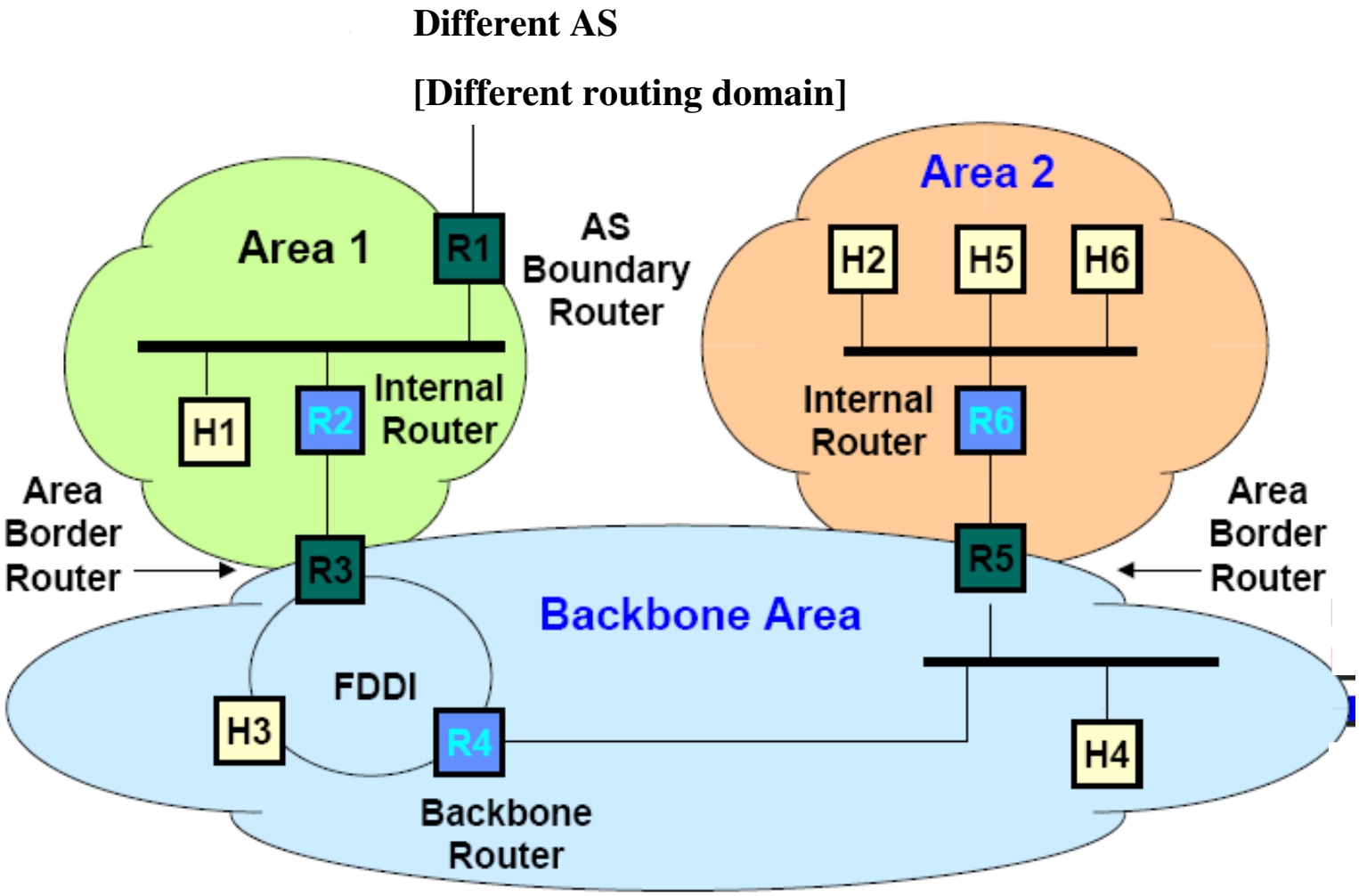


OSPF terminology/cont.

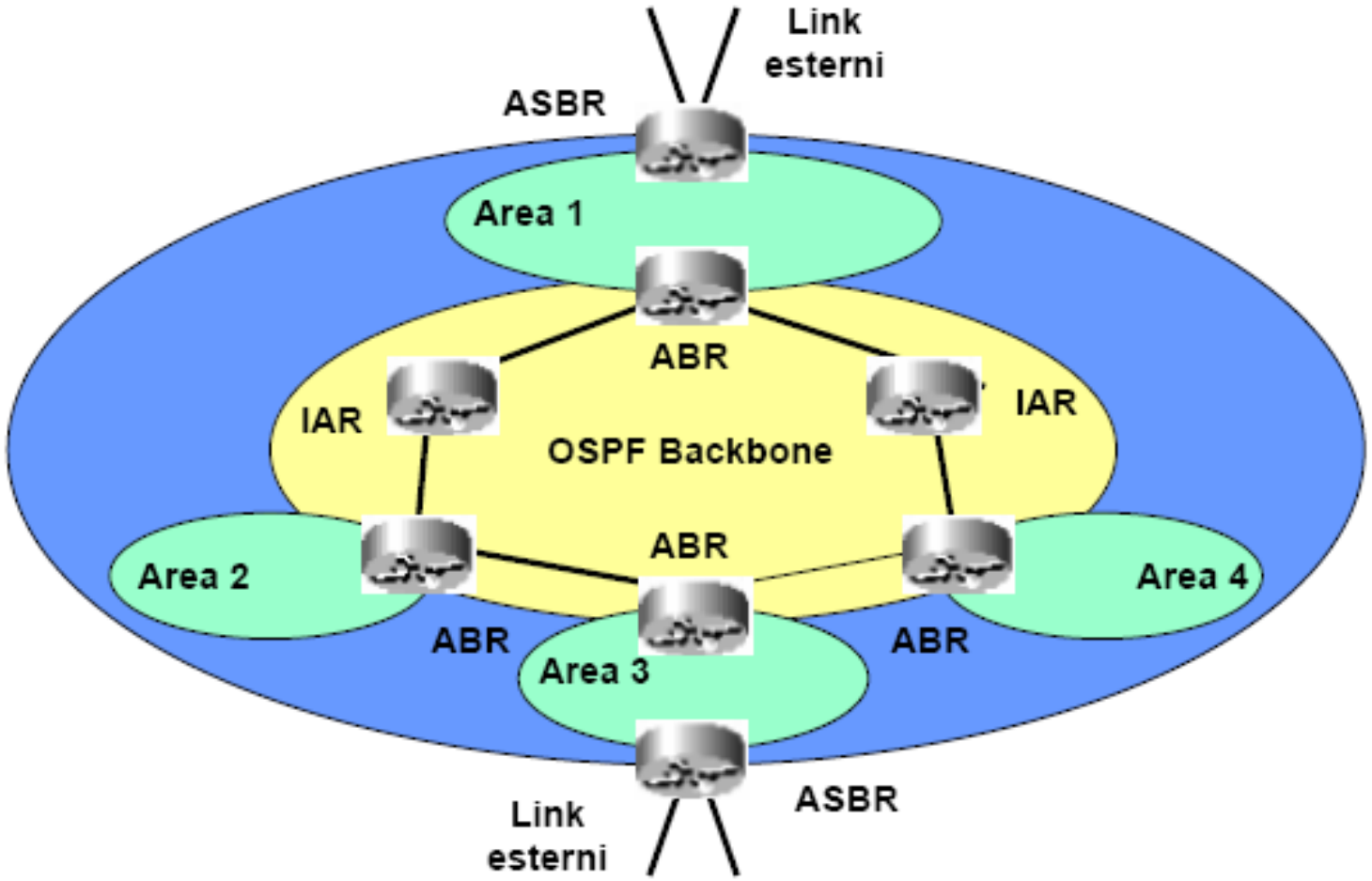


- **Physical networks:**
  - every area consists of physical networks interconnected by routers
  - 2 kinds of networks:
    - Point-to-point networks [P2P link between routers]
    - Multi-acces networks
      - ✓ Each link allows interconnection of more than 2 routers
      - ✓ May be broadcast [e.g: LAN] or non-broadcast [mesh networks]
- **Designated Router (DR):** a multi-access network has a Designated Router that:
  - Is adjacent to *all* routers connected to the multi-access network [in the sense of exchanging LSAs]
  - Sends LSAs to external routers, each LSA containing a **list of** routers connected to the multi-access network [**Network link advertisements**] to which the DR belongs

OSPF terminology/cont.



OSPF terminology



OSPF terminology

- **Type Of Service (TOS) metrics**
  - Each LSA must contain at least one metric corresponding to TOS field value "0000"
  - One LSA may contain more metrics, each relevant to a different value of the TOS field
- **Link State database**
  - Directed graph describing network link state, as seen from a router
  - Created and updated depending on LSAs received by router
- **Area ID**
  - 32 bit identifier for an area (backbone area ID = "0")
- **Router ID**
  - 32 bit identifier for a router within an AS

OSPF Terminology

- **Link State Advertisements (LSA)**
  - Packets exchanged among OSPF routers to update link state databases and inter-area and inter-AS routes
  - **Router link advertisement**
    - Contain info about state of a router's links, sent over a single area using flooding

    Topological update info
  - **Network link advertisement**
    - Generated by DRs: they contain list of routers connected to a multi-access network and are sent over area containing the DR using flooding

    Topological update info
  - **Summary link advertisement**
    - Generated by ABRs: they announce networks belonging to different areas and the costs to reach them. Flooded into every area managed by the ABR

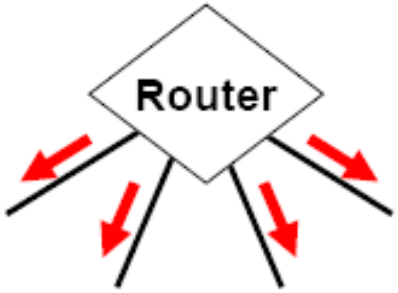
    Pre-processed info: already routing info
  - **AS external link advertisement**
    - Generated by ASBRs: announce routes to networks external to OSPF domain. e indicano i cammini verso le reti esterne al dominio OSPF. Flooded into every area managed by the ABR of an OSPF domain

    Pre-processed info: already routing info

4 LSA types

OSPF Terminology

Router Link Advertisment



State of links

leaving router

er

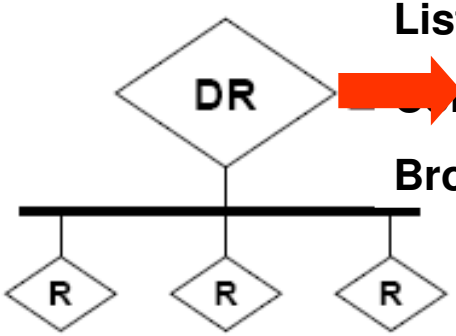
Network Link Advertisment

List of routers

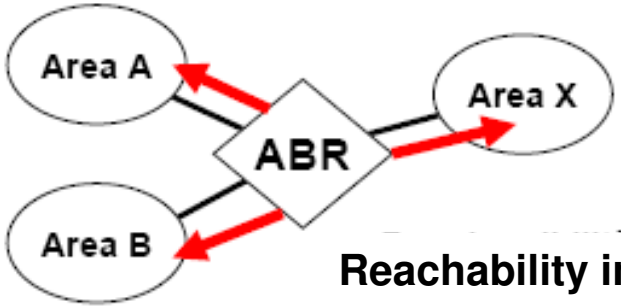
DR

nnected to a

Broadcast network



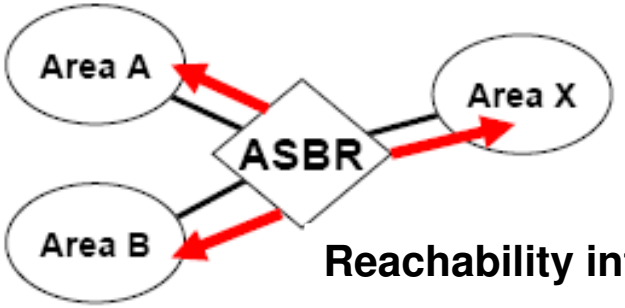
Summary Link Advertisment



Reachability info for

Networks in other areas

External Link Advertisment

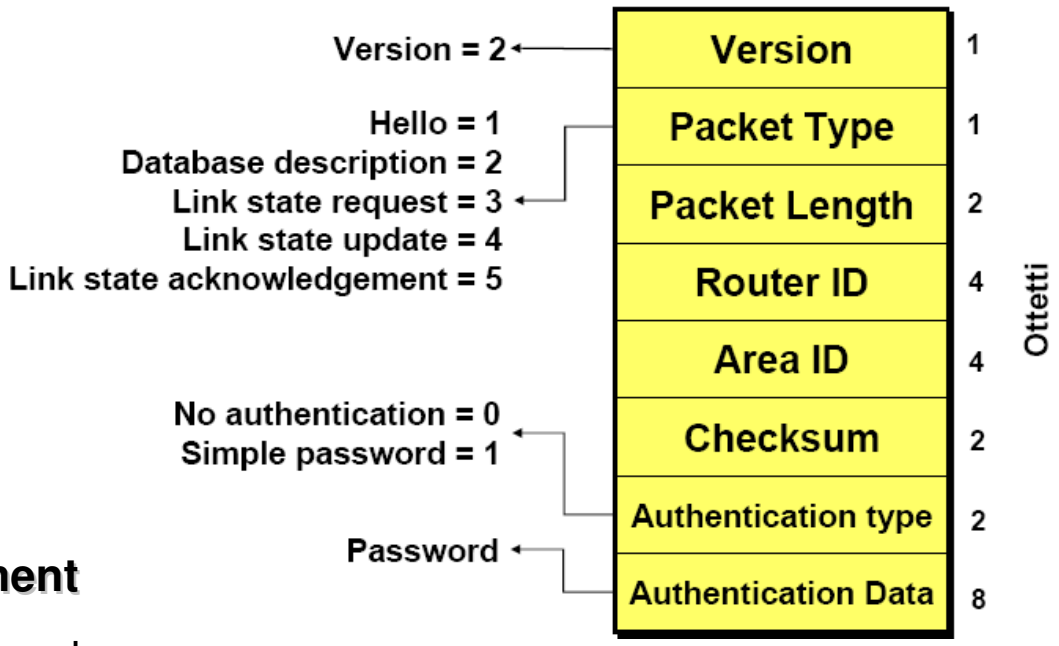


Reachability info for

Networks in other ASes

OSPF Protocol

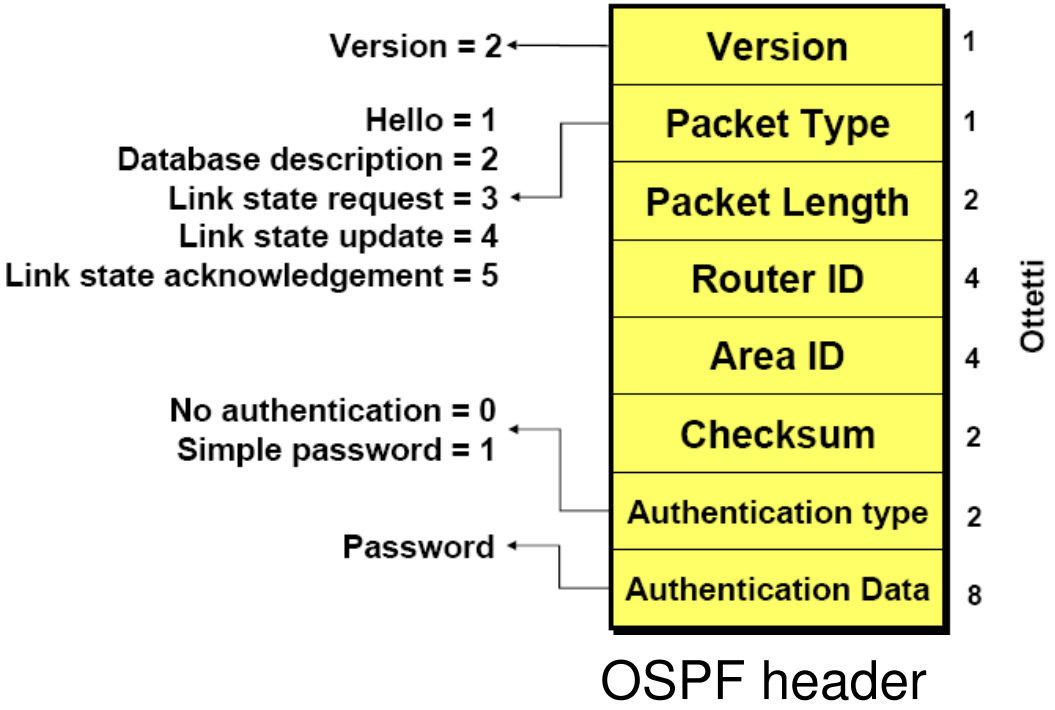
- OSPF packets encapsulated within IP datagrams
  - protocol identifier = 89
- 5 packet types:
  1. **Hello**
  2. **Database description**
  3. **Link state request**
  4. **Link state update**
  5. **Link state acknowledgement**
- All OSPF packets have same header structure
  - Header specifies packet type



OSPF header

OSPF Protocol

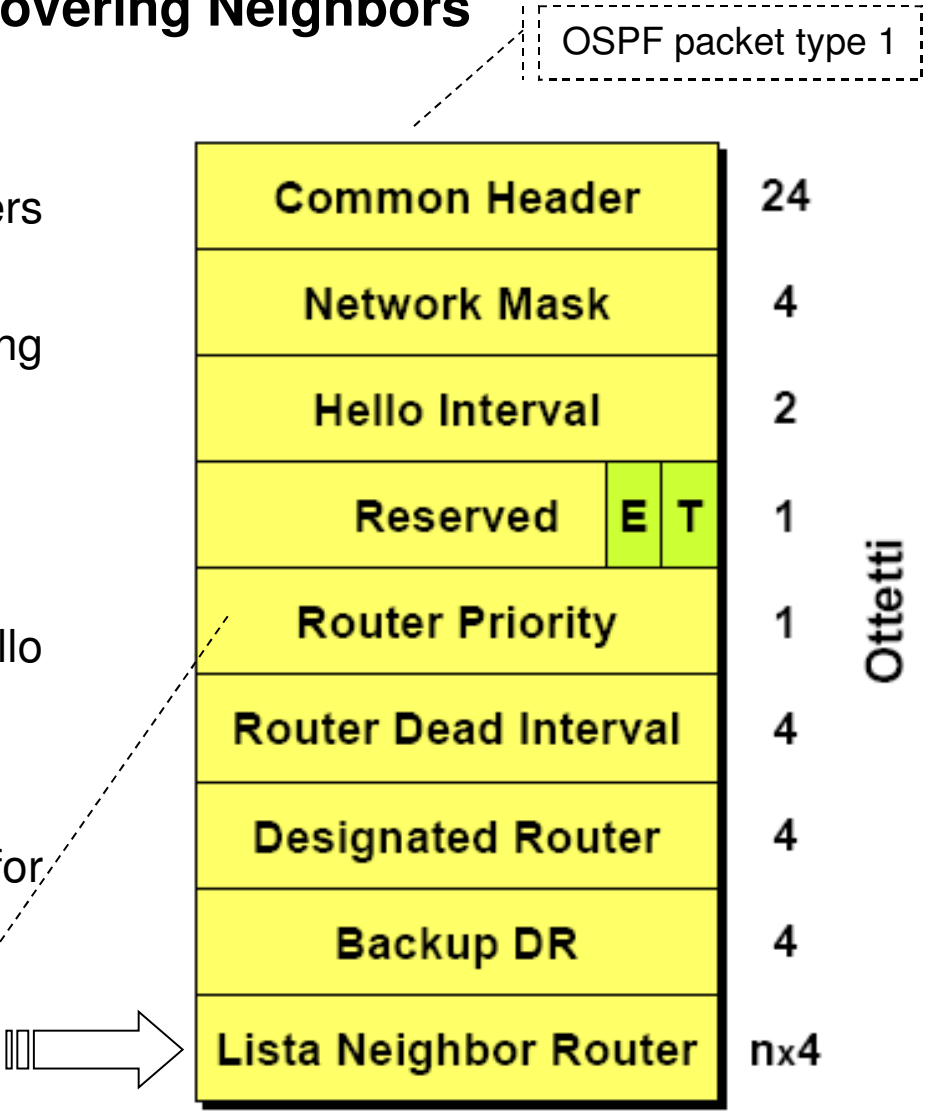
- OSPF header essentially contains:
  - ID of packet's sender router
  - Type of packet
  - Authentication info



OSPF header

Phase 1: Discovering Neighbors

- Initiated by a router to:
  - Discovering Neighbor Routers (NR) in a subnet
  - Establidhing and maintaining relationships with NRs
- Uses Hello packets
- NR list contains:
  - IDs of routers from which Hello packets were received
  - IDs of DR and backup DR
- Bit T indicates support of metrics for TOS



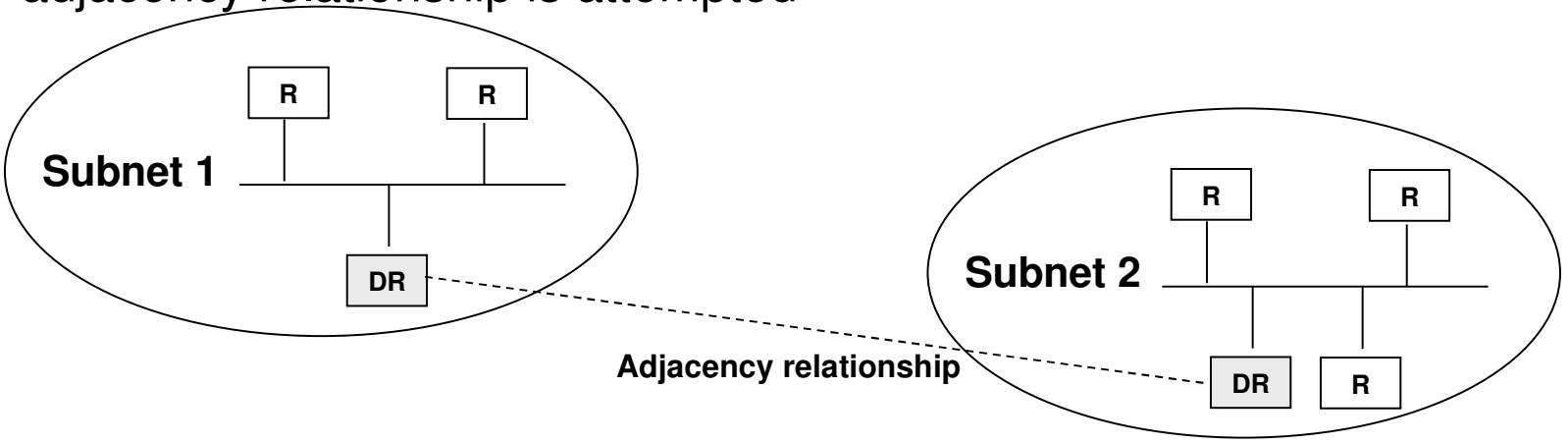
Phase 2: Electing Designated Router

- DR of a multi-access network has to:
  - Send LSAs to routers external to network segment
  - Establish and maintain adjacency relationships with other routers connected to multi-access network
- DR role assigned according to info contained in Hello packets
- Highest priority router among those exchanging Hello packets becomes DR

But belonging to same area; DR operates on behalf of all routers belonging to multi-access network segment (including itself)

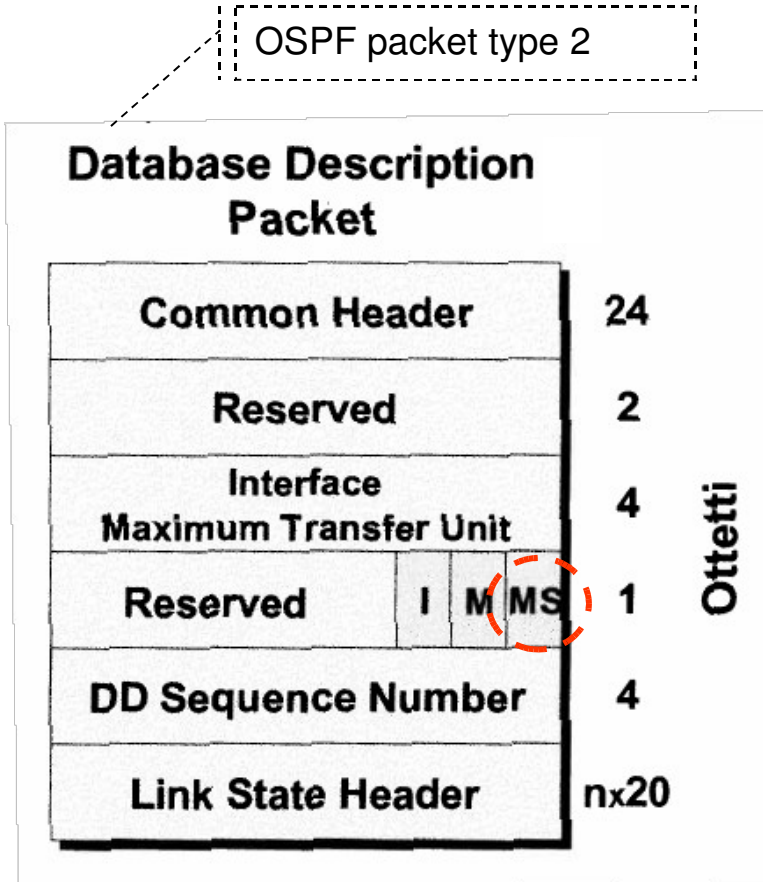
Phase 3: Establishing Adjacencies – Database Exchange

- A router establishes adjacency relationships with a subset of its neighbour routers
- Routers connected over P2P links or VLs are always adjacent
- Routers on multi-access networks are only adjacent to local DR and BDR
- Adjacent routers must possess same topology database for the common area they belong to
- Topology information exchange occurs when the establishment of an adjacency relationship is attempted



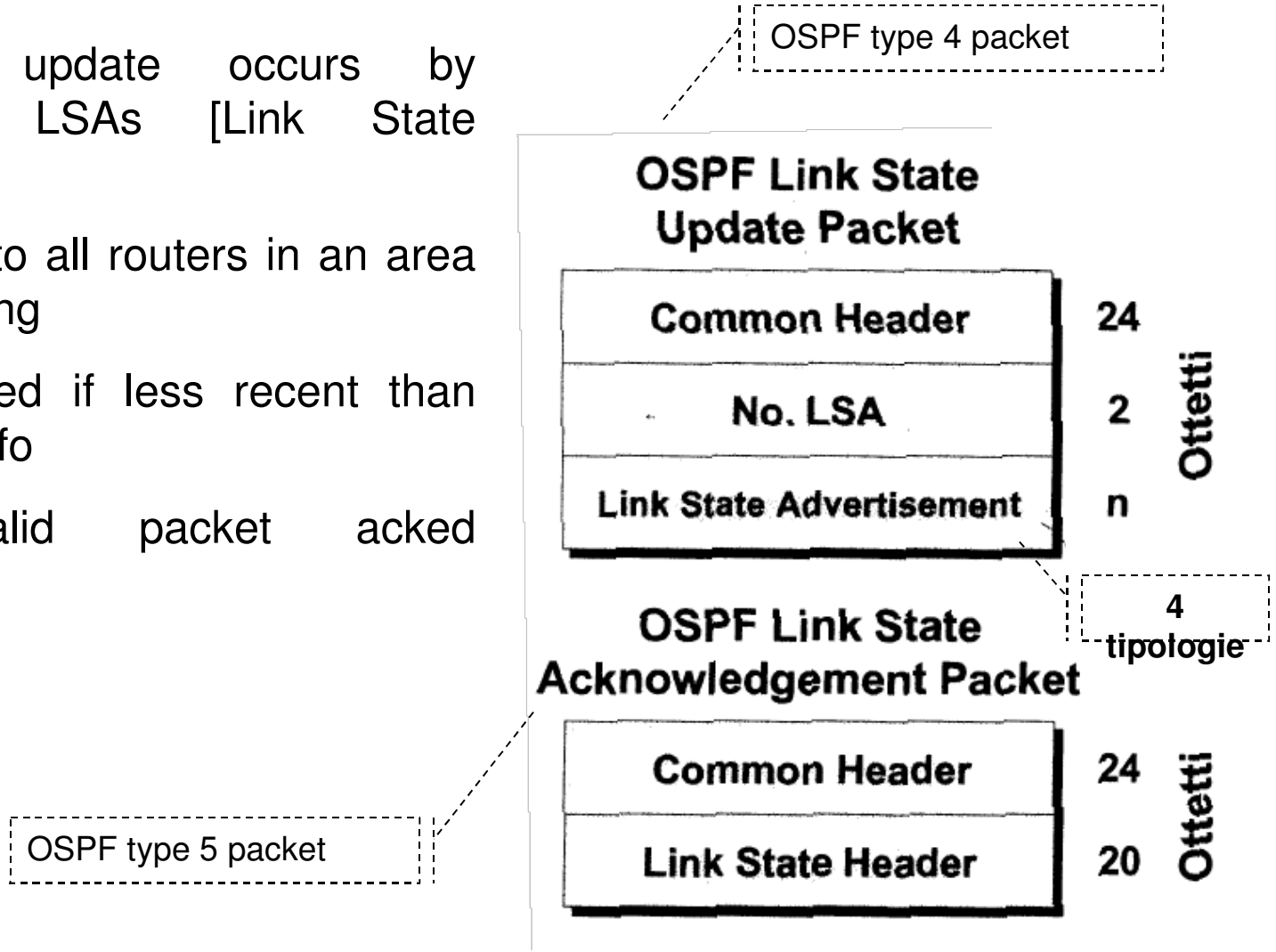
Fase 3: Establishing Adjacencies – Database Exchange/cont.

- Master Router [MS=1] sends own database
  - Master is the first to transmit
- Database can be sent using a set of packets [M=1]
- Slave router acks packets, at the same time sending its own database to Master
- If two database records are different, more recent prevails [timestamp analysis]



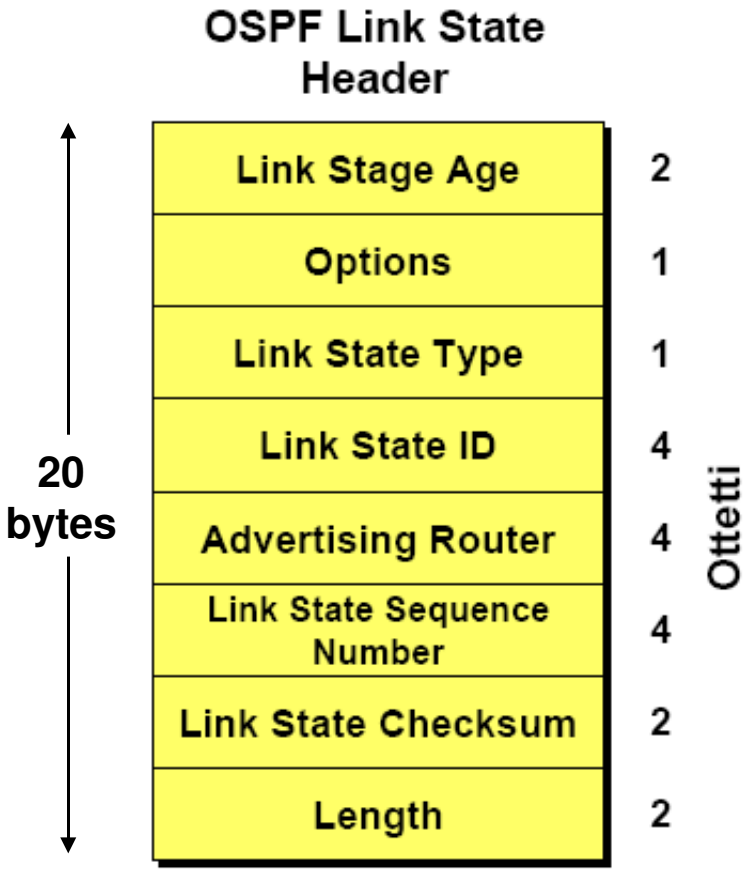
Phase 4: Link State Propagation

- Database update occurs by exchaning LSAs [Link State Packets]
- LSAs sent to all routers in an area using flooding
- LSA dropped if less recent than database info
- Every valid packet acked individually



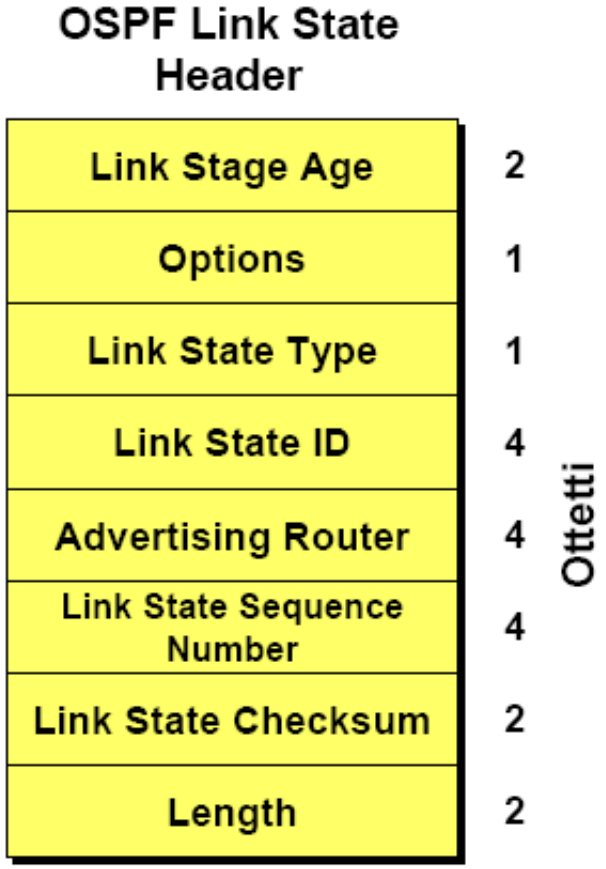
Link State Advertisment 1/4

- **All types of LSA have same header**
  - **Link Stage Age**
    - time (in secs) of advertisement since emission
  - **Link State Type**
    - 1: Router link
    - 2: Network link
    - 3: Summary link
      - *inter-area*, *intra-AS route*
    - 4: Summary link
      - *route to AS Boundary Router*
    - 5: AS External link
      - *route external networks (I.e., belonging to different ASes)*

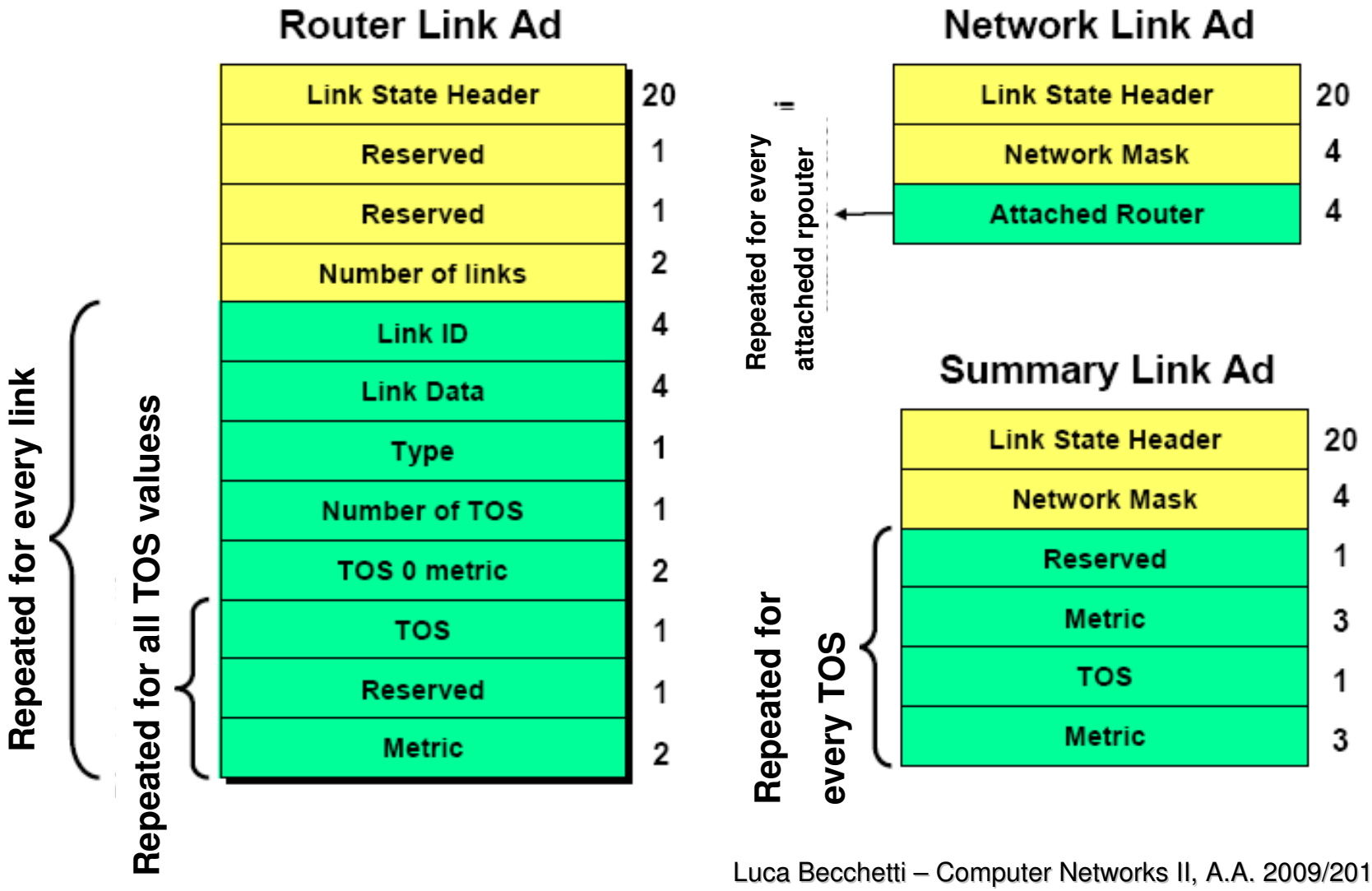


Link State Advertisment 2/4

- **Link State ID**
  - ID of link to which *this* message refers
  - *Type 1 and 4*: IP address of sending router
  - *Type 3 and 5*: IP network address of network to which message refers
  - *Type 2*: IP of sending DR
- **Advertising Router:**
  - IP address of router sending message
  - *Type 1*: same a Link State ID
  - *Type 2*: DR's IP address
  - *Types 3 and 4*: ABR's IP address
  - *Type 5*: ASBR's IP address

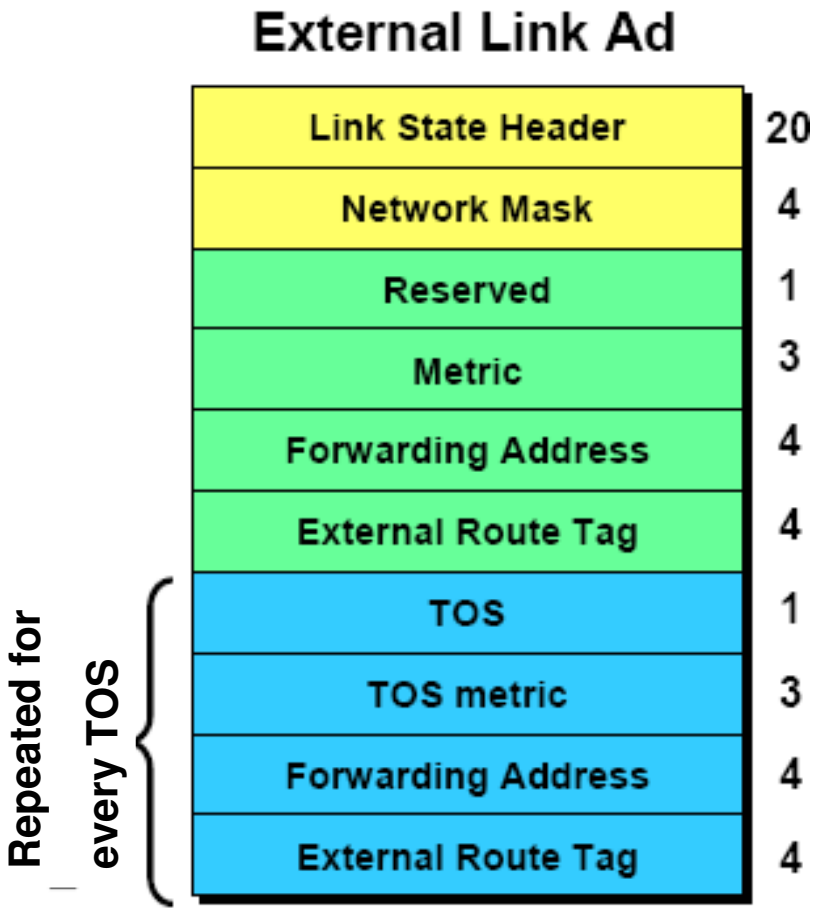


Link State Advertisment 3/4

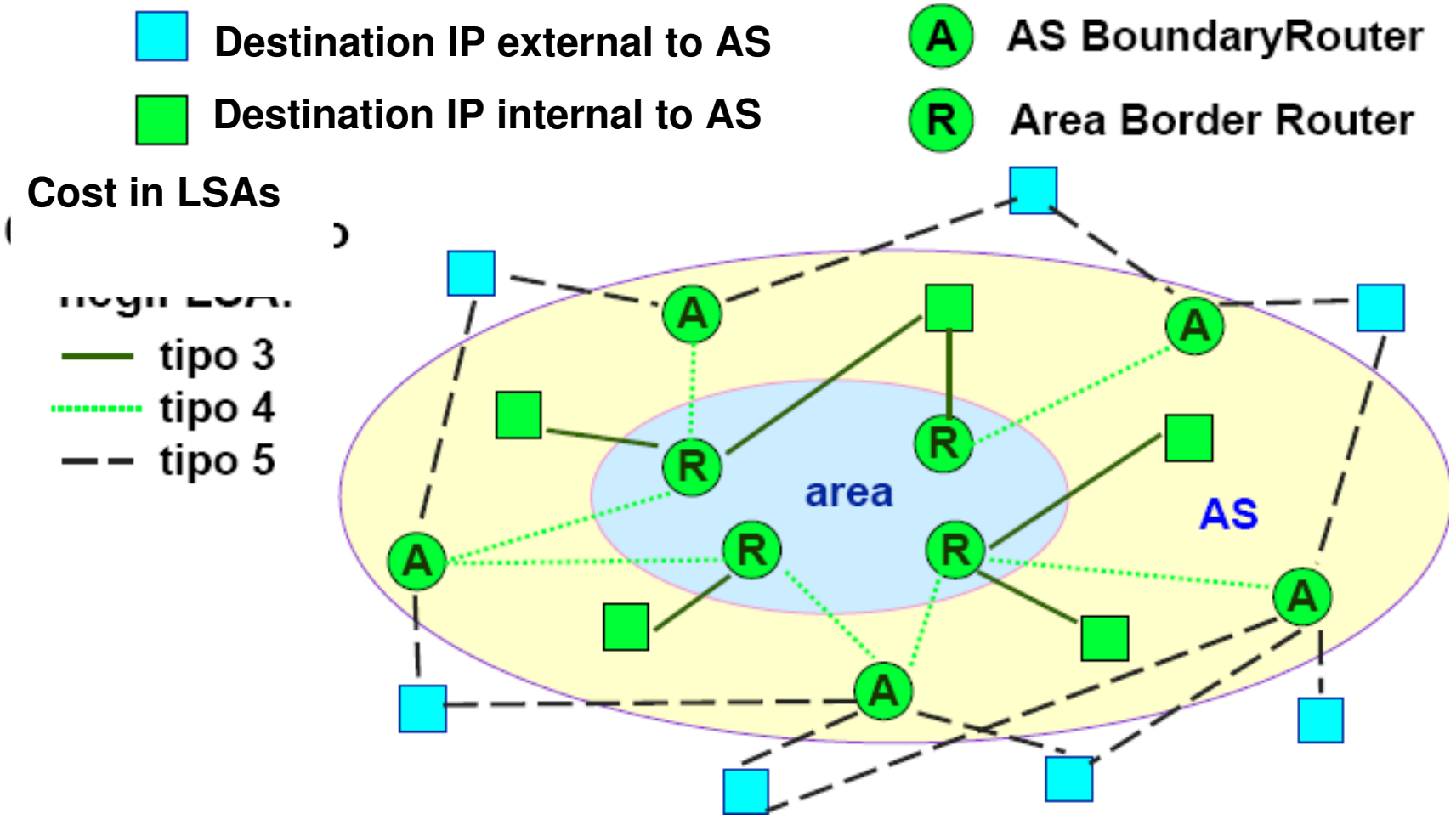


Link State Advertisment 4/4

- **Network Mask**
  - Mask identifying subnet to which packet refers. Network identification contained in header
- **Metric**
  - Cost of route
- **Forwarding Address**
  - IP address to which traffic directed to network has to be sent
- **External Route Tag**
  - ASBR dependent suffix



LSA types 3-4-5: comparison



Routing table computation

- Every router computes own Routing Table using topology database
- Shortes path tree within area occurs according to Dijkstra'a algorithm
- Some routes are added to RT:
  - Inter-area routes contained in summary link packets
  - Inter-AS contained AS external links packets
- RT recomputed at every database update

References

- Peterson & Davie, 4.1 e 4.2
- OSPF design guide (CISCO)
  - http://www.cisco.com/warp/public/104/1.pdf
- TCP/IP Guide:
- http://www.tcpipguide.com/free/t_OpenShortestPathFirstOSPF.htm
- RFC 2453 (RIPv2)
  - http://tools.ietf.org/html/rfc2453