



**La Sapienza**

Università degli Studi di Roma

Dipartimento di Informatica e Sistemistica

# RETI DI CALCOLATORI II

## BGP - **B**order **G**ateway **P**rotocol

**Emiliano Trevisani**

**[trevisani@dis.uniroma1.it](mailto:trevisani@dis.uniroma1.it)**

**A.A. 2008/2009**

# Instradamento tra Sistemi Autonomi -- BGP

Thanks to:

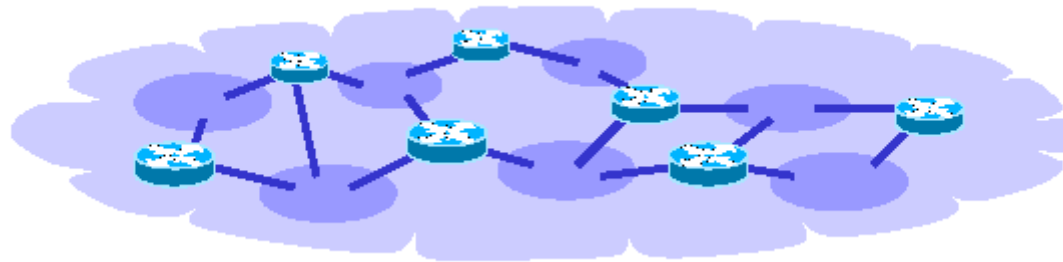
Giuseppe Di Battista, Maurizio Patrignani, Maurizio Pizzonia:  
Università di Roma Tre

**Timothy G. Griffin**

**<http://www.research.att.com/~griffin/interdomain.html>**

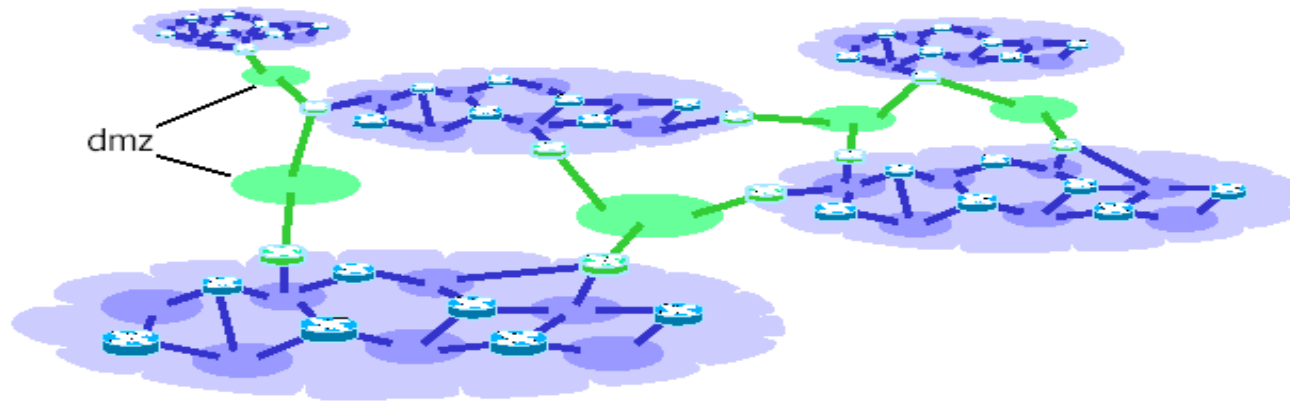
# I Sistemi Autonomi

- Ogni organizzazione è composta da un insieme di router e LAN sotto una singola amministrazione
- Un algoritmo di routing è prescelto per aggiornare automaticamente le tabelle di instradamento
- Un AS definisce in maniera coerente le politiche di instradamento all'interno della sua organizzazione



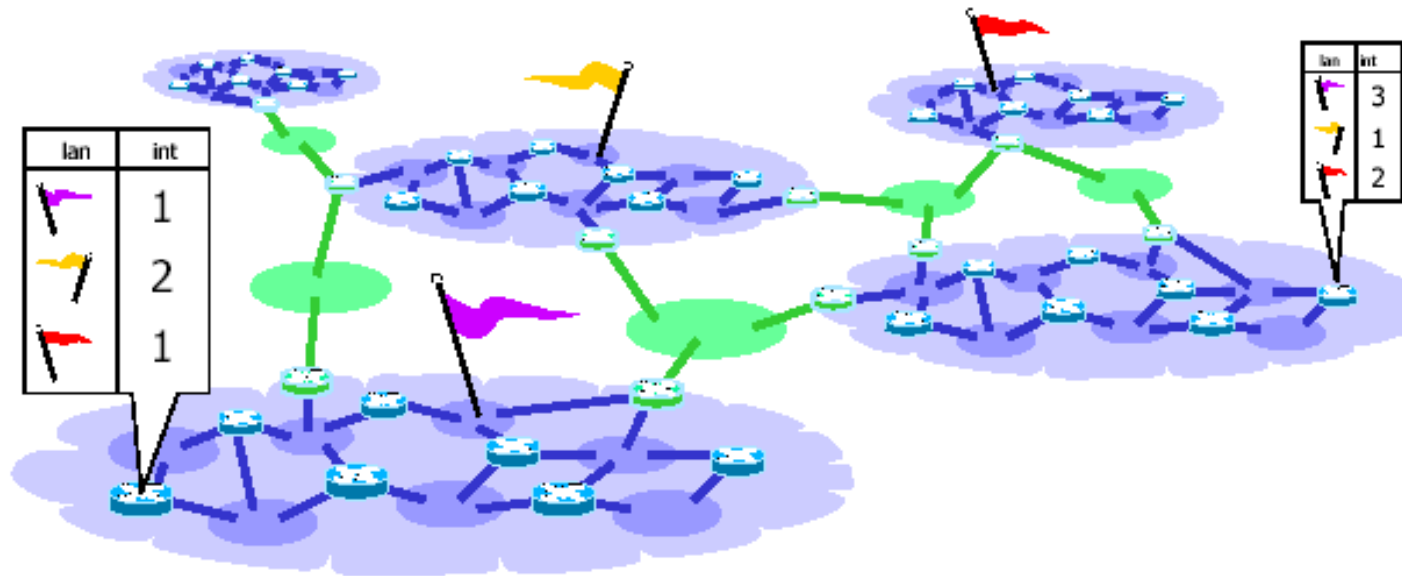
# L'interconnessione di Sistemi Autonomi

- Quando più organizzazioni si uniscono per formare una Inter-rete, occorre stabilire tra loro punti di collegamento
- Le reti che vengono aggiunte sono dette punti di demarcazione



# L'instradamento tra Sistemi Autonomi

- Ogni tabella deve avere un'entry per ogni possibile destinazione
- Questo deve valere sia per le destinazioni locali che per quelle globali



# Come aggiornare le tabelle di Instradamento?

In generale vi sono tre opzioni:

1. Eseguire un unico algoritmo di instradamento tra organizzazioni adiacenti
2. Aggiornare le tabelle di instradamento manualmente aggiungendo percorsi statici predefiniti
3. Combinare un protocollo di instradamento intra-domain con un protocollo di instradamento inter-domain: Exterior gateway protocol

# 1. Unico algoritmo di Instradamento

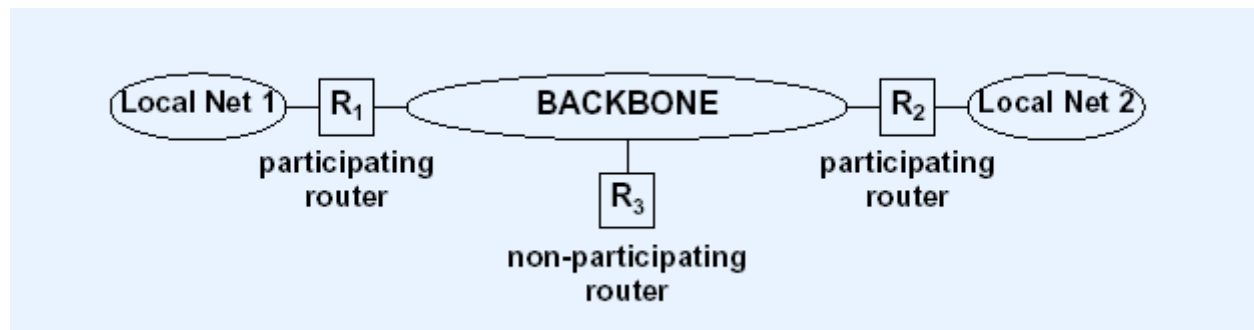
- Molti Svantaggi:
  - Ritardo di propagazione, ex: distance vector
  - Scalabilita'
  - Tutte le organizzazioni sono forzate ad usare lo stesso algoritmo
  - Un nuovo algoritmo di instradamento è di difficile adozione
  - Non considera le relazioni politiche e commerciali tra sistemi autonomi

## 2. Percorsi statici

- Si nasconde la parte interna dell'AS
- Per ogni obiettivo esterno si identifica un router alla frontiera del Sistema Autonomo di destinazione
- Informazione sul cammino da seguire per raggiungere l'obiettivo
- Svantaggi:
  - difficile da aggiornare e da correggere
  - I malfunzionamenti non sono gestiti, non si ha backup
  - Nessuna garanzia che tutti i router del percorso sono in effetti disponibili per portare il traffico a destinazione

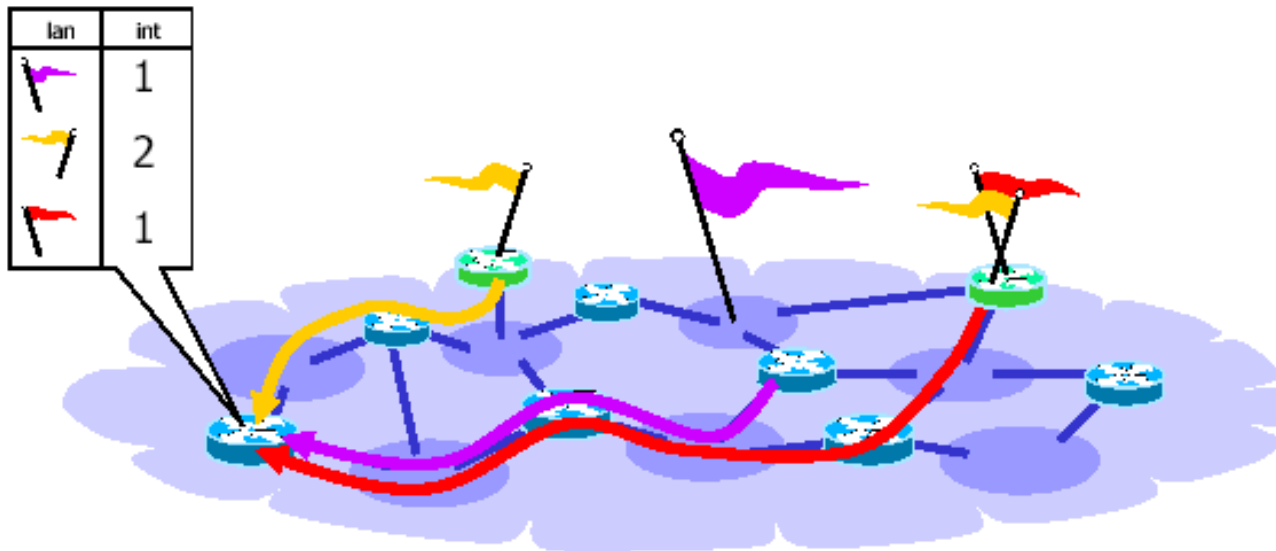
## 2. Percorsi statici

- L'instradamento può essere inefficiente
- Nell'esempio R1 ed R2 sono parte dello stesso AS. R3 invia ad R1 tutto il traffico diretto all'AS, anche quello diretto alla LAN 2.
- L'instradamento non tiene conto delle reti che si possono effettivamente raggiungere!



## 2. Percorsi statici

- L'algoritmo di instradamento diffonderà all'interno dell'AS il traffico locale e il traffico che segue i percorsi statici

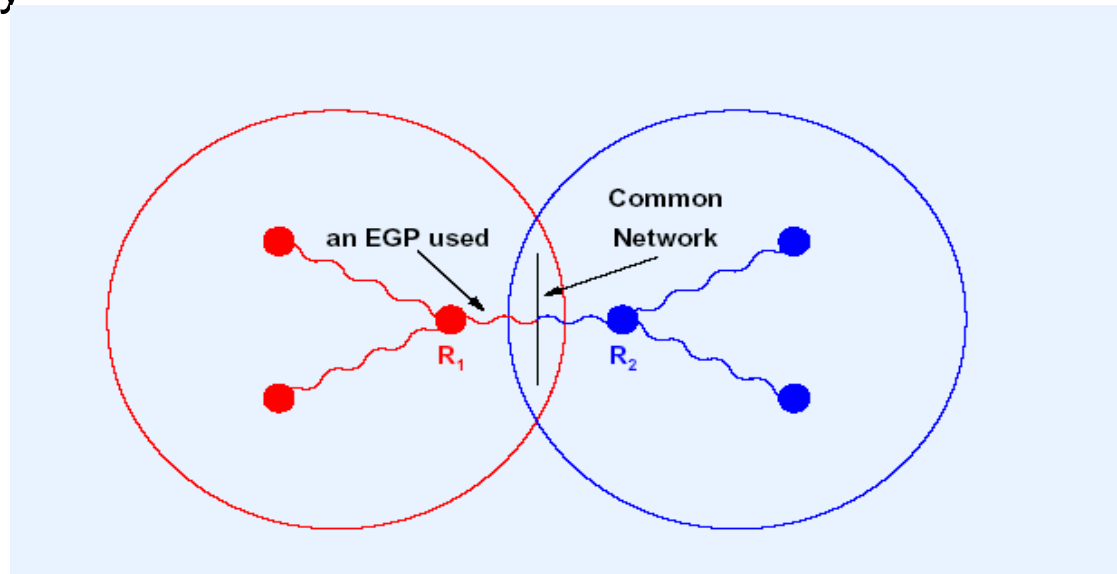


# Un approccio diverso

- Occorre avere un flusso informativo in due direzioni, sia dall'interno verso l'esterno che dall'esterno verso l'interno
- L'AS si deve far carico di garantire la consistenza degli instradamenti interni
- Occorre annunciare all'esterno quali reti interne sono raggiungibili
- Occorre assegnare le responsabilità per la diffusione delle informazioni riguardo l'instradamento

# 3. Exterior gateway protocol

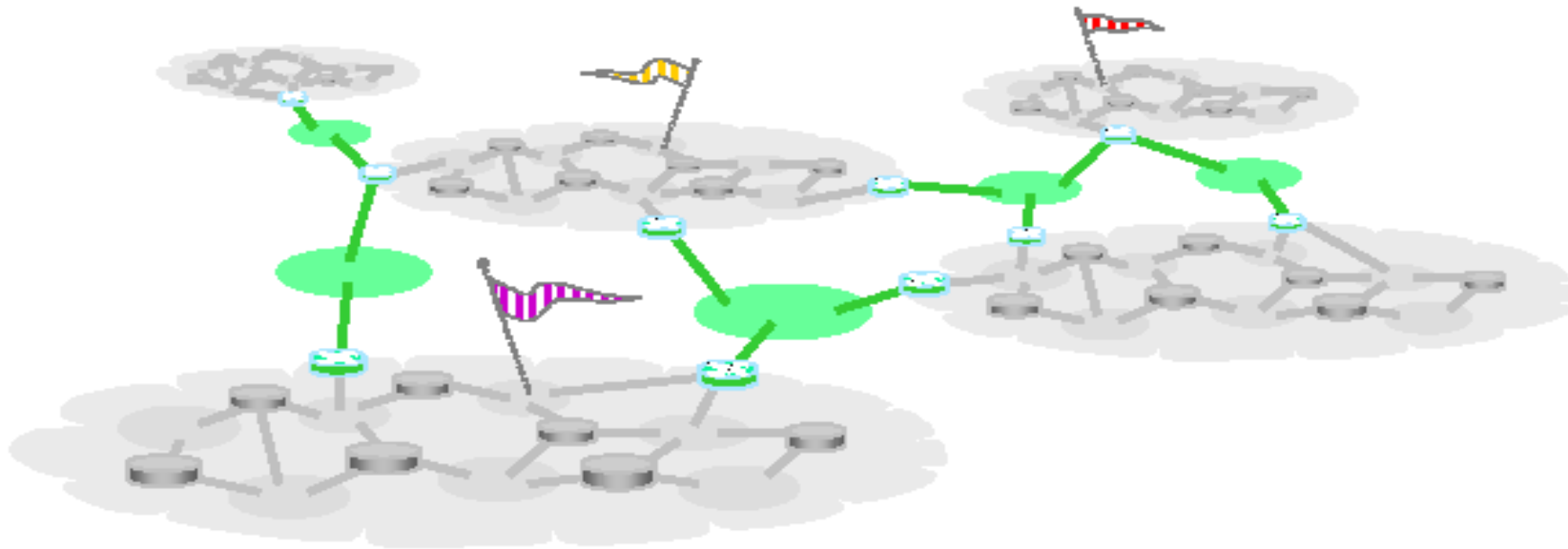
- Qualunque protocollo per lo scambio delle informazioni sull'instradamento tra Sistemi Autonomi
  - Anche un protocollo specifico anteriore a BGP
- BGP – Border Gateway Protocol
- Due AS che si scambiano informazioni di instradamento designano due router che stabiliscono una sessione di peering
- Router che partecipano a BGP sono detti Router di Confine o Gateway



# 3. Exterior Gateway Protocol

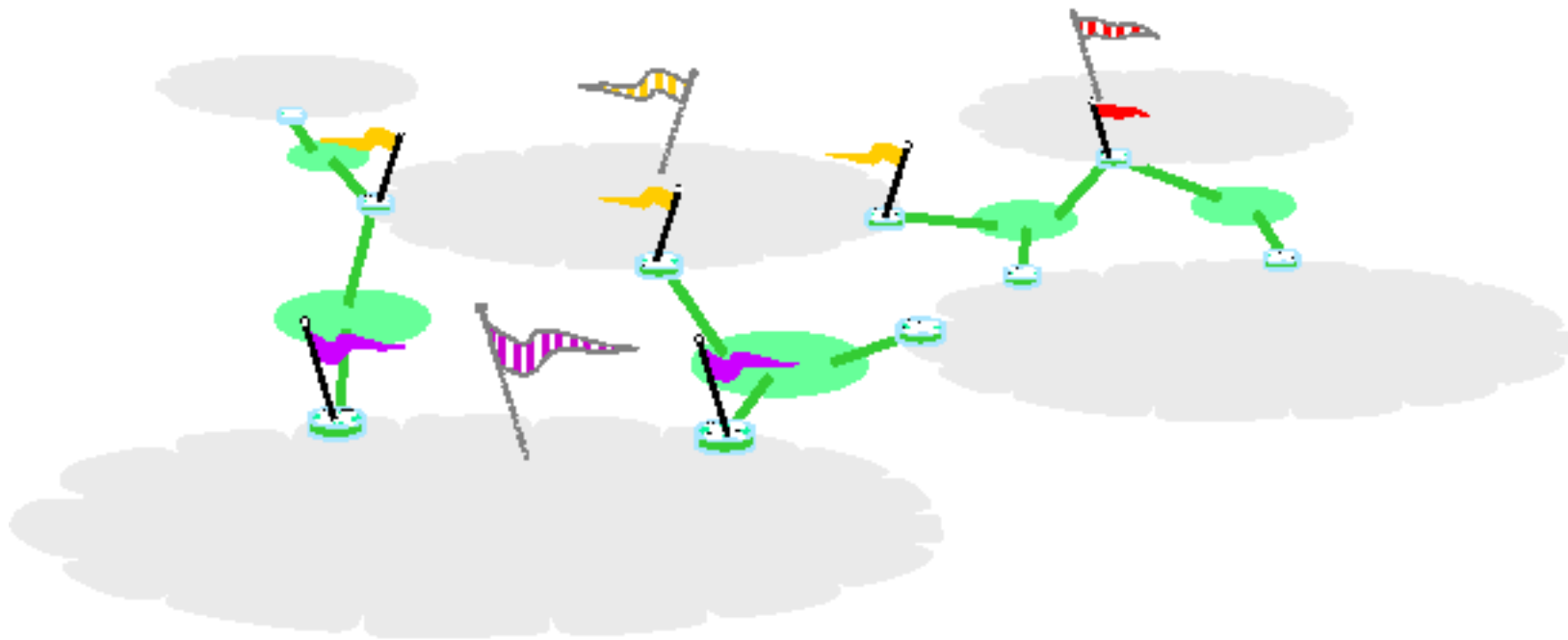
Approccio:

- Nascondi la parte interna degli AS
- Mantieni solo le zone di demarcazione e i router di frontiera degli AS



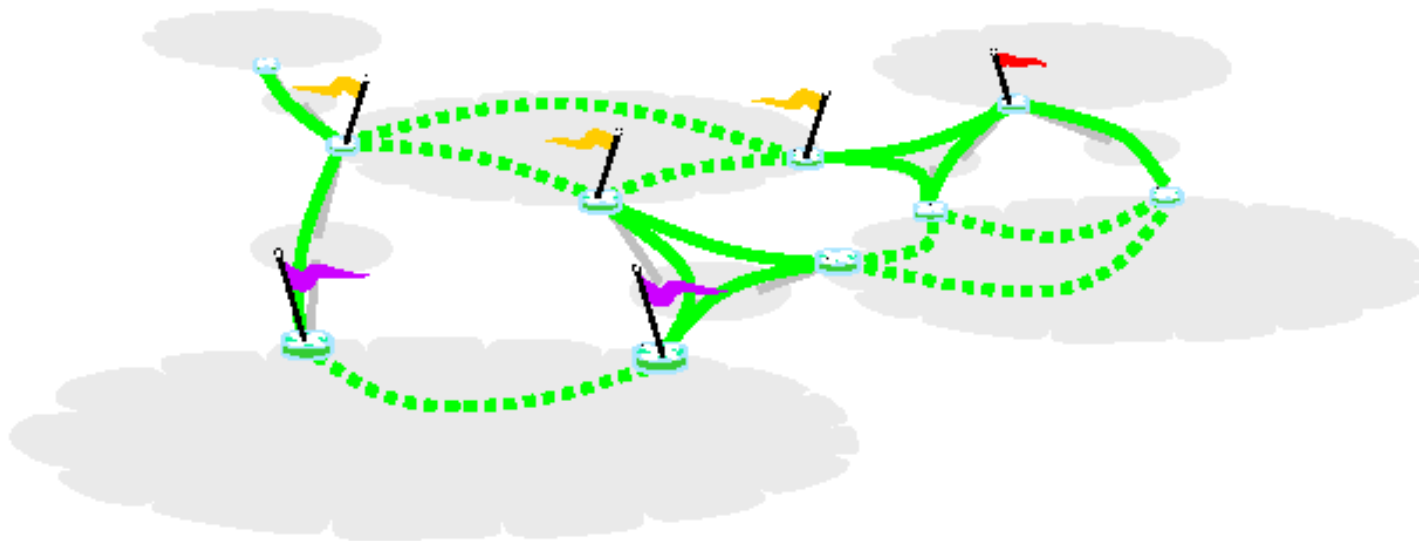
# 3. Exterior Gateway Protocol

- Ogni router di frontiera rappresenta le destinazioni interne come se fossero locali



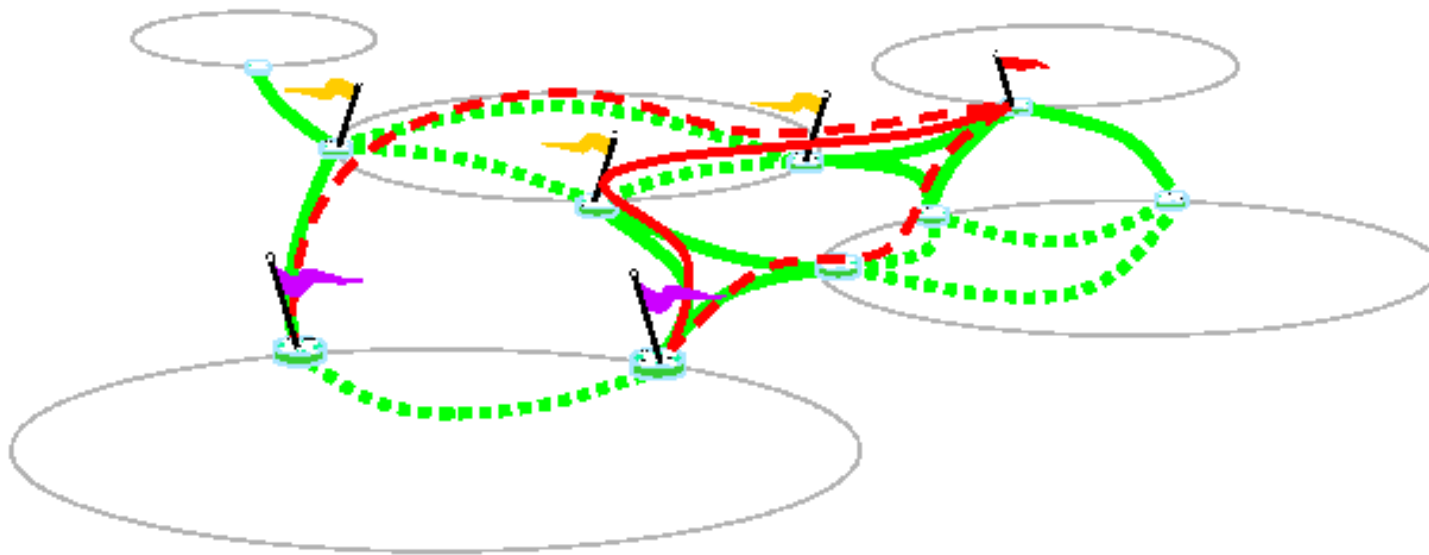
# 3. Exterior Gateway Protocol

- Semplifica il grafo considerando le informazioni sulla raggiungibilità sia interna che esterna all'AS
- Il grafo è gestito attraverso sessioni peering TCP



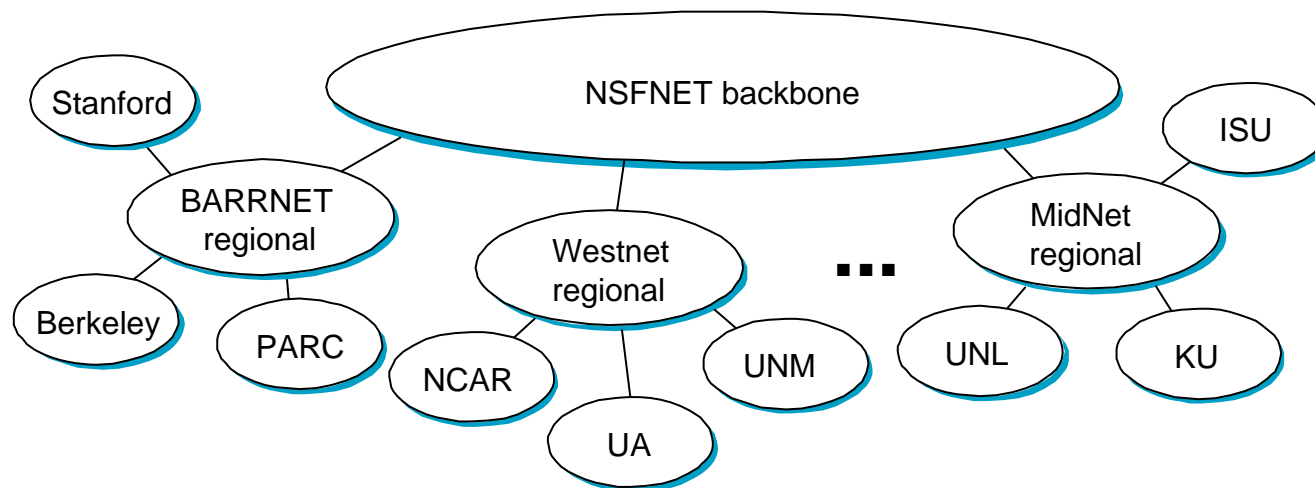
# 3. Exterior Gateway Protocol

- Definisci anche percorsi prestabiliti sulla base di considerazioni politiche



# 3. Exterior Gateway Protocol

- Concepito quando Internet aveva l'aspetto riportato sotto
- Struttura del grafo degli AS ad albero



## Exterior Gateway Protocol (EGP)

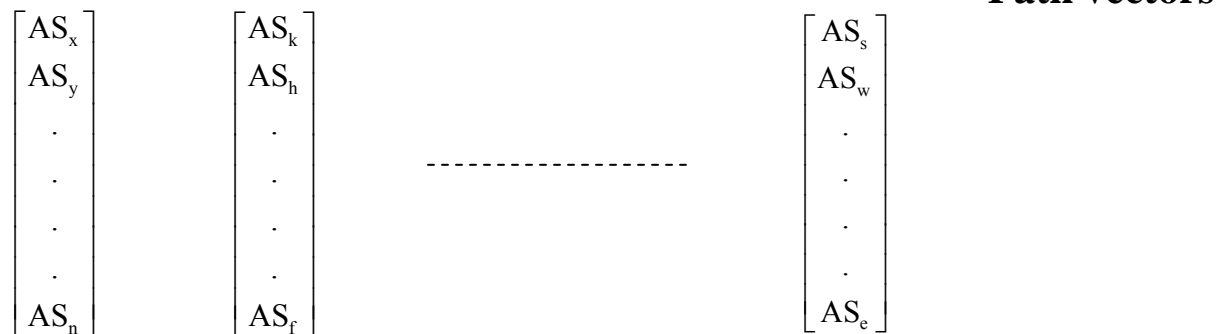
- ❑ I distance-vector routing protocol (tipo RIP) non sono adatti per essere applicati come EGP
  - assumono che tutti i router utilizzano la stessa metrica; questa condizione non è garantita tra AS diversi
  - non ci sono indicazioni dei router intermedi lungo il cammino; in un ambiente inter-AS ci possono essere transiti privilegiati o transiti proibiti
- ❑ I link-state routing protocol (tipo OSPF) non sono adatti per essere applicati come EGP
  - AS diversi possono utilizzare metriche diverse
  - La tecnica di flooding è inapplicabile tra AS diversi
- ❑ L'instradamento inter-AS non dipende solo dalle performance; spesso dipende molto di più da parametri di natura diversa [es. accordi commerciali]
  - Per questo motivo i protocolli EGP **non usano metriche ma solo info di raggiungibilità**

## Exterior Gateway Protocol (EGP)

- ❑ Nei protocolli EGP si usa la tecnica **path vector routing**
- ❑ Si utilizzano esclusivamente informazioni riguardanti:
  - quali reti possono essere raggiunte attraverso un router
  - quali AS sono attraversati lungo il cammino
- ❑ Non si utilizzano nozioni di distanza o costo
- ❑ Si determina la lista degli AS che devono essere attraversati per raggiungere una particolare rete lungo un particolare cammino
  - l'instradamento terrà conto di eventuali preferenze per alcuni AS rispetto ad altri (accordi commerciali, prestazioni, ecc.)

## Path Vector Routing Protocols

- ❑ Instradamento basato su un “vettore di cammini”
- ❑ Idea base: produrre, per ogni coppia <subnet origine, subnet destinazione> una lista di alternative ciascuna delle quali è una sequenza di AS da attraversare per raggiungere la subnet destinazione a partire da quella di origine
  - Sulla base di specifici criteri [accordi commerciali, preferenze,...] verrà selezionata una delle possibili alternative d'instradamento dall'AS che emette il pacchetto

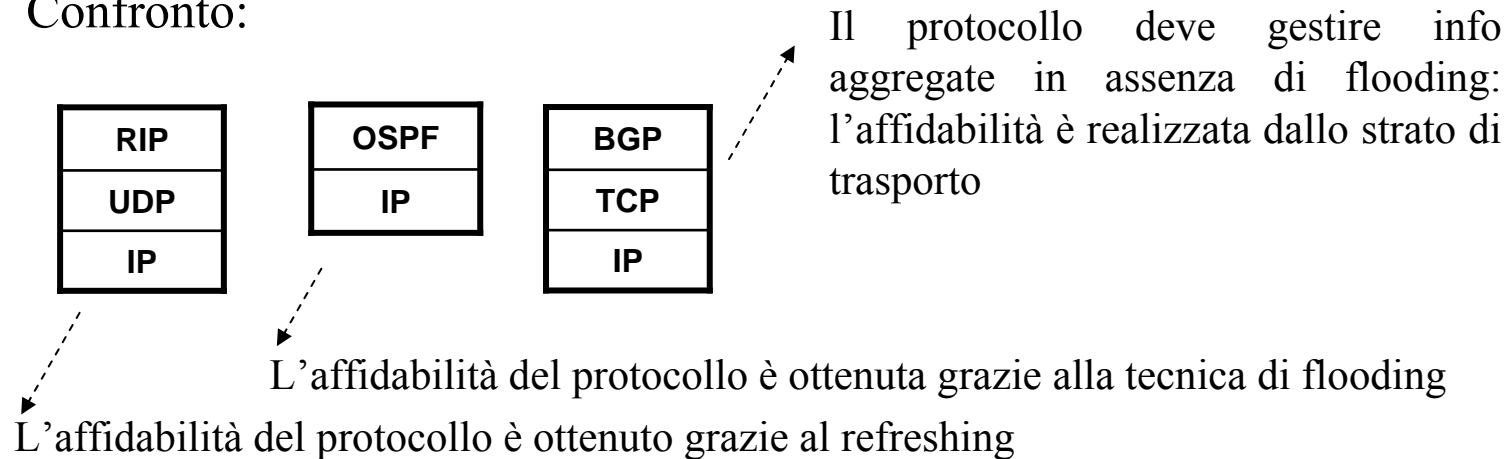


# Border Gateway Protocol (BGP)

## □ BGP:

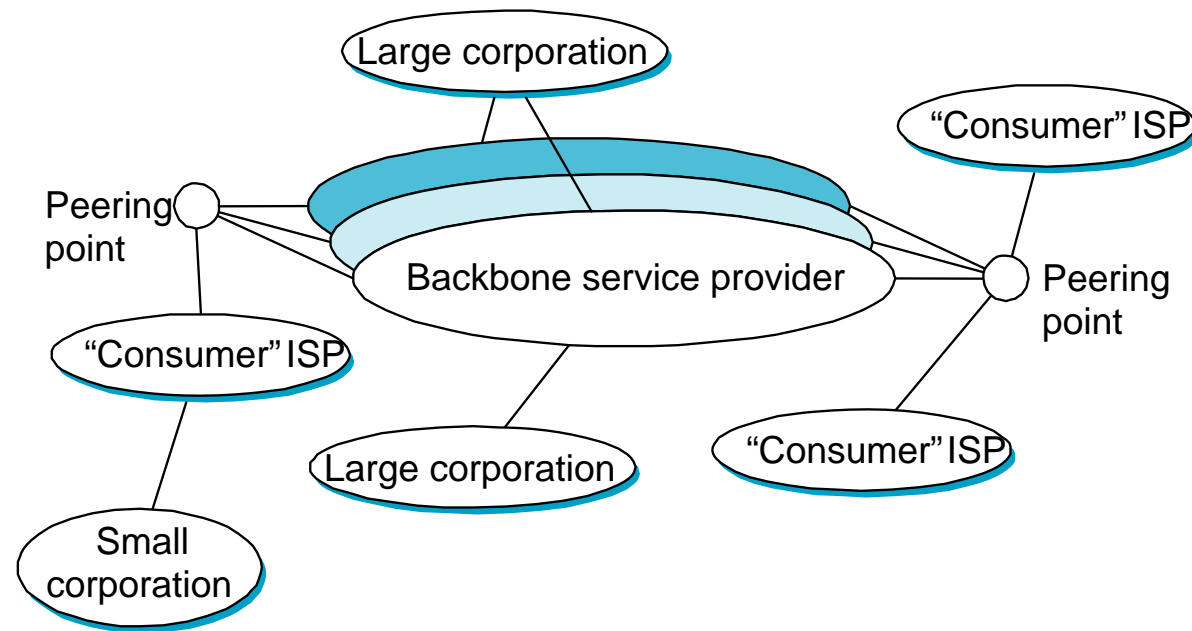
- Permette a router appartenenti a AS diversi di scambiarsi informazioni di raggiungibilità
- Supporta il CIDR [in particolare anche il subnetting a lunghezza variabile]
- Lo scambio di messaggi BGP è supportato da connessioni TCP
- La versione proposta più recente è BGPv5

## ▪ Confronto:



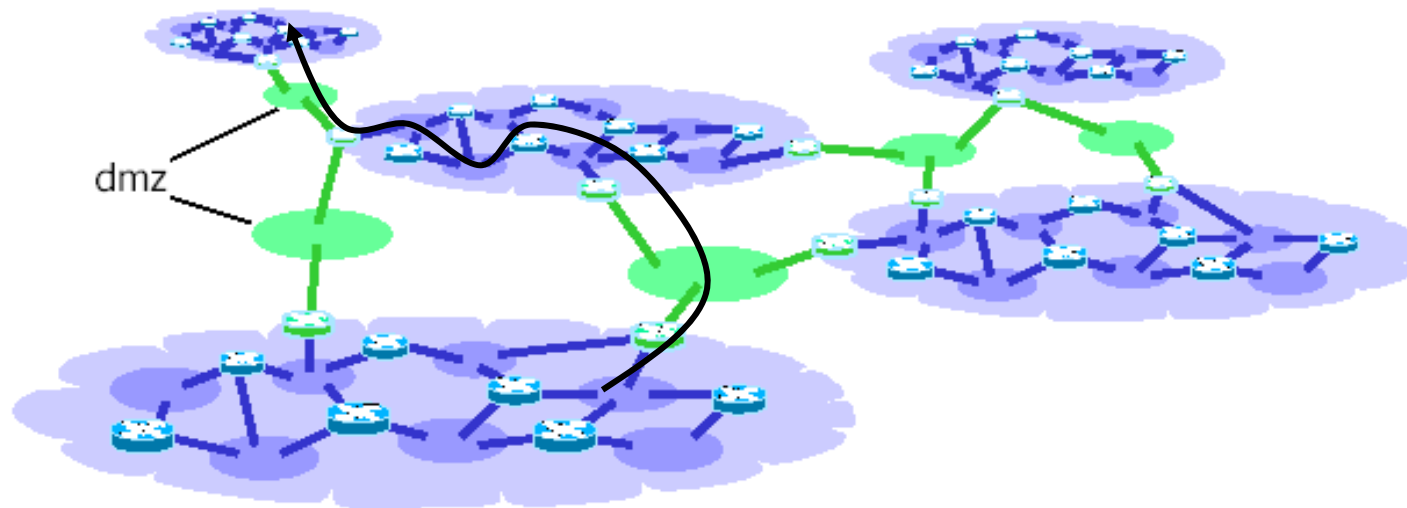
# BGP v4 – Border Gateway Protocol

- Nessuna assunzione su grafo degli AS
- Piu' reti backbone interconnesse
  - Service provider networks
- Molti SP esistono per erogare servizi



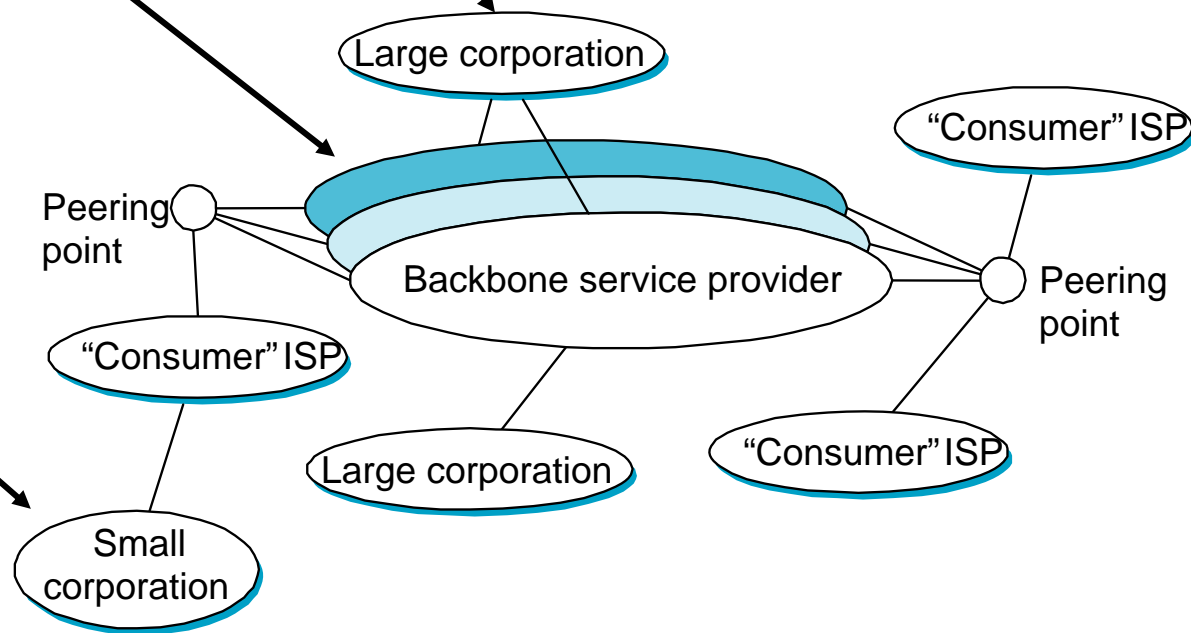
# BGP v4 – Border Gateway Protocol

- Traffico locale
  - Inizia o termina in nodi interni
- Traffico di transito
  - Varca i confini tra AS diversi



# BGP v4 – Border Gateway Protocol

- Stub AS
- Multihomed AS
- Transit AS



# BGP v4 – Border Gateway Protocol

Each AS has:

- One or more border routers
- One BGP *speaker* advertises:
  - local networks
  - other reachable networks (transit AS only)
    - C'e' comunque sempre una default route
  - gives *path* information

# Terminologia BGP

## □ BGP speaker

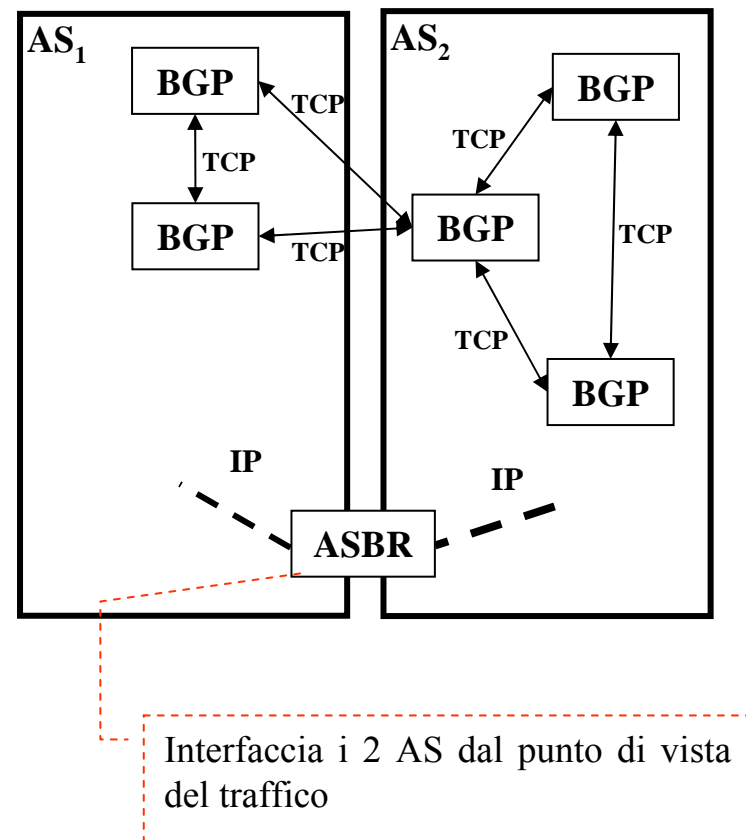
- un router che supporta il protocollo BGP
- un BGP router non necessariamente coincide con un border router (ASBR)

## □ 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



## Terminologia BGP

### ❑ AS Border Router (ASBR)

- un router che è connesso ad altri sistemi autonomi
- Interno
  - un ASBR che si trova nello stesso AS del BGP speaker
- Esterno
  - un ASBR che si trova in un altro AS rispetto a quello del BGP speaker

### ❑ AS connection

- Physical connection
  - i due AS condividono la stessa sottorete fisica
- BGP connection
  - esiste una sessione BGP tra una coppia di BGP speaker appartenenti a AS diversi

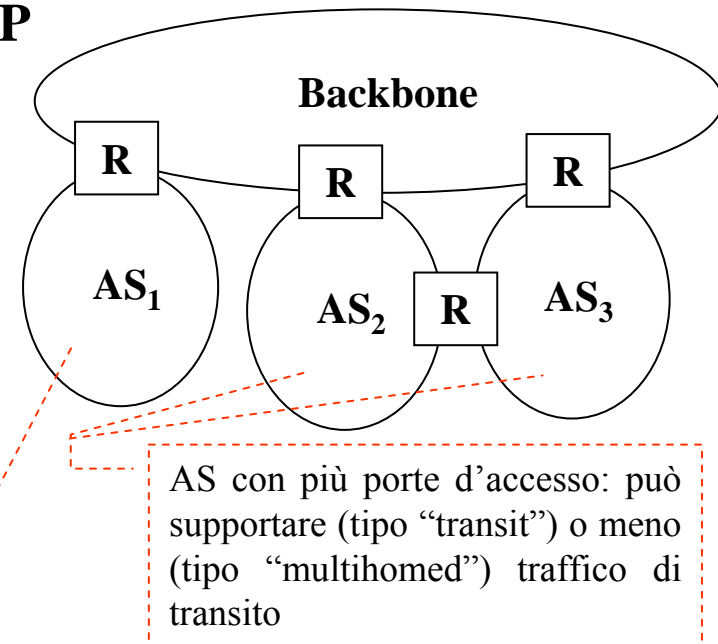
## Terminologia BGP

### □ Traffico

- Locale
  - traffico generato o terminato nell'AS
- Transit
  - traffico che non è locale

### □ AS type

- Stub
  - uno stub AS ha una singola connessione inter-AS, trasporta solo traffico locale
- Multihomed
  - ha un insieme di connessioni verso una molteplicità di altri AS, ma non trasporta traffico di transit
- Transit
  - ha un insieme di connessioni verso una molteplicità di altri AS, e trasporta anche traffico di transit



## 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

## BGP

- ❑ Il protocollo BGP impone che un AS presenti la stessa visione a tutti gli AS che usano i suoi servizi
  - questa condizione è garantita dal protocollo IGP (es. OSPF)
- ❑ Il protocollo BGP di un AS comunica ad altri AS solo cammini che lo usano come next-hop
  - conforme al classico schema di routing in IP
- ❑ Due BGP speaker, dopo aver instaurato una sessione, si scambiano i path completi verso ogni altro AS di destinazione
  - un path è indicato sotto forma di lista di AS
  - la disponibilità dell'intera lista di AS **evita l'insorgere di loop**

# BGP

## □ Neighbor Acquisition Procedure

- E' utilizzata quando due router di due AS diversi collocati sulla stessa sottorete hanno intenzione di dare inizio allo scambio di informazioni
- E' necessario l'accordo di entrambi per evitare il sovraccarico di uno di essi
- La procedura consiste nell'invio di una richiesta (Open message) e di una risposta (Keepalive message)
- La procedura può essere iniziata dal network manager

# BGP

## □ Neighbor Reachability Procedure:

- E' utilizzata per mantenere attivo il colloquio tra due router
- Ogni router si assicura che l'altro sia attivo e mantenga la relazione
- I due router si scambiano periodicamente messaggi keepalive

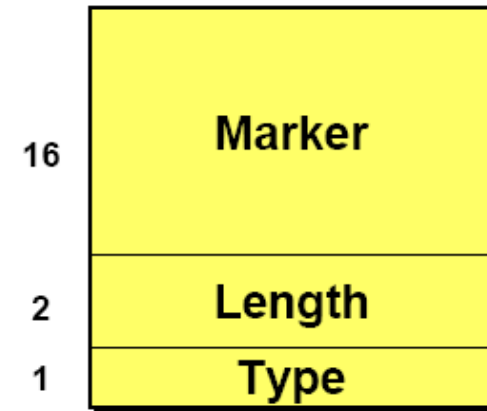
## □ Network Reachability Procedure:

- Ogni router mantiene un database delle reti raggiungibili e del cammino preferito per raggiungerle
- Quando avviene un cambio nel database il router emette un messaggio di Update verso gli altri router per comunicare l'avvenuto cambiamento

## Messaggi BGP

### □ Header (19 ottetti)

- Comune a tutti i messaggi BGP
- Marker (16 bytes)
  - realizza le funzioni di autenticazione affinché il destinatario possa verificare l'identità della sorgente
- Length (2 bytes)
  - lunghezza in ottetti del messaggio
- Type (1 byte)
  - Tipo di messaggio (Open, Update, Keepalive, Notification)

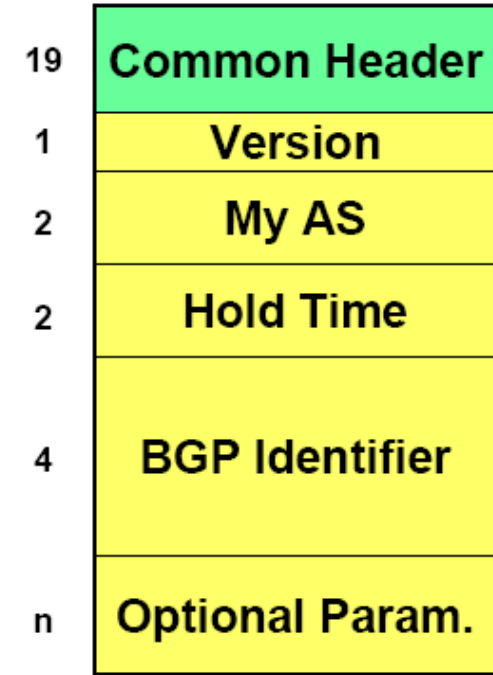


**Header BGP: 19 byte**

# Messaggi BGP

## □ Open Message

- E' utilizzato per la procedura di Neighbor Acquisition
- My AS
  - identificatore dell'AS del router
- Hold time
  - durata proposta per il timer che governa la procedura di keepalive
- BGP identifier
  - Indirizzo IP del router

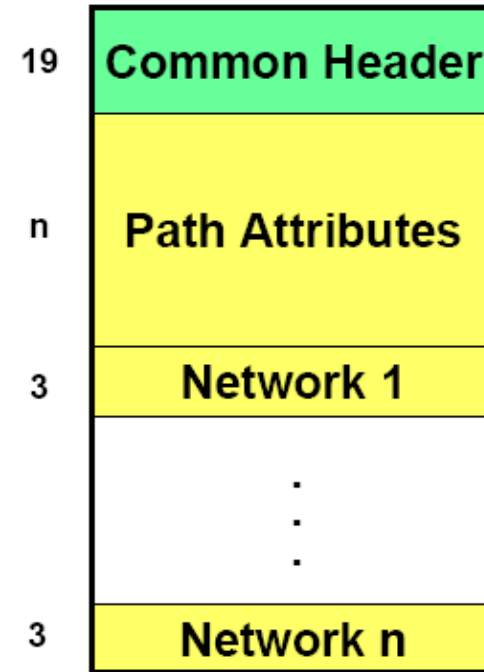


## Messaggi BGP

### □ Update Message

Messaggio “chiave” per BGP: quali reti posso raggiungere con un certo cammino, espresse come lista di AS attraversati

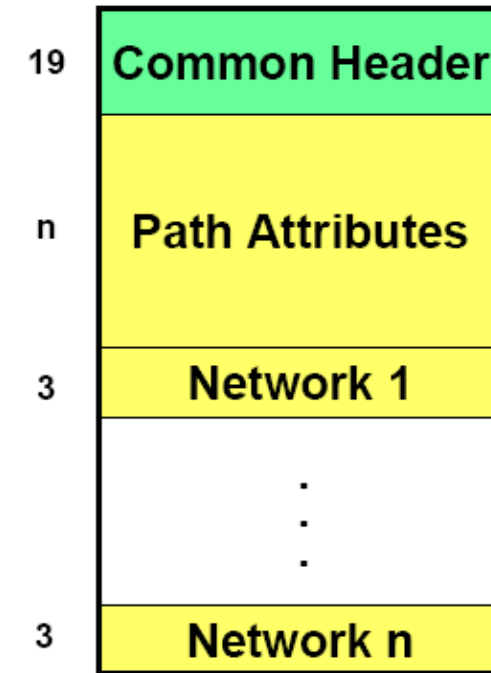
- E' usato per inviare ai router con cui esiste una relazione le informazioni di raggiungibilità relative ad un singolo cammino
- Path Attributes
  - identifica le caratteristiche del cammino
- Network 1,..., Network n
  - lista degli indirizzi delle reti raggiungibili dal cammino
  - possono essere specificati i prefissi CIDR



# Messaggi BGP

## □ Path Attributes

- Origin
  - indica se l'informazione è generata da un IGP o un EGP
- AS\_Path
- Lista degli AS attraversati
- Next\_hop
  - indirizzo IP del next hop router da usare nell'instradamento
- Multi\_Exit\_Disc
  - informazioni sul routing interno ad un AS
- Local\_Pref
  - grado di preferenza del cammino
- Atomic\_Aggregate, Aggregator
  - aggregazione di indirizzi di rete, utile nel caso di instradamento gerarchico



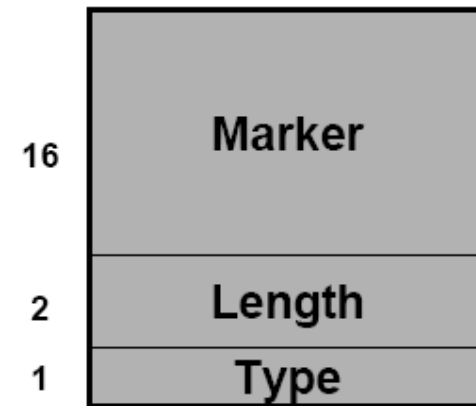
## Messaggi BGP

- ❑ I messaggi di Update sono inviati all'inizio della relazione tra due nodi e successivamente quando si verificano cambiamenti nel cammino
- ❑ Il router che riceve un messaggio di Update confronta il cammino ricevuto con quello correntemente usato
  - se il nuovo path è migliore, il vecchio è sostituito e la comunicazione è inviata agli altri router
  - se il nuovo path è meno conveniente di quello corrente non si procede a modifiche

# Messaggi BGP

## □ Keep-alive Message

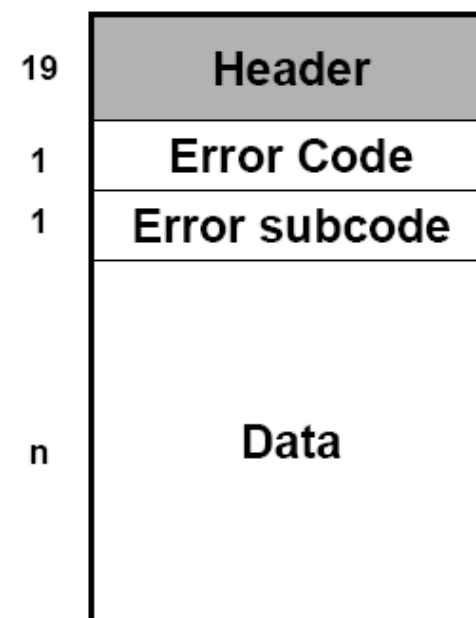
- E' usato per manifestare l'attività del router ed evita lo scadimento dell'Hold Timer
- Assicura la raggiungibilità del router emittente
- E' composto solo dai byte dell'header



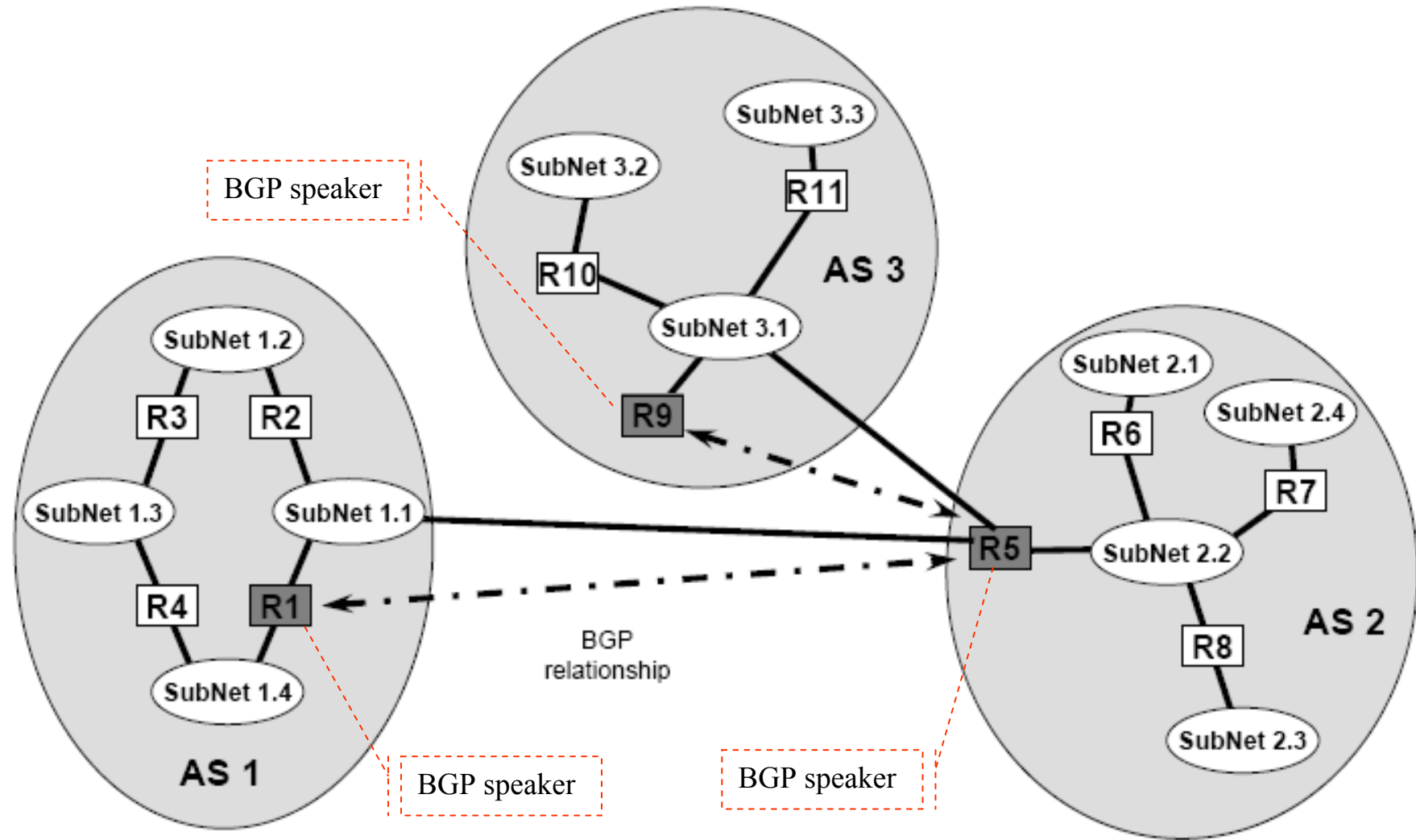
## Messaggi BGP

### □ Error Notification Message

- E' usato per inviare una notificazione di errore ai router vicini
  - Scadimento dell'Hold Timer
  - Errori procedurali e messaggi errati
  - Errori di indirizzo
  - ...



# Esempio BGP 1/3



## Esempio BGP 2/3

- ❑ I router R1 e R5 implementano sia il protocollo BGP che un protocollo IGP (es. OSPF); R1, pertanto, conosce la struttura di AS1
- ❑ Il router R1 emette il messaggio Update verso R5 con:
  - l'identità di AS1
  - IP address di R1
  - la lista delle sottoreti di AS1
- ❑ R5 memorizza che le reti di AS1 sono raggiungibili tramite R1
- ❑ R5 emetterà un Update message verso R9 contenente:
  - le identità di AS1 e AS2
  - IP address di R5
  - la lista delle sottoreti di AS1

## Esempio BGP 3/3

- ❑ Il messaggio avverte R9 che le reti di AS1 sono raggiungibili tramite il router R5 e che nel path sono attraversati sia AS2 che AS1
- ❑ A sua volta R9 invierà un Update message verso i suoi nodi vicini contenente:
  - le identità di AS1, AS2 e AS3
  - IP address di R9
  - la lista delle sottoreti di AS1
- ❑ In questo modo le informazioni di raggiungibilità si propagheranno di router in router attraverso la rete

# BGPv4 – more details

- Numerazione, peering e scambio di messaggi
  - Messaggi BGP
  - EBGP e IBGP
- Annunci BGP - Route advertisement
  - Messaggi di UPDATE
  - Attributi
- Selezione dei cammini
- Interazione con IGP
- Limitazioni BGP e soluzioni
- Architetture BGP e bilanciamento del carico

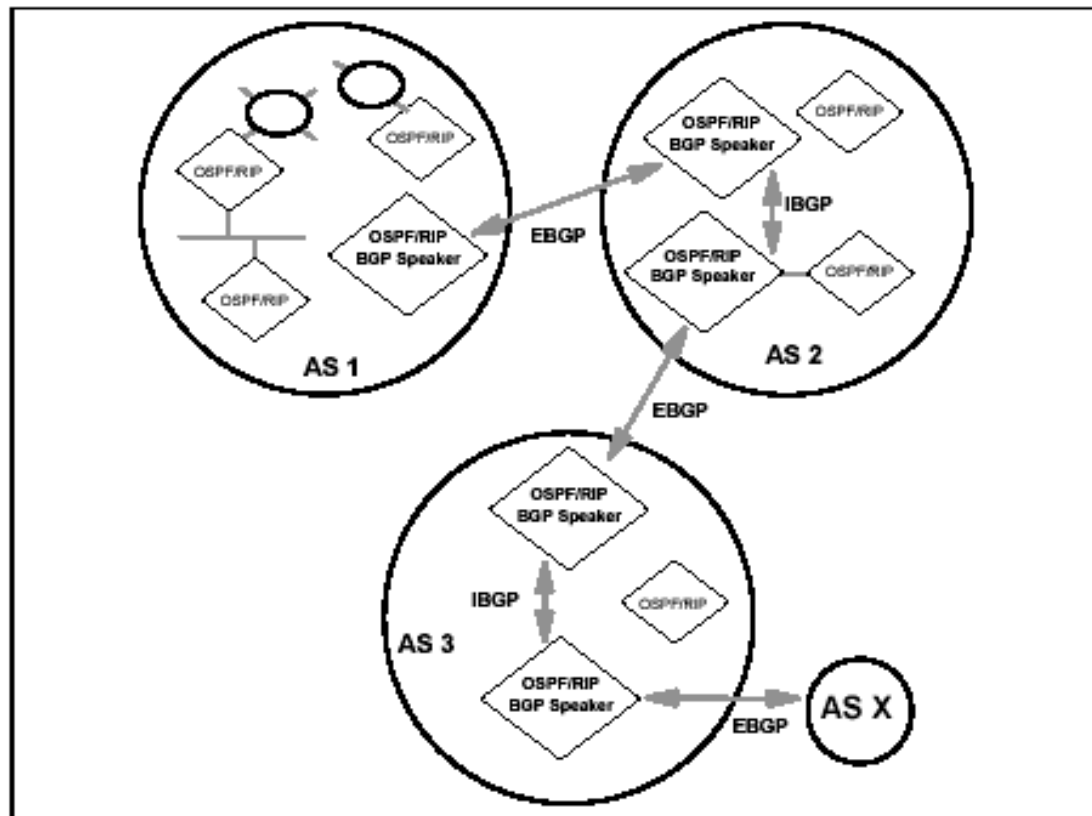
# BGP v4

## Peering e scambio di messaggi

# BGP v4 – Border Gateway Protocol

- Peer: coppia di router BGP che si scambiano informazione di instradamento
  - IBGP peer: stesso AS
  - EBGP peer: AS diversi

**Comunicazione tra peer avviene mediante connessioni TCP**

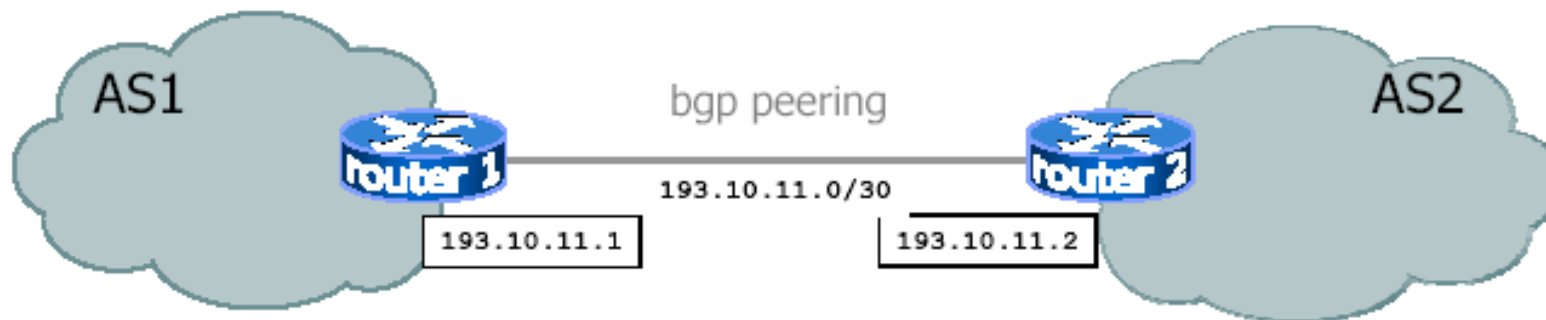


# Numerazione degli AS

- BGP richiede un numero identificativo per ogni AS (Autonomous System Number, asn) tra 1 e 65,535
- Un asn può essere ottenuto da
  - asn globale – all'autorità internet regionale: ripe, arin, apnic
  - asn privato – all'isp

# Peering tra due AS

- Le informazioni possono essere scambiate tra due AS solo se una sessione peering è attiva
- La sessione peering è una connessione TCP tra i due AS



# Funzionalità BGP

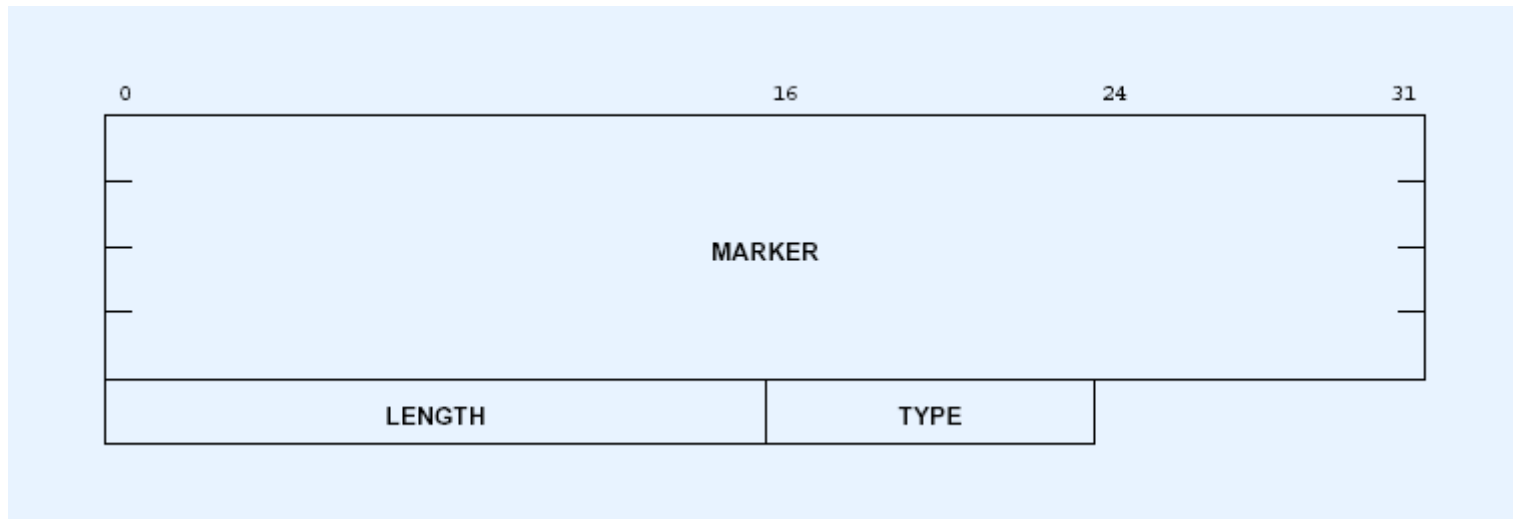
1. Apertura connessione tra peer
2. Annuncio informazioni sulla raggiungibilità
3. Verifica corretto funzionamento

Quattro tipi di messaggio BGP

Type Code	Message Type	Description
1	OPEN	Initialize communication
2	UPDATE	Advertise or withdraw routes
3	NOTIFICATION	Response to an incorrect message
4	KEEPALIVE	Actively test peer connectivity

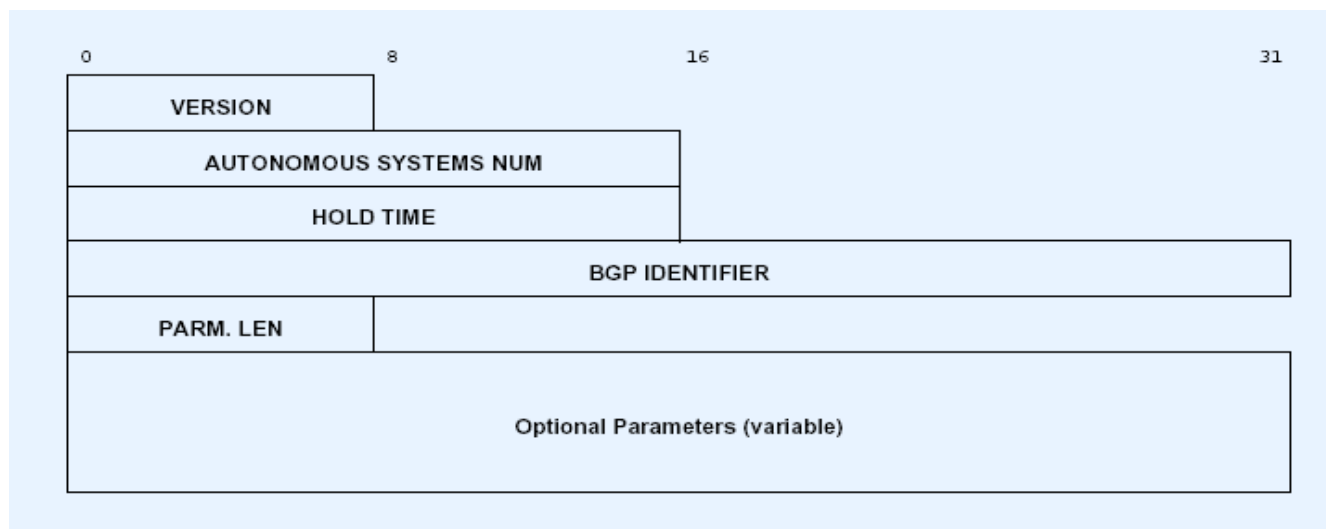
# Intestazione messaggi BGP

- Precede ogni messaggio BGP ed identifica il tipo di messaggio
- Marker (16 byte): autenticazione e sincronizzazione tra i peer
- Length (2 byte): lunghezza del messaggio tra 19 e 4096 byte
- Type: tipo di messaggio BGP



# Peering/apertura connessione

- OPEN: usato per aprire una connessione peer
- Il campo Hold specifica il massimo numero di secondi tra due messaggi successivi
- Un router bgp è caratterizzato dall'asn e da un indentificatore unico a 32 bit che deve usare per tutte le connessioni peering
- Parametri opzionali: ad esempio per l'autenticazione



# Messaggi/OPEN

- Il router destinatario di un messaggio OPEN risponde con un KEEPALIVE
- Connessione aperta quando entrambi i router hanno inviato un messaggio OPEN e ricevuto un messaggio KEEPALIVE

# Messaggi/KEEPALIVE

- Verifica periodicamente la connessione TCP tra entità peer
- Più efficiente rispetto ad inviare periodicamente messaggi di instradamento
- Intervallo KEEPALIVE ogni  $1/3$  di HOLD time, mai inferiore a 1 sec.

# Messaggi/NOTIFICATION

- Controllo o segnalazione errori
- BGP invia un messaggio di notifica e chiude la connessione TCP
- Errori:
  1. Errore nell'intestazione del messaggio
  2. Errore nel messaggio OPEN
  3. Errore nel messaggio UPDATE
  4. Timer di attesa scaduto
  5. Errore nella macchina a stati finiti
  6. Fine (connessione terminata)

# Messaggi/UPDATE

- Announcement = prefix + attributes values
- Annuncia nuove reti raggiungibili ed eventualmente l'instradamento
- Annuncia reti precedentemente annunciate non più raggiungibili

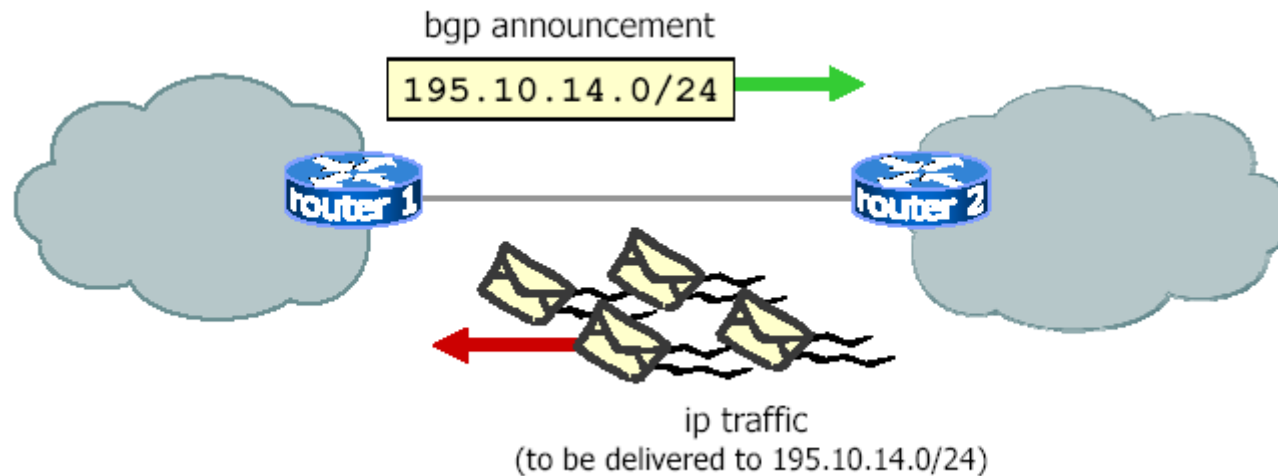
Number of Octets

19	Common Header	Type = 2
2	Unfeasible Routes Length	
Variable	Withdrawn Routes	
2	Total Path Attribute Length	
Variable	Path Attributes	
Variable	Network Layer Reachability Information	

# Annunci BGP

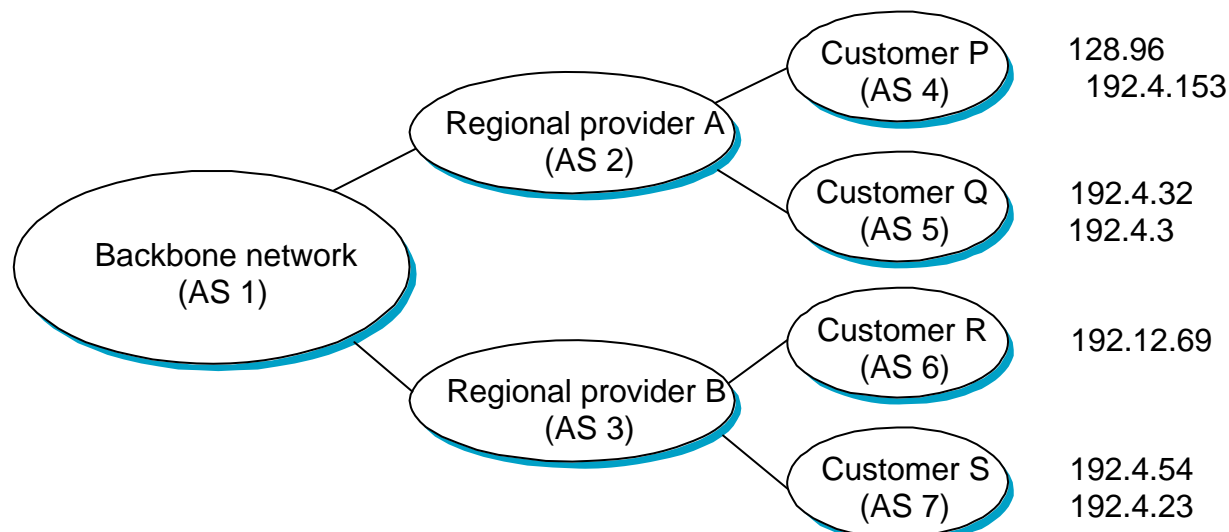
# Annunci BGP

- BGP permette ad un AS di offrire connettività ad un altro AS
- Offrire connettività significa promettere il recapito ad una specifica destinazione
- Destinazione specificata da (Netmask, Prefix)
  - Si adotta convenzione CIDR
- **Annunci BGP in messaggi UPDATE**



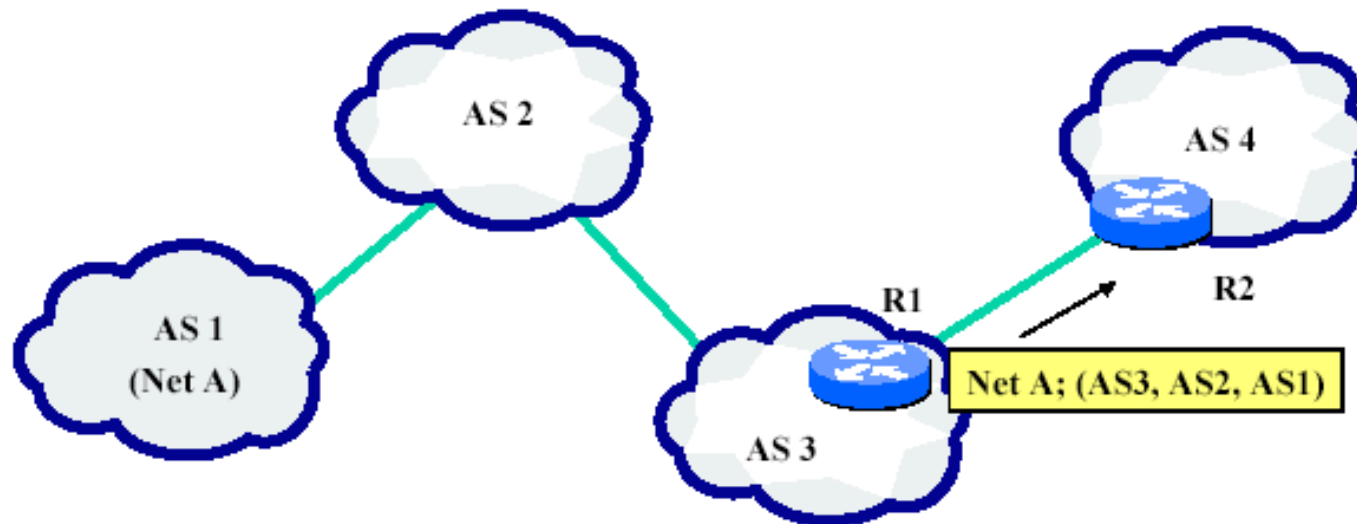
# Route BGP/Path vector

- Speaker for AS2 advertises reachability to P and Q
  - network 128.96, 192.4.153, 192.4.32, and 192.4.3, can be reached directly from AS2
- Speaker for backbone advertises
  - networks 128.96, 192.4.153, 192.4.32, and 192.4.3 can be reached along the path (AS1, AS2).
- Speaker can cancel previously advertised paths



# Path vector/cont.

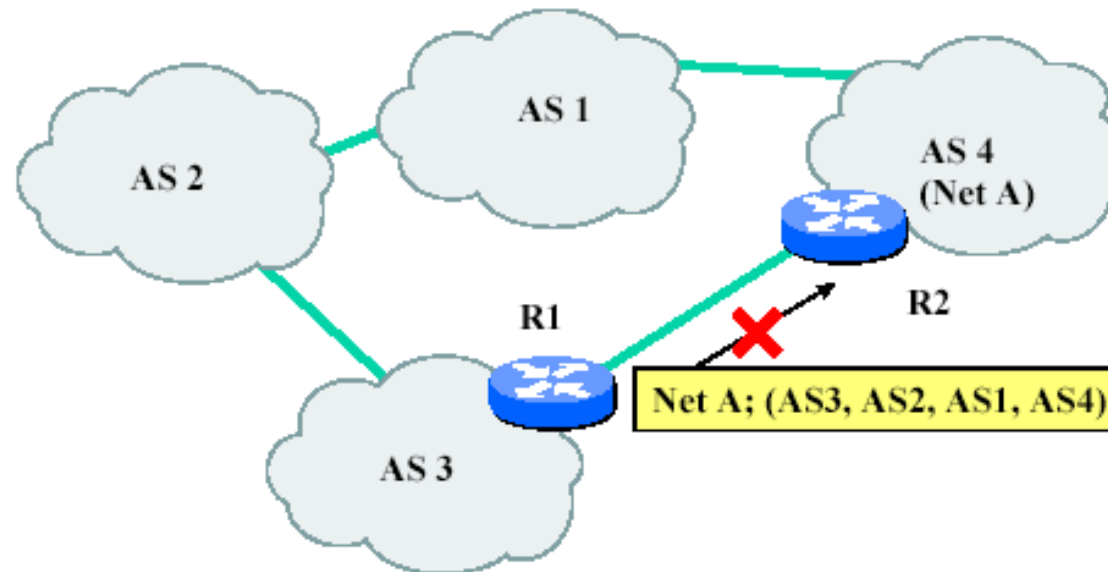
- L'informazione scambiata ha la struttura:  
DestNet:(<lista di AS>)
- Non si tratta in generale di cammini minimi



R1 dice a R2: per raggiungere la net (o subnet) A si passa per AS3, AS2 e AS1

# Gestione dei cicli

- Gli annunci contengono cammini completi
- Necessario che gli asn siano unici



R2 rifiuta la Net A se l'AS *path* associato comprende il proprio AS

# Filtro degli annunci

- Gli annunci sono inviati e/o accettati solo se alcune condizioni sono verificate
- Gli annunci possono essere filtrati sulla base di:
  - Una lista di prefissi validi
  - Una lista di numeri di AS

# Messaggi/UPDATE

- Announcement = prefix + attributes values
- Annuncia nuove reti raggiungibili ed eventualmente l'instradamento
- Annuncia reti precedentemente annunciate non più raggiungibili

Number of Octets

19	Common Header	Type = 2
2	Unfeasible Routes Length	
Variable	Withdrawn Routes	
2	Total Path Attribute Length	
Variable	Path Attributes	
Variable	Network Layer Reachability Information	

# Messaggi di UPDATE/cont.

- Withdrawn routes: lista di coppie <length, IP prefix> delle destinazioni non piu' raggiungibili
  - Length: lunghezza del prefisso in bit
- Network Layer Reachability Information (NLRI): lista di coppie <length, IP prefix> delle destinazioni annunciate
  - Length: lunghezza del prefisso in bit
- Esempio di NLRI:
  - /25, 204.149.16.128
  - /23, 206.134.32
  - /8, 10
- Il valore dell'attributo AS\_PATH e' una successione di sistemi autonomi (route) che permette di raggiungere le destinazioni descritte nella NLRI

# Attributi

- Campo variabile del pacchetto UPDATE
- Attributi sono comuni a tutte le destinazioni annunciate
- Destinazioni con attributi diversi devono essere annunciate con messaggi diversi
- Permette di individuare cicli sugli instradamenti e provenienza dei messaggi
- Categorie
  - **Well-known**: devono essere gestiti da ogni implementazione BGP - se anche **mandatory** devono essere inoltrati, eventualmente dopo modifica
  - **Optional**: non devono essere implementati, possono o meno essere inoltrati

# Path attributes - ogni route

- **AS\_PATH**
  - Elenco sistemi autonomi sul percorso - *well-known*
- **ORIGIN**
  - Origine informazione instradamento - *well-known*
- **NEXT\_HOP**
  - Indirizzo IP salto successivo - *well-known*
- **MED**: Discriminazione tra più punti di uscita AS
  - MED: MULTI\_EXIT\_DISCRIMINATOR - *optional*
  - Importante nella selezione dei cammini (v. piu' avanti)
- **LOCAL\_PREF**: Preferenza all'interno dell' AS
  - Importante nella selezione dei cammini (v. piu' avanti) - *well-known*
- Indicazione di percorsi riuniti
- ID dell' AS che ha aggregato i percorsi

# Attributo ORIGIN

- Definisce l'origine dell'informazione annunciata
- Può essere IGP, EGP o INCOMPLETE

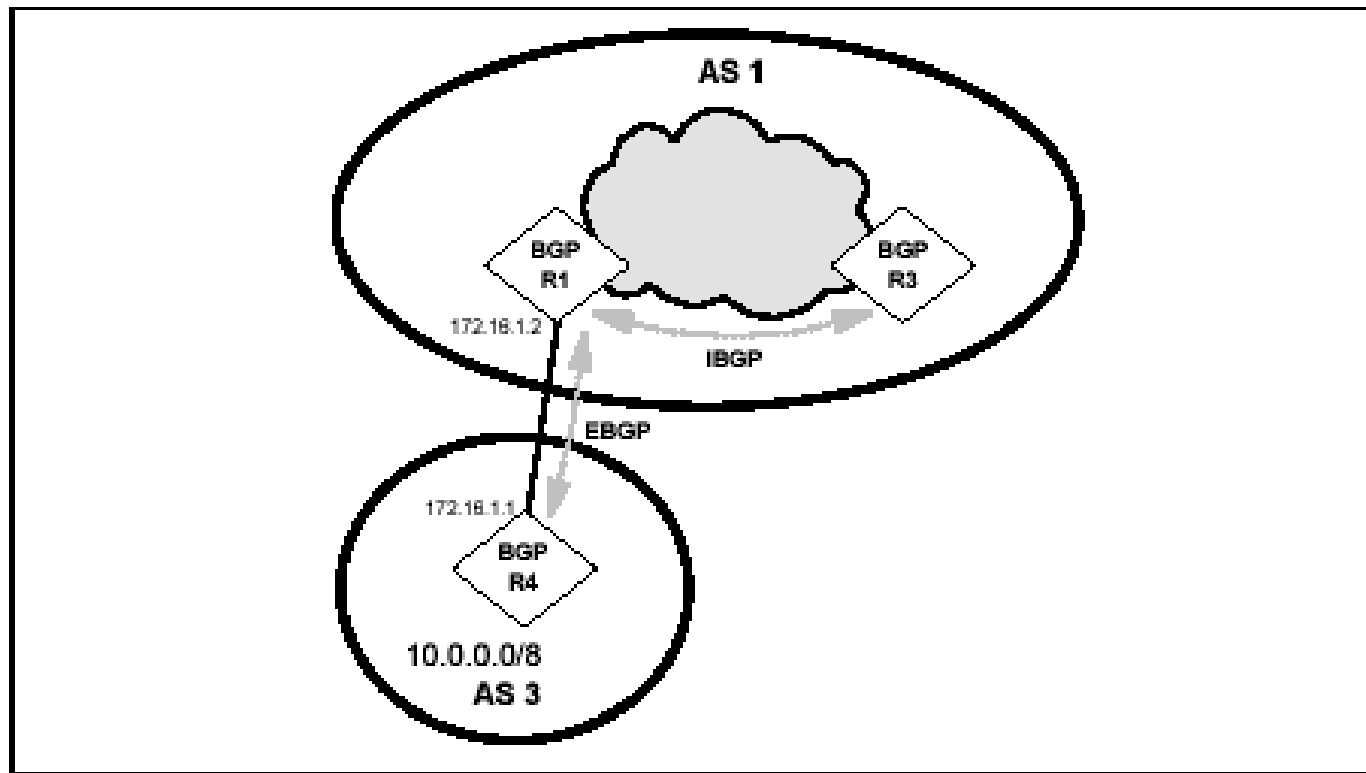
**INCOMPLETE:** si ha nel caso in cui le reti annunciate siano state inserite come route statiche nello speaker che invia l'annuncio

Number of Octets

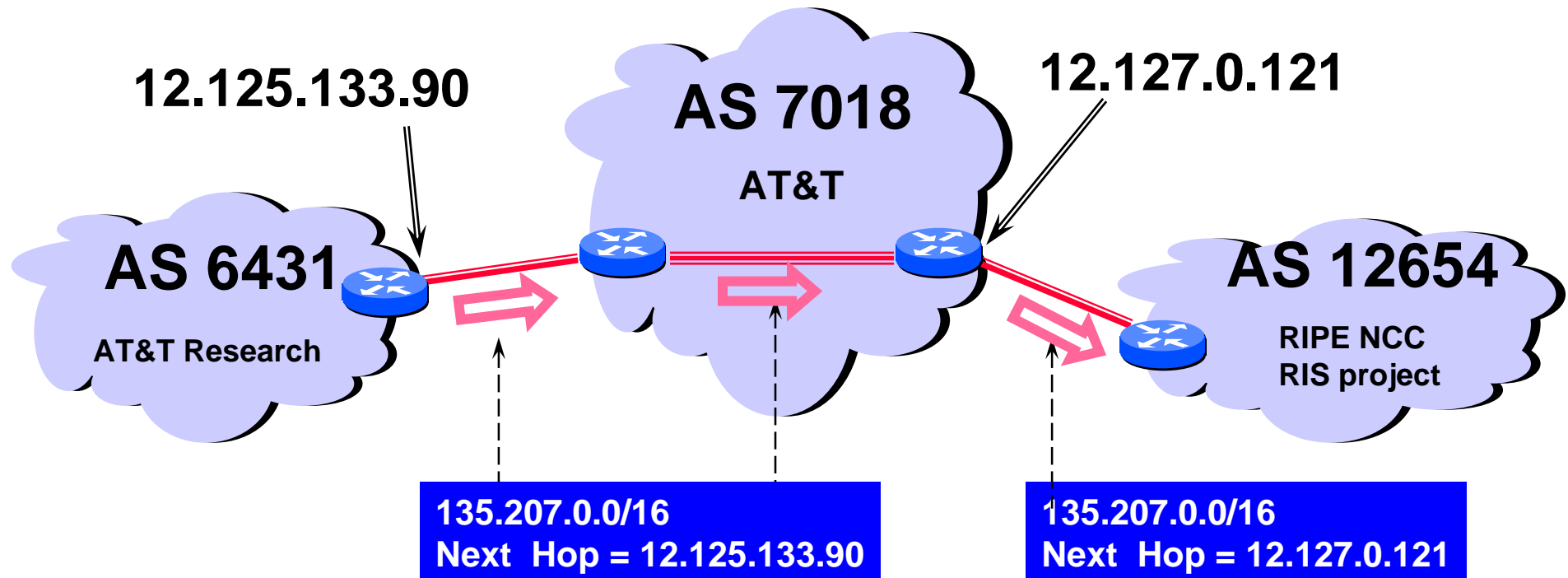
19	Common Header	Type = 2
2	Unfeasible Routes Length	
Variable	Withdrawn Routes	
2	Total Path Attribute Length	
Variable	Path Attributes	
Variable	Network Layer Reachability Information	

# Attributo NEXT\_HOP

- Indirizzo IP del next-hop nella sequenza degli AS
- Per la rete 10.0.0.0/8 R1 invia 172.16.1.1 a R3 come next hop
- R3 deve avere una route verso 172.16.1.1



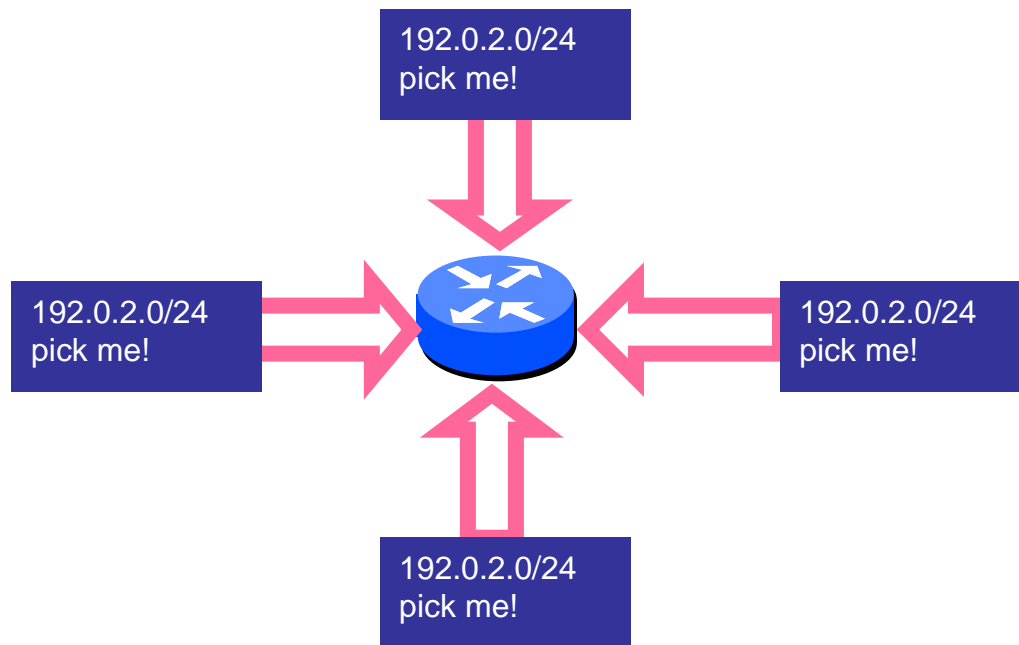
# BGP Next Hop Attribute



**Every time a route announcement crosses an AS boundary, the Next Hop attribute is changed to the IP address of the border router that announced the route.**

# Selezione dei cammini

# Attributes are Used to Select Best Routes



Given multiple routes to the same prefix, a BGP speaker must pick at most one best route

(Note: it could reject them all!)

# Route Selection Summary



**Highest Local Preference**

**Enforce relationships**

**Shortest AS\_PATH**

**Lowest MED**

**i-BGP < e-BGP**

**Lowest IGP cost  
to BGP egress**

**traffic engineering**

**Lowest router ID**

**Throw up hands and  
break ties**

# BGP Route Processing

Open ended programming.  
Constrained only by vendor configuration language

Receive  
BGP  
Updates

Apply Policy =  
filter routes &  
tweak attributes

Based on  
Attribute  
Values

Best  
Routes

Apply Policy =  
filter routes &  
tweak attributes

Transmit  
BGP  
Updates

Apply Import  
Policies

Best Route  
Selection

Best Route  
Table

Apply Export  
Policies

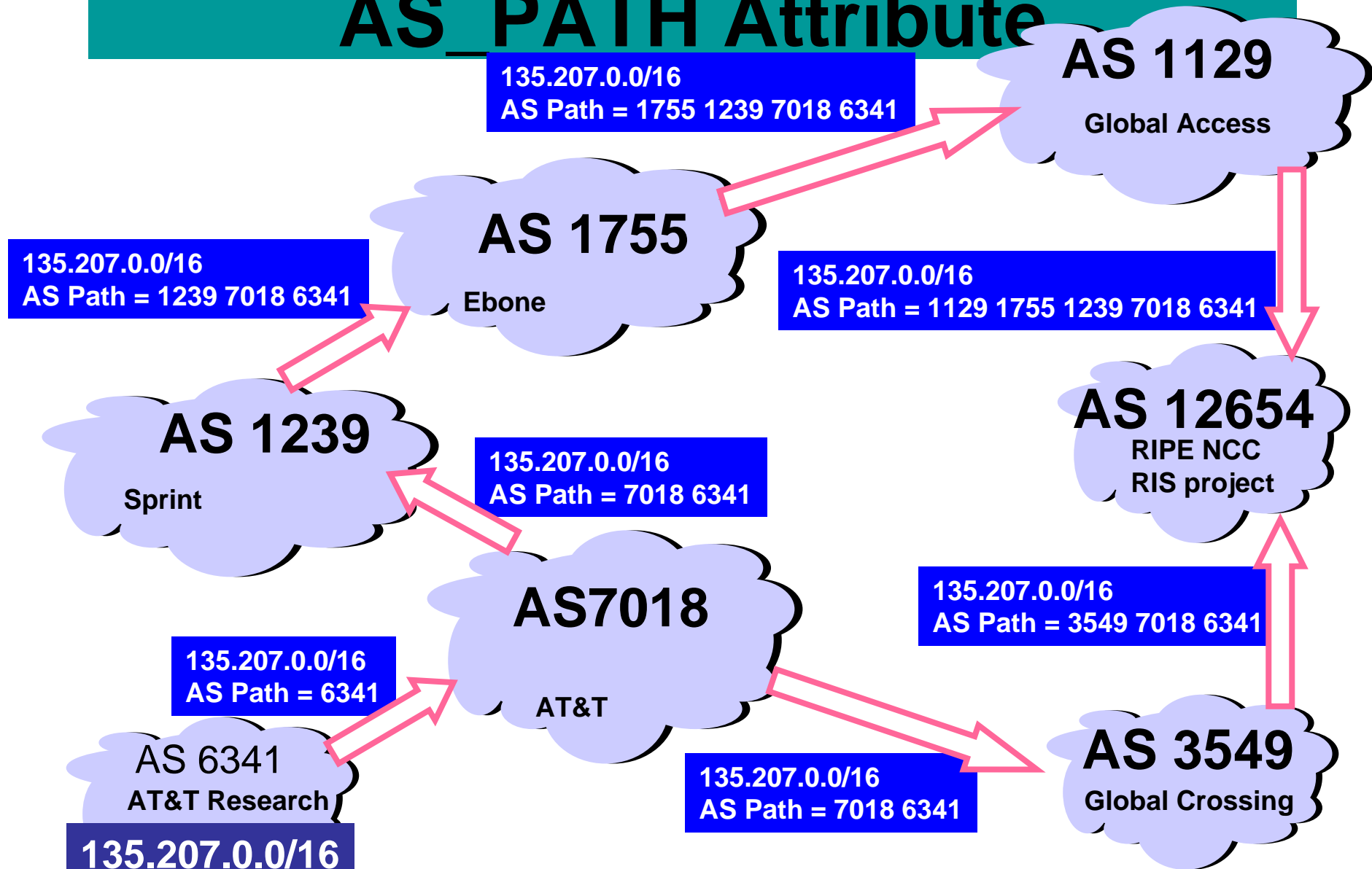
Install forwarding  
Entries for best  
Routes.

IP Forwarding Table

# Filtri sugli annunci/COMMUNITY

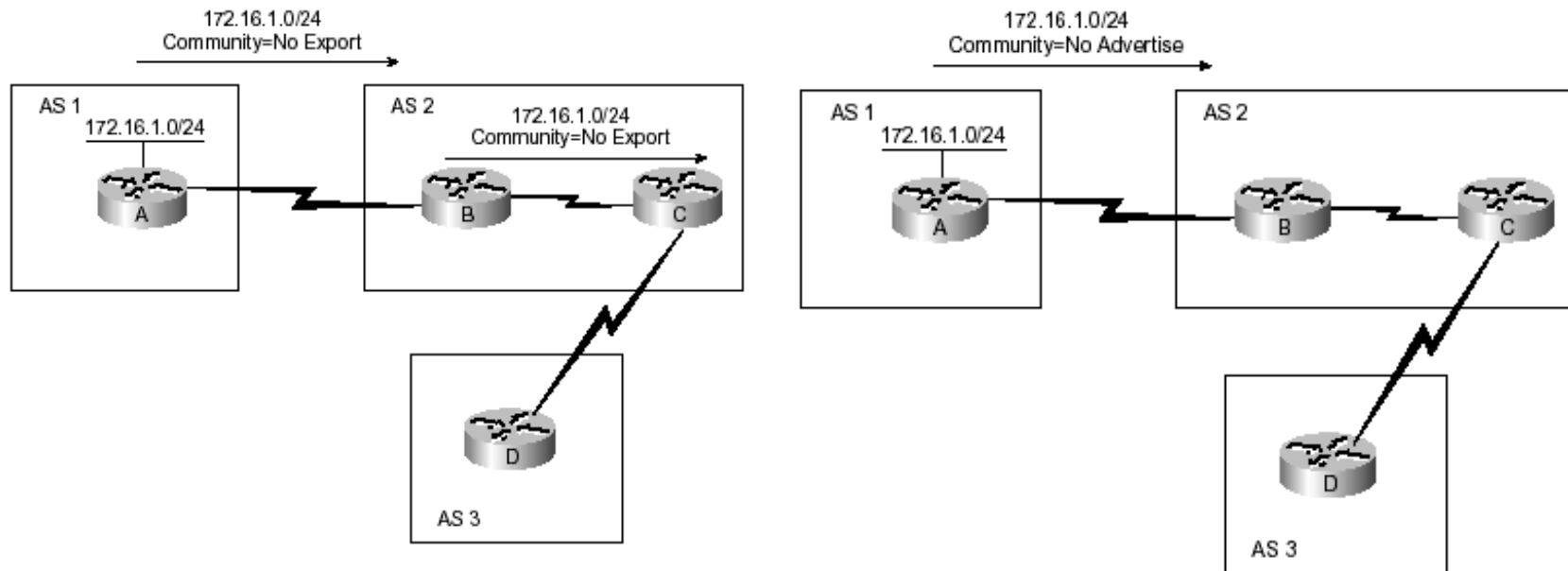
- E' possibile definire gruppi di destinazioni alle quali applicare una comune politica di inoltro
  - Tipo di politica definita dall'attributo **COMMUNITY**
- Valori predefiniti per **COMMUNITY**
  - **No-export**: la route non va annunciata ai peer BGP
  - **No-advertise**: la route non va annunciata a nessun peer
  - **Internet**: la route va annunciata a ogni peer

# AS\_PATH Attribute



Prefix Originated

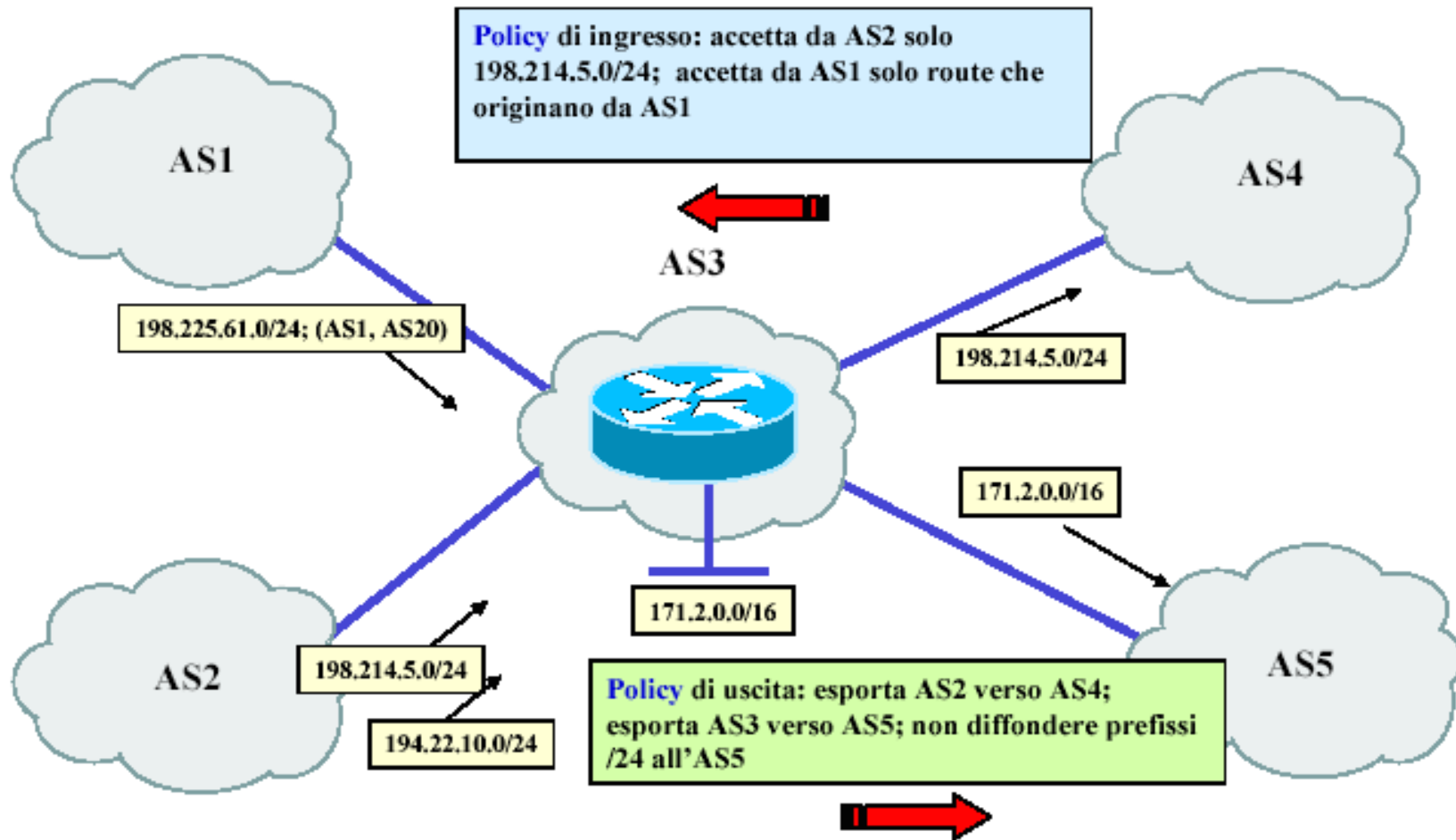
# COMMUNITY



- Se COMMUNITY = export allora C annuncia la route a D

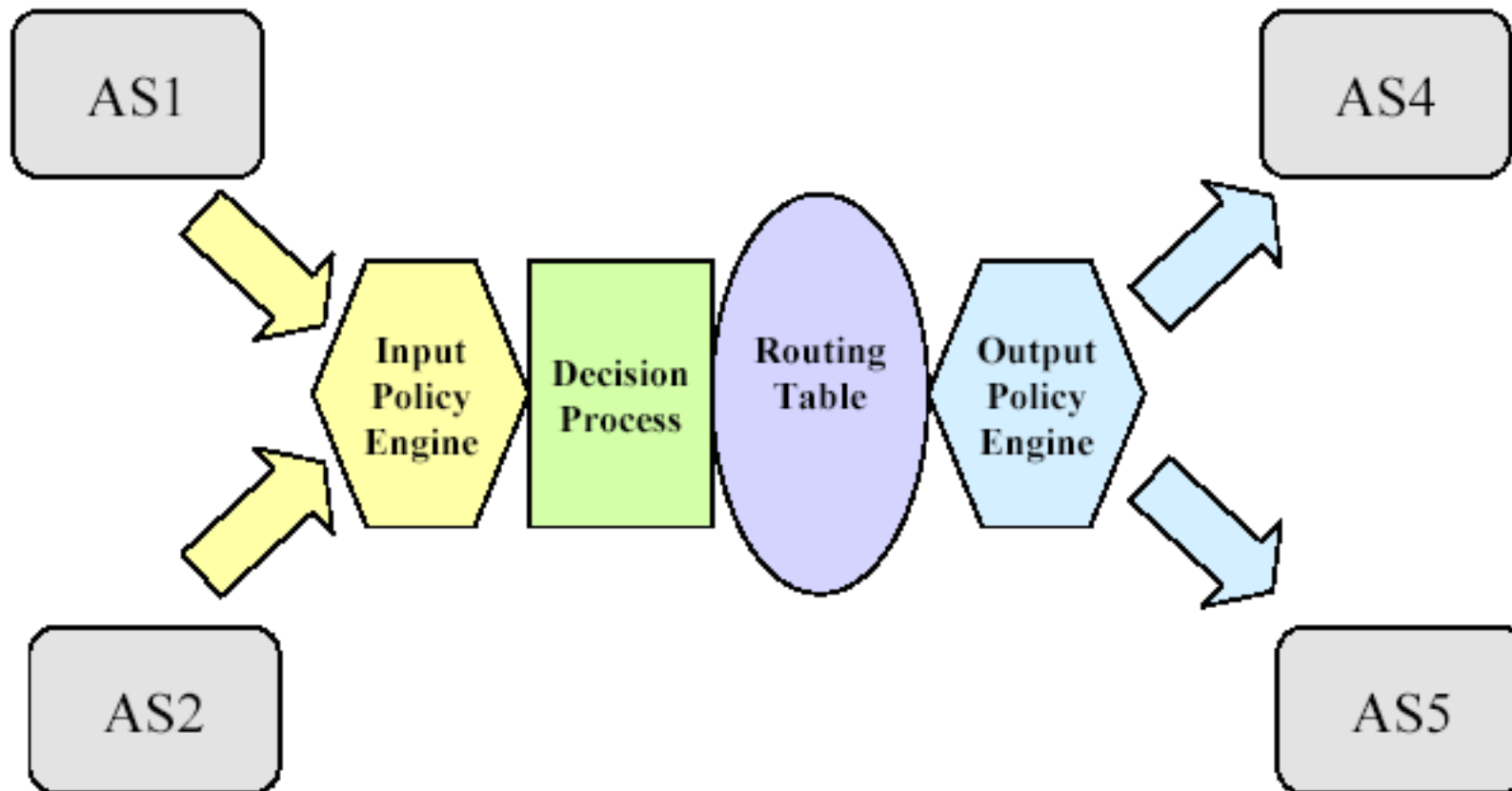
# BGP - Politiche di instradamento

- Amministratore fissa politiche di uscita/ingresso



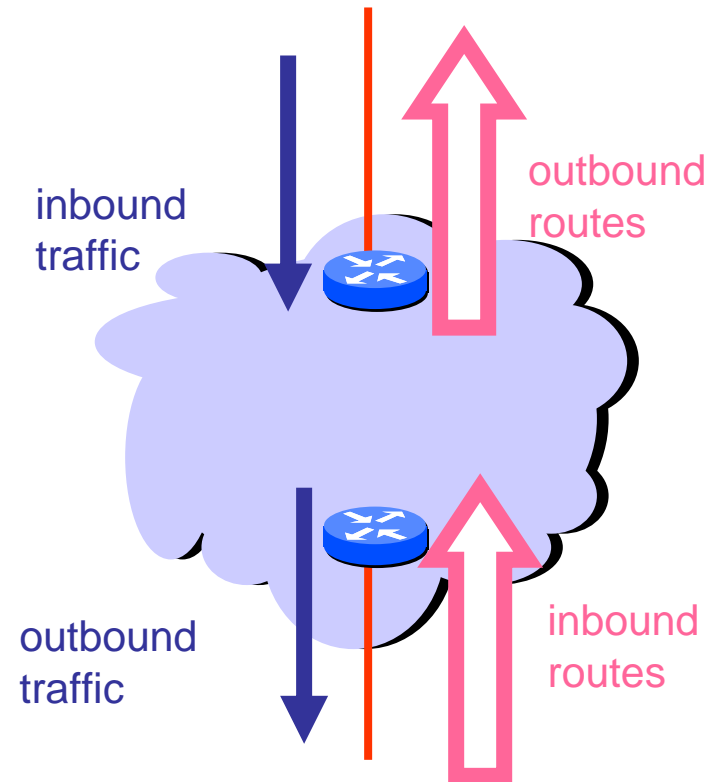
# BGP - Politiche di instradamento

- Schema architetturale



# Tweak Tweak Tweak

- For inbound traffic
  - Filter outbound routes
  - Tweak attributes on outbound routes in the hope of influencing your neighbor's best route selection
- For outbound traffic
  - Filter inbound routes
  - Tweak attributes on inbound routes to influence best route selection

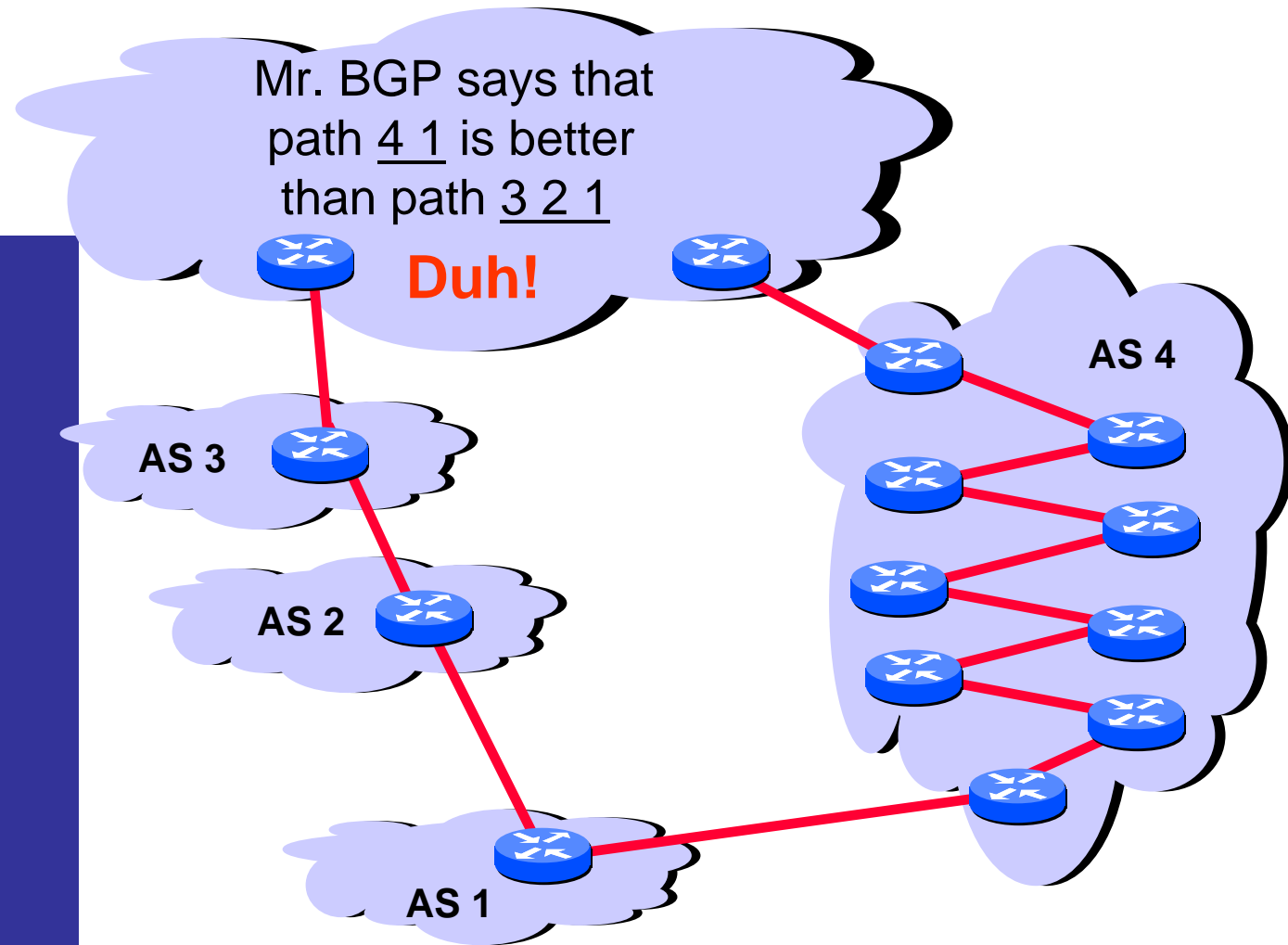


In general, an AS has more control over outbound traffic

# Shorter Doesn't Always Mean Shorter

In fairness:  
could you do  
this “right” and  
still scale?

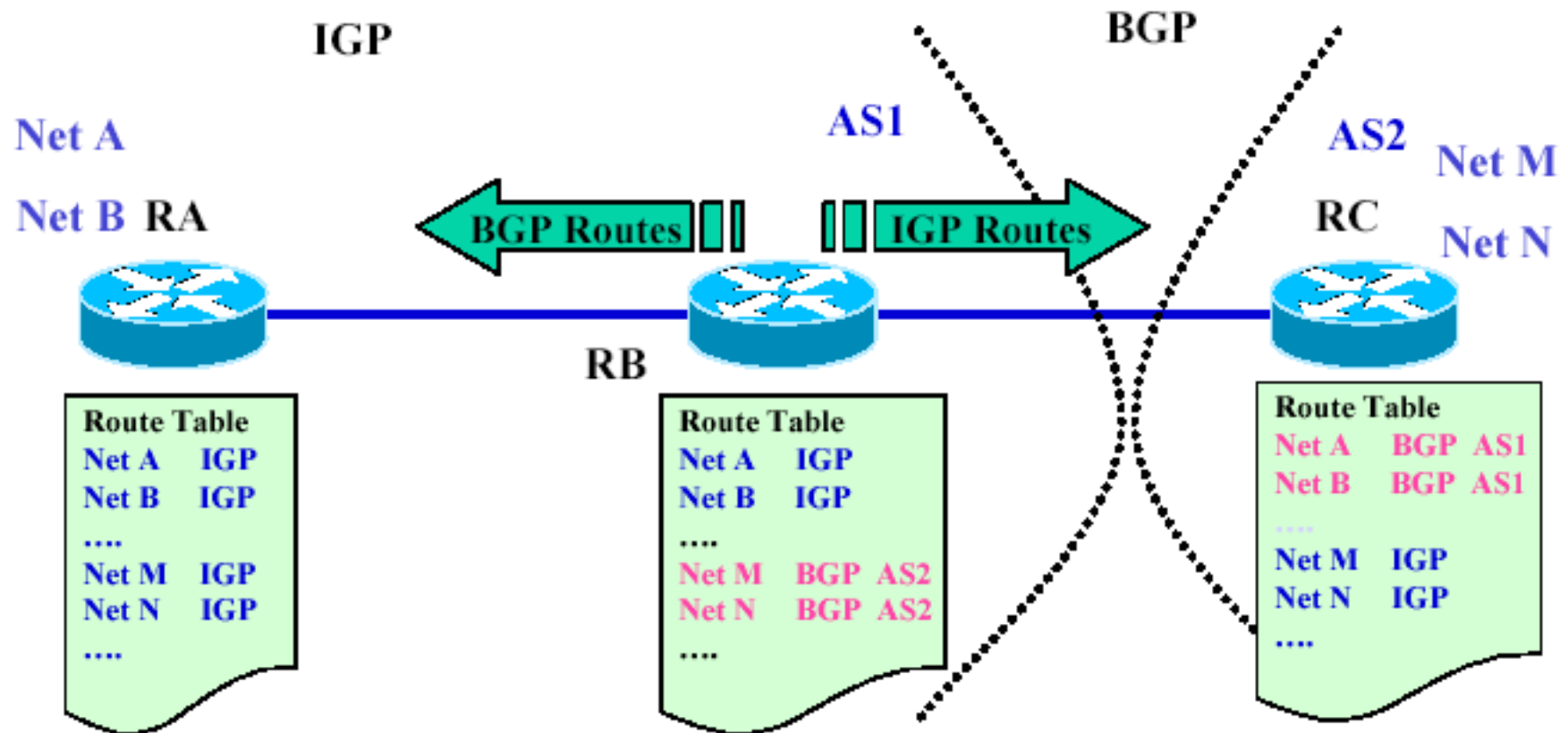
Exporting internal  
state would  
dramatically  
increase global  
instability and  
amount of routing  
state



# Interazione con IGP

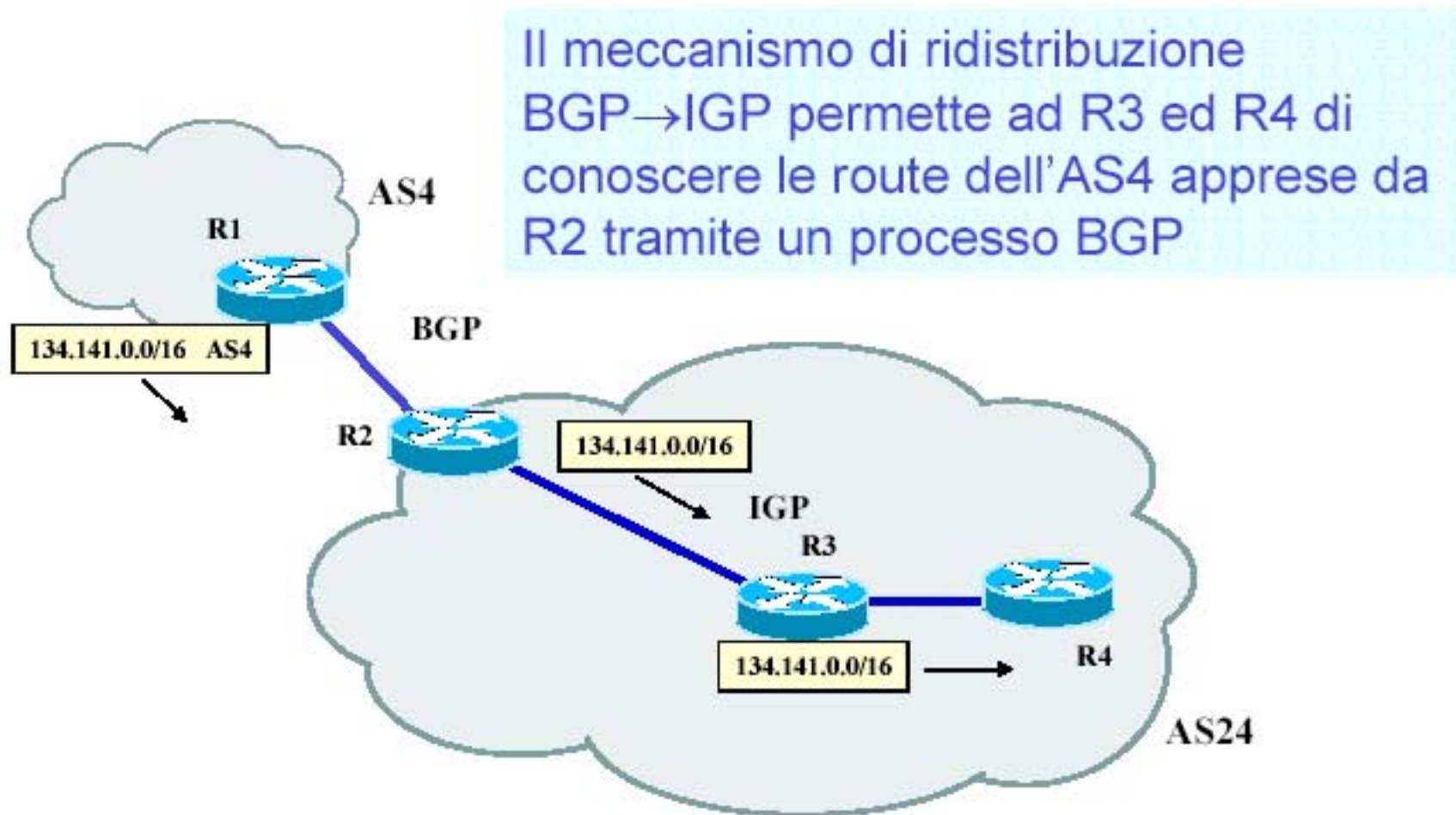
# Interazione con IGP

- Router di bordo esegue sia BGP che IGP

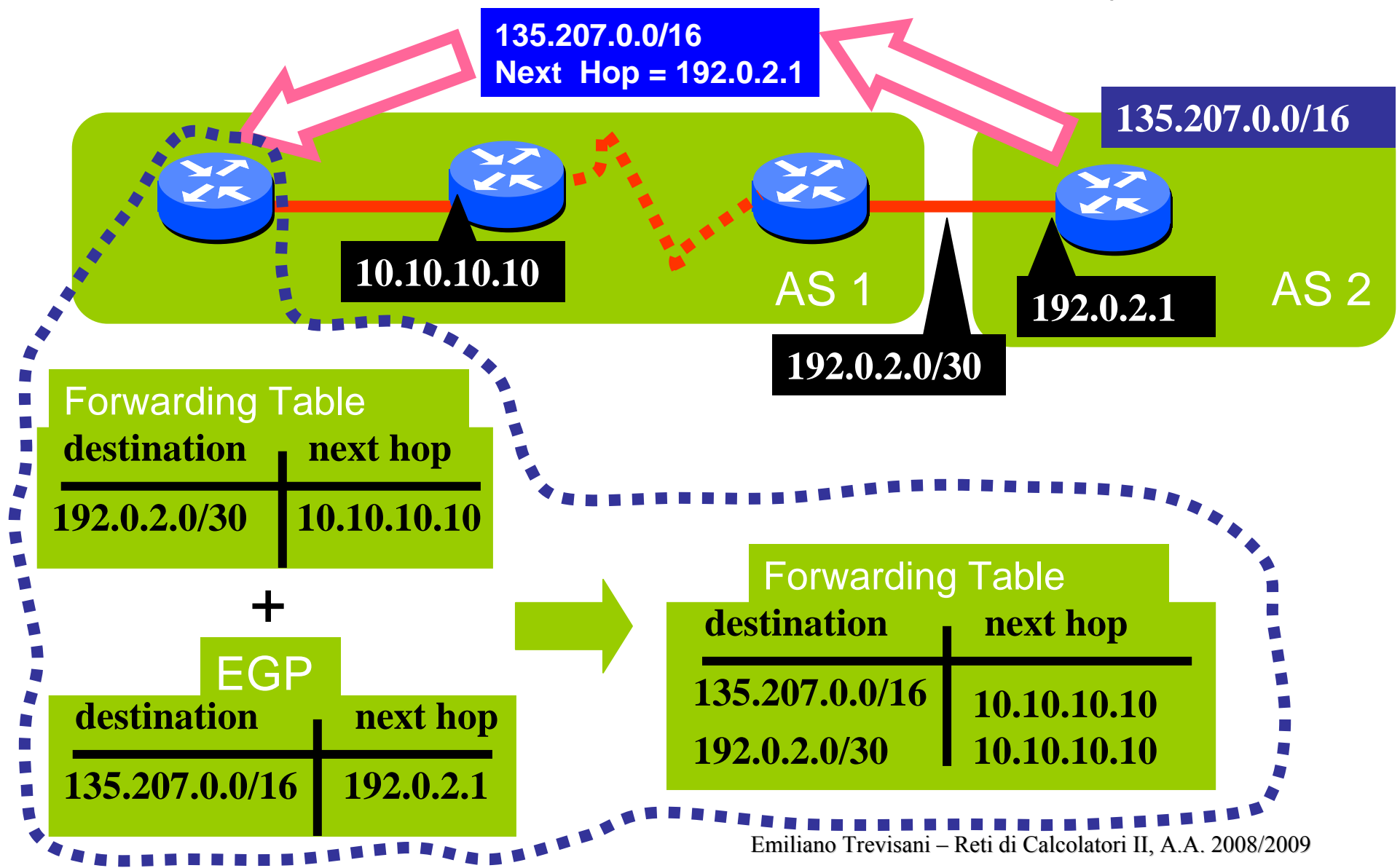


# Interazione con IGP

- Router di bordo esegue sia BGP che IGP



# Join EGP with IGP For Connectivity



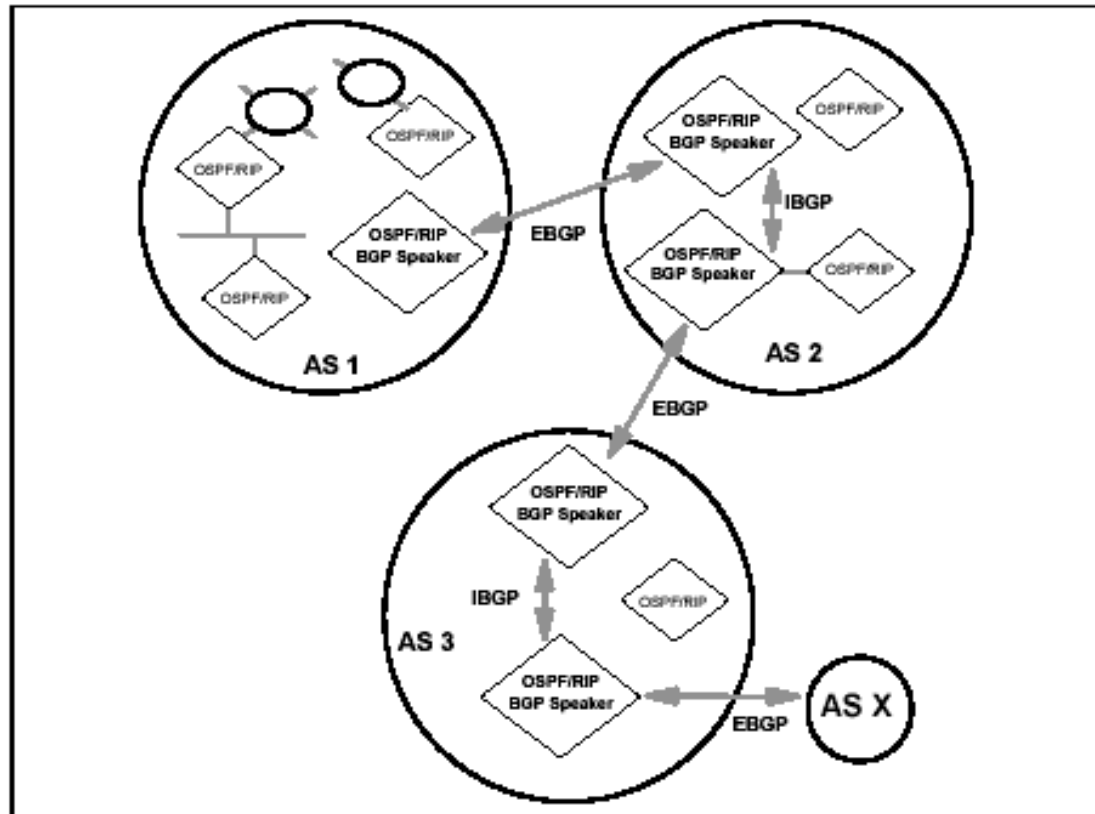
# Limiti di BGP e soluzioni

# Limiti di BGP

- BGP non può discriminare tra due percorsi sulla base della distanza o della congestione
- BGP sceglie uno dei due percorsi possibili non sulla base di una metrica di costo
- BGP permette di suddividere il carico attraverso la rete ma non in modo dinamico
- Occorre configurare manualmente quali reti sono annunciate da quali router esterni
- Tutti i sistemi autonomi devono concordare su uno schema coerente per annunciare la raggiungibilità

# Limiti di BGP/2

- Se AS2 non inoltra ad AS1 l'informazione sull'instradamento ricevuta da AS3, quest'ultimo ed ASX non saranno raggiungibili da AS1



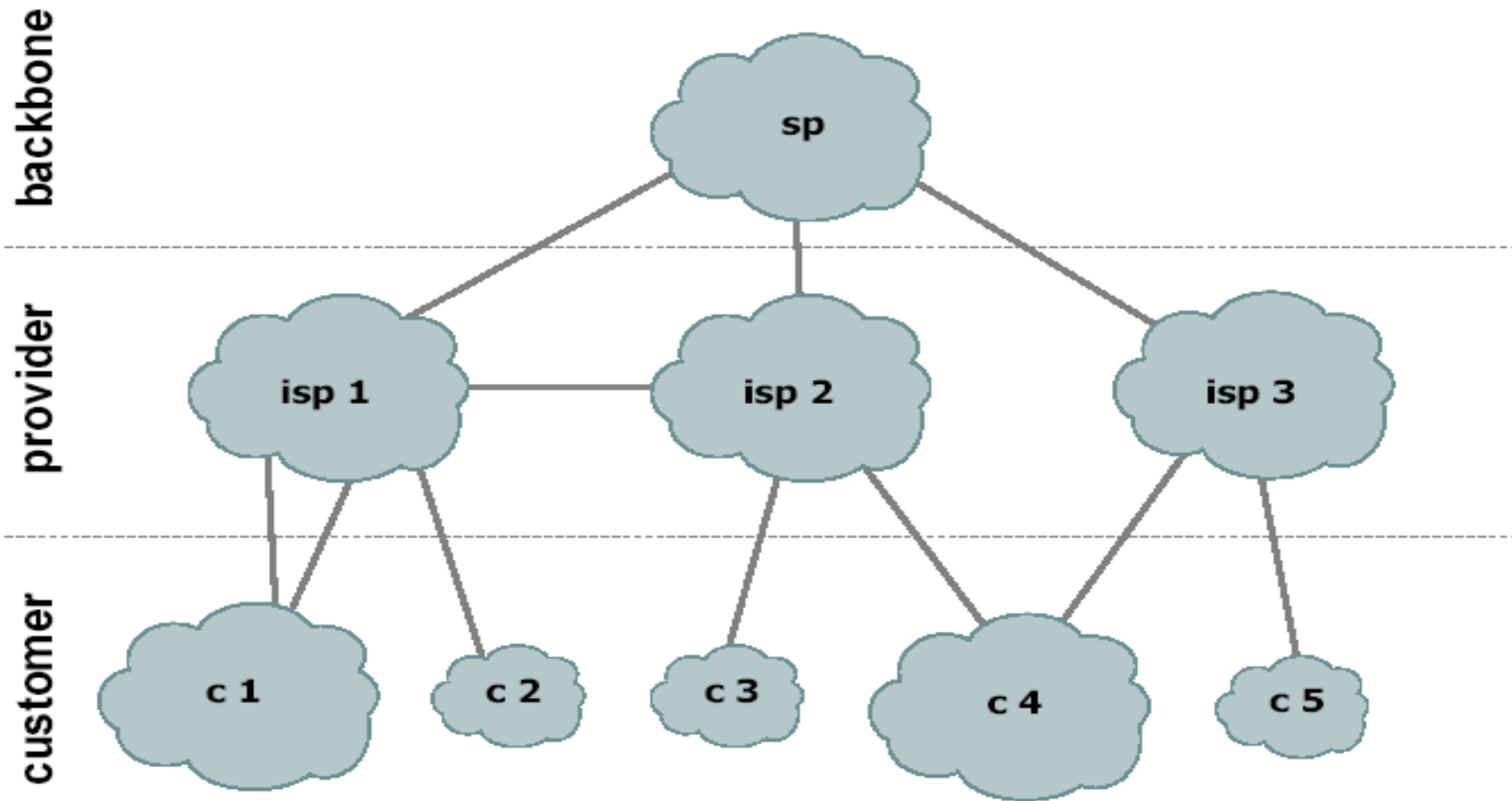
# Instradamento con arbitraggio

- Occorre un sistema per garantire la coerenza sulle informazioni di instradamento
- Database autenticato e replicato che contiene le informazioni sulla raggiungibilità
- Autenticazione: solo AS autenticati possono annunciare la raggiungibilità di una rete
- NAP sono i router di interconnessione tra ISP
- I NAP hanno un Router Server che mantiene il data base BGP ma non sono necessariamente speaker BGP
- Gli speaker BGP mantengono aperto un collegamento verso il Router Server

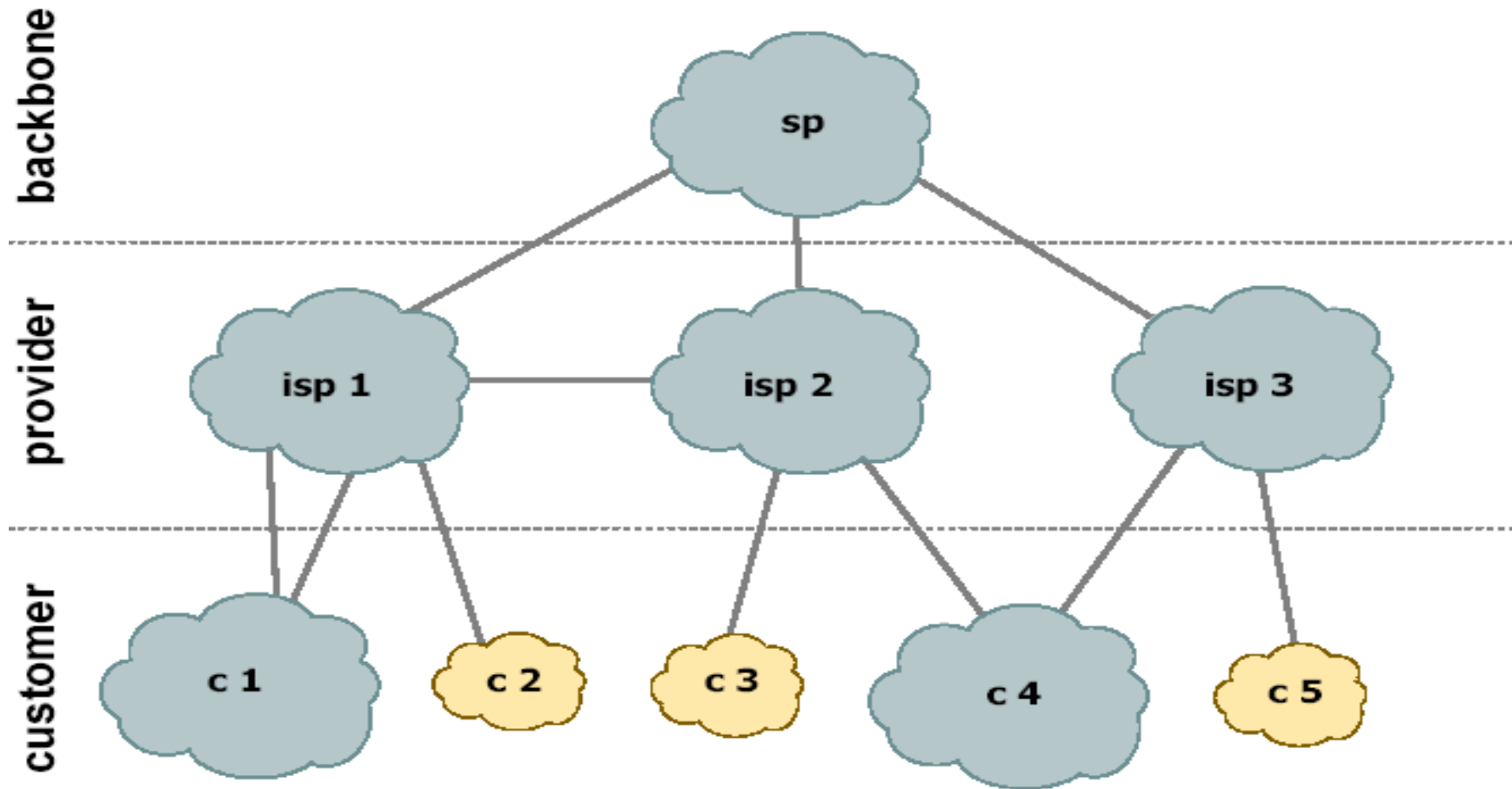
# Esempi di architetture BGP

## AS STUB e Multi-Homed

# Uno scenario BGP complesso



# Stub network

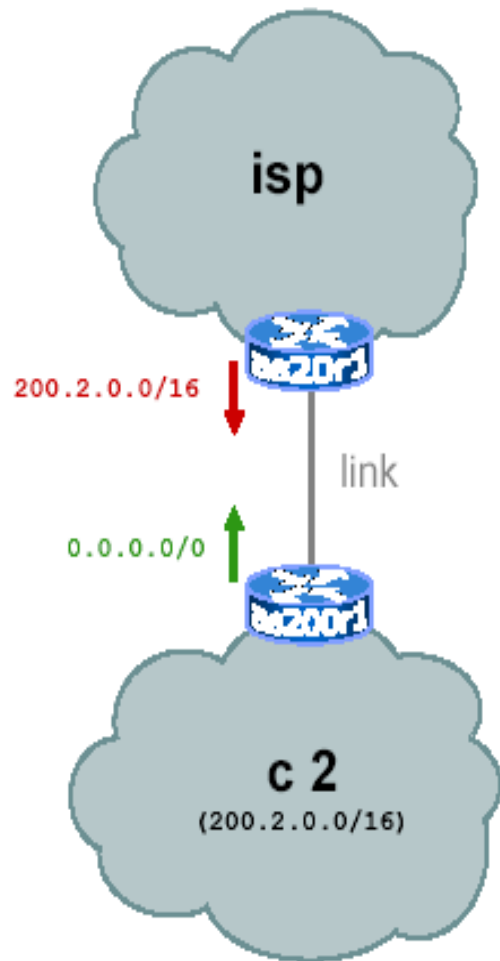


# Stub network, architettura



- Un router della rete è scelto come gateway di default e connesso ad un singolo router dell'isp con una o più connessioni
- Una singola sessione di peering in cui as200 annuncia la sua raggiungibilità e accetta l'instradamento di default sul router

# Instradamento statico per stub network

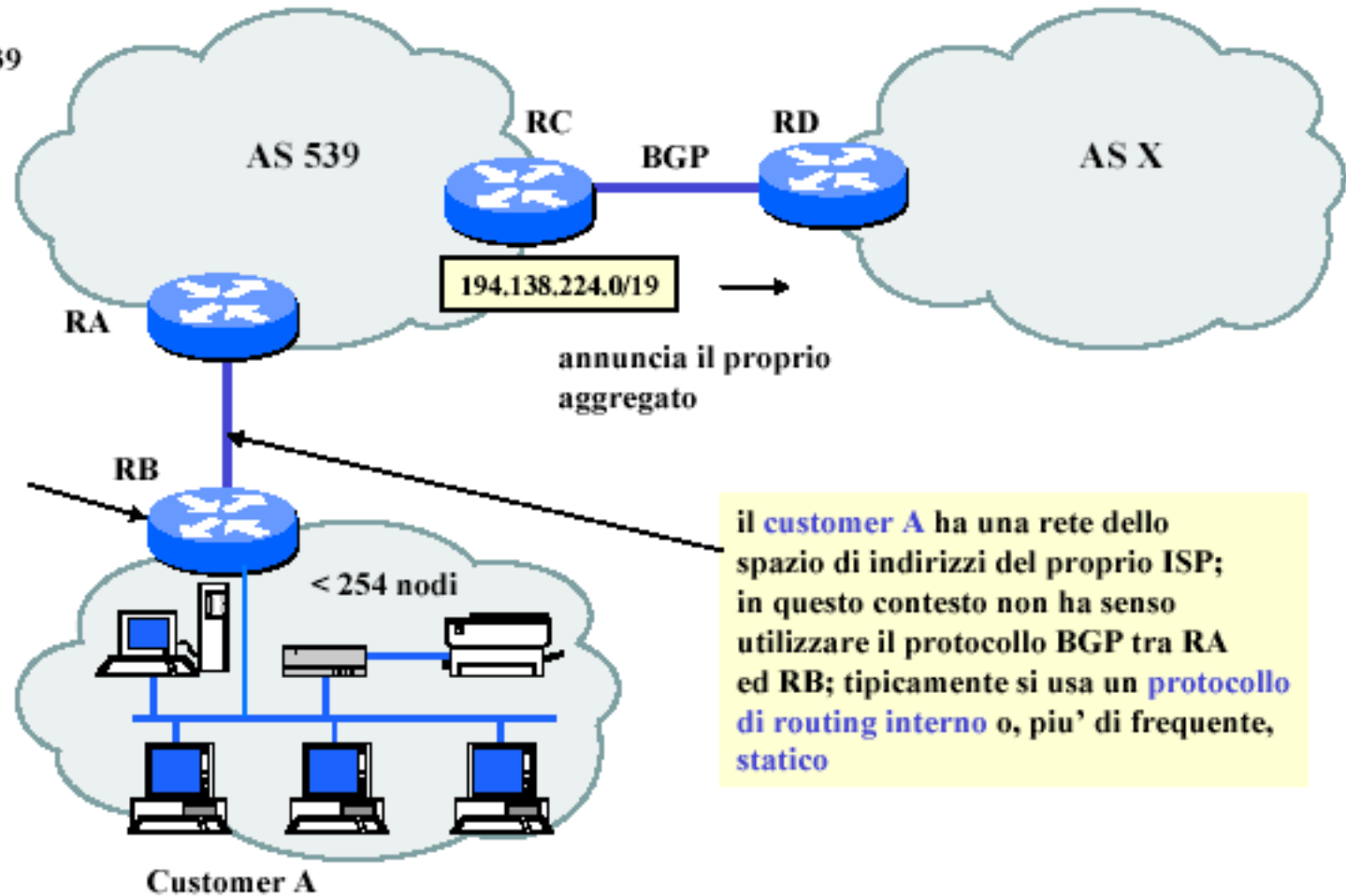


- Un instradamento statico di default è sufficiente per i pacchetti in uscita per essere inviati su internet attraverso la connessione all'isp
- Un instradamento statico è anche sufficiente per i pacchetti in ingresso per raggiungere la rete attraverso la connessione all'isp
- Non vi è alcun bisogno di BGP

# Esempio

blocco di reti assegnato dal RIPE NCC all'AS539

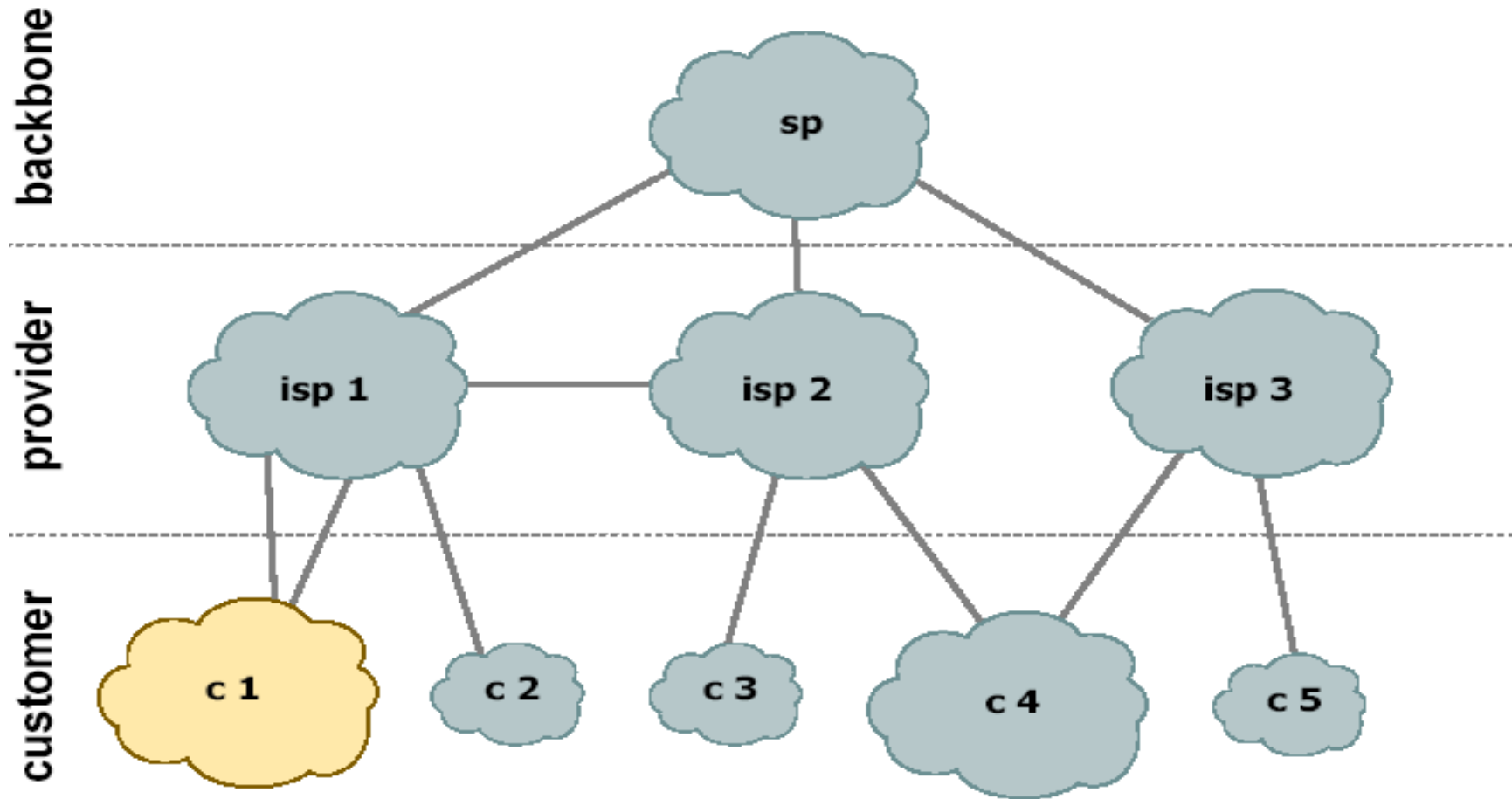
194.138.224.0/19



blocco di reti assegnato dall'ISP al proprio cliente

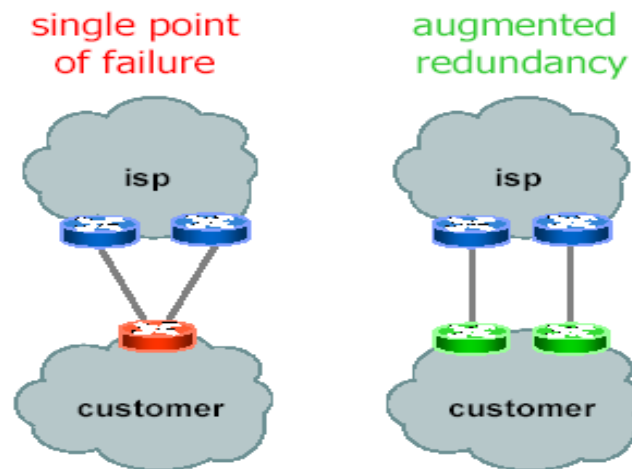
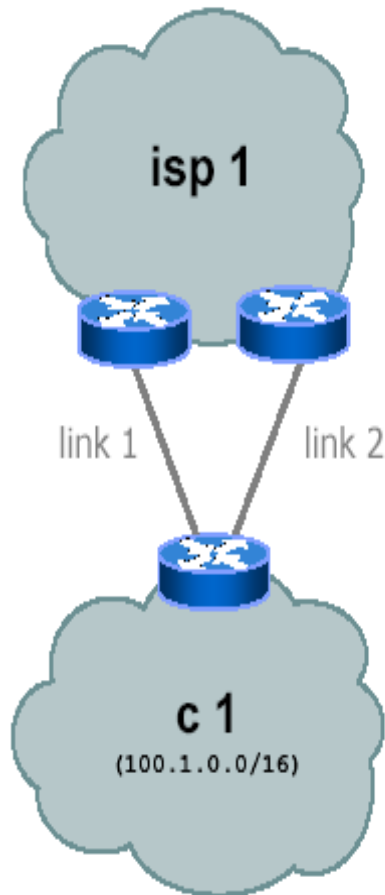
194.138.230.0/24

# Multi-homed stub networks

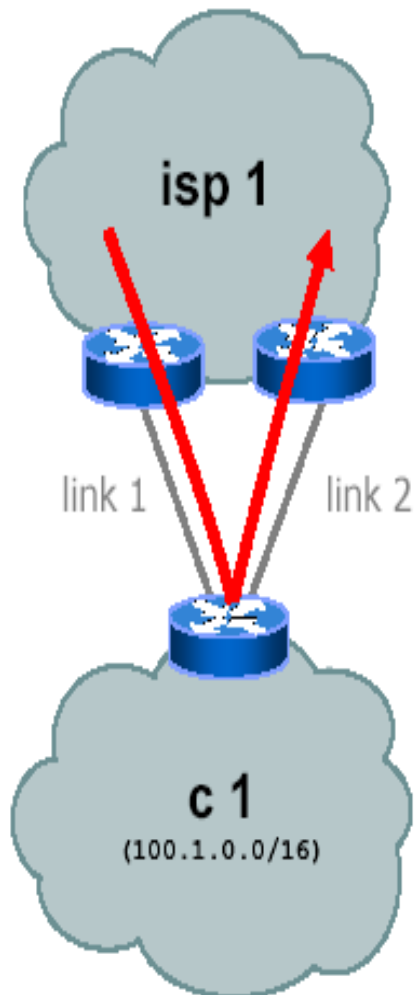


# Multi-homed stub networks

- Due collegamenti allo stesso isp
- Due router della rete customer sono di solito coinvolti

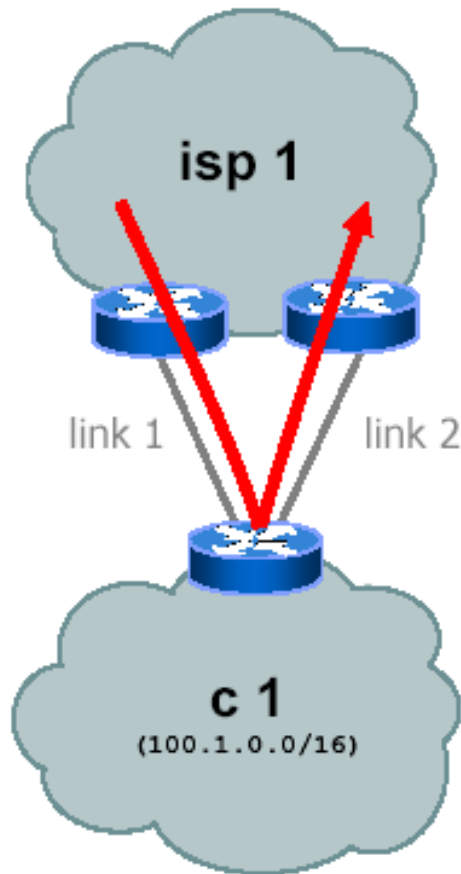


# Instradamento



- Un pacchetto diretto ad Internet può attraversare uno dei due link
- Un pacchetto proveniente da Internet può attraversare uno dei due link
- Un pacchetto in transito può attraversare entrambi i link
  - Non dovrebbe capitare negli stub

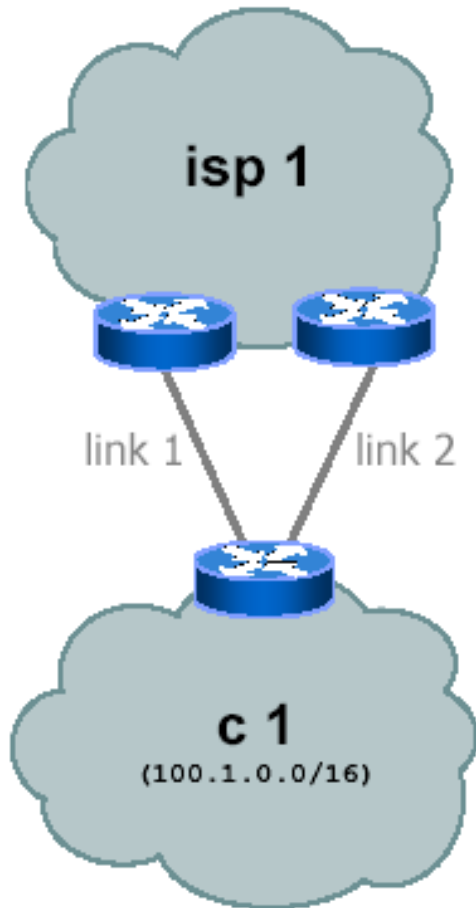
# Politiche desiderate - Backup



Esempio:

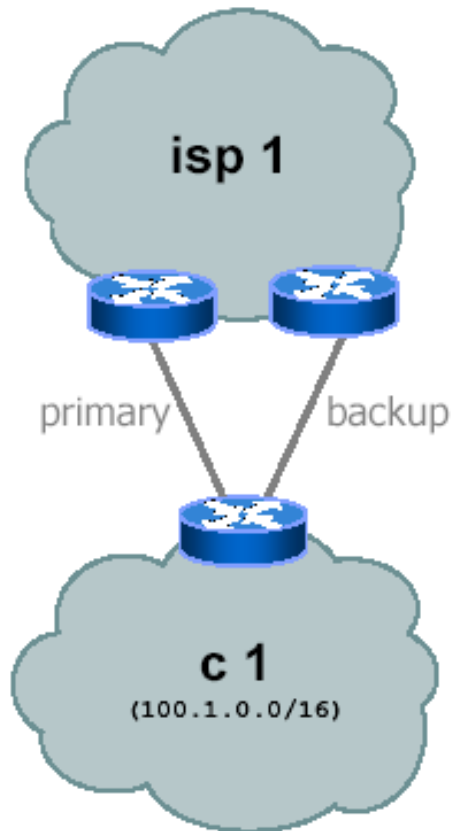
- Eliminare traffico in transito
- Traffico in ingresso:
  - Utilizzare link 1
  - Utilizzare link 2 in caso di fault su link 1
- Traffico in uscita:
  - Utilizzare link 1
  - Utilizzare link 2 in caso di fault su link 1

# Alternative a BGP



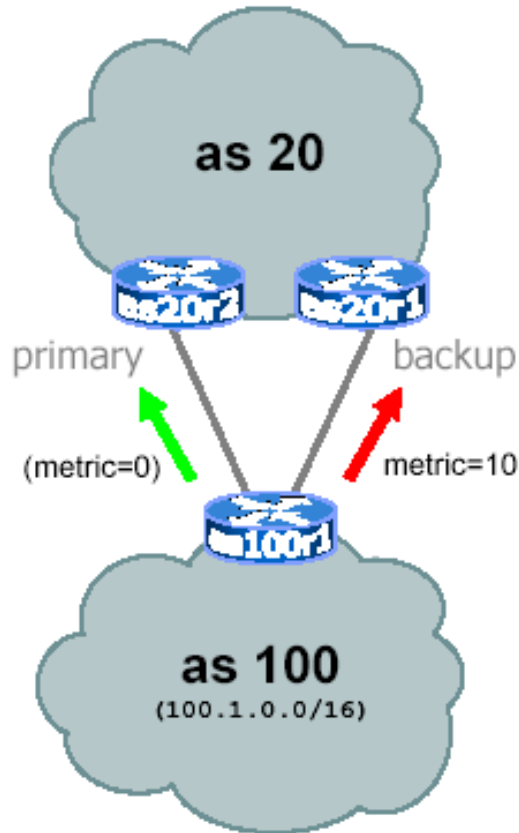
- Usare un igp:
  - Pacchetti usano link 1 o link 2 a seconda dello shortest path verso c1
  - Non è possibile escludere pacchetti in transito quando link 1 e link 2 sono sul cammino minimo tra sorgente e destinazione
- Usare cammini statici:
  - I router dell'isp e la rete devono essere configurati manualmente in modo coerente.
  - Non è possibile gestire un meccanismo di backup automatico

# La strategia usata da BGP



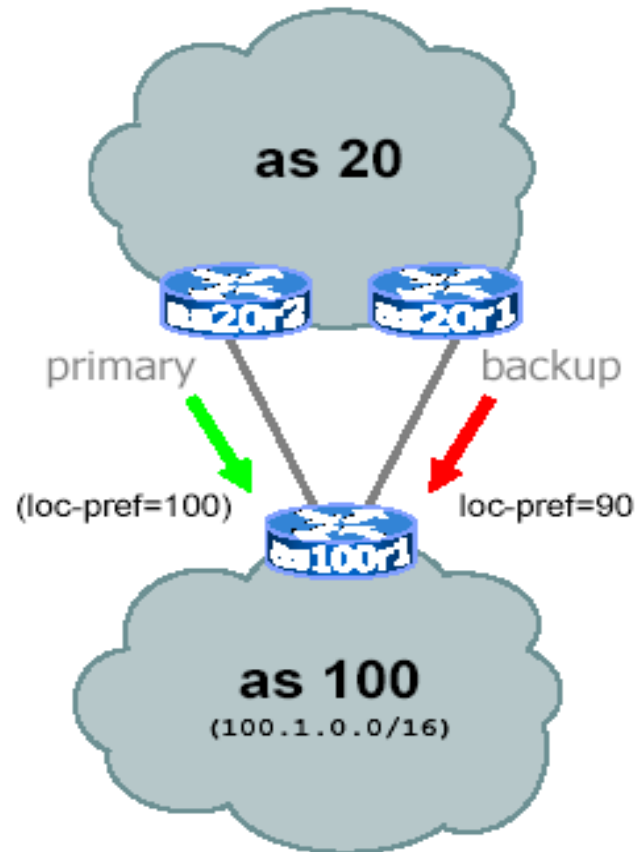
- Annuncio 100.1.0.0/16 aggregato su ogni arco:
  - Link primario invia un announcement standard
  - Il link di backup aumenta il MED sugli annunci in uscita e riduce la LOCAL\_PREF sugli annunci in ingresso
  - MED: MULTI\_EXIT\_DISCRIMINATOR
- Quando avviene un fault su un link, l'annuncio del /16 aggregato sull'altro link assicura la connettività

# Strategia BGP/MED



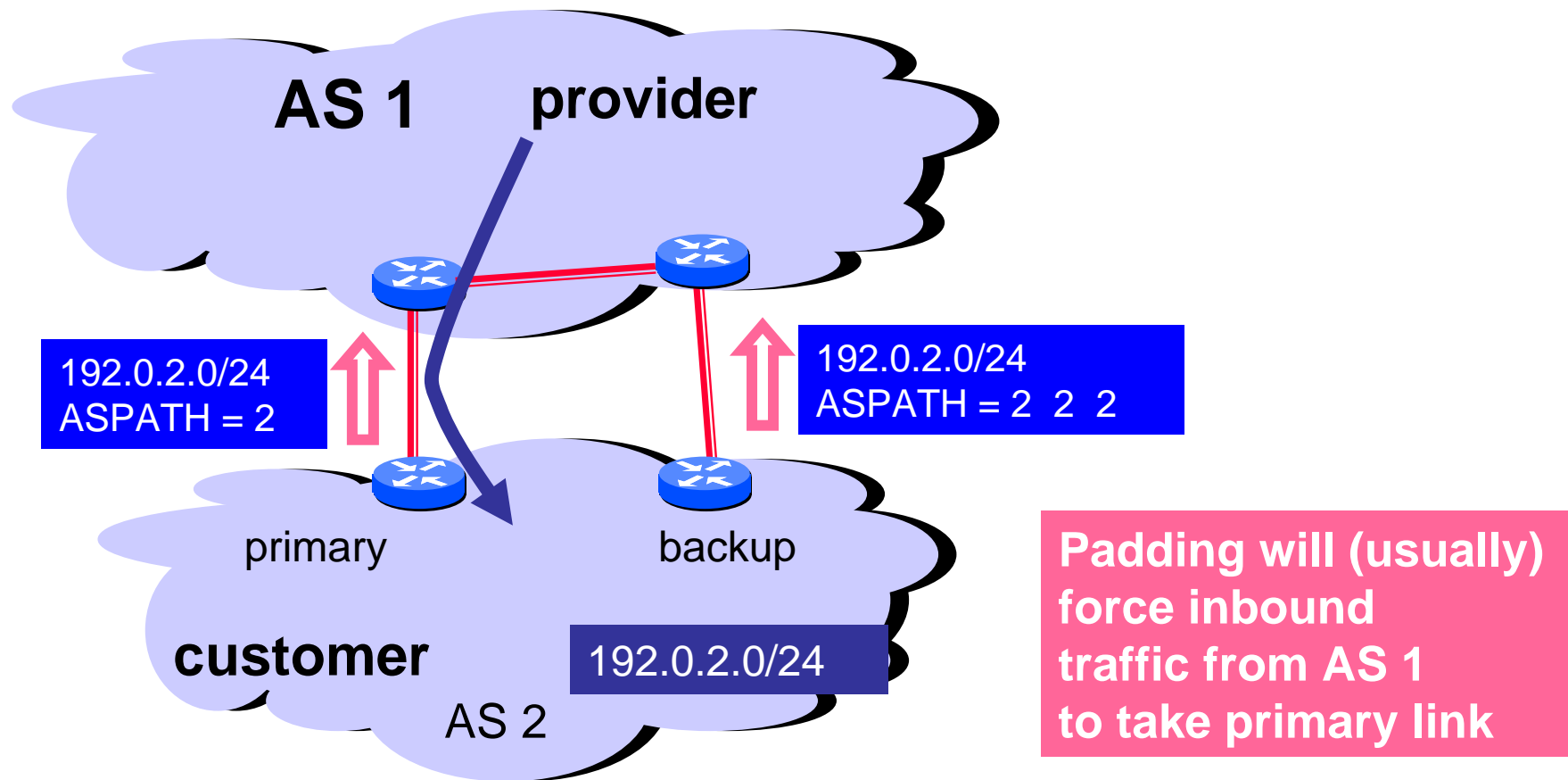
- the attribute called "metric" by the sender as, is called "multi-exit-discriminator" by the receiver as
- upon reception of the same announcement with two different meds, the provider will (hopefully) adopt the one with the smaller one
- default value is zero
- metric is set on outgoing announcements and manages inbound traffic flows

# Strategia BGP/Local Preference

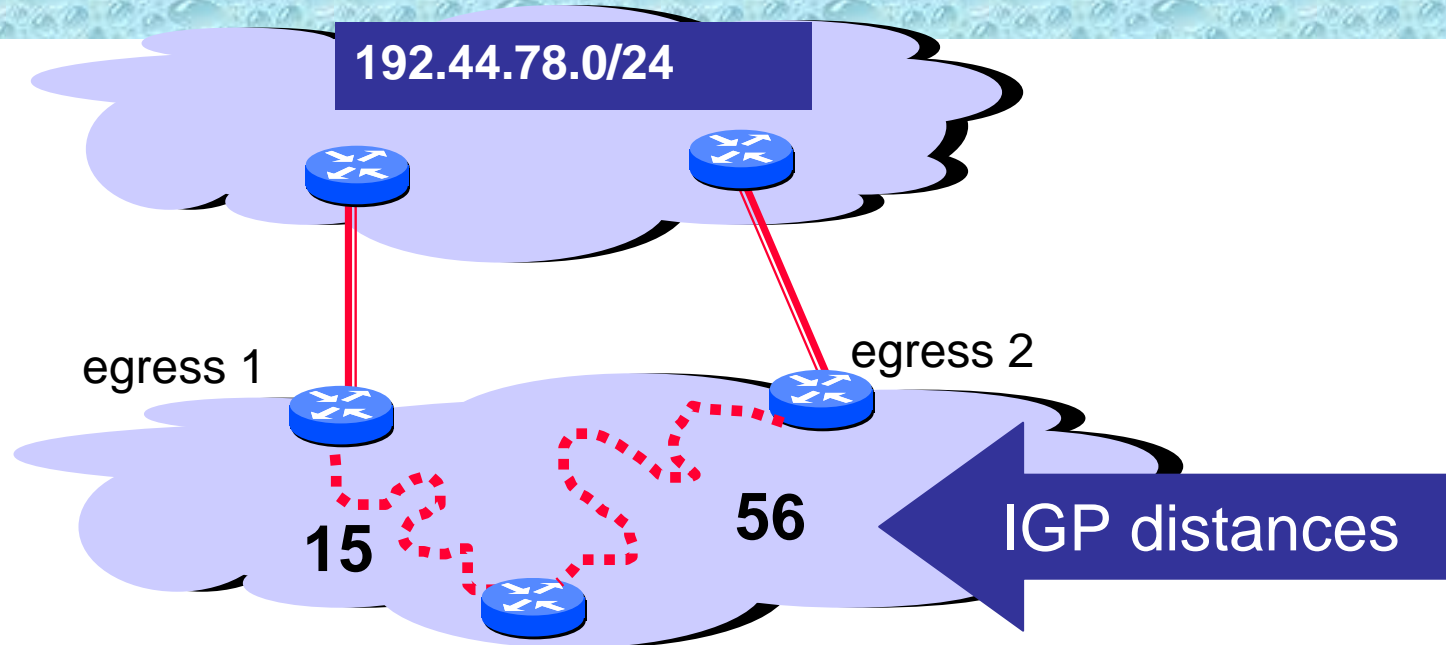


- the customer assigns a lower local-preference to the announcement coming from the backup peer
- local-preference attribute is checked before as-path length in the route selection process
- default value is 100
- local-preference applies to incoming announcements and manages outbound traffic flows

# Shedding Inbound Traffic with ASPATH Padding Hack



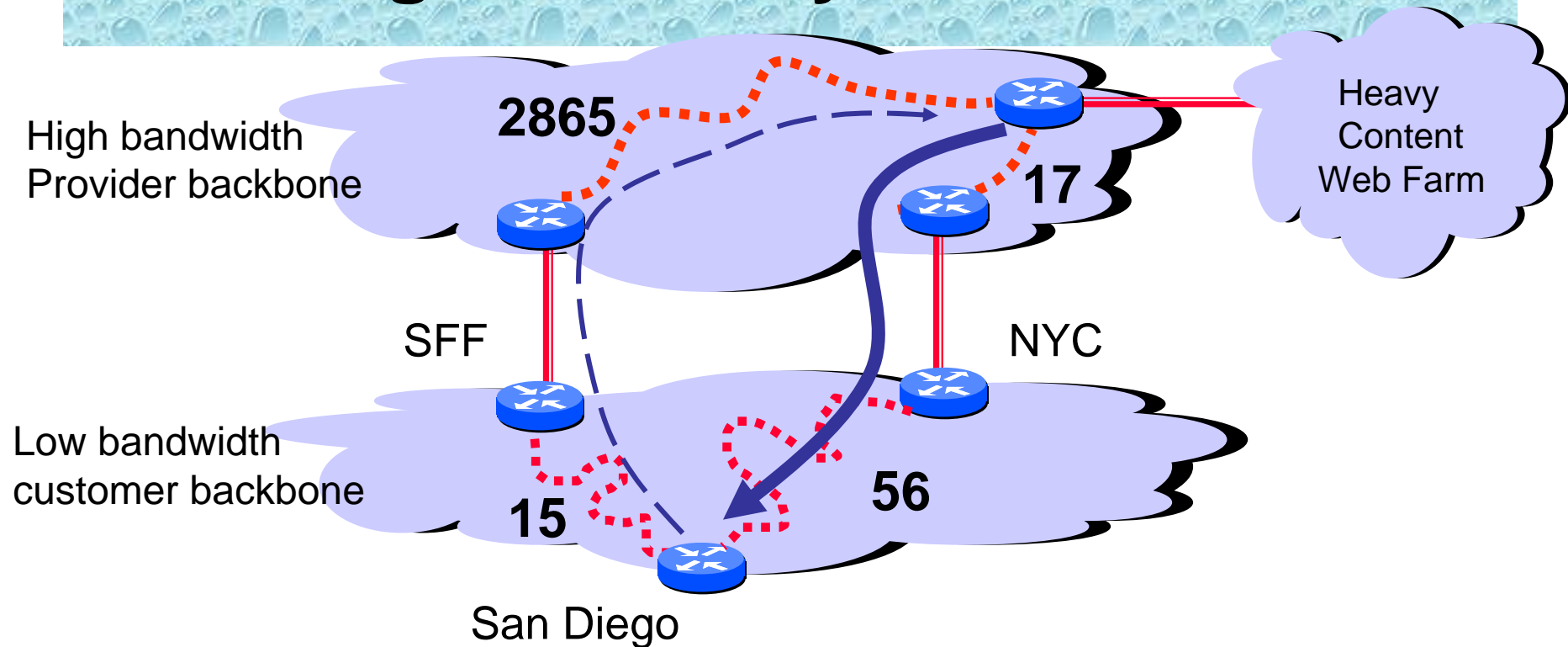
# Hot Potato Routing: Go for the Closest Egress Point



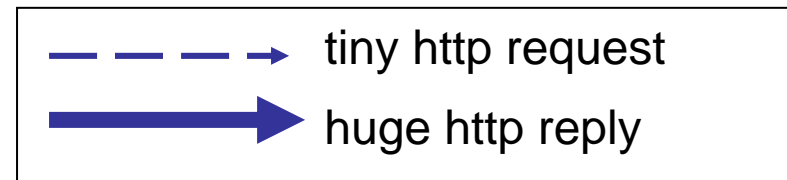
This Router has two BGP routes to 192.44.78.0/24.

Hot potato: get traffic off of your network as soon as possible. Go for egress 1!

# Getting Burned by the Hot Potato

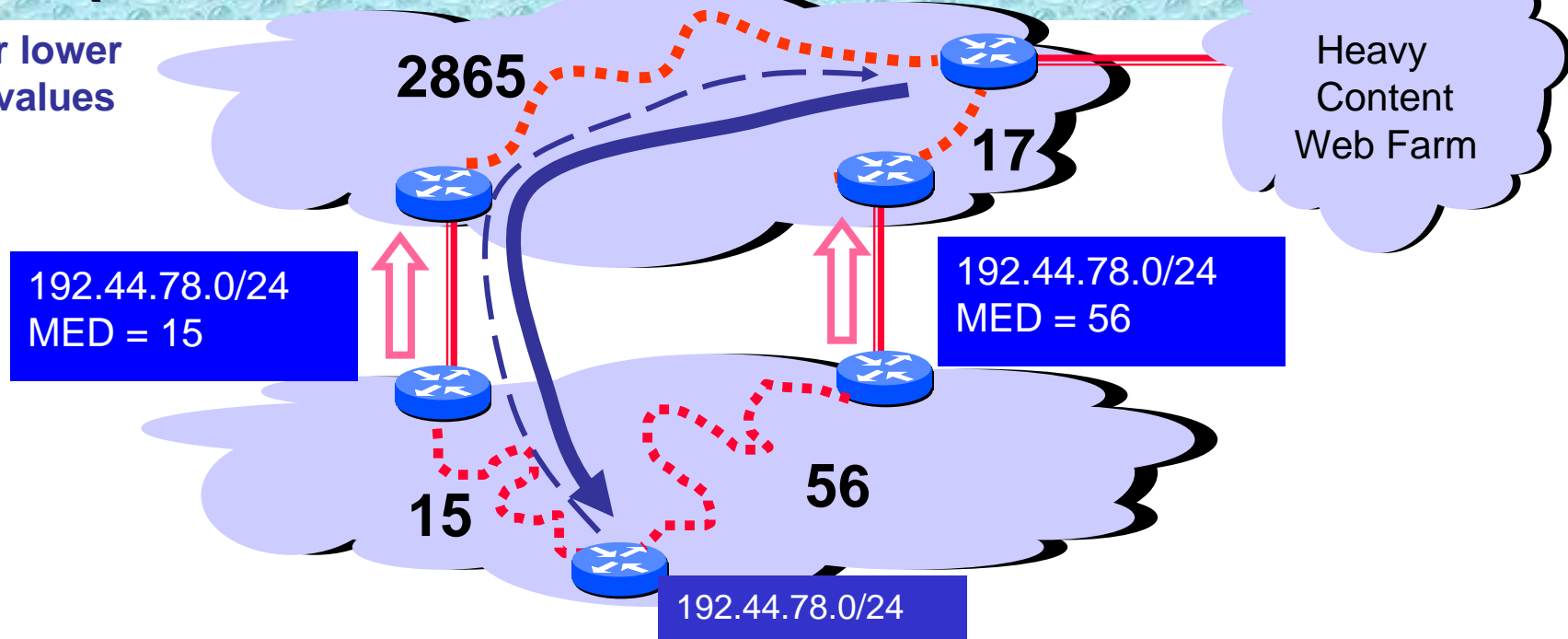


Many customers want  
their provider to  
carry the bits!



# Cold Potato Routing with MEDs (Multi-Exit Discriminator Attribute)

Prefer lower  
MED values



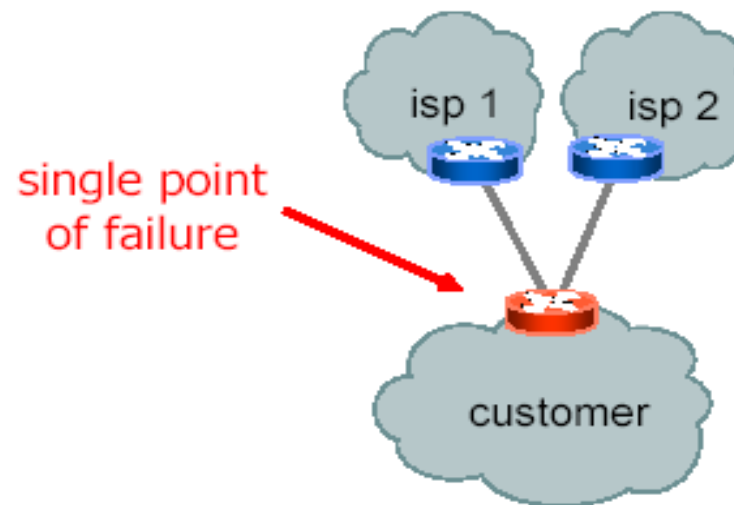
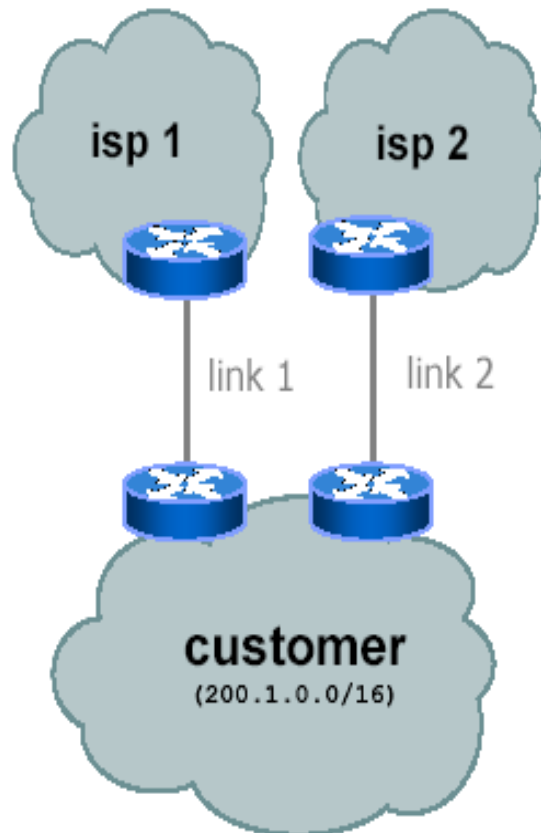
This means that MEDs must be considered BEFORE IGP distance!

Note1 : some providers will not listen to MEDs

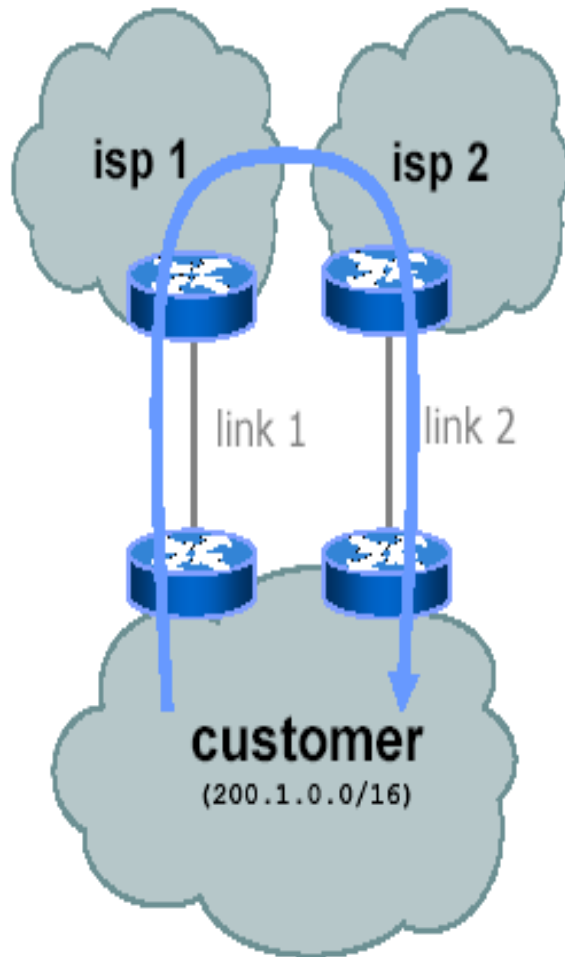
Note2 : MEDs need not be tied to IGP distance

# Multi-homed network

- Due link a due provider differenti
- In genere, due router sono coinvolti in modo tale da evitare singoli punti di rottura

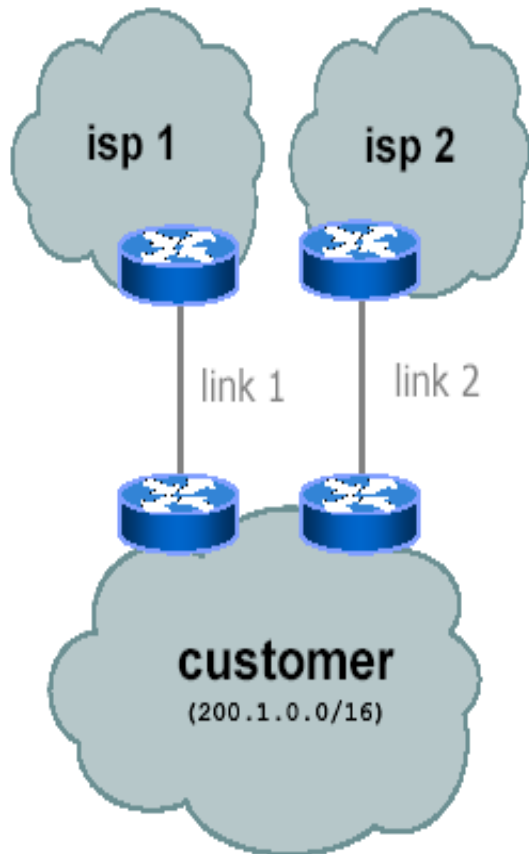


# Instradamento



- Un pacchetto in uscita può essere inviato attraverso uno dei due link per raggiungere Internet
- Un pacchetto in ingresso può usare uno dei due link per raggiungere la rete
- Un pacchetto internet può attraversare il link 1 ed il link 2
- Un pacchetto interno può attraversare entrambi i link

# Partizione del carico



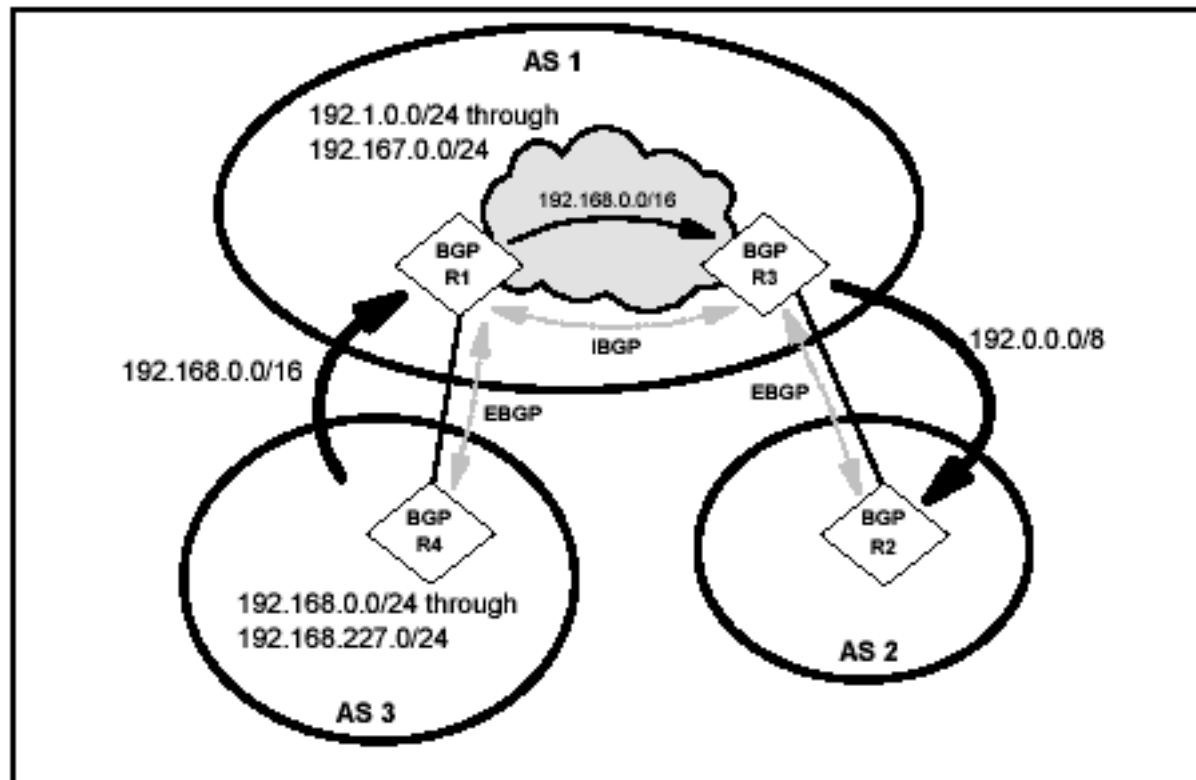
- Elimina il traffico in transito
- Traffico in uscita:
  - Metà degli host interni usano link 1,
  - l'altra metà usa link 2
- Traffico in ingresso:
  - usa link 1 per raggiungere metà degli host interni
  - Usa link 2 per l'altra metà

# Uso di BGP per il partizionamento

- Traffico in ingresso: split /16 e annuncia due /17, uno per ogni link
  - Es.: 200.1.0.0/17 su link1 e 200.1.128.0/17 su link2
  - Partizionamento approssimato del traffico in ingresso
  - Assume uguale capacità ed anche distribuzione del traffico sul blocco di indirizzi
  - Modifica lo split finché un partizionamento perfetto è ottenuto
- Traffico in uscita: accetta l'instradamento di default upstream
  - Partizionamento del traffico con instradamento verso l'uscita più vicina (igp)
  - Una buona approssimazione poiché molto del traffico è diretto verso la rete

# Route aggregation

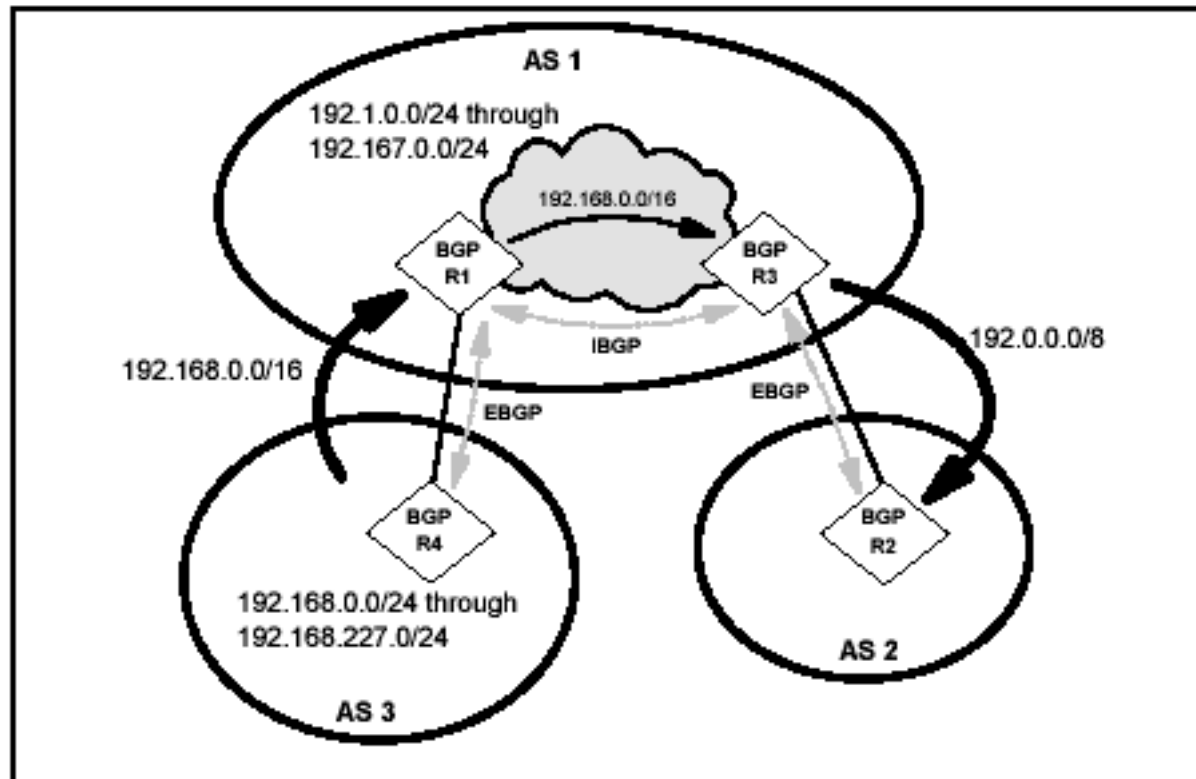
- BGP v4 usa CIDR e permette l'aggregazione delle route
- Cio' ne aumenta la scalabilita'



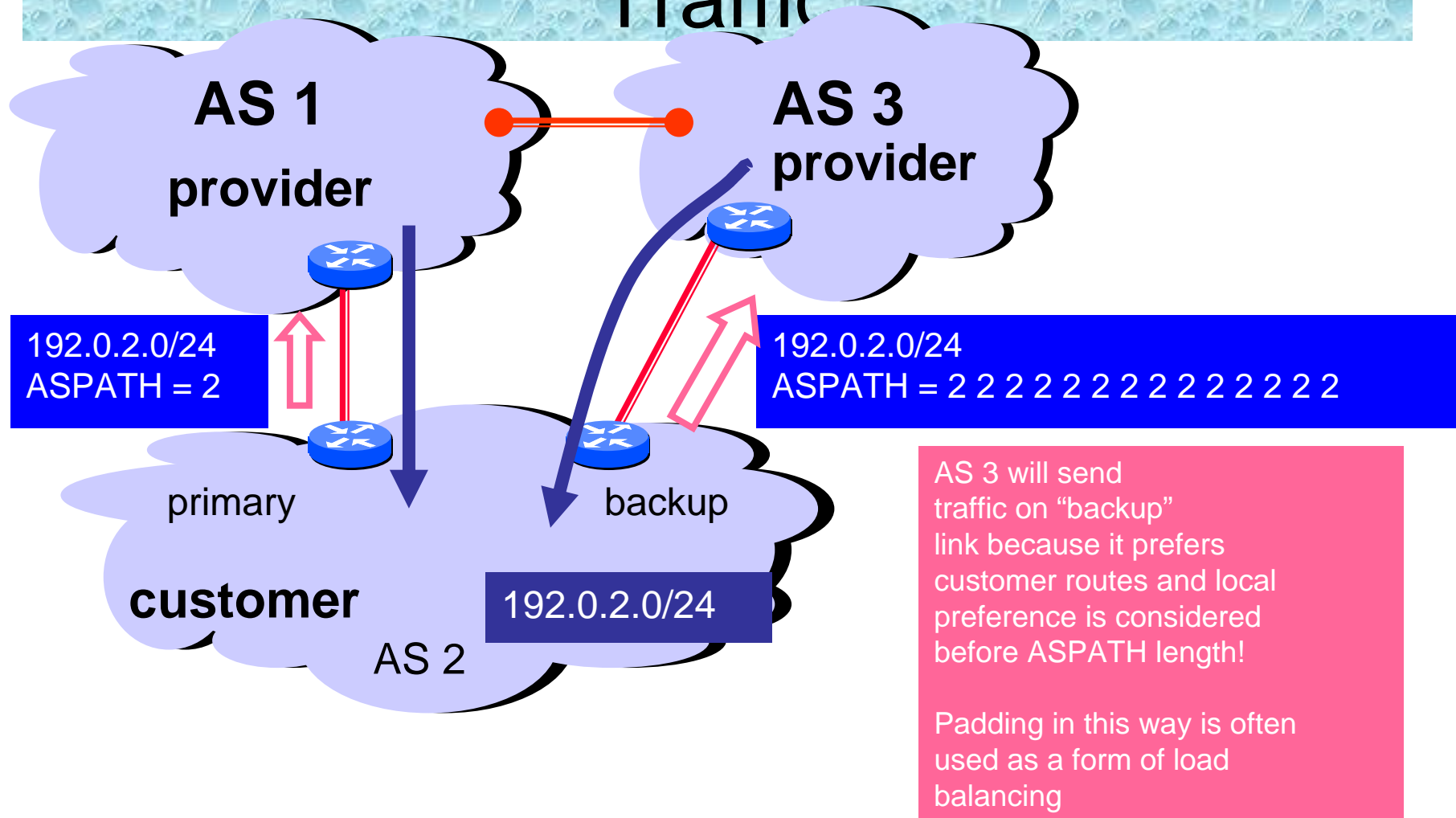
Attenzione: 182.....

# Route aggregation/cont.

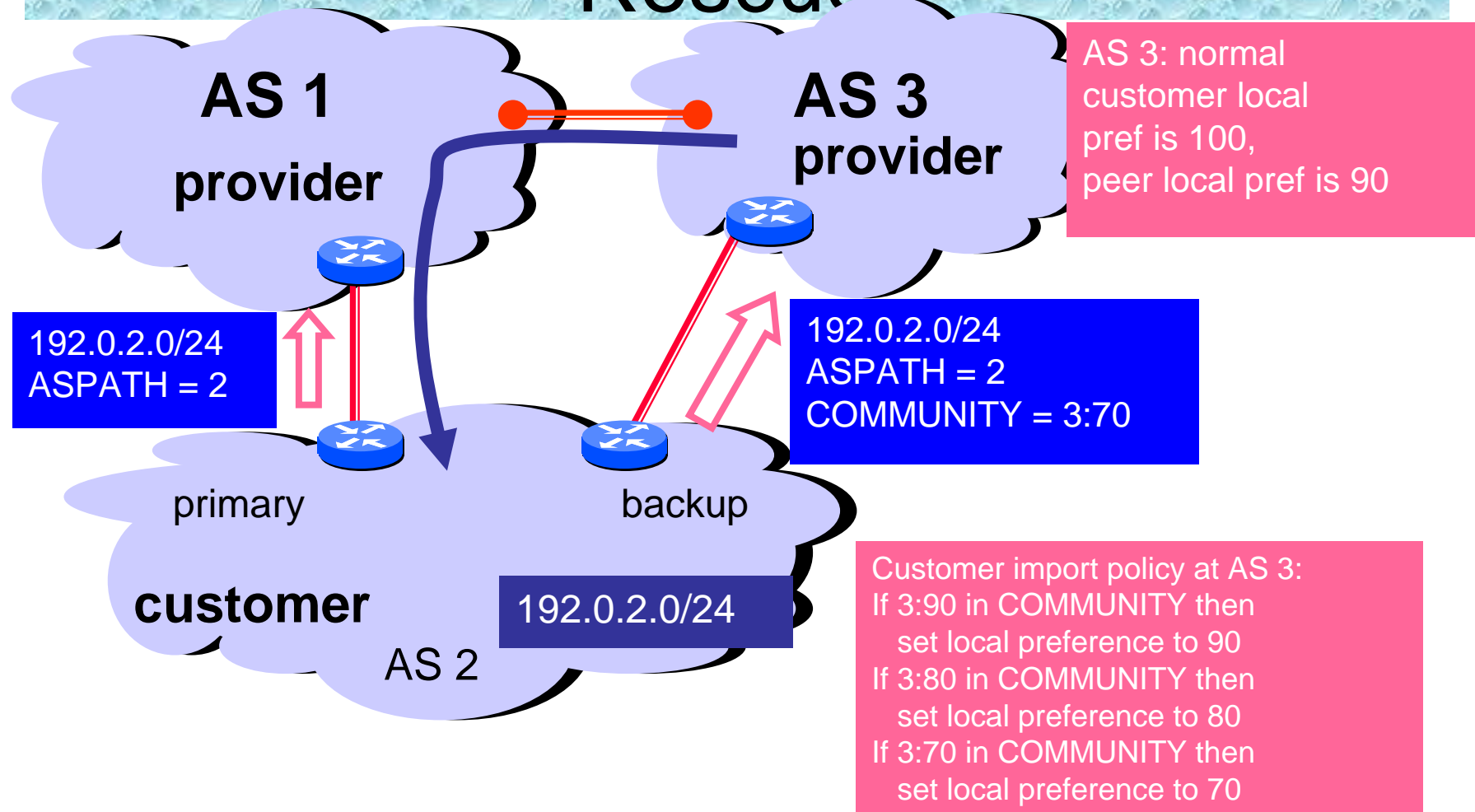
- R4 annuncia 182.168.0.0/16 <3> a R1
- R3 annuncia 182.0.0.0/8 <1 3> a R2
  - Percorso completo per evitare cicli



# Padding May Not Shut Off All Traffic



# COMMUNITY Attribute to the Rescue!



# Riferimenti

- Generali
  - Rif. 3, 5.8 e 5.9
  - White paper CISCO:  
[http://www.cisco.com/en/US/tech/tk365/technologies\\_tech\\_note09186a00800c95bb.shtml](http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a00800c95bb.shtml)
- Attributi:  
[http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito\\_doc/bgp.htm](http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/bgp.htm)
- Nota: tutti i riferimenti dati costituiscono materiale integrante del corso