



BGP

Border Gateway Protocol

Mario Baldi

**Politecnico di Torino
(Technical University of Turin)**

<http://www.mario-baldi.net>





Nota di Copyright

Questo insieme di trasparenze (detto nel seguito slide) è protetto dalle leggi sul copyright e dalle disposizioni dei trattati internazionali. Il titolo ed i copyright relativi alle slide (ivi inclusi, ma non limitatamente, ogni immagine, fotografia, animazione, video, audio, musica e testo) sono di proprietà degli autori indicati a pag. 1.

Le slide possono essere riprodotte ed utilizzate liberamente dagli istituti di ricerca, scolastici ed universitari per scopi istituzionali, non a fine di lucro. In tal caso non è richiesta alcuna autorizzazione.

Ogni altra utilizzo o riproduzione (ivi incluse, ma non limitatamente, le riproduzioni su supporti magnetici, su reti di calcolatori e stampate) in toto o in parte è vietata, se non esplicitamente autorizzata per iscritto, a priori, da parte degli autori.

L'informazione contenuta in queste slide è ritenuta essere accurata alla data dell'edizione. Essa è fornita per scopi meramente didattici e non per essere utilizzata in progetti di impianti, prodotti, reti, ecc. In ogni caso essa è soggetta a cambiamenti senza preavviso. Gli autori non assumono alcuna responsabilità per il contenuto di queste slide (ivi incluse, ma non limitatamente, la correttezza, completezza, applicabilità, aggiornamento dell'informazione).


In ogni caso non può essere dichiarata conformità all'informazione contenuta in queste slide.

In ogni caso questa nota di copyright non deve mai essere rimossa e deve essere riportata anche in utilizzi parziali.






Caratteristiche generali

- Attualmente alla versione 4 (RFC 1771)
 - Router adiacenti comunicano attraverso una connessione di livello trasporto affidabile
 - TCP
 - Protocollo Path Vector
 - Per ogni destinazione IP è fornita la sequenza di Autonomous System (AS) da attraversare
 - Ognuno è identificato con 2 ottetti
 - AS number pubblici
 - AS number privati (64512-65535)
 - Il distance vector puro indica solo il costo
 - Non c'è il problema del conteggio a infinito (“counting-to-infinity”)
- 






Caratteristiche generali

- Non definisce i criteri per la scelta delle route
 - Lasciato all'amministratore di rete che può così realizzare politiche (*policy*) di routing
 - BGP fornisce i meccanismi per
 - diffondere informazioni
 - descrivere criteri di scelta per
 - percorsi su cui inoltrare traffico
 - Per esempio: non usare percorsi che passano per l'AS x
 - informazioni da diffondere
- 






Peering Session

- Due router che scambiano informazioni con il BGP si dicono *peer*
 - Due peer hanno tra di loro una *peering session*
 - I peer sono configurati esplicitamente
 - Se due peer non sono connessi direttamente si identificano con l'indirizzo di una interfaccia di loopback
 - La peering session non è legata alla disponibilità dell'interfaccia identificante
 - Caso comune tra peer I-BGP
- 






Due tipi di utilizzo

- Un AS può avere vari punti di contatto con altri AS
 - Più di un router BGP è attivo nell'AS
 - External BGP (E-BGP)
 - Scambio di informazioni tra router appartenenti ad AS differenti
 - Internal BGP (I-BGP)
 - Router appartenenti allo stesso AS scambiano informazioni su destinazioni esterne all'AS
 - Determinare quale dei router BGP debba essere usato per il raggiungimento di ogni destinazione esterna
- 






I-BGP

- **Redistribuzione in IGP porterebbe a perdita di informazioni tipiche del BGP**
 - **Un router BGP che riceve annunci da peer I-BGP li inoltra solo a peer E-BGP**
 - **Verso peer I-BGP non modifica *AS_path*, quindi loop tra peer I-BGP non sarebbe rilevato**
 - **Ogni router deve avere una peering session con ogni altro router interno**
 - **L'uso di *BGP confederation* o *route reflector* elimina questo requisito**
 - **Può essere usato per routing intradominio**
 - **Meno efficiente di altri IGP (RIP e OSPF) perché questi scoprono da sé i loro interlocutori**
 - **Converge lentamente**
- 





Confederation

- L'AS è diviso in mini-AS
 - Ogni mini-AS ha un AS number privato
 - I mini-AS sono raggruppati in una confederation
 - Router nello stesso mini-AS sono collegati in maglia completa
 - I mini-AS sono collegati tra loro tramite E-BGP
 - Router tra mini-AS trattano gli annunci come I-BGP peers
 - Non cambiano next hop, multi exit disc e local pref
- 



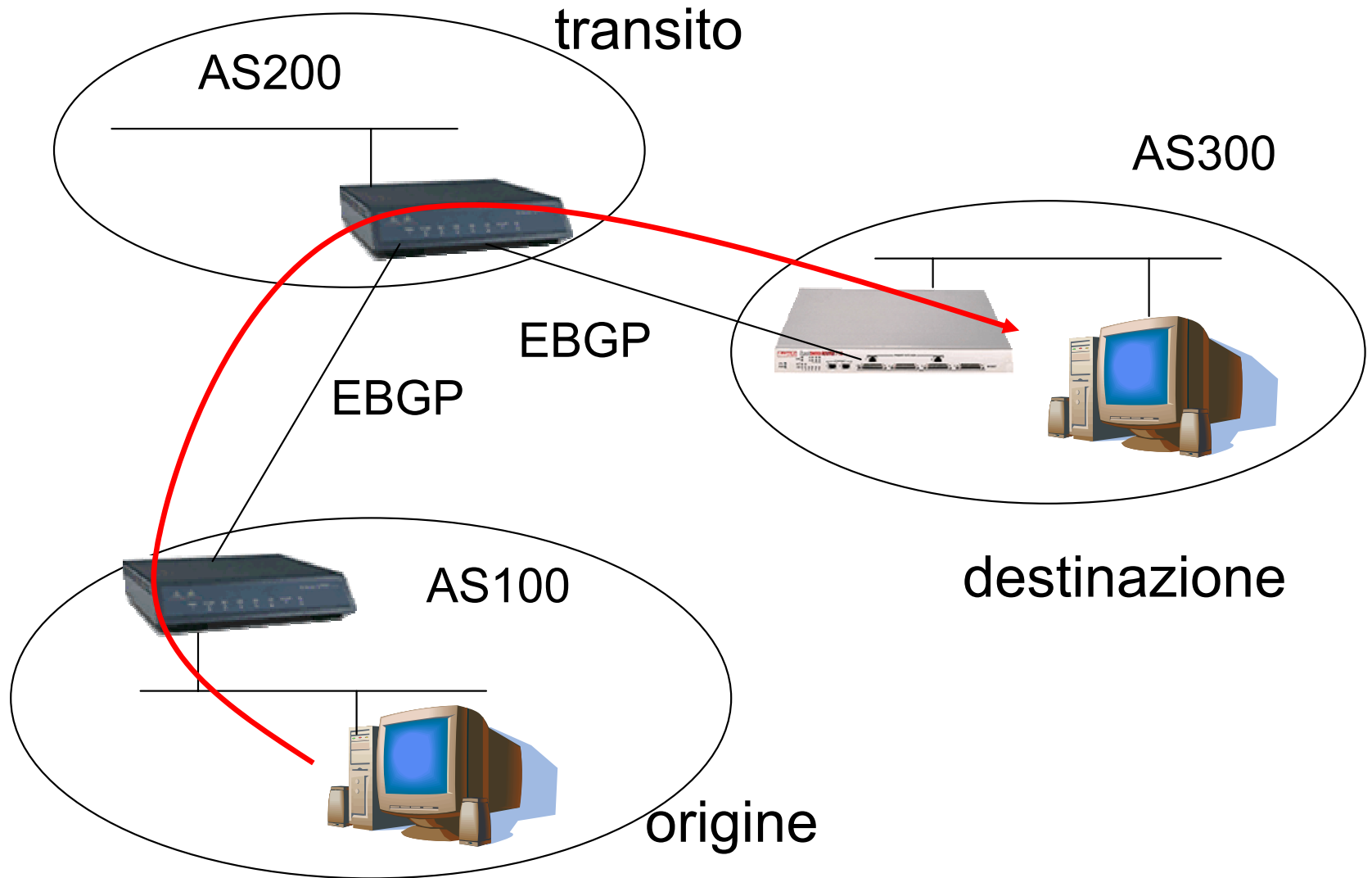


Route Reflector

- Un I-BGP router configurato da route reflector inoltra ad altri I-BGP router le route apprese tramite BGP



Trasmissione dati e AS



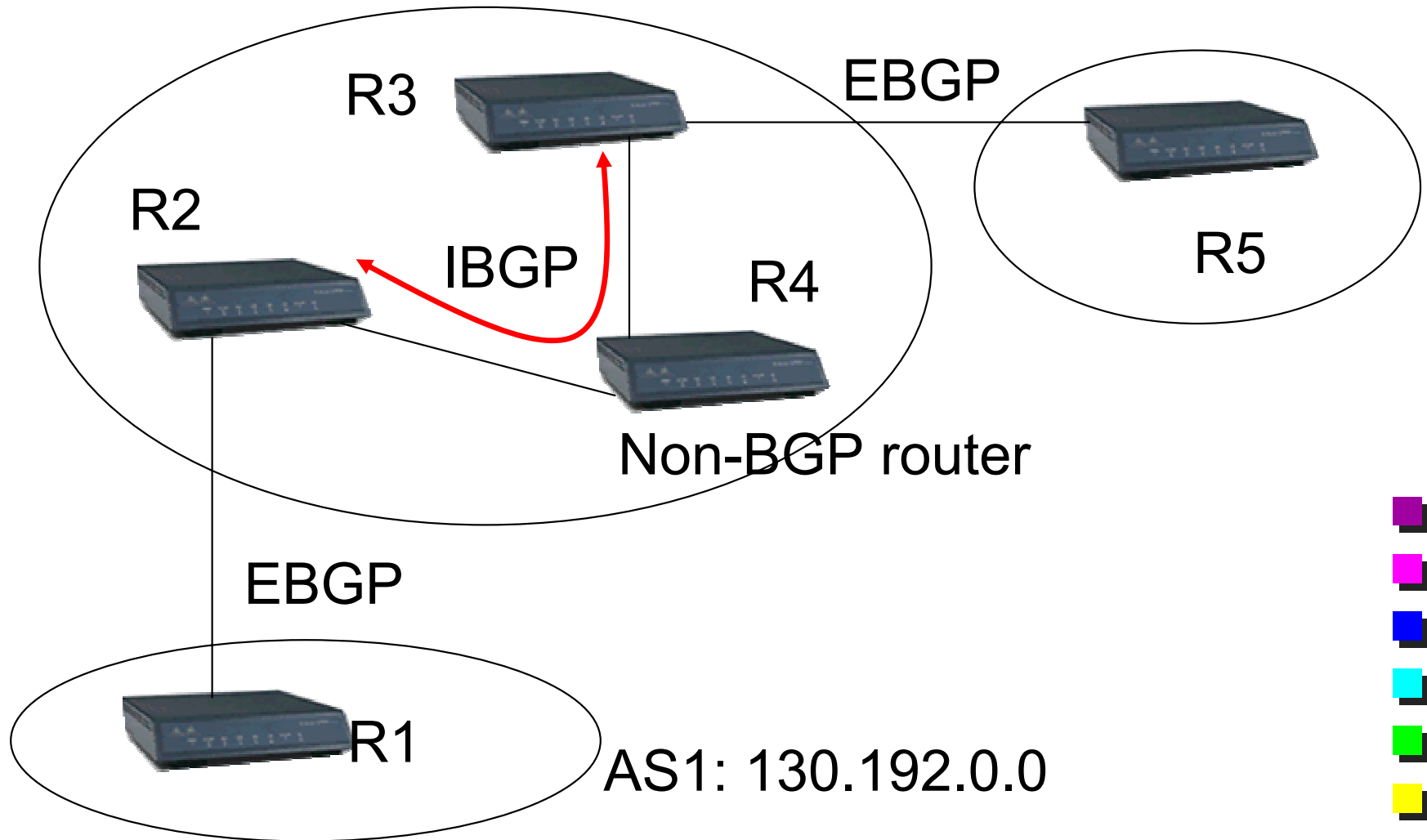


Sincronizzazione

- In un AS di transito contenente router che non parlano BGP, il traffico di transito potrebbe non essere gestito correttamente se i router non BGP non apprendono mediante un IGP gli annunci ad esso relativi
- BGP può propagare un annuncio solo quando tutti i router interni all'AS hanno ricevuto lo stesso annuncio da parte di un IGP



Sincronizzazione





Sincronizzazione

- R1 genera annunci per 130.192.0.0/24
- Gli annunci raggiungono R2, che via IBGP li propaga a R3
- R3 propaga gli annunci via EBGP a R5
- R4 può apprendere dell'esistenza della rete 130.192.0.0/24 solo mediante un IGP
- R5 inizia ad inviare a R3 il traffico diretto alla rete 130.192.0.0, e R3 lo invia a R4 per raggiungere R2 (suo peer IBGP)
- Se R4 non ha ricevuto l'annuncio per la rete 130.192.0.0/24 da parte di un IGP, non sa come gestire il traffico






Sincronizzazione

- Utilizzando la regola di sincronizzazione, il router R3 annuncia la rete 130.192.0.0/24 solo quando raggiunge via IGP l'annuncio per la stessa rete
- Può essere utile disabilitare la sincronizzazione:
 - se l'AS non è un AS di transito
 - se tutti i router dell'AS parlano BGP






Border Gateway Protocol

- Le destinazioni IP sono espresse in termini di prefissi di indirizzo
 - Classless Inter-Domain Routing (CIDR)
 - I router possono aggregare le informazioni di routing ricevute prima di propagarle
 - diminuzione del traffico di routing
 - diminuzione delle dimensioni delle basi dati nei router
 - Ogni router ha un algoritmo per fare una classifica dei percorsi alternativi: Decision Process
- 





Route sovrapposte

- Uno stesso router riceve annunci che dichiarano alcune destinazioni raggiungibili attraverso percorsi diversi
 - Percorso maggiormente specifico → valido per un sottoinsieme di destinazioni più ristretto
 - prefisso di indirizzo più lungo
 - Percorso meno specifico → valido per un sopra insieme di destinazioni
 - prefisso di indirizzo più corto
 - I router per default scelgono la route maggiormente specifica
- 





Route sovrapposte

- Se un router sceglie la route meno specifica lo segnala quando la propaga
 - si notifica che non è garantito che i pacchetti seguiranno il percorso annunciato
 - il router da cui arrivano le route sovrapposte invierà alcuni pacchetti sul percorso della route più specifica e altri su quella meno specifica (l'unica annunciata)
 - la route non può essere disaggregata
 - i router che ricevono l'annuncio non possono annunciare percorsi differenti per sottoinsiemi delle destinazioni





Gli attributi

- Per la descrizione di un percorso verso una destinazione sono usati campi detti attributi
- Non tutte le combinazioni sono ammissibili






Classificazione degli attributi

- **well-known/optional**
 - deve essere riconosciuto da ogni realizzazione
- **mandatory/discretionary**
 - deve apparire nella descrizione del percorso
- **partial**
 - un router che non riconosce l'attributo lo mantiene immutato nelle descrizioni di percorso che emette
- **transitive/nontransitive**
 - **transitive**: è mantenuto immutato da router che non lo riconoscono ed è contrassegnato come partial
 - **nontransitive**: è cancellato da router che non lo **CONOSCONO**





Origin

- **Mandatory**
 - **IGP: imparata da un protocollo di routing interno all'AS**
 - **EGP: imparata da EGP**
 - **Incomplete:**
 - **Non imparata né tramite IGP né tramite EGP**
 - **Route statica**
 - **Insieme all'attributo AS path associato ad una destinazione non più raggiungibile**
- 



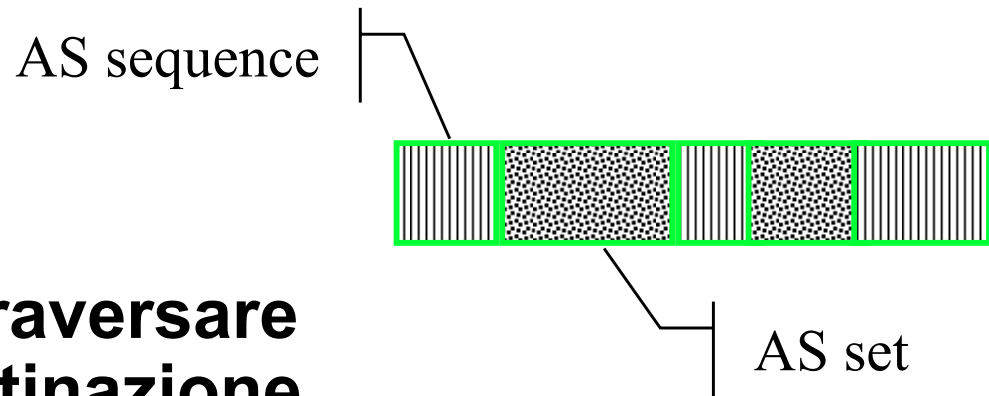
AS Path

Mandatory

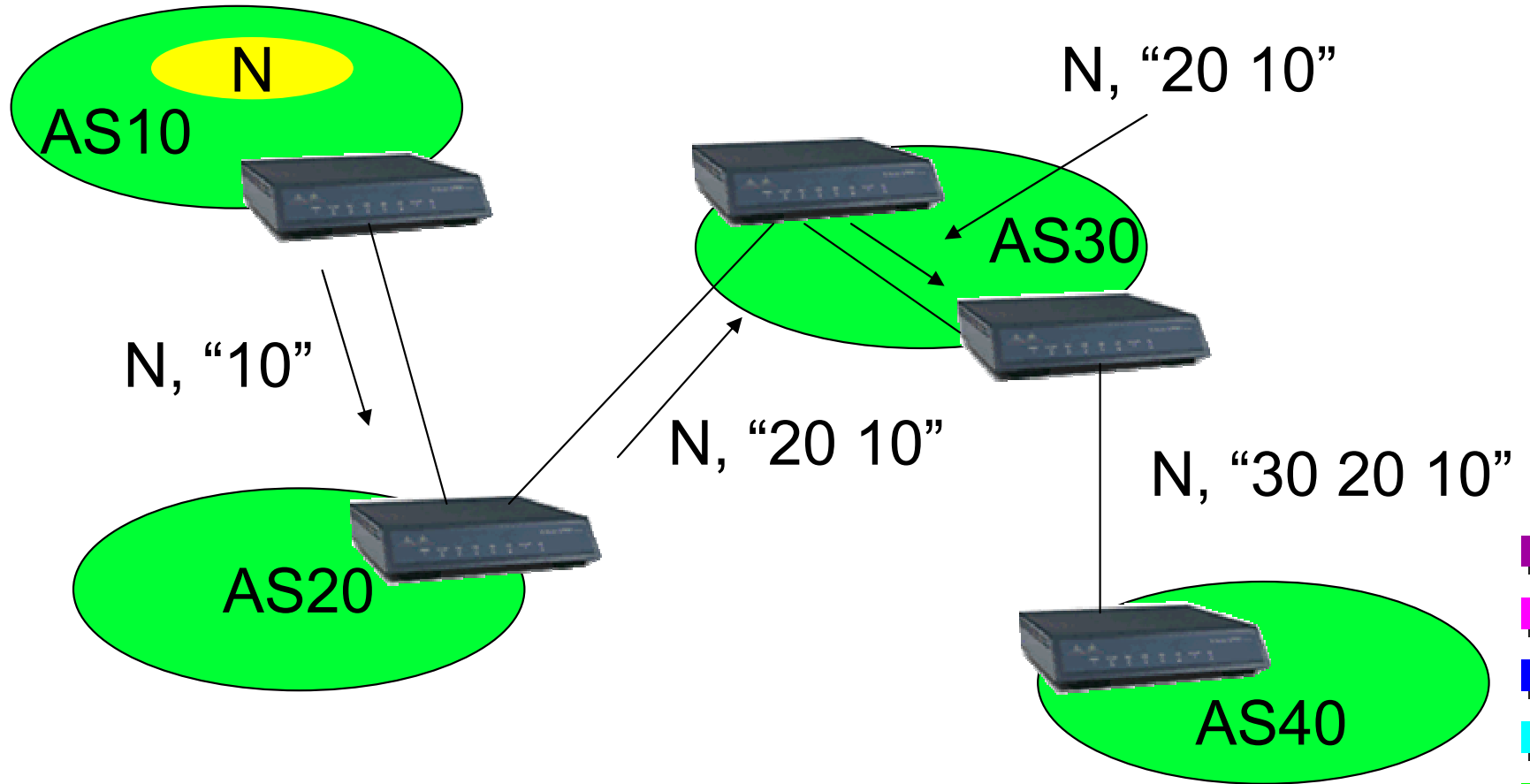
Elenco degli AS da attraversare per raggiungere la destinazione

- componenti ordinati (AS_SEQUENCE)
- componenti non ordinati (AS_SET)
- Negli annunci che un router propaga
 - se la propagazione è nel proprio AS, l'AS path non è modificato
 - se la propagazione è verso l'esterno
 - se il primo componente è ordinato, l'identificatore dell'AS è aggiunto in testa alla sequenza
 - se il primo componente non è ordinato si aggiunge un componente ordinato che contiene l'identificatore dell'AS

componenti non ordinati sono creati per aggregazione



AS Path






Next Hop

- **Mandatory**
- **Indica il router BGP (n_h) attraverso cui raggiungere la destinazione annunciata**
- **Non cambiato in annunci I-BGP**
 - n_h non appartiene all'AS a in cui è propagato
 - a deve avere una route per n_h
- **Il next hop n_h negli annunci all'interno dell'AS a è**
 - **Peer E-BGP che ha inviato la route all'AS a**
 - Ha inserito il proprio indirizzo n_h nel campo next hop
 - **Router sulla stessa sottorete del peer E-BGP che ha inviato la route all'AS a**
 - n_h è sulla route per la destinazione, ma non ha una sessione di peering con l'AS a





Multi Exit Disc

- **Optional**
 - **Usato per discriminare tra vari punti di uscita verso l'AS annunciante**
 - **Un router lo invia ad uno esterno per annunciare il costo verso il proprio AS (metrica su 4 byte)**
 - **Un router che lo riceve lo propaga all'interno del proprio AS**
 - **mai propagato all'esterno**
 - **tra route equivalenti si sceglie quella a costo minore**
- 






Local Pref

- Discretionary
- Incluso negli annunci interni (I-BGP)
 - mai propagato all'esterno
- Esprime il grado di preferenza del router nell'utilizzo della route





Atomic Aggregate

- Discretionary
 - La route propagata è la meno specifica tra più route sovrapposte
 - Aggiunto dal router che ha scelto la route meno specifica
 - Non rimosso dai router che lo ricevono
 - Non viene fatta disaggregazione delle destinazioni
- 





Aggregator

- Optional
- Aggiunto dal router che ha generato la route per aggregazione di altre
- Contiene l'identificatore di AS e l'indirizzo IP del router





Altri attributi

■ Community

- Destinazioni con lo stesso valore di community sono trattate nello stesso modo

■ Originator ID

- Router che ha originato una route propagata (da un route reflector) tramite I-BGP


■ Cluster List

- Lista di cluster attraversati dall'annuncio I-BGP nell'AS in cui sono attivi route reflector
 - Cluster: gruppo di router in cui è attivo un route reflector
- 






Politiche

- **Configurate manualmente**
 - **Permettono al gestore di rete di stabilire i criteri di scelta delle route per ogni destinazione**
 - **Le politiche riflettono accordi tra gli AS**
 - **Si possono imporre politiche molto complesse**
 - **Deve esistere un metodo locale per costruire una funzione che dati gli attributi di un percorso restituisce un grado di preferenza**
 - **Numero intero**
 - **Fase 1 del decision process**
- 



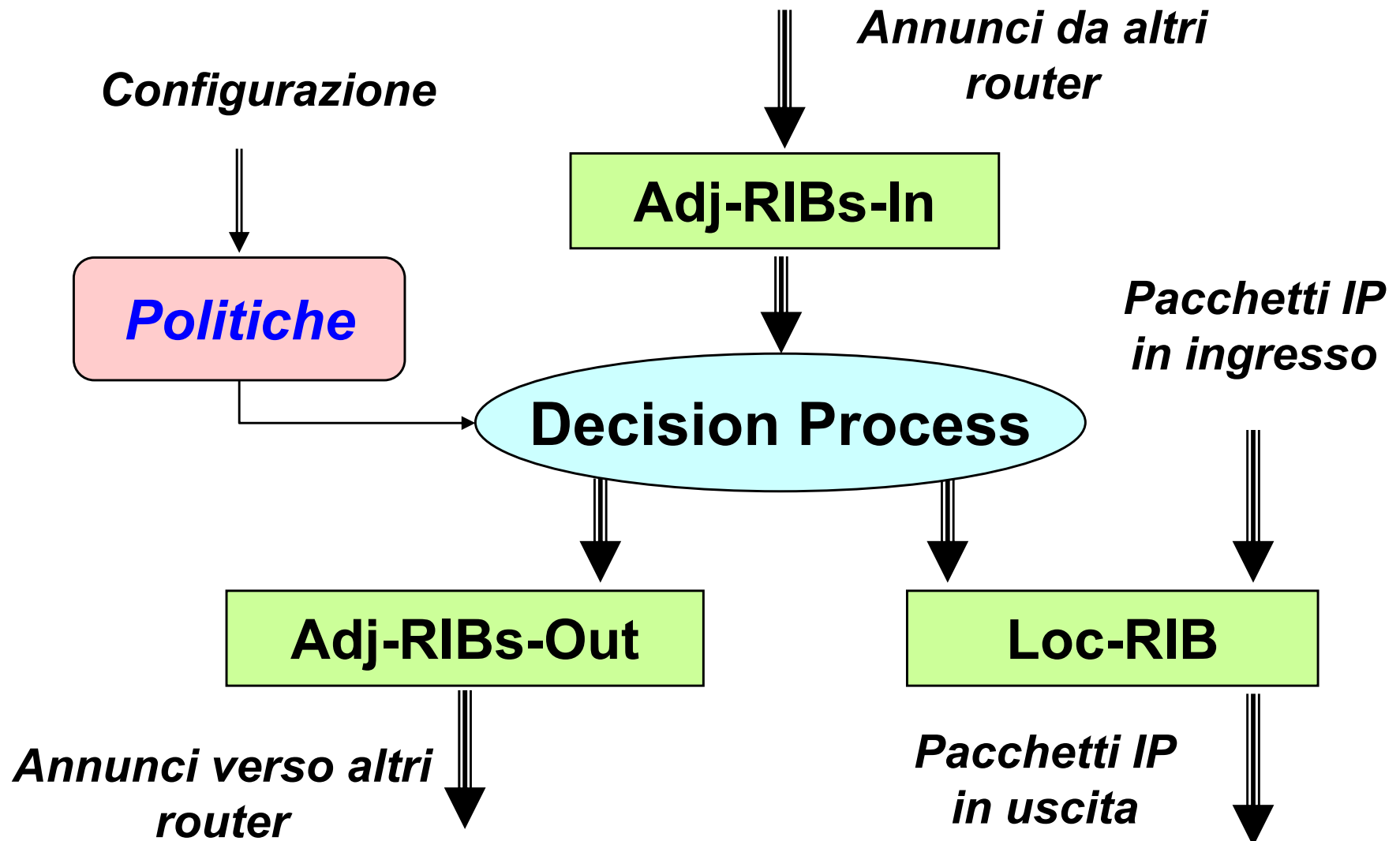


Routing Information Base (RIB)

- **Adj-RIBs-In** → informazioni imparate dagli annunci ricevuti (e non scartati)
 - Il decision process (fase 1) calcola il grado di preferenza associato ad ogni route
 - Sceglie le route con grado di preferenza più alto per annunciarle all'interno dell'AS
 - **Loc-RIB** → informazioni usate per l'instradamento
 - selezionate mediante il decision process (fase 2)
 - non contiene route con next hop non raggiungibile
 - **Adj-RIBs-Out** → informazioni da propagare
 - selezionate mediante il decision process (fase 3) a partire dalla Loc-RIB
 - Phase 3 could perform route aggregation
- 




Politiche, tabelle, annunci





Ricezione e propagazione

- Quando è ricevuto (e accettato) un annuncio
 - la destinazione si trova in Adj-RIBs-In → la nuova route rimpiazza la vecchia
 - il decision process è eseguito
 - la route è maggiormente specifica di un'altra
 - attributi diversi → il decision process è eseguito
 - parte della route meno specifica non è più valida
 - stessi attributi → la nuova route è ignorata
- 






Ricezione e propagazione

- la destinazione non è presente in Adj-RIBs-In
→ la nuova route è aggiunta
 - il decision process è eseguito
- la nuova route è meno specifica di una esistente → il decision process è eseguito
 - solo sulle destinazioni descritte dalla nuova route





Decision Process

- Applica le politiche contenute nella Policy Information Base (PIB) per selezionare le route da propagare
 - È una funzione che dati gli attributi di una route restituisce un intero (grado di preferenza)
 - Il decision process non considera
 - l'esistenza di altre route
 - la non esistenza di altre route
 - gli attributi di altre route
 - Applicata la funzione a tutte le route per una destinazione, si sceglie quella con grado di preferenza maggiore
- 






Decision Process

- Agisce su tutte le route contenute in Adj-RIB-In
- Seleziona le route da propagare all'interno dell'Autonomous System
- Selezione le route da propagare all'esterno dell'Autonomous System
- Aggrega le route e riduce le informazioni da trasmettere






Criterio di selezione di una route




Salvo diverso comportamento imposto dalle politiche, le fasi 1 e 2 del decision process portano alla selezione della route con

- Local Preference maggiore
 - se annuncio arriva tramite I-BGP
 - Path originato localmente
 - Path più corto
 - Origin minore
 - IGP < EGP < incomplete
 - Multi exit disc più basso
 - Route esterna (da E-BGP) è meglio di interna
 - Route con next hop più vicino (secondo l'IGP)
 - Path annunciato dal router con ID più basso
- 





Dampening Route Flaps

- **Route flap: una route annunciata è ritirata (withdrawn) e poi annunciata nuovamente**
 - **Succede, per esempio, quando il collegamento verso il proprio ISP è interrotto e ripristinato**
 - **Route flap consumano notevole tempo di CPU sui router di Internet**
 - **Un provider può smettere di annunciare una route che fa più di uno o due flap**
 - **È talvolta meglio non essere annunciati (in modo specifico) dal proprio provider**
- 






Mesaggi BGP

■ Open

- primo messaggio trasmesso quando un link diviene operante
- negoziazione della versione

■ Update

- contiene informazioni di routing
 - distribuisce una sola route
 - può annullare (withdraw) molte route
 - può includere molte destinazioni (per cui è valida la stessa route)
- 






Mesaggi BGP

■ Notification

- ultimo messaggio trasmesso prima di abbattere un link

■ Keepalive

- indica al router adiacente che il mittente è ancora attivo
 - usato quando non si hanno informazioni di routing da trasmettere
- 



Formato dei messaggi BGP

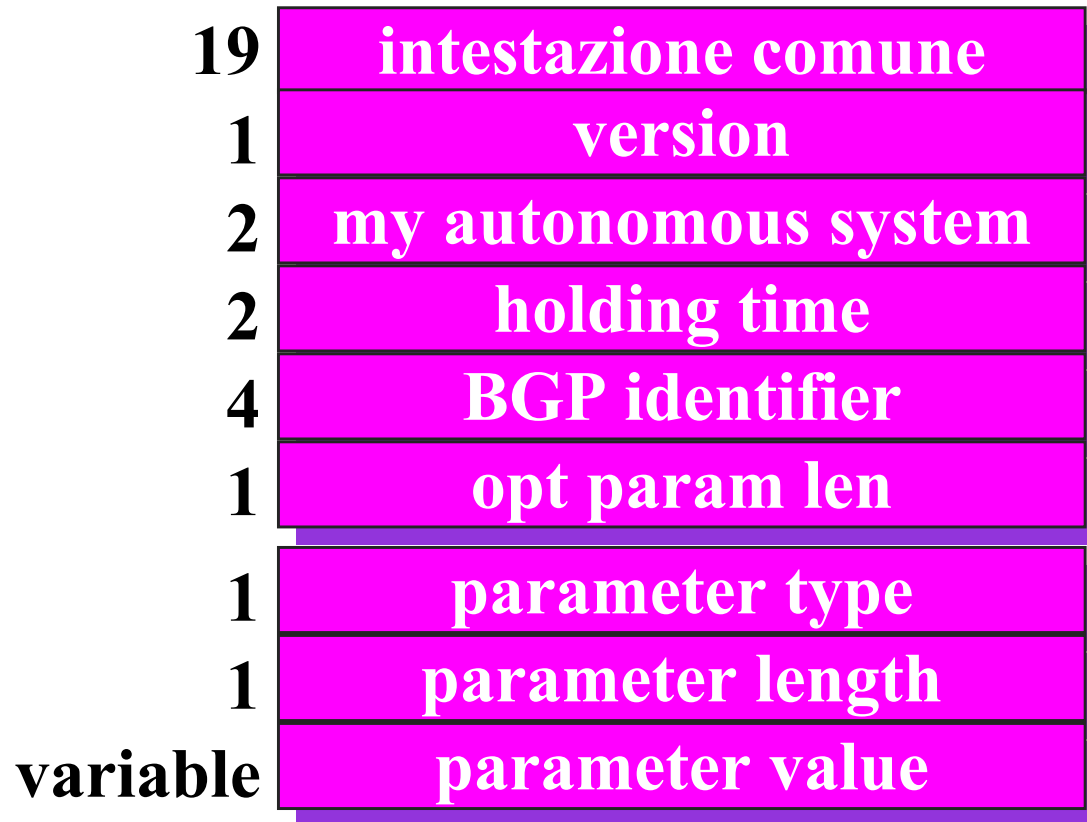
■ Intestazione comune

- type = 1 → open
- type = 2 → update
- type = 3 → notification
- type = 4 → keepalive



Formato dei messaggi BGP

■ Messaggio open

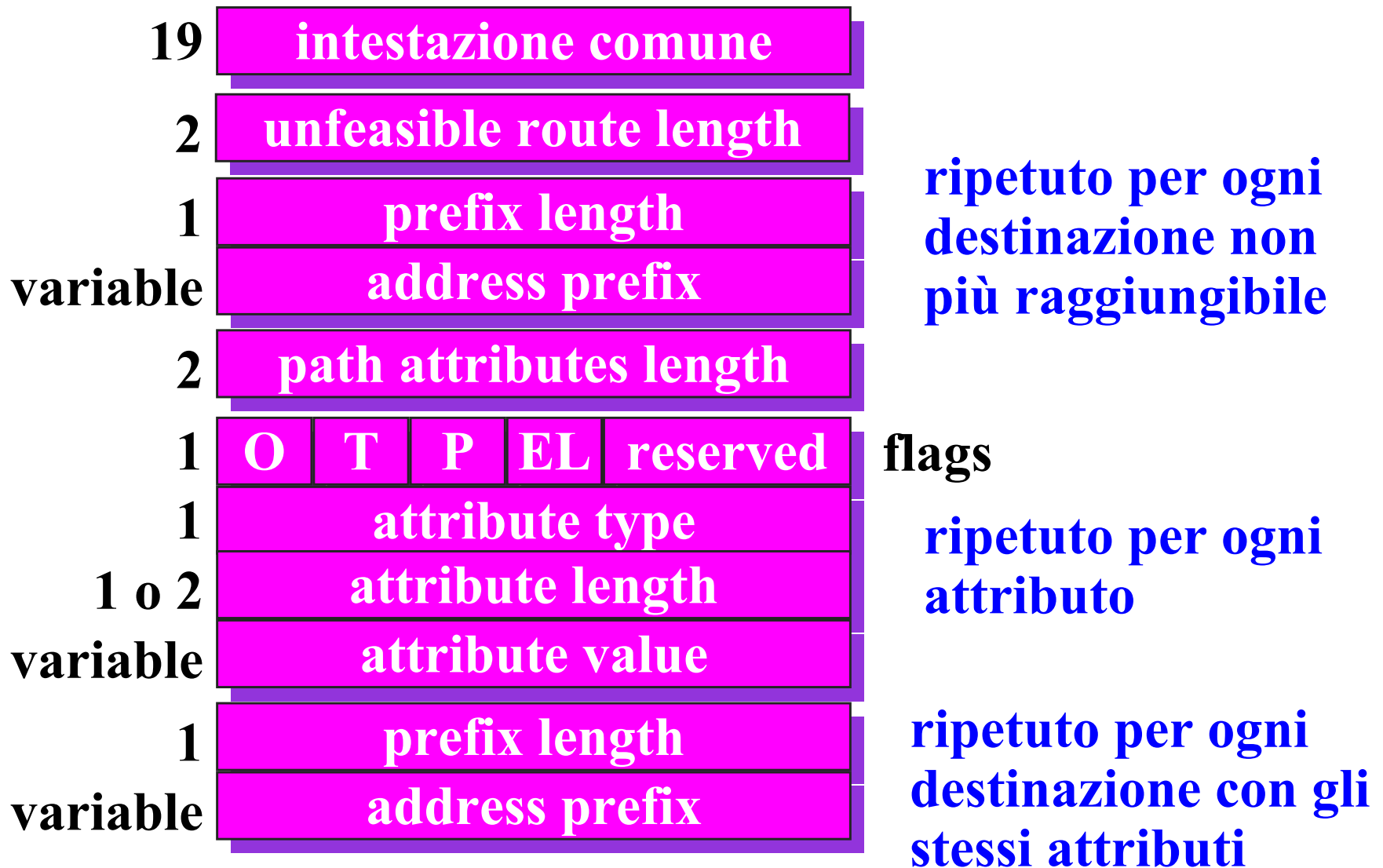


- authentication information: parameter type = 1



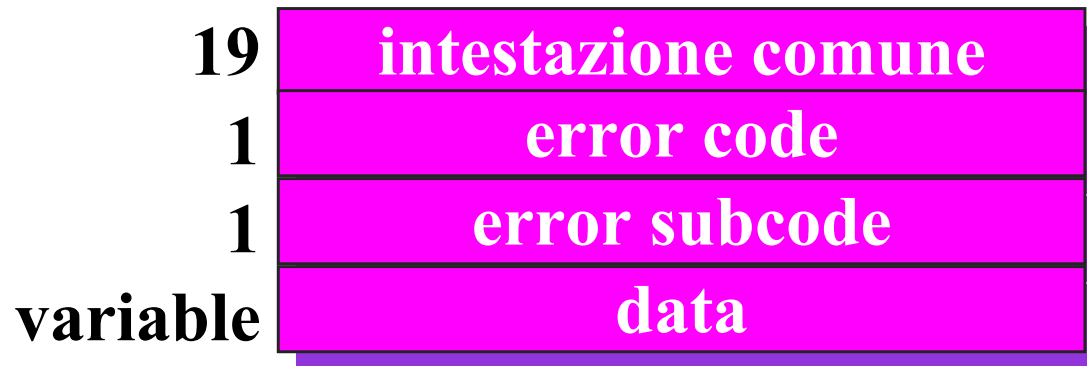
Formato dei messaggi BGP

■ Messaggio update



Formato dei messaggi BGP

■ Messaggio notification



■ Messaggio keepalive

