Marco Listanti

Strato di rete

Protocolli "RIP", "OSPF" e "BGP"

Protocolli di instradamento

Protocolli di instradamento

- I protocolli d'instradamento intra-AS sono noti anche come Interior Gateway Protocol (IGP)
- I protocolli IGP più comuni sono:
  - RIP (Routing Information Protocol)
  - OSPF (Open Shortest Path First)
  - IGRP (Interior Gateway Routing Protocol) (protocollo proprietario Cisco)

Routing Information Protocol
RIP

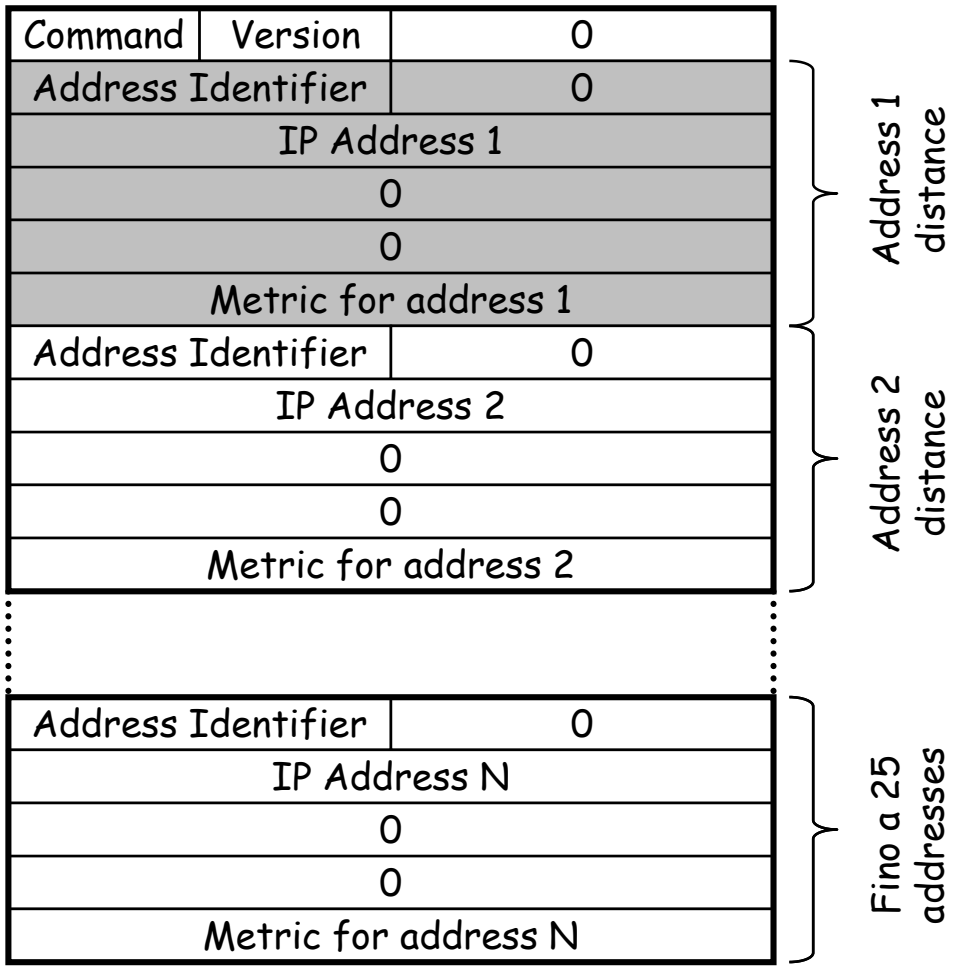
RIP

- **RFC 1058**
- **Distance Vector Routing Protocol**
  - la metrica dei rami dipende normalmente dal loro stato (sano/guasto)
  - Conteggio degli hop come metrica di costo (max = 15 hop)
- **E' utilizzato in reti di piccole-medie dimensioni**
- **E' molto semplice, ma**
  - la convergenza è lenta
  - lo stato di equilibrio può essere un sub-ottimo

RIP

- I messaggi RIP sono trasportati dal protocollo UDP (port number 520)
- Due tipi di messaggi
  - *Request* per chiedere ai vicini il distance vector
  - *Response* per annunciare il distance vector
- Ogni messaggio *Response* contiene un elenco comprendente fino a 25 sottoreti e la distanza tra l'origine del messaggio e ciascuna di queste
- I router adiacenti si scambiano periodicamente gli aggiornamenti d'instradamento
  - Valore di default 30 secondi

RIP v1



- **Header**
  - Command
    - request
    - response
  - Version
- **Block**
  - IP address
    - rete, sottorete o host
  - Metric
    - distanza dalla rete indicata nell'IP address

RIP v2

Command	Version	Reserved
Address Identifier		Reserved
IP Address 1		
Subnet Mask		
Next Hop		
Metric for address 1		
Address Identifier		Reserved
IP Address 2		
Subnet Mask		
Next Hop		
Metric for address 2		
⋮		
Address Identifier		Reserved
IP Address N		
Subnet Mask		
Next Hop		
Metric for address N		

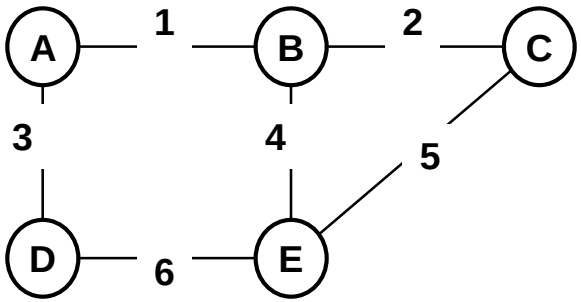
Address 1 distance

Address 2 distance

Fino a 25 addresses

- IP address
  - rete, sottorete o host
- Subnet Mask
  - specifica come interpretare i bit dell'indirizzo
- Next Hop
  - indica a quale next hop router il router emittente il messaggio RIP invierà i pacchetti diretti all'indirizzo specificato
- Metric
  - distanza dalla rete indicata nell'IP address

Esempio RIP: Inizializzazione (1)



- **Condizione iniziale**
  - Routing table vuote
- **Metrica**
  - Distanza

Routing Table

A

Destinazione	A	B	C	D	E
Distanza	0	?	?	?	?
Link	local	?	?	?	?

B

Destinazione	A	B	C	D	E
Distanza	?	0	?	?	?
Link	?	local	?	?	?

C

Destinazione	A	B	C	D	E
Distanza	?	?	0	?	?
Link	?	?	local	?	?

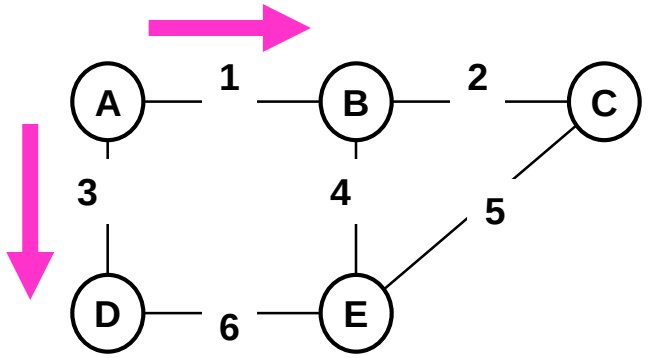
D

Destinazione	A	B	C	D	E
Distanza	?	?	?	0	?
Link	?	?	?	local	?

E

Destinazione	A	B	C	D	E
Distanza	?	?	?	?	0
Link	?	?	?	?	local

Esempio RIP: Inizializzazione (2)



- **Step 2:**
  - A emette un messaggio verso B e D

A	Address	A	---	---	---	---
	Metric	0	---	---	---	---

Routing Table

A	Destinazione	A	B	C	D	E
	Distanza	0	?	?	?	?
	Link	local	?	?	?	?

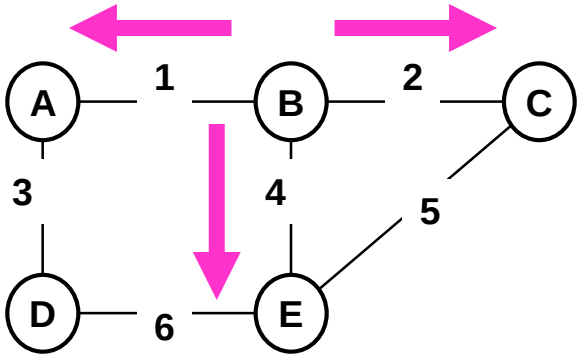
B	Destinazione	A	B	C	D	E
	Distanza	1	0	?	?	?
	Link	1	local	?	?	?

C	Destinazione	A	B	C	D	E
	Distanza	?	?	0	?	?
	Link	?	?	local	?	?

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

E	Destinazione	A	B	C	D	E
	Distanza	?	?	?	?	0
	Link	?	?	?	?	local

Esempio RIP: Inizializzazione (3)



- **Step 3:**
  - B emette un messaggio verso A, C e E

B	Address	A	B	---	---	---
	Metric	1	0	---	---	---

Routing Table

A	Destinazione	A	B	C	D	E
	Distanza	0	1	?	?	?
	Link	local	1	?	?	?

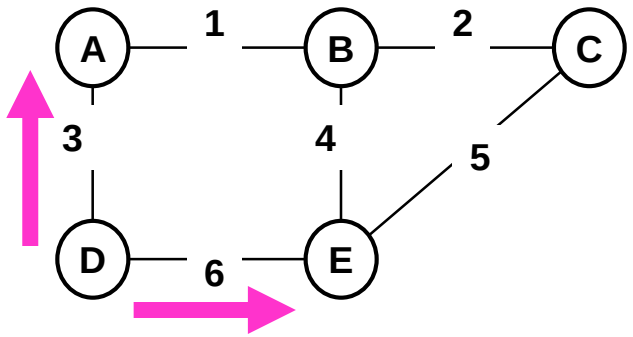
B	Destinazione	A	B	C	D	E
	Distanza	1	0	?	?	?
	Link	1	local	?	?	?

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

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

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

Esempio RIP : Inizializzazione (4)



- **Step 4:**
  - D emette un messaggio verso A e E

D	Address	A	---	---	D	---
	Metric	1	---	---	0	---

Routing Table

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

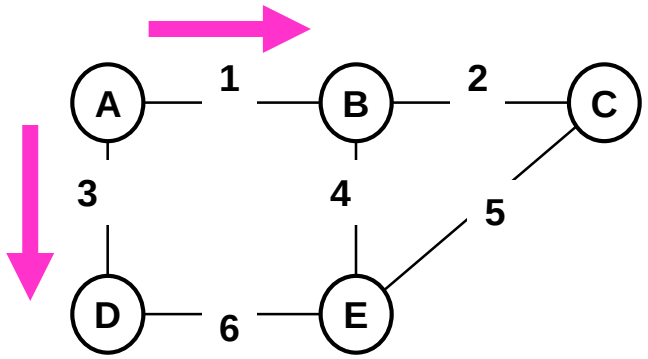
B	Destinazione	A	B	C	D	E
	Distanza	1	0	?	?	?
	Link	1	local	?	?	?

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

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

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

Esempio RIP : Inizializzazione (5)



- **Step 5:**
  - A emette un messaggio verso B e D

A	Address	A	B	---	D	---
	Metric	0	1	---	1	---

Routing Table

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

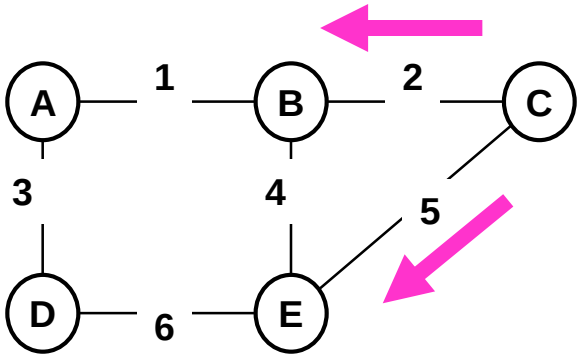
B	Destinazione	A	B	C	D	E
	Distanza	1	0	?	2	?
	Link	1	local	?	1	?

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

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

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

Esempio RIP : Inizializzazione (6)



- **Step 6:**
  - C emette un messaggio verso B e E

C	Address	A	B	C	---	---
	Metric	2	1	0	---	---

Routing Table

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

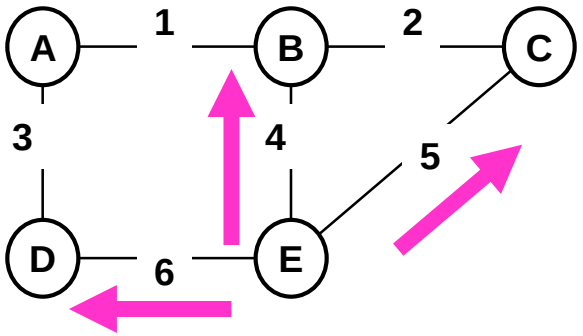
B	Destinazione	A	B	C	D	E
	Distanza	1	0	1	2	?
	Link	1	local	2	1	?

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

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

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

Esempio RIP : Inizializzazione (7)



- **Step 7:**
  - E emette un messaggio verso B, C e D

E	Address	A	B	C	D	E
	Metric	2	1	1	1	0

Routing Table

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

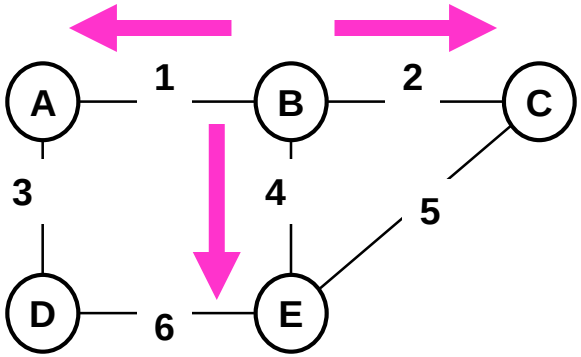
B	Destinazione	A	B	C	D	E
	Distanza	1	0	1	2	1
	Link	1	local	2	1	4

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

Esempio RIP : Inizializzazione (8)



- **Step 8:**
  - B emette un messaggio verso A, C e E

B	Address	A	B	C	D	E
	Metric	1	0	1	2	1

Routing Table

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

B	Destinazione	A	B	C	D	E
	Distanza	1	0	1	2	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

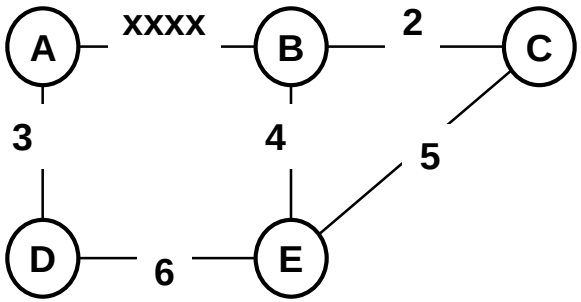
D	Destinazione	A	B	C	D	E
	Distanza	1	2	2	0	1
	Link	3	3	6	local	6

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

RIP: guasto sul collegamento e recupero

- **Se un router non riceve messaggi da un nodo adiacente per un intervallo di 180 sec, il nodo adiacente viene considerato spento o guasto**
  - RIP modifica la tabella d'instradamento locale
  - Propaga l'informazione mandando annunci ai router adiacenti
  - I nodi adiacenti inviano nuovi messaggi (se la loro tabella d'instradamento è cambiata)
  - L'informazione che il collegamento è guasto si propaga su tutta la rete
  - L'utilizzo dell'inversione avvelenata (poison reverse) evita i loop (distanza infinita = 16 hop)

Esempio RIP : Guasto di un ramo (1)



- **Condizione iniziale**
  - rete a regime
  - guasto del ramo AB
- **Metrica**
  - Distanza

Routing Table

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

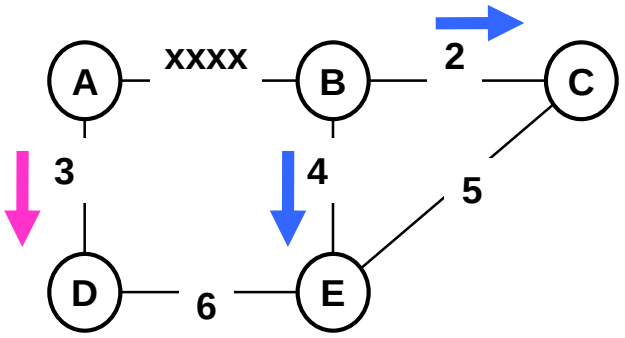
B	Destinazione	A	B	C	D	E
	Distanza	inf	0	1	inf	1
	Link	1	local	2	1	4

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

Esempio RIP : Guasto di un ramo (2)



- Step 1
  - Messaggio di A verso D

A	Address	A	B	C	D	E
	Metric	0	inf	inf	1	inf

  - Messaggio di B verso C ed E

B	Address	A	B	C	D	E
	Metric	inf	0	1	inf	1

Routing Table

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

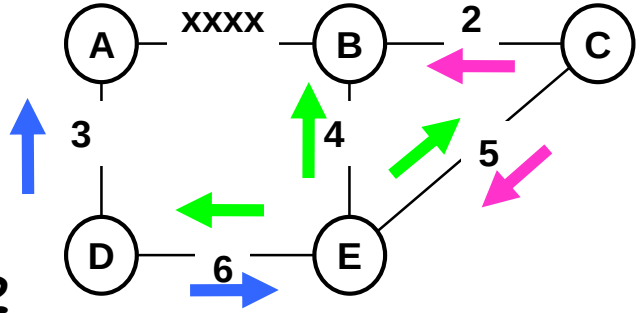
B	Destinazione	A	B	C	D	E
	Distanza	inf	0	1	inf	1
	Link	1	local	2	1	4

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

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

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

Esempio RIP : Guasto di un ramo (3)



- **Step 2**
  - Messaggio di C verso B, E

C	Address	A	B	C	D	E
	Metric	inf	1	0	2	inf

  - Messaggio di D verso A, E

D	Address	A	B	C	D	E
	Metric	1	inf	2	0	1

  - Messaggio di E verso B, C, D

E	Address	A	B	C	D	E
	Metric	inf	1	1	1	0

Routing Table

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

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

C	Destinazione	A	B	C	D	E
	Distanza	inf	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	6	6	local	6

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

Esempio RIP : Guasto di un ramo (4)

- **Step 3**
  - Messaggio di A verso D

A	Address	A	B	C	D	E
	Metric	0	inf	3	1	2

  - Messaggio di B verso E, C

B	Address	A	B	C	D	E
	Metric	inf	0	1	2	1

  - Messaggio di D verso A, E

D	Address	A	B	C	D	E
	Metric	1	2	2	0	1

  - Messaggio di E verso B, C, D

E	Address	A	B	C	D	E
	Metric	2	1	1	1	0

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

Open Shortest Path First
OSPF

OSPF (Open Shortest Path First)

- **RFC 2328**
- **È un protocollo "link state"**
  - Utilizza il flooding di informazioni sullo stato dei link
    - Messaggi **Link State Advertisement** (**LSA**)
  - Utilizza l'algoritmo di Dijkstra per la determinazione del percorso a costo minimo
- **Al momento di un cambiamento di stato di un link, il router emette un LSA verso tutti gli altri router**
- **Gli LSA sono trasferiti nel sistema autonomo utilizzando ilmeccanismo di flooding**
  - I messaggi OSPF vengono trasportati direttamente in pacchetti IP
    - Non è utilizzato un protocollo di trasporto (TCP o UDP)
  - "Rapida" convergenza in caso di cambiamenti di stato

Vantaggi di OSPF

- **Sicurezza**
  - gli scambi tra router sono autenticati
- **Multipath**
  - quando più percorsi verso una destinazione hanno lo stesso costo, OSPF consente di usarli senza doverne scegliere uno, come invece avveniva in RIP
  - **Equal Path Cost Multipath** (**ECMP**)
- **Su ciascun collegamento, vi possono essere più metriche di costo per differenti TOS**
  - es: il costo di un link via satellite sarà "basso" per un pacchetto best effort; "elevato" per un pacchetto real time
- **Supporto integrato per l'instradamento unicast e multicast**
  - Per consentire l'instradamento multicast viene impiegato MOSPF (OSPF multicast) che utilizza il database topologico di OSPF
- **Supporto alle gerarchie in un dominio d'instradamento**

Link State Routing

- **Gli LSA sono emessi**
  - quando un router contatta un nuovo router adiacente
  - quando un link si guasta
  - quando il costo di un link varia
  - periodicamente ogni fissato intervallo di tempo
- **La rete trasporta gli LSA mediante la tecnica di *flooding***
  - un LSA è rilanciato da un router su tutte le sue interfacce tranne quella da cui è stato ricevuto
  - gli LSA trasportano dei riferimenti temporali (time stamp) o numeri di sequenza per
    - evitare il rilancio di pacchetti già rilanciati
    - consentire un corretto riscontro dal ricevente

Tecnica Flooding

- **Obiettivi di OSPF**
  - Tutti i router di una rete abbiano un database topologico contenente lo stato della rete
  - Tutti i router di una rete abbiano le stesse informazioni sullo stato dei link
- **Alla ricezione di un LSP**
  - un router esamina i campi di un LSP: link identifier, metrica, time stamp o numero di sequenza
  - se il dato non è contenuto nel database, viene memorizzato e l'LSP è rilanciato su tutte le interfacce del router tranne quella di ricezione
  - se il dato ricevuto è più recente di quello contenuto nel database, il suo valore è memorizzato e l'LSP è rilanciato su tutte le interfacce del router tranne quella di ricezione
  - se il dato ricevuto è più vecchio di quello contenuto nel database, viene rilanciato un LSP con il valore contenuto nel database esclusivamente sull'interfaccia di arrivo dell'LSP
  - se i due dati sono della stessa età non viene eseguita alcuna operazione

Tecnica Flooding

- **La tecnica *flooding* ha i seguenti vantaggi**
  - esplora tutti i possibili cammini tra origine e destinazione
  - è estremamente affidabile e robusta
  - almeno una copia di ogni LSP seguirà la via a minor costo
- **Il traffico di controllo generato dipende dalle dimensioni della rete e può essere molto elevato**

Suddivisione di grandi reti in aree

- **Se la rete è di grandi dimensioni**
  - Cresce il numero di record del database e quindi la memoria necessaria in ogni router
  - Cresce il tempo necessario al calcolo dei percorsi
  - Cresce il traffico di segnalazione dovuto all'invio degli LSP
- **OSPF supporta un instradamento di tipo gerarchico**
  - Una rete è suddivisa in aree
    - Sezioni indipendenti di rete
    - Database separati
    - Meccanismi di flooding indipendenti
  - Le singole aree sono interconnesse da un area di backbone

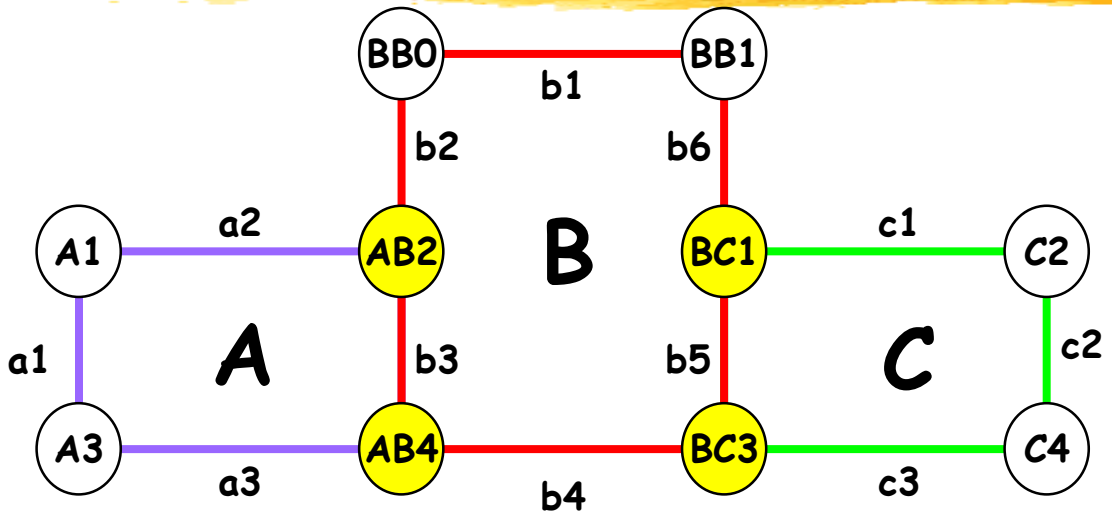
Suddivisione di grandi reti in aree (2)

- **I router che interconnettono aree diverse sono detti Area Border Router – ABR**
  - Gli ABR appartengono ad aree diverse
  - Ogni area ha almeno un ABR
  - Ogni area è almeno connessa all'area di backbone
    - Almeno un ABR è connesso all'area di backbone
- **Un ABR**
  - contiene i database di tutte le aree a cui appartiene
  - emette degli appositi messaggi (summary records) che contengono la lista delle sottoreti raggiungibili attraverso le aree a cui appartiene

Instradamenti esterni

- Un AS è connesso ad altri AS attraverso uno o più "AS Border Router" (ASBR)
- Se un area ha un unico ASBR è sufficiente indicare a tutti i router interni all'area l'instradamento di default verso l'esterno
- Se gli ASBR sono più di uno, ognuno di essi indicherà ai router interni il costo della via verso l'esterno
  - External record

Suddivisione di grandi reti in aree (3)



- **Il database di un router dell'area A conterrà**
  - I record dei link a1, a2, a3, comunicati dai router A1, A3, AB2, AB4
  - I summary record relativi alle sottoreti comprese nell'area di backbone e nell'area C, comunicati dai router AB2 e AB4
    - Ad ogni sottorete sarà associato il costo di raggiungimento
    - Analogia con i protocolli distance vector
  - Gli external record emessi dai router BB0 e BB1 e rilanciati dai router AB2 e AB4
    - Ad ogni destinazione sarà associato il costo di raggiungimento

Open Shortest Path First

- **OSPF è il protocollo IGP più utilizzato nelle di grandi dimensioni:**
  - è basato sullo scambio di LSP detti Link State Advertisement (LSA)
  - supporta metriche relativi a diversi valori del campo TOS
  - supporta l'uso del concetto di variable length subnet mask (CIDR)
  - supporta il servizio di autenticazione tra router
  - supporta l'indicazione di specific routes
  - riduzione delle dimensione delle tabelle di routing con l'uso del concetto di Designated Router (DR)
  - supporta l'indicazione di virtual link per l'interconnessione di aree non contigue

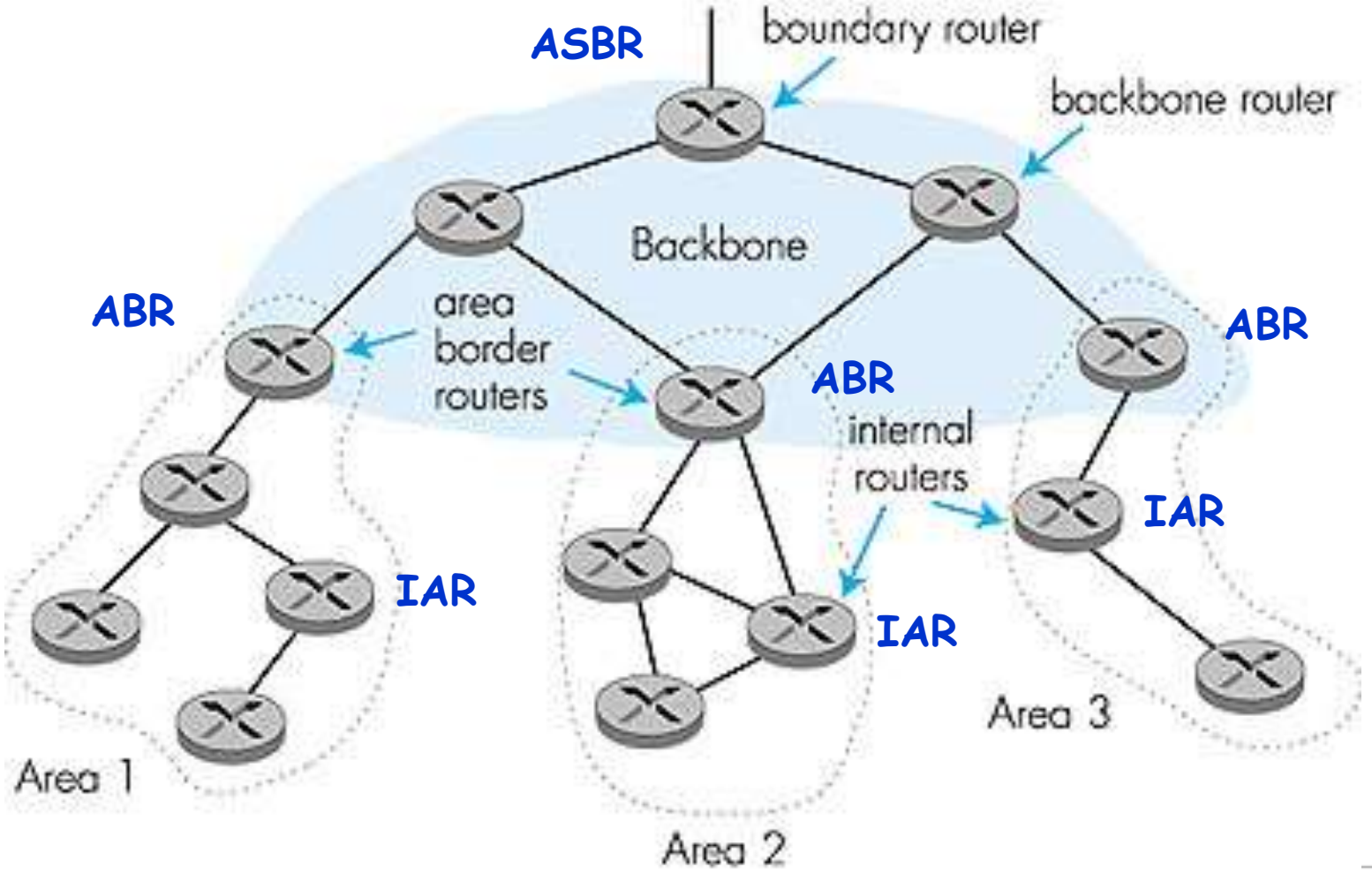
Terminologia OSPF

- **Area**
  - è un insieme logico di reti e di router (es. geografico, amministrativo, ...)
  - ha lo scopo di limitare la dimensione dei database di descrizione della topologia di rete all'interno dei router
  - all'interno di un area i router devono avere database identici che descrivono la topologia di rete
  - informazioni sulla parte di rete esterna all'area sono contenute dagli Area Border Router (ABR)
  - un Area Border Router trasmette LSA contenenti informazioni sulle reti esterne all'interno dell'area (costo di raggiungimento)
  - tutte le reti OSPF devono essere composte da almeno un area, denominata area di backbone

Terminologia OSPF

- **Intra-Area Router (IAR)**
  - sono i router che sono situati all'interno di una area OSPF
  - scambiano LSA con tutti gli altri router dell'area
  - gestiscono il database relativo alla topologia dell'area
- **Area Border Router (ABR)**
  - sono i router che sono connessi a due o più aree OSPF
  - gestiscono i database topologici di tutte le aree a cui sono connessi
  - trasmettono all'interno di ogni area LSA relativi alle reti presenti in ogni area
- **AS Boundary Router (ASBR)**
  - sono i router che sono situati a bordi del dominio OSPF
  - scambiano LSA contenenti informazioni di raggiungibilità di reti di altri AS
  - inviano LSA all'interno del dominio con informazioni sui percorsi esterni

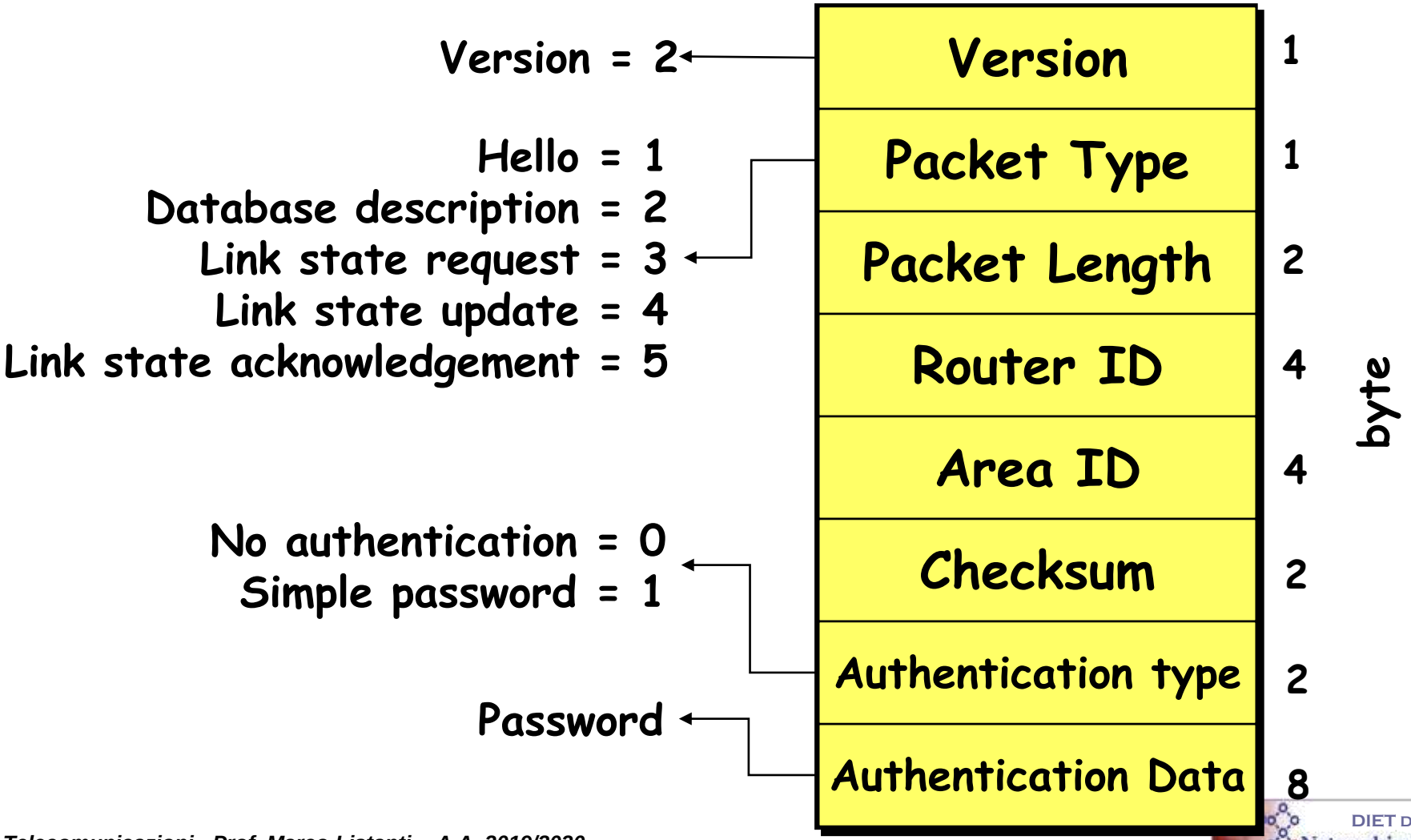
OSPF strutturato gerarchicamente



Tipologie di LSA

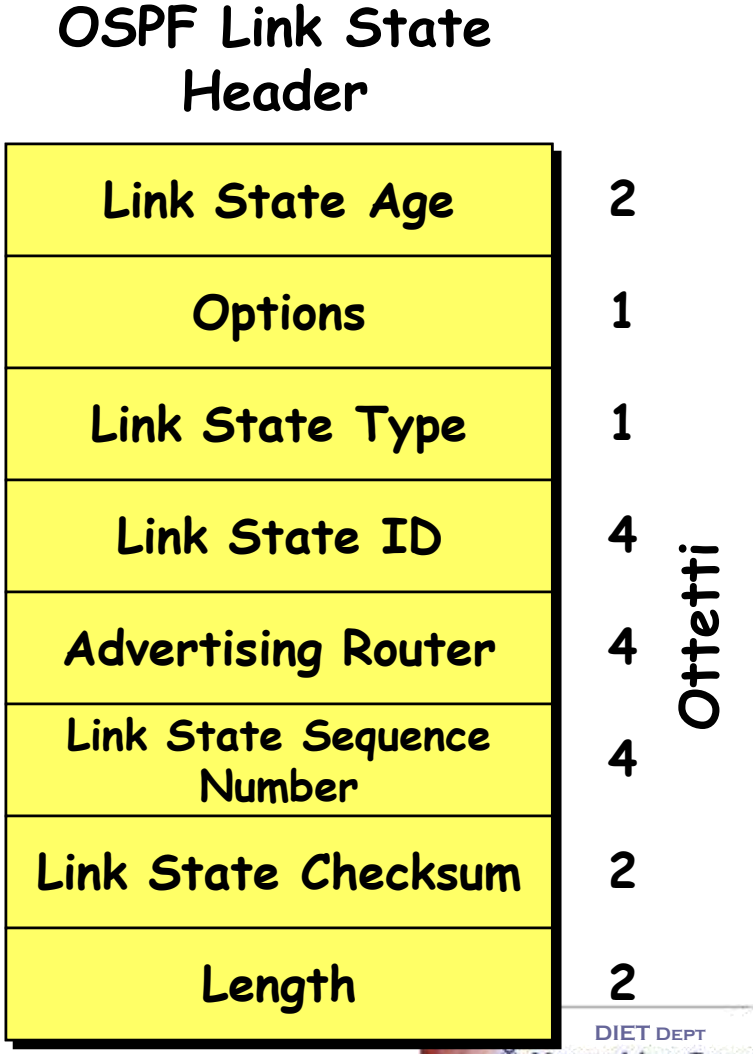
- **Link State Advertisements (LSA)**
  - sono i messaggi scambiati tra router OSPF per aggiornare i link state database e i percorsi inter-area e inter-AS
  - Router link advertisement
    - indicano lo stato dei link uscenti da un router, sono inviati all'interno di una singola area
  - Summary link advertisement
    - sono generati dagli ABR e individuano le reti contenute nelle altre aree ed i relativi costi di raggiungimento, sono inviati all'interno di tutte le aree gestite da un ABR
  - AS external link advertisement
    - sono generati dagli ASBR e indicano i cammini verso le reti esterne al dominio OSPF, sono inviati all'interno di tutte le aree di un dominio OSPF

Header pacchetti OSPF



Link State Advertisment (1)

- **Tutti i tipi di LSA hanno lo stesso header**
  - Link State Age
    - indica il tempo (in secondi) di emissione dell'advertisement
  - Link State Type
    - 1: Router link
    - 2: Network link
    - 3: Summary link
      - **inter-area, intra-AS route**
    - 4: Summary link
      - **route verso l'AS Boundary Router**
    - 5: AS External link
      - **route verso reti esterne all'AS**



Link State Advertisment (2)

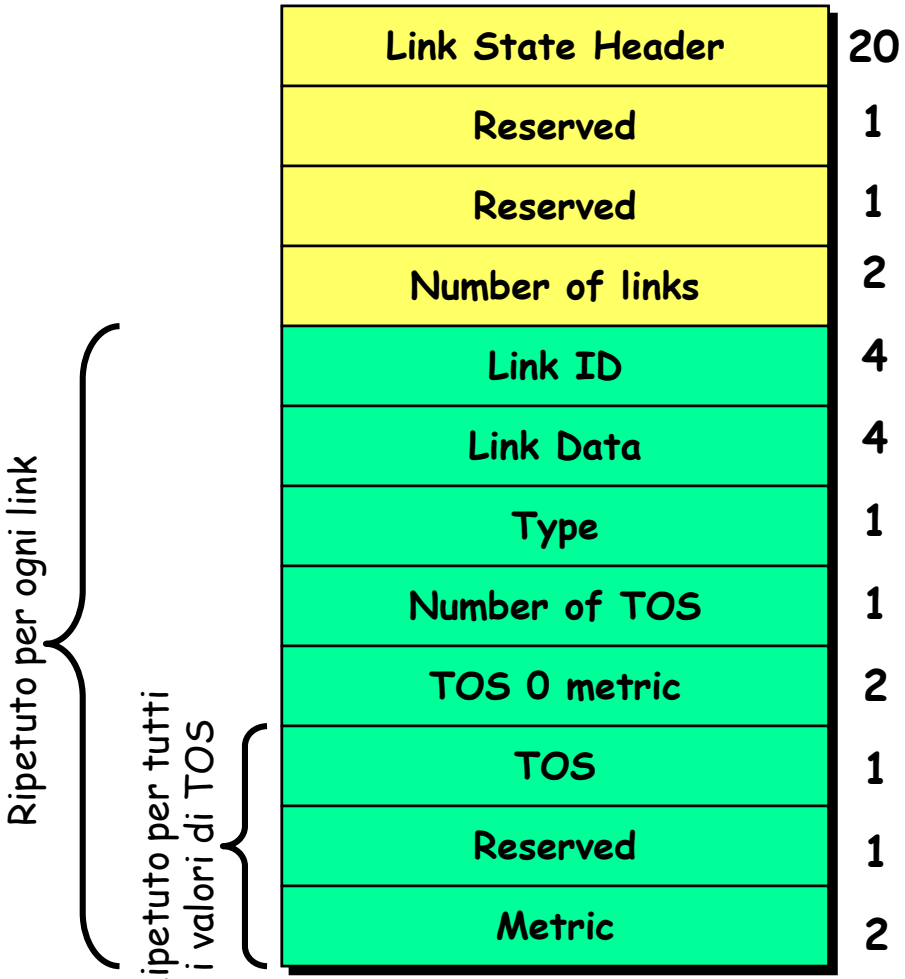
- Link State ID
  - Indica il tipo di link a cui si riferisce il messaggio
  - Tipo 1 e 4: indirizzo IP del Router emittente
  - Tipo 3 e 5: indirizzo IP della rete a cui si riferisce il messaggio
  - Tipo 2: indirizzo IP del DR emittente
- Advertising Router
  - Indirizzo IP del router che ha emesso il messaggio
  - Tipo 1 : identico al campo Link State ID
  - Tipo 2: indirizzo IP del DR
  - Tipo 3 e 4: indirizzo IP del ABR
  - Tipo 5: indirizzo IP del ASBR

OSPF Link State Header

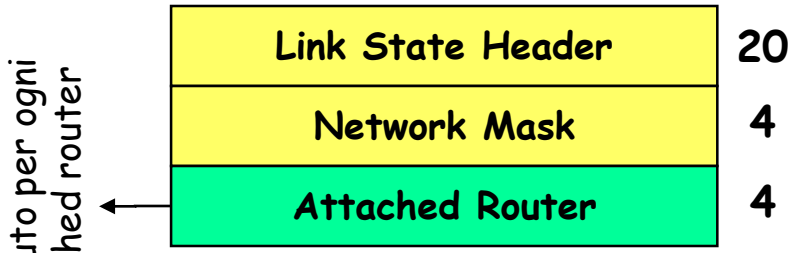
	Ottetti
Link Stage Age	2
Options	1
Link State Type	1
Link State ID	4
Advertising Router	4
Link State Sequence Number	4
Link State Checksum	2
Length	2

Link State Advertisement (3)

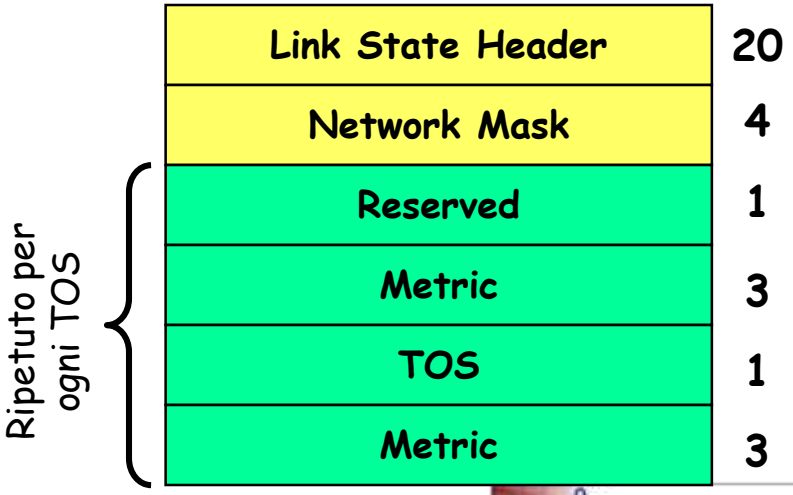
Router Link Ad



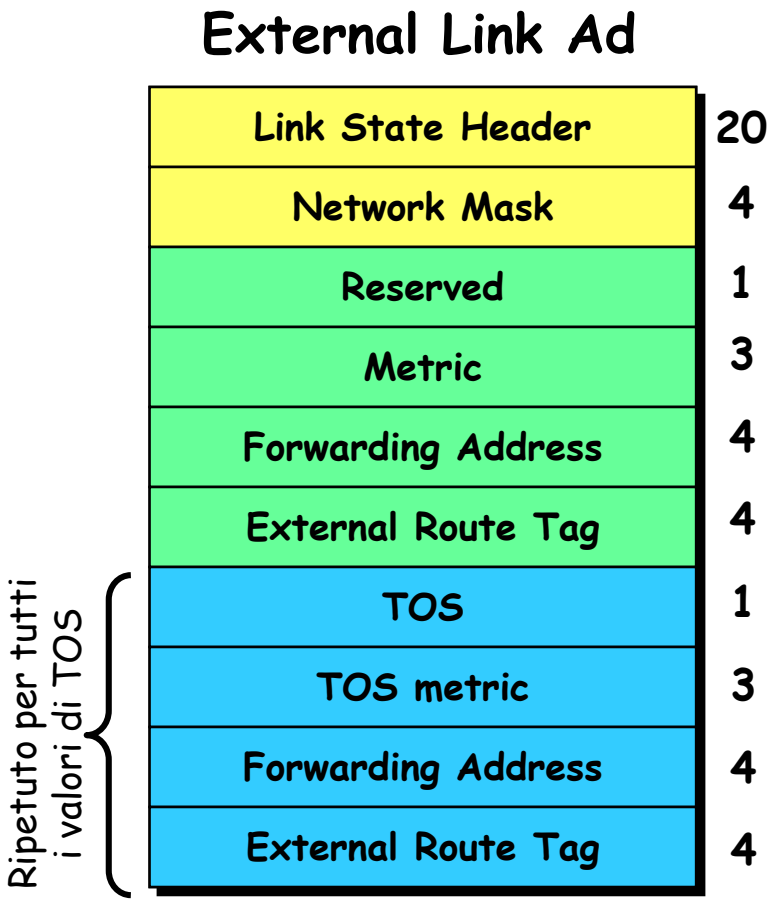
Network Link Ad



Summary Link Ad



Link State Advertisement (4)



- **Network Mask**
  - maschera della rete a cui si riferisce il pacchetto, l'indicazione della rete è contenuta nell'header
- **Metric**
  - costo del cammino
- **Forwarding Address**
  - Indirizzo IP a cui deve essere inviato il traffico diretto alla rete indicata
- **External Route Tag**
  - suffisso ad uso degli ASBR

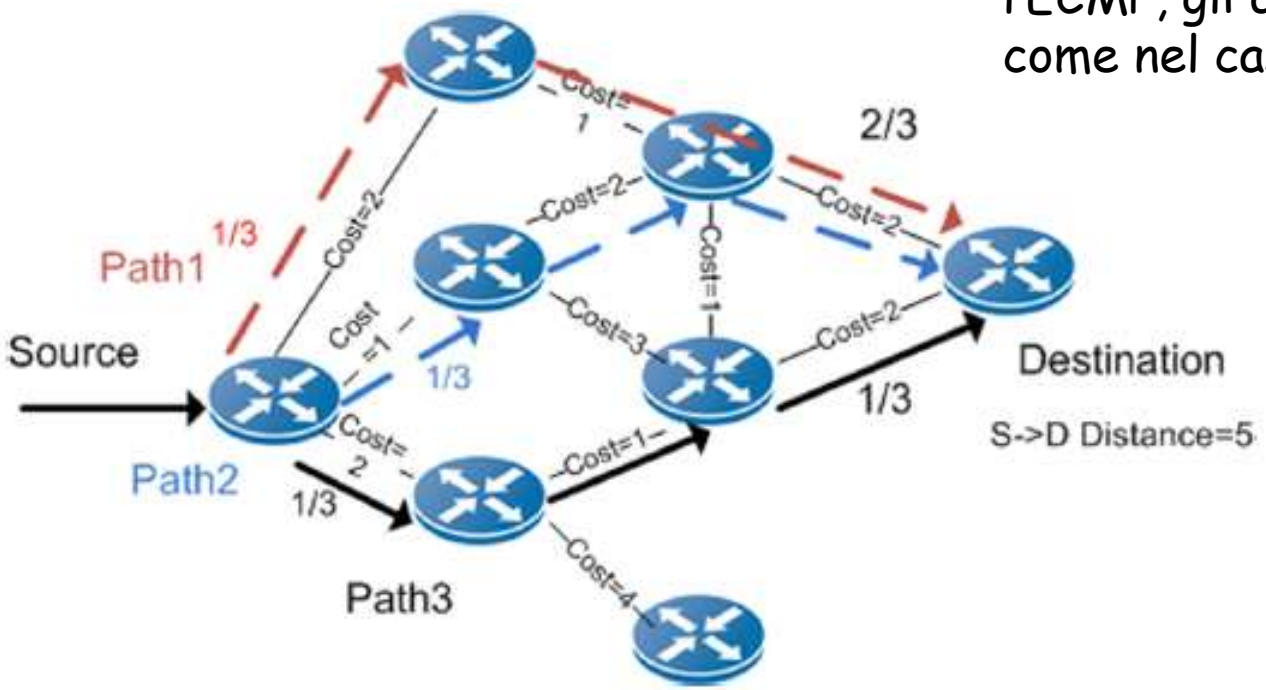
Equal Cost Multi Path (ECMP)

ECMP

- ECMP è una tecnica che rende possibile l'uso di "equal cost path" tra le sottoreti sorgente e destinazione tra cui distribuire il traffico
- Gli "equal cost path" calcolati dall'algoritmo di Dijkstra sono memorizzati nella tabella di bilanciamento del carico (Load Balancing Table)
- Il throughput di rete aumenta di un valore variabile tra il 50% e 110%
- I percorsi alternativi possono essere utilizzati come backup reciproco in caso di guasto in rete

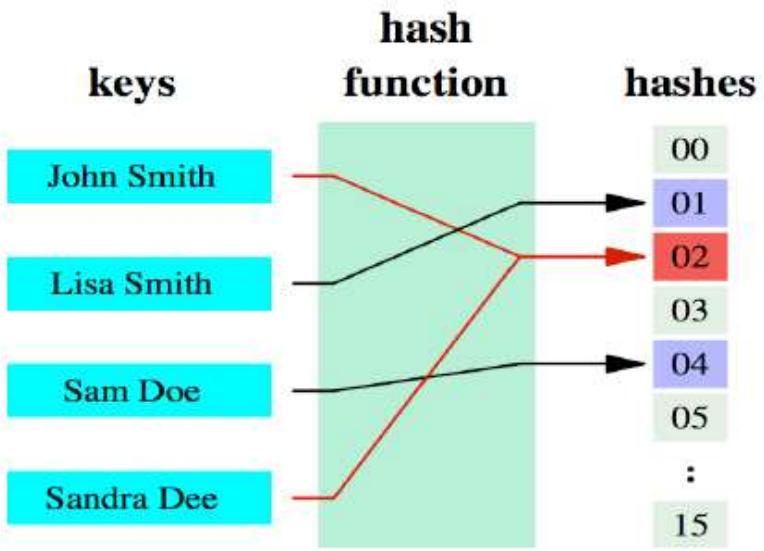
ECMP: esempio

- Il traffico tra S e D è ripartito sui tre cammini in modo uguale
- Solo il nodo sorgente supporta l'ECMP, gli altri nodi si comportano come nel caso si single path



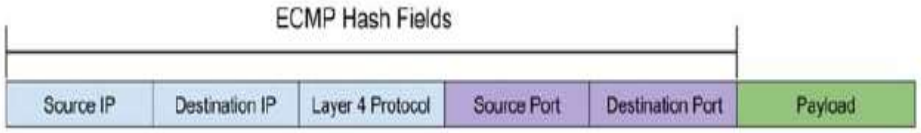
Flow Hashing ECMP

- **Funzione Hash**
  - Funzione che trasforma dati di grandi dimensioni e di lunghezza variabile in una stringhe di dimensioni piccole e di lunghezza costante
  - Spesso l'obiettivo è quello di utilizzare l'hash di un dato come indice di accesso ad per il lookup in una tabella
- **Flow Hashing ECMP**
  - Esegue la ripartizione del traffico sui cammini disponibili a costo uguale per bilanciare il carico in rete
  - Calcola l'hash dell'header del pacchetto che viene utilizzato come input per il lookup della output port dello switch
  - E' conservata la corretta sequenza dei pacchetti
    - I pacchetti di uno stesso flusso (es. stessa coppia sorgente, destinazione) sono instradati sullo stesso path



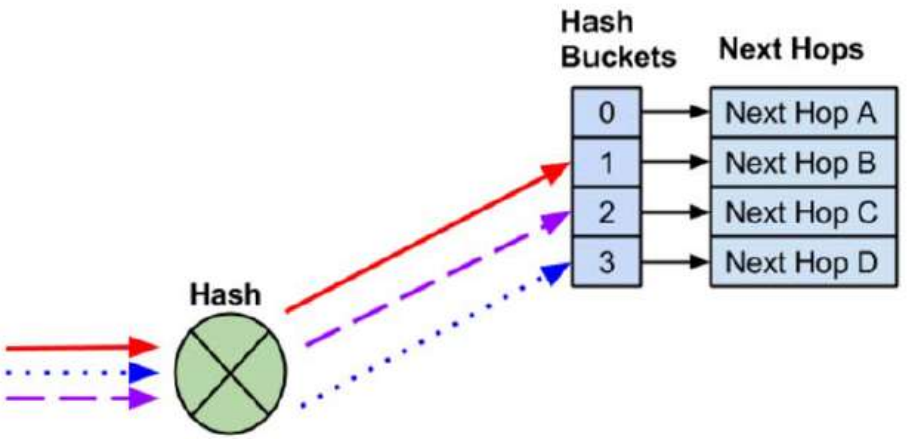
Flow Hashing ECMP: implementazione

- La tabella di routing contiene per ogni destinazione entry multipli associati ai path con costo uguale
- L'hash dell'header di un pacchetto è usato come indice per l'accesso alla tabella di routing per la decisione della porta di uscita dal router
- Oltre all'informazione dell'interfaccia di ingresso, i campi di un pacchetto IP/TCP utilizzati per la funzione di hash sono normalmente
  - IP protocol
  - Source & Destination IP
  - Source & Destination Port
- Poichè la funzione hash è deter
  stesso per un dato input, tutti i pacchetti appartenenti allo stsso flusso (stesso header) saranno instradati sullo stesso path



Hashing ECMP: implementazione

- Poichè la funzione hash è deterministica l'output è lo stesso per lo stesso per un dato input
- Tutti i pacchetti appartenenti allo stesso flusso (header identico) saranno instradati sullo stesso path
  - Esempio
    - 4 next hop
    - 3 flussi
    - Ogni next hop è associato ad un valore dell'hash

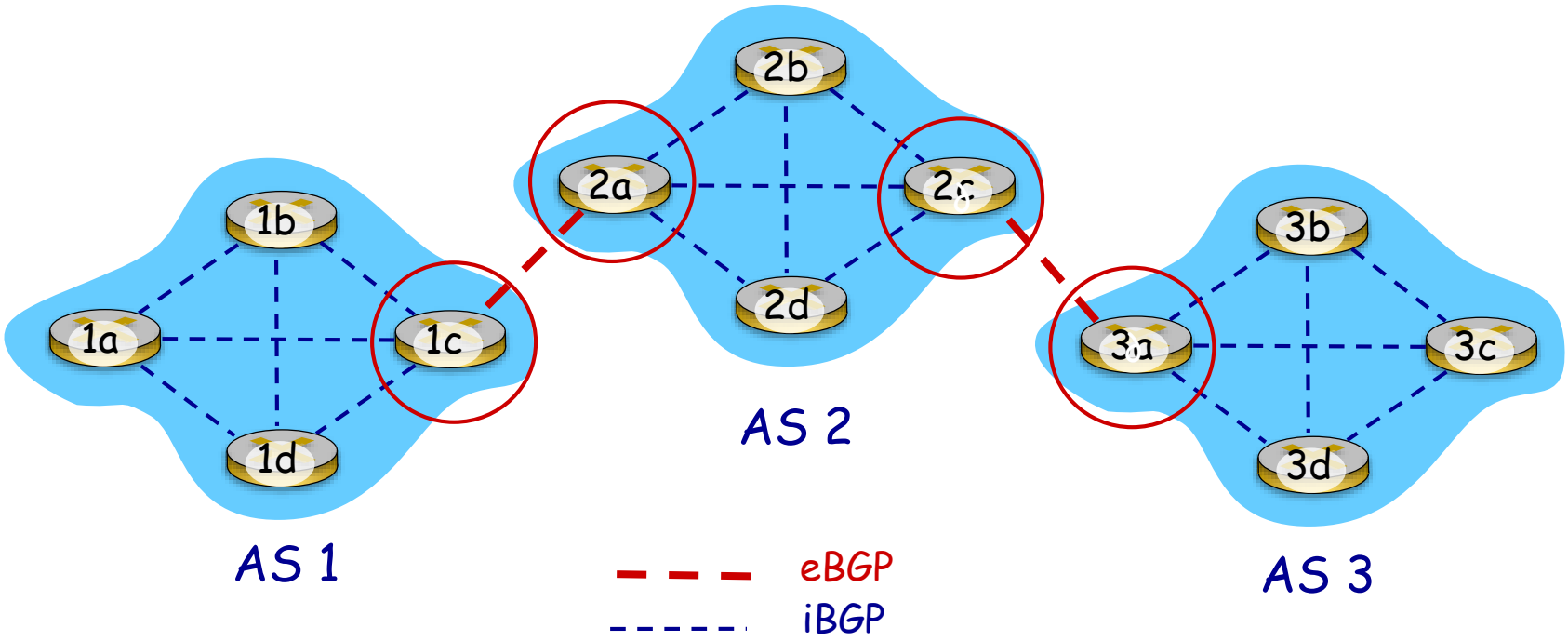


Border Gateway Protocol
BGP

BGP

- **Rappresenta lo standard dei protocolli EGP**
- **BGP mette a disposizione di ciascun AS un modo per**
  - ottenere informazioni sulla raggiungibilità delle sottoreti da parte di AS confinanti (**iBGP**)
  - propagare le informazioni di raggiungibilità a tutti i router interni di un AS (**eBGP**)
  - determinare percorsi "buoni" verso le sottoreti sulla base delle informazioni di raggiungibilità e delle politiche dell'AS
- **BGP consente a ciascuna sottorete di comunicare la propria esistenza al resto di Internet**
- **BGP è un protocollo path vector**
  - annuncia i cammini (path) verso le sottoreti (prefissi) di destinazione

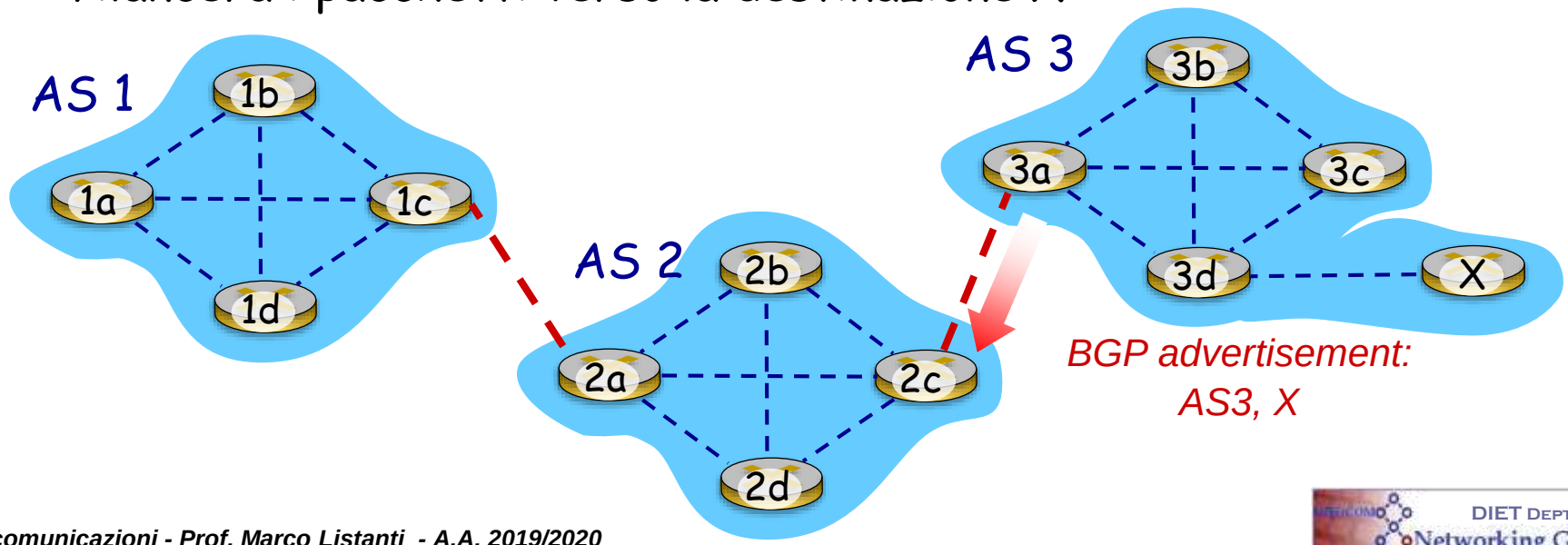
Connessioni eBGP e iBGP



I gateway router eseguono ambedue i protocolli eBGP and iBGP

BGP: generalità

- **BGP session**
  - due BGP router (peers)
    - scmbiano i messaggi BGP utilizzando una connessione TCP
    - Annunciano i cammini verso le reti di destinazione (network prefix)
- Quando il router 3a (AS3 gateway router) annuncia il cammino AS3,X a rouer 2c (AS2 gateway router) assicura che AS3 rilancerà i pacchetti verso la destinazione X

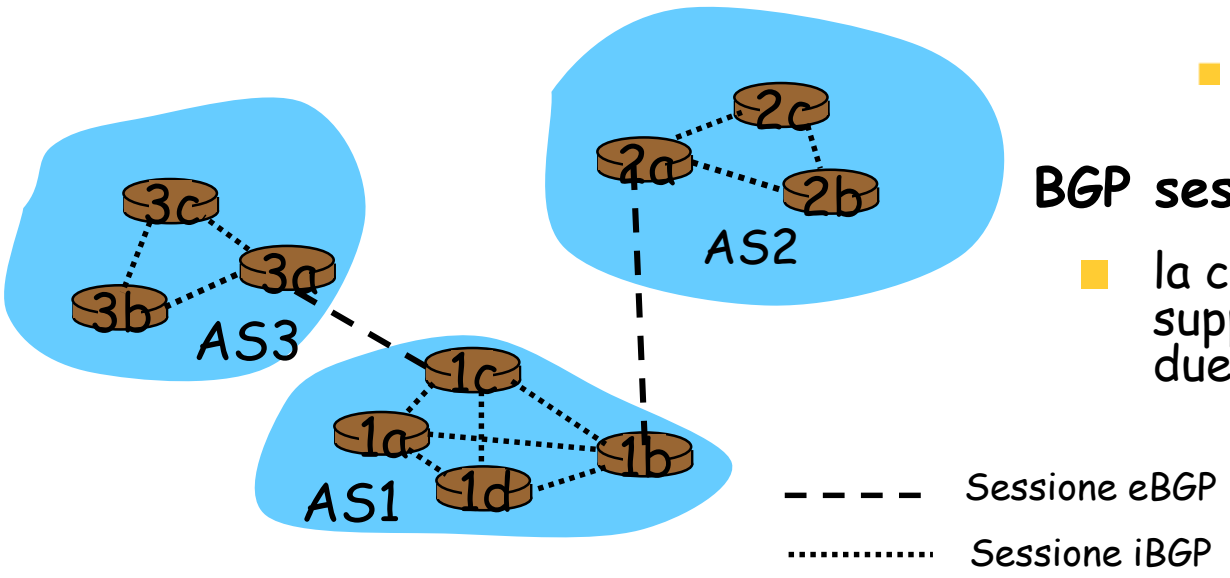


Terminologia BGP

- **BGP speaker**
  - un router che supporta il protocollo BGP
  - un BGP router non necessariamente coincide con un border router

- **BGP Neighbors**
  - una coppia di BGP speaker che si scambiano informazioni di instradamento inter-AS
  - possono essere di due tipi
    - *Interni*: se appartengono allo stesso AS
    - *Esterni*: se appartengono ad AS diversi

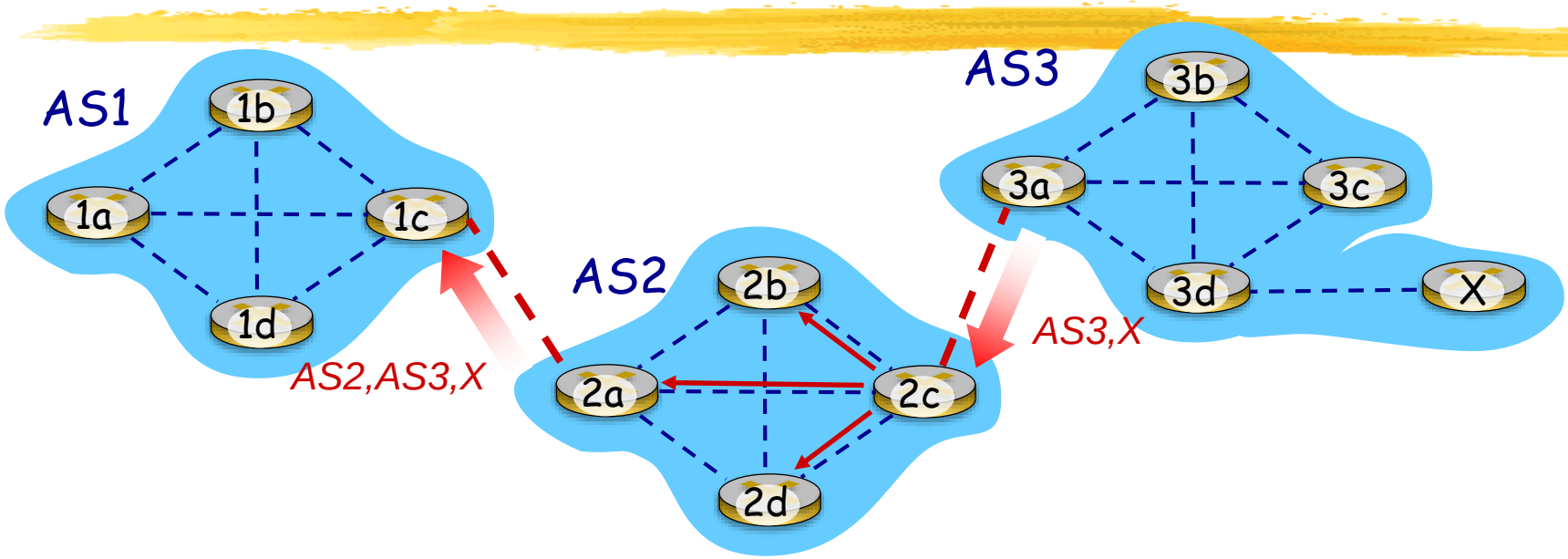
- **BGP session**
  - la connessione TCP che supporta il colloquio tra due BGP speaker



Attributi del percorso e rotte BGP

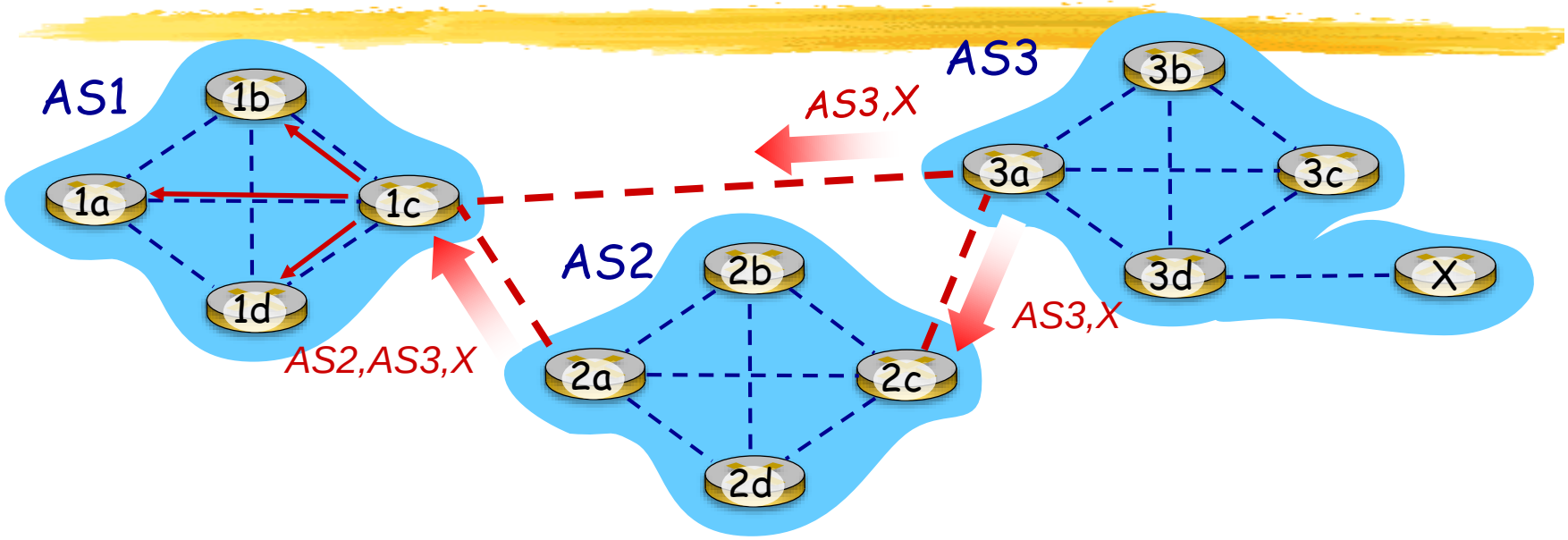
- **Quando un router annuncia un prefisso per una sessione BGP, include anche un certo numero di attributi BGP**
  - prefisso + attributi = "rotta"
- **Due dei più importanti attributi sono:**
  - AS-PATH: elenca i sistemi autonomi attraverso i quali è passato l'annuncio del prefisso
  - NEXT-HOP: quando si deve inoltrare un pacchetto tra due sistemi autonomi, questo potrebbe essere inviato su uno dei vari collegamenti fisici che li connettono direttamente
- **Quando un router gateway riceve un annuncio di rotta, utilizza le proprie politiche d'importazione per decidere se accettare o filtrare la rotta**

BGP path advertisement



- Il router 2c riceve un messaggio di path advertisement AS3,X (protocollo eBGP) dal router 3a
- In base alla policy di AS2, il router 2c accetta il cammino AS3,X, e lo rilancia (protocollo iBGP) a tutti i router di AS2
- In base alla policy di AS2, il router 2a annuncia (protocollo eBGP) il cammino AS2, AS3, X al router 1c

BGP path advertisement



- I gateway router possono apprendere l'esistenza di path multipli verso una destinazione
  - Il router 1c acquisisce il path AS2,AS3,X dal router 2a
  - Il router 1c acquisisce il path AS3,X dal router 3a
  - In base alla policy di AS1, il gateway router 1c sceglie il path AS3,X e annuncia il path all'interno di AS1 (protocollo iBGP)

Politiche d'instradamento BGP

- A, B, C sono reti di provider
- W, Y sono reti d'utente
- X è una rete stub dual-homed (interconnessa a due reti)
  - X non vuole supportare il traffico di transito da B a C
  - X non annuncerà a B la rotta verso C

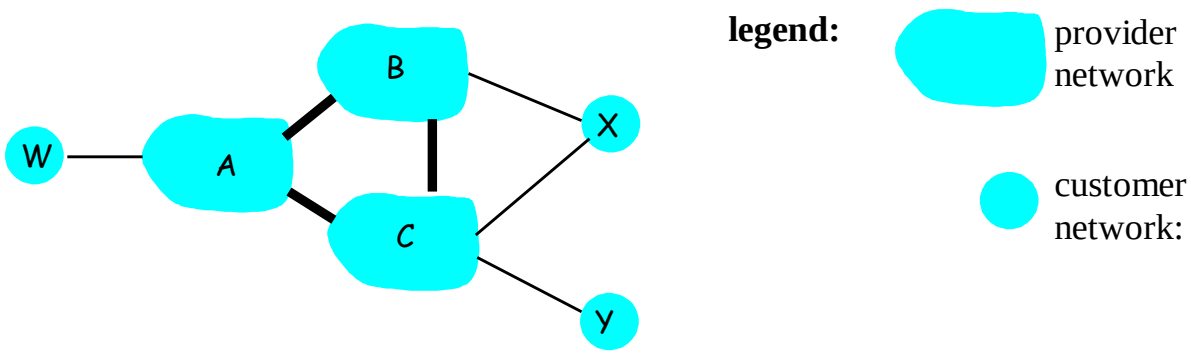


Figure 4.5-BGPnew: a simple BGP scenario

Politiche d'instradamento BGP

- **A annuncia a B e C del percorso AW**
- **B annuncia a X del percorso BAW**
- **B non annuncia a C del percorso BAW se:**
  - B non ha nessun "interesse commerciale" nella rotta CBAW poiché nessuna tra le reti A, C e W è cliente di B
  - B vuole costringere C ad instradare verso W attraverso A
  - B vuole instradare solo da/verso i suoi clienti

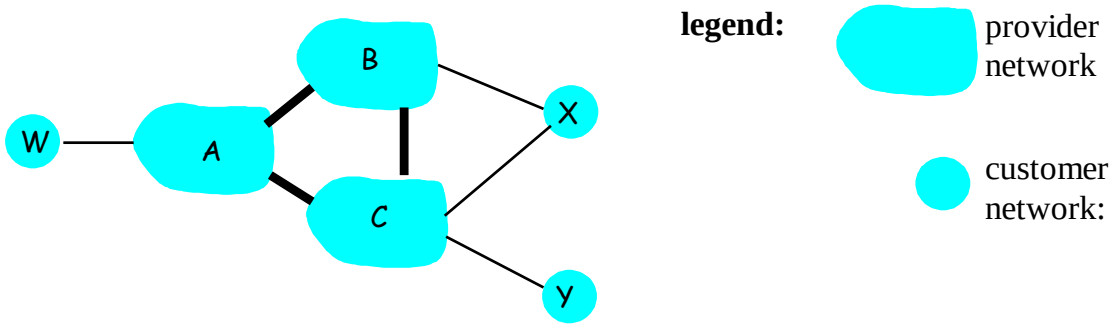
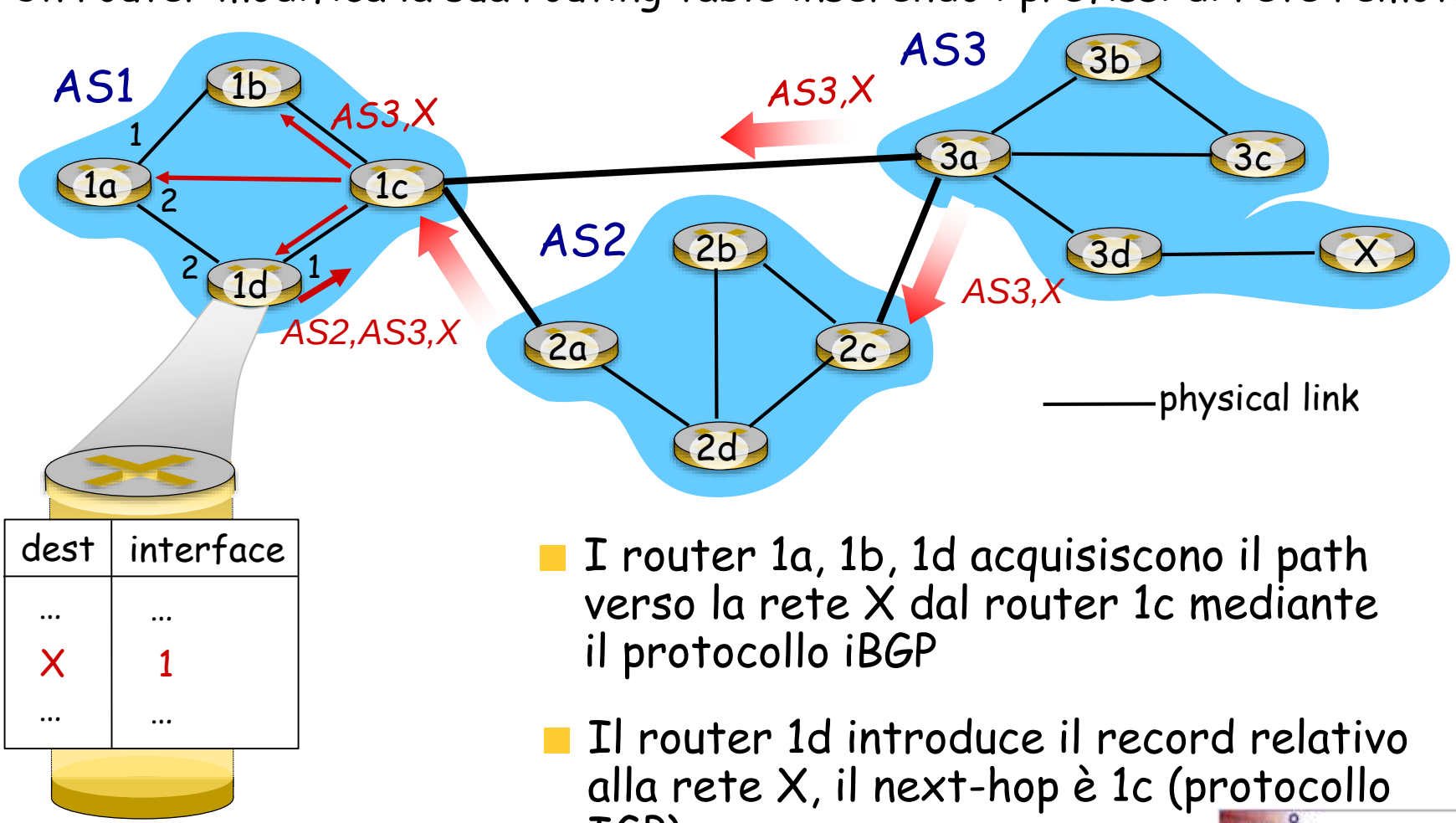


Figure 4.5-BGPnew: a simple BGP scenario

Routing table BGP e OSPF (1)

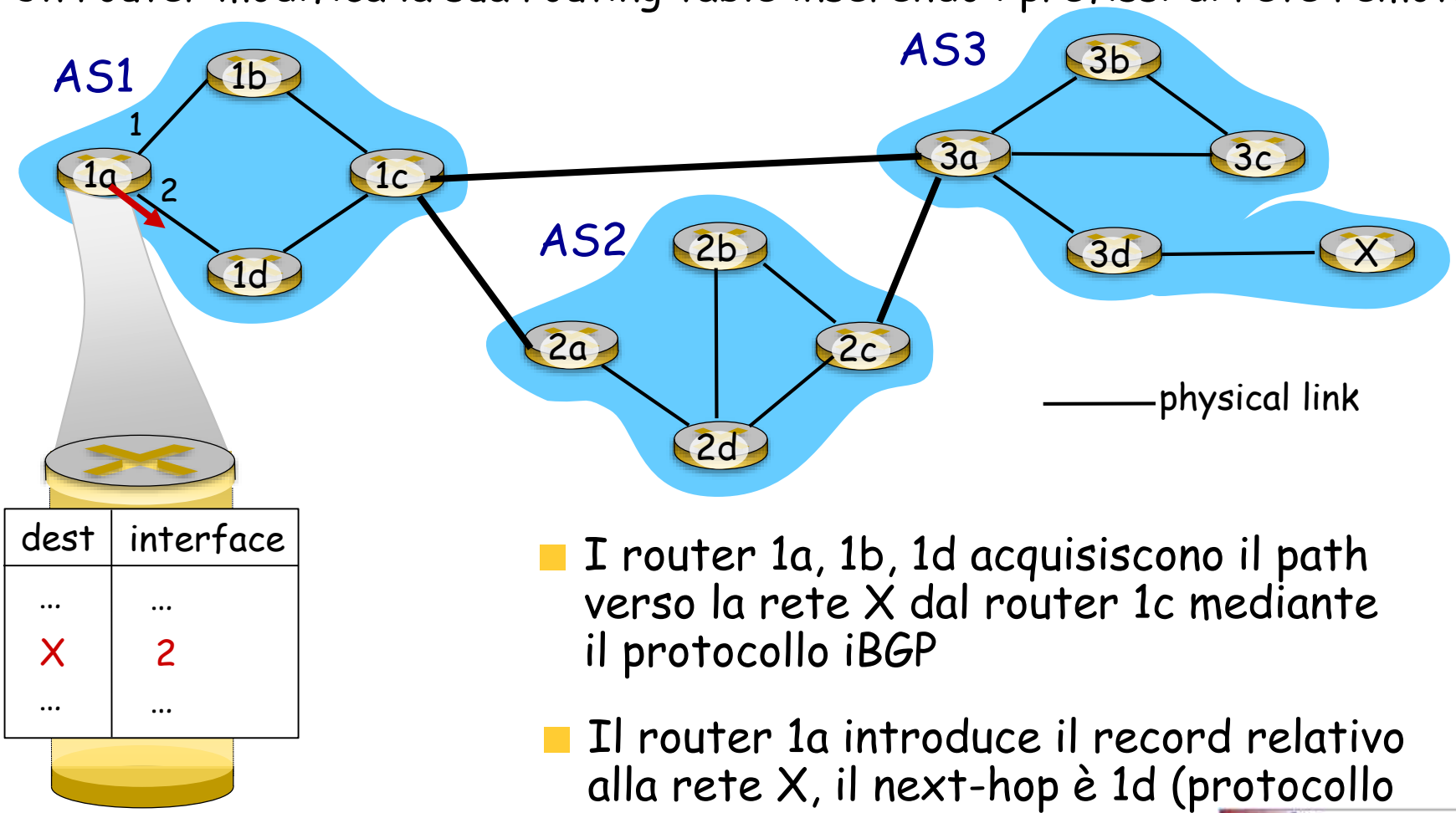
- Un router modifica la sua routing table inserendo i prefissi di rete remoti



- I router 1a, 1b, 1d acquisiscono il path verso la rete X dal router 1c mediante il protocollo iBGP
- Il router 1d introduce il record relativo alla rete X, il next-hop è 1c (protocollo IGP)

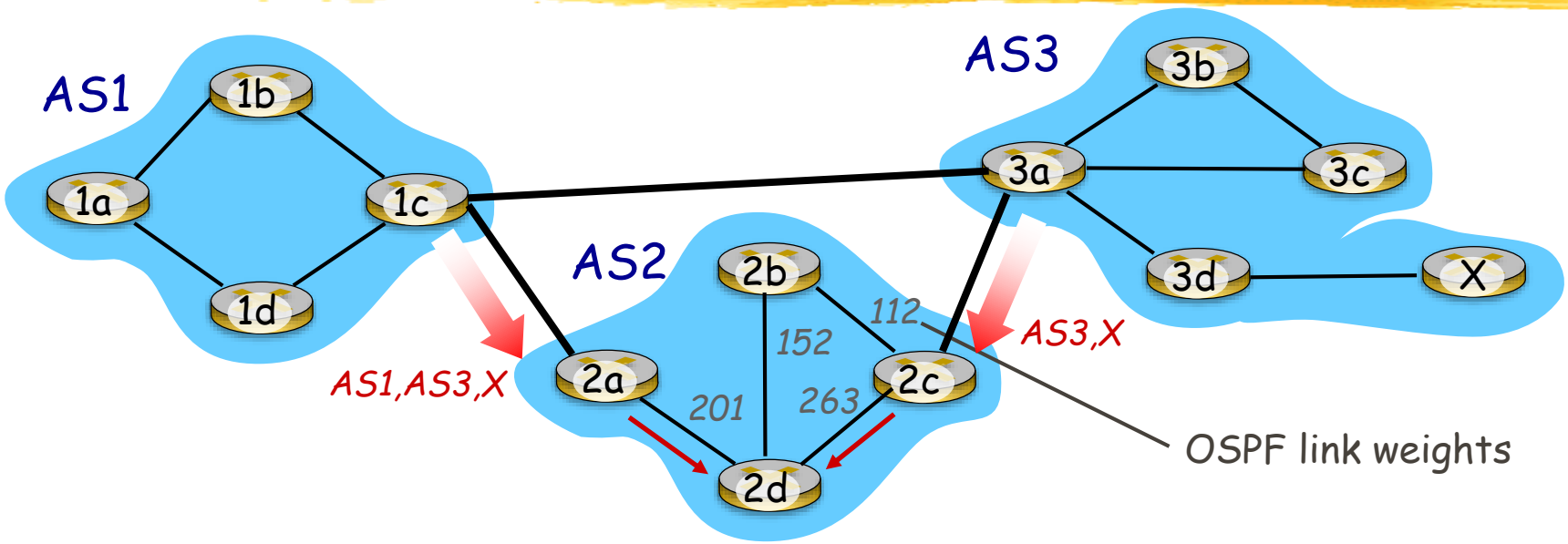
Routing table BGP e OSPF (2)

- Un router modifica la sua routing table inserendo i prefissi di rete remoti



- I router 1a, 1b, 1d acquisiscono il path verso la rete X dal router 1c mediante il protocollo iBGP
- Il router 1a introduce il record relativo alla rete X, il next-hop è 1d (protocollo IGP)

Routing Hot Potato



- Il router 2d apprende (via iBGP) i routing alternativi verso X: router 2a o 2c
- *Hot potato routing*
  - Si sceglie il router gateway (es. il router 2a) che viene raggiunto dal cammino con costo minore, indipendentemente dal costo dei cammini inter-AS

Terminologia BGP

- **AS number**
  - identificatore a 16-bit che identifica univocamente un AS
- **AS path**
  - è la lista di AS che sono attraversati in un cammino
- **Politiche di routing**
  - nel protocollo BGP non sono definite regole fisse per la scelta dei cammini inter-AS, ma le regole sono definite dal gestore di ogni AS
    - un AS multi-homed può rifiutare di operare come AS di transito
    - un AS multi-homed può operare come AS di transito solo per alcuni AS
    - un AS può scegliere a quale altro AS affidare il traffico di transito
  - Tra le possibili scelte un BGP speaker sceglie quella da preferire in base alla politica di routing fissata dal gestore
  - In caso di cammini alternativi, un BGP speaker li mantiene tutti ma ne comunica uno solo agli altri AS

Selezione dei percorsi BGP

- **Un router può ricavare più di una rotta verso un determinato prefisso, e deve quindi sceglierne una**
- **Regole di eliminazione**
  - Alle rotte viene assegnato come attributo un valore di preferenza locale. Si selezionano quindi le rotte con i più alti valori di preferenza locale
  - Si seleziona la rotta con valore AS-PATH più breve
  - Si seleziona quella il cui router di NEXT-HOP è più vicino: instradamento "hot potato"
  - Se rimane ancora più di una rotta, il router si basa sugli identificatori BGP

Messaggi BGP

- I messaggi BGP vengono scambiati attraverso connessioni TCP
- Messaggi BGP
  - OPEN
    - apre la connessione TCP tra router BGP e autentica il mittente
  - UPDATE
    - annuncia un nuovo percorso (o cancella quello vecchio)
  - KEEPALIVE
    - mantiene la connessione attiva in mancanza di messaggi UPDATE
  - NOTIFICATION
    - riporta gli errori in un precedente messaggio
    - usato anche per chiudere il collegamento

Protocolli inter-AS vs. protocolli intra-AS

Politiche

- Inter-AS: il controllo amministrativo desidera avere il controllo su come il traffico viene instradato e su chi instrada attraverso le sue reti.
- Intra-AS: unico controllo amministrativo, e di conseguenza le questioni di politica hanno un ruolo molto meno importante nello scegliere le rotte interne al sistema

Scala

- L'instradamento gerarchico fa "risparmiare" sulle tabelle d'instradamento, e riduce il traffico dovuto al loro aggiornamento

Prestazioni

- Intra-AS: orientato alle prestazioni
- Inter-AS: le politiche possono prevalere sulle prestazioni