Reti di calcolatori e Internet: Un approccio top-down

7^a edizione Jim Kurose, Keith Ross

Pearson Paravia Bruno Mondadori Spa

- Introduzione
- Algoritmi di instradamento
 - Stato del collegamento
 - Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

- Introduzione
- Algoritmi di instradamento
 - Stato del collegamento
 - Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

Funzionalità del livello Rete

Il livello rete ha due funzioni:

- forwarding: sposta i pacchetti dal router di input ad un appropriato router di output
- routing: determina il percorso che devono seguire i pacchetti dalla sorgente alla destinazione

Piano dei dati

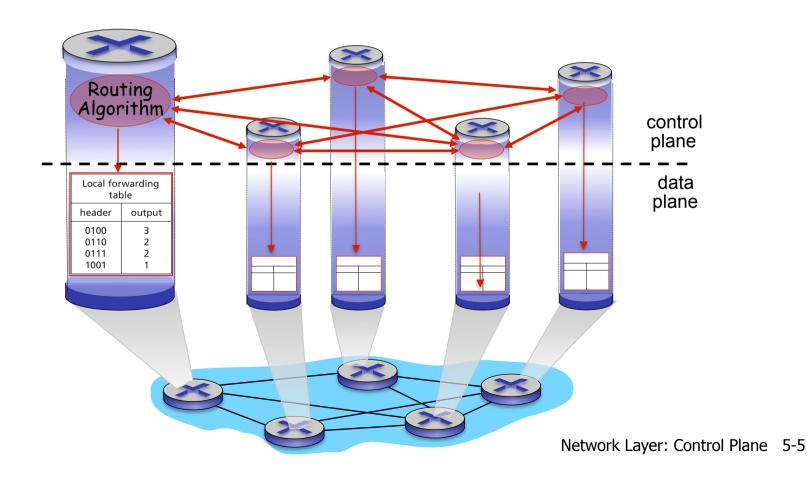
Piano di controllo

Due approcci per il piano di controllo:

- Controllo per-router (tradizionale)
- Controllo centralizzato (software defined networks)

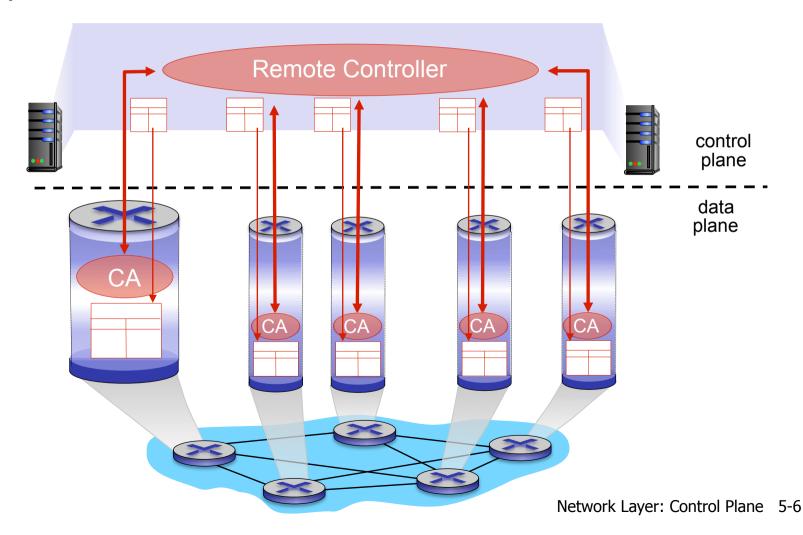
Piano di controllo Per-router

In ogni router viene eseguito un algoritmo di routing. I router interagiscono tra loro per calcolare le tabelle di inoltro



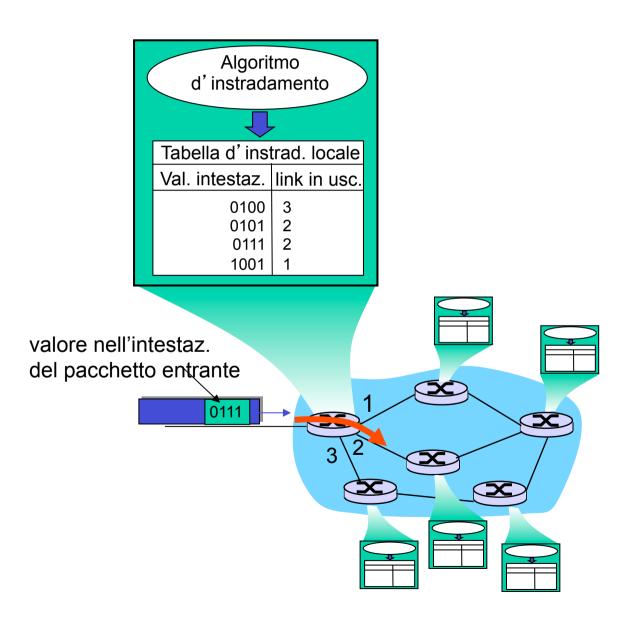
Piano di controllo centralizzato

Un controllore remoto interagisce con gli agenti di controllo locali presenti nei router, indicando le tabelle di inoltro

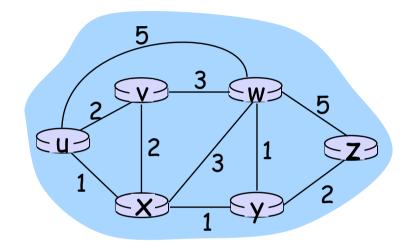


- □ Introduzione
- Algoritmi di instradamento
 - Stato del collegamento
 - Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

Algoritmi d'instradamento



Grafo di una rete di calcolatori



Grafo: G = (N,E)

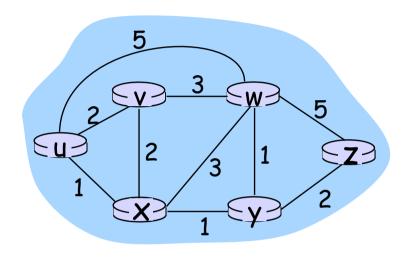
N = insieme di nodi (router) = { u, v, w, x, y, z }

 $E = insieme di archi (collegamenti) = \{ (u,v), (u,x), (v,x), (v,w), (x,w), (x,y), (w,y), (w,z), (y,z) \}$

N.B.: Il grafo è un'astrazione utile anche in altri contesti di rete

Esempio: P2P, dove N è un insieme di peer ed E è un insieme di collegamenti TCP

Grafo di una rete: costi



- c(x,x') = costo del collegamento (x,x')
 - es., c(w,z) = 5
- il costo di un cammino è semplicemente la somma di tutti i costi degli archi lungo il cammino

Costo di un cammino $(x_1, x_2, x_3, ..., x_p) = c(x_1, x_2) + c(x_2, x_3) + ... + c(x_{p-1}, x_p)$

Domanda: Qual è il cammino a costo minimo tra u e z?

Algoritmo d'instradamento: determina il cammino a costo minimo.

Classificazione degli algoritmi d'instradamento

Globale o decentralizzato?

- L'algoritmo riceve in ingresso tutti i collegamenti tra i nodi e i loro costi.
- Algoritmi a stato del collegamento (link-state algorithm).

Decentralizzato:

- Ogni nodo elabora un vettore di stima dei costi (distanze) verso tutti gli altri nodi nella rete.
- Il cammino a costo minimo viene calcolato in modo distribuito e iterativo.
- Algoritmo a vettore distanza (VC, distance-vector algorithms)

Statico o dinamico?

Statico:

 I cammini cambiano molto raramente.

Dinamico:

- Determinano gli instradamenti al variare di:
 - Volume di traffico
 - Topologia della rete

- □ Introduzione
- Algoritmi di instradamento
 - Stato del collegamento (Link State)
 - Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

Algoritmo d'instradamento a stato del collegamento (LS)

Algoritmo di Dijkstra:

- La topologia di rete e tutti i costi dei collegamenti sono noti a tutti i nodi
 - attraverso il "link-state broadcast".
 - tutti i nodi dispongono delle stesse informazioni
- □ Calcola il cammino a costo minimo da un nodo (origine) a tutti gli altri nodi della rete.
 - Crea una tabella d'inoltro per quel nodo
- ☐ È iterativo: dopo la k-esima iterazione i cammini a costo minimo sono noti per k nodi di destinazione.

Definiamo la seguente notazione:

- c(x,y): costo del collegamenti dal nodo x al nodo y; = ∞ se non sono adiacenti.
- D(v): costo del cammino dal nodo origine alla destinazione v per quanto riguarda l'iterazione corrente.
- p(v): immediato predecessore di v lungo il cammino.
- N': sottoinsieme di nodi per cui il cammino a costo minimo dall'origine è definitivamente noto.

Algoritmo di Dijsktra

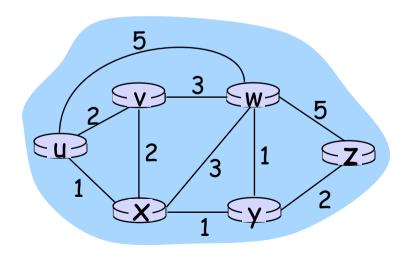
```
Inizializzazione:
   N' = \{u\}
  per tutti i nodi v
    se v è adiacente a u
       allora D(v) = c(u,v)
6
    altrimenti D(v) = ∞
   Ciclo
    determina un w non in N' tale che D(w) sia minimo
    aggiungi w a N'
    aggiorna D(v) per ciascun nodo v adiacente a w e non in N':
12 D(v) = min(D(v), D(w) + c(w,v))
13 /* il nuovo costo verso v è il vecchio costo verso v oppure
14 il costo del cammino minimo noto verso w più il costo da w a v */
15 Finché N' = N
```

Algoritmo di Dijkstra: esempio

Ste	p N '	D(v) p(v)	D(w) p(w)	D(x) p(x)	D(y) p(y)	D(z) p(z)		
0	u	7,u	3,u	5,u	∞	∞		
1	uw	6,w		5,u) 11,W	∞		
<u>2</u> 3	uwx	6,w)		11,W	14,x		
3	uwxv				10,	14,x		
4	uwxvy					12,		
5	uwxvyz							X
							5	4
						u	3	*** \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \

Algoritmo di Dijkstra: esempio

pass	so	N'	D(v),p(v)	D(w),p(w)	D(x),p(x)	D(y),p(y)	D(z),p(z)
	0	u	2,u	5,u	1,u	∞	∞
	1	ux←	2,u	4,x		2,x	∞
	2	uxy	2,u	3,y			4,y
	3	uxyv		3,y			4,y
	4	uxyvw 🗲		_			4,y
	5	uxyvwz 🕶					



Algoritmo di Dijkstra: esempio

Cammino a costo minimo da u:

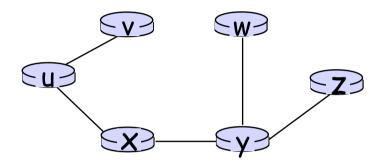


Tabella d'inoltro in u:

destinazione	collegamento		
V	(u,v)		
X	(u,x)		
У	(u,x)		
w	(u,x)		
z	(u,x)		

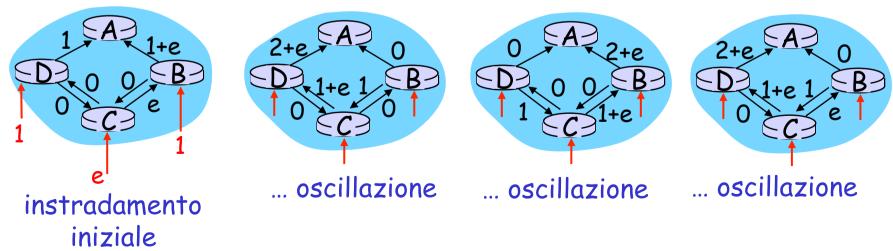
Algoritmo di Dijkstra: discussione

Complessità dell'algoritmo: n nodi

- Ciascuna iterazione: controllo su tutti i nodi, w, non in N
- \Box n(n+1)/2: $O(n^2)$
- □ La più efficiente implementazone possibile: O(nlogn)

Possibili oscillazioni:

□ Es. costo del collegamento = quantità di traffico trasportato



- □ Introduzione
- Algoritmi di instradamento
 - Stato del collegamento (Link State)
 - Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

Algoritmo d'instradamento con vettore distanza (DV)

<u>Formula di Bellman-Ford (programmazione dinamica)</u> definisce

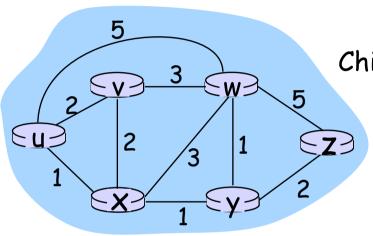
 $d_x(y) := il costo del percorso a costo minimo dal nodo <math>x$ al nodo y.

Allora:

$$d_{x}(y) = \min_{x} \{c(x, y) + d_{y}(y)\}$$

dove min, riguarda tutti i vicini di x.

Formula di Bellman-Ford: esempio



Chiaramente, $d_v(z) = 5$, $d_x(z) = 3$, $d_w(z) = 3$

La formula B-F definisce:

$$d_{u}(z) = \min \{ c(u,v) + d_{v}(z), \\ c(u,x) + d_{x}(z), \\ c(u,w) + d_{w}(z) \} \\ = \min \{ 2 + 5, \\ 1 + 3, \\ 5 + 3 \} = 4$$

Algoritmo con vettore distanza

- $D_x(y)$ = stima del costo del percorso a costo minimo da se stesso al nodo y.
- \square Vettore distanza: $D_x = [D_x(y): y \text{ in } N]$
- \square Il nodo x conosce il costo verso ciascun vicino v: c(x,v)
- \square II nodo x mantiene $D_x = [D_x(y): y \text{ in } N]$
- □ Il nodo x mantiene anche i vettori distanza di ciascuno dei suoi vicini
 - Per ciascun vicino v, x mantiene $D_v = [D_v(y): y \text{ in } N]$

Algoritmo con vettore distanza

Idea di base:

- Ogni nodo invia una copia del proprio vettore distanza a ciascuno dei suoi vicini.
- Quando un nodo x riceve un nuovo vettore distanza, DV, da qualcuno dei sui vicini, lo salva e usa la formula B-F per aggiornare in proprio vettore distanza come segue:

$$D_{x}(y) \leftarrow \min_{v} \{c(x,v) + D_{v}(y)\}$$
 per ciascun nodo y in N.

Finché tutti i nodi continuano a cambiare i propri DV in maniera asincrona, ciascuna stima dei costi Dx(y) converge a $d_x(y)$.

Algoritmo con vettore distanza

Iterativo, asincrono: ogni iterazione locale è causata da:

- cambio del costo di uno dei collegamenti locali
- ricezione da qualche vicino di un vettore distanza aggiornato.

Distribuito:

- Ogni nodo aggiorna i suoi vicini solo quando il suo DV cambia.
 - i vicini avvisano i vicini solo se necessario.

Ciascun nodo:

Attende (un messaggio del cambio del costo da parte del suo vicino)

Effettua il calcolo

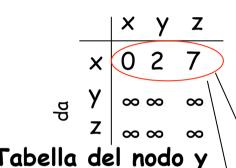
Se il DV cambia,
lo notifica ai suoi vicini.

$$\begin{split} D_x(y) &= \min\{c(x,y) + D_y(y), c(x,z) + D_z(y)\} \\ &= \min\{2+0, 7+1\} = 2 \end{split}$$

da

 $D_{x}(z) = \min\{c(x,y) +$ $D_{y}(z)$, $c(x,z) + D_{z}(z)$ = min{2+1, 7+0} = 3

Tabella del nodo x

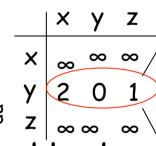


costo verso

costo versø

	X	Y	/z
X	0	2	3
У	2	0	1
Z	7	1	0

Tabella del nodo y costo verso



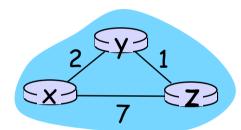
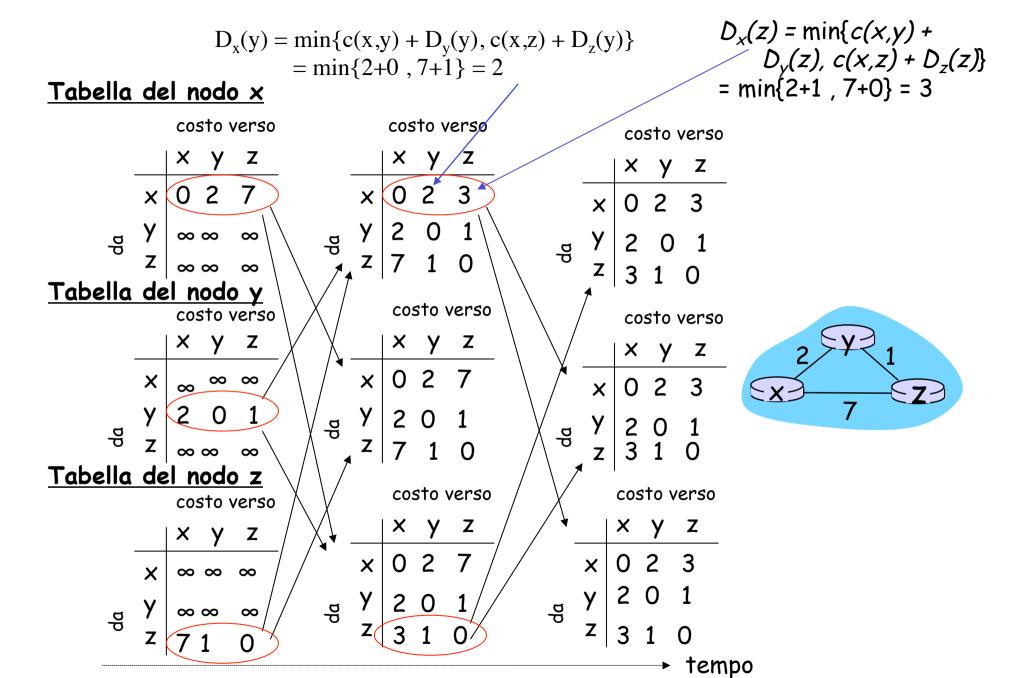


Tabella del nodo z

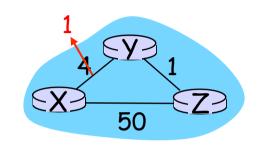
costo verso Z $\infty \infty \infty$

tempo



Distance Vector: cambio del costo dei link

- □ Il nodo rileva che un link a cui è direttamente connesso cambia il costo
- Aggiorna il vettore distnza
- Se il costo è cambiato nel percorso minimo, lo notifica ai vicini

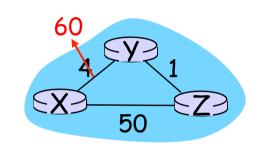


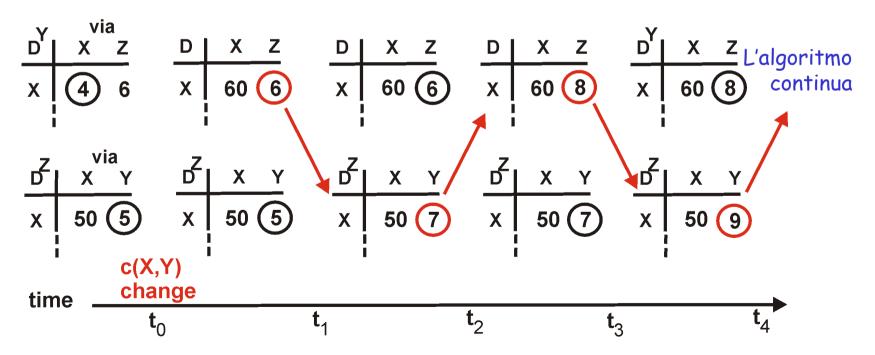
"le buone $x \mid 4 \mid 6$ $x \mid 1 \mid 6$ $x \mid 1 \mid 6$ $x \mid 1 \mid 3$ notizie viaggiano $\frac{z}{x} \mid \frac{via}{x} \mid y$ $\frac{z}{x} \mid 50 \mid 5$ $x \mid 50 \mid 5$ $x \mid 50 \mid 2$ $x \mid 50 \mid 2$ $x \mid 50 \mid 2$ $x \mid 50 \mid 2$

L'algoritmo termina

Distance Vector: cambio del costo dei link

 □ Le cattive notizie viaggiano lentamente - problema del "conteggio all'infinito"

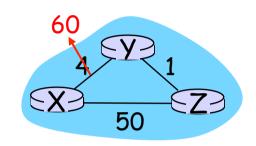


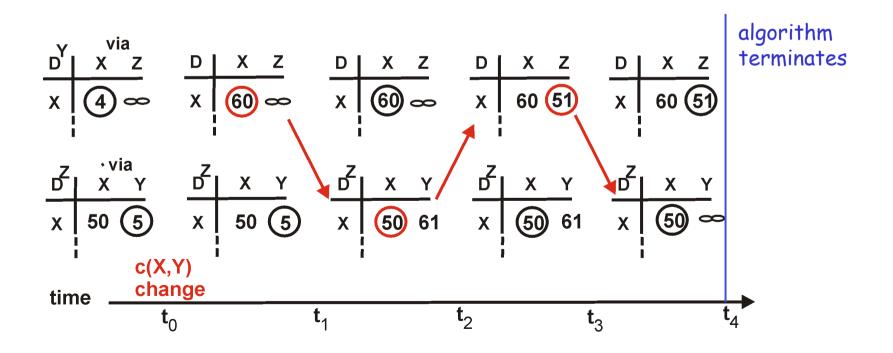


Distance Vector: inversione avvelenata

If Z passa da Y per andare X:

 Z dice a Y che la sua distanza (di Z) per X è infinita (così Y non vuole passare da Z per raggiungere X)





Confronto tra gli algoritmi LS e DV

Complessità dei messaggi:

- LS: con n nodi, E collegamenti, implica l'invio di O(nE) messaggi.
- <u>DV:</u> richiede scambi tra nodi adiacenti.

Velocità di convergenza:

- □ <u>LS:</u> l'algoritmo O(n²) richiede O(nE) messaggi.
- DV: può convergere lentamente.
 - può presentare cicli d'instradamento.
 - può presentare il problema del conteggio all'infinito.

Robustezza: cosa avviene se un router funziona male?

LS:

- un router può comunicare via broadcast un costo sbagliato per uno dei suoi collegamenti connessi (ma non per altri).
- i nodi si occupano di calcolare soltanto le proprie tabelle.

DV:

- un nodo può comunicare cammini a costo minimo errati a tutte le destinazioni.
- la tabella di ciascun nodo può essere usata degli altri.
 - Un calcolo errato si può diffondere per l'intera rete.

- □ Introduzione
- Algoritmi di instradamento
 - Stato del collegamento (Link State)
 - Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

Instradamento gerarchico

Abbiamo fin qui visto la rete come una collezione di router interconnessi

- Ciascun router era indistinguibile dagli altri
- Visione omogenea della rete

... nella pratica le cose non sono così semplici

Scala: con 200 milioni di destinazioni:

- Archiviare le informazioni d'instradamento su ciascun host richiederebbe un'enorme quantità di memoria.
- □ Il traffico generato dagli aggiornamenti LS non lascerebbero banda per i pacchetti di dati!

Autonomia amministrativa:

- 🗖 Internet = la rete delle reti
- Da un punto di vista ideale, ciascuno dovrebbe essere in grado di amministrare la propria rete nel modo desiderato, pur mantenendo la possibilità di connetterla alle reti esterne.

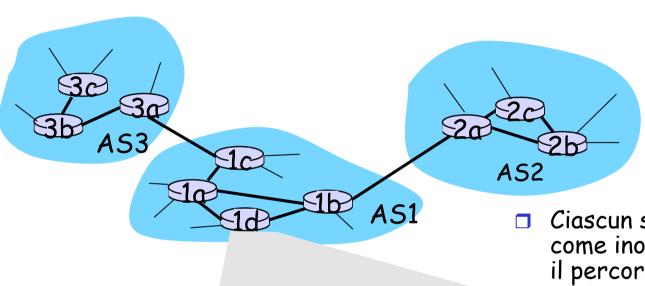
Instradamento gerarchico

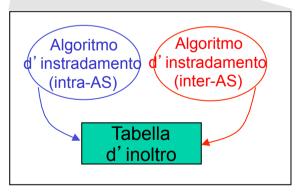
- Organizzazione di router in sistemi autonomi (AS, autonomous system).
- □ I router di un gruppo autonomo eseguono lo stesso algoritmo d'instradamento.
 - Protocollo d'instradamento interno al sistema autonomo (intra-AS).
 - I router appartenenti a differenti AS possono eseguire protocolli d'instradamento intra-AS diversi

Router gateway

□ Hanno il compito aggiuntivo d'inoltrare pacchetti a destinazioni esterne.

Sistemi autonomi interconnessi



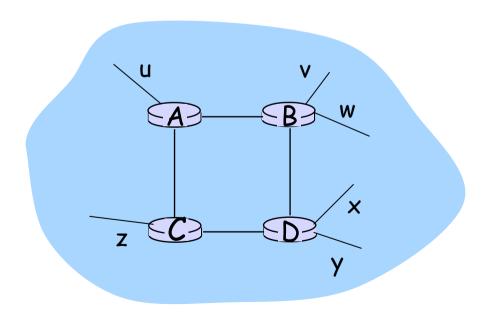


- Ciascun sistema autonomo sa come inoltrare pacchetti lungo il percorso ottimo verso qualsiasi destinazione interna al gruppo
 - I protocolli d'instradamento dei tre sistemi autonomi non sono necessariamente gli stessi
 - I router 1b, 1c, 2a e 3a sono gateway

- □ Introduzione
- □ Algoritmi di instradamento
 - O Stato del collegamento (Link State)
 - O Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

RIP (Routing Information Protocol)

- □ È un protocollo a vettore distanza.
- □ È tipicamente incluso in UNIX dal 1982.
- □ Conteggio degli hop come metrica di costo (max = 15 hop) (include la sottorete di destinazione)



Dal router A alle varie sottoreti:

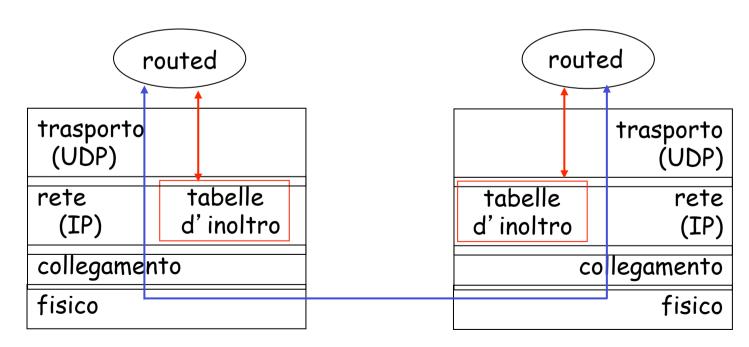
destinazione	hop
u	1
V	2
W	2
×	3
У	3
Z	2

Annunci RIP

- □ In RIP, i router adiacenti si scambiano gli aggiornamenti d'instradamento ogni 30 secondi circa utilizzando un messaggio di risposta RIP, noto anche come annuncio RIP (RIP advertisement).
- □ Ogni messaggio contiene un elenco comprendente fino a 25 sottoreti di destinazione all'interno del sistema autonomo nonché la distanza del mittente rispetto a ciascuna di tali sottoreti.

Tabella d'instradamento RIP

- Un processo chiamato routed esegue RIP, ossia mantiene le informazioni d'instradamento e scambia messaggi con i processi routed nei router vicini.
- Poiché RIP viene implementato come un processo a livello di applicazione, può inviare e ricevere messaggi su una socket standard e utilizzare un protocollo di trasporto standard.



Capitolo 5: Livello di rete (Piano di Controllo)

- □ Introduzione
- □ Algoritmi di instradamento
 - O Stato del collegamento (Link State)
 - O Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - o BGP
- □ ICMP: Internet Control Message Protocol

OSPF (Open Shortest Path First)

- open": le specifiche del protocollo sono pubblicamente disponibili.
- □ È un protocollo a stato del collegamento:
 - Utilizza il flooding di informazioni di stato del collegamento
 - Utilizza l'algoritmo di Dijkstra per la determinazione del percorso a costo minimo.
- Con OSPF, ogni volta che si verifica un cambiamento nello stato di un collegamento, il router manda informazioni d'instradamento a tutti gli altri router.
- Invia messaggi OSPF all'intero sistema autonomo, utilizzando il flooding.
- I messaggi OSPF vengono trasportati direttamente da IP (e non da TCP o UDP) (codice protocollo di livello superiore 89)
 - OSPF deve implementare i meccanismi di trasferimento affidabile

Capitolo 5: Livello di rete (Piano di Controllo)

- □ Introduzione
- □ Algoritmi di instradamento
 - O Stato del collegamento (Link State)
 - O Vettore distanza
 - Instradamento gerarchico
- Instradamento in Internet
 - O RIP
 - OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

Instradamento inter-AS in Internet: BGP

- □ BGP (Border Gateway Protocol): rappresenta l'attuale standard de facto.
- Basato su Distance-Vector
- 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
 - 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.

Instradamento tra sistemi autonomi

- □ Supponiamo che un router in AS1 riceva un datagramma la cui destinazione ricade al di fuori di AS1
 - Il router dovrebbe inoltrare il pacchetto verso uno dei due gateway. Ma quale??

AS1 deve:

- Sapere quali destinazioni sono raggiungibili attraverso AS2 e quali attraverso AS3
- 2. Informare tutti i router all'interno del sistema in modo che ciascuno possa configurare la propria tabella d'inoltro per gestire destinazioni esterne

Ed ecco a voi l'instradamento inter-AS!

altre reti

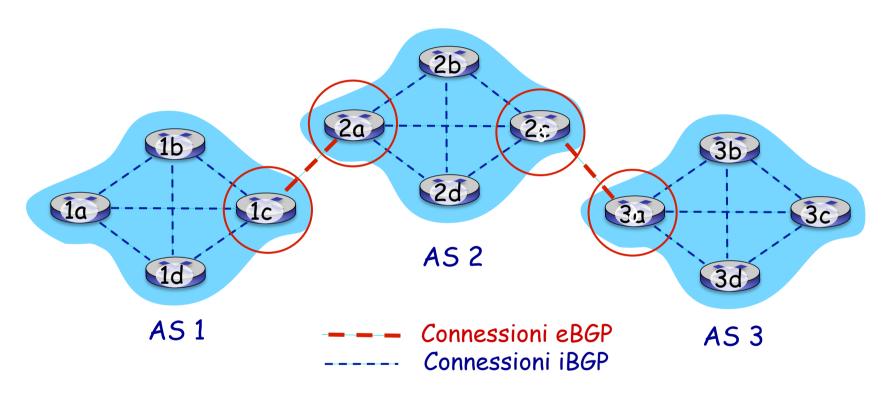
AS1

AS2

Ed ecco a voi l'instradamento inter-AS!

altre reti

eBGP, iBGP connections

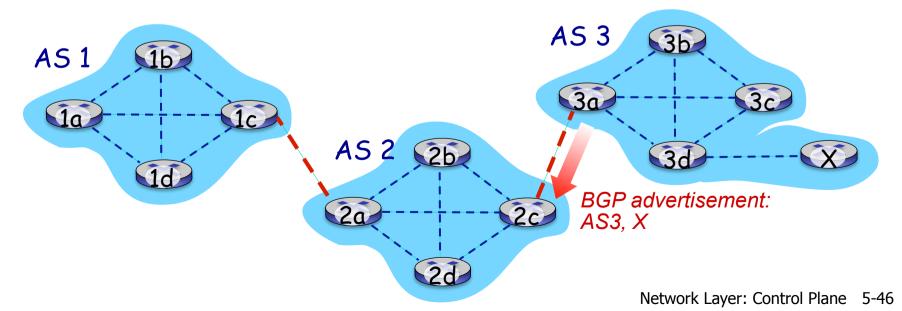




I router gateway eseguono entrambi i protocolli eBGP e iBGP

BGP basics

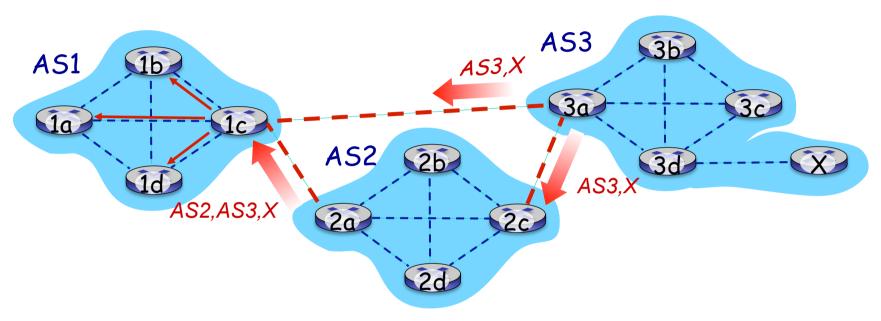
- Sessione BGP: due router BGP si scambiano messaggi tramite connessioni TCP:
 - Pubblicizzano un *percorso* verso differenti prefissi di rete (destinazione)
- □ Quando il router gateway 3a del AS3 pubblicizza il percorso AS3,X verso il gateway router 2c di AS2:
 - ○AS3 promette a AS2 che inoltrerà i datagrammi verso X



Attributi dei percorsi

- □ I messaggi rappresentano I percorsi come
 - Prefisso + attributi = "route"
- Due importanti attributi:
 - AS-PATH: lista degli AS da attraversare
 - NEXT-HOP: router interno al AS che rappresenta il nexthop
- □ Policy-based routing:
 - o i router gateway decidono se accettare o declinare un percorso in base alla *import policy* (e.g., mai instradare attraversio l'AS Y).
 - La policy determina anche a quali AS pubblicizzare i propri percorsi

BGP distribuzione delle informazioni

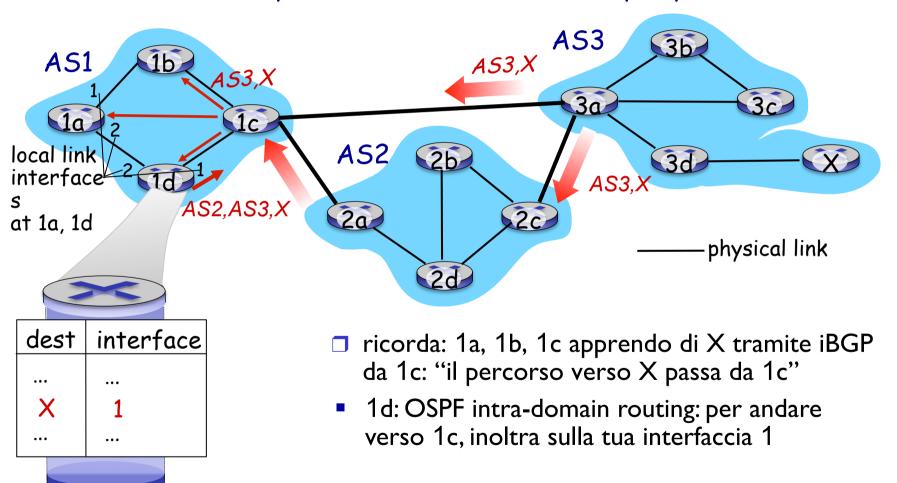


Il router gateway può apprendere percorsi multipli verso una destinazione

- □ Il router gateway 1c del AS1 apprende il percorso AS2,AS3,X da 2a
- Il router gateway 1c del AS1 apprende il percorso AS3,X da 3a
- In base alle policy del AS1 il router gateway 1c sceglie il percorso AS3,X, e annuncia il percorso all'interno di AS1 tramite iBGP

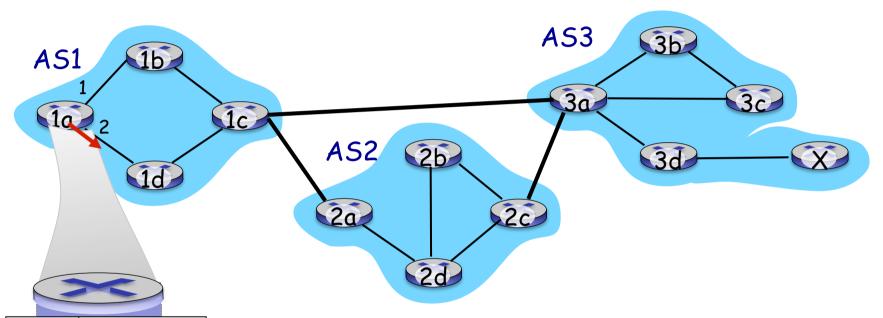
BGP, OSPF, forwarding table entries

Q: come i router impostano le tabelle di inoltro per prefissi distanti?



BGP, OSPF, forwarding table entries

Q: come i router impostano le tabelle di inoltro per prefissi distanti?



dest	interface
X	2
•••	

- □ ricorda: 1a, 1b, 1c apprendo di X tramite iBGP da 1c: "il percorso verso X passa da 1c"
- 1d: OSPF intra-domain routing: per andare verso 1c, inoltra sulla tua interfaccia 1
- 1a: OSPF intra-domain routing: per andare verso 1c, inoltra sulla tua interfaccia 2

Capitolo 5: Livello di rete (Piano di Controllo)

- Introduzione
- □ Algoritmi di instradamento
 - O Stato del collegamento (Link State)
 - O Vettore distanza
 - Instradamento gerarchico
- □ Instradamento in Internet
 - O RIP
 - O OSPF
 - O BGP
- □ ICMP: Internet Control Message Protocol

Internet Control Message Protocol (ICMP)

	Viene usato da host e router per scambiarsi informazioni a	<u>Tipo</u>	Codice	<u>Descrizione</u>
	livello di rete.	0	0	Risposta eco (a ping)
	 report degli errori: host, 	3	0	rete destin. irraggiungibile
	rete, porta, protocollo	3	1	host destin. irraggiungibile
	irraggiungibili.	3	2	protocollo dest. irraggiungibile
	echo request/reply	3	3	porta destin. irraggiungibile
	(usando il programma ping).	3	6	rete destin. sconosciuta
	Livello di rete "sopra" IP:	3	7	host destin. sconosciuto
	 ICMP è considerato parte 	4	0	riduzione (controllo
	di IP.			di congestione)
П	Messaggi ICMP: hanno un	8	0	richiesta eco
	campo tipo e un campo codice,	9	0	annuncio del router
	e contengono l'intestazione e i	10	0	scoperta del router
	primi 8 byte del datagramma	11	0	TTL scaduto
	IP.	12	0	errata intestazione IP

Traceroute e ICMP

- Il programma invia una serie di datagrammi IP alla destinazione.
 - Il primo pari a TTL =1
 - Il secondo pari a TTL=2, ecc.
 - Numero di porta improbabile
- Quando l'n-esimo datagramma arriva all'n-esimo router:
 - Il router scarta il datagramma.
 - Invia all'origine un messaggio di allerta ICMP (tipo 11, codice 0).
 - Il messaggio include il nome del router e l'indirizzo IP.

- Quando il messaggio ICMP arriva, l'origine può calcolare RTT
- Traceroute lo fa per 3 volte

Criteri di arresto dell'invio

- Quando un segmento UDP arriva all'host di destinazione.
- ☐ L'host di destinazione restituisce un messaggio ICMP di porta non raggiungibile (tipo 3, codice 3).
- Quando l'origine riceve questo messaggio ICMP, si blocca.