Reti di Telecomunicazioni

Network Layer

Routing Internet

Autori

Queste slides sono state scritte da

MicheleMichelotto: michele.michelotto@pd.infn.it

che ne detiene i diritti a tutti gli effetti

Copyright Notice

Network layer

Introduzione
Layer: Modello OSI e TCP/IP
Physics Layer
Data Link Layer
MAC sublayer
Network Layer
Transport Layer
Application Layer

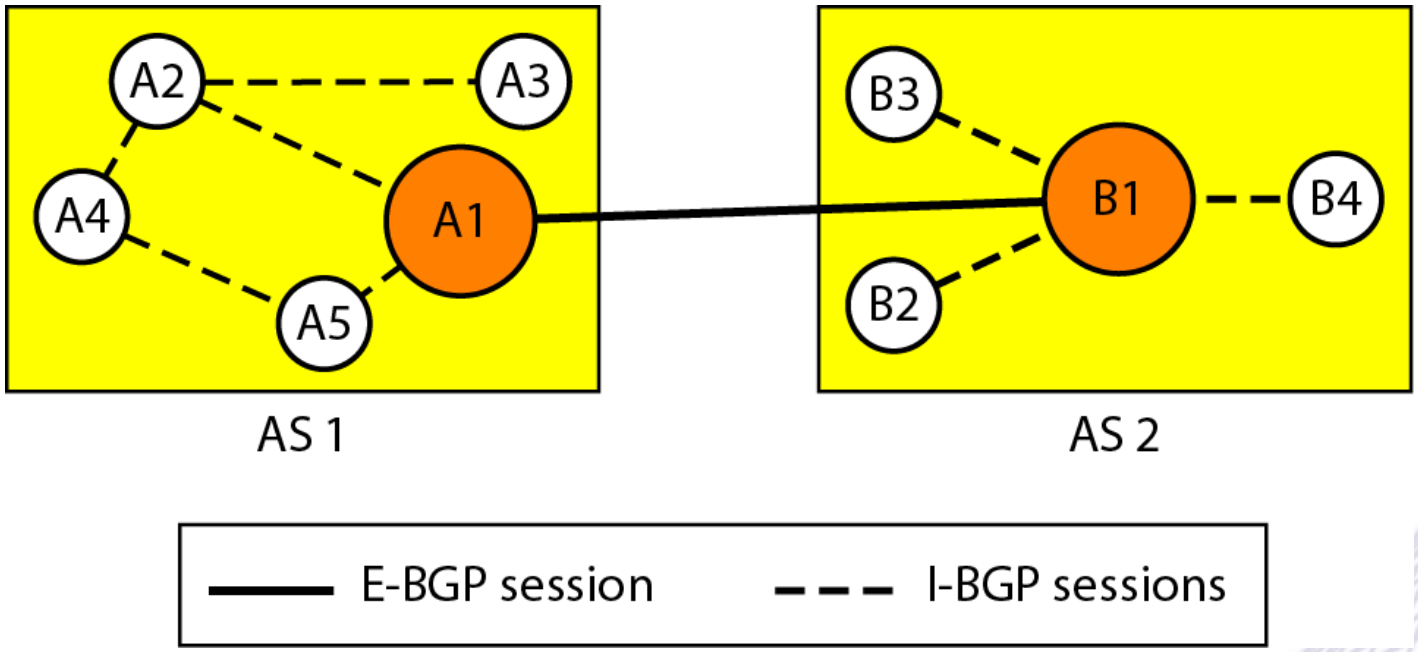
Routing IP

- Dopo aver visto il problema in generale del routing di tipo Distance Vector e di tipo Link State vediamo cosa succede nelle reti di Internet

Routing di Internet

- Internet è costituita da diversi Autonomous Systems (AS)
  - Ogni AS viene gestito da una diversa organizzazione.
  - All'interno di un AS l'organizzazione può scegliere il suo algoritmo di routing. Tuttavia se ne viene scelto uno standard poi risulta più facile il routing nei confini verso gli altri AS
  - Un algoritmo di routing entro un AS si dice "Interior Gateway Protocol"
  - Un algoritmo per il routing tra AS si chiama "Exterior Gateway Protocol"

Autonomous System



RIP

- RIP
  - Il primo protocollo interno era un protocollo “distance vector” chiamato RIP e basato su un algoritmo “Bellmann-Ford” eredidato da Arpanet
  - Funziona bene in piccoli AS
  - Soffre del problema del count-to-infinity e converge lentamente
  - Venne rimpiazzato nel 1979 da un protocollo link state
  - Nel 1988 IETF inizia a lavorare su di un successore chiamato OSPF (RFC 2328) che diventa standard nel 1990

OSPF

- Open Shortest Path First. Requirements:
  - Open. Deve essere un protocollo non proprietario
    - problema sentito dato che che il 90% dei router sono (erano) CISCO con protocolli proprietari
  - Deve supportare diverse “distance metrics”:
    - Distanza fisica, delay, etc…
  - Deve essere un algoritmo dinamico che si adatti, e velocemente ai cambi di topologia
  - Supportare Routing basato su tipo di servizio
    - Es. Routing di traffico real time da una parte e altro traffico dall’altra.
    - NB Nessuno usa il campo QoS ma nessun protocollo lo usa. Lo stess vale per OSPF, infatti lo hanno poi tolto

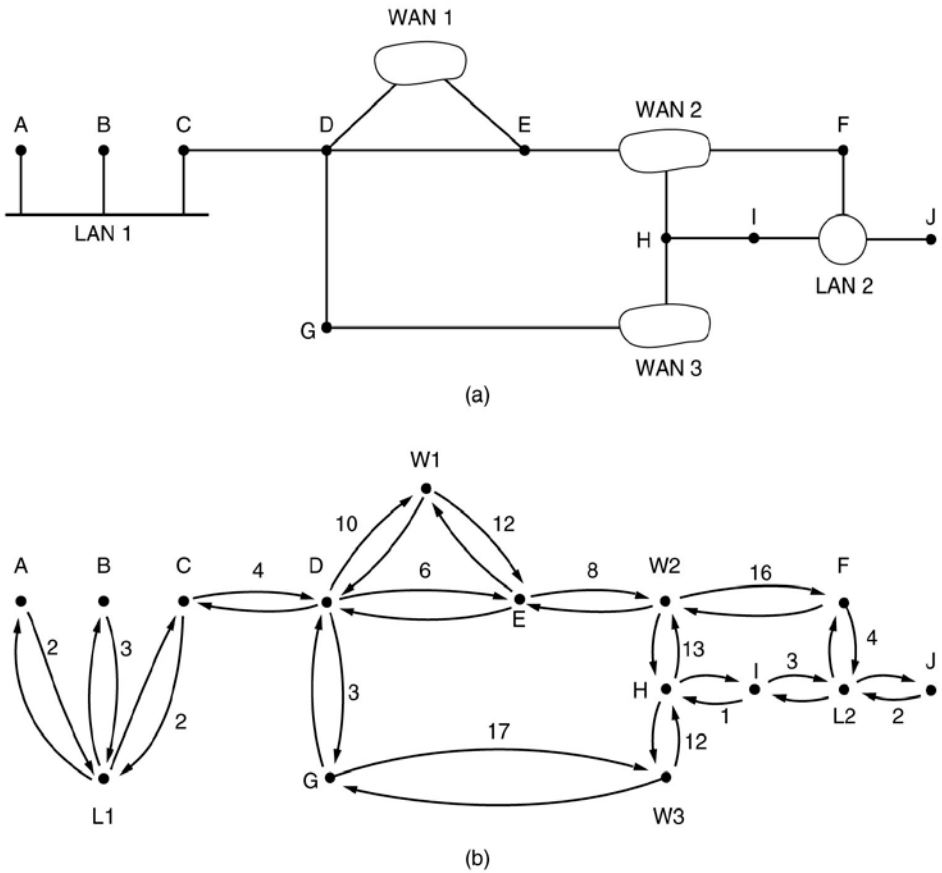
OSPF

- Open Shortest Path First. Requirements:
  - Deve fare “Load Balancing”, separando il carico su diversi link. Di solito viene usato solo la best route
  - Supporto per sistemi gerarchici.
    - Un router non deve conoscere l’intera gerarchia di internet che nel 1988 era già cresciuta a dimensioni enormi
  - Elementi di security
    - Per evitare che studenti in vena di scherzi possano indurre i router a cambiare le informazioni di routing
  - Gestione di router connessi a Internet via tunneling

Tipi di rete supportati

- OSPF supporta tre tipi di rete
  - Link punto a punto tra esattamente due router
  - Reti multiple access con broadcasting (quasi tutte le LAN)
  - Reti multiple access senza broadcasting (molte WAN packet switched)
- Una rete multiple access ha diversi router
  - Ognuno dei quali può comunicare direttamente con tutti gli altri. Tutte le LAN e le WAN lo possono fare.

Rete Multi Access

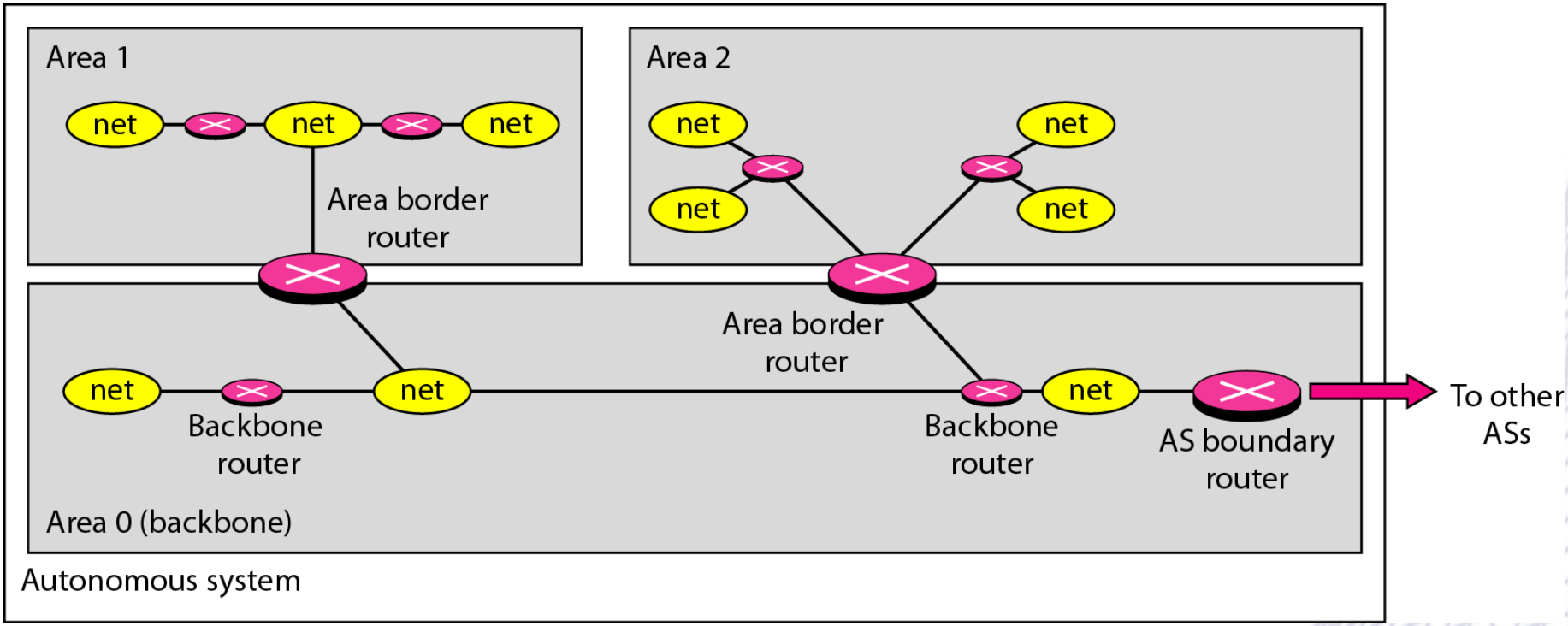


- A fianco sono presenti i 3 tipi di rete
- Gli host non hanno alcun ruolo, solo i router
- OSPF fa un'astrazione della rete con un grafo in cui ad ogni arco viene assegnato un costo (distanza, delay, etc.)
- Una connessione tra due router è una coppia di archi, uno in ogni direzione. I pesi possono essere diversi
- Una rete multiaccess è rappresentata da un nodo per la rete e un nodo per ogni router

Area

- Molti AS di Internet sono a loro volta grandi e quindi non facili da gestire
  - Sono allora divisi in aree numerate
  - Un'**area** è una rete o un insieme di reti contigue
  - Le reti non si possono sovrapporre ma non devono essere esaustive. Ci possono essere router che non sono in alcuna area. Al di fuori di un area non si vedono i dettagli e la topologia interna di un'area

Un AS con aree OSPF



Backbone

- Ogni AS ha un area speciale l'area di **backbone** chiamata area 0
- Le aree sono connesse al backbone da speciali router di confine. Questo permette di andare da un'area dell'AS ad un altra attraverso il backbone
- Anche un tunnel viene rappresentato nel grafo come un arco ed ha un costo
- Ogni router che è connesso a due o più di due aree è parte del backbone.

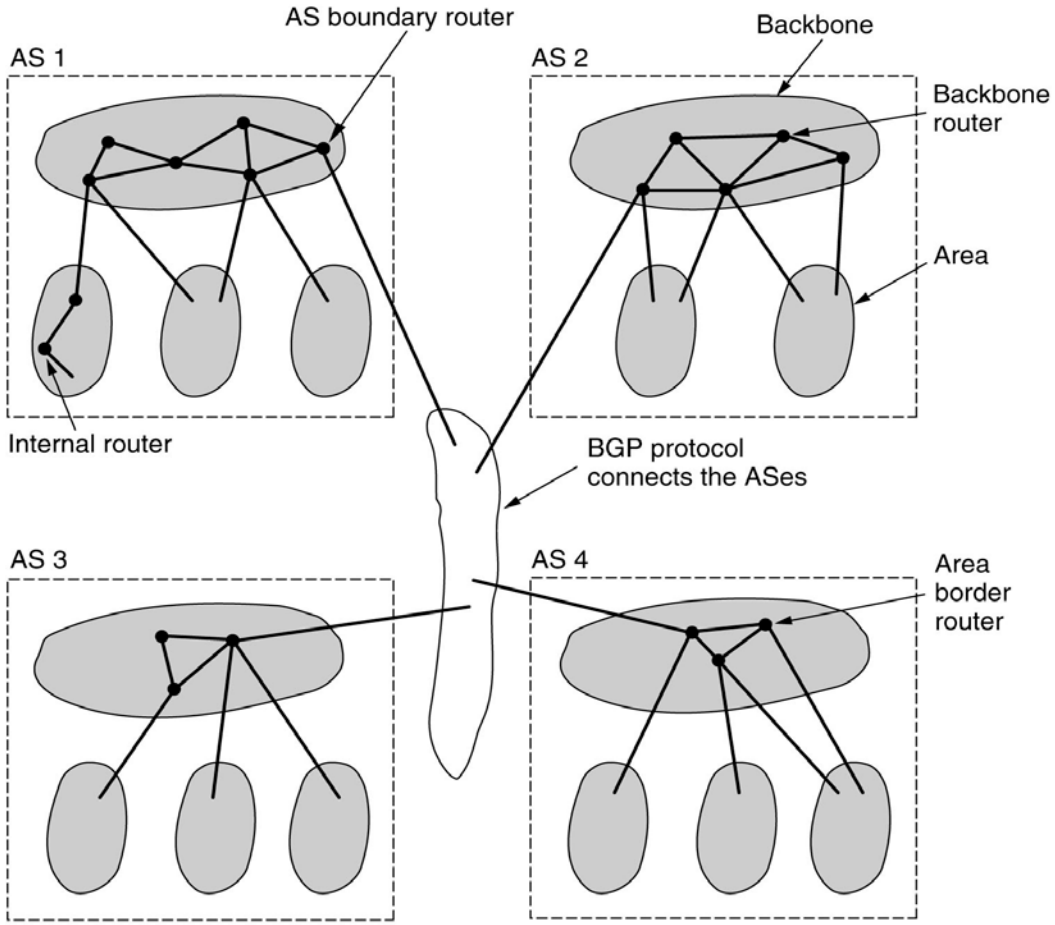
Link state database

- All'interno di un'area ogni router ha lo stesso link state database e usa lo stesso algoritmo "shortest path"
  - L'algoritmo ha la funzione di calcolare lo shortest path dal router stesso ad ogni altro router dell'area, compreso quello di backbone, di cui ce ne deve essere almeno uno
  - Un router che si connette a due aree deve avere i database di entrambe le aree e usare gli algoritmi separatamente per le due aree

Tre tipi di routes

- Intra-area
  - Le più facili dal momento che un router mittente conosce lo shortest path verso il destinatario
- Interarea
  - In tre passi: dal router al backbone, attraverso il backbone e fino al router di destinazione
  - Questo Impone una configurazione a stella con il backbone come hub e le aree come raggi
- Inter-AS

AS, backbone e aree



- Router interni completamente dentro un'area
- Router di bordo area che connettono due o più aree
- Router di backbone
- Router sui confini dell'AS che parlano con router in altri AS

Annuncio

- Al momento del Boot, il router manda un messaggio “HELLO” su tutte le linee punto-punto e in multicast agli altri router sulla LAN
- Su WAN ha bisogno di qualche configurazione per sapere chi contattare
- Dalle risposte che ottiene il router impara chi sono i suoi vicini
- I Router sulla LAN sono tutti suoi vicini

Router adiacenti

- Lo scambio di informazioni è tra router adiacenti che non è lo stesso che router vicini
- Sarebbe inefficiente che ogni router della LAN parlasse con tutti gli altri router della LAN.
- Un router viene eletto “**designated router**” e viene detto adiacente a tutti gli altri router della LAN e scambia informazioni con essi
- Router vicini che non sono adiacenti non si scambiano informazioni
- Per facilitare la transizione nel caso il designated router dovesse cadere viene anche eletto un designated router di backup

Link State Update

- Ogni router manda periodicamente dei messaggi “LINK STATE UPDATE” ai router adiacenti
- Questi messaggi forniscono gli stati e i costi usati nel database della topologia della rete
- Per aumentare l’affidabilità ognuno di questi messaggi ha un acknowledgement
- Inoltre ha un numero di sequenza per capire se è più vecchio di quelli che ha in tabella
- Un Router manda questi messaggi anche quando una linea cambia di stato (up – down) o il suo costo cambia

Database Description

- I messaggi “DATABASE DESCRIPTION” forniscono i numeri in sequenza di tutte le entries “link state” in possesso del mittente
- Comparando i suoi valori con quelli del mittente, il ricevente determina chi ha i valori più aggiornati
- Sono mandati quando una linea passa da down a up

Link State Request

- Un router può richiedere informazioni di link state da altri router mandandogli un messaggio “LINK STATE REQUEST”
- Il risultato è che ogni coppia di router adiacenti controlla chi dei due ha le informazioni più recenti
- Le nuove informazioni si diffondono in tutta l’area in questo modo

I messaggi

- I cinque tipi di messaggio vengono tramessi come pacchetti IP

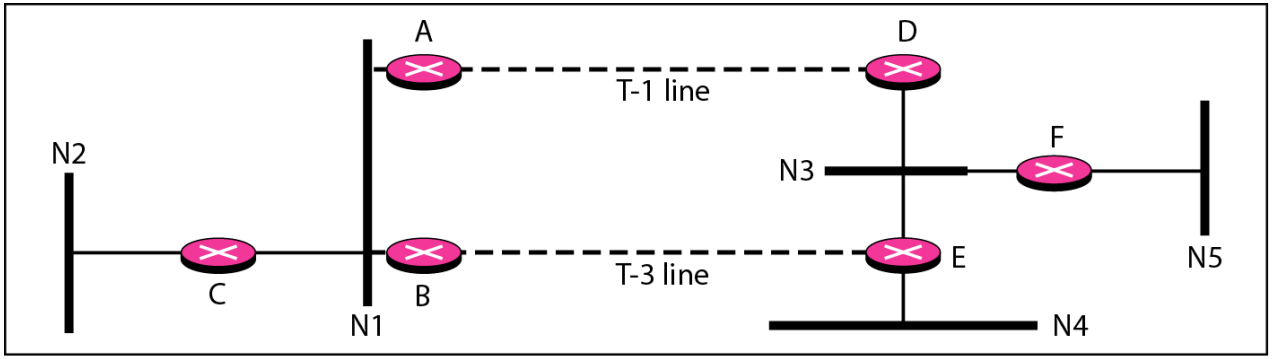
Message type	Description
Hello	Used to discover who the neighbors are
Link state update	Provides the sender's costs to its neighbors
Link state ack	Acknowledges link state update
Database description	Announces which updates the sender has
Link state request	Requests information from the partner

Mettiamo tutto insieme

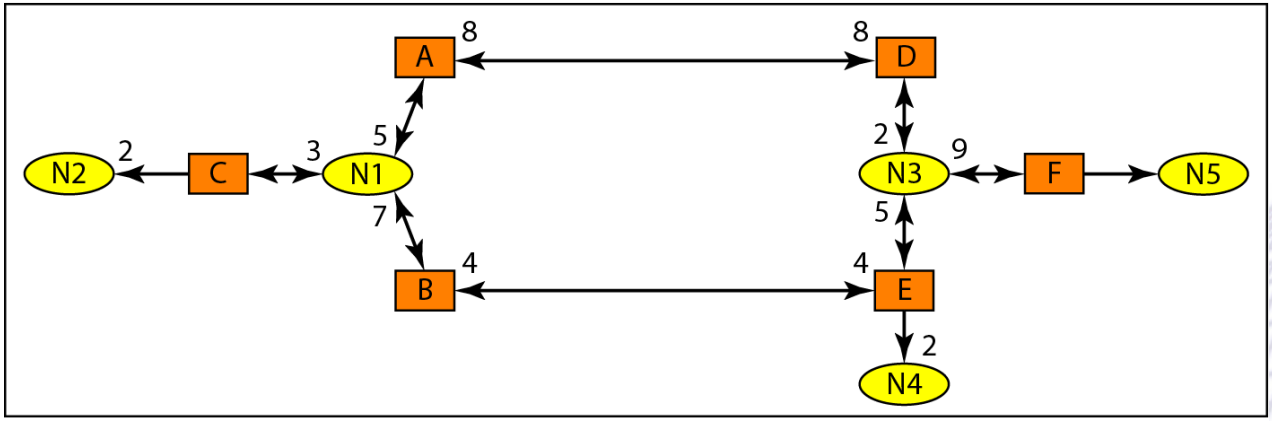
- Con il flooding del HELLO ogni router informa gli altri router nell'area dei suoi vicini e dei loro costi
- Queste informazioni permettono ad ogni router di costruirsi un grafo per la sua area e calcolare il shortest path
- Inoltre i router di backbone accettano informazioni dai router di bordo-area per calcolare la migliore route da ogni router di backbone ad ogni altro router
- Queste informazioni poi vengono propagate indietro ai router di bordo-area che le distribuiscono all'interno dell'area
- Usando queste informazioni un router che sta mandando un pacchetto interarea può scegliere il miglior router di uscita verso il backbone

Rappresentazione grafica

- Un Autonomous system
  - Costituito da 7 reti e 6 router
  - Nella parte in basso la stessa rete come vista da OSPF



a. Autonomous system



b. Graphical representation

BGP

- Tra diversi AS viene usato un protocollo diverso da OSPF chiamato BGP
  - Border Gateway Protocol: RFC 1771 e RFC 1774
- Perché i requirements di un protocollo di gateway esterno sono diversi da quelli di un protocollo di gateway interno
- Nel primo caso sono più importanti le policies mentre nel secondo l'efficienza di routing dei pacchetti tra i router

Scelte politiche

- Un AS di una impresa vuole poter mandare e ricevere pacchetti verso e da qualsiasi sito Internet
- Tuttavia potrebbe non volere fare da transito tra due AS estranei, anche se per caso il suo AS dovesse essere nello shortest path tra i due AS estranei
- D'altra parte potrebbe voler fare da transito per i suoi vicini o anche per altri AS se viene pagato per questo servizio
  - Per esempio le compagnie telefoniche potrebbero voler fare da transito per i suoi utenti ma non per altri

Esempi di policy

- Comprendono considerazioni politiche, sociali ed economiche
  - Non transitare attraverso alcuni AS (da un ISP non passare attraverso un ISP concorrente
  - Non mettere l'Iran in una route che parte dal Pentagono
  - Non passare attraverso gli USA per andare da Ontario a British Columbia
  - Non passare per l'Albania se ci sono alternative (condizionale)
  - Traffico che si origina in IBM non deve transitare per Microsoft
- Le policies sono configurate a mano in ogni router BGP ma non sono parte del protocollo

Tipi di rete

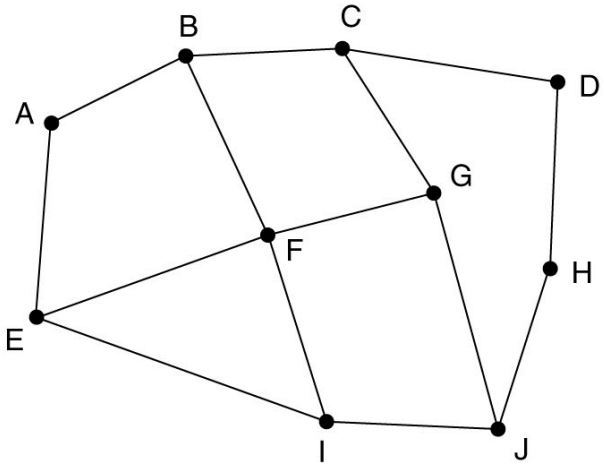
- Un router BGP vede un mondo fatto di AS e di linee che li connettono
  - Cioè se esiste un link tra un border router di ognuno dei due AS
- Si possono dividere in tre categorie di rete
  - **Stub networks**: Ha un'unica connessione al grafo BGP. Non possono essere usati per transito perché non c'è nulla all'altro lato
  - **Multiconnected network**: Possono essere usate per transitare traffico eccetto il traffico che rifiutano
  - **Transit network**: come i backbone che sono disposti a gestire traffico di terzi, forse con restrizioni e di solito a pagamento

Funzionamento

- Coppie di router BGP si parlano stabilendo connessioni TCP
  - In questo modo comunicano in modo affidabile senza vedere i vari dettagli delle reti che attraversano
  - Si tratta di un protocollo “distance vector” ma è molto diverso da RIP
  - Invece di tenere solo i costi per ogni destinazione ogni router BGP tiene traccia anche del path usato
  - Analogamente invece di mandare periodicamente ad ogni vicino il costo stimato per ogni destinazione ogni router BGP dice ai suoi vicini il path esatto che sta usando

Esempio

- Vediamo la tabella di F che usa FGCD per andare a D
- Quando i router vicini danno le loro tabelle, forniscono i loro path completi (b)
- F riesamina i diversi path e scarta subito quelli di I ed E perché passano per F stesso
- Quindi sceglie tra B e G



(a)

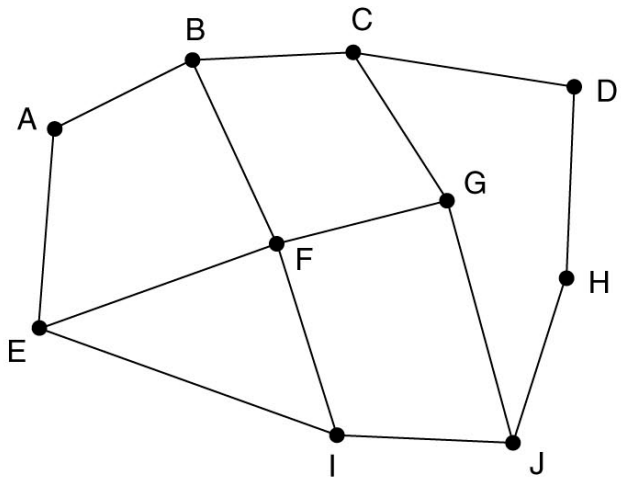
Information F receives from its neighbors about D

From B: "I use BCD"
From G: "I use GCD"
From I: "I use IFGCD"
From E: "I use EFGCD"

(b)

Punteggio

- Ogni router contiene un modulo che esamina le routes per una certa destinazione ed assegna un punteggio per la “distanza” verso quella destinazione
- Ogni route che viola una policy prende un punteggio infinito
- Alla fine il router sceglie la distanza minore
- La funzione che assegna i punteggi non fa parte del protocollo BGP e può essere una qualsiasi funzione scelta dall’amministratore



(a)

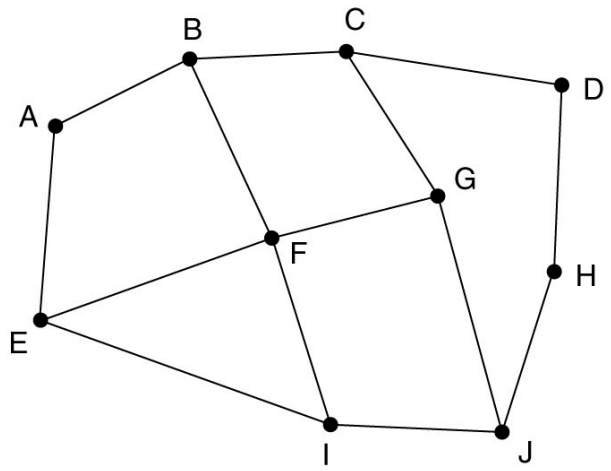
Information F receives from its neighbors about D

From B: "I use BCD"
From G: "I use GCD"
From I: "I use IFGCD"
From E: "I use EFGCD"

(b)

Convergenza

- BGP risolve il problema di conteggio all'infinito che affligge gli altri algoritmi "distance vector"
- Se G va in crash o la line FG va down allora F riceve la route dagli altri vicini rimasti, BCD, IFGCD e EFGCD
- Vede subito che le ultime due sono inutili visto che passano per F stesso quindi sceglie FBCD come nuova route
- Altri algoritmi distance vector spesso fanno la scelta sbagliata perché non sanno quale dei vicini ha rotte indipendenti verso la destinazione e quali non ne hanno



(a)

Information F receives
from its neighbors about D

From B: "I use BCD"
From G: "I use GCD"
From I: "I use IFGCD"
From E: "I use EFGCD"

(b)

Internet Multicasting

- Per alcune applicazioni un processo potrebbe voler mandare gli stessi pacchetti a destinazioni multiple
  - Es update di sistema, database distribuiti, dati di borsa, teleconferenze digitali, distribuzione di notizie, tele-learning
- IP supporta multicasting con gli indirizzi di classe D
  - 28 bit sono disponibili per identificare gruppi quindi possono esistere più di 250 Milioni
  - Quando un processo manda un pacchetto a un indirizzo di classe D, viene fatto un tentativo best-effort di consegnarlo a tutti i membri del gruppo indirizzato, ma senza garanzie. Alcuni membri potrebbero non ricevere il pacchetto

Gruppi permanenti

- Ci sono due tipi di indirizzi:
- Permanenti che ci sono sempre e non devono essere creati
- Ogni gruppo permanente ha un indirizzo di gruppo permanente. Per Es:
  - 224.0.0.1 Tutti i sistemi in una LAN
  - 224.0.0.2 Tutti i router di una LAN
  - 224.0.0.5 Tutti i router OSPF di una LAN
  - 244.0.0.6 Tutti i designated router di una LAN

Gruppi Temporanei

- I gruppi Temporanei devono essere creati prima di essere usati
  - Un processo può chiedere al suo host di unirsi ad uno specifico gruppo o di lasciare un gruppo
  - Quando l'ultimo processo di un host lascia un gruppo, il gruppo cessa di esistere in quell'host
  - Ogni host tiene traccia di quali gruppi i suoi processi facciano parte

Router Multicast

- Il Multicast viene implementato da speciali router multicast che possono essere co-locati con i normali router
- Una volta al minuto ogni multicast router manda un multicast hardware (cioè a livello data link) agli host nella sua LAN (224.0.0.1) chiedendo di rispondere a quali gruppi i loro processi appartengano
- Ogni host risponde con tutti gli indirizzi di classe D a cui sono interessati

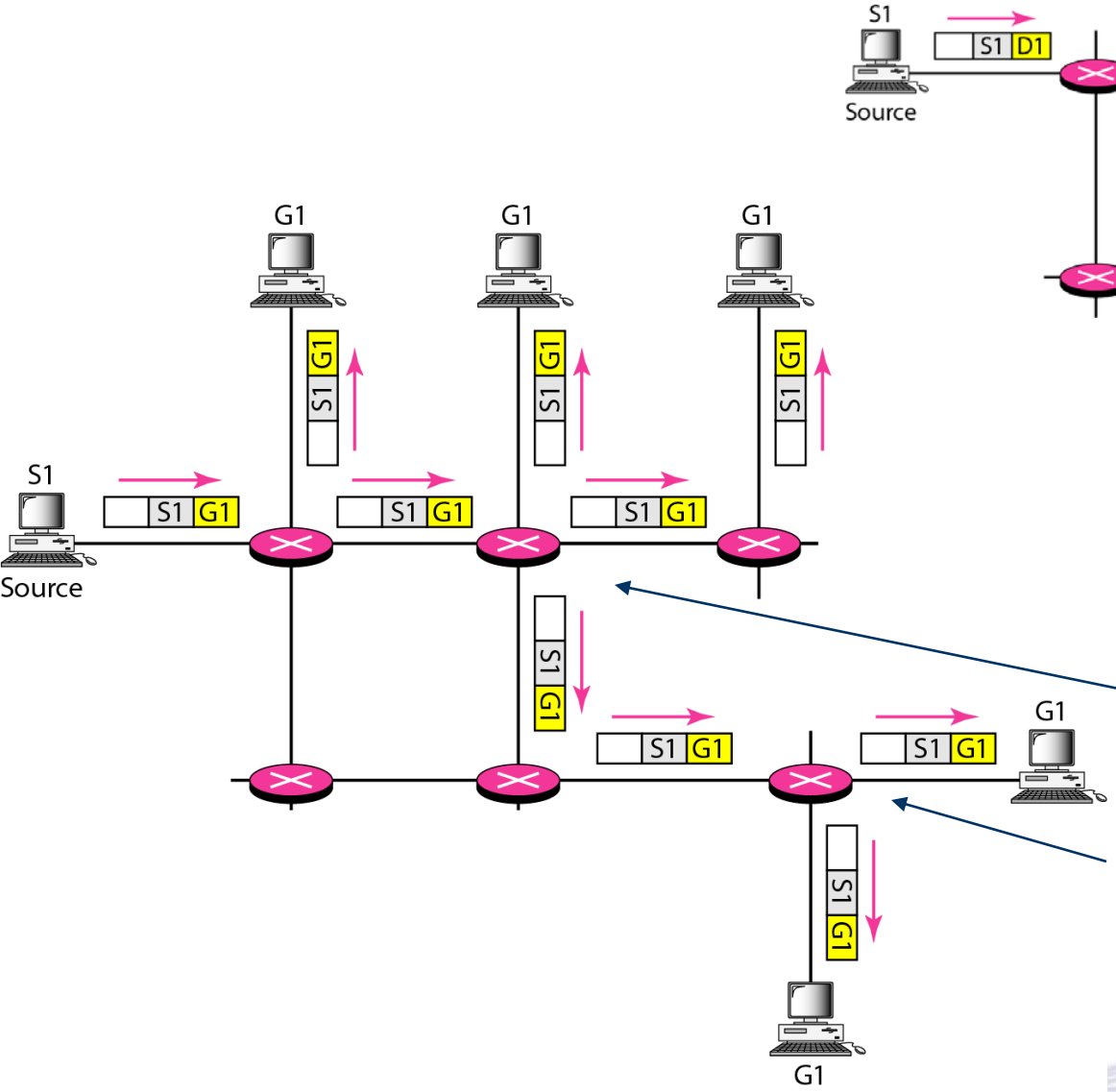
IGMP: RFC 1112

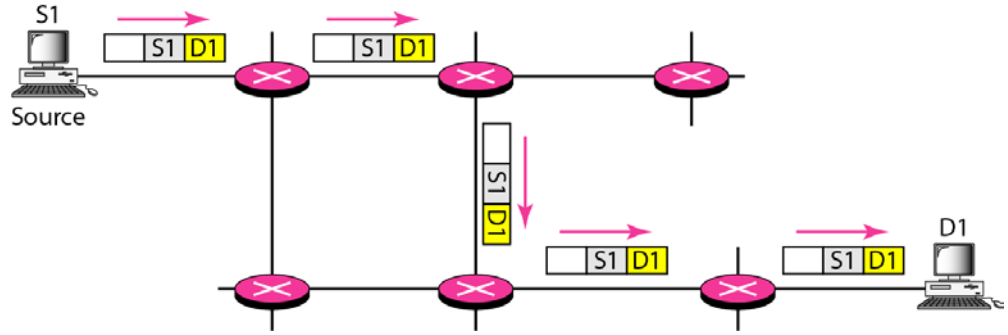
- Questi pacchetti di richiesta e risposta usano un protocollo chiamato IGMP (Internet Group Management Protocol) che somiglia vagamente a ICMP
- Ha solo due tipi di pacchetti: **query** e **response**, ognuno con un formato fisso e molto semplice, che contiene alcune informazioni di controllo nella prima parola del payload e un indirizzo di classe D nella seconda

Routing Multicast

- Il routing multicast viene fatto usando spanning tree. Ogni multicast router scambia informazioni con i suoi vicini, usando un protocollo **distance vector** modificato in modo che ognuno si possa costruire uno spanning tree per ogni gruppo
- Ci sono diverse ottimizzazioni per potare gli alberi ed eliminare i router e le reti che non sono interessate ad un particolare gruppo
- Il protocollo usa pesantemente il tunneling per non disturbare nodi che non sono nello spanning tree

Routing multicast



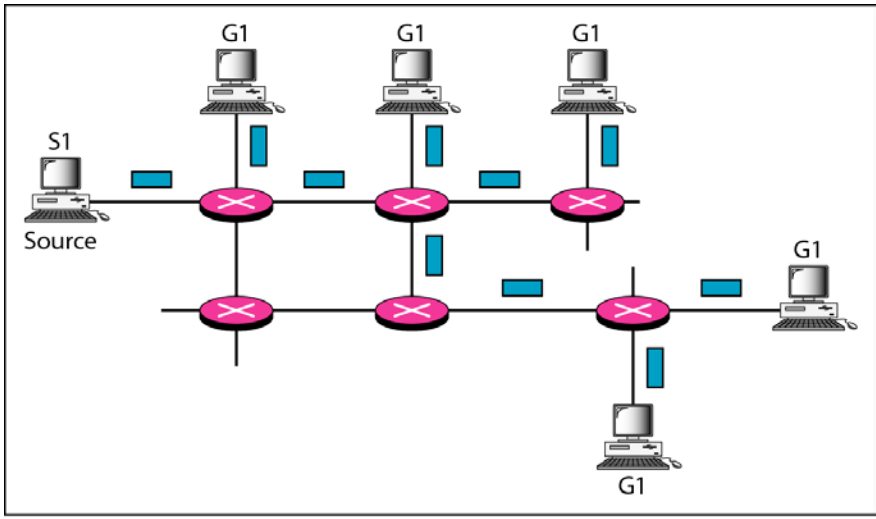


Nel protocollo unicast il router manda il pacchetto su di un'unica interfaccia

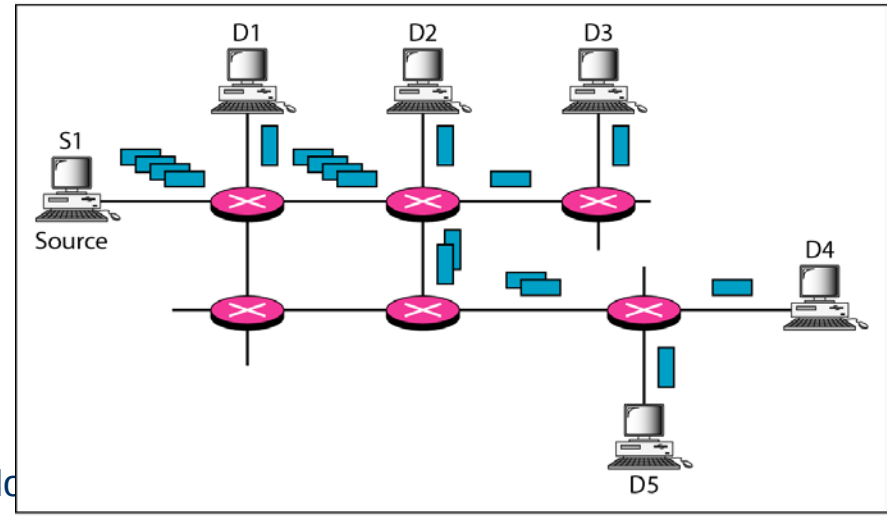
Nei protocolli multicast un router può mandarlo su una o più interfacce

Multicast e unicast multiplo

- Con il multicast si manda un unico pacchetto a molti destinatari
- Posso ottenere lo stesso scopo mandando pacchetti multipli in modo che ne arrivi uno ad ogni destinatario
- Quindi il multicast è più efficiente in termini di utilizzo della banda
- Inoltre tra le varie copie di un pacchetto ci sarà un pur minimo ritardo. Se il numero di copie è elevato il ritardo di spedizione potrebbe essere inaccettabile
- Invece con multicast non c'è ritardo perché viene spedito un unico pacchetto



a. Multicasting



b. Multiple unicasting