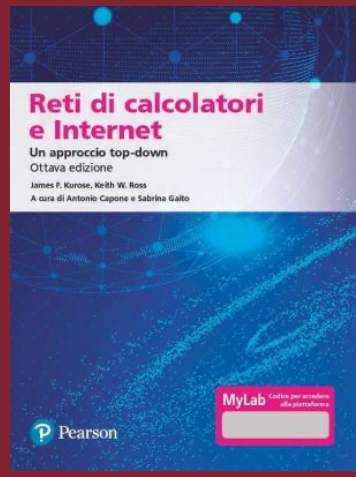
Reti di Elaboratori

Livello di Rete: Instradamento intra e inter ISP



SAPIENZA
UNIVERSITÀ DI ROMA

Alessandro Checco
alessandro.checco@uniroma1.it

Capitolo 5

Livello di rete – piano di controllo: sommario

- introduzione
- algoritmi di instradamento
  - link state
  - distance vector
- instradamento intra-ISP: RIP e OSPF
- instradamento tra ISP: BGP
- piano di controllo SDN
  - gestione della rete, configurazione
    - SNMP
    - NETCONF/YANG

Rendere l'instradamento scalabile

il nostro studio del routing finora - idealizzato

- tutti i router identici
- rete "piatta"

… non è realistico

problema di scala: miliardi di destinazioni:

- non è possibile memorizzare tutte le destinazioni nelle tabelle di instradamento!
- lo scambio di tabelle di instradamento intaserebbe i collegamenti!

autonomia amministrativa:

- Internet: una rete di reti
- ogni amministratore di rete potrebbe voler controllare il routing nella propria rete

Approccio Internet al routing scalabile

aggregare i router in regioni note come "sistemi autonomi"
(AS: autonomous systems, noti anche come domini)

intra-AS (intra-dominio):

instradamento *all'interno dello stesso AS ("rete")*

- tutti i router nell'AS devono eseguire lo stesso protocollo intra-dominio
- i router in diversi AS possono scegliere il proprio protocollo di instradamento
- gateway router: sul bordo del proprio AS, è collegato ai router di altri AS

inter-AS (inter-dominio):

instradamento *tra* reti diverse (diversi AS)

- i gateway eseguono l'instradamento tra diversi domini (e partecipano anche all'instradamento intra-dominio)

Sistemi autonomi interconnessi

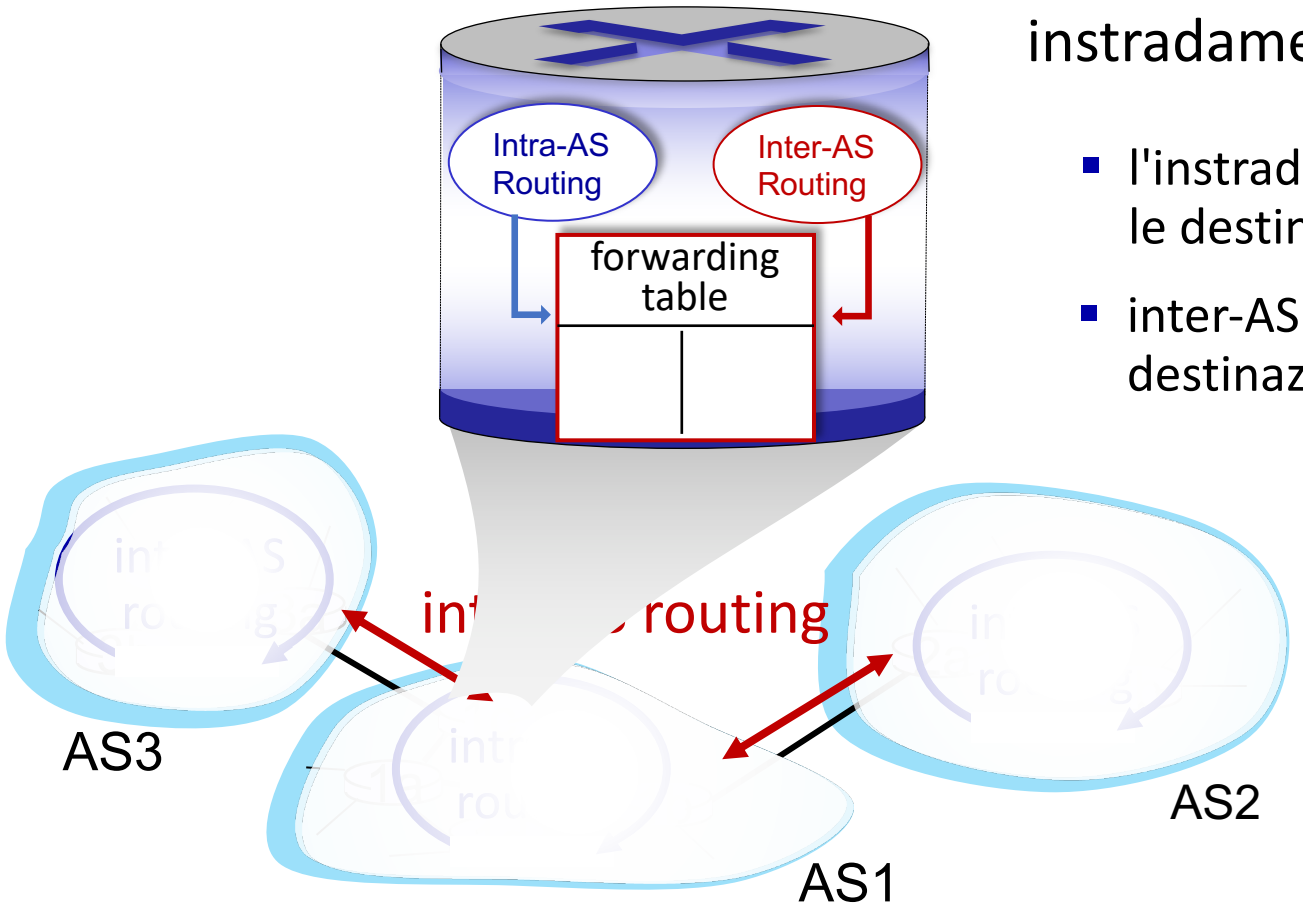


tabella di inoltro configurata da algoritmi di instradamento intra e inter-dominio

- l'instradamento intra-AS determina i campi per le destinazioni all'interno del dominio
- inter-AS & intra-AS determinano i campi per le destinazione esterne

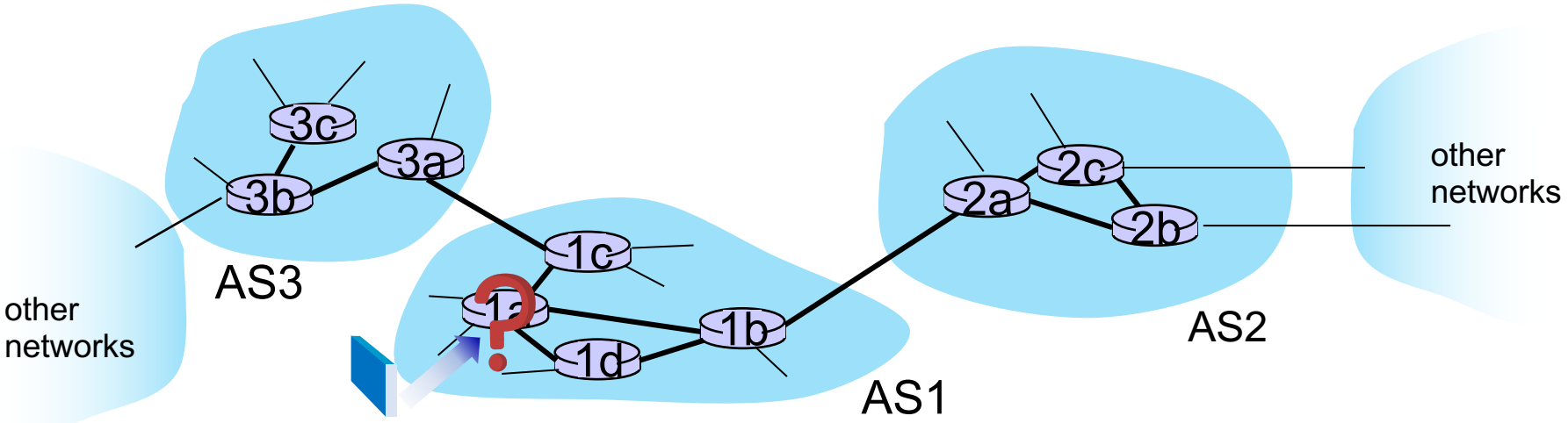
Il ruolo di intra-AS nell'instradamento inter-AS

- supponiamo che il router in AS1 riceva un datagramma destinato all'esterno di AS1:

? • il router dovrebbe inoltrare il pacchetto ad un gateway router in AS1, ma quale?

L'instradamento tra domini (inter-AS) di AS1 deve:

1. apprendere quali destinazioni siano raggiungibili tramite AS2, quali tramite AS3
2. propagare queste informazioni sulla raggiungibilità a tutti i router in AS1



Routing intra-dominio: instradamento all'interno di un AS

protocolli di instradamento intra-AS più comuni:

- RIP: protocollo di informazioni di instradamento [RFC 1723]
  - Usato dagli anni ‘80 a ‘00
  - basato su Distance Vector classico: DV scambiati ogni 30 secondi
  - non più ampiamente utilizzato
- EIGRP: Enhanced Interior Gateway Routing Protocol
  - Basato su DV
  - di proprietà di Cisco per decenni
  - Diventato aperto nel 2013 [RFC 7868])
- OSPF: Open Shortest Path First [RFC 2328]
  - instradamento link-state (stile Dijkstra)
  - Protocollo IS-IS (standard ISO, non RFC) essenzialmente uguale a OSPF (Intermediate System to Intermediate System)

RIP (Routing Information Protocol)

- È un protocollo a vettore distanza (distance vector)
- È tipicamente incluso in UNIX BSD dal 1982
- Metrica di costo: distanza misurata in hop (max = 15 hop, il valore 16 indica l'infinito)
  - Ogni link ha costo unitario

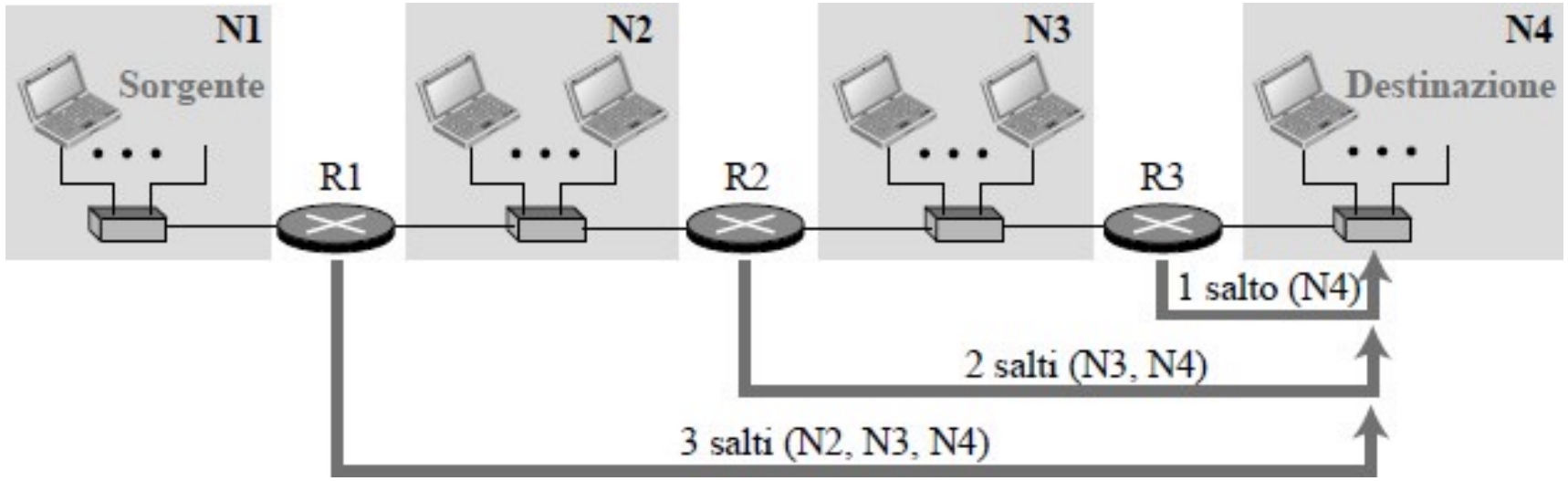


Tabelle di routing

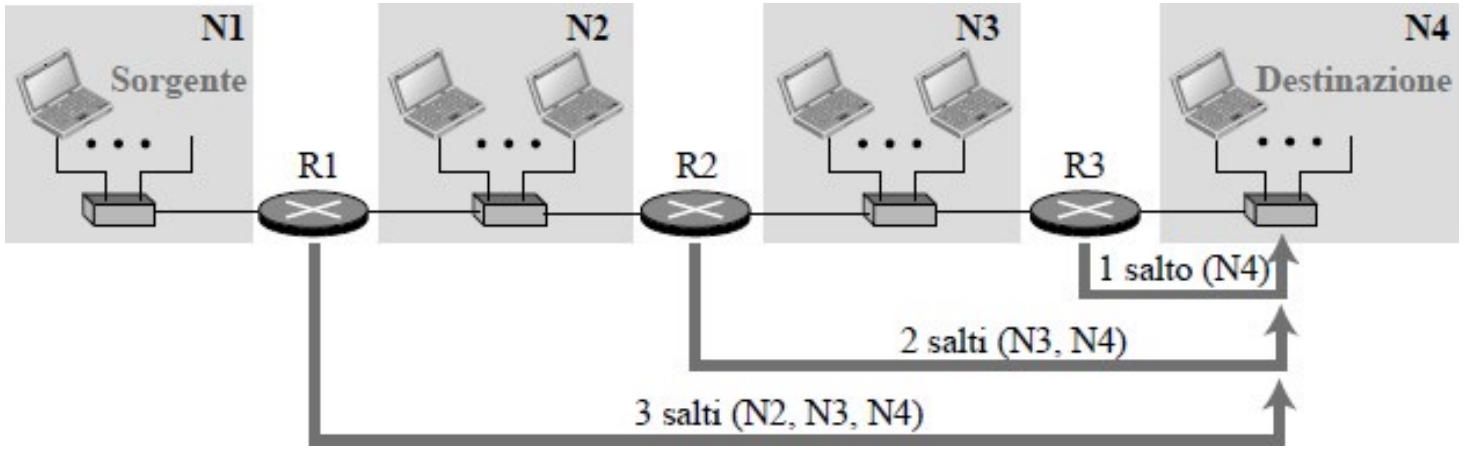


Tabella d'inoltro per R1

Rete di destinazione	Prossimo router	Costo (in hop)
N1	——	1
N2	——	1
N3	R2	2
N4	R2	3

Tabella d'inoltro per R2

Rete di destinazione	Prossimo router	Costo (in hop)
N1	R1	2
N2	——	1
N3	——	1
N4	R3	2

Tabella d'inoltro per R3

Rete di destinazione	Prossimo router	Costo (in hop)
N1	R2	3
N2	R2	2
N3	——	1
N4	——	1

L'informazione nella tabella di routing è sufficiente per raggiungere la destinazione

RIP protocol

- ***RIP Route Determination Algorithm***

  On a regular basis (= periodically), each router running RIP will send out its routing table entries (distance vector) to provide information to other routers about the networks and hosts it knows how to reach. Any routers on the same network as the one sending out this information will be able to update their own tables based on the information they receive. Any router that receives a message from another router on the same network saying it can reach network *X* at a cost of *N*, knows it can reach network *X* at a cost of *N+1* by sending to the router it received the message from.

N.B. Invece di inviare solo vettori di distanza, i router inviano più informazioni (vedi struttura messaggi più avanti)

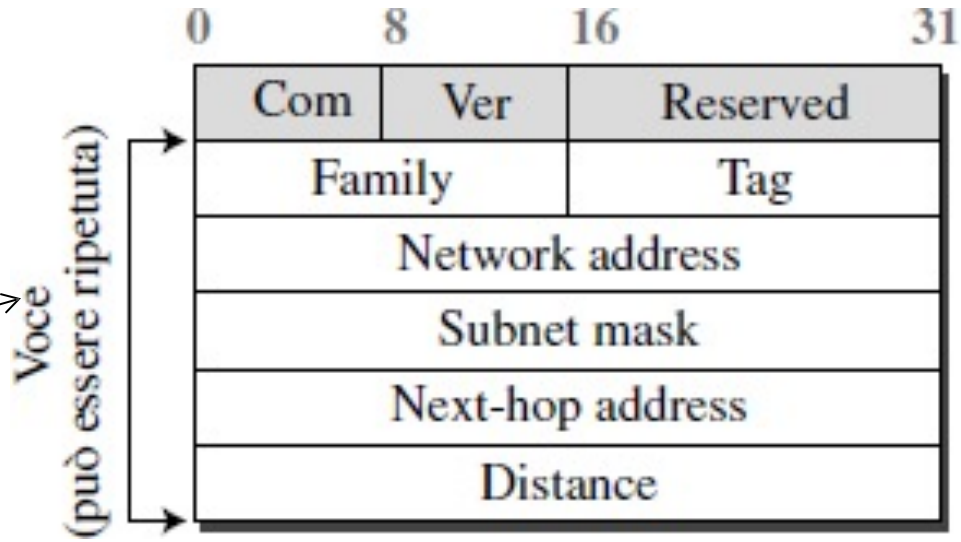
Messaggi RIP

- RIP si basa su una coppia di processi client-server e sul loro scambio di messaggi, sopra UDP porta 520
- RIP Request:
  - Quando un nuovo router viene inserito nella rete invia una RIP Request per ricevere immediatamente informazioni di routing dai vicini
  - Fini diagnostici (richiedere una voce specifica)
- RIP Response (o advertisement):
  - In risposta a una Request (solicited response)
  - Peridicamente ogni 30 sec (unsolicited response)

- 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

Struttura messaggi RIP

Corrisponde a una entry della tabella di routing



Campi

Com: comando, richiesta(1), risposta (2)
Ver: versione, la versione corrente è 2
Family: famiglia del protocollo, per il TCP/IP il valore è 2
Tag: informazioni sul sistema autonomo
Network address: indirizzo di destinazione
Subnet mask: maschera di sottorete (lunghezza del prefisso)
Next-hop address: indirizzo del prossimo hop
Distance: numero di hop fino alla destinazione

Timer RIP

- *Timer periodico*
  - Controlla invio messaggi di aggiornamento (randomizzato 25-35 secondi per evitare effetti di sincronizzazione e congestione)
- *Timer di scadenza*
  - Regola la validità dei percorsi (180 secondi, randomizz 150-210)
  - Se entro lo scadere del timer non si riceve aggiornamento, il percorso viene considerato scaduto e il suo costo impostato a 16
- *Timer per garbage collection*
  - Elimina percorsi dalla tabella (120 secondi)
  - Dopo il timer di scadenza il router continua ad annunciare il percorso con costo pari a 16, e allo scadere del timer garbage collection rimuove il percorso e non lo annuncia più

RIP: guasto sul collegamento e recupero

Se un router non riceve notizie dal suo vicino per ~180 sec --> il nodo adiacente/il collegamento viene considerato spento o guasto.

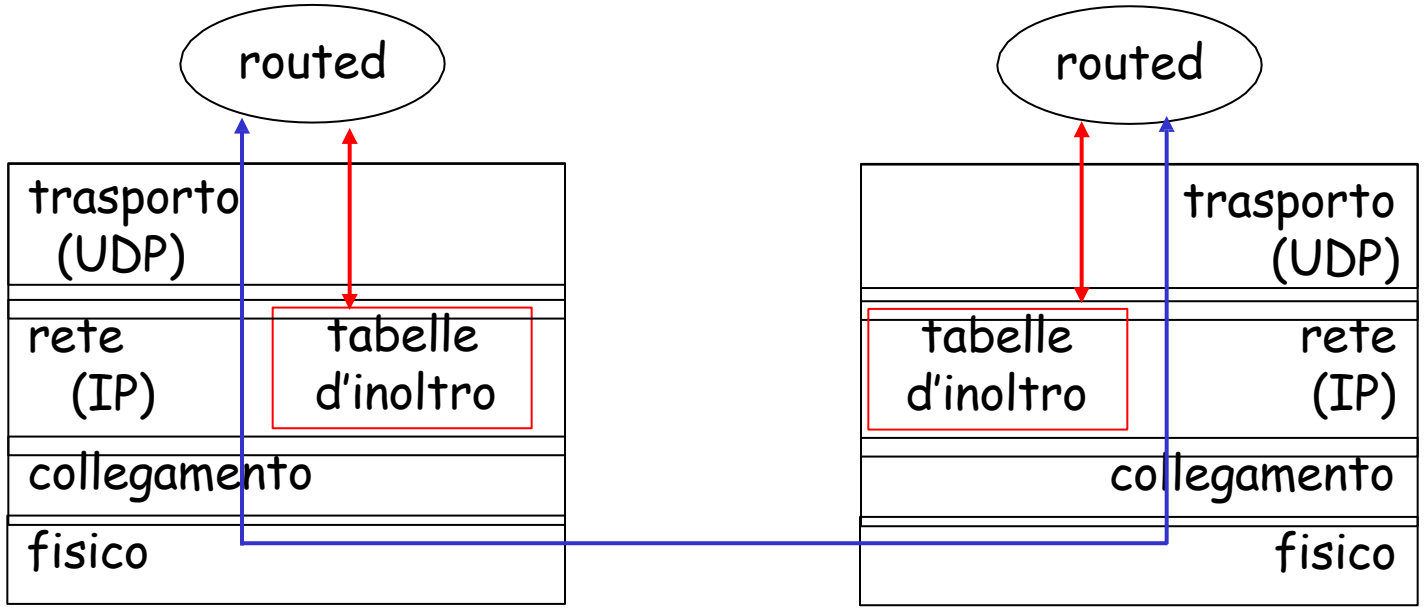
- RIP modifica la tabella d'instradamento locale (16)
- Propaga l'informazione mandando annunci ai router vicini
- I vicini inviano nuovi messaggi (se la loro tabella d'instradamento è cambiata)
- L'informazione che il collegamento è fallito si propaga rapidamente su tutta la rete
- L'utilizzo dell'inversione avvelenata evita i loop (distanza infinita = 16 hop per percorsi che passano dal vicino a cui sto inviando l'aggiornamento)

Caratteristiche di RIP

- *Split horizon with poisoned reverse* (inversione avvelenata)
  - Split horizon: non invia rotte apprese da A indietro ad A stesso. Serve per evitare che un router invii rotte non valide al router da cui ha imparato la rotta (evitare cicli).
  - Poisoned reverse: Si mette a infinito (16) il costo della rotta che passa attraverso il vicino a cui si manda advertisement. Quando un link cade (16) si può inviare questa informazione a tutti (compreso il router da cui si è appresa questa informazione)
- *Triggered updates*
  - Riduce il problema della convergenza lenta
  - Quando cambia una rotta si inviano immediatamente informazioni ai vicini senza attendere il timeout
- *Hold-down*
  - Fornisce robustezza
  - Quando si riceve una informazione di una rotta non più valida, si avvia un timer e tutti gli advertisement riguardanti quella rotta che arrivano entro il timeout vengono tralasciati/ignorati. Ad esempio se il router B risulta non valido, allora anche le informazioni riguardanti B ricevute da altri router vengono considerate non valide

Implementazione di RIP

- Implementato come applicazione sopra UDP porta 520
- Un processo chiamato routed (route daemon) 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.



Instradamento OSPF (Open Shortest Path First)

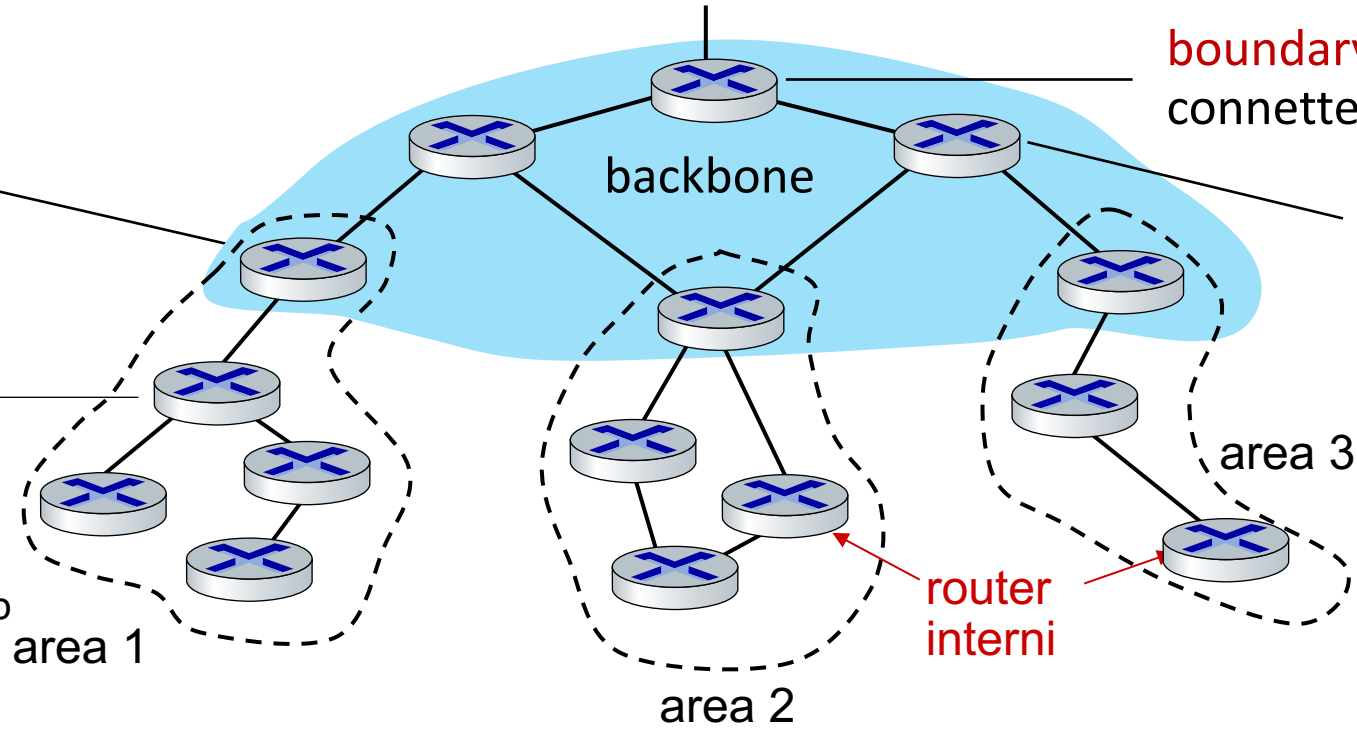
- "aperto": pubblicamente disponibile
- link-state classico
  - ogni router esegue il flooding degli annunci link-state OSPF (direttamente come payload di IP [prot.n. 89] anziché utilizzare TCP/UDP) a tutti gli altri router del AS
  - più metriche dei costi di collegamento possibili: larghezza di banda, ritardo, etc.
  - ogni router conosce la topologia completa, utilizza l'algoritmo di Dijkstra per calcolare la tabella di inoltro
- *sicurezza:* tutti i messaggi OSPF autenticati (per prevenire intrusioni malevole)

OSPF gerarchico

- gerarchia a due livelli: local area, backbone.
  - annunci link-state propagati (flooded) solo nell'area locale o nella dorsale
    - permette di ridurre la quantità di messaggi in base alla gerarchia
  - ogni nodo conosce la topologia dettagliata della propria area (o backbone), mentre conosce solo la direzione per raggiungere le altre destinazioni

area border router
(router di bordo d'area):
"riepiloga" le distanze verso altre destinazioni nella propria area, e pubblicizza queste informazioni nel backbone

router locali:
- flooding del LS solo nell'area
- calcolano l'instradamento all'interno dell'area
- inoltrano i pacchetti verso l'esterno tramite il router di bordo d'area



boundary router: si connette ad altri AS

backbone router: esegue OSPF limitato al backbone (flooding all'interno della backbone)

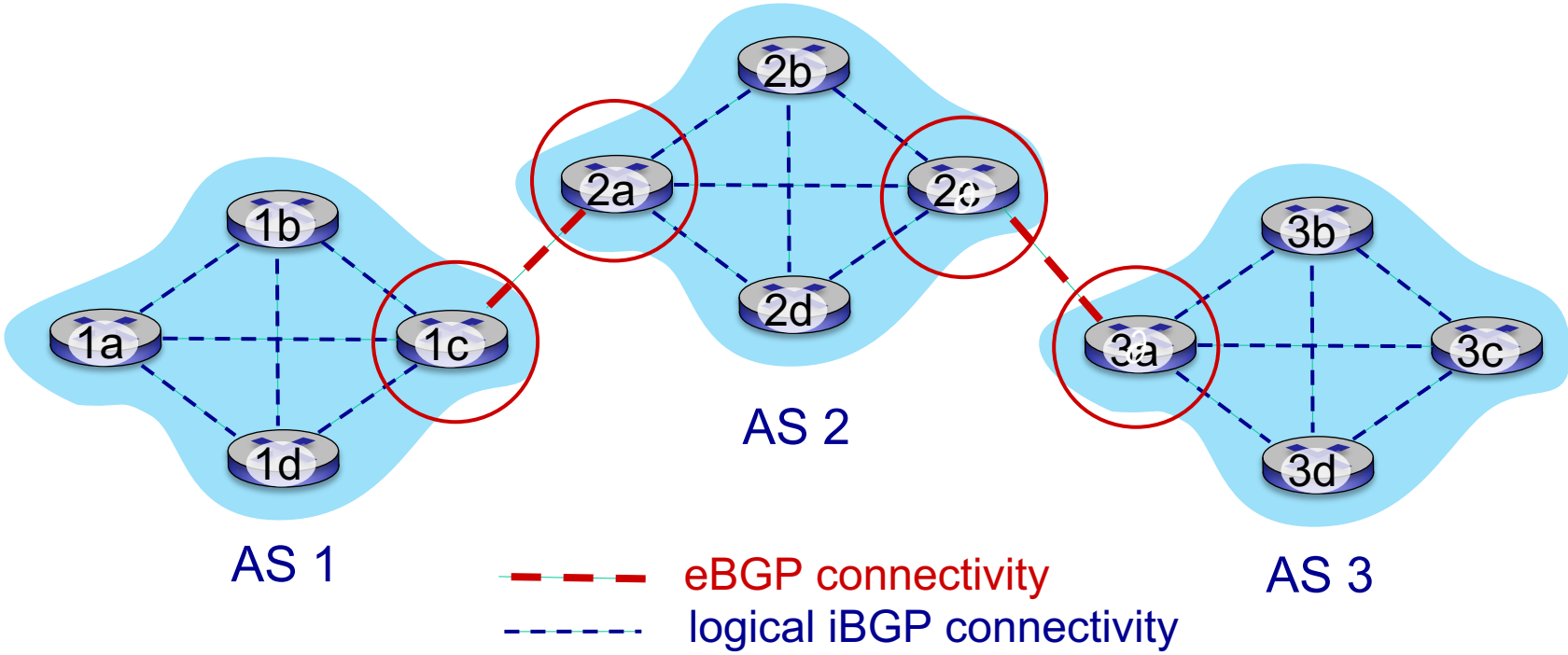
Livello di rete – piano di controllo: sommario

- introduzione
- algoritmi di instradamento
  - link state
  - distance vector
- instradamento intra-ISP: RIP e OSPF
- instradamento tra ISP: BGP
- piano di controllo SDN
  - gestione della rete, configurazione
    - SNMP
    - NETCONF/YANG

Instradamento Internet tra AS: BGP

- BGP (Border Gateway Protocol): *il* protocollo di instradamento interdominio più usato
  - "colla che tiene insieme Internet"
- consente alla sottorete di pubblicizzare la propria esistenza, e le destinazioni che può raggiungere, al resto di Internet: *" Sono qui, ecco chi posso raggiungere e come (tramite quale percorso)"*
- BGP fornisce ad ogni AS un mezzo per:
  - eBGP: ottenere informazioni sulla raggiungibilità di una sottorete tramite gli AS vicini
  - iBGP: propaga le informazioni sulla raggiungibilità a tutti i router interni all'AS
  - determinare percorsi "buoni" verso altre reti in base alle informazioni di raggiungibilità e alla *politica* interna del AS (ad esempio evitare un certo paese)
  - **pubblicizzare** (alle reti confinanti) le proprie informazioni di raggiungibilità (o anche non farlo/pubblicizzare solo una parte)

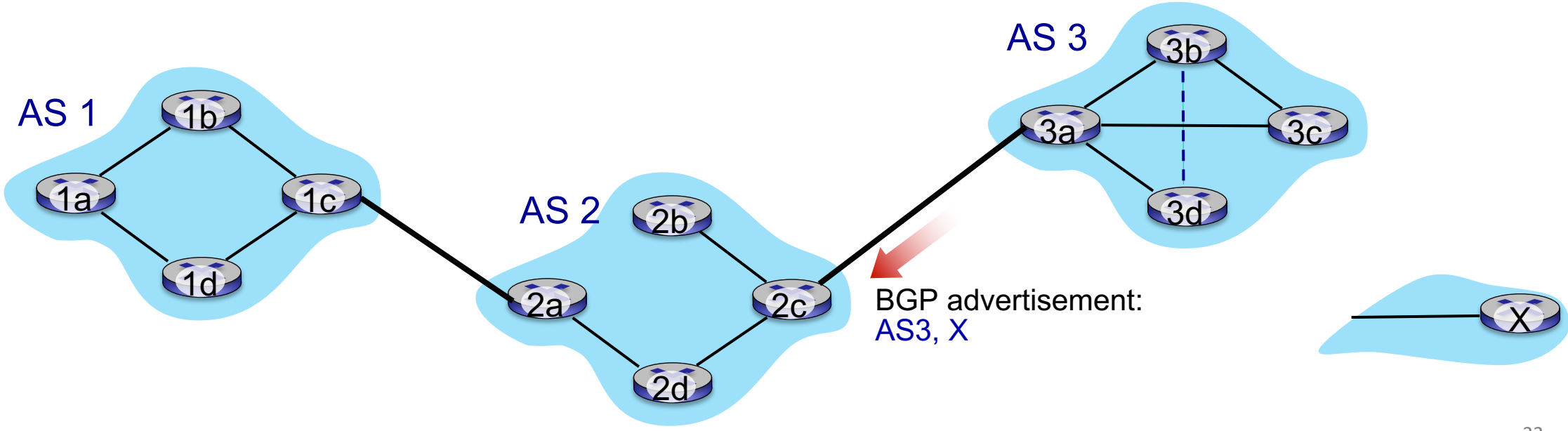
connessioni eBGP e iBGP



gateway routers run both eBGP and iBGP protocols

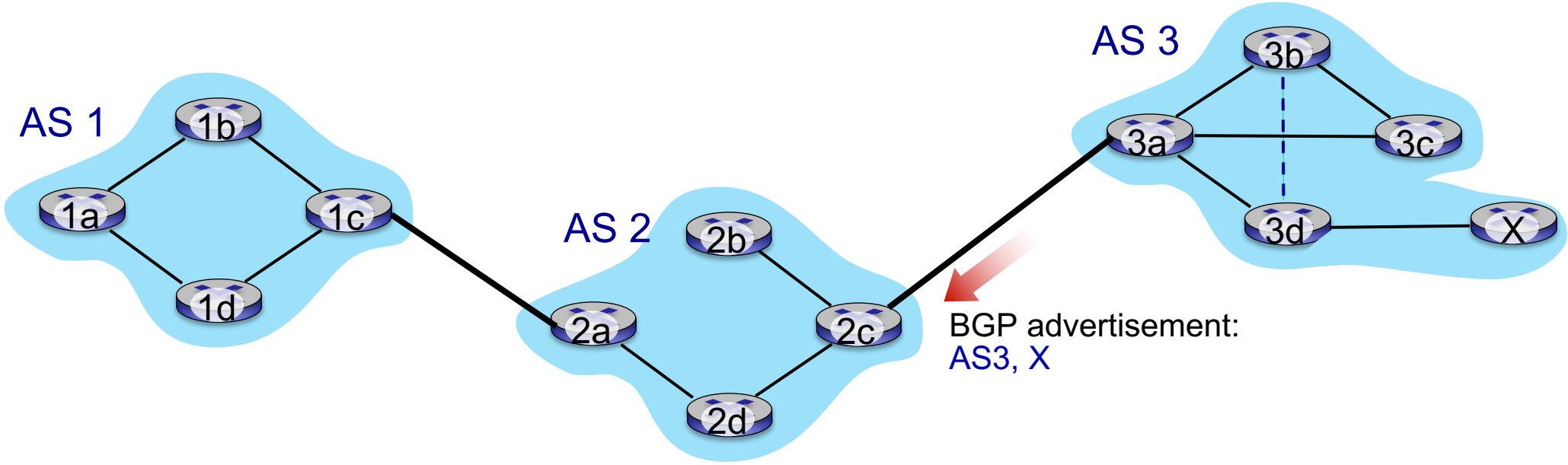
Nozioni di base su BGP

- Sessione BGP: due router BGP ("peer") si scambiano messaggi BGP su una connessione TCP semipermanente:
  - vengono pubblicizzati *percorsi* a diversi prefissi di rete di destinazione (ad es. a una rete /16. BGP è un protocollo "path vector")
- quando il gateway AS3 3a annuncia il percorso AS3,X al gateway AS2 2c:
  - AS3 *promette* ad AS2 che inoltrerà i datagrammi verso X



Nozioni di base su BGP

- Sessione BGP: due router BGP ("peer") si scambiano messaggi BGP su una connessione TCP semipermanente:
  - vengono pubblicizzati *percorsi* a diversi prefissi di rete di destinazione (ad es. a una rete /16. BGP è un protocollo "path vector")
- quando il gateway AS3 3a annuncia il percorso AS3,X al gateway AS2 2c:
  - AS3 *promette* ad AS2 che inoltrerà i datagrammi verso X



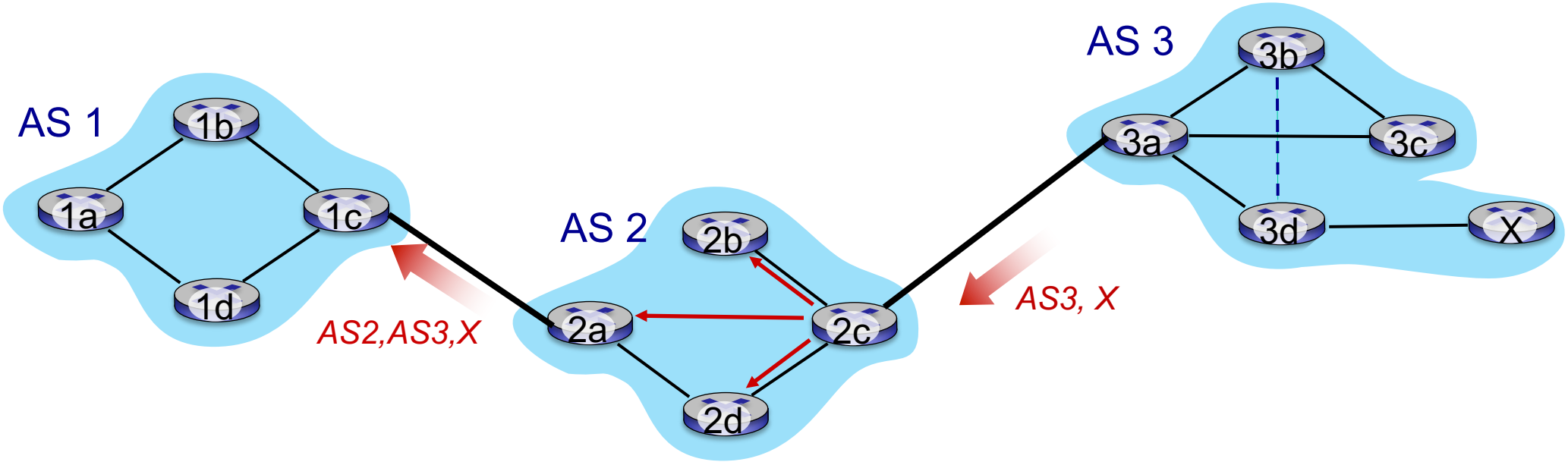
Messaggi del protocollo BGP

- Messaggi BGP scambiati tra peers sopra connessioni TCP
- Struttura dei messaggi BGP [RFC 4371]:
  - OPEN: apre la connessione TCP al peer BGP remoto e autentica il peer BGP che apre la connessione
  - UPDATE: pubblicizza nuovo percorso (o ritira vecchio)
  - KEEPALIVE: mantiene viva la connessione in assenza di UPDATEs; serve anche per ACK la richiesta OPEN
  - NOTIFICATION: segnala gli errori nei msg precedenti; utilizzato anche per chiudere la connessione

Attributi di percorso e rotte BGP

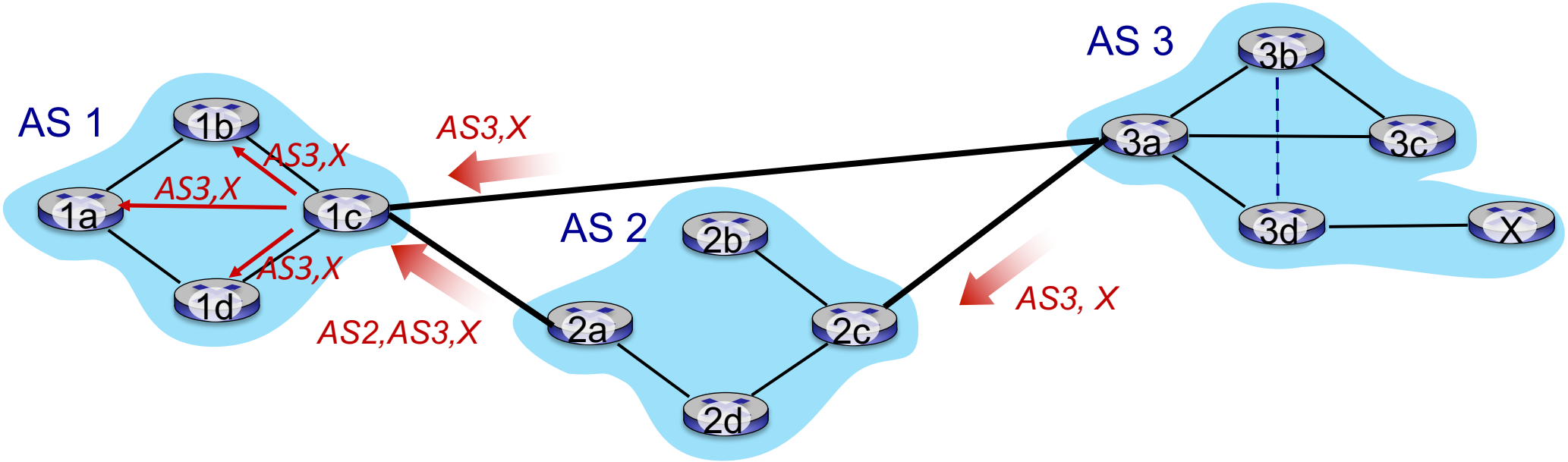
- Percorso pubblicizzato BGP: prefisso + attributi
  - path prefix: destinazione che viene pubblicizzata (ciderized x.x/24)
  - due attributi importanti:
    - AS-PATH: elenco di AS attraverso i quali è passato l'annuncio del prefisso
    - NEXT-HOP: indica un router interno all'AS che è il prossimo hop per raggiungere la destinazione (egress router, 3d nell'esempio precedente)
- policy-based routing:
  - il router che riceve l'advertising di un percorso verso X usa una **policy di AS** per accettare/rifiutare il percorso (ad es. non passare mai tramite AS X o paese Z, per questo serve AS-PATH)
  - il router usa tale policy anche per decidere se pubblicizzare un percorso agli AS vicini (potrebbe causare traffico di transito e potrei non volerlo)

Esempio BGP path advertisement



- Il router AS2 2c riceve l'annuncio del percorso AS3, X (tramite eBGP) dal router AS3 3a
- basandosi sulla politica AS2, il router AS2 2c accetta il percorso AS3, X, e lo propaga (tramite iBGP) a tutti i router AS2
- in base alla politica AS2, il router AS2 2a annuncia (tramite eBGP) il percorso AS2, AS3, X al router AS 1c

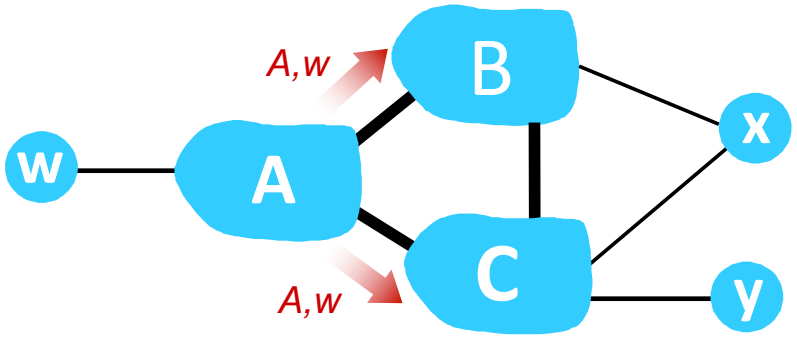
BGP path advertisement (multiple paths)

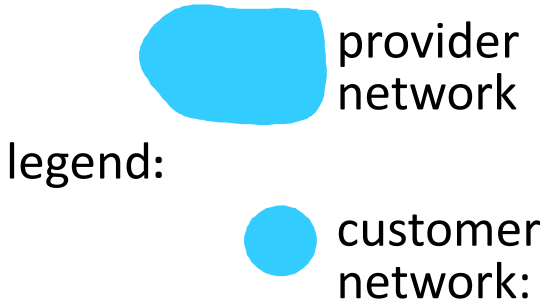


il router gateway può apprendere più percorsi verso la destinazione:

- Il router gateway 1c di AS1 apprende il percorso AS2, AS3, X da 2a
- Il router gateway 1c di AS1 apprende il percorso AS3, X da 3a
- basandosi sulla policy di AS, Il router gateway 1c di AS1 sceglie il percorso AS3, X e annuncia il percorso all'interno di AS1 tramite iBGP

BGP: creare politiche tramite l'advertising

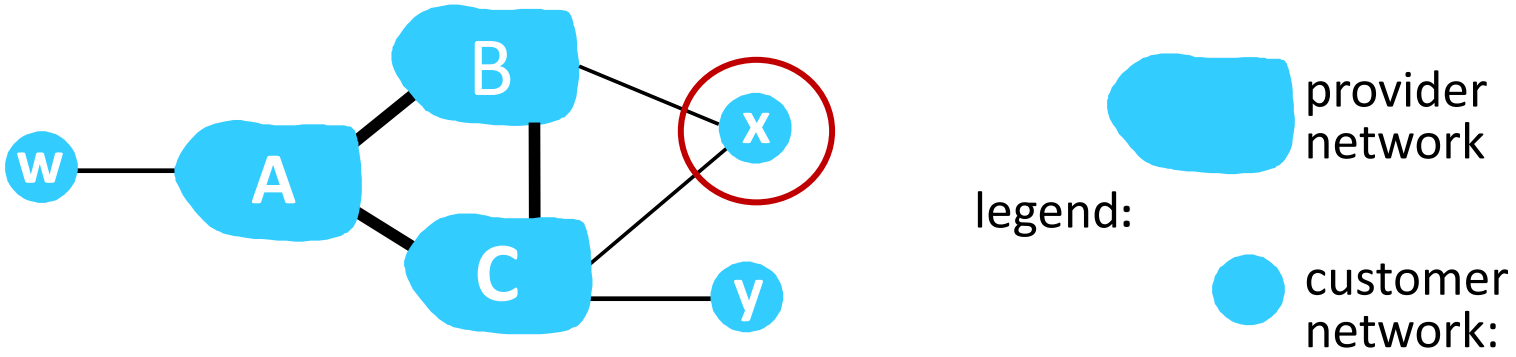




L'ISP vuole solo instradare il traffico da/verso le reti dei suoi clienti (non vuole trasportare il traffico di transito tra altri ISP - una tipica politica del "mondo reale")

- A annuncia il percorso A,w verso B e verso C
- B **sceglie di non pubblicizzare** B,A,w a C!
  - B non ottiene alcun "ricavo" per l'instradamento C,B,A,w, poiché C, A, w non sono clienti di B
  - C non apprende il percorso C,B,A,w
- C instraderà sul percorso C,A,w (non usando B) per arrivare a w

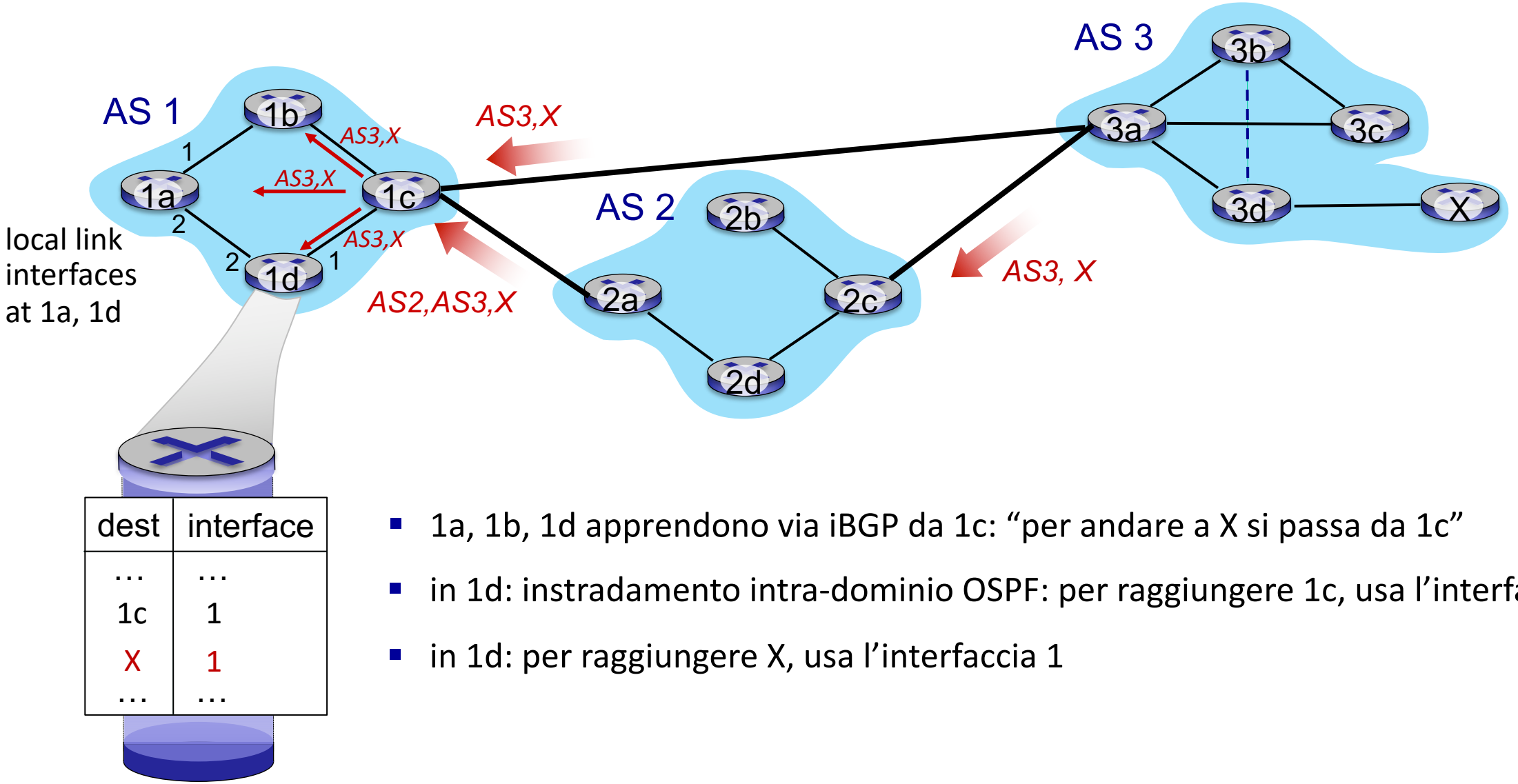
BGP: creare politiche tramite l'advertising



L'ISP vuole solo instradare il traffico da/verso le reti dei suoi clienti (non vuole trasportare il traffico di transito tra altri ISP - una tipica politica del "mondo reale")

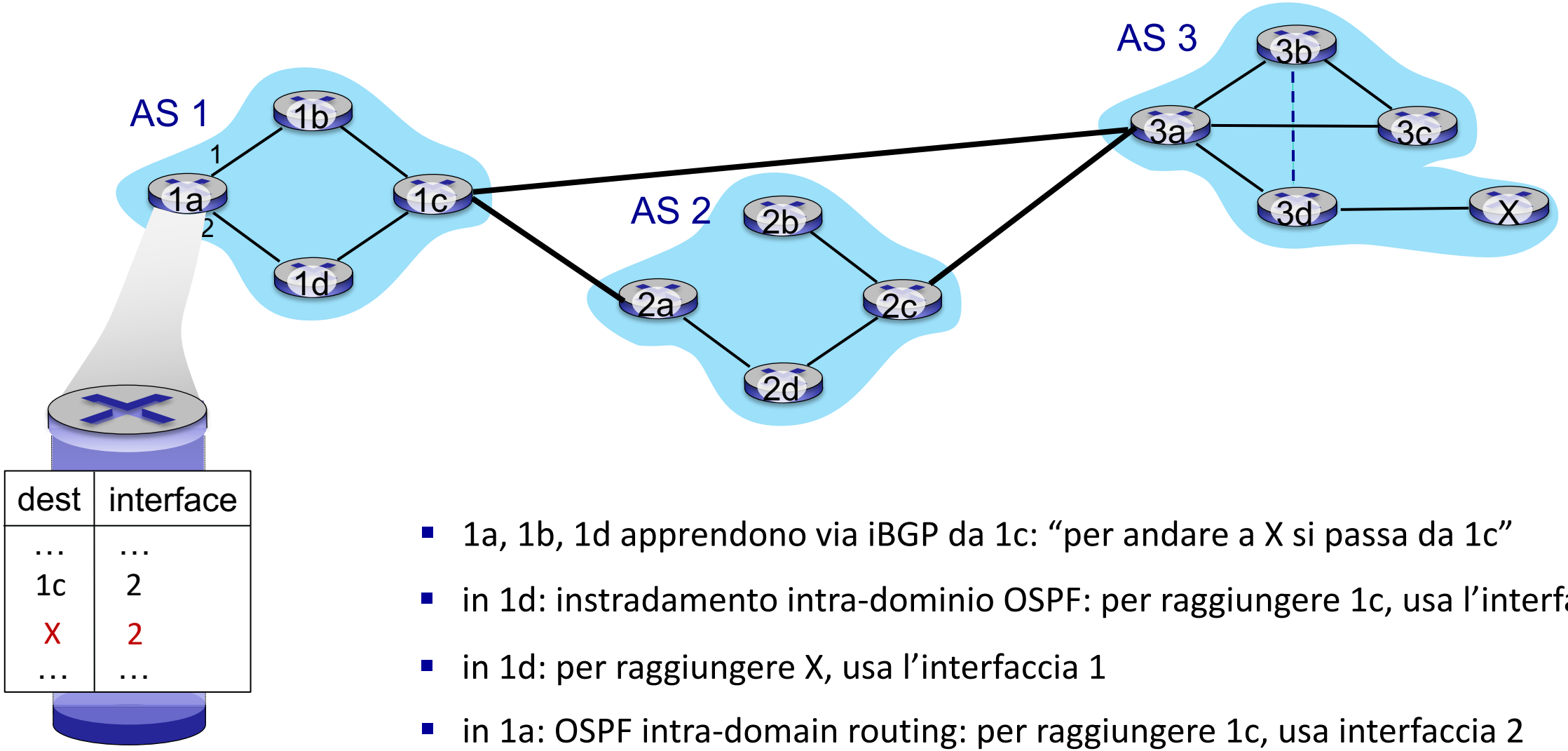
- A, B, C sono **reti provider**
- x,w,y sono customer (delle reti provider)
- x è dual-homed: collegato a due reti
- *policy da applicare:* x non vuole instradare da B a C tramite x (piccola rete! Rischio traffico di transito B-C)
  - .. quindi x non pubblicizzerà a B un percorso per C e viceversa

BGP path advertisement (dettaglio)



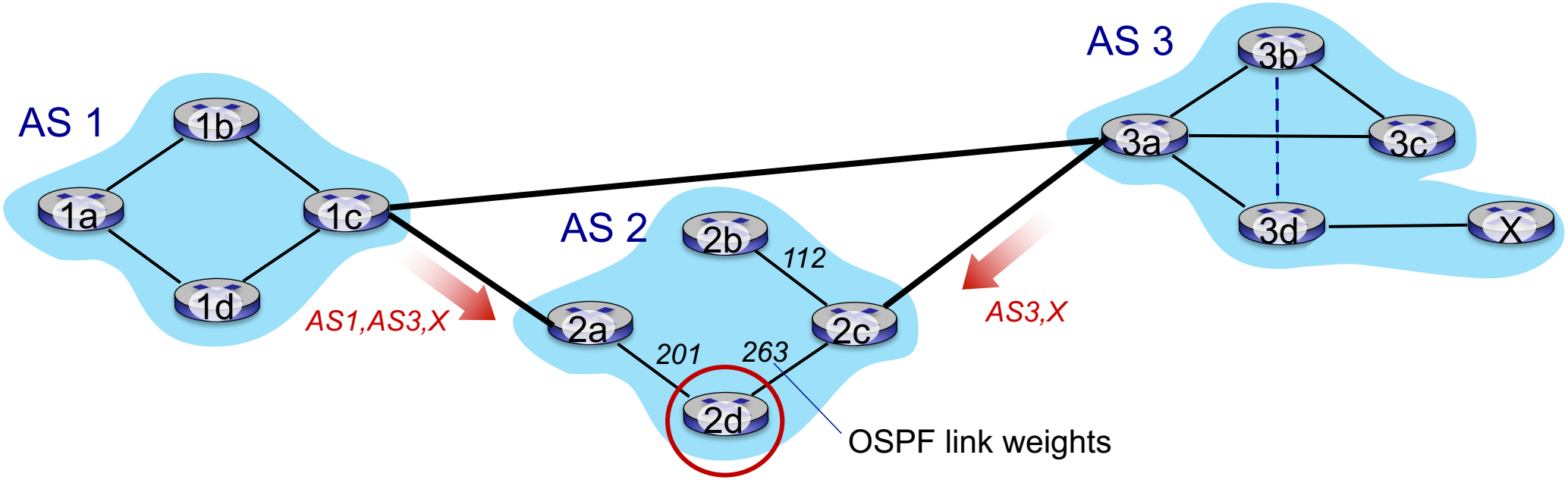
- 1a, 1b, 1d apprendono via iBGP da 1c: “per andare a X si passa da 1c”
- in 1d: instradamento intra-dominio OSPF: per raggiungere 1c, usa l’interfaccia 1
- in 1d: per raggiungere X, usa l’interfaccia 1

BGP path advertisement (dettaglio)



- 1a, 1b, 1d apprendono via iBGP da 1c: “per andare a X si passa da 1c”
- in 1d: instradamento intra-dominio OSPF: per raggiungere 1c, usa l’interfaccia 1
- in 1d: per raggiungere X, usa l’interfaccia 1
- in 1a: OSPF intra-domain routing: per raggiungere 1c, usa interfaccia 2
- in 1a: per raggiungere X, usa l’interfaccia 2

Hot potato routing (patata bollente)



- 2d apprende (tramite iBGP) che può inoltrare a X tramite 2a o 2c
- hot potato routing: scegli il gateway locale che ha il minor costo intra-dominio (ad esempio, 2d sceglie 2a, anche se poi ci sono più hop per raggiungere X): non ci preoccupiamo del costo inter-dominio!
  - vantaggi: minor traffico interno
  - svantaggi: minor controllo sul traffico esterno (cold potato se è importante)

Selezione del percorso BGP

- il router potrebbere apprendere più di un percorso verso la destinazione AS, seleziona il percorso in base a (una combinazione di):
  1. attributo del valore di preferenza locale: decisione di policy
  2. AS-PATH più breve
  3. router NEXT-HOP più vicino: hot potato routing
  4. criteri aggiuntivi

Sommario: Perché un diverso routing Intra-, Inter-AS?

policy:

- inter-AS: l'amministratore vuole controllare come viene instradato il suo traffico, chi passa attraverso la sua rete
- intra-AS: singolo amministratore, quindi policy meno problematica, si può usare un protocollo più semplice e "automatico"

scala:

- l'instradamento gerarchico ricuce le dimensioni della tabella, riduce il traffico necessario per l’aggiornamento delle tabelle

prestazioni:

- intra-AS: può concentrarsi sulle prestazioni
- inter-AS: la politica prevale sulla performance