# Livello di Rete: Algoritmo PV, Protocollo BGP multicast

Gaia Maselli
maselli@di.uniroma1.it

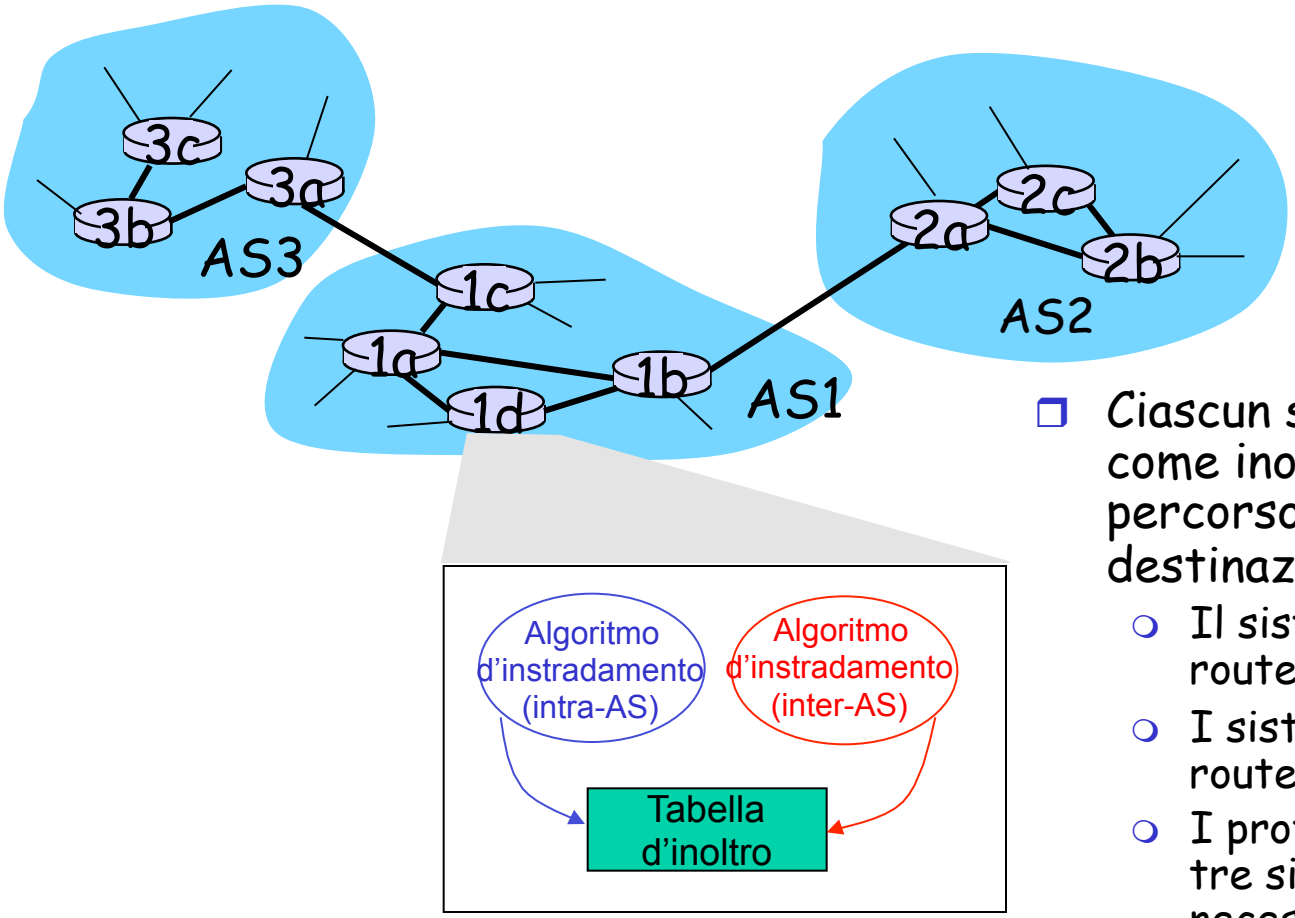
Parte di queste slide sono state prese dal materiale associato ai libri:

1) B.A. Forouzan, F. Mosharraf – Reti di calcolatori. Un approccio top-down. Copyright © 2013 McGraw-Hill Education Italy srl. Edizione italiana delle slide a cura di Gabriele D'Angelo e Gaia Maselli
2) Computer Networking: A Top Down Approach , 6th edition. All material copyright 1996-2009 J.F Kurose and K.W. Ross, All Rights Reserved

# Sistemi autonomi

- Ogni ISP è un sistema autonomo
- Ad ogni AS viene assegnato un numero identificativo univoco di 16 bit (autonomous number – ASN) dall'ICANN
- Gli AS possono avere diverse dimensioni
- Gli AS sono classificati in base al modo in cui sono connessi ad altri AS
  - *AS stub*: ha un solo collegamente verso un altro AS. Il traffico è generato o destinato allo stub ma non transita attraverso di esso (es. grande azienda)
  - *AS multihomed*: ha più di una connessione con altri AS ma non consente transito di traffico (azienda che usa servizi di più di un network provider ma non fornisce connettività agli altri AS)
  - *AS di transito*: è collegato a più AS e consente il traffico (network provider e dorsali)

# Sistemi autonomi interconnessi




- Ciascun sistema autonomo sa come inoltrare pacchetti lungo il percorso ottimo verso qualsiasi destinazione interna al gruppo
  - Il sistema AS1 ha quattro router
  - I sistemi AS2 e AS3 hanno tre router ciascuno
  - I protocolli d'instradamento dei tre sistemi autonomi non sono necessariamente gli stessi
  - I router 1b, 1c, 2a e 3a sono gateway

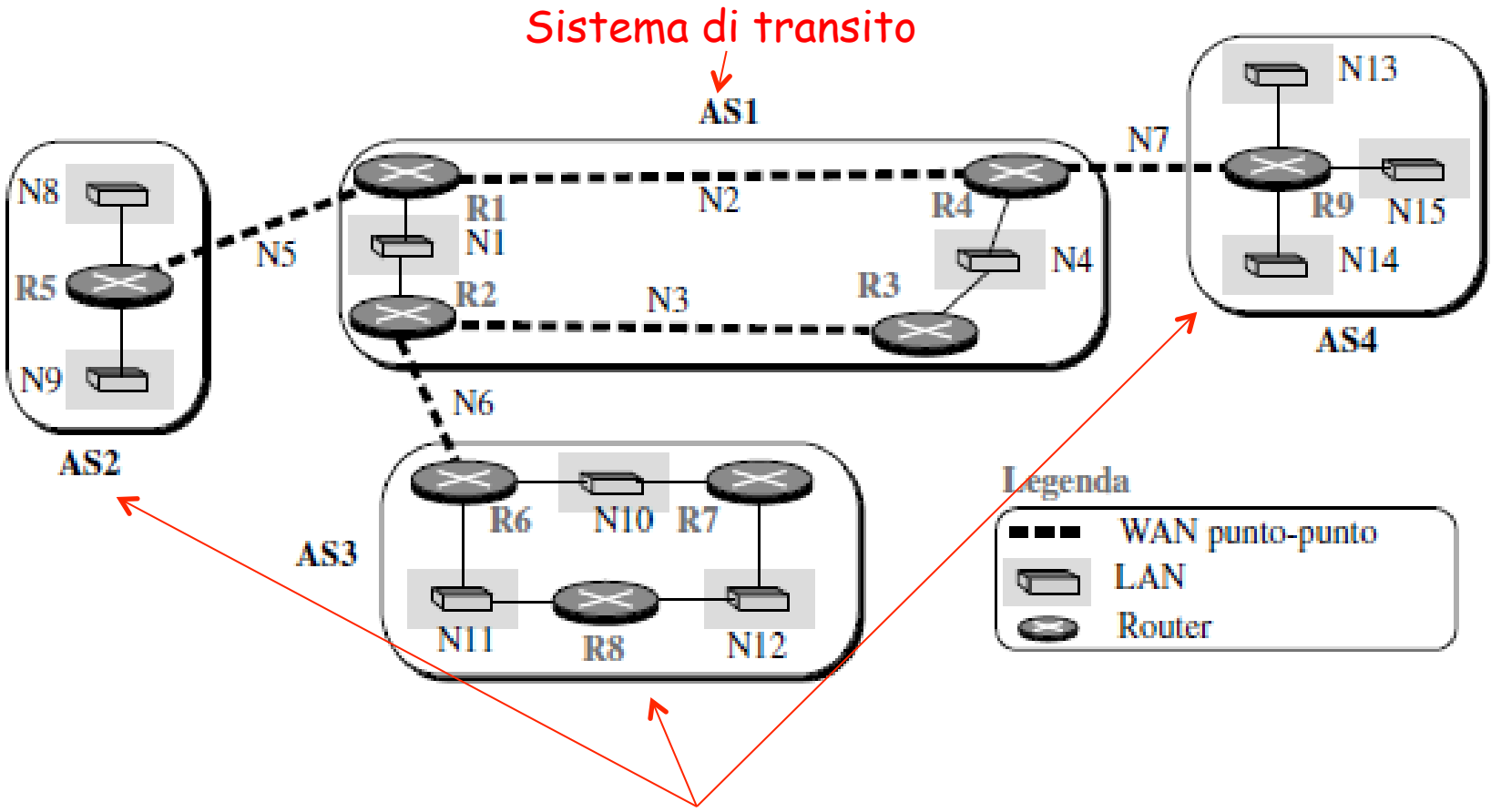
# Routing intra-dominio

- RIP: Routing Information Protocol
- OSPF: Open Shortest Path First

# Routing inter-dominio

- BGP: Border Gateway Protocol

# Instradamento inter-AS




Stub AS – lo scambio di dati tra questi AS passa attraverso AS1

Ogni router all'interno degli AS sa come raggiungere tutte le reti che si trovano nel suo AS ma non sa come raggiungere una rete che si trova in un altro AS

# Border Gateway Protocol

- RIP e OSPF vengono utilizzati per determinare i percorsi ottimali per le coppie origine-destinazione interne a un sistema autonomo
- BGP (Border Gateway Protocol) – proprietà di CISCO
  - Usato per determinare percorsi per le coppie origine-destinazione che interessano più sistemi autonomi
  - rappresenta l'attuale standard *de facto.*
- BGP è un protocollo path vector (distance vector con percorsi)
- BGP mette a disposizione di ciascun AS un modo per:
  1. ottenere informazioni sulla raggiungibilità delle sottoreti da parte di AS confinanti
  2. propagare le informazioni di raggiungibilità a tutti i router interni di un AS
  3. 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.
- In BGP le destinazioni non sono host ma prefissi CIDR che rappresentano una sottorete o una collezione di sottoreti

# Path-vector routing

- Sia LS che DS si basano sul costo minimo
- Tuttavia ci sono casi in cui il costo minimo non è l'obiettivo prioritario

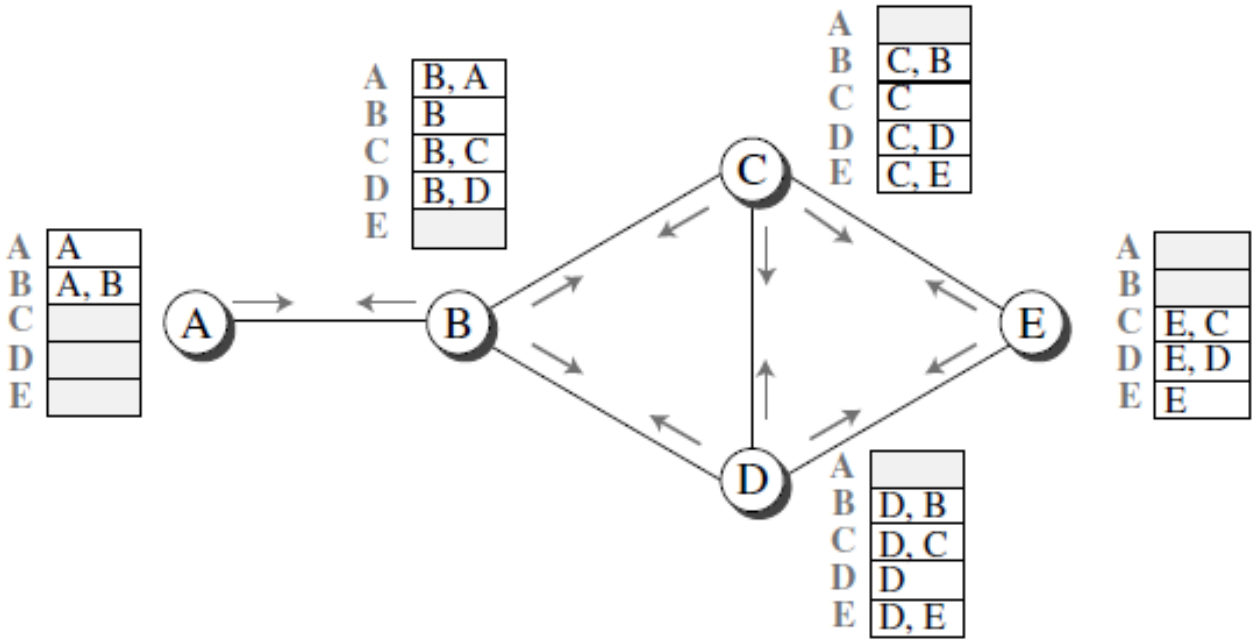
Esempio: un mittente non vuole che i suoi pacchetti passino attraverso determinati router

- Il routing a costo minimo non consente di applicare questo tipo di politiche nella scelta del percorso
- Path-vector routing (routing a vettore di percorso): la sorgente può controllare il percorso
  - Politiche: minimizzare il numero di hop
  - Evitare alcuni nodi

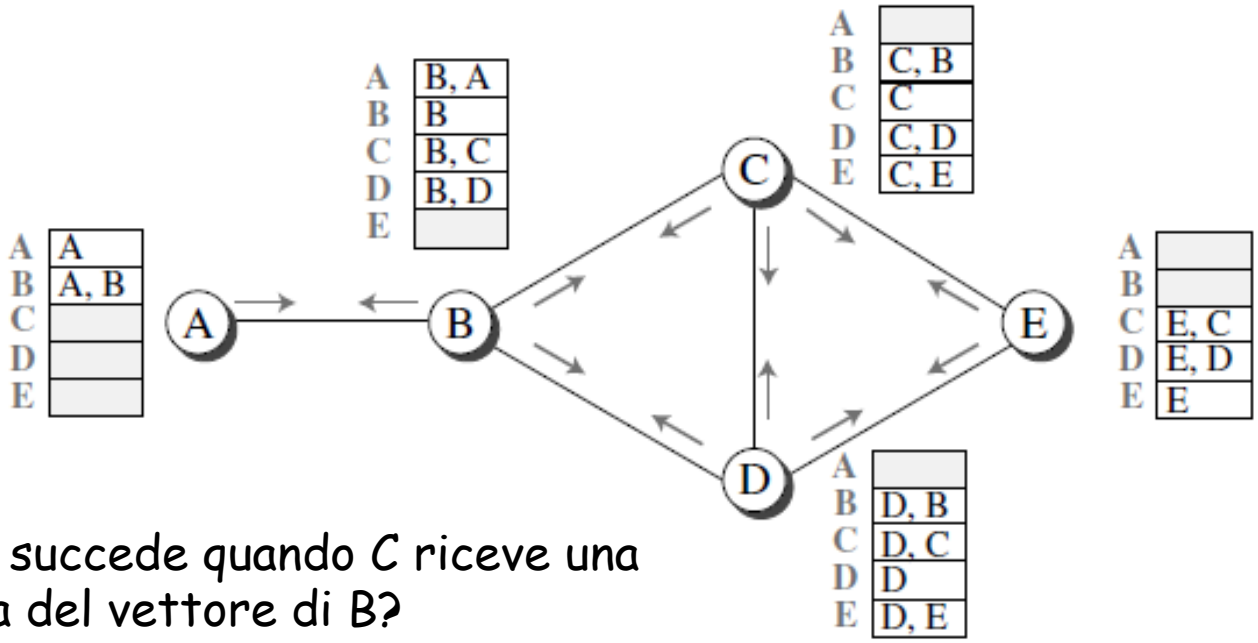
# Path vector routing

- Simile a distance vector ma vengono inviati percorsi invece che solo destinazioni
- Ogni nodo quando riceve un path vector da un vicino, aggiorna il suo path vector applicando la sua politica invece del costo minimo

Inizializzazione



# Aggiornamento dei path vector




Cosa succede quando C riceve una Copia del vettore di B?

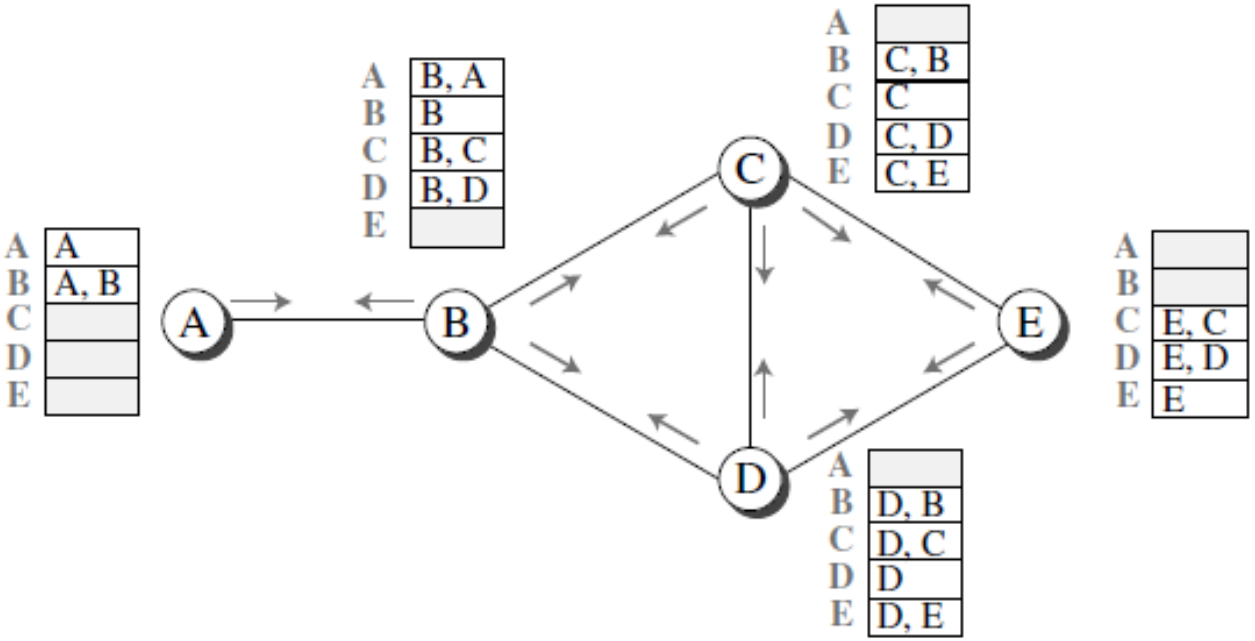
| | Nuovo C | Vecchio C | B |
|---|---|---|---|
| A | C, B, A | | B, A |
| B | C, B | C, B | B |
| C | C | C | B, C |
| D | C, D | C, D | B, D |
| E | C, E | C, E | |

C[ ] = migliore (C[ ], C + B[ ])

Evento 1: C riceve una copia del vettore di B

**Nota:**
X [ ]: vettore X
Y: nodo Y

# Aggiornamento dei path vector




Nessun cambiamento

| | Nuovo C | Vecchio C | D |
|---|---|---|---|
| A | C, B, A | C, B, A | |
| B | C, B | C, B | D, B |
| C | C | C | D, C |
| D | C, D | C, D | D |
| E | C, E | C, E | D, E |

C[ ] = migliore (C[ ], C + D[ ])

Evento 2: C riceve una copia del vettore di D

# Algoritmo path-vector

```
1  Path_Vector_Routing ( )
2  {
3     // Inizializzazione
4     for (y = 1 to N)
5     {
6        if (y è me_stesso)
7           Path[y] = me_stesso
8        else if (y è un vicino)
9           Path[y] = me_stesso + il_nodo_vicino
10       else
11          Path[y] = vuoto
12    }
13    Spedisci il vettore {Path[1], Path[2], …, Path[y]} a tutti i vicini
14    // Aggiornamento
15    repeat (sempre)
16    {
17       wait (un vettore Path_w da un vicino w)
18       for (y = 1 to N)
19       {
20          if (Path_w comprende me_stesso)
21             scarta il percorso                    // Evita ogni ciclo
22          else
23             Path[y] = il_migliore_tra {Path[y], (me_stesso + Path_w[y])}
24       }
25       If (c'è un cambiamento nel vettore)
26          Spedisci il vettore {Path[1], Path[2], …, Path[y]} a tutti i vicini
27    }
28 }  // Fine del path-vector
```

# eBGP e iBGP

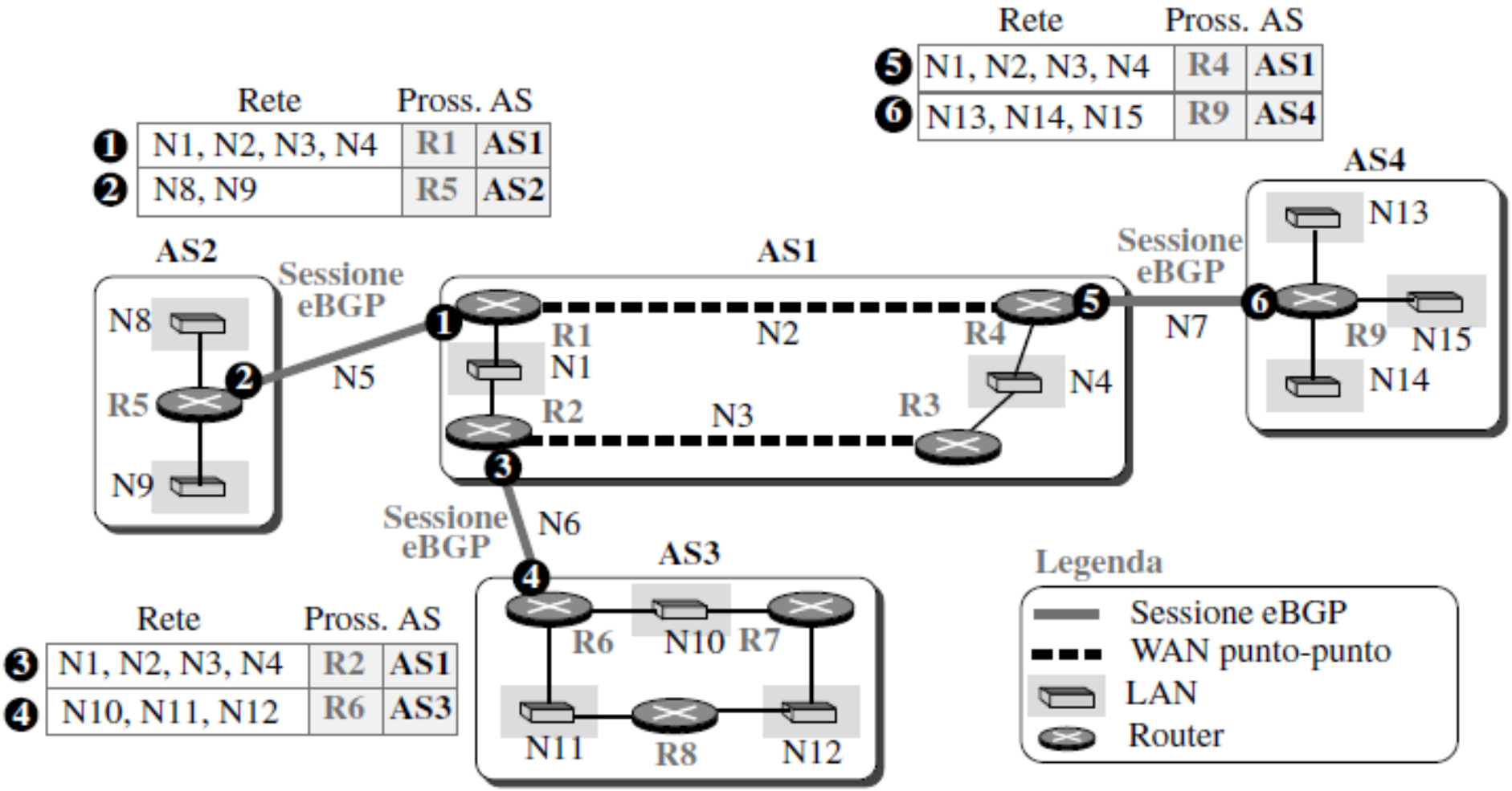
- Per permettere ad ogni router di instradare correttamente i pacchetti, qualsiasi sia la destinazione, è necessario installare su tutti i router di confine (border router) dell'AS una variante del BGP chiamata *BGP esterno* (*external BGP* o *eBGP*)
- Tutti i router (non solo quelli di confine) dovranno invece usare la seconda variante del BGP, chiamata *BGP interno* (*internal BGP* o *iBGP*)
- I router di confine devono eseguire tre protocolli di routing (intra-dominio, eBGP, iBGP) e tutti gli altri router ne eseguono due (intra-dominio e iBGP)

# Fondamenti di BGP

- Coppie di router si scambiano informazioni di instradamento su connessioni TCP usando la porta 179
- I router ai capi di una connessione TCP sono chiamati peer BGP, e la connessione TCP con tutti i messaggi BGP che vi vengono inviati è detta sessione BGP.
- Notiamo che  le linee di sessione BGP non sempre corrispondono ai collegamenti fisici.

# eBGP

- Due router di confine che si trovano in due diversi AS formano una coppia di peer BGP e si scambiano messaggi



# eBGP

- I messaggi scambiati durante le sessioni eBGP servono per indicare ad alcuni router come instradare i pacchetti destinati ad alcune reti, ma le informazioni di raggiungibilità non sono complete
- Problemi da risolvere

1. I router di confine sanno instradare pacchetti solo ad AS vicini
2. Nessuno dei router non di confine (interno agli AS) sa come instradare un pacchetto destinato alle reti che si trovano in altri AS

- Soluzione: iBGP

# iBGP

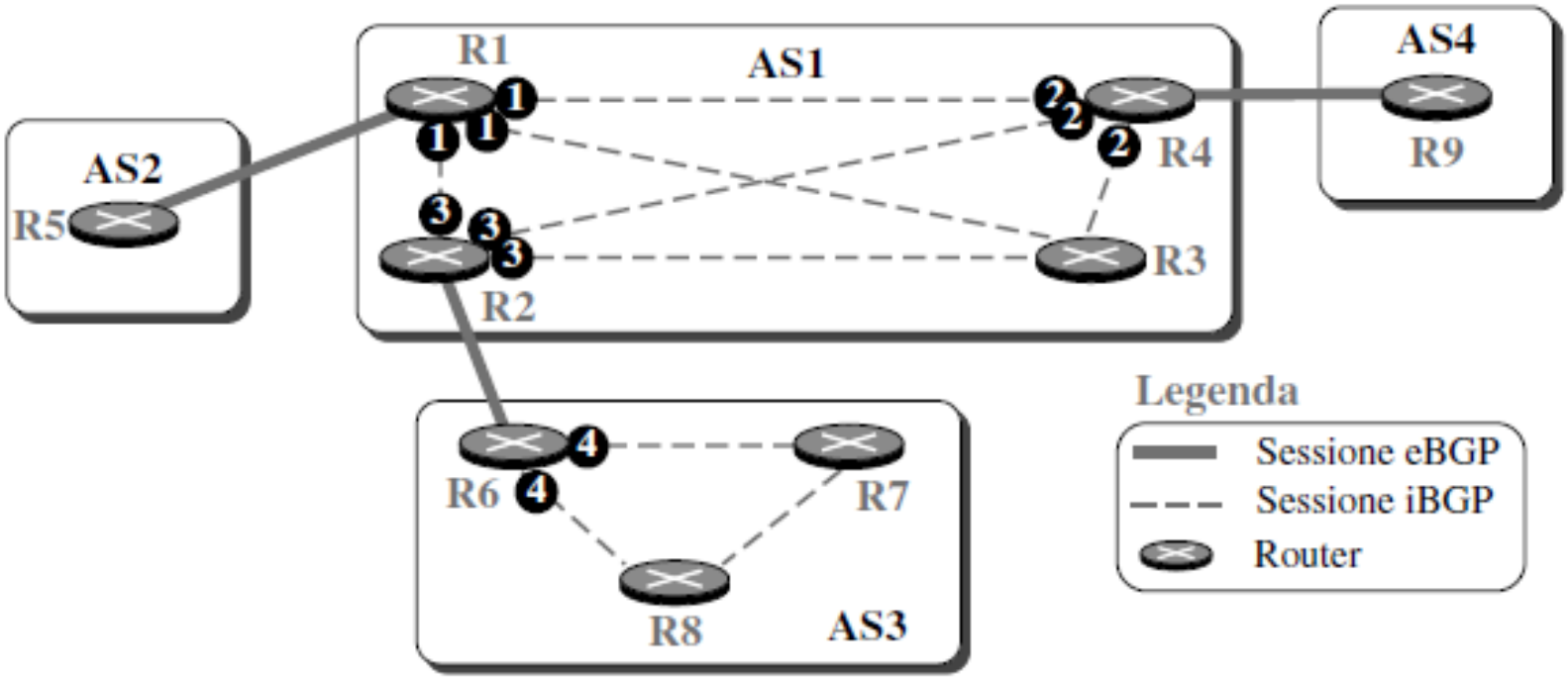
- Crea una sessione tra ogni possibile coppia di router all'interno di un AS
- Non tutti i nodi hanno messaggi da inviare (es. R3, R7, R8), ma tutti ricevono

| | Rete | Pross. | AS |
|---|---|---|---|
| 1 | N8, N9 | R1 | AS1, AS2 |

| | Rete | Pross. | AS |
|---|---|---|---|
| 2 | N13, N14, N15 | R4 | AS1, AS4 |

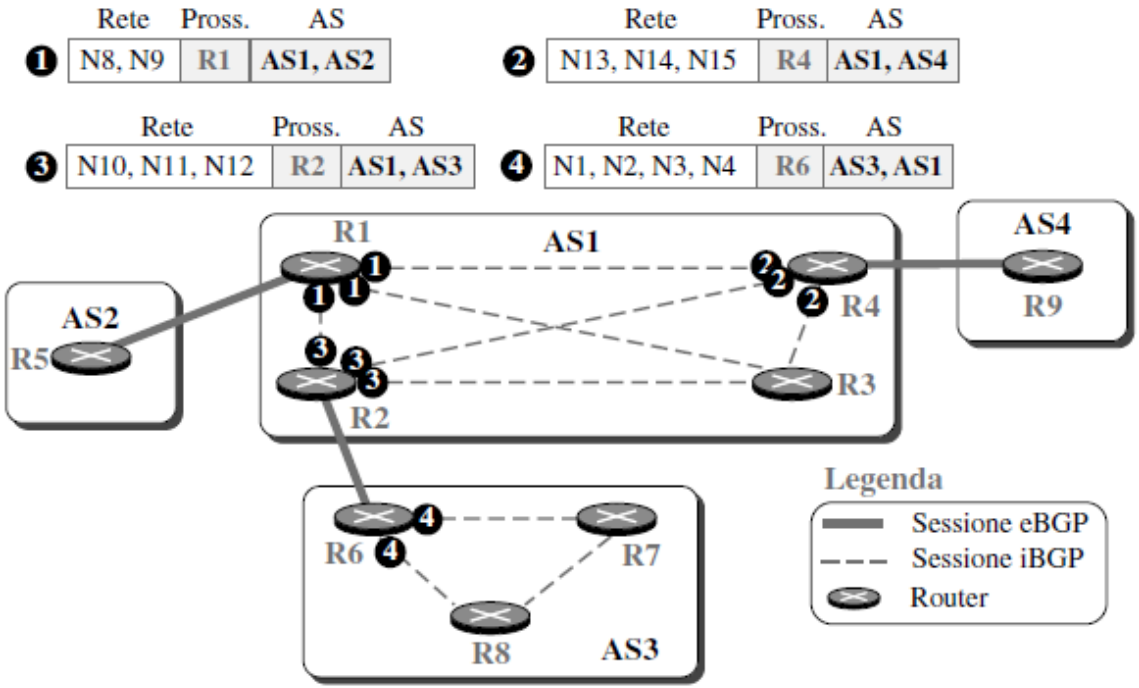
| | Rete | Pross. | AS |
|---|---|---|---|
| 3 | N10, N11, N12 | R2 | AS1, AS3 |

| | Rete | Pross. | AS |
|---|---|---|---|
| 4 | N1, N2, N3, N4 | R6 | AS3, AS1 |



# Scambio di messaggi

- Il processo di aggiornamento non termina dopo il primo scambio di messaggi
- Esempio: R1 dopo che ha ricevuto il messaggio di aggiornamento di R2, combina le informazioni circa la raggiungibilità di AS3 con quelle che già conosceva relativamente a AS1 e invia un nuovo messaggio d'aggiornamento a R5 (che quindi sa come raggiungere AS1 e AS3)



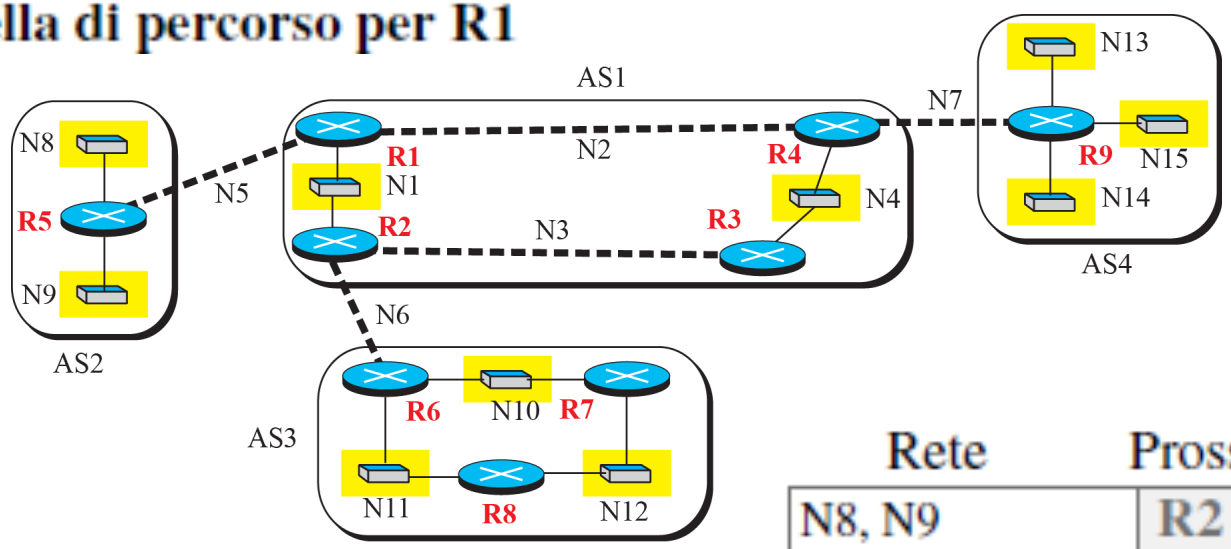
# Scambio di messaggi

- Il processo di aggiornamento continua finché non ci sono più aggiornamenti
- Le informazioni ottenute da eBGP e iBGP vengono combinate per creare le tabelle dei percorsi

# Tabelle di percorso

| Rete | Pross. | Percorso |
|---|---|---|
| N8, N9 | R5 | AS1, AS2 |
| N10, N11, N12 | R2 | AS1, AS3 |
| N13, N14, N15 | R4 | AS1, AS4 |

**Tabella di percorso per R1**




| Rete | Pross. | Percorso |
|---|---|---|
| N8, N9 | R1 | AS1, AS2 |
| N10, N11, N12 | R6 | AS1, AS3 |
| N13, N14, N15 | R1 | AS1, AS4 |

**Tabella di percorso per R2**

| Rete | Pross. | Percorso |
|---|---|---|
| N8, N9 | R2 | AS1, AS2 |
| N10, N11, N12 | R2 | AS1, AS3 |
| N13, N14, N15 | R4 | AS1, AS4 |

**Tabella di percorso per R3**

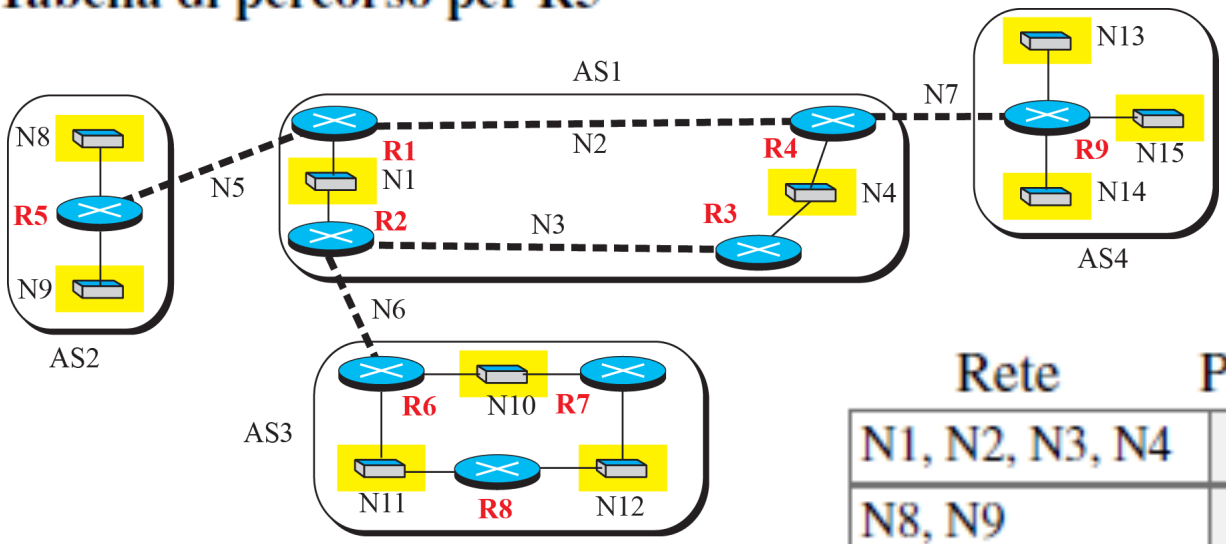
# Tabelle di percorso

| Rete | Pross. | Percorso |
|---|---|---|
| N1, N2, N3, N4 | R1 | AS2, AS1 |
| N10, N11, N12 | R1 | AS2, AS1, AS3 |
| N13, N14, N15 | R1 | AS2, AS1, AS4 |

**Tabella di percorso per R5**

| Rete | Pross. | Percorso |
|---|---|---|
| N8, N9 | R1 | AS1, AS2 |
| N10, N11, N12 | R1 | AS1, AS3 |
| N13, N14, N15 | R9 | AS1, AS4 |

**Tabella di percorso per R4**




| Rete | Pross. | Percorso |
|---|---|---|
| N1, N2, N3, N4 | R2 | AS3, AS1 |
| N8, N9 | R2 | AS3, AS1, AS2 |
| N13, N14, N15 | R2 | AS3, AS1, AS4 |

**Tabella di percorso per R6**

# Tabelle di routing

- Le tabelle di percorso ottenute da BGP non vengono usate di per sé per l'instradamento dei pacchetti bensì inserite nelle tabelle di routing intra-dominio (generate da RIP o OSPF)
- Nel caso di stub, l'unico router di confine dell'area aggiunge una regola di default alla fine della sua tabella di routing e definisce come prossimo router quello che si trova dall'altro lato della connessione eBGP
- Nel caso di AS di transito, il contenuto della tabella di percorso deve essere inserito nella tabella di routing ma bisogna impostare il costo (RIP e OSPF usano metriche differenti)
  - Si imposta il costo pari a quello per raggiungere il primo AS nel percorso

# Tabelle d'inoltro dopo l'aggiunta delle informazioni BGP

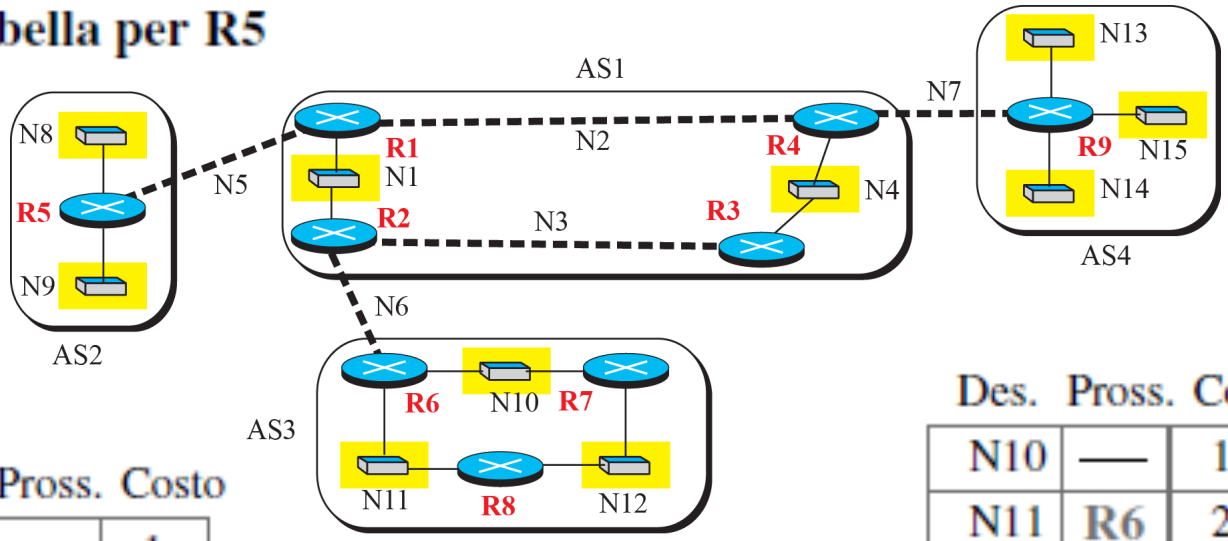
| Des. | Pross. | Costo |
|---|---|---|
| N8 | — | 1 |
| N9 | — | 1 |
| **0** | **R1** | 1 |

**Tabella per R5**

Nel caso di stub, l'unico router di confine dell'area aggiunge una regola di default alla fine della sua tabella di routing e definisce come prossimo router quello che si trova dall'altro lato della connessione eBGP

| Des. | Pross. | Costo |
|---|---|---|
| N13 | — | 1 |
| N14 | — | 1 |
| N15 | — | 1 |
| **0** | **R4** | 1 |

**Tabella per R9**




| Des. | Pross. | Costo |
|---|---|---|
| N10 | — | 1 |
| N11 | — | 1 |
| N12 | R7 | 2 |
| **0** | **R2** | 1 |

**Tabella per R6**

| Des. | Pross. | Costo |
|---|---|---|
| N10 | R6 | 2 |
| N11 | — | 1 |
| N12 | — | 1 |
| **0** | **R6** | 2 |

**Tabella per R8**

| Des. | Pross. | Costo |
|---|---|---|
| N10 | — | 1 |
| N11 | R6 | 2 |
| N12 | — | 1 |
| **0** | **R6** | 2 |

**Tabella per R7**

# Tabelle d'inoltro dopo l'aggiunta delle informazioni BGP

| Des. | Pross. | Costo |
|---|---|---|
| N1 | — | 1 |
| N4 | R4 | 2 |
| N8 | R5 | 1 |
| N9 | R5 | 1 |
| N10 | R2 | 2 |
| N11 | R2 | 2 |
| N12 | R2 | 2 |
| N13 | R4 | 2 |
| N14 | R4 | 2 |
| N15 | R4 | 2 |

**Tabella per R1**

| Des. | Pross. | Costo |
|---|---|---|
| N1 | — | 1 |
| N4 | R3 | 2 |
| N8 | R1 | 2 |
| N9 | R1 | 2 |
| N10 | R6 | 1 |
| N11 | R6 | 1 |
| N12 | R6 | 1 |
| N13 | R3 | 3 |
| N14 | R3 | 3 |
| N15 | R3 | 3 |

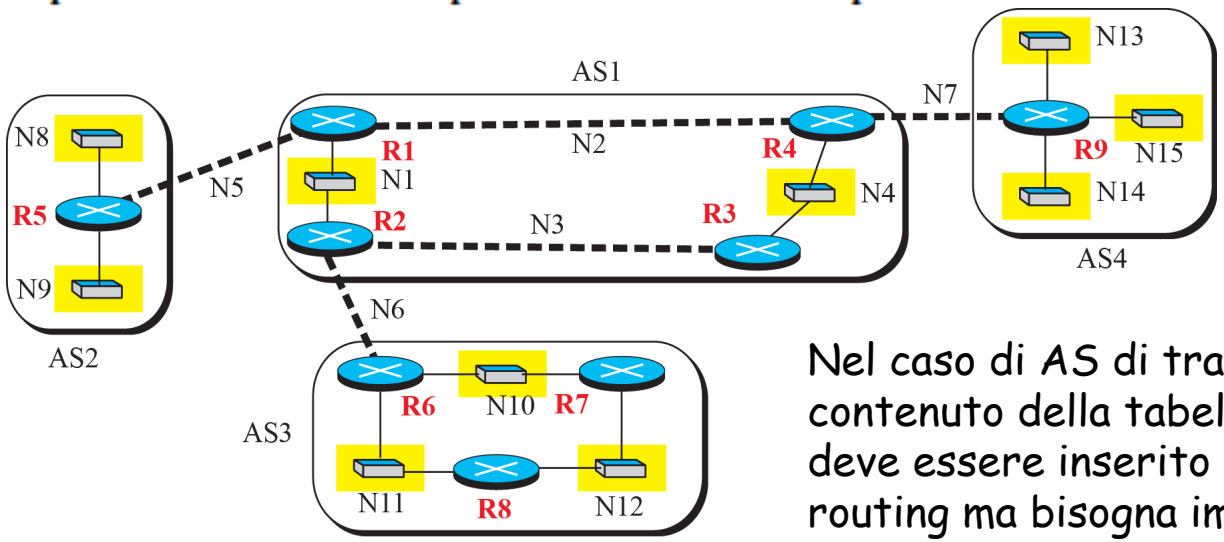
**Tabella per R2**

| Des. | Pross. | Costo |
|---|---|---|
| N1 | R2 | 2 |
| N4 | — | 1 |
| N8 | R2 | 3 |
| N9 | R2 | 3 |
| N10 | R2 | 2 |
| N11 | R2 | 2 |
| N12 | R2 | 2 |
| N13 | R4 | 2 |
| N14 | R4 | 2 |
| N15 | R4 | 2 |

**Tabella per R3**

| Des. | Pross. | Costo |
|---|---|---|
| N1 | R1 | 2 |
| N4 | — | 1 |
| N8 | R1 | 2 |
| N9 | R1 | 2 |
| N10 | R3 | 3 |
| N11 | R3 | 3 |
| N12 | R3 | 3 |
| N13 | R9 | 1 |
| N14 | R9 | 1 |
| N15 | R9 | 1 |

**Tabella per R4**




Nel caso di AS di transito, il contenuto della tabella di percorso deve essere inserito nella tabella di routing ma bisogna impostare il costo

# Attributi del percorso e rotte BGP

- Quando un router annuncia una rotta per un prefisso (di rete) per una sessione BGP, include anche un certo numero di attributi BGP.
  - prefisso + attributi = "rotta"
- Due dei più importanti attributi sono:
  - AS-PATH: serve per selezionare i percorsi. Elenca i sistemi autonomi attraverso i quali è passato l'annuncio del prefisso (e quindi gli hop intermedi della rotta). Ogni sistema autonomo non stub ha un identificativo univoco

    (in questo modo si evitano cicli)
  - NEXT-HOP: indirizzo IP dell'interfaccia su cui viene inviato il pacchetto. (Un router ha più indirizzi IP, uno per ogni interfaccia)
- Quando un router gateway riceve un annuncio di rotta, utilizza le proprie politiche d'importazione per decidere se accettare o filtrare la rotta.
  - Il sistema autonomo può non voler inviare traffico su uno degli AS presenti nel AS-PATH
  - Router conosce rotta migliore

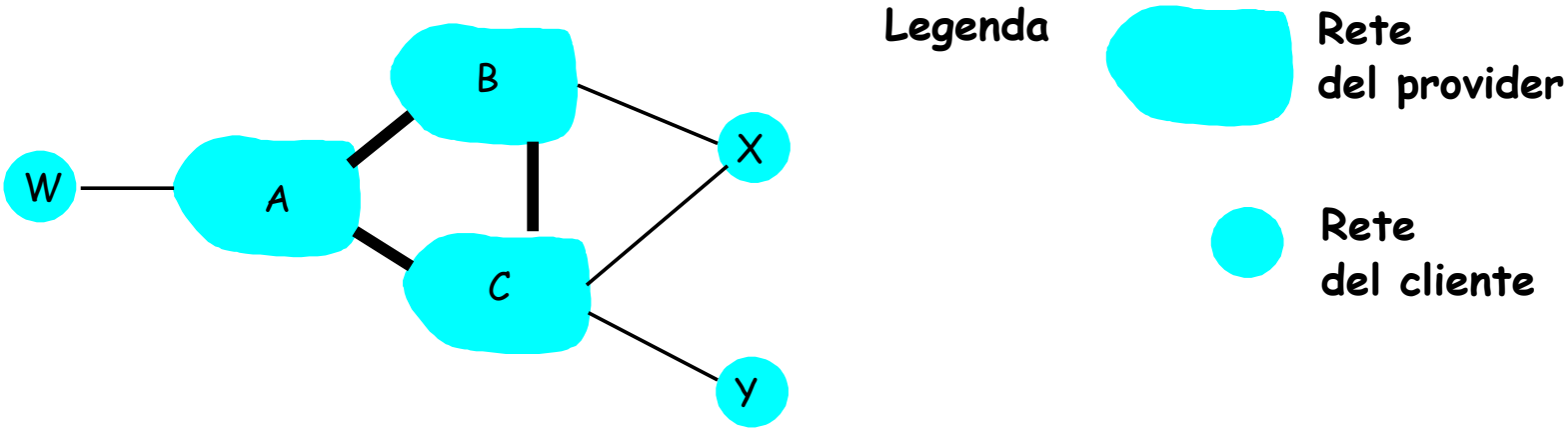
# Selezione dei percorsi BGP

- Un router può ricavare più di una rotta verso una destinazione (percorsi multipli), e deve quindi sceglierne una.
- Regole di eliminazione:
  1. Alle rotte viene assegnato come attributo un valore di preferenza locale. Si selezionano quindi le rotte con i più alti valori di preferenza locale (riflette la politica imposta dall'amministratore)
  2. Si seleziona la rotta con valore AS-PATH più breve.
  3. Si seleziona quella il cui router di NEXT-HOP ha costo minore: hot-potato routing.
  4. Se rimane ancora più di una rotta, il router si basa sugli identificatori BGP.

# Messaggi BGP

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

# Politiche d'instradamento BGP




- A, B, C sono reti di provider di dorsale.
- X, W, Y sono reti stub
- X è una rete stub a più domicili
  - X non vuole che il traffico da B a C le passi attraverso
  - ... e così X non annuncerà a B la rotta verso C

# Perché i protocolli d'instradamento inter-AS sono diversi da quelli 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

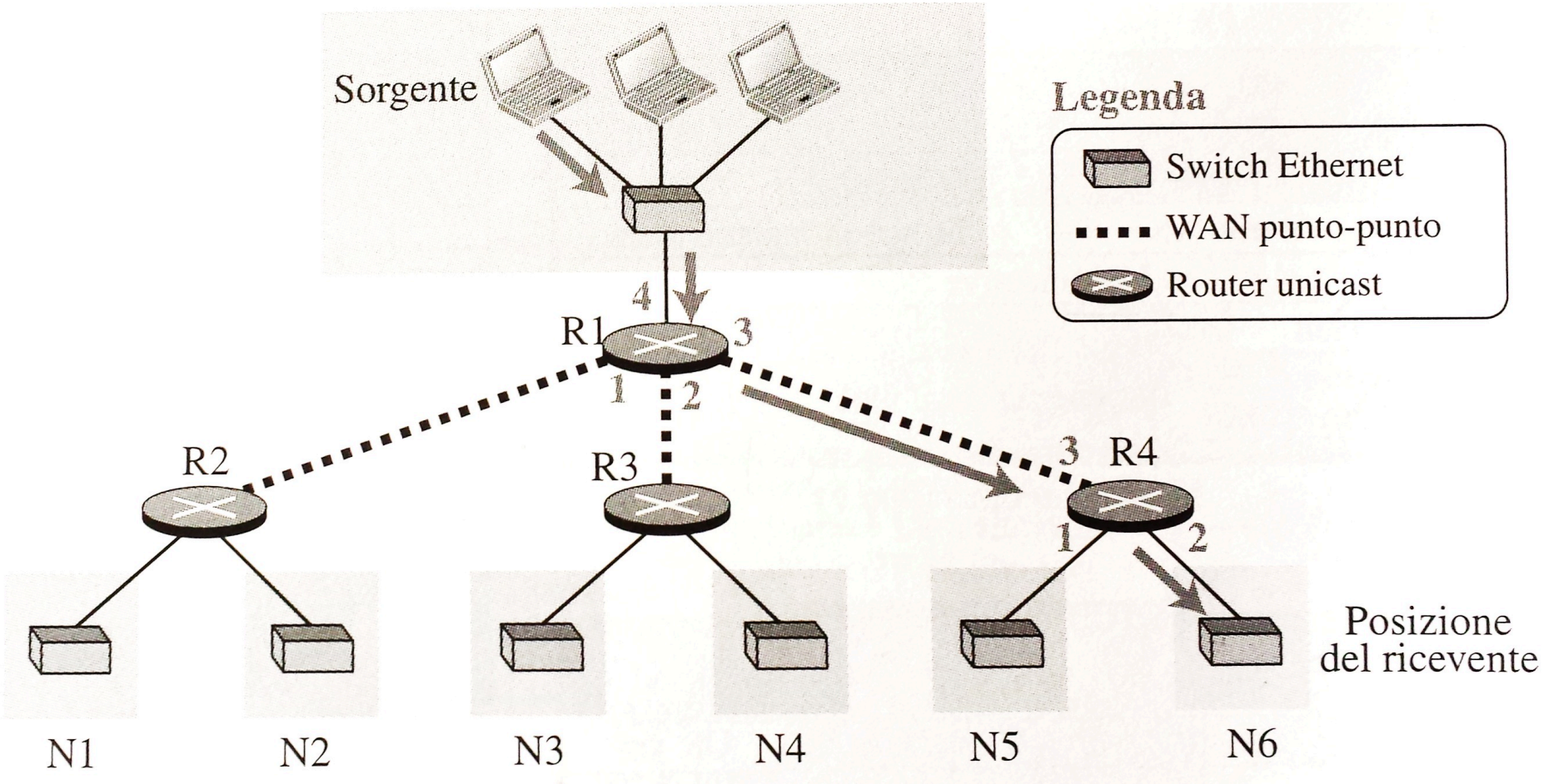
Prestazioni:

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

# Routing multicast

# Unicast

- UNICAST: comunicazione tra una sorgente e una destinazione
  - Indirizzo IP sorgente – indirizzo IP destinazione

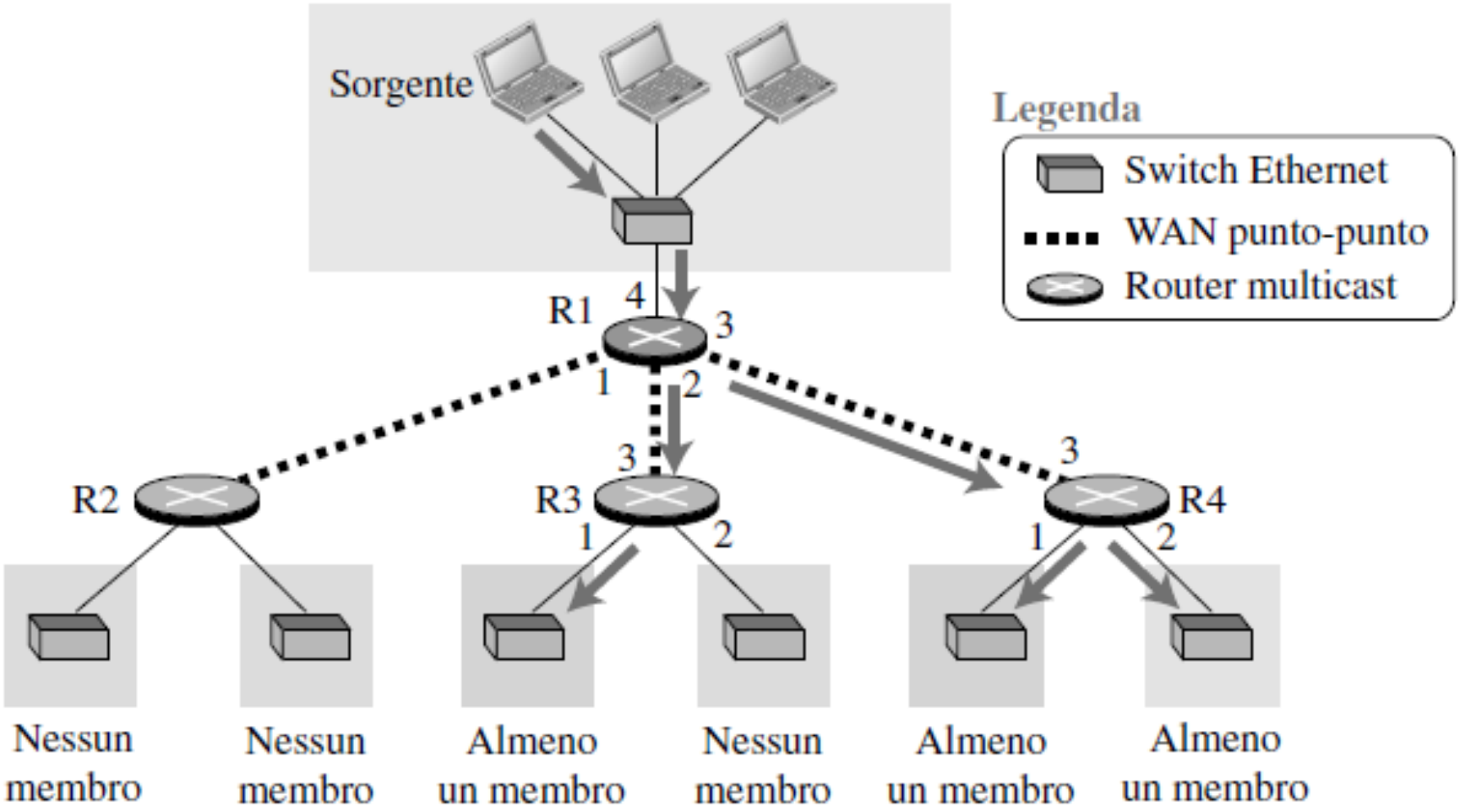


# Broadcast

- BROADCAST: invio di un pacchetto da un nodo sorgente a TUTTI i nodi della rete
  - Comunicazione 1 a N, N: tutti I nodi della rete
  - Indirizzo IP sorgente – indirizzo broadcast di destinazione

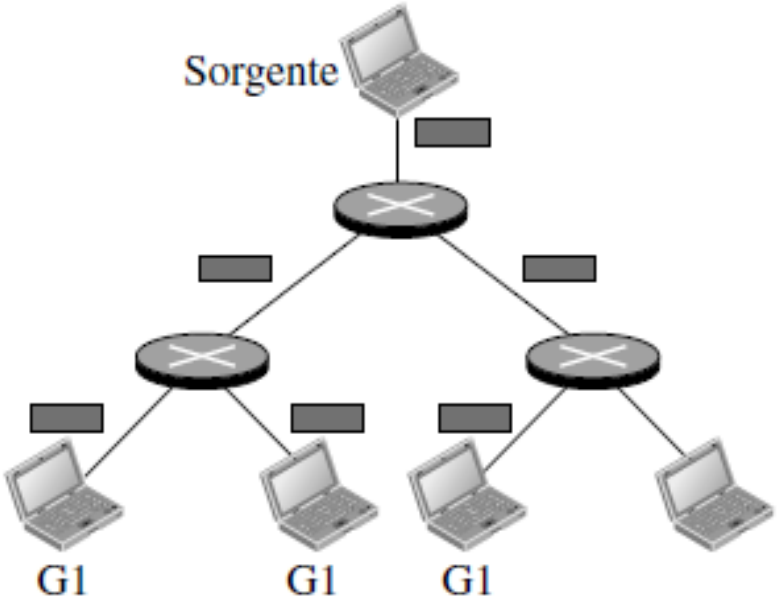
# Multicast

- MULTICAST: comunicazione tra una sorgente e un gruppo di destinazioni



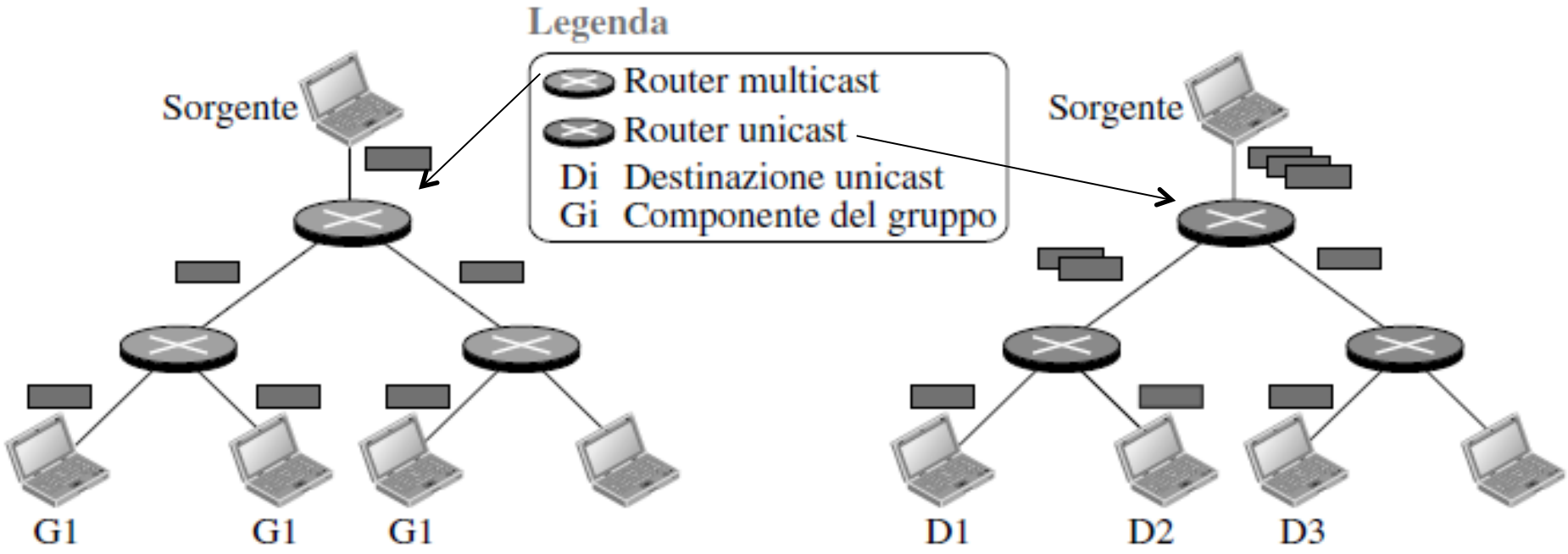
# Multicast

- Il multicast ha inizio con un solo datagramma spedito dalla sorgente che viene poi duplicato dai router
- L'indirizzo di destinazione presente in ogni datagramma è lo stesso per tutti i duplicati




a. Invio in multicast

# Confronto tra multicast e unicast multiplo




a. Invio in multicast

b. Invio di unicast multipli

- Un solo datagramma alla sorgente

Inefficiente e aggiunge ritardi

# Instradamento multicast

- Molte applicazioni richiedono il trasferimento di pacchetti da uno o più mittenti ad un gruppo di destinatari
  - trasferimento di un aggiornamento SW su un gruppo di macchine
  - streaming (audio/video) ad un gruppo di utenti o studenti
  - applicazioni con dati condivisi (lavagna elettronica condivisa da più utenti)
  - aggiornamento di dati (andamento di borsa)
  - giochi multi-player interattivi

# Problema dell'indirizzamento

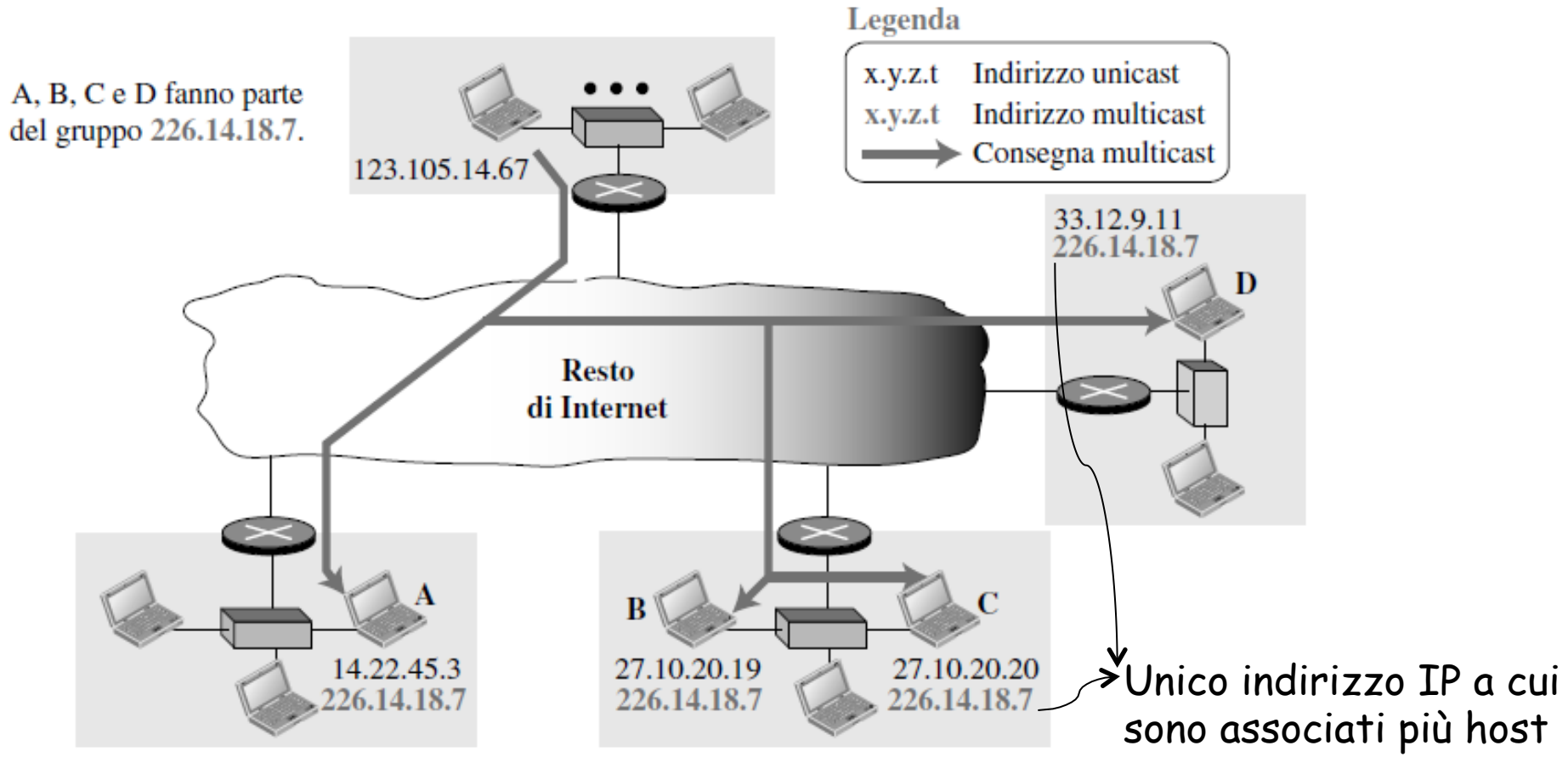
- Come è possibile comunicare con host che partecipano a un gruppo ma appartengono a reti diverse?
- ES. Un gioco multi-player interattivo può coinvolgere host appartenenti a continenti diversi
- L'indirizzo di destinazione nell'IP può essere uno solo
- Soluzione: unico indirizzo per tutto il gruppo ovvero *indirizzo multicast*

# Gruppo multicast

Indirizzi multicast:

| 1110 | group identifier |
|---|---|

First byte: 224 to 239




I router devono sapere quali host sono associati a un gruppo multicast !!!

# Indirizzi multicast

- Blocco di indirizzi riservati per il multicast

- In IPv4
  - 224.0.0.0/4
  - 1110---identificatore del gruppo---

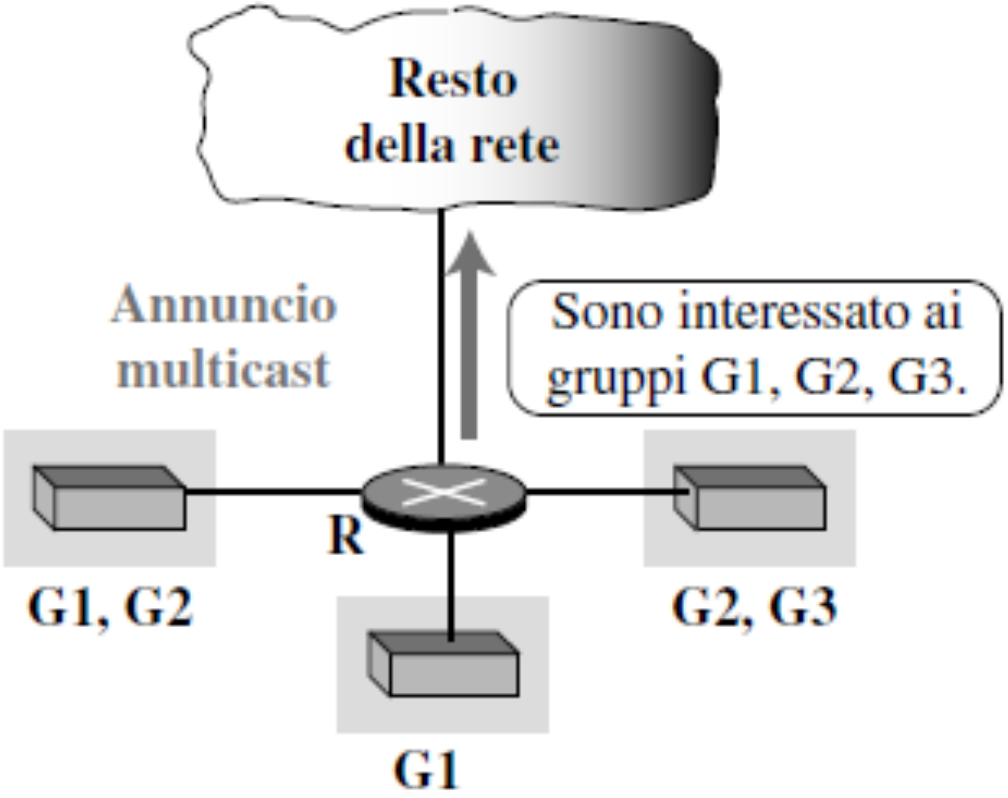
    (da 224.0.0.0 a 239.255.255.255)
  - Numero di gruppi: $2^{28}$

# Gruppi multicast

- L'appartenenza a un gruppo non ha alcuna relazione con il prefisso associato alla rete
- Un host che appartiene ha un gruppo ha un indirizzo multicast separato e aggiuntivo rispetto al primario
- L'appartenenza non è un attributo fisso dell'host (periodo di appartenenza puo' essere limitato)
- Come può un router sapere quali host appartengono a un gruppo?

# Gruppi multicast

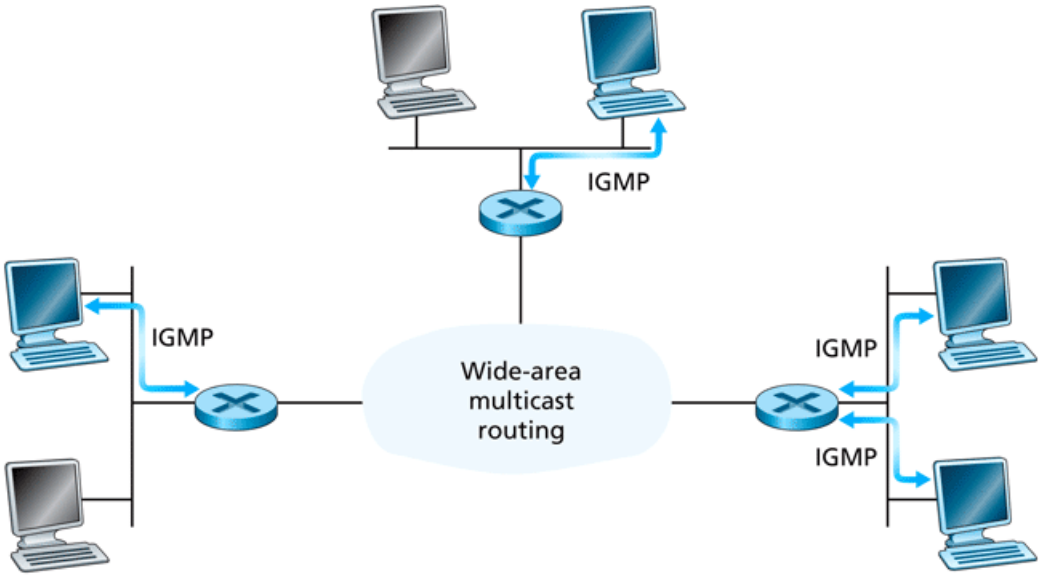
- Un router deve scoprire quali gruppi sono presenti in ciascuna delle sue interfacce
- Il router deve propagare le informazioni agli altri router



# Internet Group Management Protocol (IGMP)

- Lavora **tra** un **host** e il **router** che gli è direttamente connesso

1. Offre agli host il mezzo di informare i router ad essi connessi del fatto che un'applicazione in esecuzione vuole aderire ad uno specifico gruppo multicast
2. È necessario un protocollo che coordini i router multicast in Internet (instradare pacchetti multicast dalla sorgente alla destinazione)



# IGMP

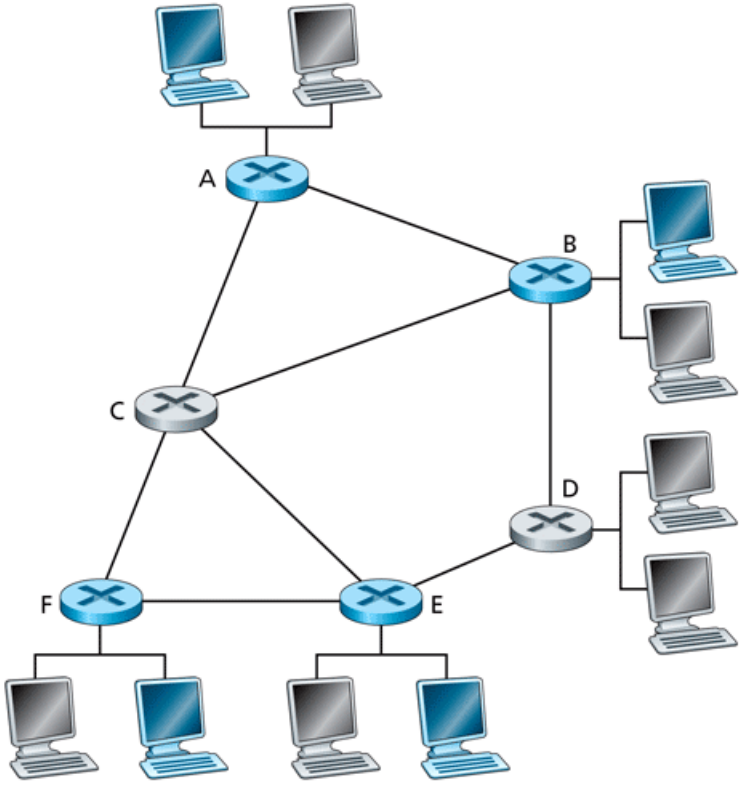
- Messaggi incapsulati in datagrammi IP, con IP protocol number 2
  - Mandati con TTL a 1
- Messaggi IGMP
  - **Membership query**: router → host, per determinare a quali gruppi hanno aderito gli host su ogni interfaccia (inviati periodicamente)
  - **Membership report**: host → router, per informare il router su un'adesione, anche non inseguito a una query (al momento dell'adesione)
  - **Leave group**: host → router, quando si lascia un gruppo
- Il leave group è opzionale: il router può capire che non ci sono più host associati a un gruppo quando non riceve report in risposta a query

# IGMP

- Un router multicast tiene una lista per ciascuna sottorete dei gruppi multicast (multicast group membership → almeno un elemento del gruppo fa parte della sottorete) con un timer per membership
  - la membership deve essere aggiornata da report inviati prima della scadenza del timer
  - può essere anche aggiornata tramite messaggi di leave espliciti

# Problema del routing multicast

- Fra la popolazione complessiva di router solo alcuni (quelli collegati a host del gruppo multicast) dovranno ricevere traffico multicast



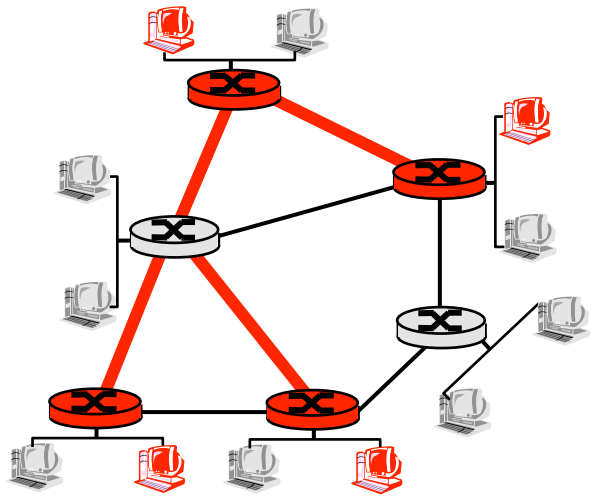

A,B,E,F sono router che devono ricevere traffico multicast

***Obiettivo:*** trovare un albero che colleghi tutti i router connessi ad host che appartengono al gruppo multicast.
I pacchetti verranno instradati su questo albero

# Approcci per determinare albero d'instradamento multicast

*Albero condiviso dal gruppo:*

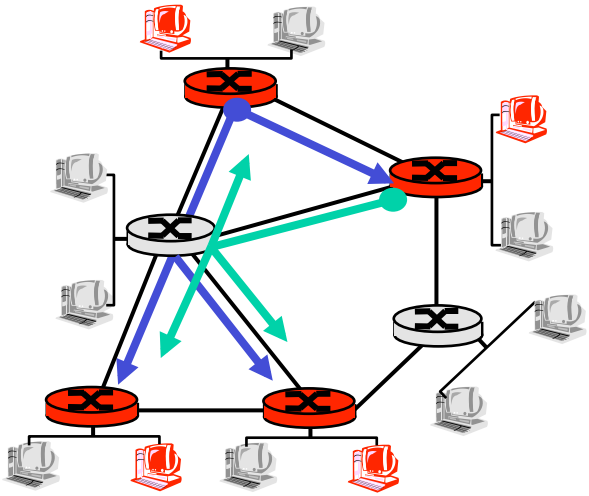
- viene costruito un singolo albero d'instradamento condiviso da tutto il gruppo multicast
- Un router agisce da rappresentante del gruppo
- Se il mittente del traffico multicast non è il centro, allora esso invierà il traffico in unicast al centro, e il centro provvederà a inviarlo al gruppo



Albero condiviso dal gruppo

# Approcci per determinare albero d'instradamento multicast

- *Albero basato sull'origine:* viene creato un albero per ciascuna origine nel gruppo multicast
  - Ci sono tanti alberi quanti sono i mittenti del gruppo multicast
  - Per la costruzione si usa un algoritmo basato su inoltro su percorso inverso, con pruning (potatura)



Albero basato sull'origine

# Instradamento multicast in Internet

Intra-dominio multicast (interno a un sistema autonomo)

- DVMRP: distance-vector multicast routing protocol
- MOSPF: multicast open shortest path first
- PIM: protocol independent multicast

Inter-dominio multicast (tra sistemi autonomi)

- MBGP: multicast border gateway protocol