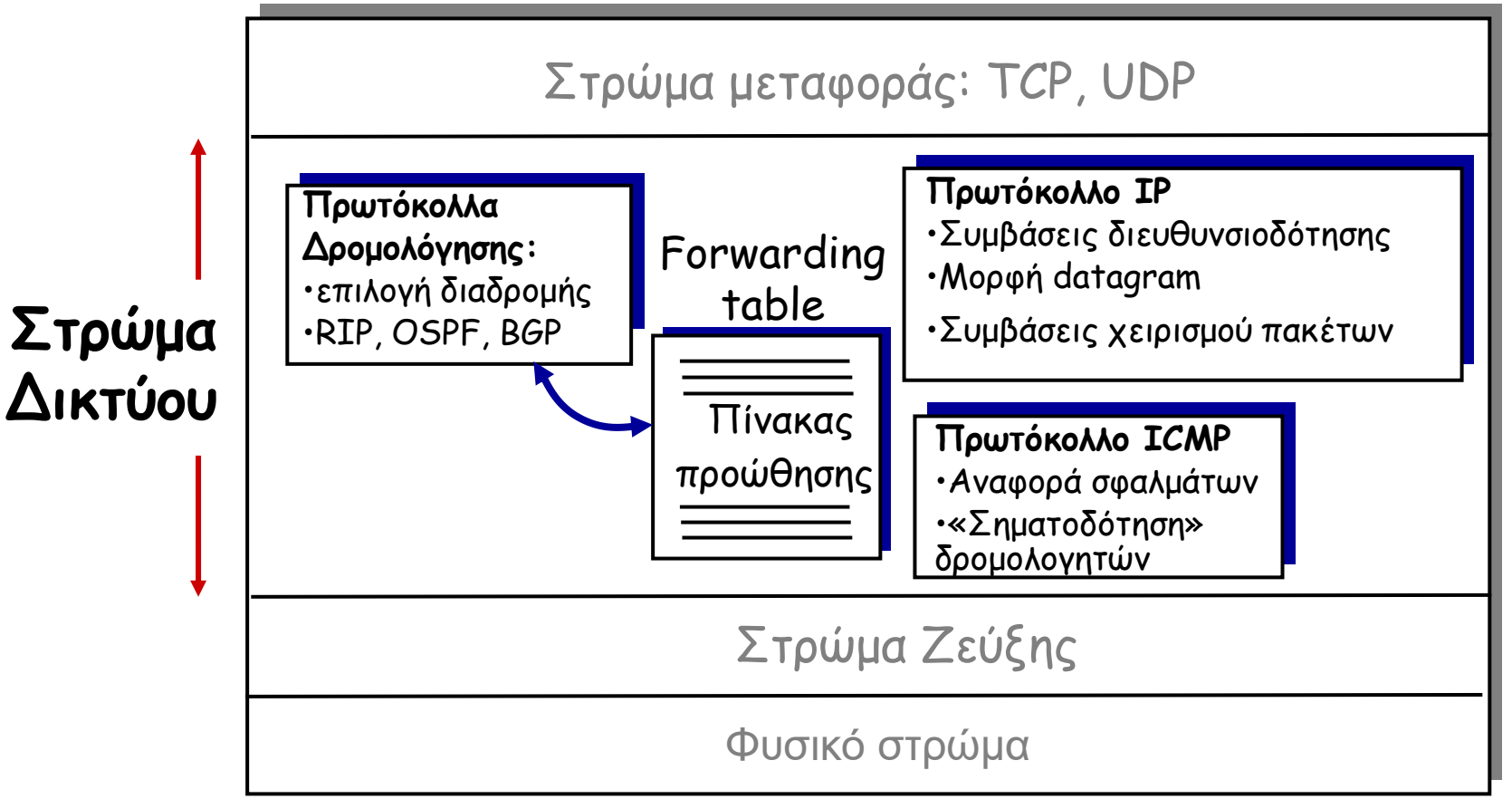
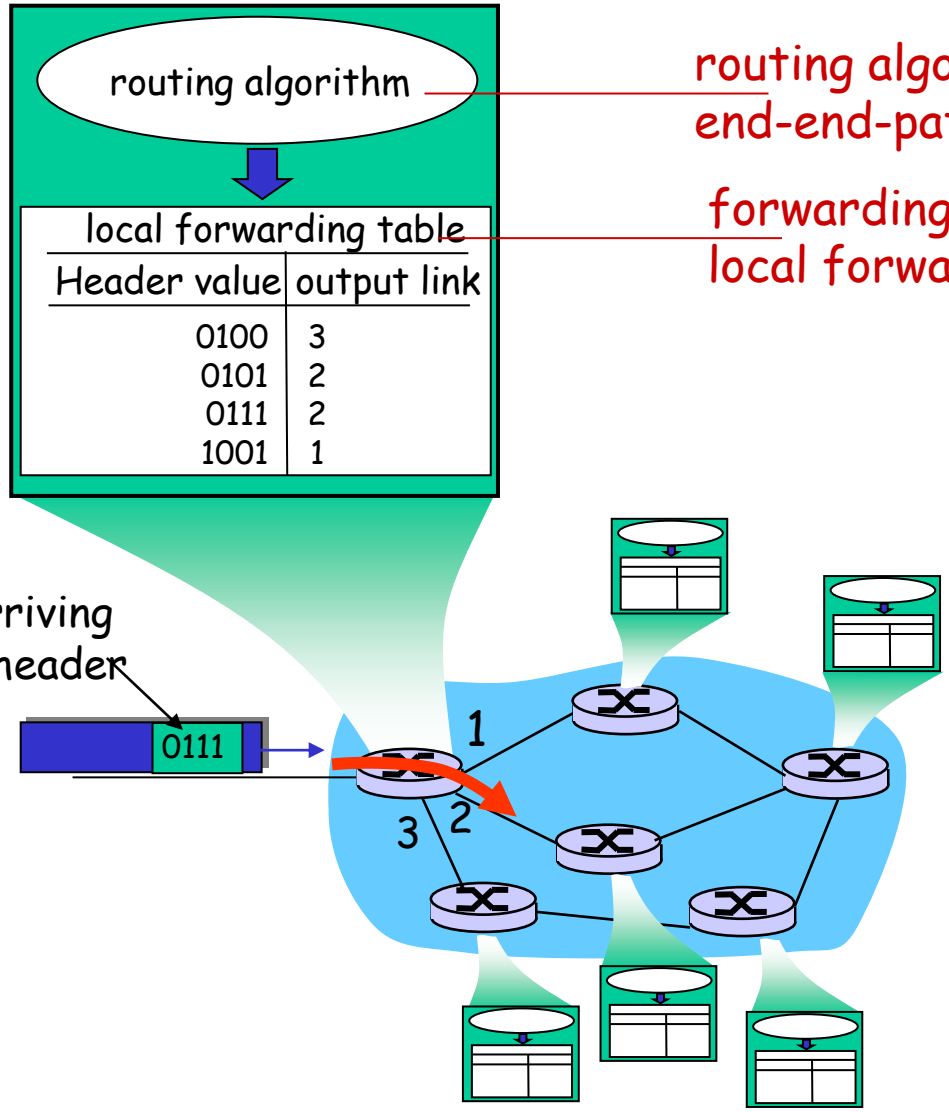
Το στρώμα δικτύου του Διαδικτύου

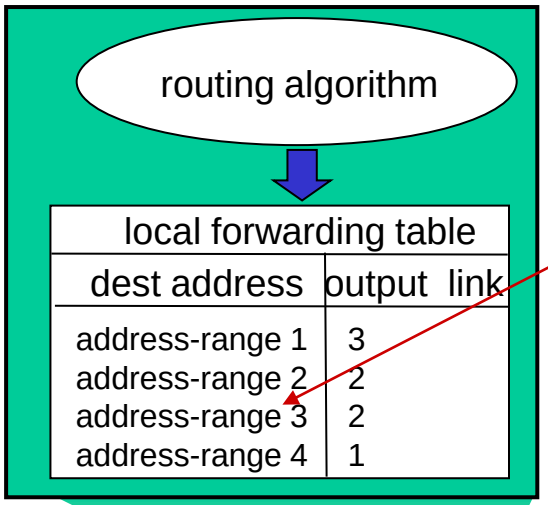


Interplay between routing and forwarding

Να θυμάσαι!!!



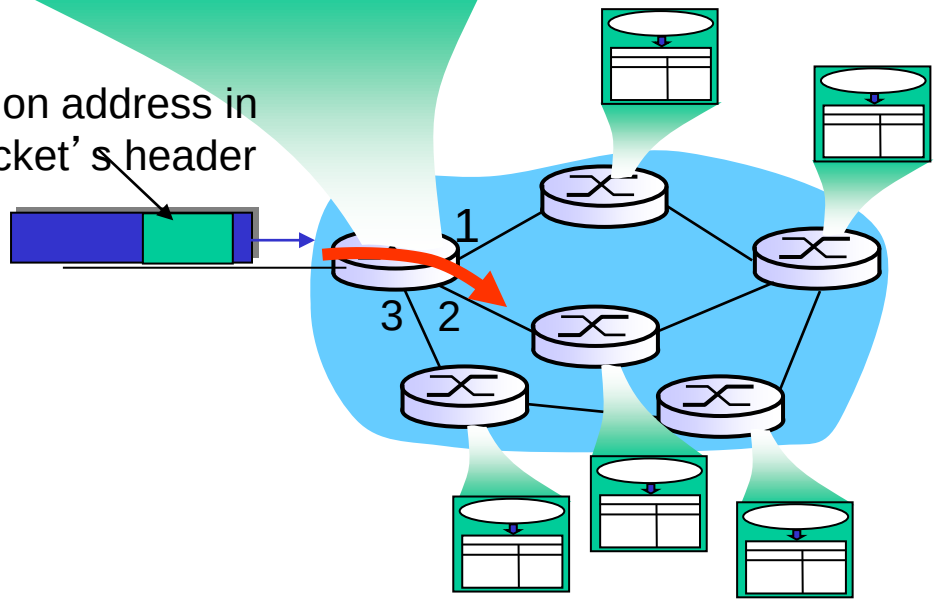
Datagram forwarding table



4 billion IP addresses, so rather than list individual destination address list *range* of addresses (aggregate table entries)

Prefix - CIDR
Longest Prefix Matching

IP destination address in arriving packet's header



ICMP: Internet Control Message Protocol

Upper layer Protocol Number = 1

- used by hosts, routers, gateways to communication network-level information
  - error reporting: unreachable host, network, port, protocol
  - echo request/reply (used by ping)
- network-layer "above" IP:
  - ICMP msgs carried in IP datagrams
- ICMP message: type, code plus first 8 bytes of IP datagram causing error

Type	Code	description
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	**0**	**TTL expired**
12	0	bad IP header

Traceroute: Μια εφαρμογή με βάση TTL - ICMP

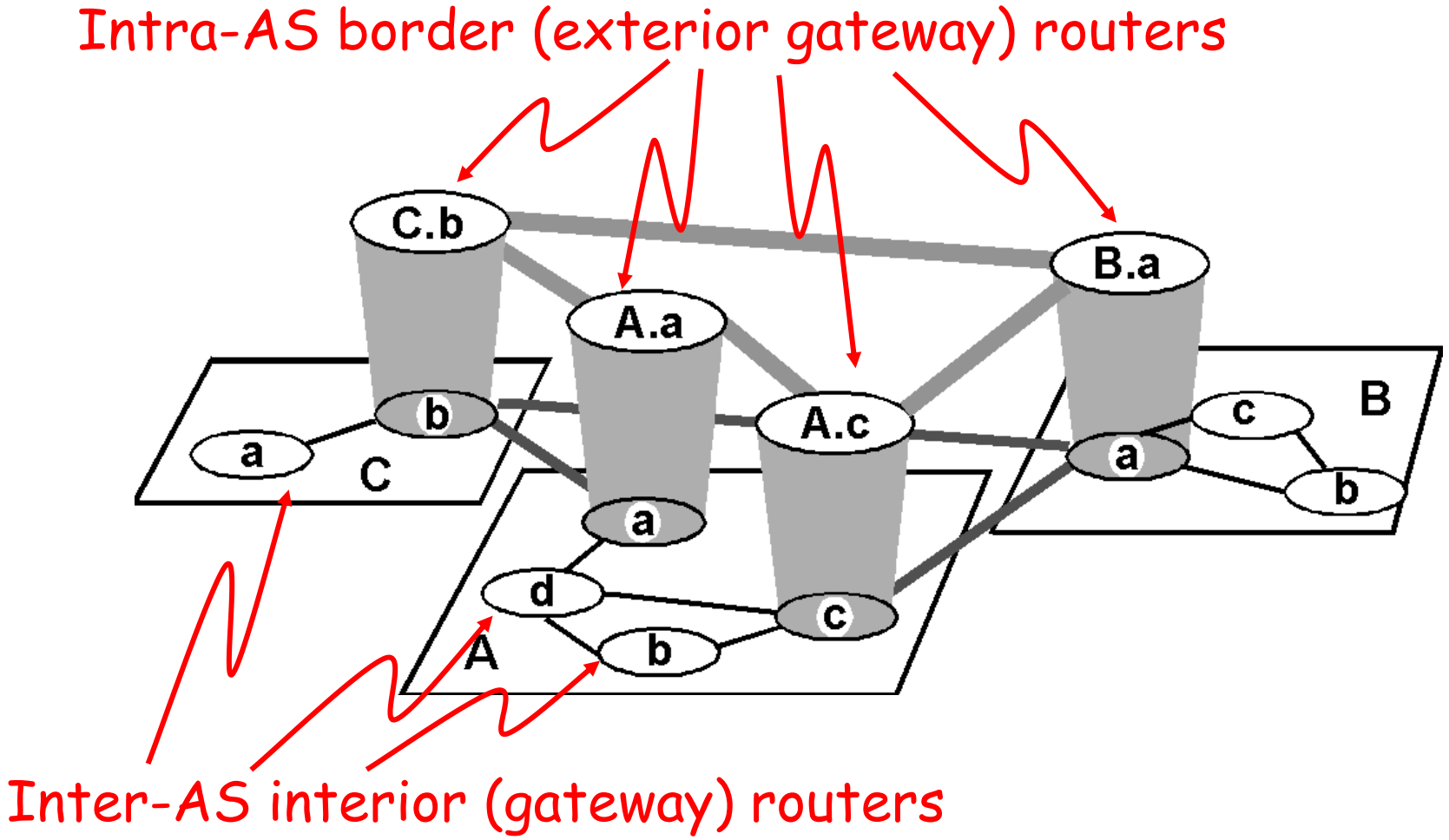
- Το Traceroute μας βοηθάει σε διαδικασίες test ή debugging (διαδικασία εντοπισμού και διόρθωσης λαθών). Λειτουργεί ως εξής:
- Στέλνει ένα IP πακέτο με TTL=1 σε ένα κόμβο παραλήπτη.
- Ο 1ος Router στην διαδρομή μεταξύ αποστολέα-παραλήπτη θα μειώσει το TTL στην τιμή 0, και επομένως θα στείλει ένα μήνυμα ICMP «TTL expired» στον αποστολέα.
- Ακολούθως, ο αποστολέας θα στείλει πάλι το ίδιο πακέτο στον παραλήπτη με TTL=2, οπότε ο 2ος Router στην διαδρομή μεταξύ αποστολέα-παραλήπτη θα είναι αυτός που θα μηδενίσει το TTL και θα στείλει το μήνυμα ICMP «TTL expired» στον αποστολέα.
- Η διαδικασία αυτή επαναλαμβάνεται από τον αποστολέα μέχρι το πακέτο να φθάσει στον παραλήπτη (αυξάνοντας κάθε φορά το TTL κατά 1).
- Έτσι, ο αποστολέας μπορεί να μάθει από ποιους Routers πέρασε το datagram.
- Για καλύτερη εκτίμηση των συνολικών χρόνων RTT (Round Trip Time), κάθε φορά το datagram στέλνεται εις τριπλούν.

Routing in the Internet

- The Global Internet consists of Autonomous Systems (AS) interconnected with each other:
  - **Stub AS**: small corporation
  - **Multihomed AS**: large corporation (no transit)
  - **Transit AS**: provider

- Two-level routing:
  - **Intra-AS**: administrator is responsible for choice
  - **Inter-AS**: unique standard

Internet AS Hierarchy



Intra-AS Routing

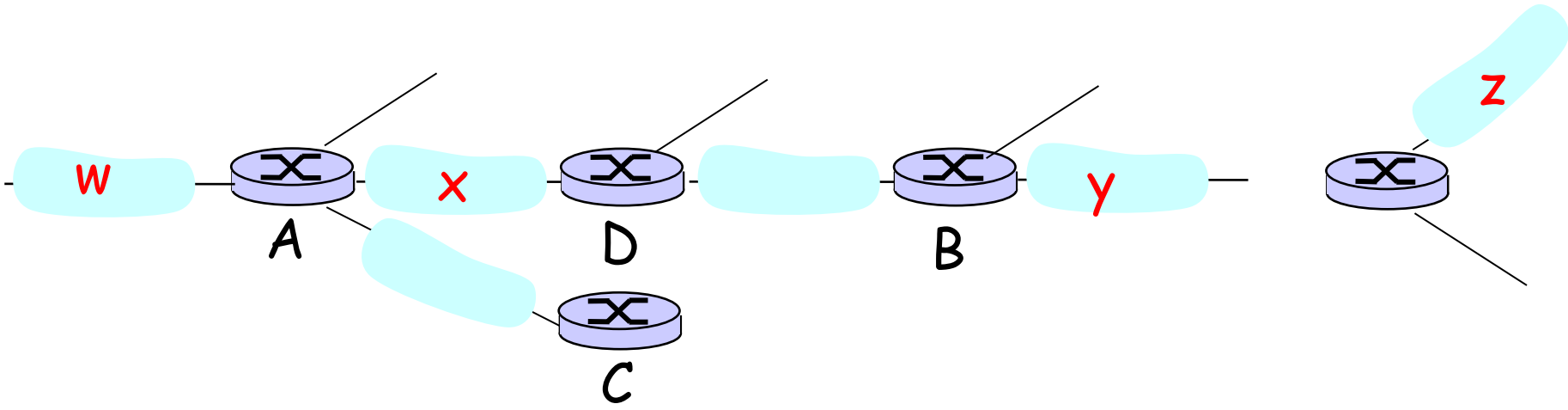
- Συνώνυμο: Interior Gateway Protocols (IGP)
- Τα πιο γνωστά IGPs:
  - **RIP**: Routing Information Protocol (Distance Vector algorithm)
  - **OSPF**: Open Shortest Path First (Link State algorithm)
  - **IGRP**: Interior Gateway Routing Protocol (Distance Vector algorithm)
    - Cisco proprietary protocol
    - Όμοιο με RIP
    - Σε αντίθεση προς το RIP χρησιμοποιεί το TCP για routing updates

RIP (Routing Information Protocol)

- Distance vector algorithm
- Included in BSD-UNIX Distribution in 1982
- Distance metric: # of hops (**max = 15 hops**)

  (*δηλαδή θεωρείται ότι κάθε ζεύξη έχει κόστος 1*)
  - *Can you guess why?*
- Distance vectors: exchanged every **30 sec** via Response Message (also called **advertisement**)
- Each advertisement: route to up to 25 **destination nets**

RIP (Routing Information Protocol)



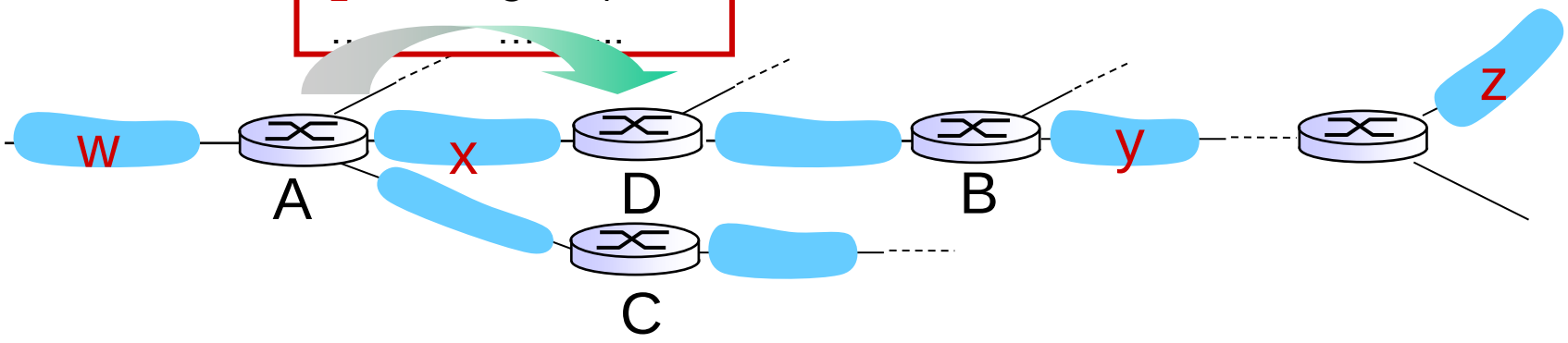
Routing table in D

Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	B	7
x	--	1
….	….	....

Παράδειγμα δημοσιοποίησης (RIP - advertisement)

Δημοσιοποίηση από το A-στο-D

dest	next	hops
w	-	1
x	-	1
z	C	4
..	...	...



Αλλαγή στον πίνακα δρομολόγησης του δρομολογητή D

Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

A

5

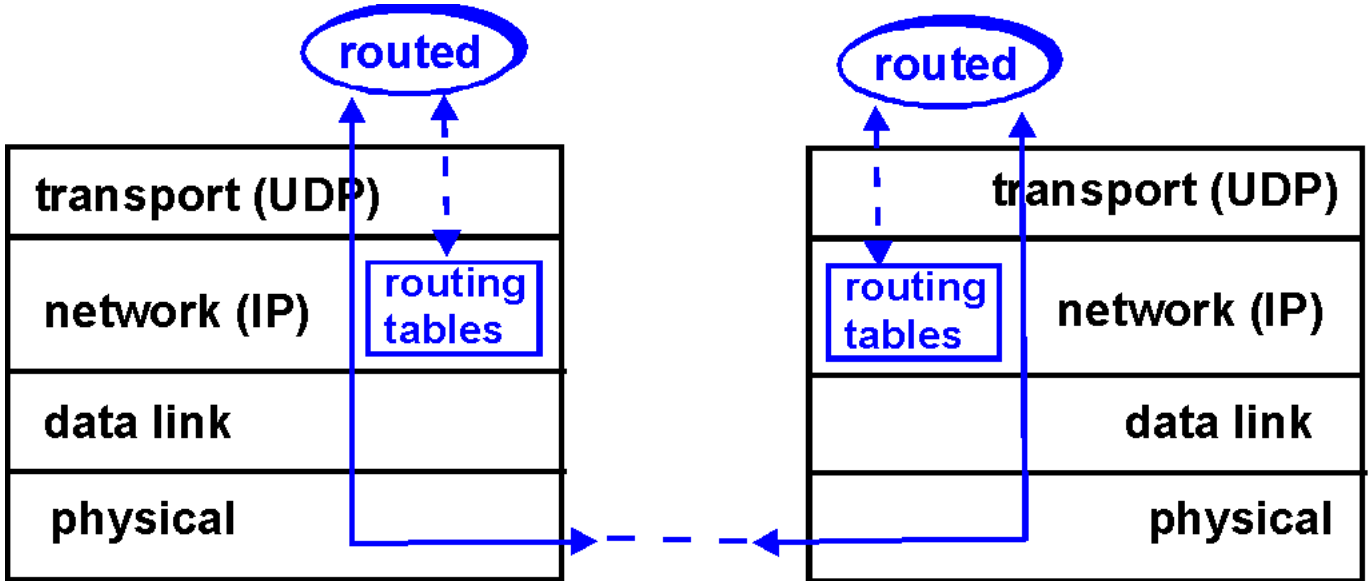
RIP: Link Failure and Recovery

If no advertisement heard after **180 sec** -->
neighbor/link declared dead

- routes via neighbor invalidated
- new advertisements sent to neighbors
- neighbors in turn send out new advertisements (if tables changed)
- link failure info quickly propagates to entire net
- poison reverse (αθώα ψέματα) used to prevent ping-pong loops (**infinite distance = 16 hops**)

RIP Table processing

- RIP routing tables managed by a**pplication-level** process called route-d (daemon)
- advertisements sent in UDP packets, periodically repeated



RIP Table example (continued)

Router: *giroflee.eurocom.fr*

Destination	Gateway	Flags	Ref	Use	Interface
127.0.0.1	127.0.0.1	UH	0	26492	lo0
192.168.2.	192.168.2.5	U	2	13	fa0
193.55.114.	193.55.114.6	U	3	58503	le0
192.168.3.	192.168.3.5	U	2	25	qaa0
224.0.0.0	193.55.114.6	U	3	0	le0
default	193.55.114.129	UG	0	143454	

- **Three** attached **class C** networks (LANs)
- Router only knows routes to attached LANs
- **Default** router used to "go up"
- Route multicast address: 224.0.0.0 (class D)
- Loopback interface (for debugging) - 127.0.0.1

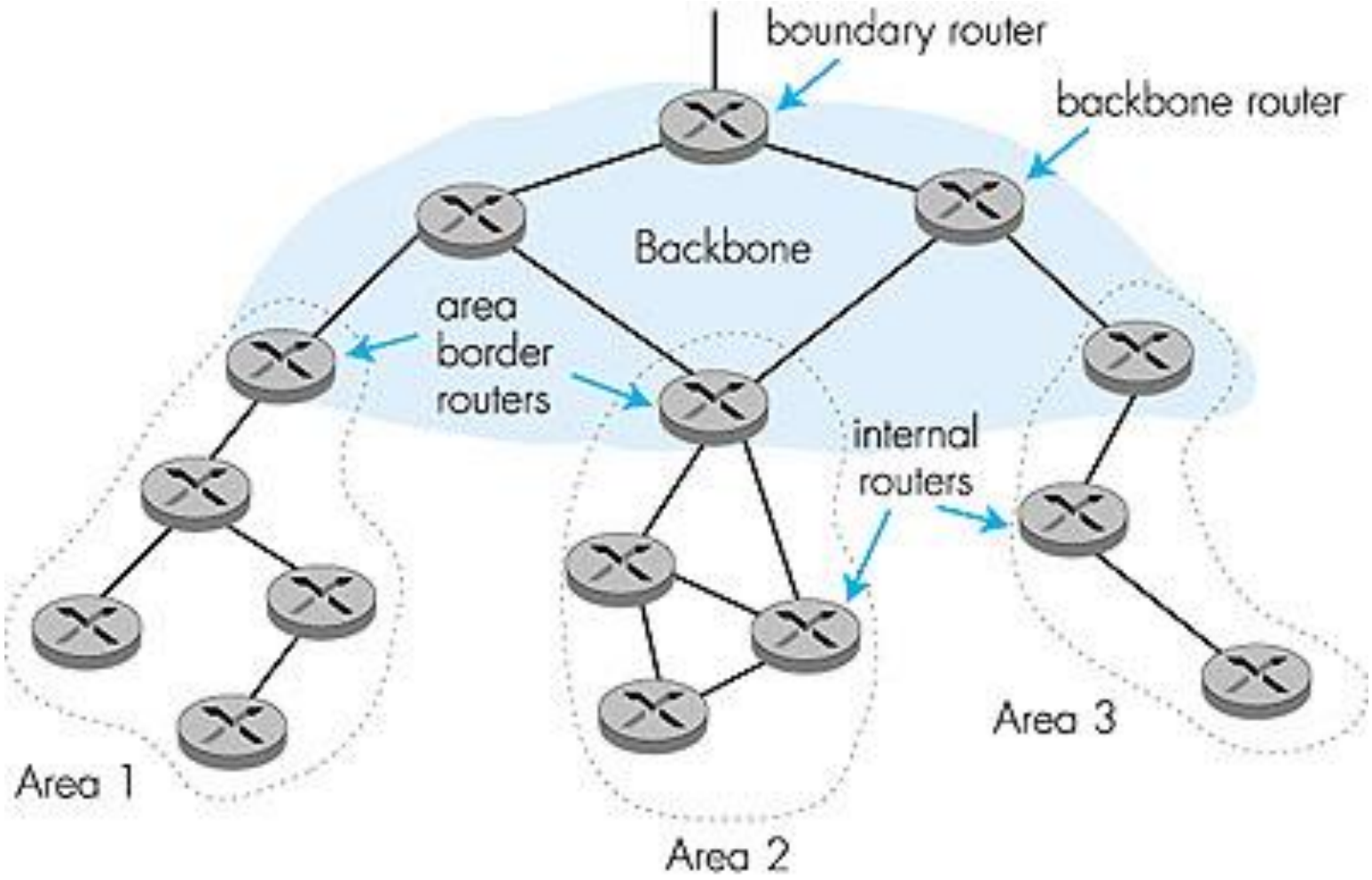
OSPF (Open Shortest Path First)

- "open": publicly available
- Uses Link State algorithm
  - LS packet dissemination
  - Topology map at each node
  - Route computation using Dijkstra's algorithm
- **Hello** messages every **30 min.**
- Advertisements disseminated to entire AS (via flooding - πλημμύρας) (Δημοσιοποίηση διαδρομών προς όλους τους άλλους Routers).
- Η δημοσιοποίηση του OSPF μεταφέρεται μέσω του IP (upper layer 89).

OSPF "advanced" features (not in RIP)

- Security: all OSPF messages authenticated (to prevent malicious intrusion); TCP connections used
- Multiple same-cost paths allowed (only one path in RIP)
- For each link, multiple cost metrics for different TOS (eg, satellite link cost set "low" for best effort; high for real time)
- Hierarchical OSPF in large domains **(large AS)**.

Hierarchical OSPF



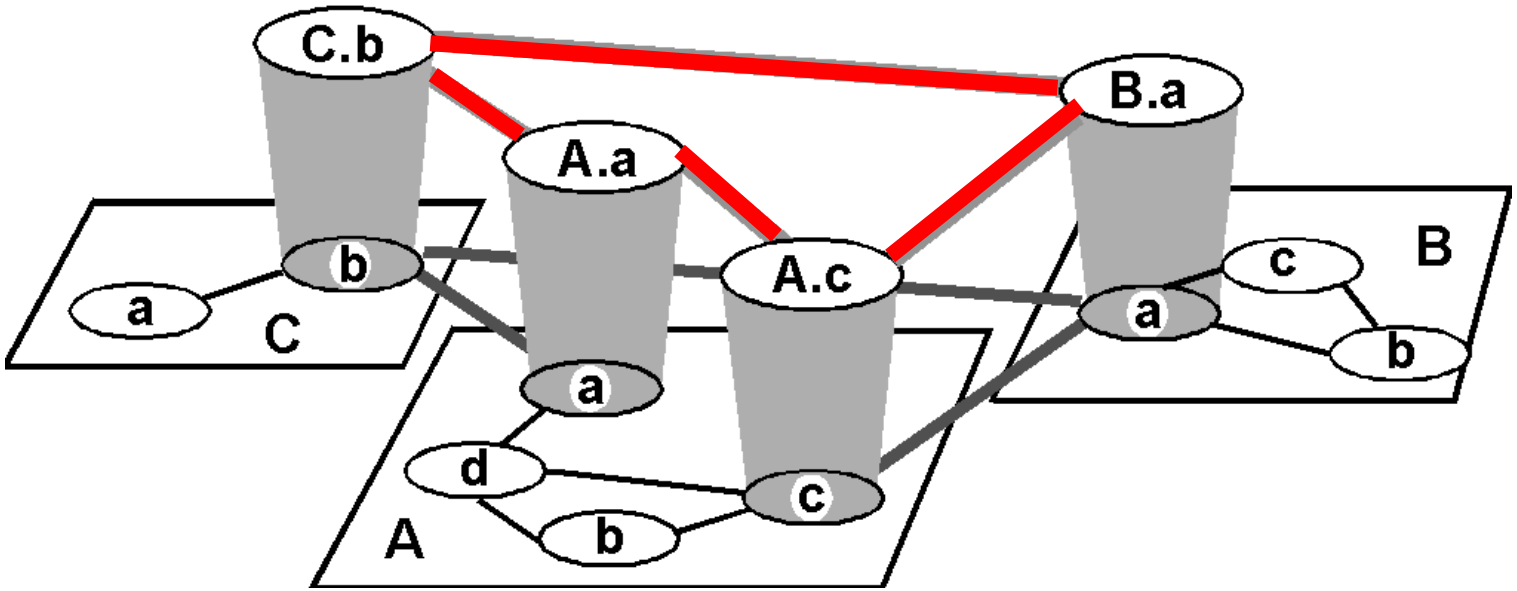
Hierarchical OSPF

- Two-level hierarchy: local area, backbone.
  - Link-state advertisements only in area
  - each nodes has detailed area topology; only know direction (shortest path) to nets in other areas.
- **Area border routers:** "summarize" distances to nets in own area, advertise to other Area Border routers.
- **Backbone routers:** run OSPF routing limited to backbone.
- **Boundary routers:** connect to other ASs.

IGRP (Interior Gateway Routing Protocol)

- CISCO proprietary; successor of RIP (mid 80s)
- Distance Vector, like RIP
- several cost metrics (delay, bandwidth, reliability, load etc)
- uses TCP to exchange routing updates
- Loop-free routing via Distributed Updating Alg. (DUAL) based on *diffused computation*

Inter-AS routing



Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): *the* de facto standard
- **Path Vector** protocol:
  - similar to Distance Vector protocol
  - each Border Gateway broadcast to neighbors (peers) *entire path* (i.e., sequence of ASs) to destination
  - E.g., Gateway X may send its path to dest. Z:

    Path (X,Z) = X,Y1,Y2,Y3,…,Z

Internet inter-AS routing: BGP

Suppose: gateway X send its path to peer gateway W

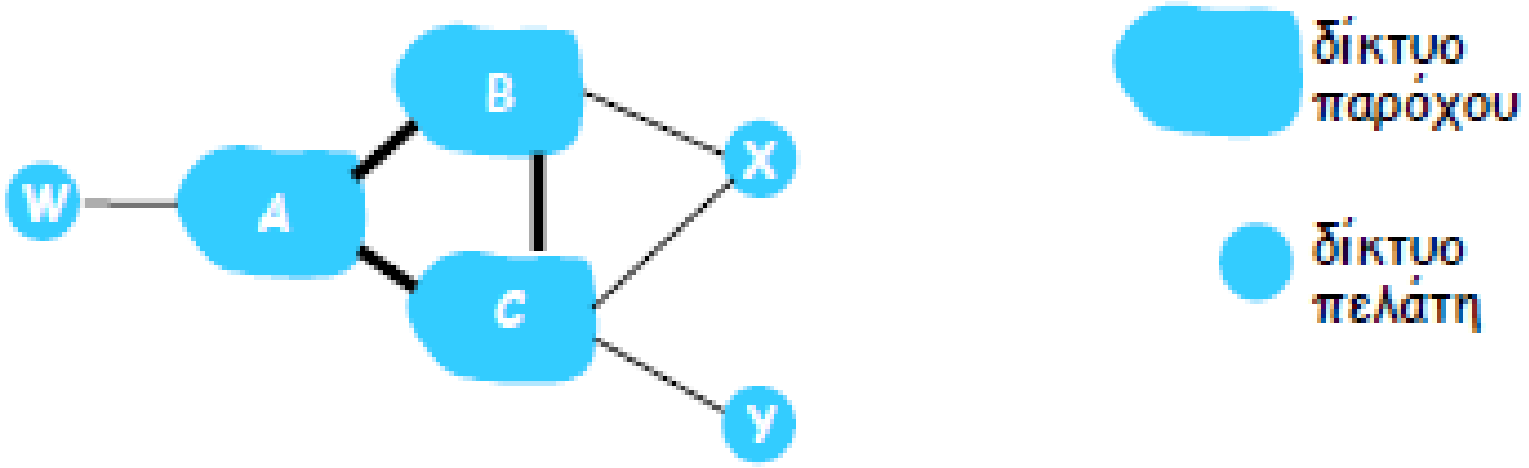
- W may or may not select path offered by X
  - cost, policy (don't route via competitors AS), loop prevention reasons.
- If W selects path advertised by X, then:

  Path (W,Z) = w, Path (X,Z)
- Note: X can control incoming traffic by controling it route advertisements to peers:
  - e.g., don't want to route traffic to Z -> don't advertise any routes to Z

Internet inter-AS routing: BGP

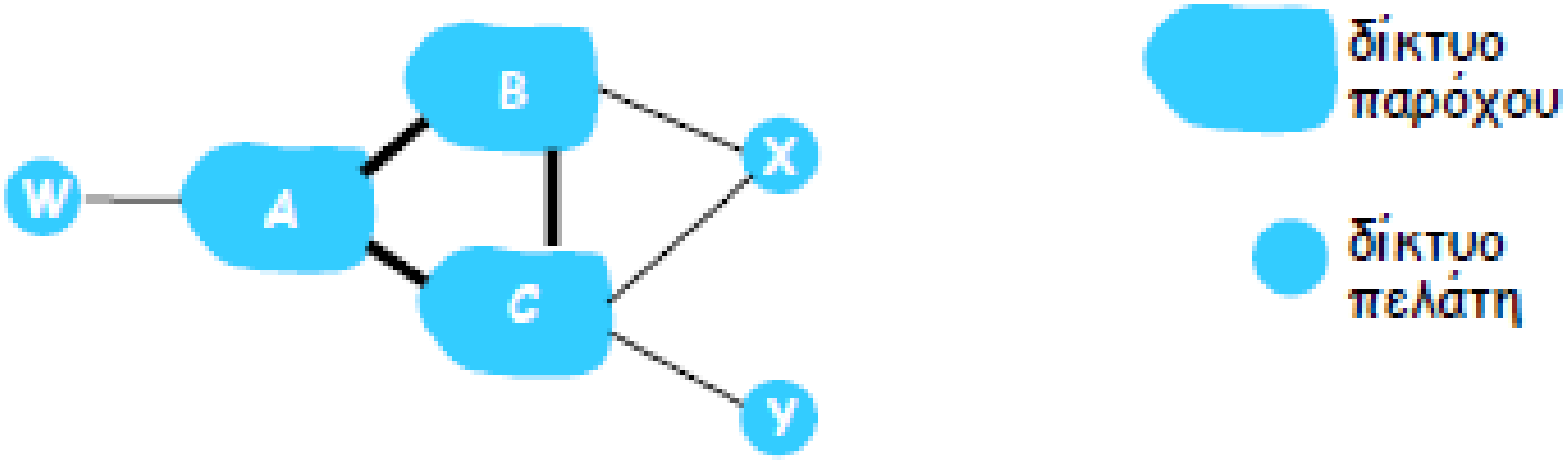
- Το BGP επιτρέπει σε κάθε υποδίκτυο να διαφημίζει την ύπαρξή του στο υπόλοιπο Διαδίκτυο: **"Είμαι εδώ"**
- Το BGP παρέχει σε κάθε AS ένα τρόπο για:
  - **eBGP**: να λαμβάνει πληροφορίες προσέγγισης υποδικτύου από γειτονικά AS (external).
  - **iBGP**: να διαδίδει τις πληροφορίες προσέγγισης σε όλους τους δρομολογητές που είναι εσωτερικοί στο AS (internal).
  - Να καθορίζει τις "καλές" διαδρομές προς άλλα δίκτυα με βάση τις πληροφορίες προσέγγισης και μια πολιτική δρομολόγησης.

BGP – Πολιτική Δρομολόγησης



- Τα A,B,C είναι δίκτυα παρόχων (provider networks)
- Τα X,W,Y είναι πελάτες (των δικτύων παρόχων)
- Το X είναι διεστιακό (dual-homed): συνδέεται σε δύο δίκτυα
  - Το X δεν θέλει να δρομολογεί από το B μέσω του X προς το C
  - .. έτσι το X δεν θα δημοσιοποιήσει στο B μια διαδρομή προς το C

BGP – Πολιτική Δρομολόγησης (συνέχεια)



- Το A διαφημίζει τη διαδρομή AW στο B
- Το B διαφημίζει τη διαδρομή BAW στο X
- Θα πρέπει το B να διαφημίσει τη διαδρομή BAW στο C;
  - Σε καμία περίπτωση! Το B δεν έχει όφελος από τη δρομολόγηση CBAW, καθώς ούτε το W ούτε το C είναι πελάτες του B
  - Το B θέλει να εξαναγκάσει το C να δρομολογεί προς το W μέσω του A
  - Το B θέλει να δρομολογεί *μόνο* προς/από τους πελάτες του!

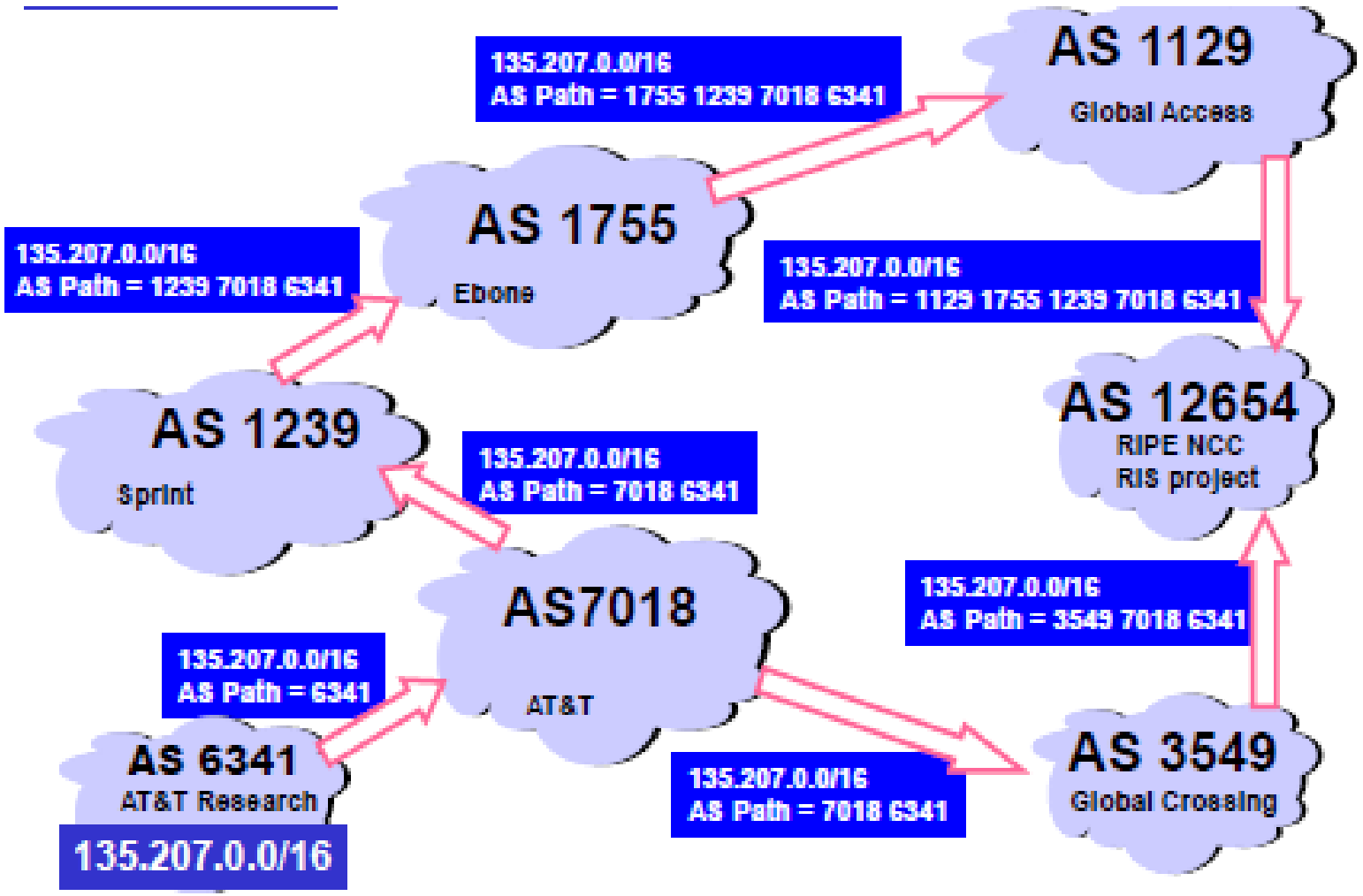
Subnet – Prefix – BGP route

- Στο Διαδίκτυο, το **subnet** (υποδίκτυο στην ελληνική, αλλά αποφεύγετε: subnetwork) είναι ένα μέρος ενός μεγαλύτερου δικτύου.
  - Το subnet ΔΕΝ περιλαμβάνει δρομολογητή (Router).
  - Τα όριά του, ορίζονται από τον Router και τα host interfaces.
- Το **prefix** (πρόθεμα) είναι ένα μέρος (το αρχικό) μιας διεύθυνσης IP.
  - Παραπέμπει στην μορφή CDIR της διεύθυνσης IP.
  - Αν έχουμε a.b.c.d/x το prefix είναι τα x πρώτα bits.
  - Ένα prefix περιλαμβάνει ένα ή περισσότερα subnets.
- Ο BGP route περιλαμβάνει ένα prefix + ιδιοχαρακτηριστικά διαδρομής
  - Όταν ένα gateway «διαφημίζει» ένα prefix σε μια επικοινωνία μέσω του πρωτοκόλλου BGP, το prefix ακολουθείται από ορισμένα χαρακτηριστικά.
  - Μπορούμε να λέμε απλά route (αντί BGP route).
  - Χρήσιμη έκφραση στην Αγγλική: "The route of the path"

BGP – Ιδιοχαρακτηριστικά διαδρομής

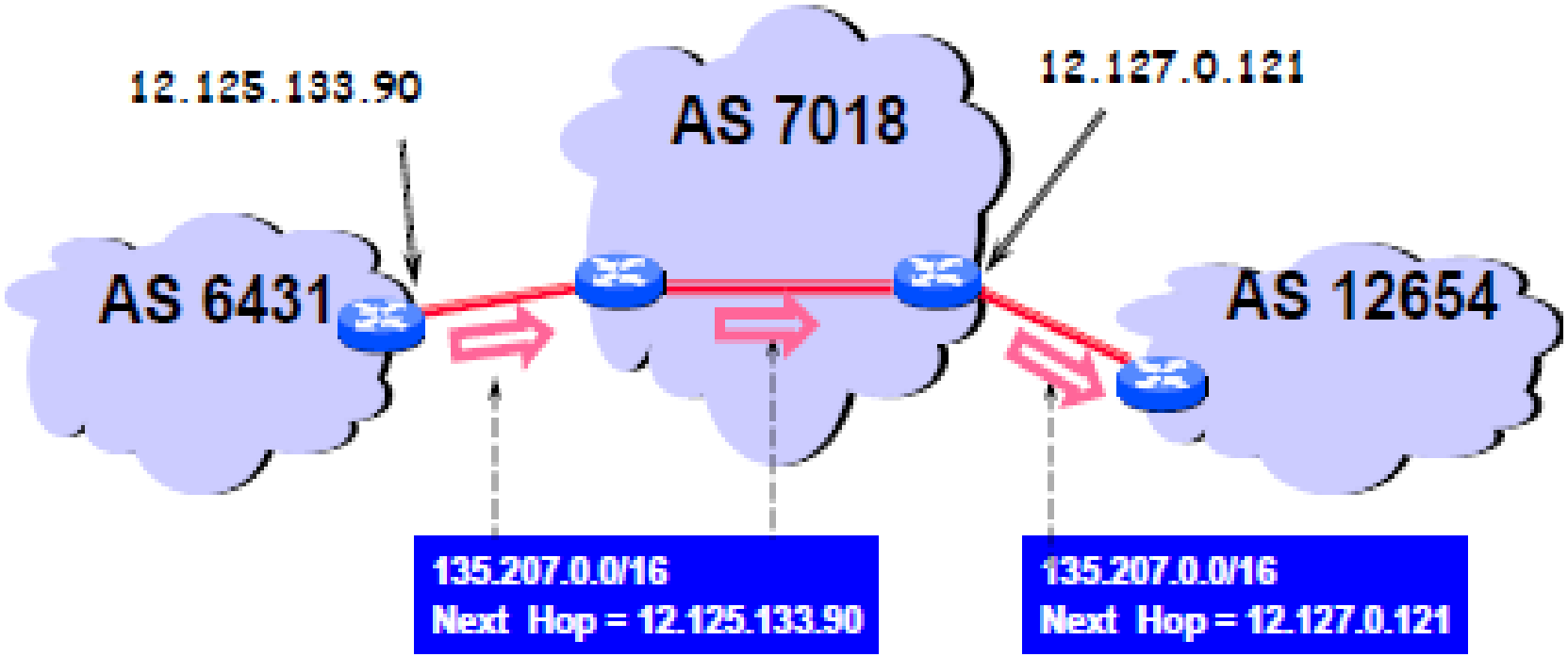
- Δύο σημαντικά ιδιοχαρακτηριστικά:
- **AS Path**: περιλαμβάνει τα AS μέσω των οποίων έχει περάσει η διαφήμιση για το πρόθεμα: π.χ., AS 67, AS 17
- **Next Hop**: indicates specific internal-AS router to next-hop AS (may be multiple links from current AS to next-hop-AS)
- Όταν ένας gateway λαμβάνει τη διαφήμιση διαδρομής, αποφασίζει αν θα την αποδεχθεί ή όχι, ανάλογα με την πολιτική εισαγωγής (import policy) διαδρομών που θέλει (π.χ. Ποτέ δεν δρομολογώ κίνηση μέσω του AS X).

Παράδειγμα AS Path



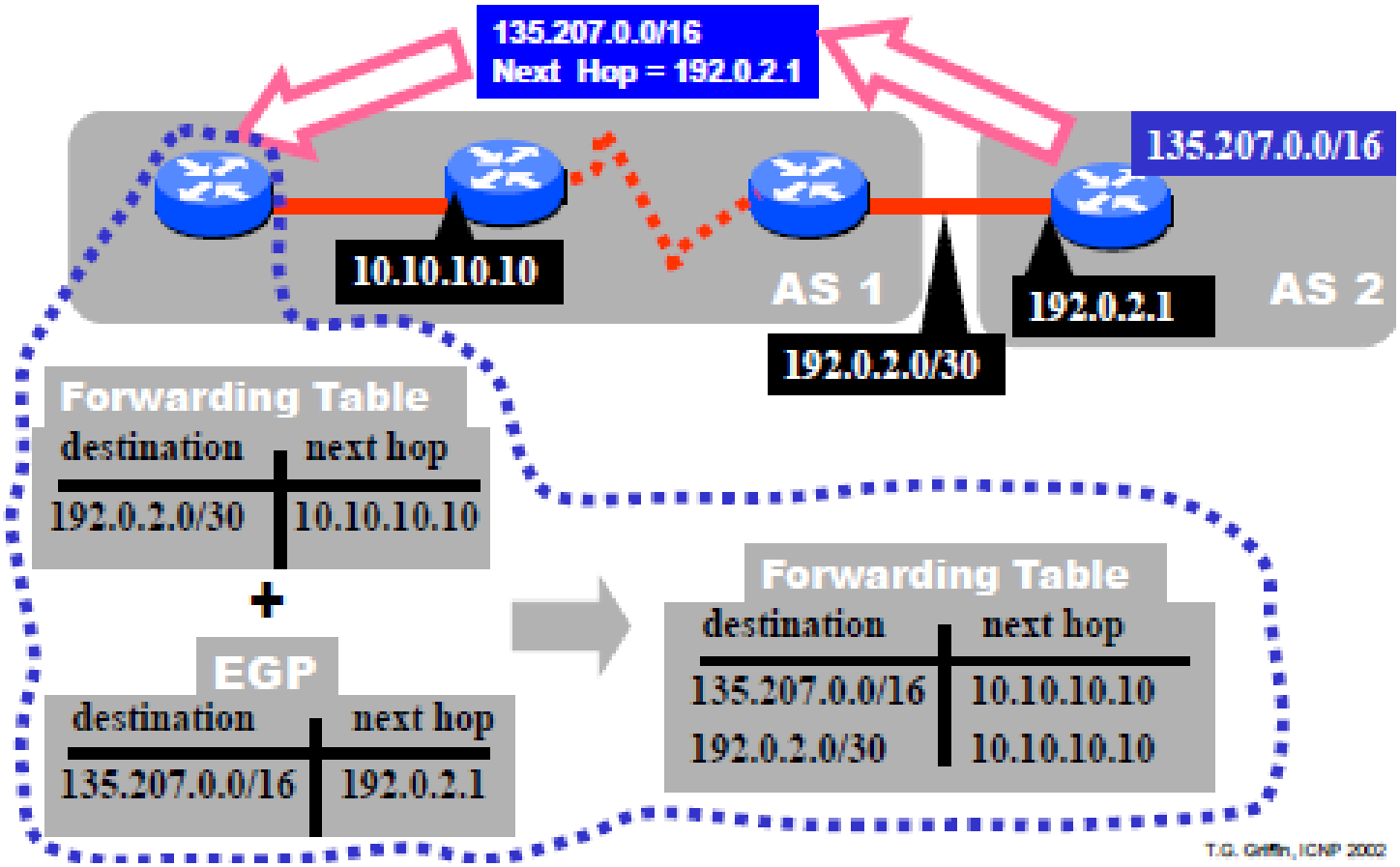
Αφετηρία προθέματος

Παράδειγμα Next Hop



- Όταν μια ανακοίνωση prefix φεύγει από ένα AS μέσω ενός συνοριακού δρομολογητή, το Next Hop παίρνει την τιμή της διεύθυνσης IP του συνοριακού δρομολογητή (του interface μέσω του οποίου εστάλη η ανακοίνωση).

Παράδειγμα ενημέρωσης πίνακα προώθησης



- Στον πίνακα προώθησης ενός Router συνδυάζονται οι πληροφορίες δρομολόγησης Intra-AS και Inter-AS routing.

BGP Sessions

- Το BGP παρέχει σε κάθε AS ένα τρόπο για:
  - **eBGP**: να λαμβάνει πληροφορίες προσέγγισης υποδικτύου από γειτονικά AS (external) – Δημιουργία session με εξωτερικά gateways.
  - **iBGP**: να διαδίδει τις πληροφορίες προσέγγισης σε όλους τους δρομολογητές που είναι εσωτερικοί στο AS (internal). Δημιουργία εσωτερικών sessions.
  - Να καθορίζει τις "καλές" διαδρομές προς άλλα δίκτυα με βάση τις πληροφορίες προσέγγισης και μια πολιτική δρομολόγησης.

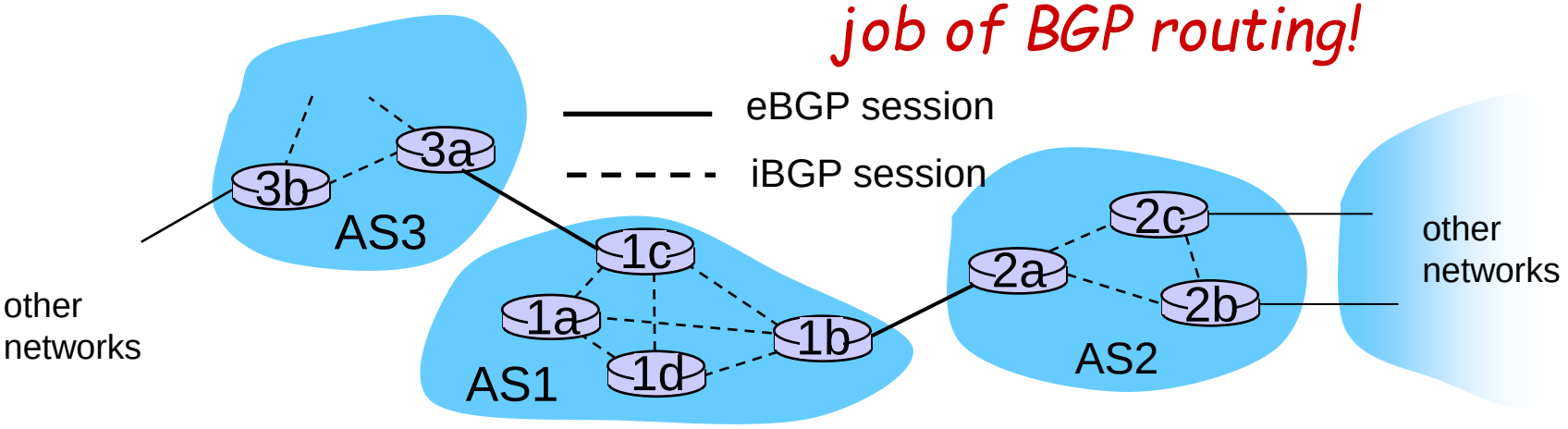
Inter-AS (BGP) tasks

- suppose router (e.g. 1d) in AS1 receives datagram destined outside of AS1:
  - router should forward packet to gateway router, but which one? (1c or 1b?)

AS1 must:

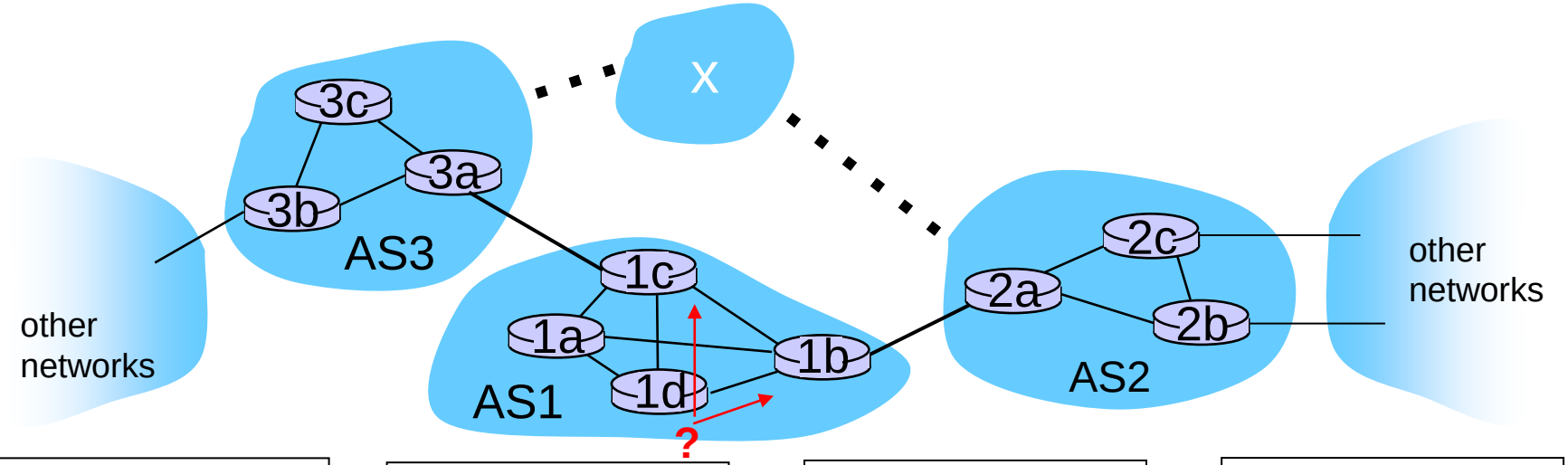
1. learn which dests are reachable through AS2, which through AS3 (eBGP)
2. propagate this reachability info to all routers in AS1 (iBGP)

job of BGP routing!



Example: choosing among multiple ASes

- now suppose AS1 learns from inter-AS protocol that subnet *x* is reachable from AS3 *and* from AS2.
- to configure forwarding table, router 1d must determine towards which gateway it should forward packets for dest *x*
- *hot potato routing: αποστολή στον πλησιέστερον Router ώστε να βγει γρήγορα εκτός AS1*



learn from inter-AS protocol that subnet *x* is reachable via multiple gateways

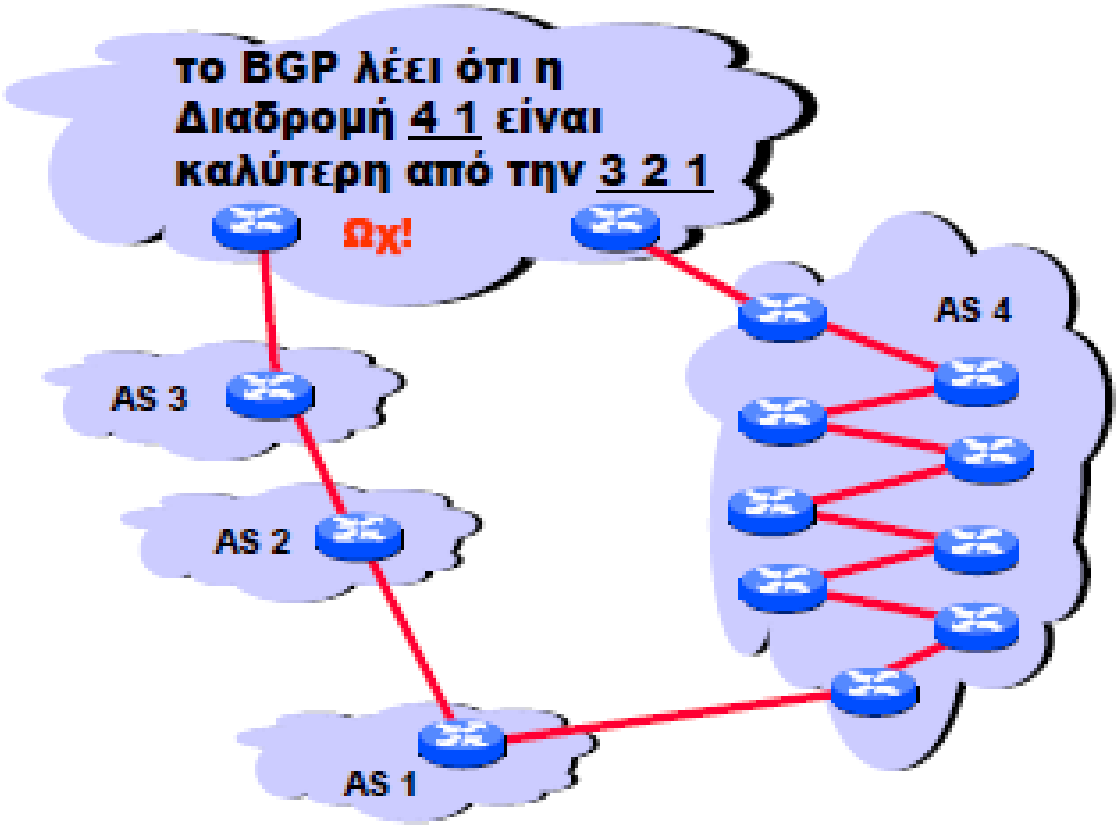
→ use routing info from intra-AS protocol to determine costs of least-cost paths to each of the gateways

→ hot potato routing: choose the gateway that has the smallest least cost

→ determine from forwarding table the interface *I* that leads to least-cost gateway. Enter *(x,I)* in forwarding table

Πολιτική Hot Potato

- Δεν ενδιαφέρεται για την επιλογή της καλύτερης διαδρομής
- Ενδιαφέρεται να διώξει το datagram εκτός AS το συντομέτερο δυνατόν



Internet inter-AS routing: BGP

- BGP messages exchanged using TCP.
- BGP messages:
  - OPEN: opens TCP connection to peer and authenticates sender
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: reports errors in previous msg; also used to close connection

Why different Intra- and Inter-AS routing ?

Policy:

- Inter-AS: admin wants control over how its traffic routed, who routes through its net.
- Intra-AS: single admin, so no policy decisions needed

Scale:

- hierarchical routing saves table size, reduced update traffic

Performance:

- Intra-AS: can focus on performance
- Inter-AS: policy may dominate over performance