

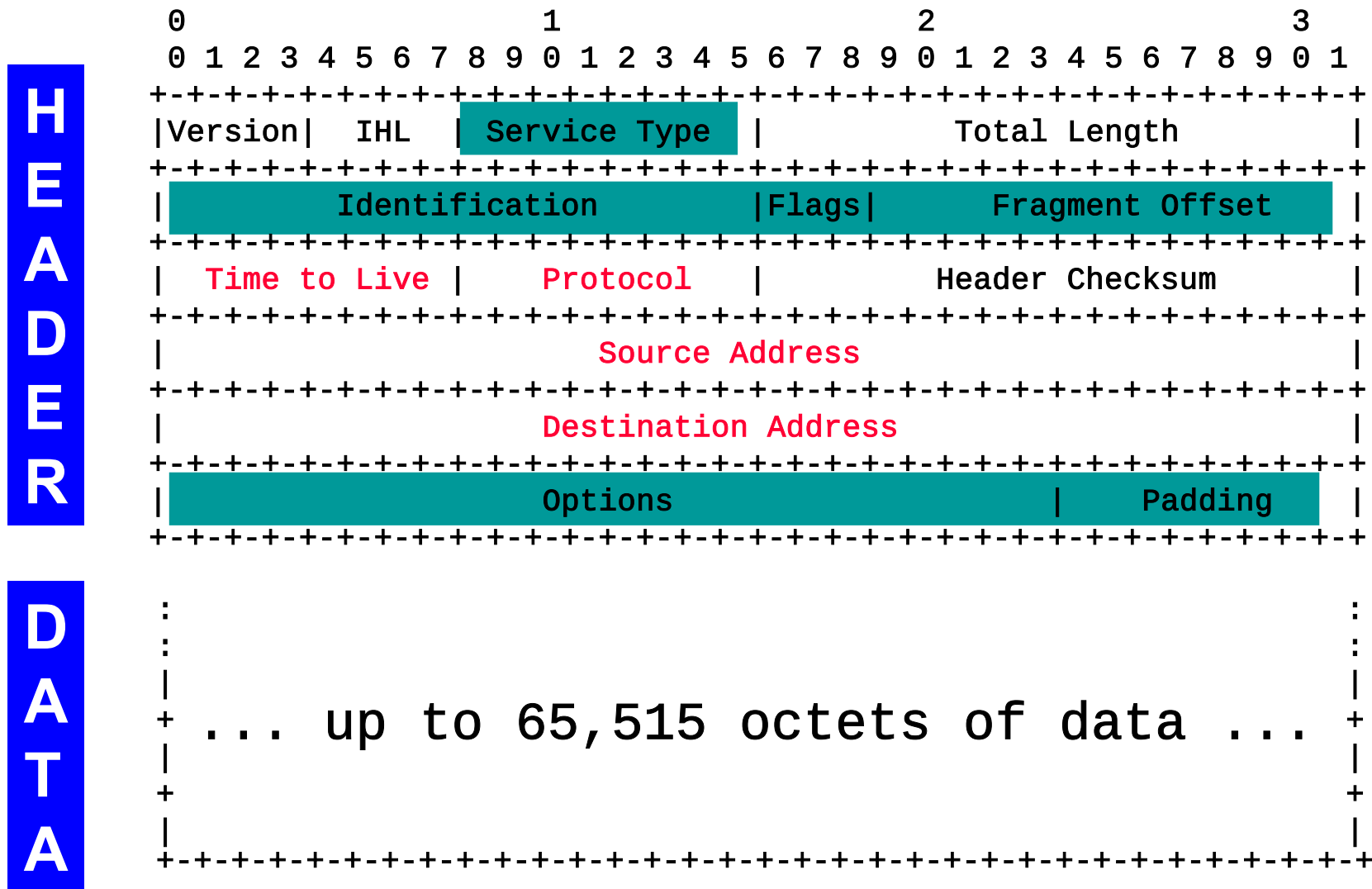
BGP, Border Gateway Protocol

Τελευταία ενημέρωση: 09/09/2004

- Στοιχεία Δρομολόγησης
- Εισαγωγή στο BGP
- Λειτουργία
- Μηνύματα
- Ιδιότητες Διαδρομών (Attributes)
- Επιλογή Διαδρομής
- Τακτική (Policy)
- Δυνατότητες (Capabilities)
- Ευστάθεια-Δυναμική (Stability-Dynamics)
- Κλιμάκωση (Scaling)
- Κατασκευαστές
- Βιβλιογραφία
- Διάρθρωση (...μέρος 2)

Στοιχεία Δρομολόγησης

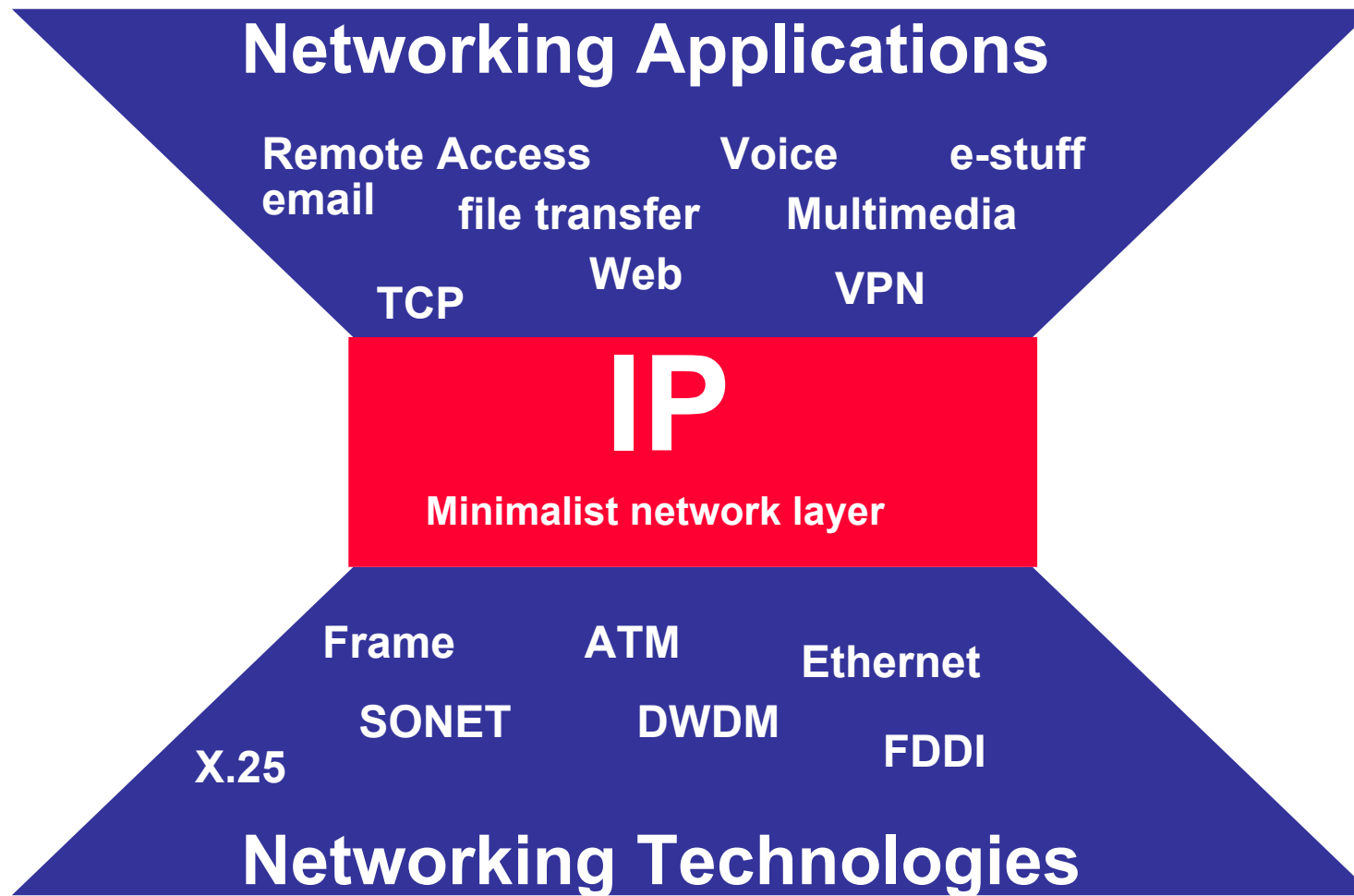
IP πακέτο (datagram)



μικρή χρήση

1981, RFC 791

Όλα (θα) περιστρέφονται γύρω από το IP;



- Πάνω στο IP περνάνε ένα πλήθος από εφαρμογές
- Το IP μπορεί να πατάει σε ένα πλήθος από διαφορετικές τεχνολογίες μετάδοσης

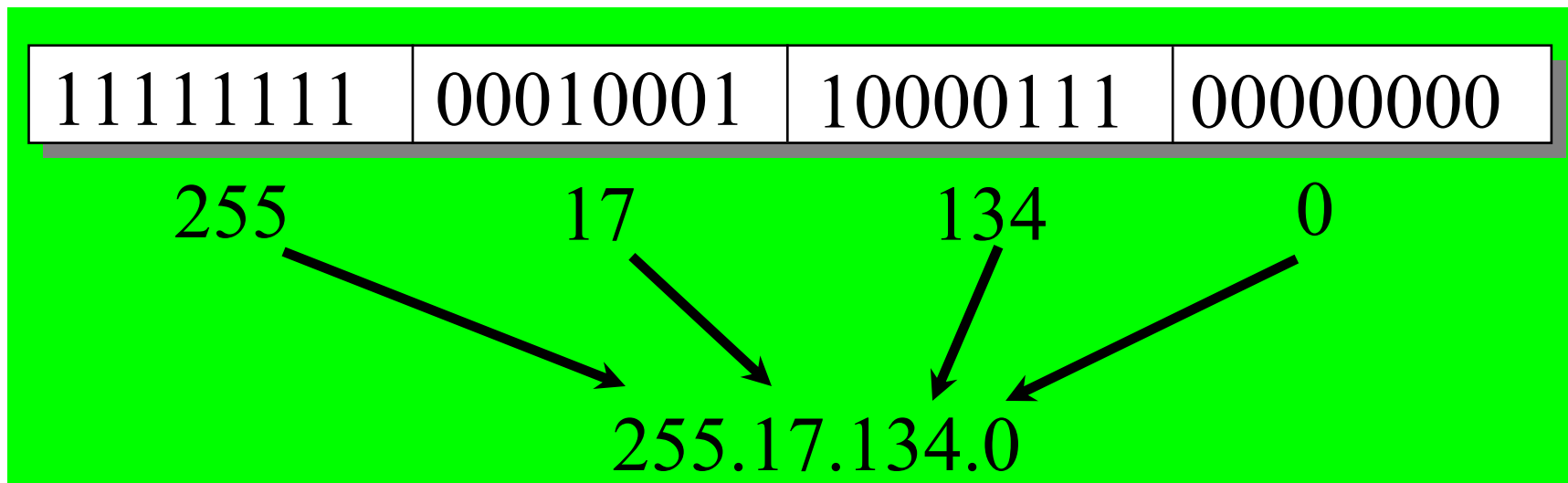
Best Effort Connectivity



- Η βασική υπηρεσία που παρέχεται σήμερα από τα IP δίκτυα είναι απλή σύνδεση χωρίς εγγυήσεις για τις **παραμέτρους** επικοινωνίας (delay, jitter, bandwidth, packet loss) και η οποία ονομάζεται **best effort** ή **no guarantee** ή **sent and pray**

IP Διευθυνσιοδότηση

- Στο internet χρησιμοποιείται προς το παρόν **IPv4** διευθυνσιοδότηση, όπου οι διευθύνσεις έχουν μέγεθος **32bits** (4.5δισ διευθύνσεις) , ενώ μελλοντικά με την **IPv6** διευθυνσιοδότηση θα έχουμε διευθύνσεις μήκους **128bits**
- Οι διευθύνσεις γράφονται σε μορφή **Φ.X.Υ.Ω** όπου Φ, X, Ψ, Ω είναι 8bit αριθμοί
- Οι διευθύνσεις **224.0.0.0-239.255.255.255** (multicast) και **240.0.0.0-255.255.255.255** (μελλοντική χρήση) είναι **δεσμευμένες** ([RFC 3330](#))
- Επίσης και τα δίκτυα 0.0.0.0/8, 14.0.0.0/8, 127.0.0.0/8 είναι **δεσμευμένα**
- Οι υπόλοιπες **1.0.0.0** ως **223.255.255.255** είναι **διαθέσιμες**



Subnetting

- Η κάθε διεύθυνση έχει ένα τμήμα που ορίζει το **δίκτυο** και το οποίο θα ονομάζουμε **πρόθεμα** ή **δίκτυο** ή **πρόθεμα δικτύου** και ένα που ορίζει την **συσκευή**
- Το μέγεθος τους καθορίζεται από τη **μάσκα**

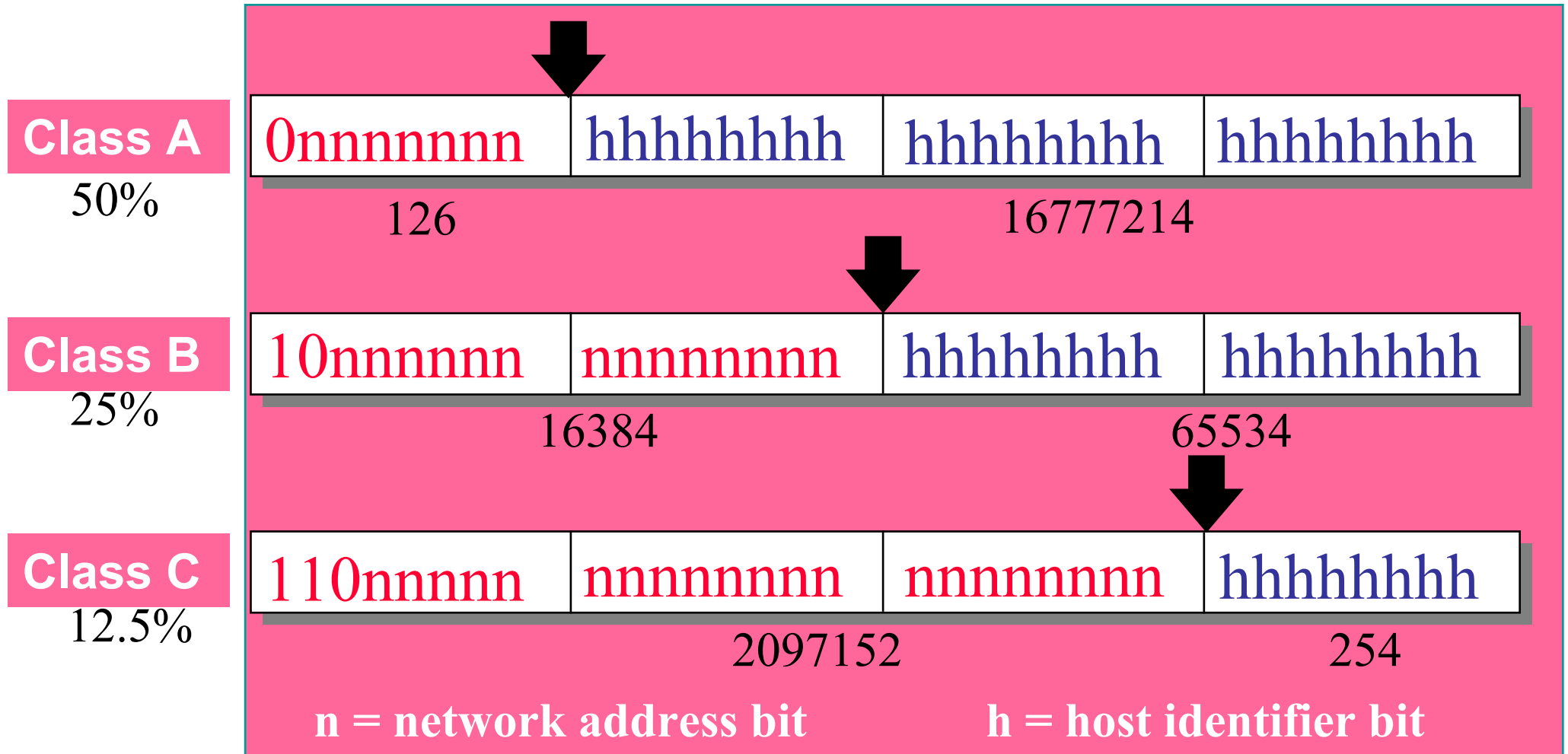
```
8.20.15.1 = 00001000.00010100.00001111.00000001
255.0.0.0 = 11111111.00000000.00000000.00000000
-----
net id |      host id
```

```
netid = 00001000 = 8
```

```
hostid = 00010100.00001111.00000001 = 20.15.1
```

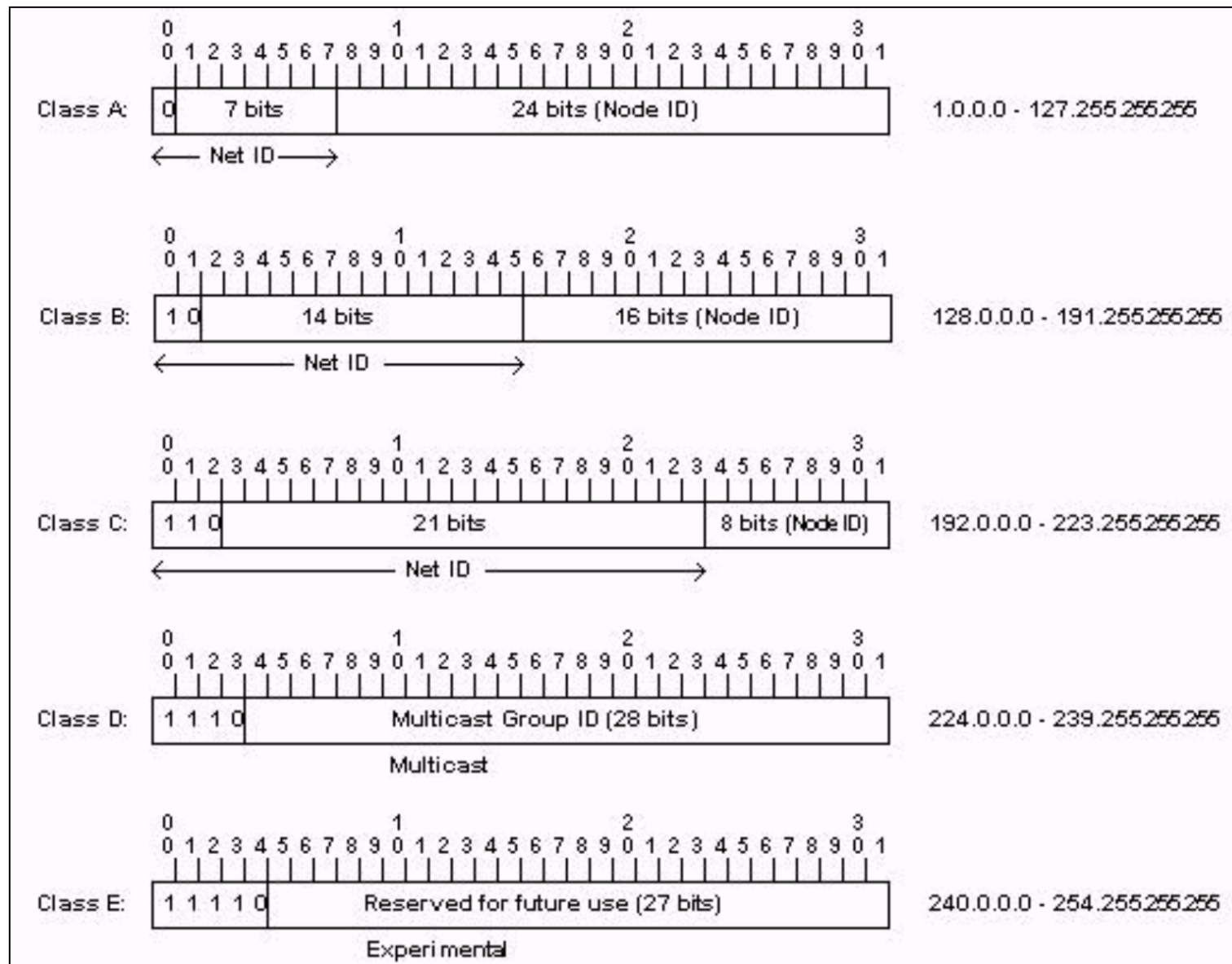
Τάξεις δικτύων

- Ανάλογα το πώς χωρίζεται η διεύθυνση έχουμε πέντε τάξεις δικτύων: A, B, C, D και E



- Σπάταλη**, μη αποτελεσματική ανάθεση διευθύνσεων
- Σπάνια αναφέρονται πλέον λόγω της χρήσης του **CIDR**

Τάξεις δικτύων

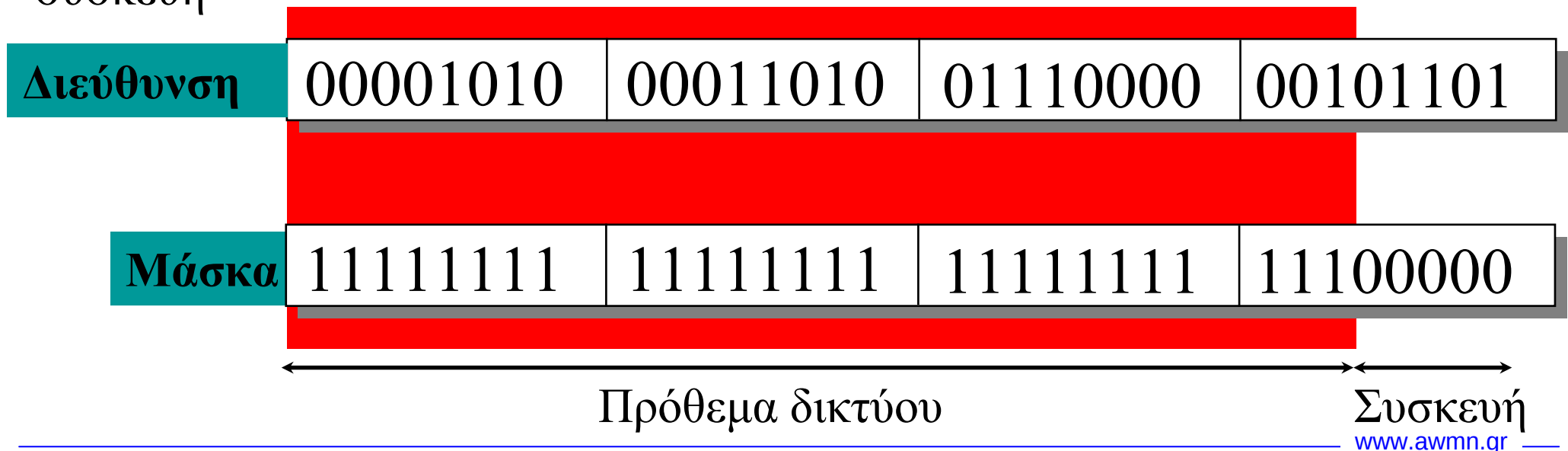


RFC 1518,1519: Classless Inter-Domain Routing (CIDR)

- Μπορούμε να κόψουμε μία περιοχή διευθύνσεων με όποιο τρόπο θέλουμε, άροντας τον περιορισμό των κλάσεων και κάνοντας πιο αποτελεσματική **αξιοποίηση** του χώρου των διευθύνσεων
- Σχετικά στατιστικά, αναφορές : <http://www.cidr-report.org/>

Παράδειγμα

- 10.26.122.45 με μάσκα 255.255.255.224 ή αλλιώς 10.26.122.45/27 σημαίνει ότι τα 27 πρώτα bits ορίζουν το δίκτυο, ενώ τα υπόλοιπα την συσκευή



Subnetting σε ένα C τάξεως υποδίκτυο

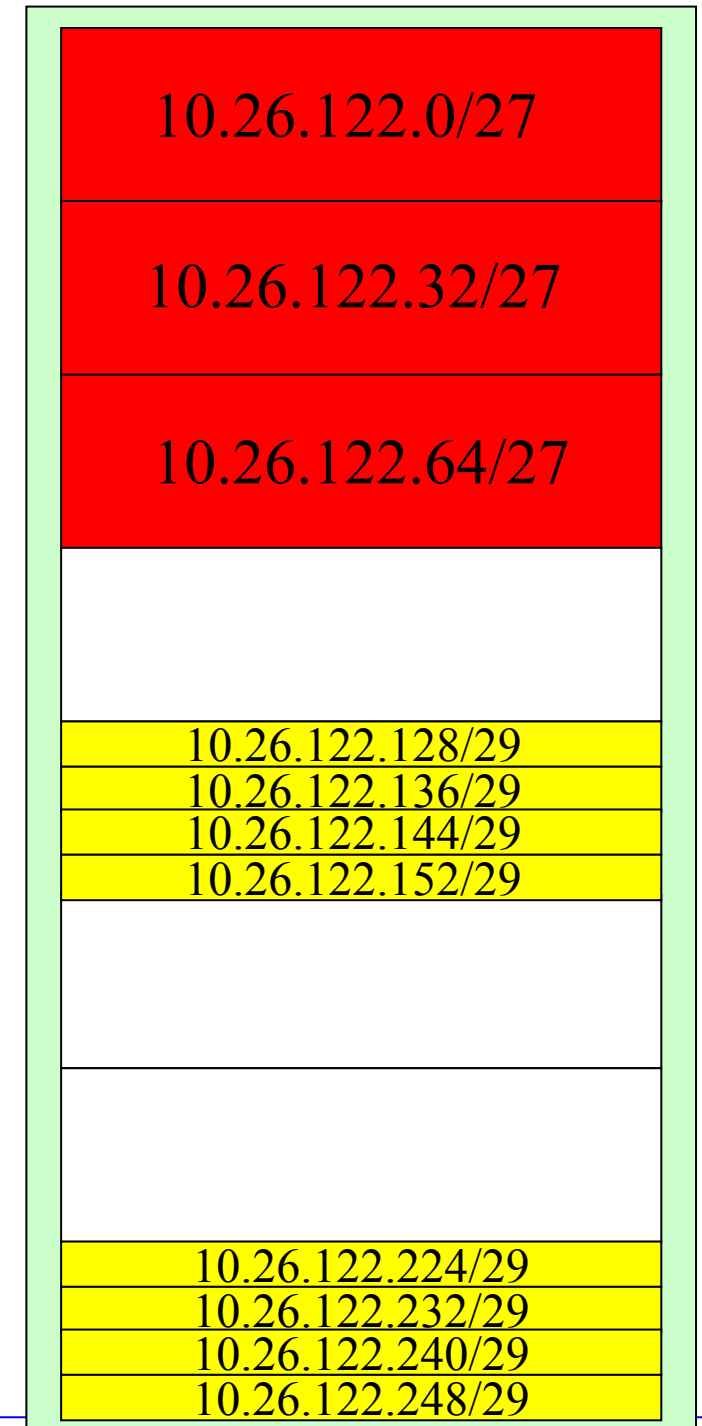
- Μπορούμε να το κόψουμε σε πολλά διαφορετικού μεγέθους υποδικτυα, αρκεί αυτά να μην **επικαλύπτονται**
- Έτσι **αξιοποιούμε** καλύτερα το διαθέσιμο χώρο διευθύνσεων

Μάσκα		Αριθμός διευθύνσεων	Αριθμός συσκευών
128 (/25)	10000000	128	126
192 (/26)	11000000	64	62
224 (/27)	11100000	32	30
240 (/28)	11110000	16	14
248 (/29)	11111000	8	6
252 (/30)	11111100	4	2

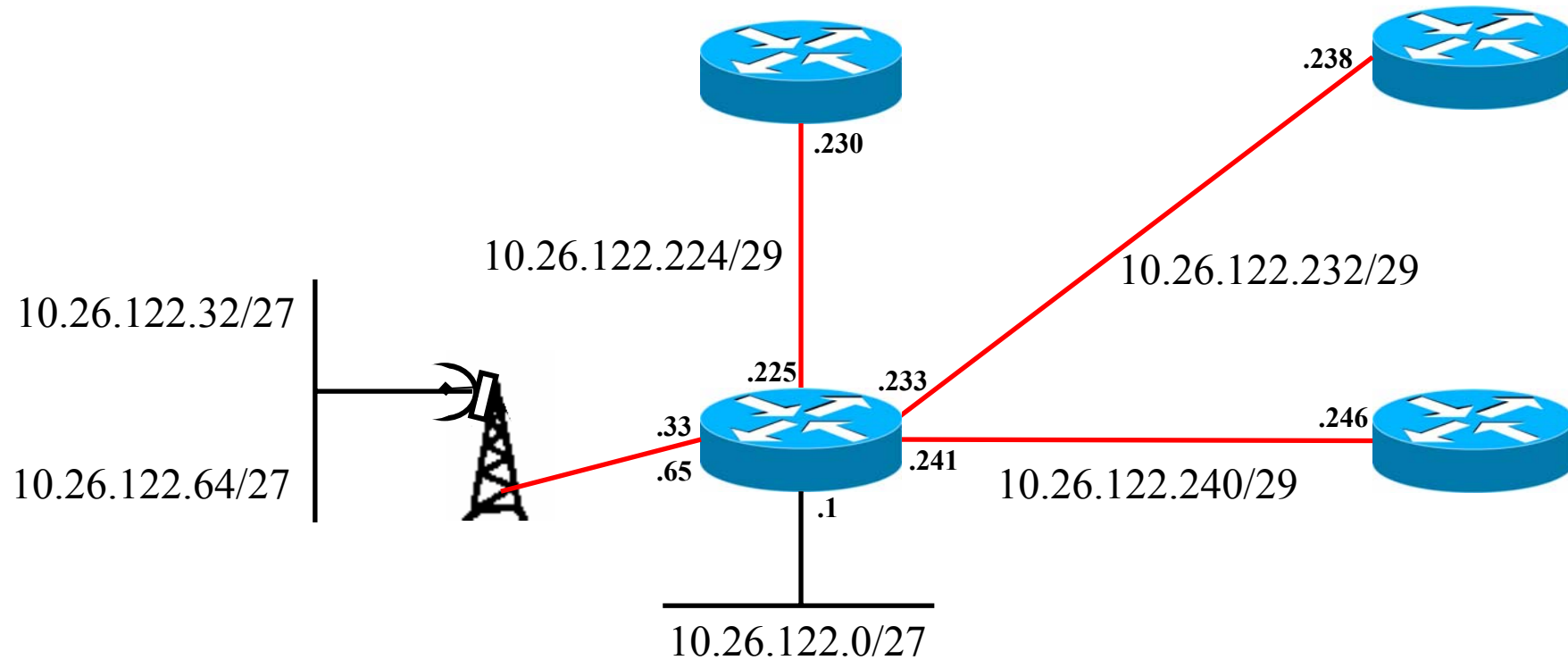
```
204.15.5.0 - 11001100.00001111.00000101.00000000
255.255.255.224 - 11111111.11111111.11111111.11100000
                  -----| sub |-----
```

VLSM, Variable Length Subnetting Mask, Παράδειγμα

- Επιτρέπει να χρησιμοποιούμε διαφορετικές μάσκες στα υποδίκτυα
- Αυτά πρέπει να μην επικαλύπτονται



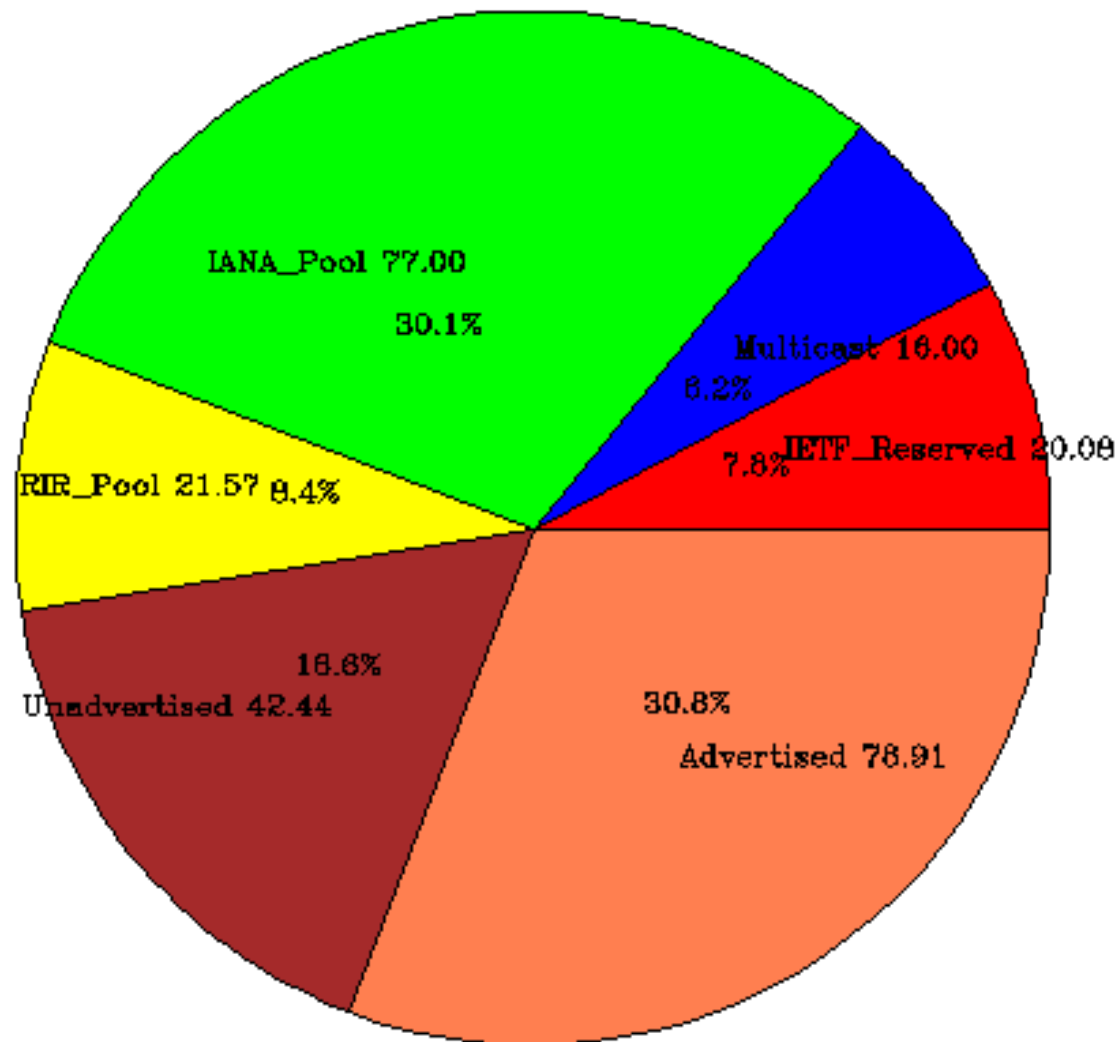
VLSM, Variable Length Subnetting Mask, Παράδειγμα



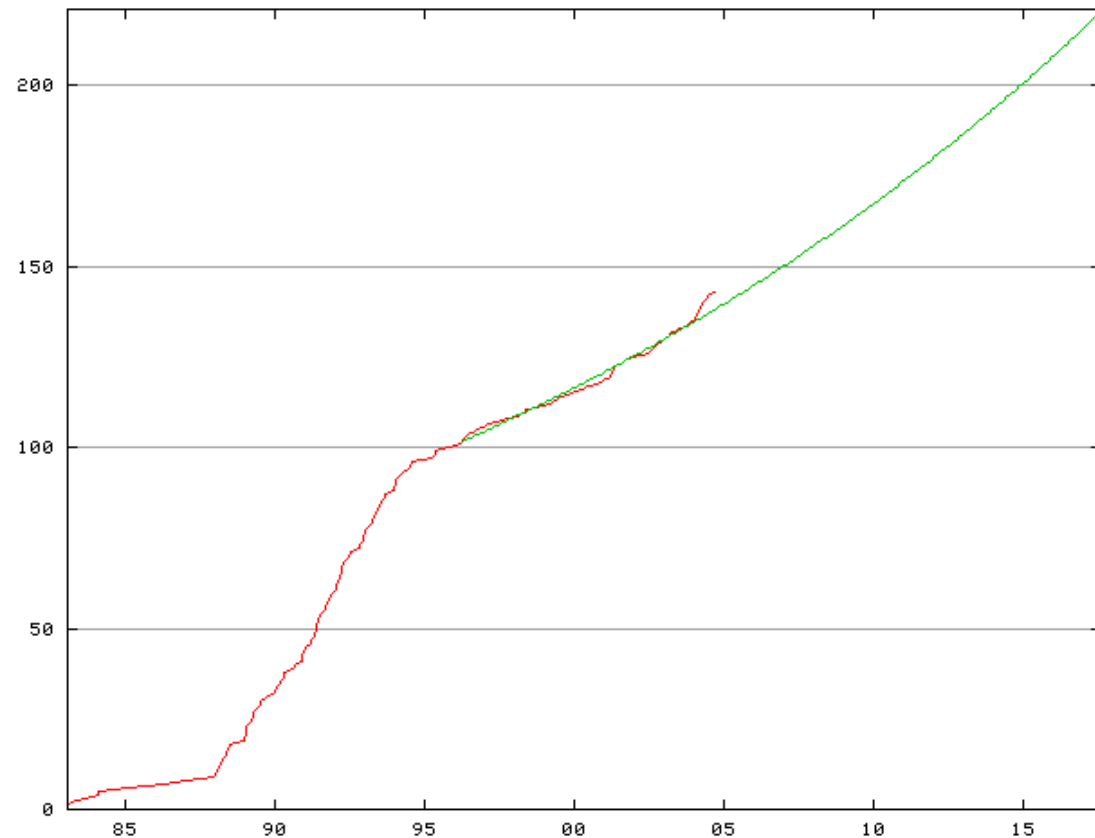
Τρέχουσα (09/04) χρήση του χώρου των IPv4 διευθύνσεων

- Εκφράζουμε το πλήθος των IP διευθύνσεων σε μονάδα μέτρησης, ισοδύναμων **A τάξης δικτύων** (16,777,216 διευθύνσεις το κάθε ένα)
 - Ο συνολικός χώρος των δεσμευμένων διευθύνσεων είναι ισοδύναμος με 20,09 /8 τμήματα
 - Ο **συνολικός** χώρος των διαθέσιμων διευθύνσεων είναι ισοδύναμος με 219,92 /8 τμήματα
- Από αυτές τις **219,92**:
- **77 /8** τμήματα είναι **αδιάθετα** από την IANA
 - **142,92 /8** έχουν ανατεθεί από την IANA σε **RIRs** ή κατευθείαν σε **πελάτες**
- Από αυτές τις 142,92 /8:
- **21,57 /8** είναι **αδιάθετες** από τις RIRs
 - **121,35 /8** έχουν **ανατεθεί** σε τελικούς πελάτες
- Από αυτές τις 121,35:
- **76,91 /8** **ανακοινώνονται** στο διαδίκτυο
 - **42,44 /8** **δεν ανακοινώνονται** στο δίδικτυο

Τρέχουσα (09/04) χρήση του χώρου των IPv4 διευθύνσεων

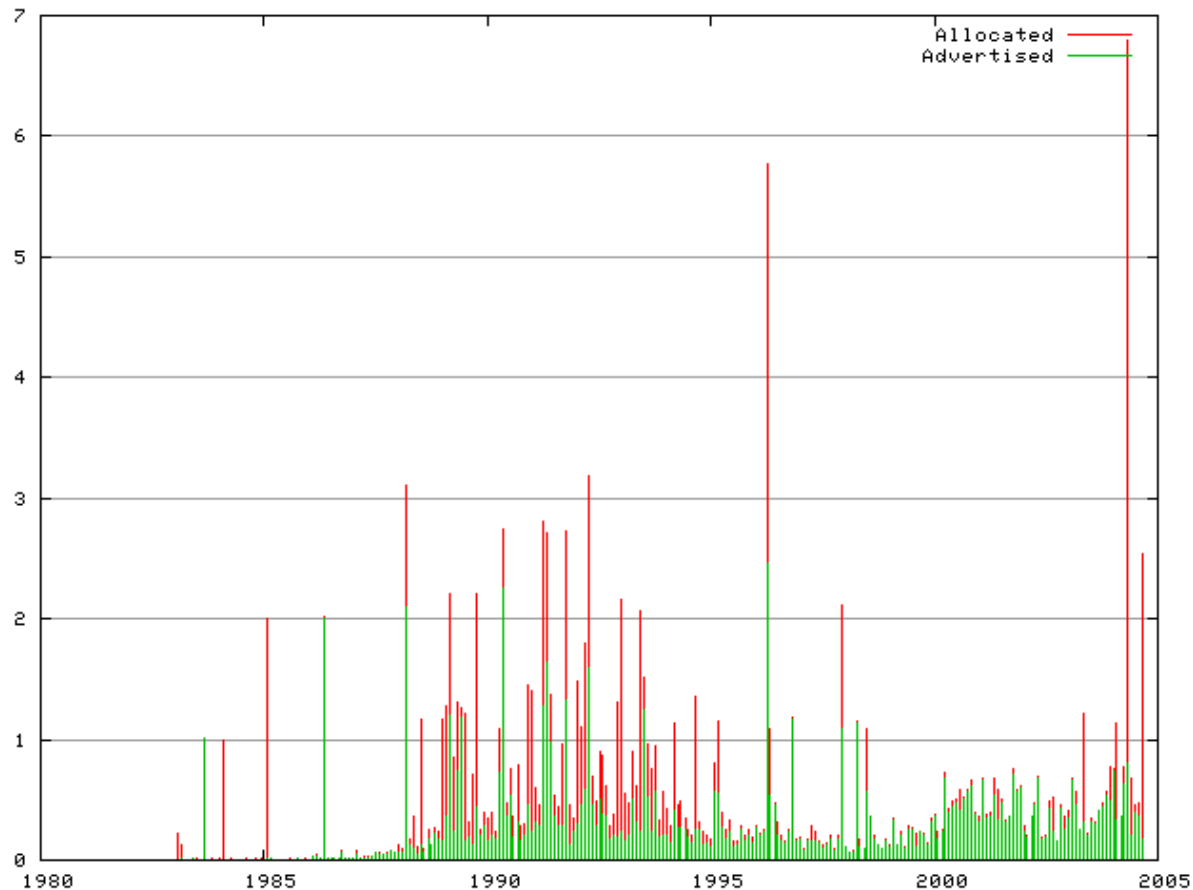


Εξέλιξη του χώρου των IPv4 διευθύνσεων σε χρήση



- Το μετράμε σε /8 δίκτυα
- Πριν το 1995 βλέπουμε μια μεγάλη αυξητική τάση σαν επακόλουθο της ισχυρής ζήτησης σε B δίκτυα
- Από τότε μέχρι σήμερα ο ρυθμός αύξηση έχει μειωθεί σαν αποτέλεσμα της χρήσης του **CIDR**
- Με βάση το μοντέλο αυτό προβλέπεται **εξάντληση** το **2016**

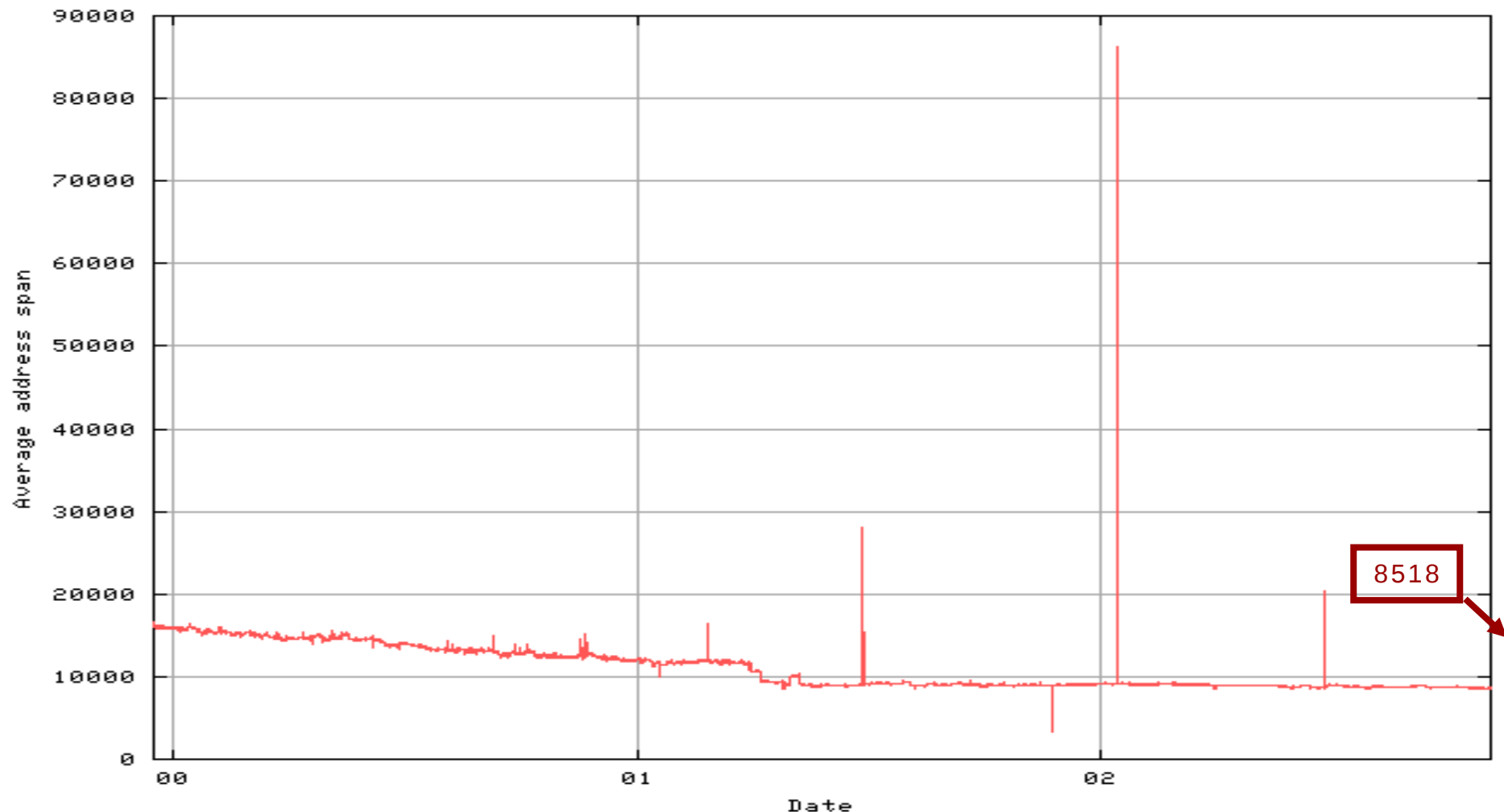
Αποτελεσματικότητα χρήσης IPv4 διευθύνσεων



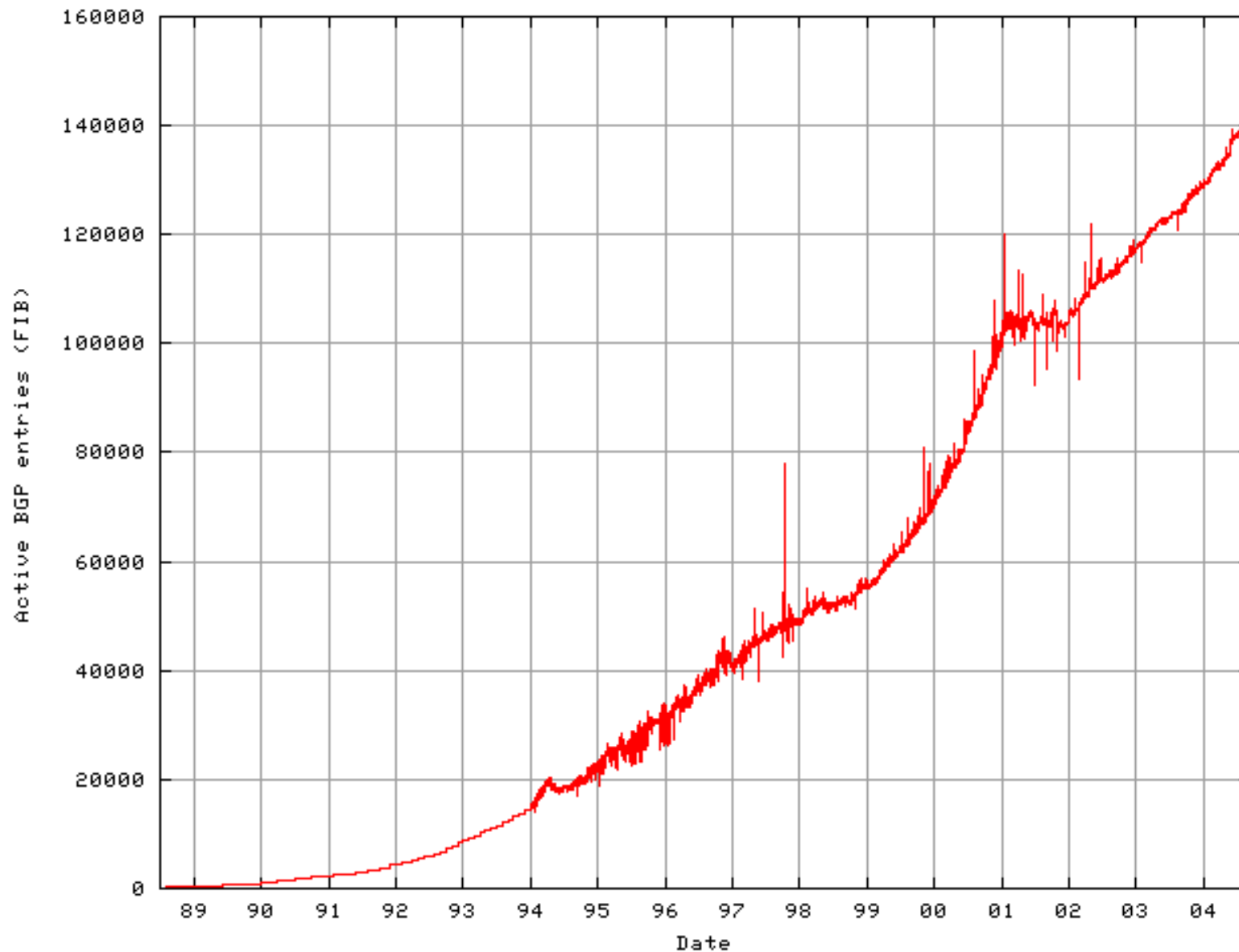
- Ένα μέγεθος που μας δείχνει το πόσο **αποτελεσματικά** ανατίθενται οι διευθύνσεις, είναι το ποσοστό των **ανακοινώσιμων** διευθύνσεων στο διαδίκτυο, προς τις διευθύνσεις που έχουν **διατεθεί**
- Η αποτελεσματικότητα αυτή έχει αυξηθεί σημαντικά από το **1996**

Μέσο εύρος των BGP προθεμάτων

- Όσο μεγαλύτερο το εύρος των προθεμάτων που ανακοινώνονται, τόσο μικρότερο το πλήθος τους και τόσο λιγότερες οι εγγραφές στους πίνακες δρομολόγησης

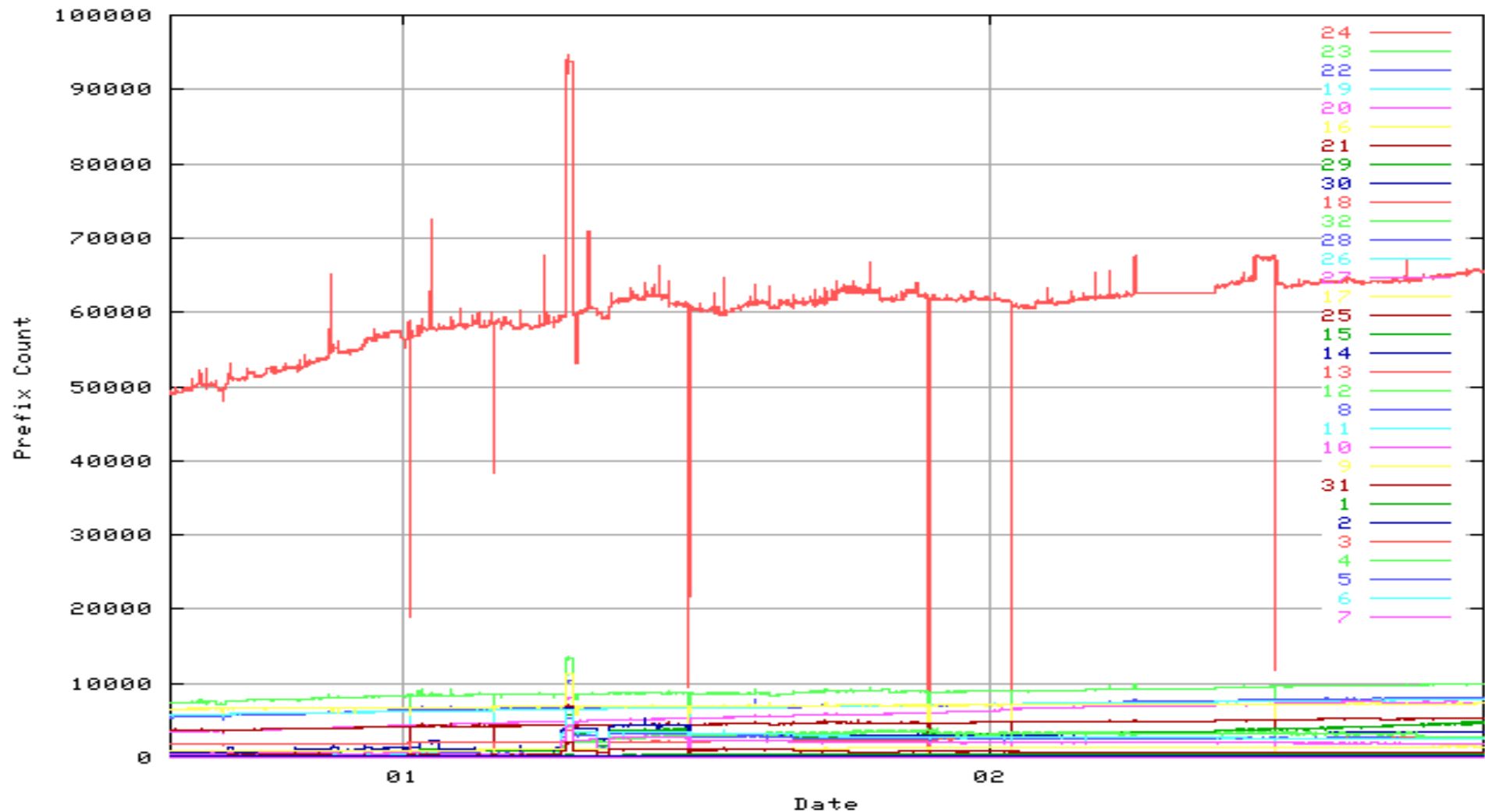


Εγγραφές στον πίνακα FIB ενός BGP δρομολογητή

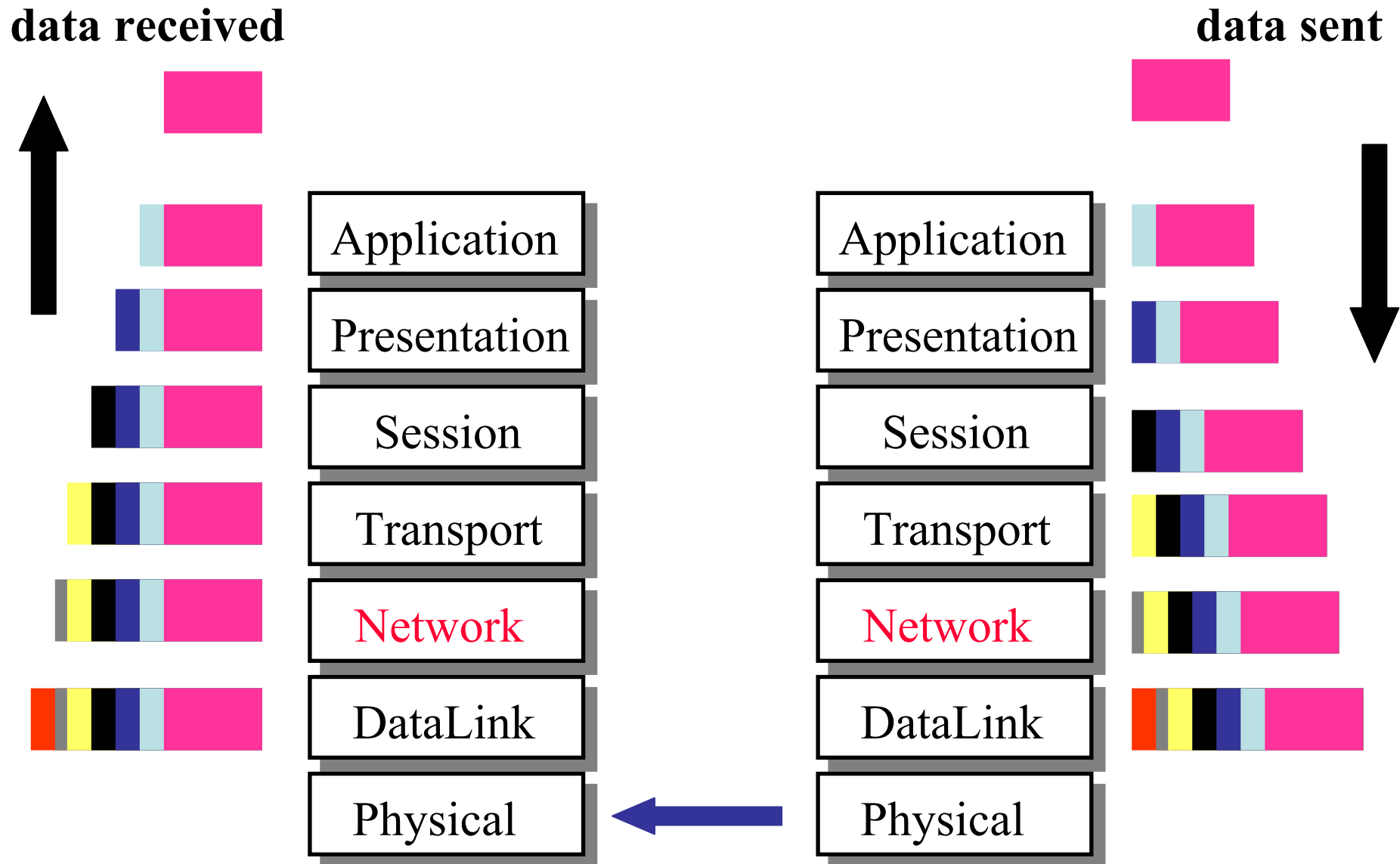


Αριθμός προθεμάτων που ανακοινώνονται στο διαδίκτυο

- Τα περισσότερα προθέματα – δίκτυα που ανακοινώνονται είναι C τάξεως



Μοντέλο αναφοράς OSI (Open Systems Interconnection)



Μοντέλο αναφοράς OSI (Open Systems Interconnection)

- Αναπτύχθηκε από τον οργανισμό **ISO** (International Organization for Standards)
- Στον κόσμο του IP τα τρία πρώτα επίπεδα **ομαδοποιούνται** σε ένα, το **επίπεδο εφαρμογής**
- Κάθε επίπεδο παρέχει **υπηρεσίες** στο παρακάτω και στο παραπάνω

Μεταφοράς (transport)

- Αξιόπιστη μετάδοση από άκρο σε άκρο

Δικτύου (network)

- Χρησιμοποιεί πλαίσια για να στείλει τα πακέτα (datagrams) με κατάλληλες διευθύνσεις δικτύου

Σύνδεσης δεδομένων (data link)

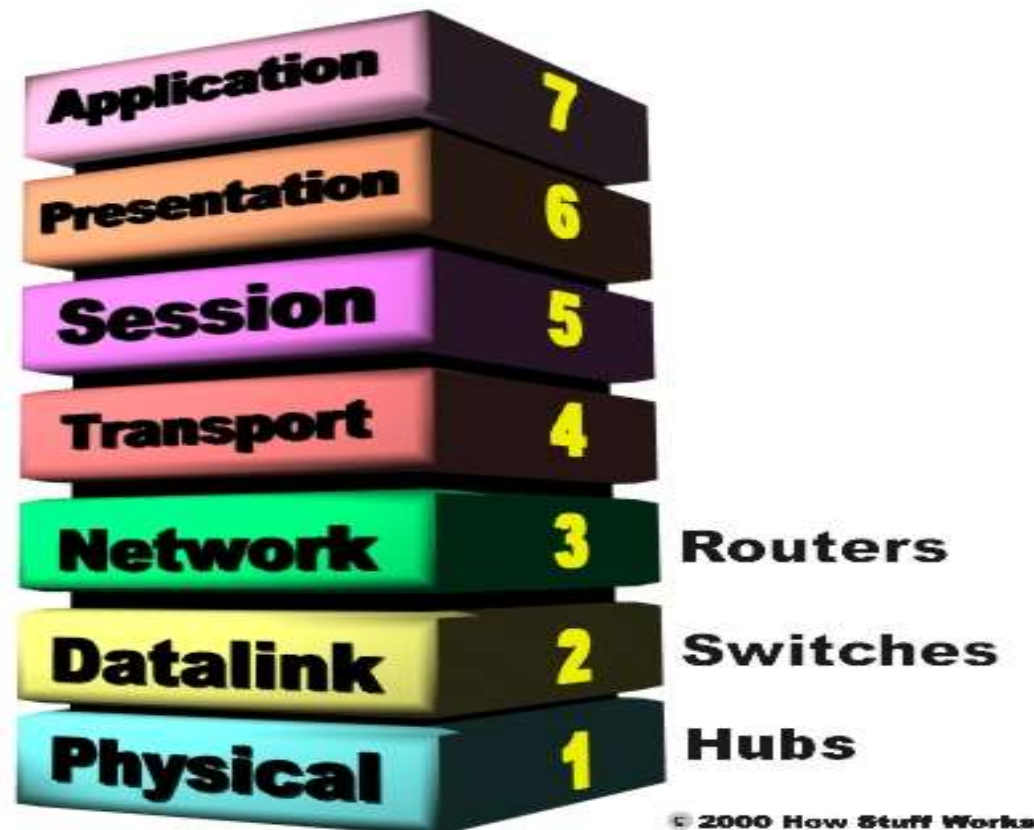
- Πλαισίωση δεδομένων

Φυσικό (physical)

- Μεταφορά της πληροφορίας σε κάποιο φυσικό μέσο

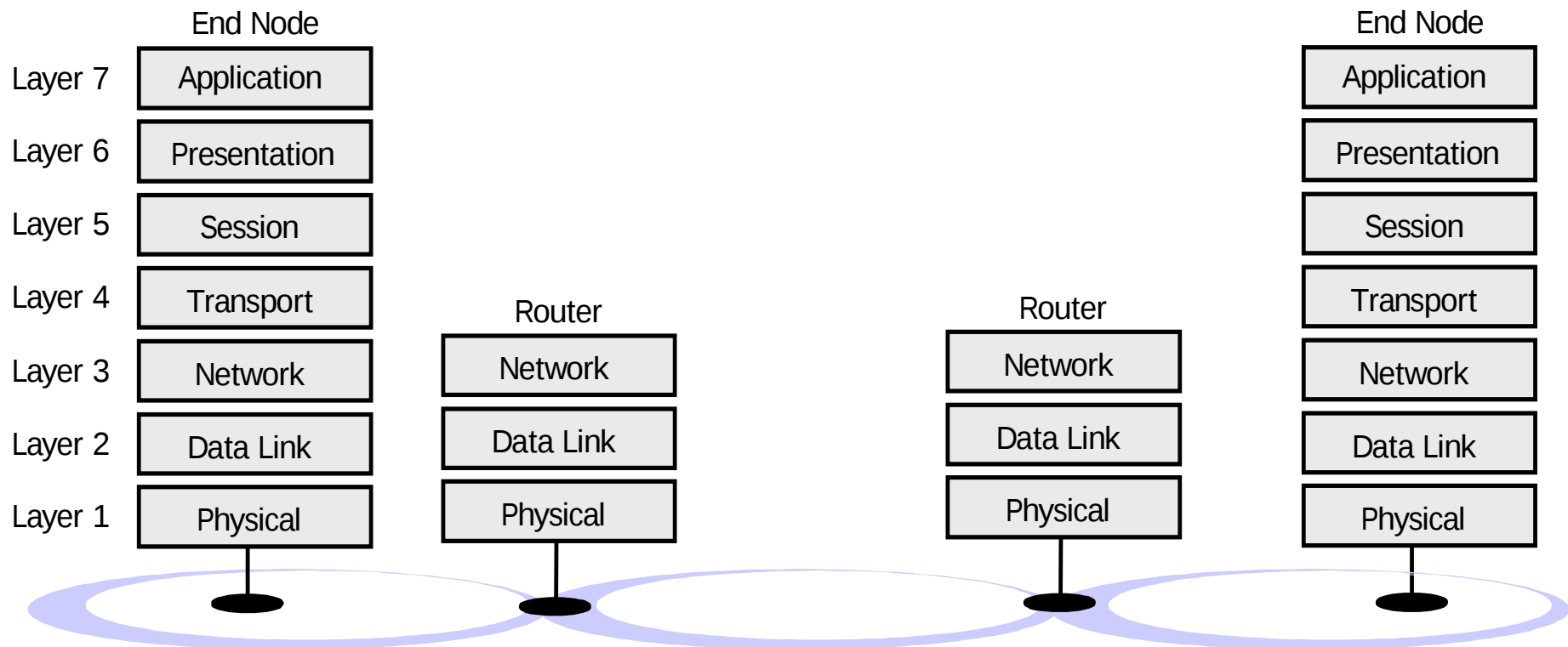
Στοιχεία δικτύου

- Ο δρομολογητής (**router**) δουλεύει μέχρι και το **επίπεδο δικτύου** του OSI
- Ο μεταγωγέας (**switch**) εργάζεται μέχρι και το **επίπεδο σύνδεσης δεδομένων**
- Το **hub** εργάζεται στο **φυσικό επίπεδο**

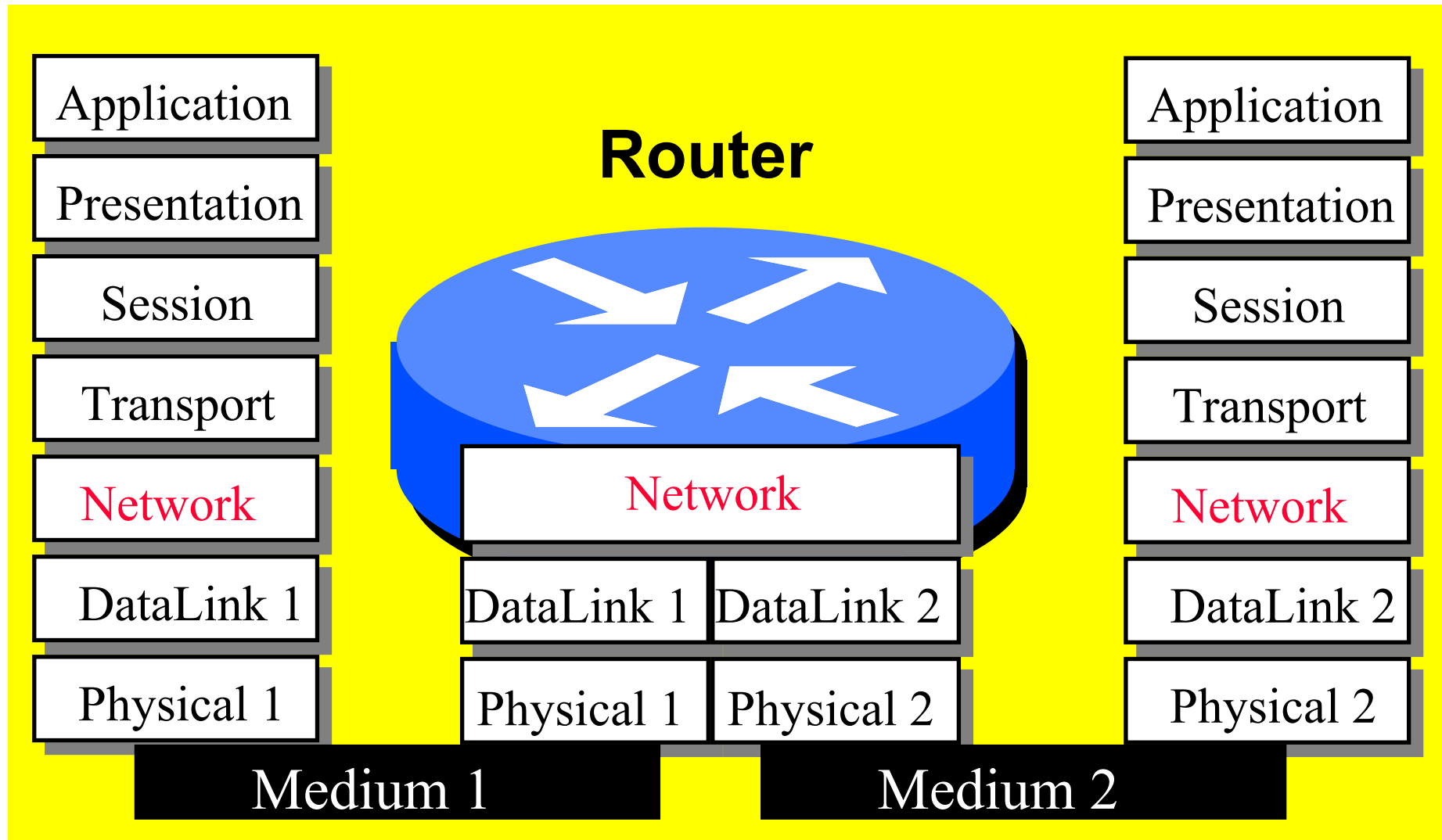


Στοιχεία δικτύου και επίπεδα OSI

- Στα άκρα βρίσκεται ο χρήστης ο οποίος μπορεί να έχει υλοποιημένη όλη τη στίβα των επιπέδων του OSI
- Οι δρομολογητές αναλαμβάνουν την προώθηση της πληροφορίας από άκρο σε άκρο υλοποιούν τα **τρία πρώτα επίπεδα**



Στοιχεία δικτύου και επίπεδα OSI



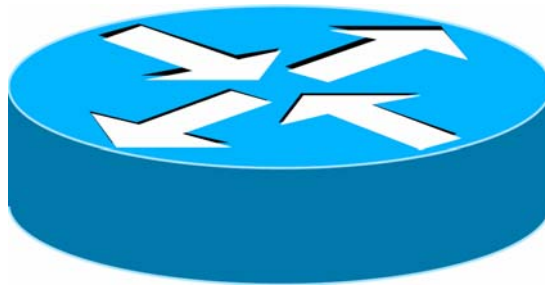
Δρομολογητής (Router)

- Ο δρομολογητής είναι ένα στοιχείο δικτύου το οποίο έχει διάφορες διεπαφές στις οποίες λαμβάνονται και εκπέμπονται IP πακέτα
- Ο δρομολογητής εξετάζει το **κάθε ένα πακέτο** ξεχωριστά
- Προσπαθεί να βρεί τον πιο κατάλληλο δρόμο-διαδρομή για κάθε ένα πακέτο (**Δρομολόγηση – Routing**)
- Προωθεί το πακέτο αυτό προς την κατάλληλη κατεύθυνση (**Προώθηση – Forwarding**)



Δρομολογητής (Router)

- Οι πρώτοι IP δρομολογητές ήταν **UNIX** μηχανάκια με ειδικές διεπαφές
- Σταδιακά, εξεικευμένα προϊόντα με **βελτιστοποιημένα λειτουργικά** συστήματα αναπτύχθηκαν
- Οι δρομολογητές ονομάζοντας στη γλώσσα του OSI, **Intermediate Systems**, ενώ μερικές φορές αποκαλούνται **Gateways**



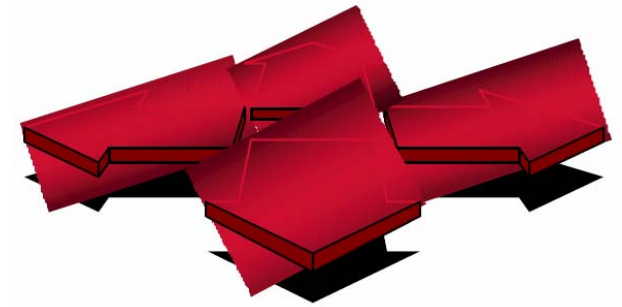
Δρομολόγηση (Routing)



- Εύρεση **διαδρομών – μοναπατιών**
- Βασίζεται σε πληροφορία που ελήφθει από ένα πρωτόκολλο δρομολόγησης
- Περισσότερες της μιας **εναλλακτικές διαδρομές** μπορεί να υπάρχουν
- Οι αλλαγές συμβαίνουν **περιοδικά** ή/και όταν **αλλάζει η τοπολογία**
- Οι αποφάσεις στηρίζονται σε **κριτήρια** όπως η τοπολογία, η πολιτική δρομολόγησης, οι μετρικές (αριθμός hop, καθυστέρηση, εύρος..)
- Σκοπός είναι η **βελτιστοποίηση κάποιας παραμέτρου**, δεδομένου του δικτύου
- Χρησιμοποιείται πληροφορία από όλο το δίκτυο για να δημιουργηθούν **δομές δεδομένων** που θα χρησιμοποιηθούν για την **προώθηση** της πληροφορίας
- Πλήρης **αυτοματοποίηση** δεν είναι δυνατή, **μερική** επιτυγχάνεται μέσω **δυναμικών** πρωτόκολλων δρομολόγησης
- Συχνά απαιτεί **πολύπλοκη διάρθρωση**

Προώθηση (Forwarding)

- **Προώθηση** των πακέτων στις κατάλληλες θύρες ακολουθώντας τις εντολές δρομολόγησης
- Αφορά τη χρήση **τοπικής** πληροφορίας (πίνακες, δομές δεδομένων) ώστε να αποφασιστεί η μεταχείριση στην κίνηση
 - Ταξινόμηση της κίνησης (traffic classification)
 - Αντιμετώπιση βασισμένη στην ταξινόμηση
 - Τυχόν **απόρριψη** της κίνησης
 - Που και πως θα προωθηθεί η κίνηση
- Ο πίνακας προώθησης (forwarding table) δημιουργείται από τους **αλγόριθμους δρομολόγησης**
- Η **προώθηση** των πακέτων στις διεπαφές του δρομολογητή γίνεται με βάση:
 - Τη **διεύθυνση προορισμού** του πακέτου
 - Την **τάξη υπηρεσίας** (fair queuing, precedence, ..)
 - Τοπικές **απαιτήσεις** (φιλτράρισμα πακέτων)
- Η προώθηση μπορεί να γίνει με ειδικό **υλικό** το οποίο να επιταχύνει τη διαδικασία



Εγγραφές του Πίνακα προώθησης

Ενημερώνονται:

Στατικά

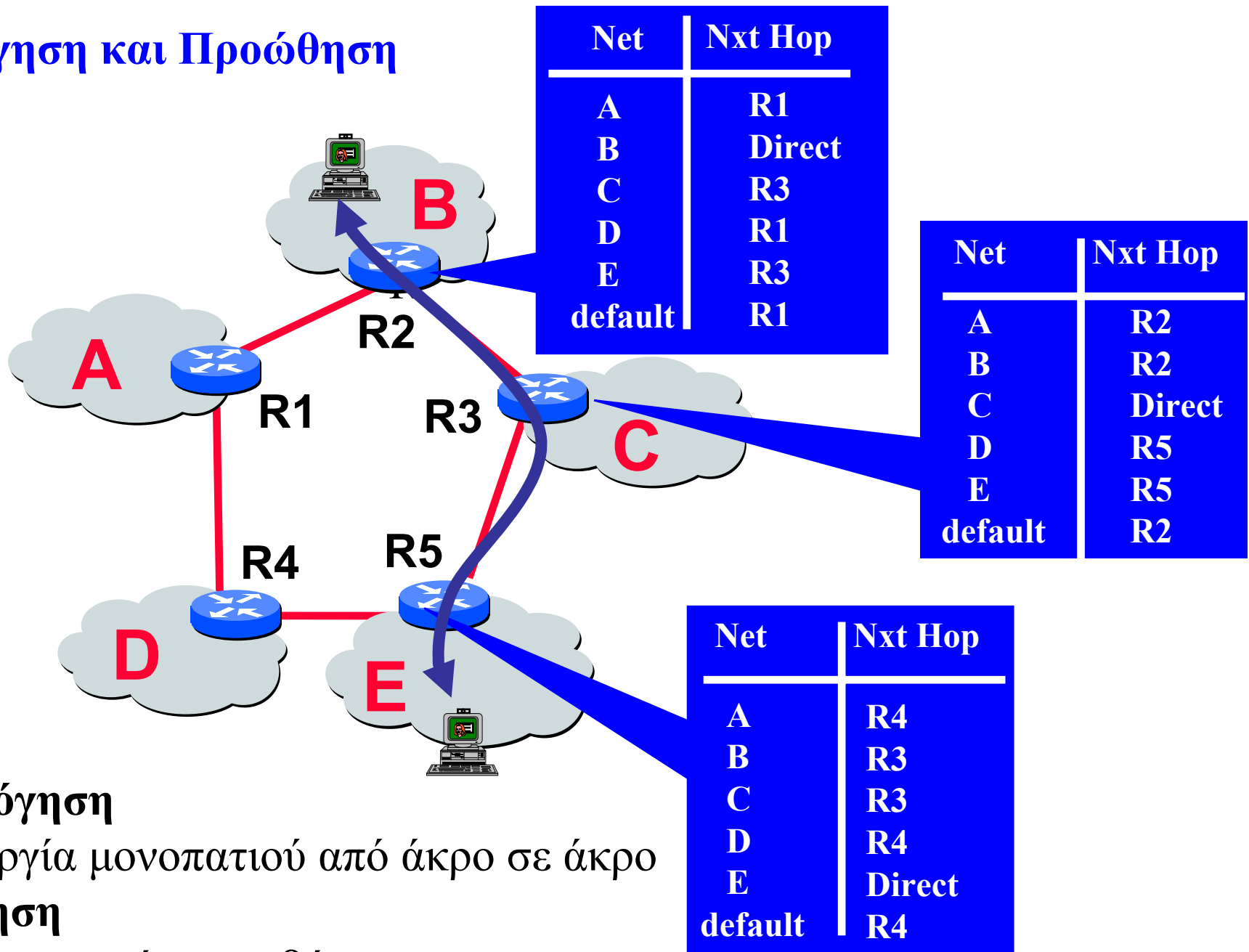
- Ο διαχειριστής εισάγει χειροκίνητα κάποιες εγγραφές
- Περισσότερος έλεγχος
- Δεν περιορίζεται στην προώθηση με βάση τον προορισμό
- Δεν επιτρέπει κλιμάκωση
- Αργή προσαρμογή στις βλάβες – μεταβολές του δικτύου

Δυναμικά

- Οι δρομολογητές ανταλλάσσουν πληροφορία σχετικά με τα δίκτυα που μπορούν να προσεγγίσουν, χρησιμοποιώντας τα πρωτόκολλα δρομολόγησης
- Στη συνέχεια υπολογίζουν την καλύτερη διαδρομή
- Γρήγορη προσαρμογή στις αλλαγές της τοπολογίας
- Εύκολη κλιμάκωση
- Πολυπλοκότητα, κατανάλωση πόρων όπως CPU, μνήμη, εύρος
- Δύσκολη ανίχνευση και διόρθωση προβλημάτων
- Τα τρέχοντα πρωτόκολλα βασίζονται στην διεύθυνση προορισμού

Στην πράξη χρησιμοποιούνται και οι δύο τρόποι, με τη στατική δρομολόγηση να κυριαρχεί στα **άκρα** του δικτύου

Δρομολόγηση και Προώθηση



Δρομολόγηση

- Δημιουργία μονοπατιού από άκρο σε άκρο

Προώθηση

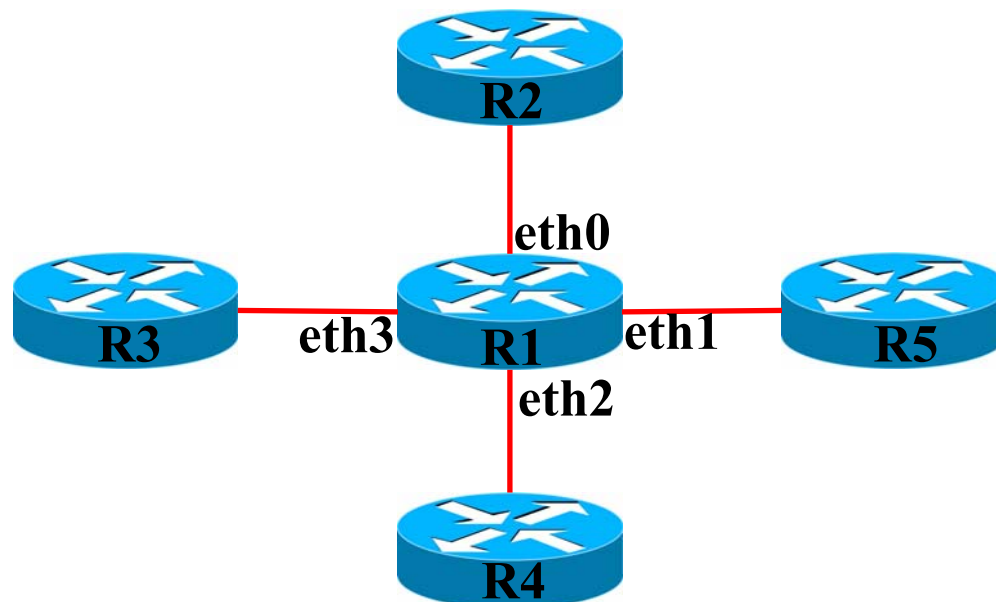
- Εύρεση του επόμενου βήματος

Αναζήτηση στον πίνακα δρομολόγησης

- Η δρομολόγηση βασίζεται στην **IP** διεύθυνση του **προορισμού**
- Ισχύει η αρχή του “**longest match**” δηλαδή θα προτιμηθεί η διαδρομή η οποία έχει το πιο **καλό ταιρίασμα** με τον προορισμό
- Ελέγχεται **ανά πακέτο** αν υπάρχουν εγγραφές που να ταιριάζουν στον προορισμό
- Από αυτές που ταιριάζουν θα προτιμηθεί αυτή που έχει τα περισσότερα bits κοινά
- Επειδή η αναζήτηση γίνεται ανά πακέτο, αν ο πίνακας προώθησης είναι μεγάλος έχουμε μεγαλύτερο **επεξεργαστικό φόρτο**
- Η αυξημένη υπολογιστική ισχύς είναι το βασικό πρόβλημα **κλιμάκωσης** στα δίκτυα
- Η λύση είναι η εύρεση νέων πρωτοκόλλων, τρόπων δρομολόγησης, για παράδειγμα το **MPLS**

Αναζήτηση στον πίνακα δρομολόγησης

- Έστω IP πακέτο με προορισμό 10.26.122.10 φτάνει στον R1
- Ο δρομολογητής ανακαλύπτει ότι πρώτη εγγραφή του έχει 0 bits κοινά, η δεύτερη 8, η τρίτη 24, και η τέταρτη 27
- Άρα το πακέτο θα προωθηθεί προς τη θύρα eth3



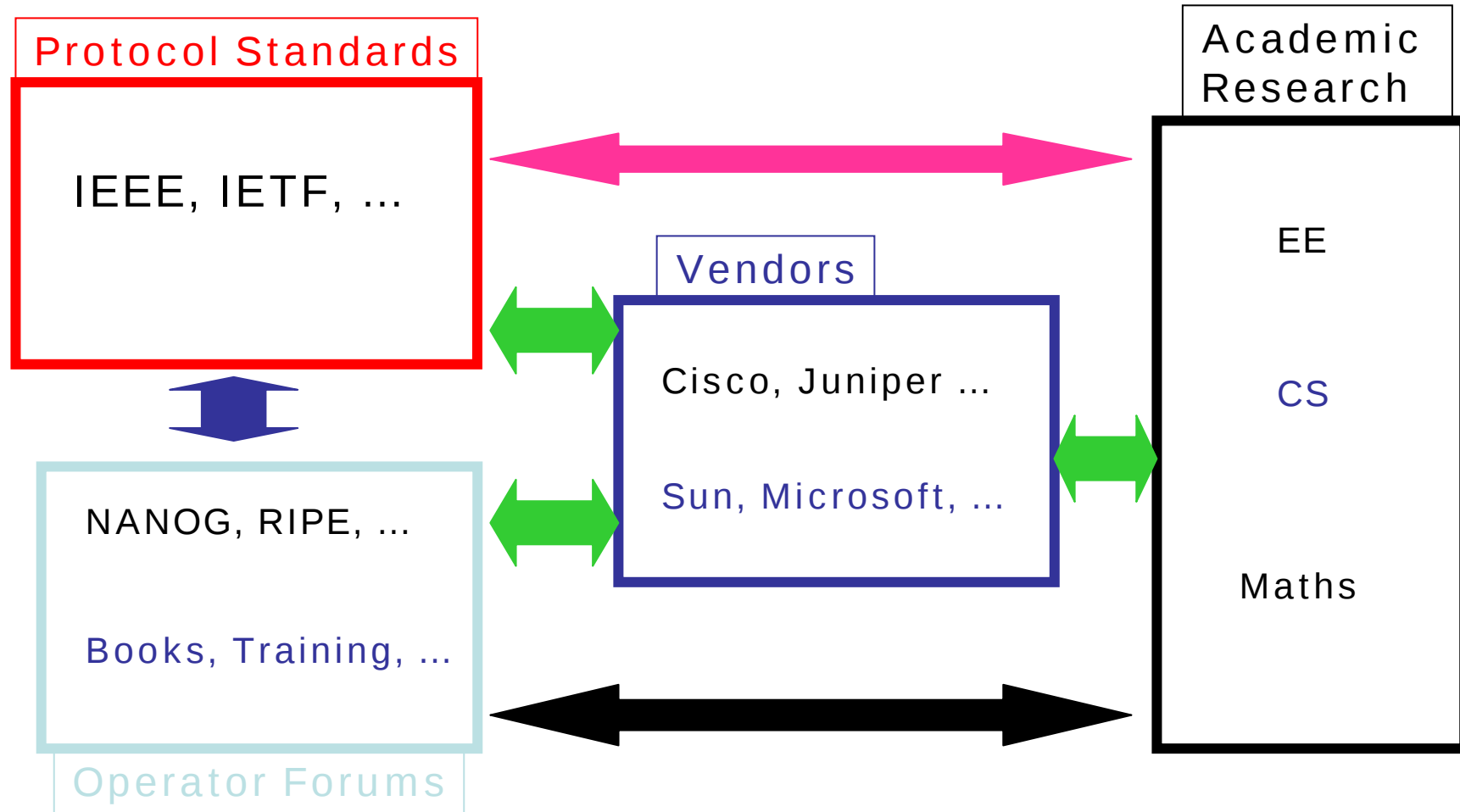
Πρόθεμα	Διεπαφή
0.0.0.0/0	eth0
10.0.0.0/8	eth1
10.26.122.0/24	eth2
10.26.122.0/27	eth3

Σύγκριση προορισμού με πρόθεμα

	00001010	00011010	01111010	00000000
10.26.122.0/27	11111111	11111111	11111111	11100000
10.26.122.34	00001010	00011010	01111010	00 1 000010

- Το 10.26.122.34 δεν ανήκει στο 10.26.122.0/27
- Η εργασία αυτή πρέπει να γίνει για κάθε πακέτο ξεχωριστά, κατά συνέπεια σε δρομολογητές με **πολλές χιλιάδες εγγραφές** και με διεπαφές όπου οι **ρυθμοί μετάδοσης** είναι μεγάλοι, συνεπάγεται ότι οι απαιτήσεις σε **CPU** και **μνήμη** είναι αντίστοιχα μεγάλες
- Τούτο εξηγεί και το μεγάλο **κόστος** παρόμοιων δρομολογητών

Ποιοι εξελίσσουν την δρομολόγηση;

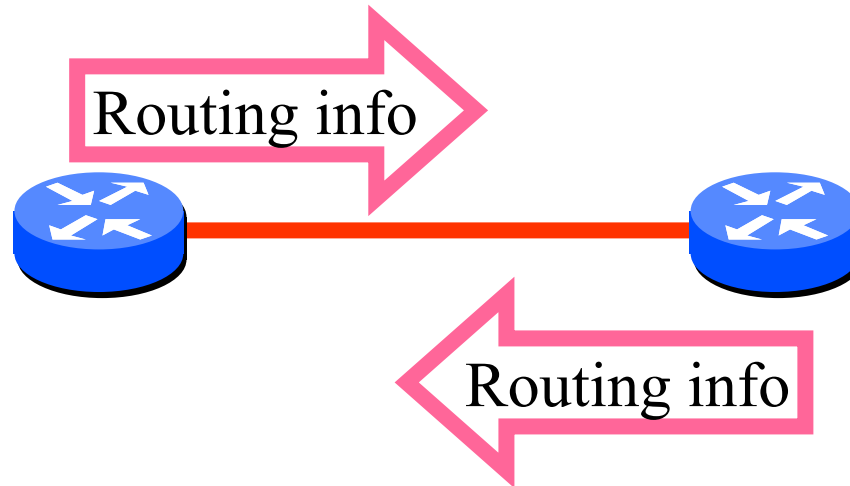


- Εμπλέκονται οργανισμοί προτυποποίησης, πάροχοι υπηρεσιών, πωλητές εξοπλισμού, ακαδημαϊκά ιδρύματα έρευνας

Προκλήσεις

- Τα δίκτυα είναι **πολυπληθή** και **ετερογενή** αποτελούμενα από διαφορετικές τεχνολογίες με διαφορετικές **τακτικές δρομολόγησης**
- Η IP δρομολόγηση περιλαμβάνει διαρκώς **καινούργια πρωτόκολλα** και επεκτάσεις τους, καινούργιους κόμβους, και νέους τρόπους χρησιμοποίησης των υπαρχόντων τεχνολογιών

Επικοινωνία Δρομολογητών

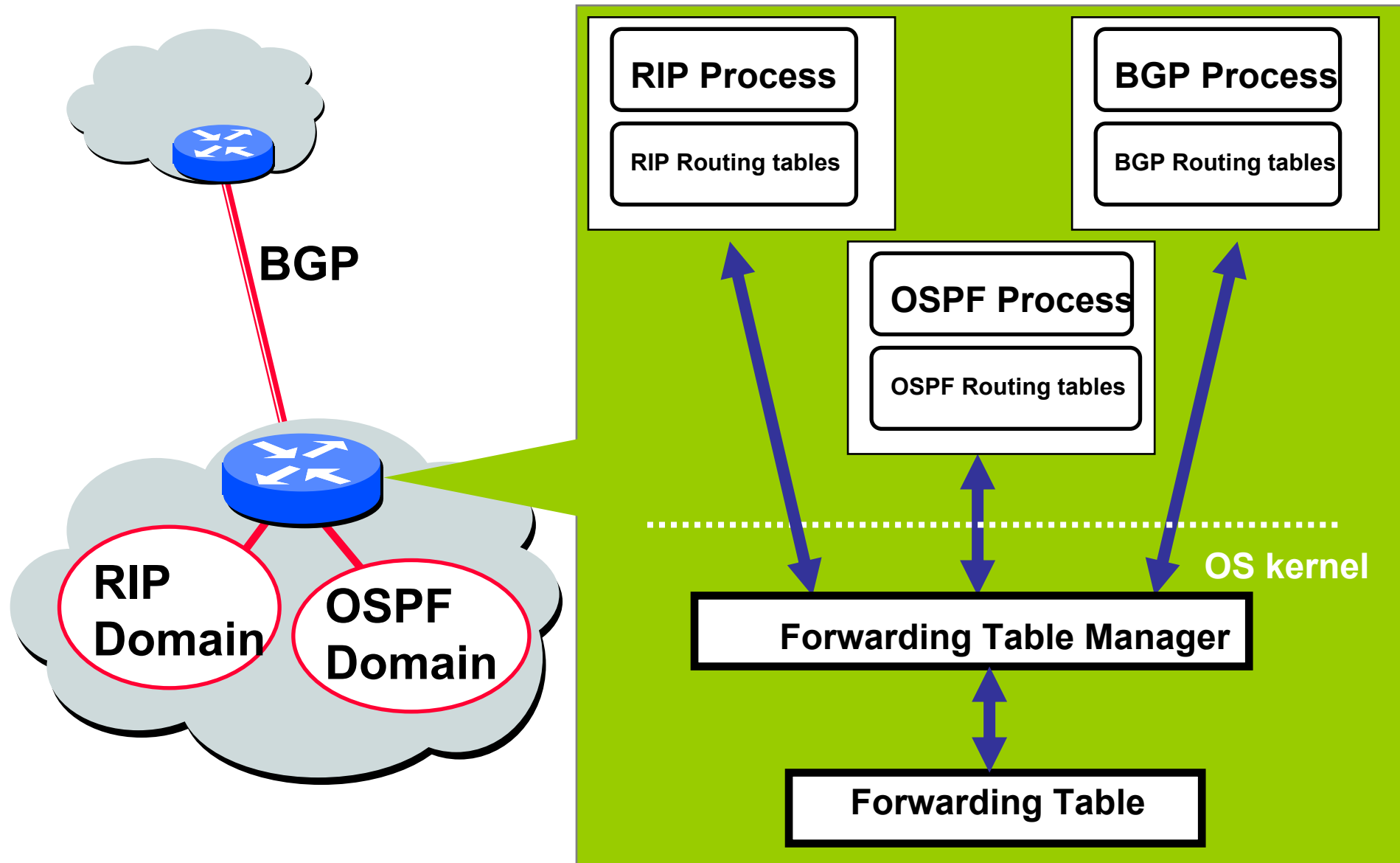


- Ο υπολογισμός των διαδρομών γίνεται ανάμεσα σε δρομολογητές εντός του ίδιου κράτους δρομολόγησης (**routing domain**)
- Ο υπολογισμός του **επόμενου βήματος** είναι η διαδικασία η οποία απαιτεί τους περισσότερους πόρους σε ένα δρομολογητή
- Τα μηνύματα δρομολόγησης συνήθως **δεν δρομολογούνται**, αλλά ανταλλάσσονται μέσω διαδικασιών δευτέρου επιπέδου και ανάμεσα σε γειτονικούς δρομολογητές (IBGP και multi-hop EBGP είναι εξαιρέσεις)

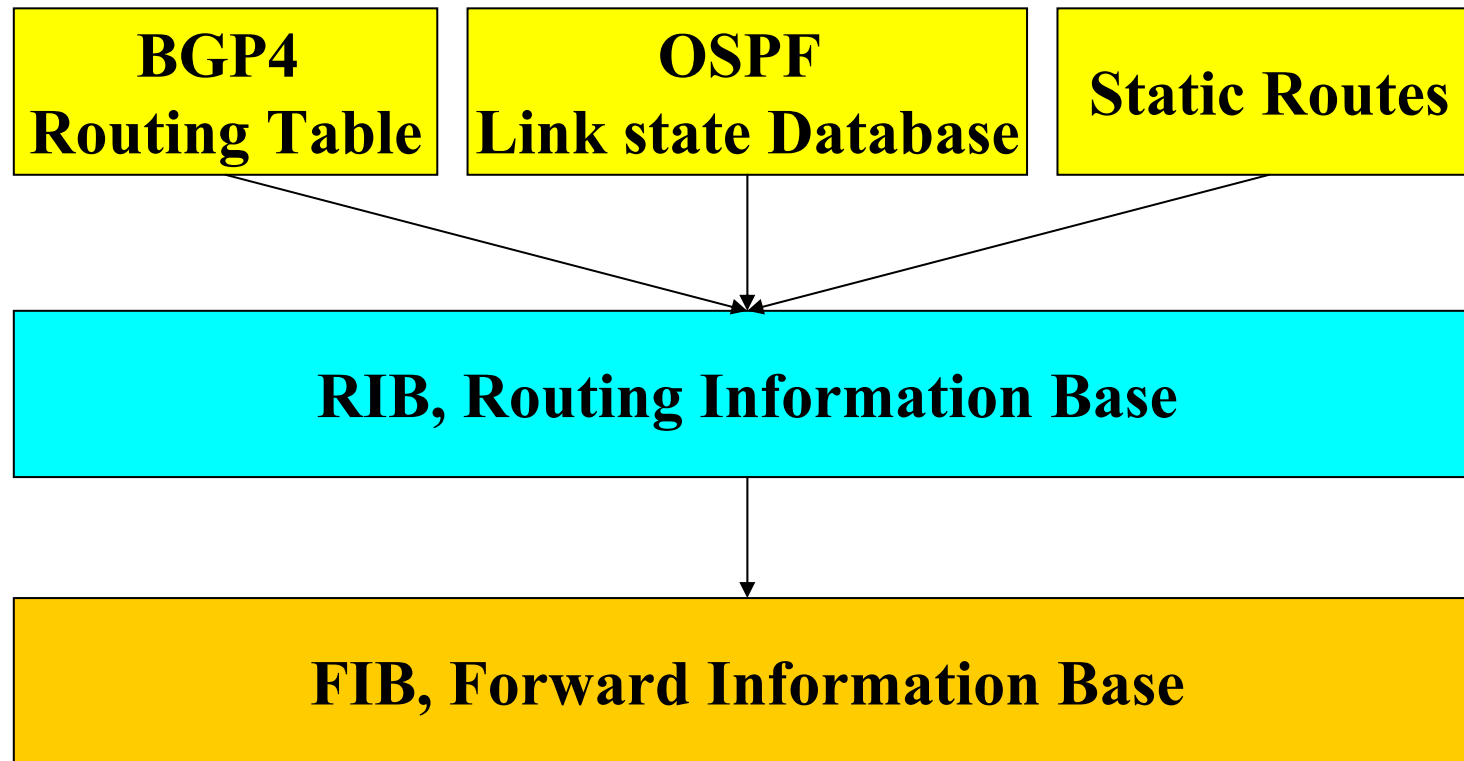
Συμφόρηση δικτύου

- Είναι η κατάσταση στην οποία ένας δρομολογητής δέχεται κίνηση την οποία δεν μπορεί να εξυπηρετήσει για διαφόρους λόγους με αποτέλεσμα να **απορρίπτει πακέτα**
- Δημιουργείται ως αποτέλεσμα της μορφής που έχει η κίνηση (**bursty**)
- Τα IP πρωτόκολλα δρομολόγησης δεν μπορούν να **ανιχνεύσουν** και ούτε να **αναδρομολογήσουν** την κίνηση σε μια τέτοια περίπτωση
- Ο λόγος είναι ότι οι αλλαγές στην δρομολόγηση είναι **πολύ πιο αργές** από τις μεταβολές στο φόρτο και ότι η προσπάθεια για αλλαγές στη δρομολόγηση για αντιστάθμιση των μεταβολών του φόρτου θα οδηγούν σε ένα έντονα **ασταθές** και **ταλαντούμενο** σύστημα

Σε ένα μόνο δρομολογητή μπορούν να τρέχουν πολλά πρωτόκολλα δρομολόγησης



Οι πίνακες δρομολόγησης τροφοδοτούν τον πίνακα προώθησης



- Το κάθε πρωτόκολλο έχει το **δικό** του πίνακα δρομολόγησης
- Μία **κεντρική** διαδικασία επιλέγει εγγραφές από κάθε πίνακα και ενημερώνει τις βάσεις **RIB, FIB** του **πυρήνα**

Αυτόνομο κράτος δρομολόγησης – Autonomous Routing Domain, ARD

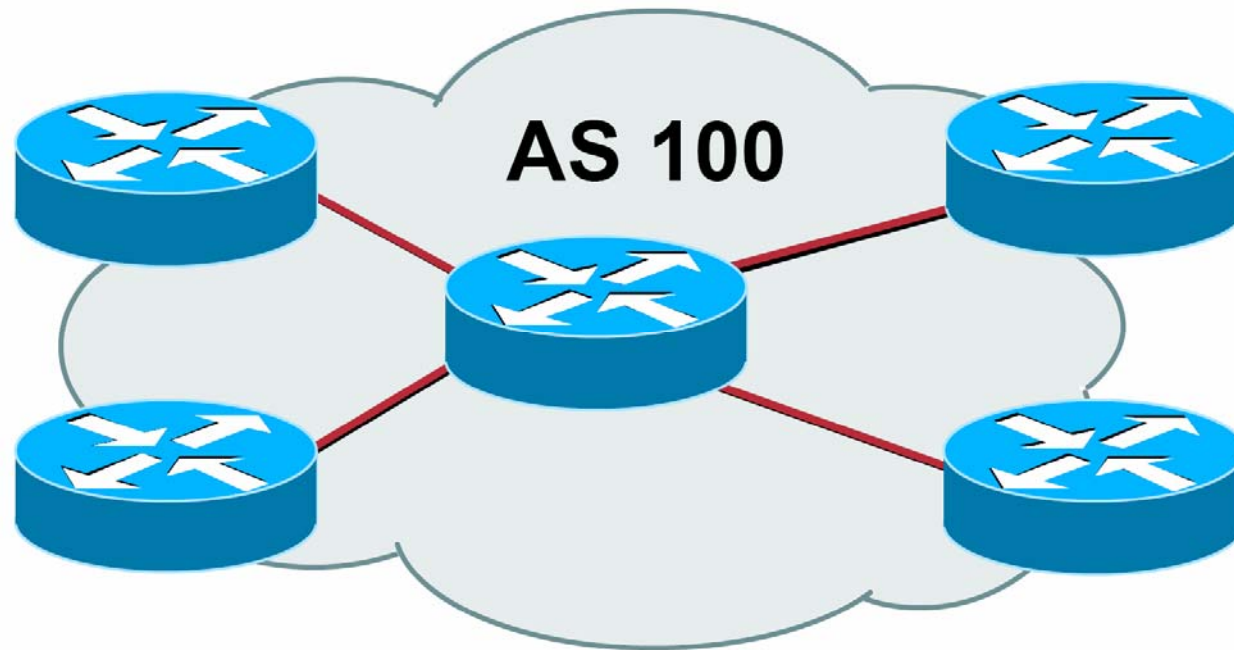
- Είναι μια συλλογή από φυσικά δίκτυα ενωμένα, τα οποία έχουν **ενοποιημένη τακτική δρομολόγησης**

Παράδειγμα

- Campus
- Εταιρικά δίκτυα
- Εσωτερικά δίκτυα των ISP
-

Αυτόνομο σύστημα - AS, Autonomous system

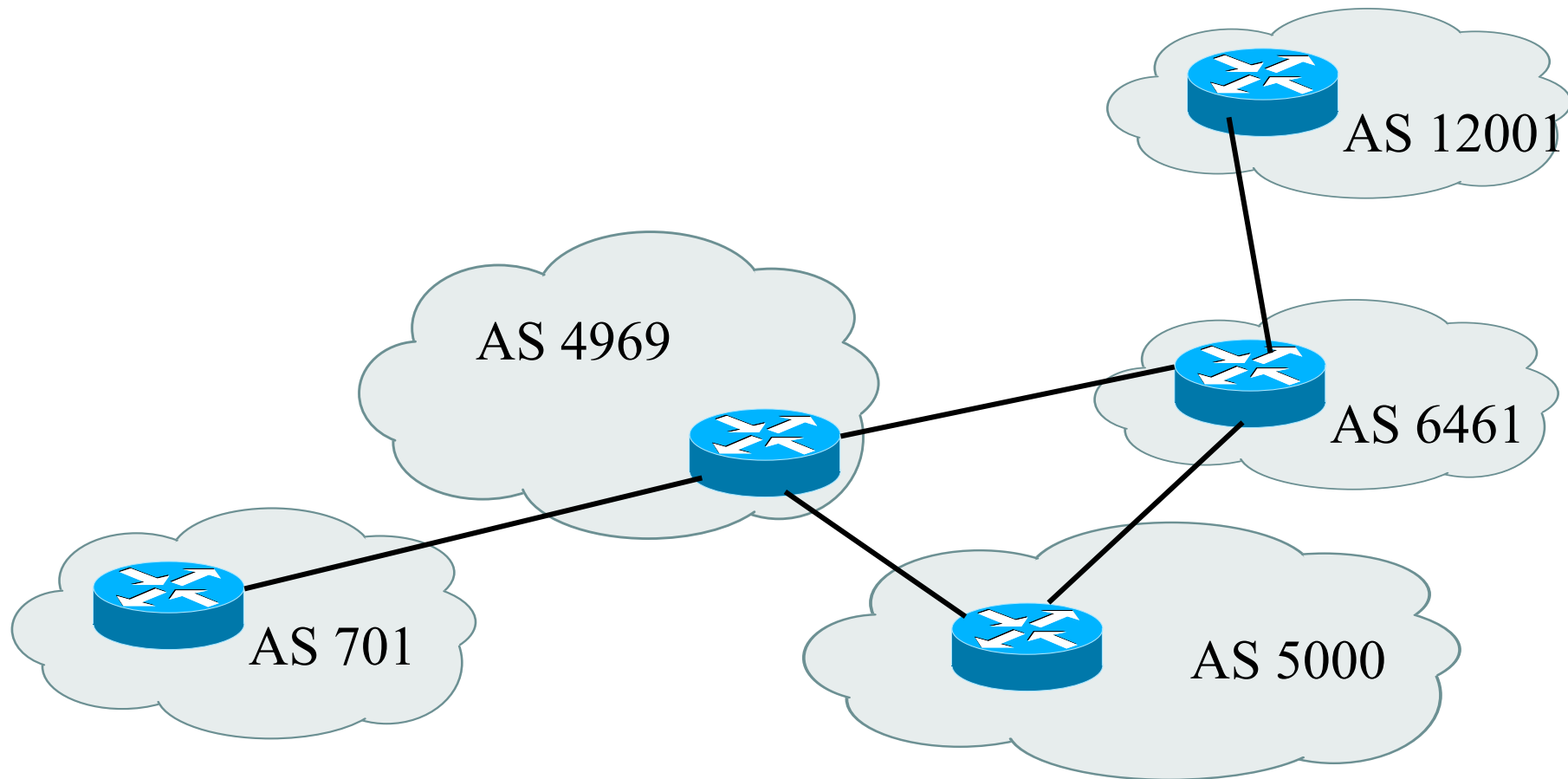
- Πρόκειται για μια συλλογή από δίκτυα συνήθως κάτω από κοινή διαχειριστική αρχή και με την ίδια πολιτική δρομολόγησης
- Ένα ή περισσότερα πρωτόκολλα δρομολόγησης IGP εφαρμόζονται
- Προσδιορίζεται από ένα μοναδικό αριθμό



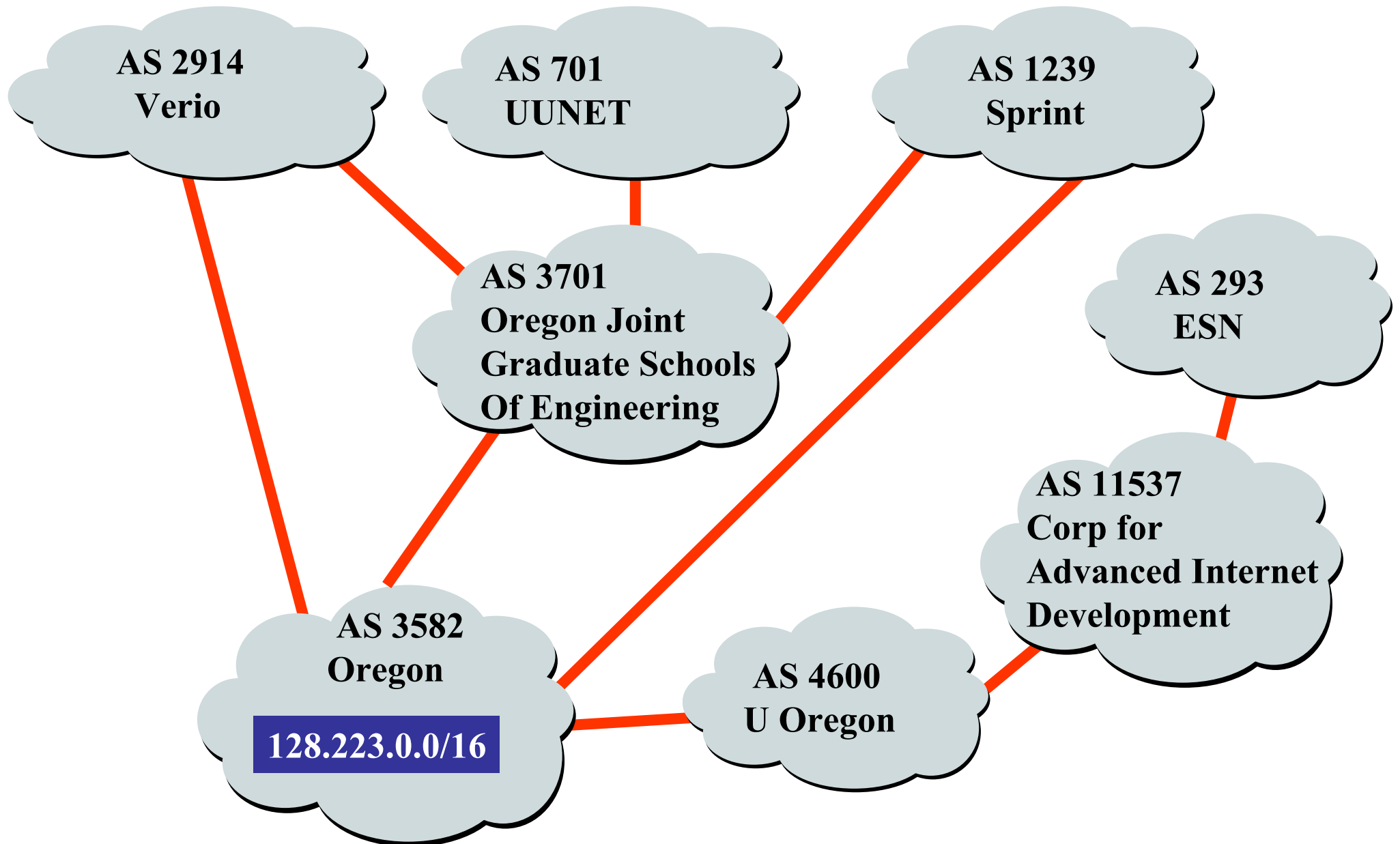
Αυτόνομο σύστημα – Autonomous System, AS

- Ένα **AS** είναι ένα **ARD** στο οποίο έχει ανατεθεί ένας αριθμός Αυτόνομου Συστήματος, **ASN**
- Η διαχείριση ενός AS μοιάζει στα άλλα AS να έχει ένα συνεχή εσωτερικό **σχέδιο δρομολόγησης** και παρουσιάζει μια **συνεπή** εικόνα σχετικά με το ποια δίκτυα είναι προσεγγίσιμα μέσω αυτού
- **RFC1930**: Οδηγίες για τη **δημιουργία, επιλογή και εγγραφή** ενός αυτόνομου συστήματος

Το διαδίκτυο αποτελείται από AS διασυνεδεμένα μεταξύ τους



Παράδειγμα: Σύνδεση του Oregon's AS



Αριθμός Αυτόνομου Συστήματος - ASN, Autonomous System Number

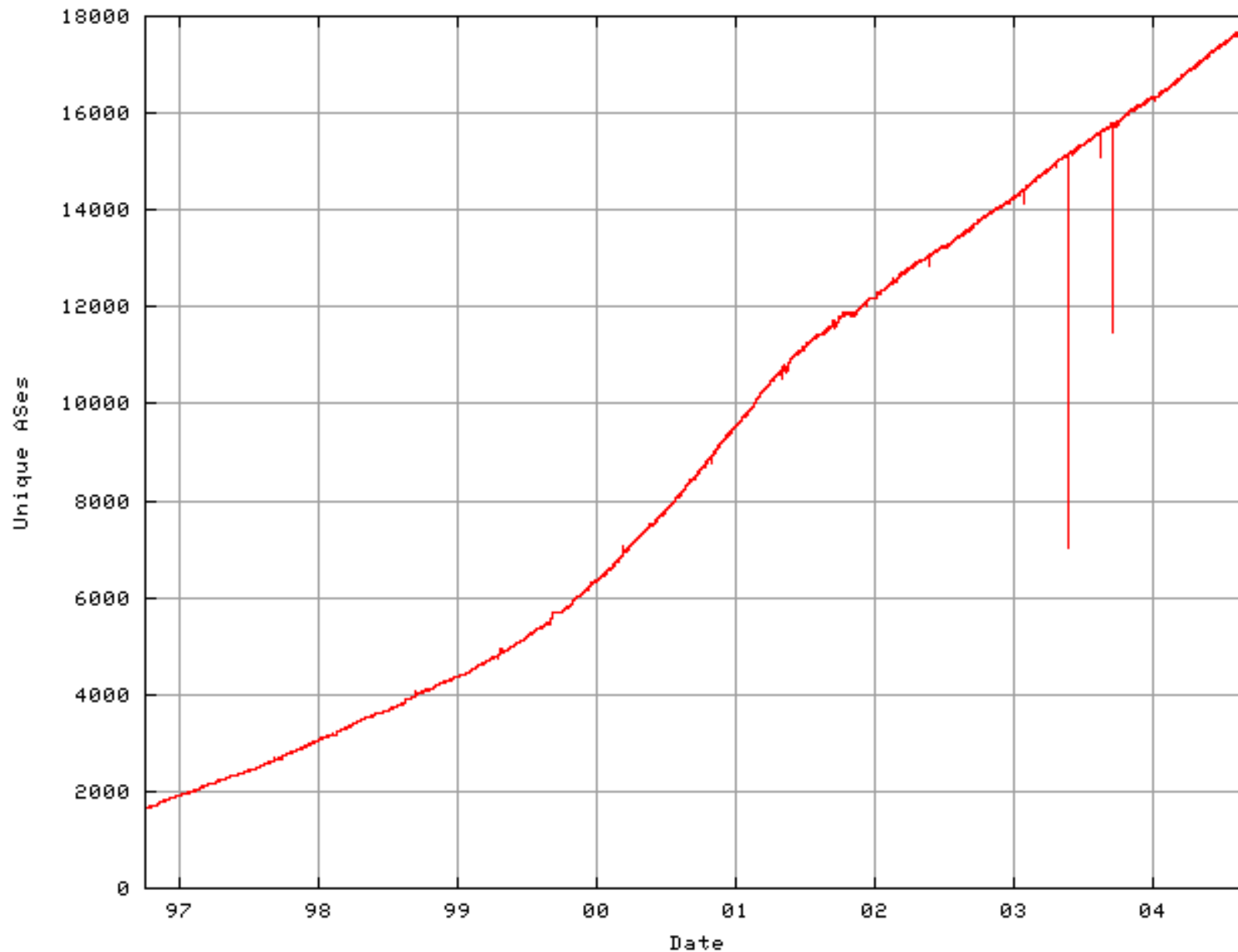
- Ένας **16bit** αριθμός μοναδικός για κάθε AS
- Η ανάθεση και ο έλεγχος γίνεται από την **IANA**, Internet Assigned Numbers Authority (<http://www.iana.org/>)
- Οι τιμές **1-64511** ανατίθενται από τους **RIRs**, Regional Internet Registries
- Επίσης μπορεί να είναι διαθέσιμες από **ISP** που αποτελούν μέλη μιας **RIP**
- Οι τιμές **64512-65534** είναι για **ιδιωτική** χρήση και δεν πρέπει **ποτέ** να **εμφανίζονται** στο διαδίκτυο
- Οι τιμές 0 και 65535 είναι **δεσμευμένες**
- Σήμερα έχουν ανατεθεί περίπου **33000** αριθμοί στους RIR και περίπου **16000** είναι σε χρήση
- Ένας αριθμός ASN υπονοεί μια **ενιαία τακτική δρομολόγησης**

Αριθμός Αυτόνομου συστήματος - ASN, Autonomous System Number

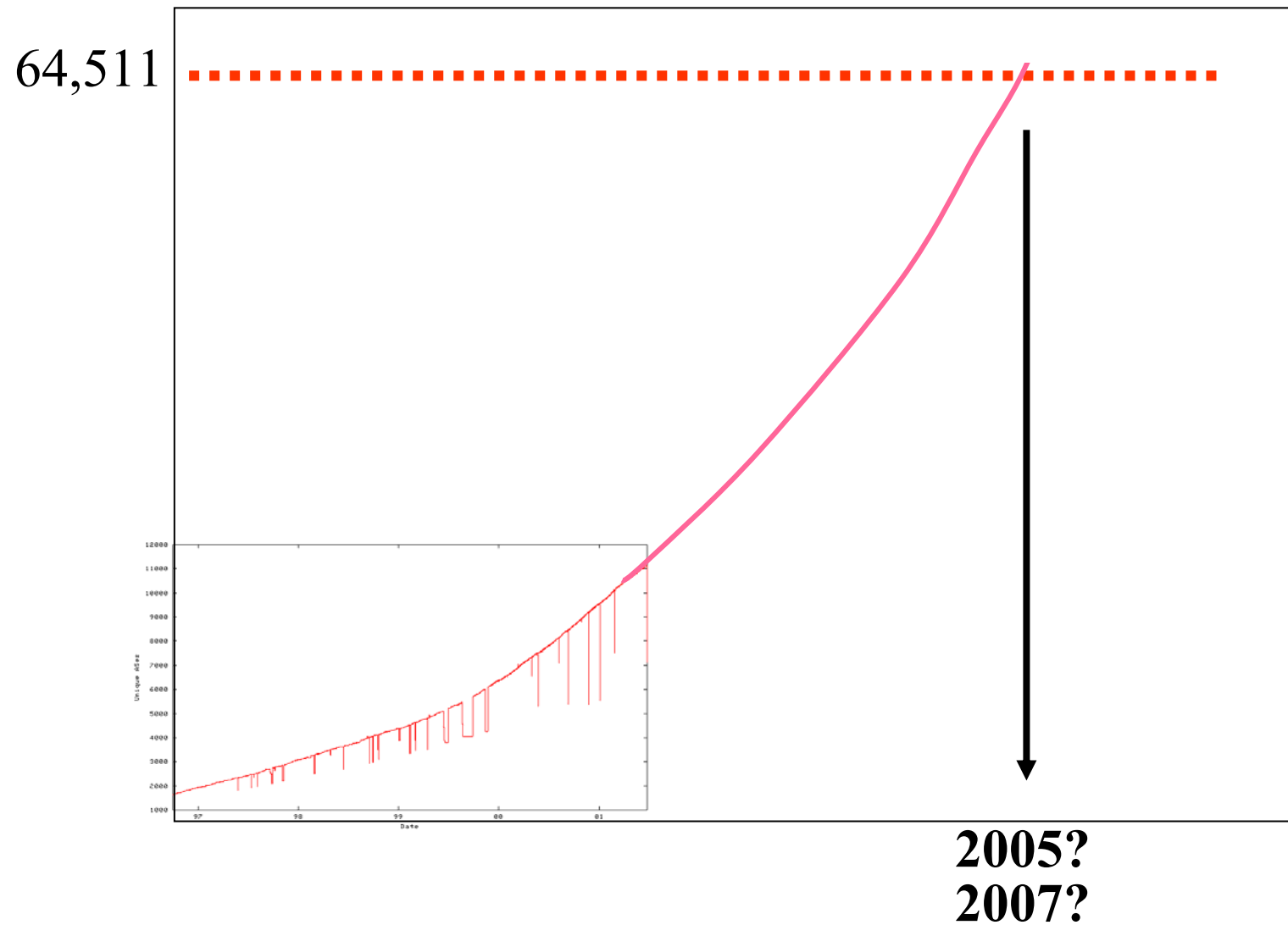
Παραδείγματα

- Genuity: 1
- MIT: 3
- JANET: 786
- UC San Diego: 7377
- AT&T: 7018, 6341, 5074, ...
- UUNET: 701, 702, 284, 12199, ...
- Sprint: 1239, 1240, 6211, 6242, ...

Πόσα Αυτόνομα Συστήματα υπάρχουν;



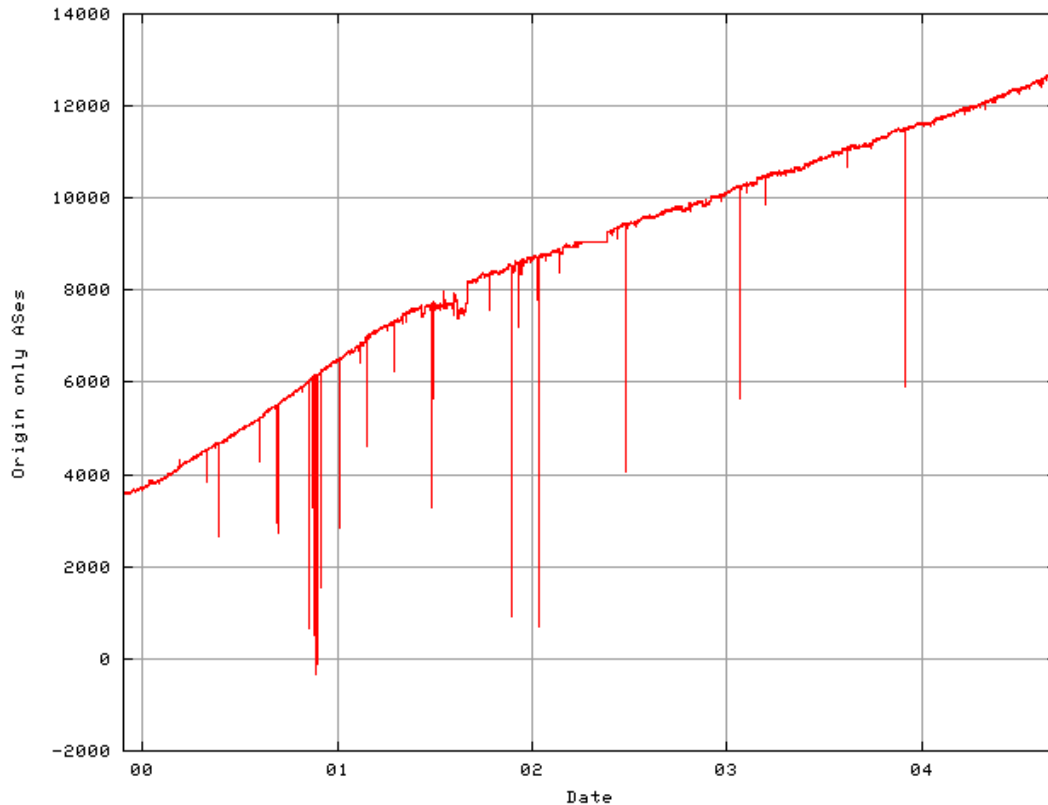
Πότε θα τελειώσουν οι αριθμοί ASN;



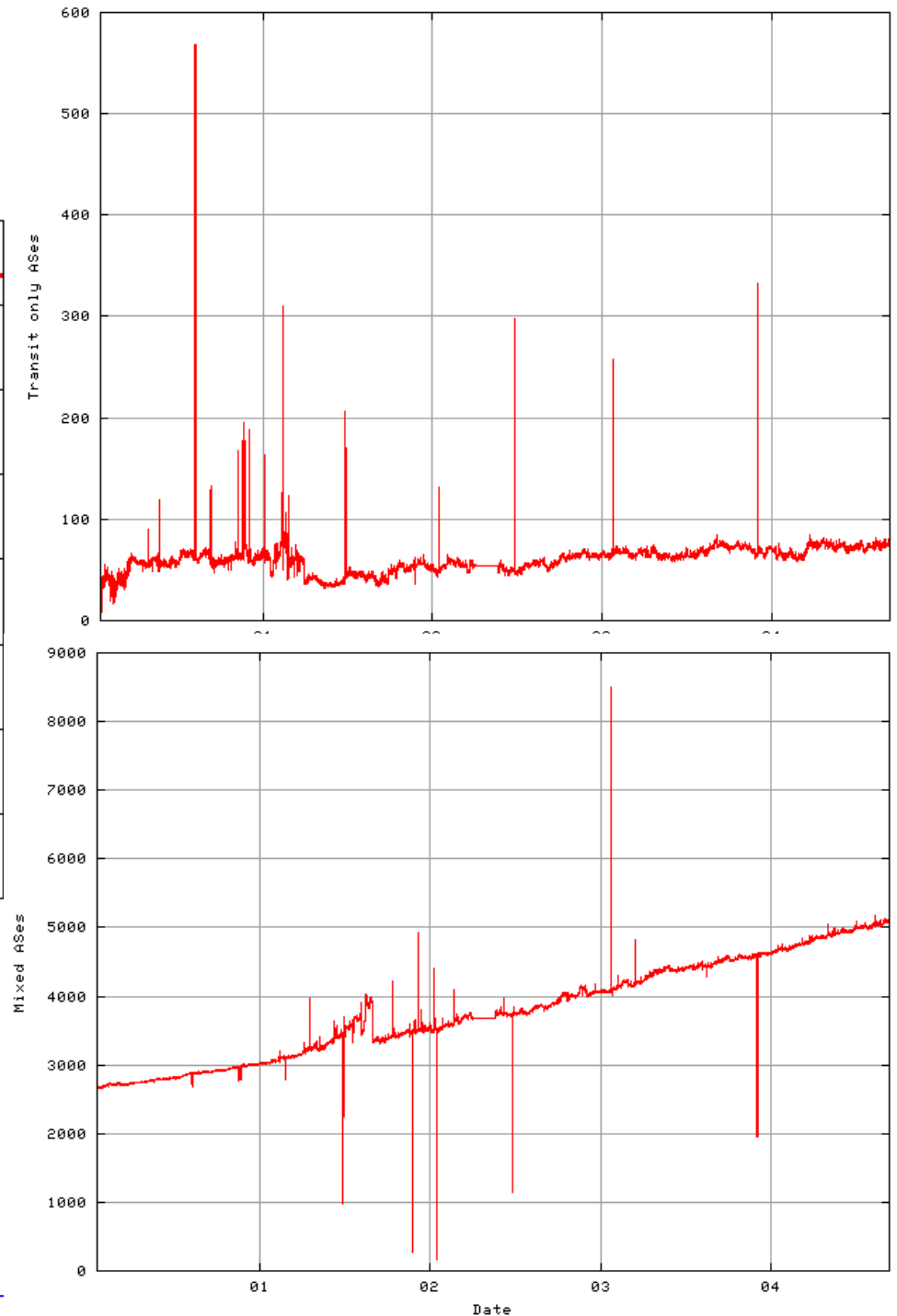
Τι θα κάνουμε;

- **Μεγαλύτεροι ASNs αριθμοί**
 - Σύντομα οι αριθμοί θα είναι μήκους **32bit** για να αντιμετωπιστεί η μεγάλη ζήτηση
 - Internet Draft: “BGP support for four-octet AS number space” (<http://www.ietf.org/internet-drafts/draft-ietf-idr-as4bytes-07.txt>)
 - Απαιτεί αλλαγή πρωτοκόλλων και ευρεία ανάπτυξη
- **Αλλαγή στον τρόπο που χρησιμοποιούμε τους ASN**
 - Να επιτρέψουμε σε δίκτυα **multihomed, non-transit** να χρησιμοποιούν **ιδιωτικούς** ASNs (όπως έγινε με τις διευθύνσεις)
 - **ASE** (AS number Substitution on Egress)
 - Internet Draft: “Autonomous System Number Substitution on Egress” (draft-jhaas-ase-00.txt)
 - Απαιτεί εργασία στα άκρα του δικτύου, απαιτεί αλλαγή πρωτοκόλλου (για αποτροπή βρόχων)
 - Κάνει κάποιες βλάβες πιο δύσκολα ανιχνεύσιμες

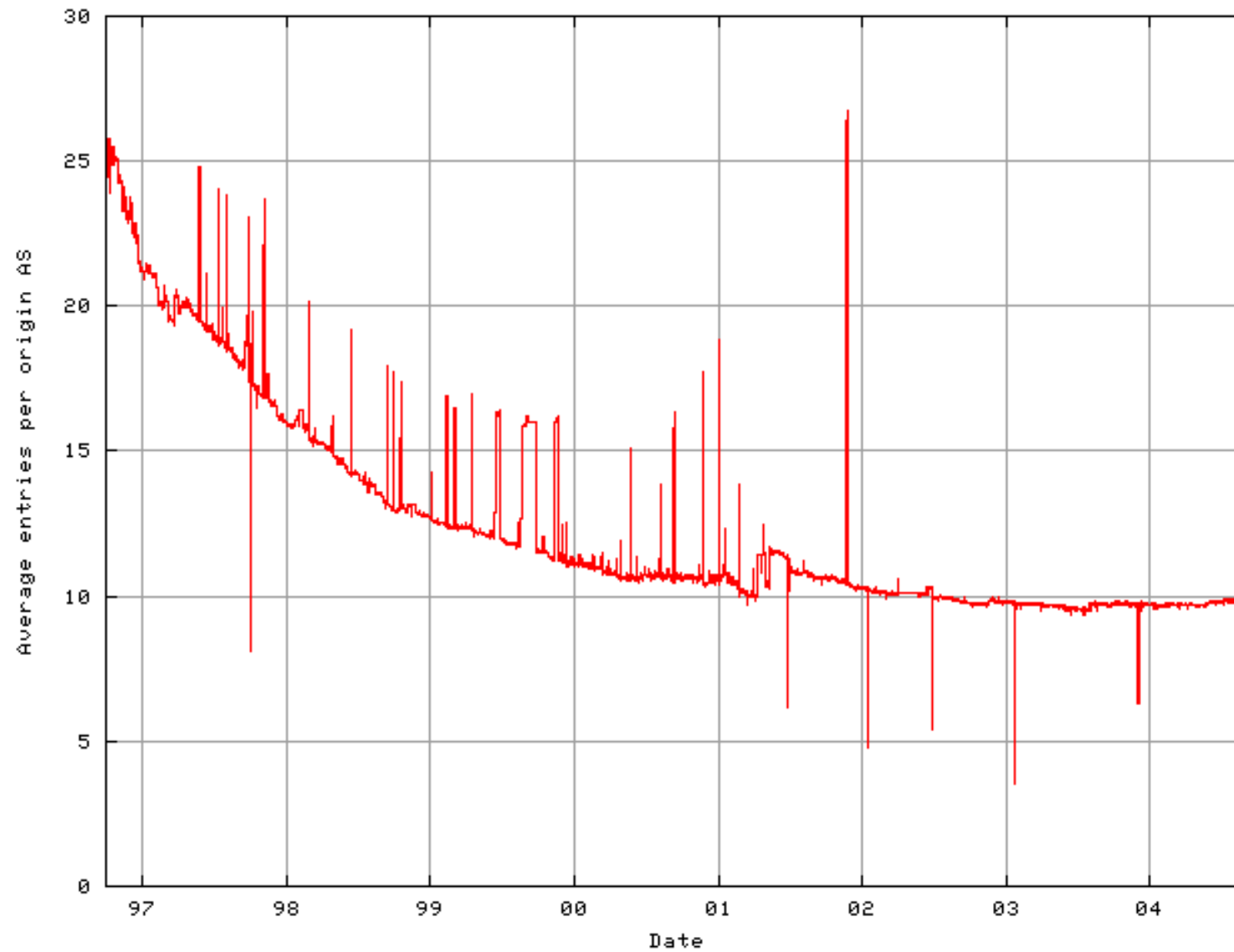
Τι είδους Αυτόνομα Συστήματα υπάρχουν;



- Τα περισσότερα είναι μόνο γενεσιουργά (**originating**), ενώ πολύ λίγα είναι καθαρά διαβιβαστικά (**transit**)



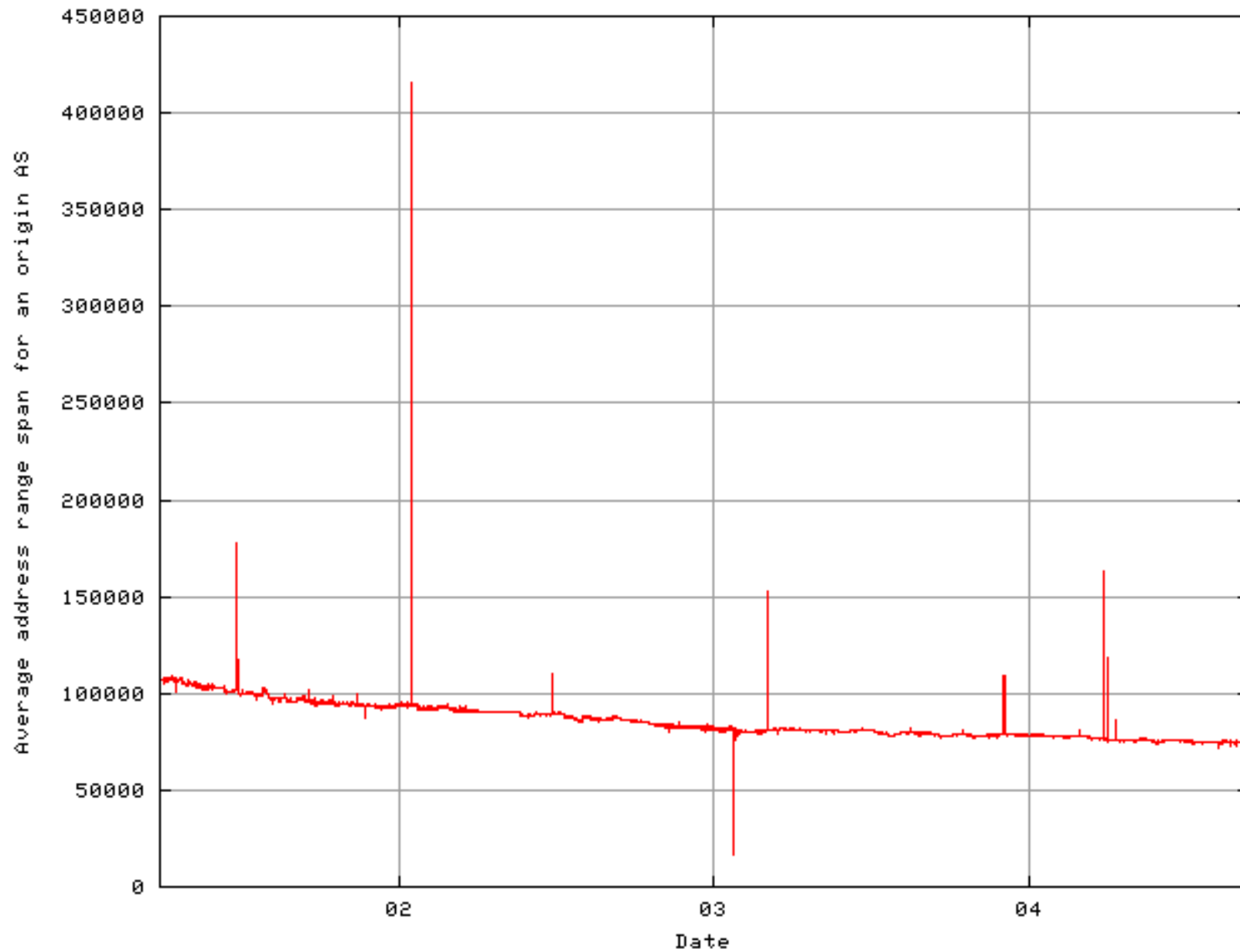
Μέσος όρος διαδρομών που παράγονται ανά AS



Υπάρχουν και AS που παράγουν πολλές διαδρομές (top 20)

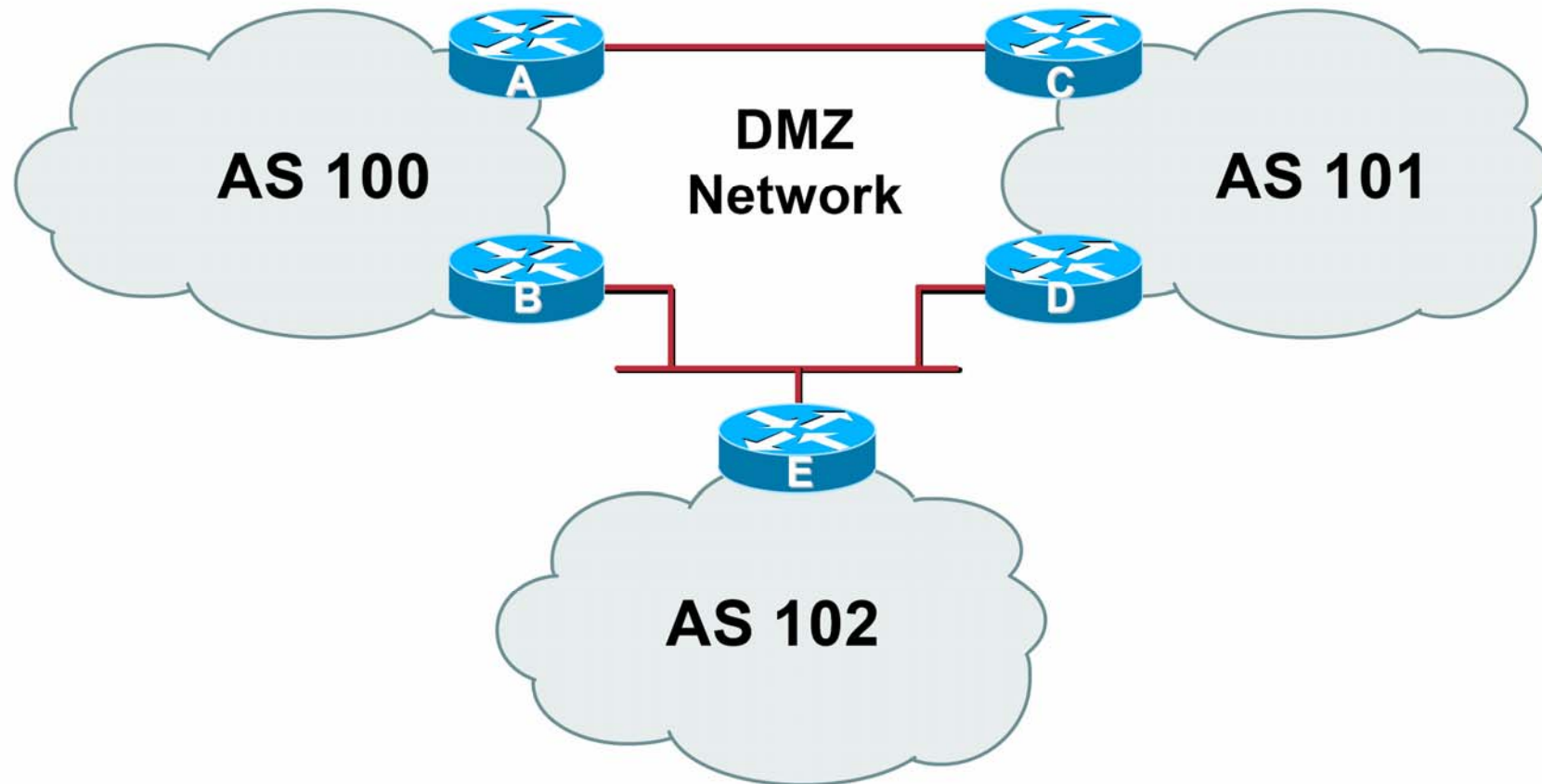
•1382	<u>AS7018</u>	ATTW AT&T WorldNet Services
•1249	<u>AS701</u>	UU UUNET Technologies, Inc.
•950	<u>AS1239</u>	SPRN Sprint
•826	<u>AS2386</u>	ADCS-1 AT&T Data Communications Services
•781	<u>AS4134</u>	CHINANET-BACKBONE No.31,Jin-rong Street
•778	<u>AS4323</u>	TWTC Time Warner Telecom
•739	<u>AS18566</u>	CVAD Covad Communications
•717	<u>AS6197</u>	BNS-14 BellSouth Network Solutions, Inc
•715	<u>AS721</u>	DNIC DoD Network Information Center
•641	<u>AS702</u>	AS702 MCI EMEA - Commercial IP service provider in Europe
•621	<u>AS209</u>	QWEST-4 Qwest
•582	<u>AS20115</u>	CC04 Charter Communications
•580	<u>AS3356</u>	LEVEL3 Level 3 Communications
•545	<u>AS4766</u>	KIXS-AS-KR Korea Telecom
•539	<u>AS6347</u>	SAVV SAVVIS Communications Corporation
•531	<u>AS9583</u>	SATYAMNET-AS Satyam Infoway Ltd.,
•510	<u>AS852</u>	AGTA Telus Advanced Communications
•488	<u>AS7843</u>	ADELPH-13 Adelphia Corp.
•466	<u>AS3561</u>	CWU Cable & Wireless USA
•451	<u>AS6198</u>	BNS-14 BellSouth Network Solutions, Inc

Μέσο εύρος διευθύνσεων ανά AS



DMZ, Demarcation Zone

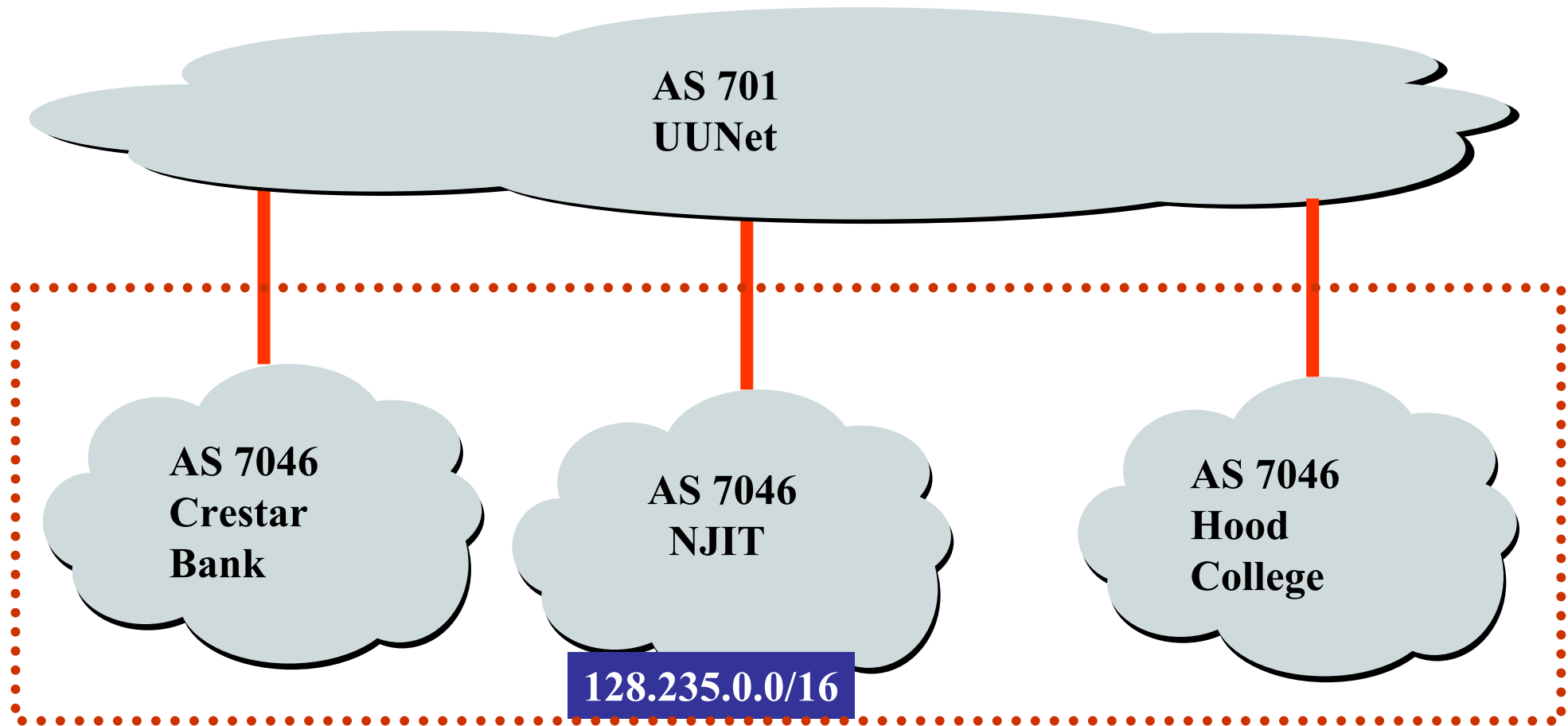
- Ορίζεται ως το **διαμοιραζόμενο-κοινό** δίκτυο μεταξύ αυτόνομων συστημάτων



AS != ARD

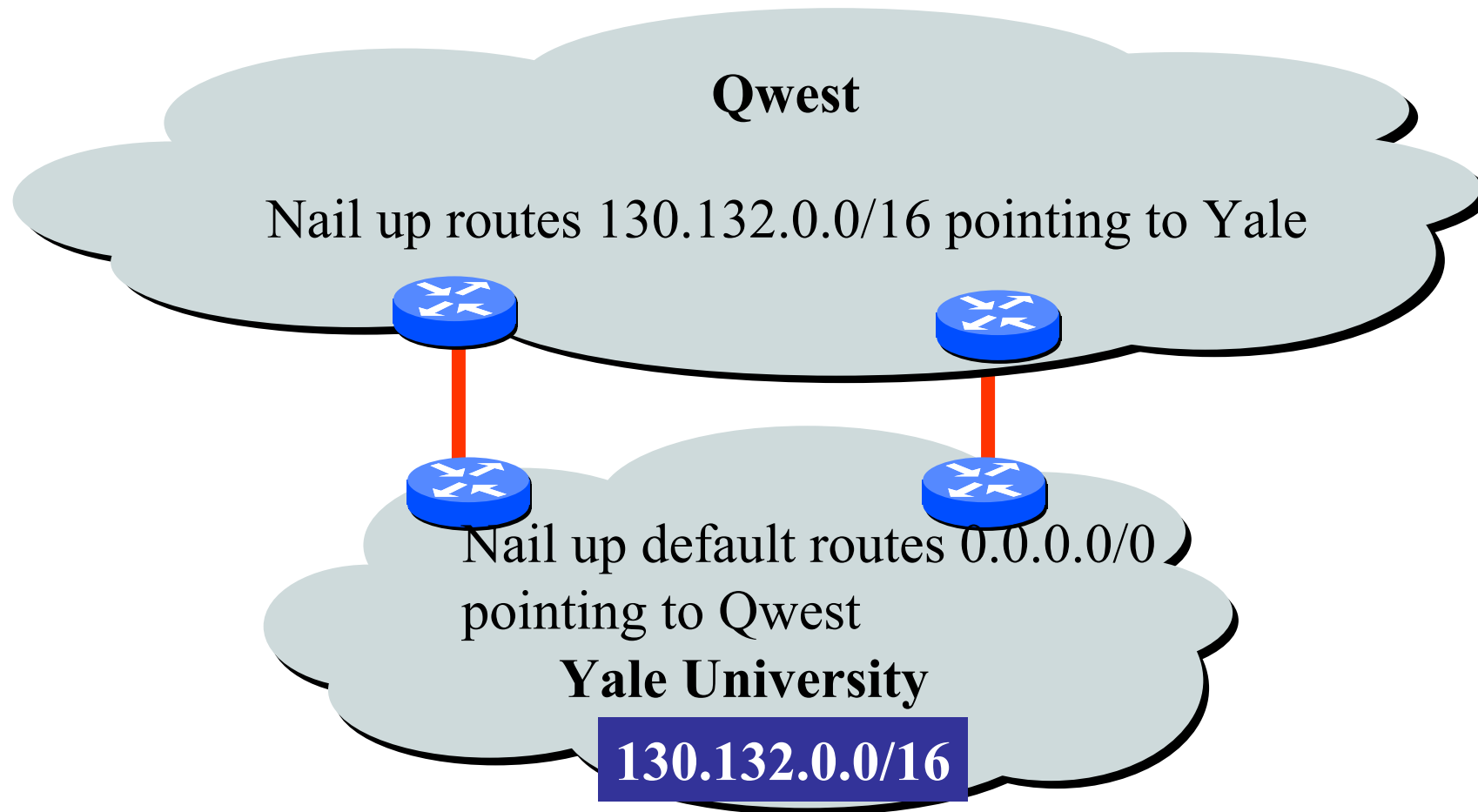
- Τα περισσότερα αυτόνομα κράτη δρομολόγησης, **ARD** δεν έχουν **ASN** διότι συνδέονται με **στατική δρομολόγηση** και βρίσκονται στο άκρο του διαδικτύου
- Μερικά ασυσχέτιστα ARD **μοιράζονται το ίδιο ASN (RFC2270)**
- Μερικά ARD υλοποιούνται με **πολλαπλά ASN**

Τα ARD μπορεί να μοιράζονται το ίδιο ASN (RFC 2270)



- Το ASN7046 έχει ανατεθεί στο UUNET
- Χρησιμοποιείται από πελάτες που φιλοξενούνται στο UUNET, αλλά χρειάζονται BGP για κάποιους άλλους λόγους (διαμοιρασμό φόρτου, ..)

Τα ARD μπορεί να συνδέονται με στατική δρομολόγηση



- Η στατική δρομολόγηση είναι ο πιο συνήθης τρόπος για να συνδέσουμε ένα **ARD** στο διαδίκτυο
- Αυτός είναι ένας λόγος για τον οποίο το BGP είναι ένα μυστήριο για αρκετούς

Τύποι κίνησης

Εξερχόμενη κίνηση (Egress Traffic)

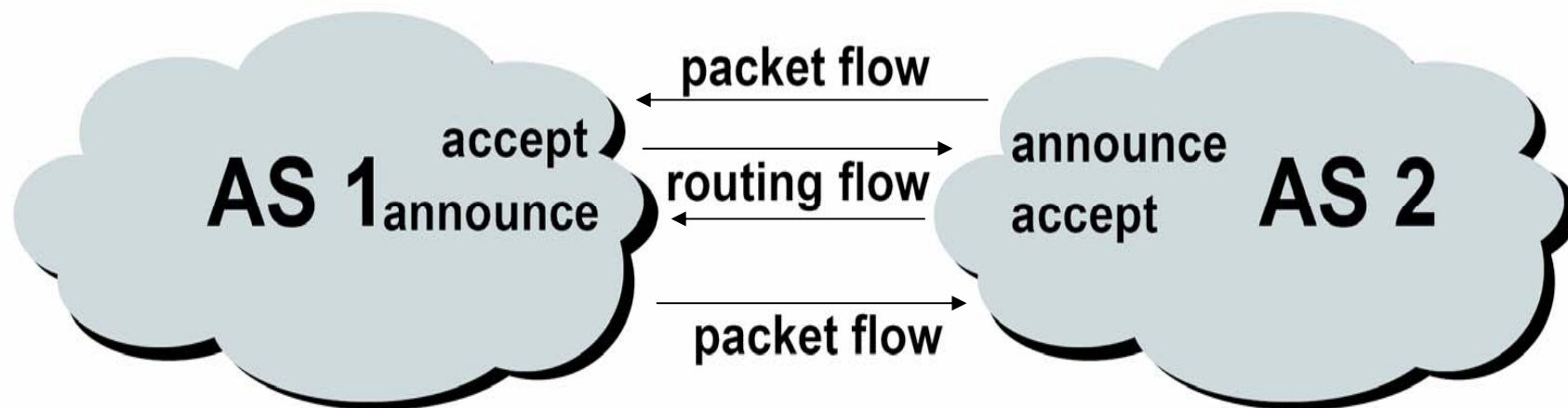
- Πως τα πακέτα φεύγουν από το δίκτυο
- Εξαρτάται από:
 - Διαθεσιμότητα διαδομών (τι μας στέλνουν οι άλλοι)
 - Αποδοχή διαδρομών (τι αποδεχόμαστε από τους άλλους)
 - Τακτική δρομολόγησης (τι κάνουμε με τις διαδρομές που μας στέλνουν)
 - Συμφωνίες για ομότιμες σχέσεις (peering) ή για διέλευση (transit)

Εισερχόμενη κίνηση (Ingress Traffic)

- Πως τα πακέτα εισέρχονται στο δίκτυο μας και σε αυτό των πελατών μας
- Εξαρτάται από:
 - Τι πληροφορία δρομολόγησης-συνδεσιμότητας στέλνουμε και σε ποιον
 - Βασίζεται στην διευθυνσιοδότηση μας και τα AS's
 - Βασίζεται στην τακτική των άλλων (τι δέχονται από εμάς και τι κάνουν με αυτό)

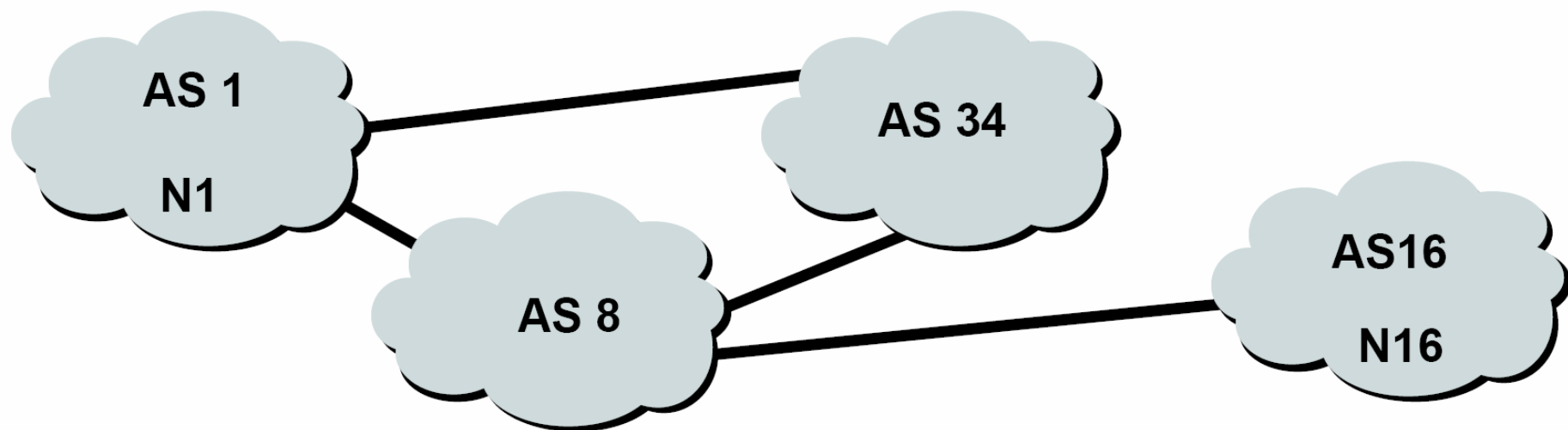
Ροή πληροφορίας δρομολόγησης – Ροή κίνησης

- Το AS1 πρέπει να **στείλει** πληροφορία δρομολόγησης στο AS2
- Το AS2 πρέπει να **αποδεχτεί** αυτή την πληροφορία
- **Μετά** μπορεί να γίνει ροή πληροφορίας από το AS2 στο AS1
- Το ίδιο πρέπει να γίνει από το AS2 στο AS1
- Η ροή της **κίνησης** γίνεται **αντίθετα** στη ροή της **πληροφορίας** δρομολόγησης
- Έτσι **φιλτράροντας** την **εξερχόμενη** πληροφορία δρομολόγησης αποτρέπουμε **εισερχόμενη** κίνηση και το αντίθετο



Ροή πληροφορίας δρομολόγησης – Ροή κίνησης

- Στο σχήμα για να υπάρξει ροή πληροφορίας από το δίκτυο **N1** στο δίκτυο **N16** πρέπει το **AS16** να **διαφημίσει** το δίκτυο **N16** προς το **AS8**
- Αυτό πρέπει να **αποδεχθεί** την ανακοίνωση και να το **αναγγείλει** προς τα **AS1** ή **AS34**
- Το **AS1** πρέπει να **αποδεχθεί** την ανακοίνωση από το **AS8** ή το **AS34**
- Για τη ροή πληροφορίας προς την αντίστροφη κατεύθυνση αντίστοιχα πρέπει να ισχύουν



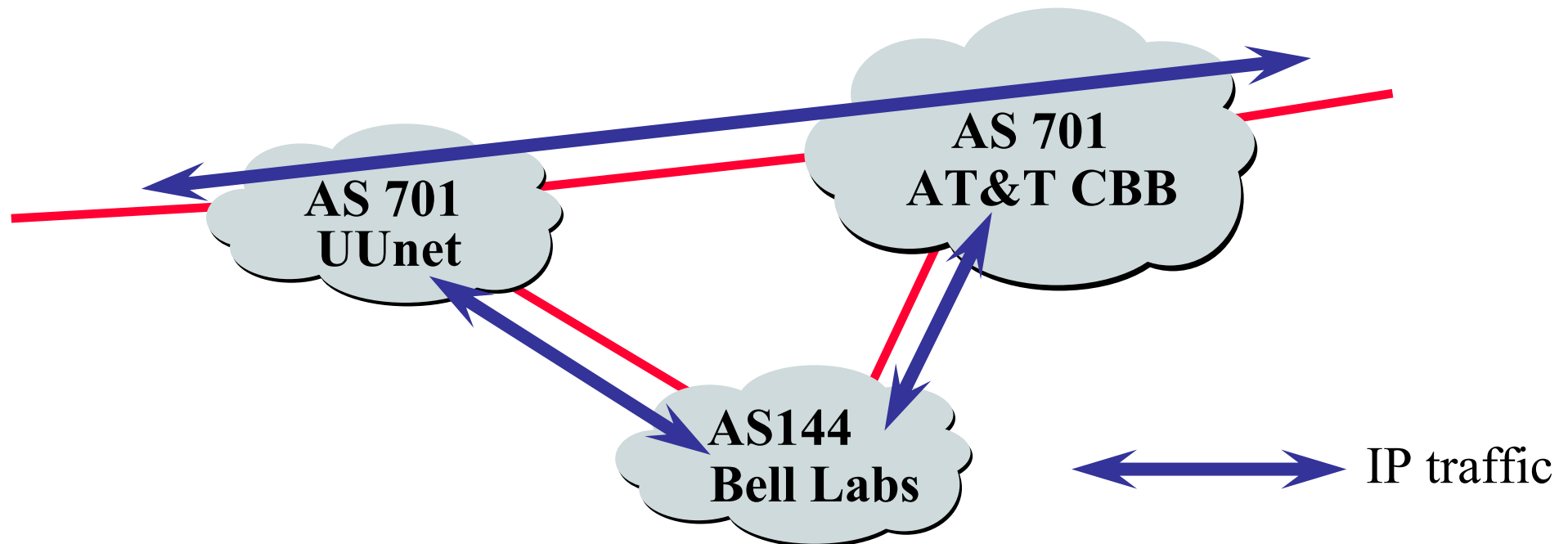
Μη Διαβιβαστικά και Διαβιβαστικά AS (Nontransit vs Transit AS)

Transit AS

- Επιτρέπει σε κίνηση η οποία **δεν παράγεται** ούτε έχει **προορισμό** εντός του AS να διέλθει **διαμέσω** του δικτύου

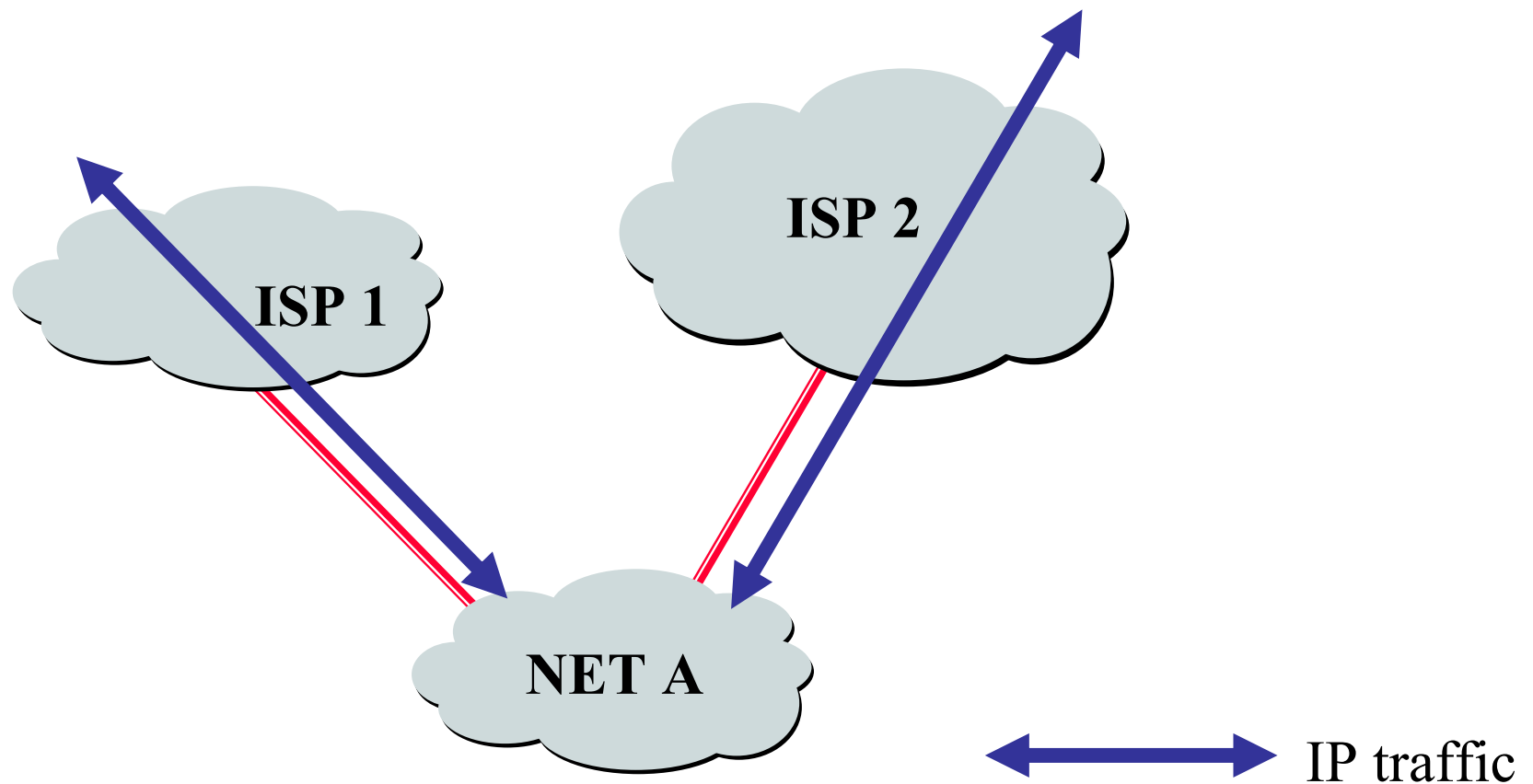
Nontransit AS

- Επιτρέπει μόνο την κίνηση η οποία έχει **παραχθεί** ή έχει **προορισμό** εντός του AS



Μη Διαβιβαστικά και Διαβιβαστικά AS (Nontransit vs Transit AS)

- Οι **ISP** έχουν συνήθως διαβιβαστικά AS
- Ένα **εταιρικό δίκτυο** ή ένας **πάροχος περιεχομένου** (content provider) μπορεί να έχει ένα μη διαβιβαστικό AS

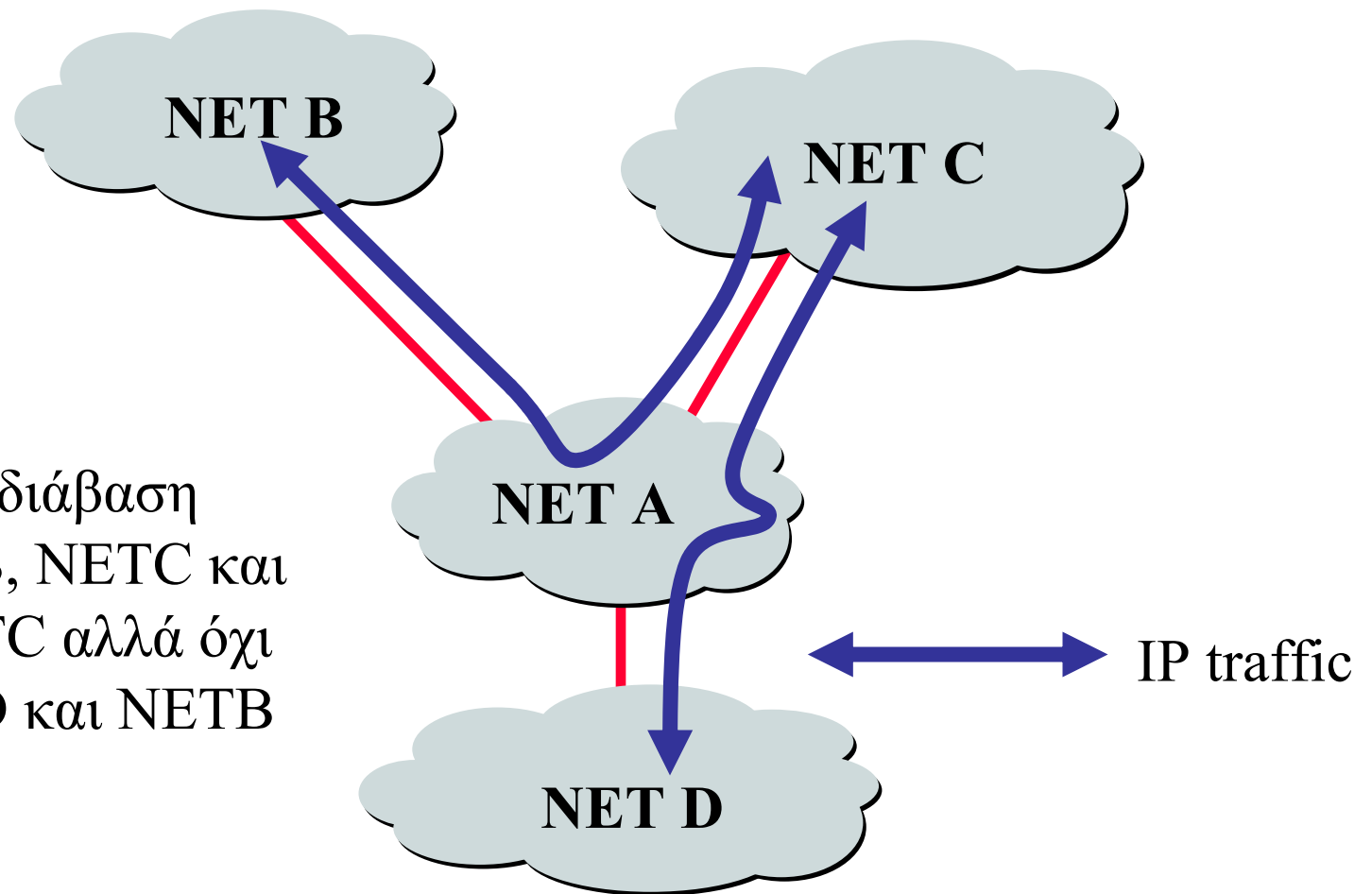


Επιλεκτικώς διαδιβαστικά ΑΣ (SelectiveTransit AS)

- Τα περισσότερα διαβιβαστικά παρέχουν **επιλεκτική** διάβαση

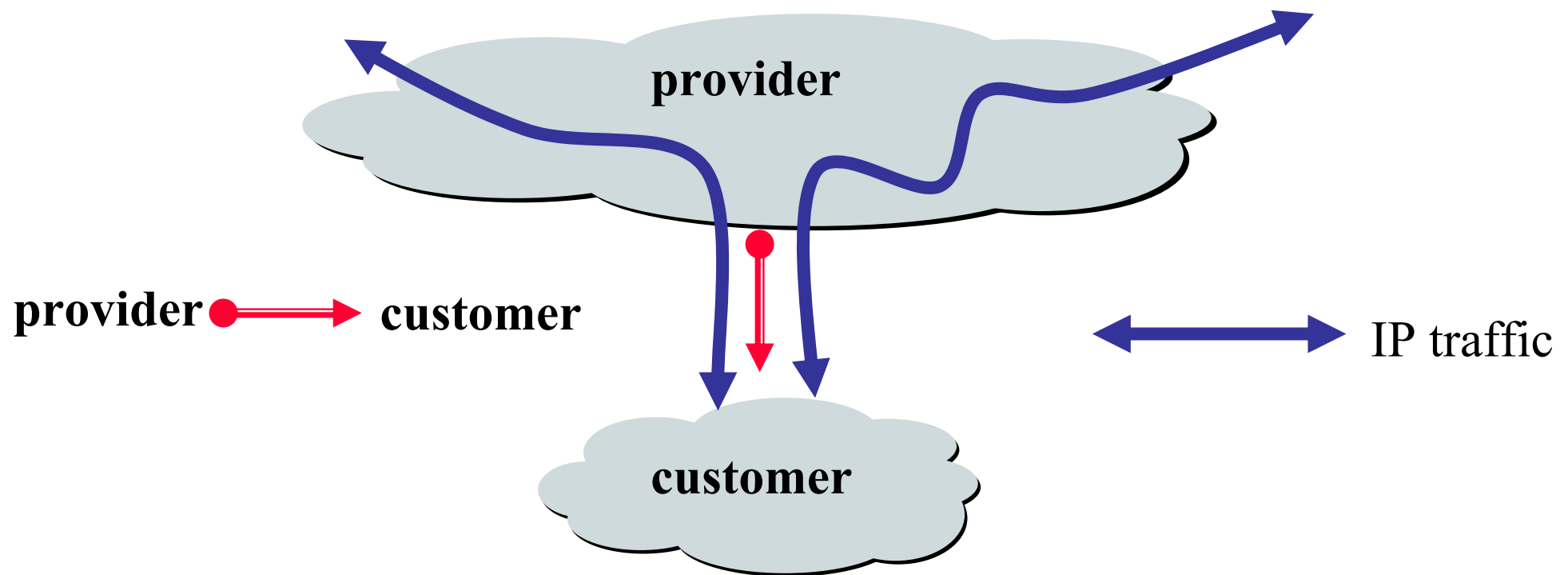
Παράδειγμα

- Το NETA παρέχει διάβαση ανάμεσα στα NETB, NETC και στα NETD και NETC αλλά όχι ανάμεσα στα NETD και NETB

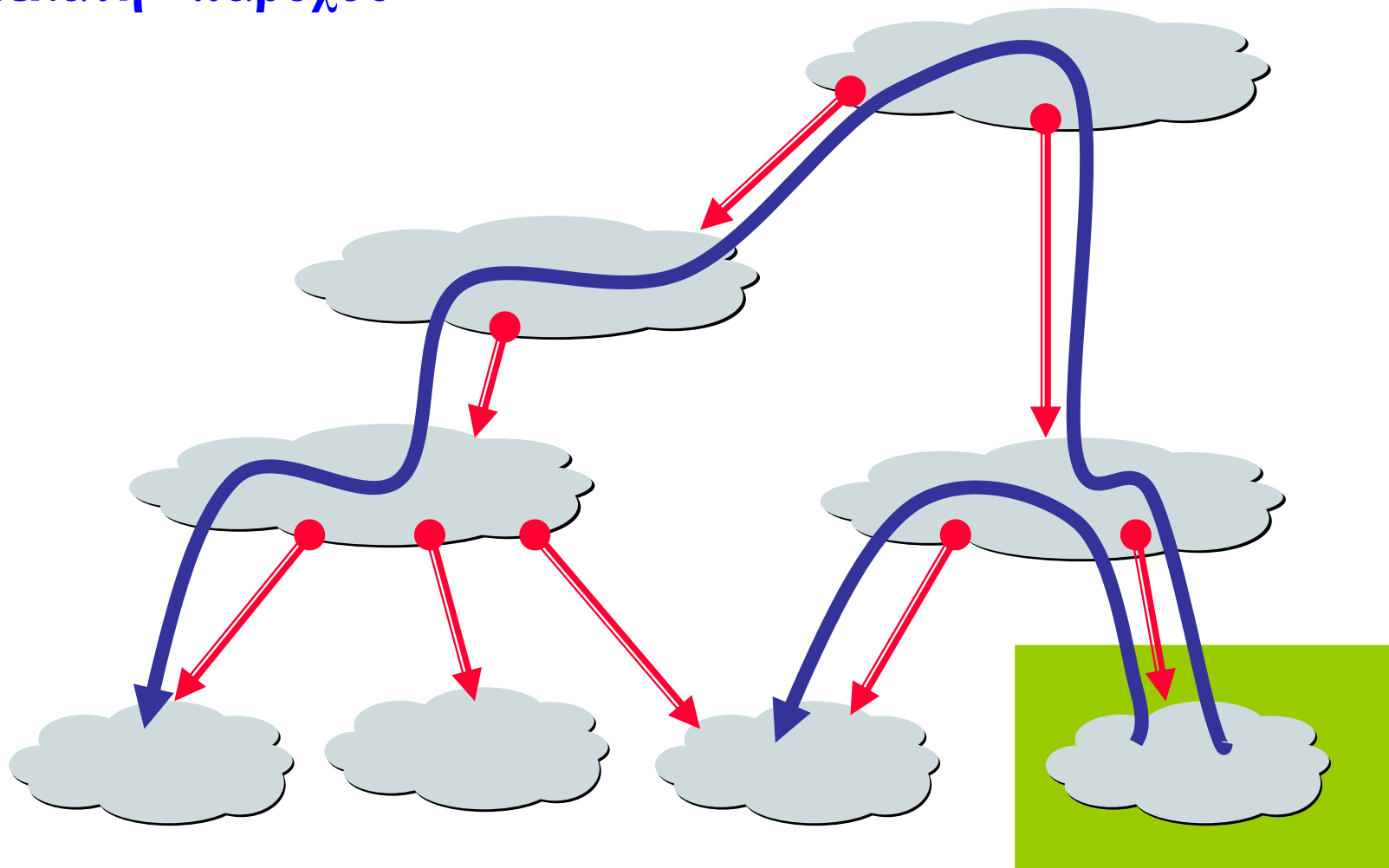


Πελάτης -Πάροχος

- Ο πελάτης πληρώνει τον πάροχο για την πρόσβαση στο διαδίκτυο



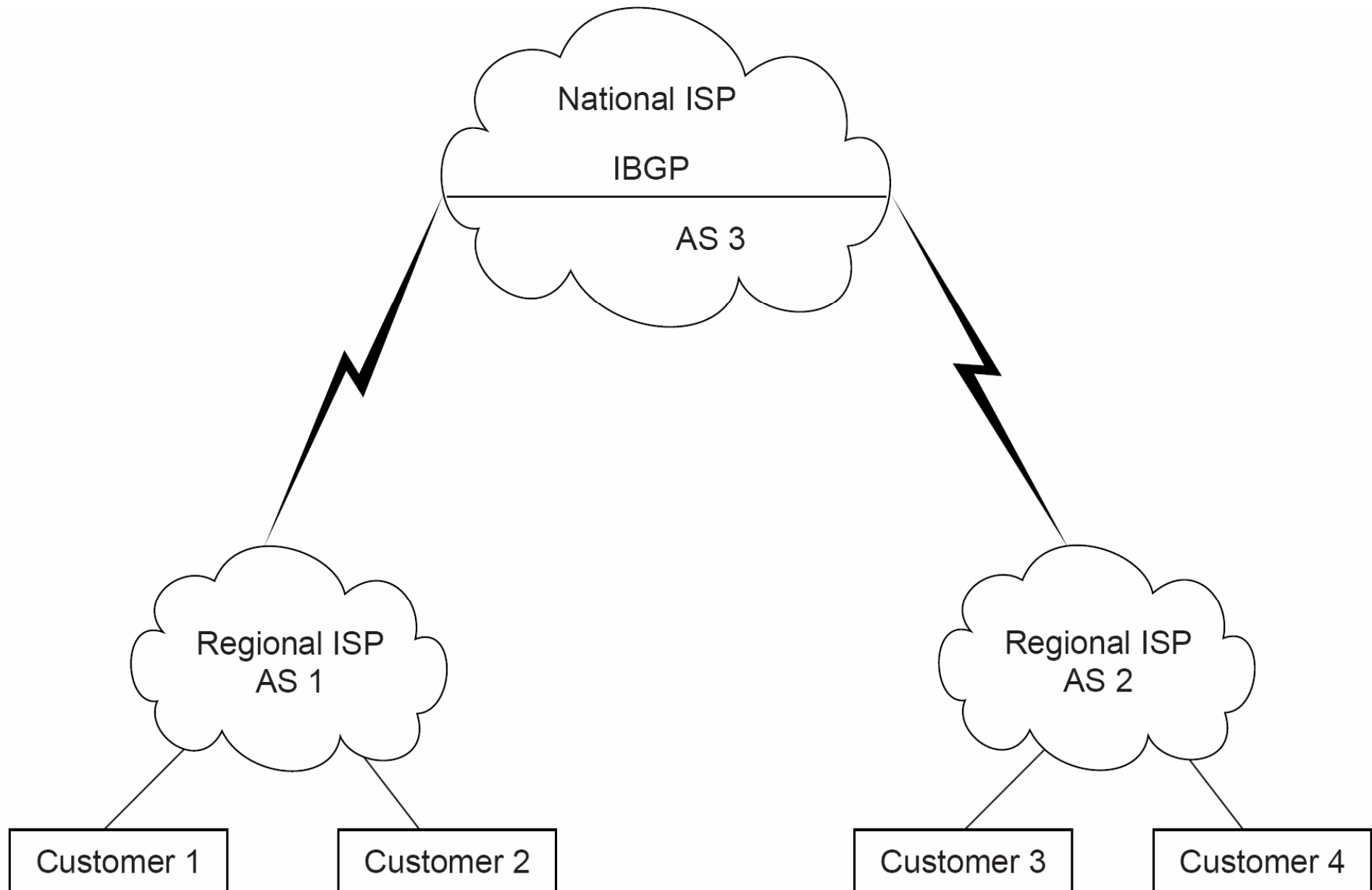
Ιεραρχία πελάτη - πάροχου



provider ● → customer

← IP traffic

Ιεραρχία πελάτη - πάροχου

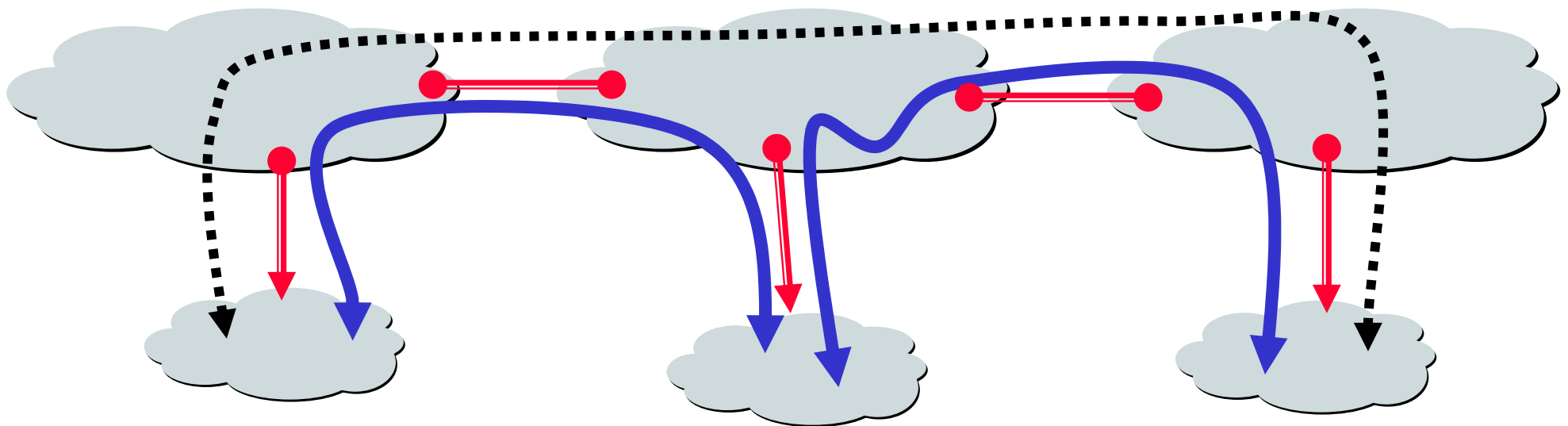


Τι πρωτόκολλα χρησιμοποιούνται και που;

- Το **BGP** χρησιμοποιείται για την ανταλλαγή πληροφορίας δρομολόγησης για το **διαδίκτυο** και είναι το πρωτόκολλο που χρησιμοποιείται ανάμεσα στους πάροχους internet (ISP)
- Τα δίκτυα των **πελατών** όπως πανεπιστήμια, εταιρικά δίκτυα συνήθως χρησιμοποιούν ένα πρωτόκολλο Interior Gateway Protocol (**IGP**) όπως το RIP ή το OSPF για την ανταλλαγή της πληροφορίας δρομολόγησης εντός των δικτύων τους
- Οι **πελάτες** συνδέονται στους **ISPs** και οι ISPs χρησιμοποιούν **BGP** για την ανταλλαγή των διαδρομών των πελατών ή των ISP

Ομότιμες σχέσεις (Peering)

- Οι ομότιμοι παρέχουν **διάβαση** ανάμεσα στους πελάτες τους
- Δεν παρέχουν διάβαση ανάμεσα σε **άλλους** ομότιμους
- Η συνήθης πρακτική είναι να **μην** ανταλλάσσουν **χρήματα** μεταξύ τους

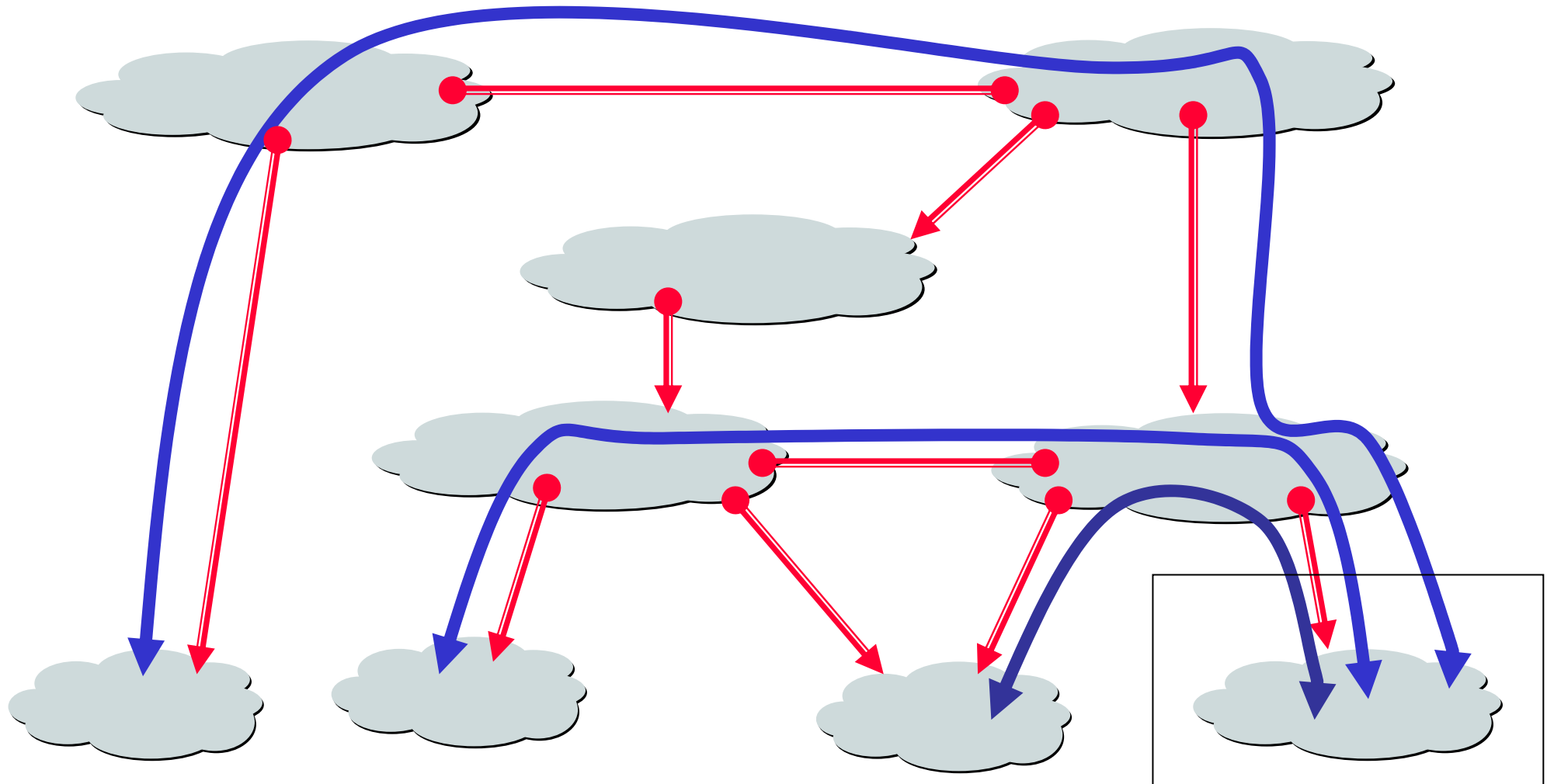


peer ●———● peer
provider ●———▶ customer

↔
traffic
allowed

⋯↔
traffic NOT
allowed

Οι Ομότιμες σχέσεις παρέχουν συντομεύσεις στις διαδρομές



peer ●—● peer
provider ●→ customer

Οι Ομότιμες σχέσεις συμφέρουν ή όχι;

ΝΑΙ

- Μείωση του upstream κόστους **διάβασης**
- Αυξημένη **επίδοση** άκρο σε άκρο
- Μπορεί να είναι η **μοναδική οδός** προς κάποιο τμήνα του διαδικτύου

ΟΧΙ

- Είναι προτιμότερο να έχουμε **πελάτες** παρά **ομότιμους**
- Οι ομότιμοι είναι συνήθως οι **ανταγωνιστές**
- Οι ομότιμες σχέσεις μπορεί να απαιτούν περιοδική **επαναδιαπραγμάτευση**
- Συχνά οι ομότιμες συμφωνίες έχουν **εμπιστευτικό** χαρακτήρα
- Από τα πλέον **σημαντικά** θέματα στο χώρο των ISP

Υλοποίηση σχέσης πελάτη-πάροχου και ομότιμης σχέσης

- Υλοποιούνται με την προσεκτική υλοποίηση κάποιας συγκεκριμένης **τακτικής δρομολόγησης**
- Το BGP επιτρέπει σαν πρωτόκολλο **τακτικής** (policy based) την εύκολη υλοποίηση τους

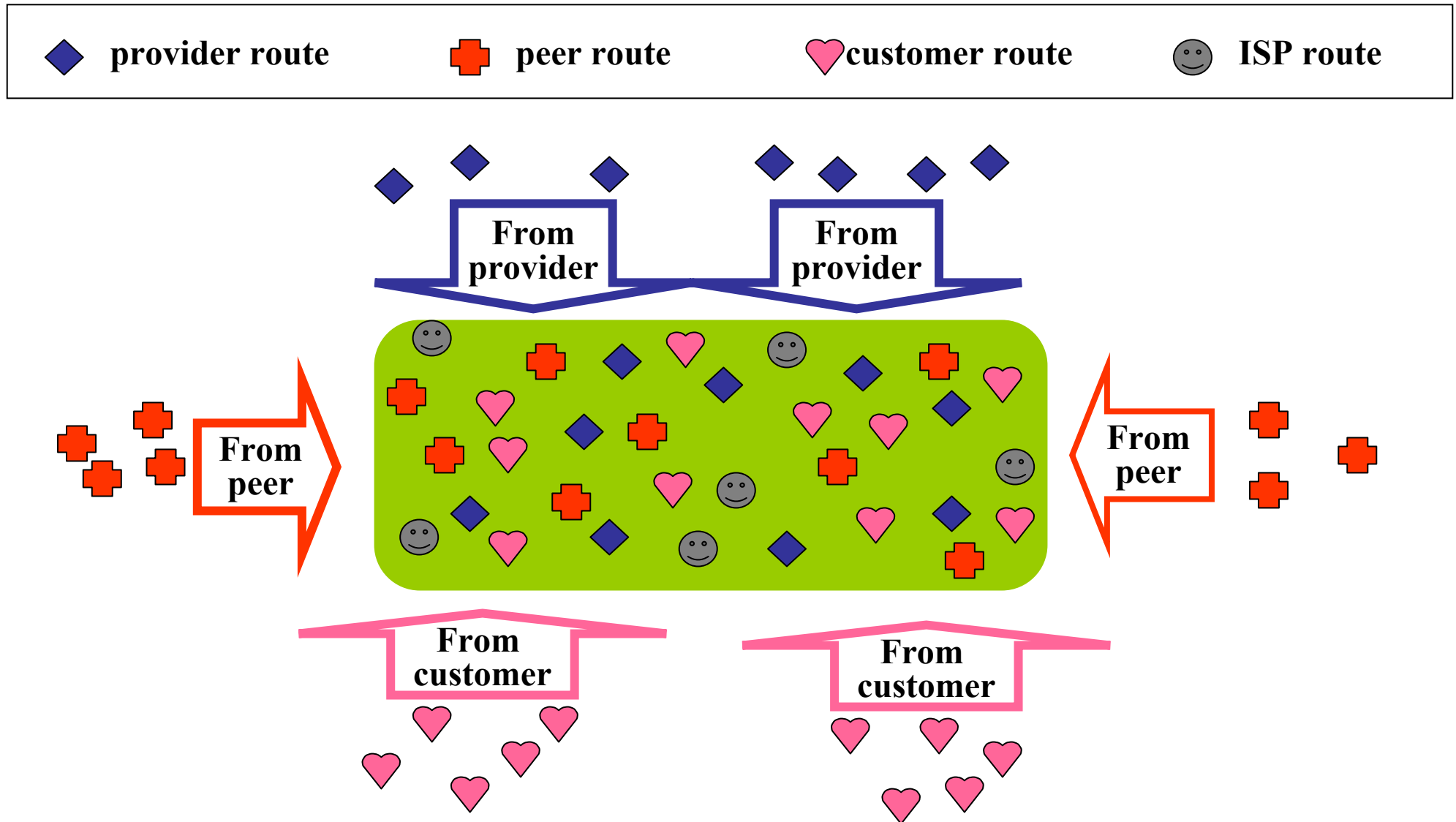
Σχέσεις διέλευσης

- Φίλτράρισμα στις εξερχόμενες διαδρομές

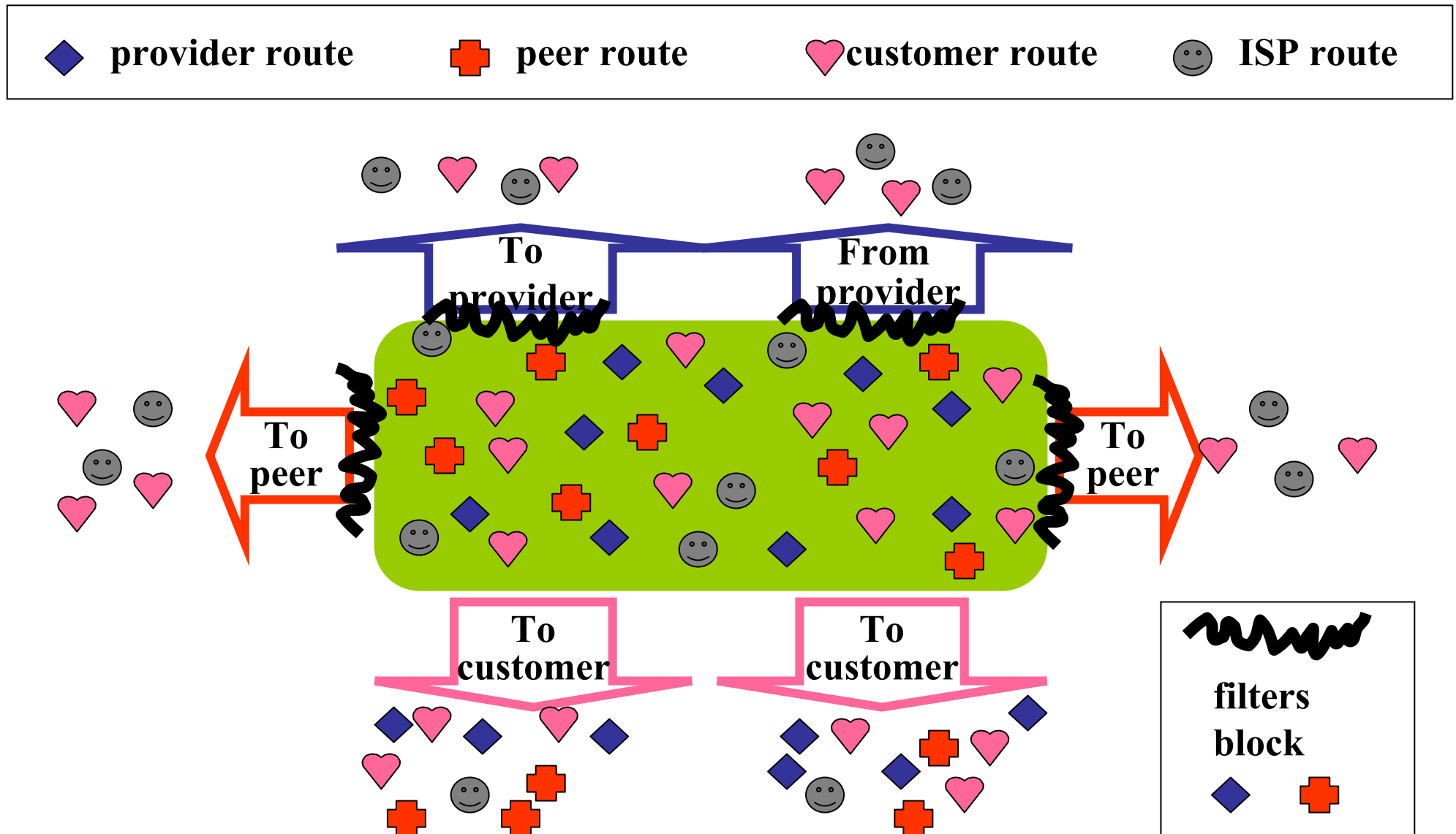
Διάταξη προτίμησης διαδρομών

- Πάροχος < Ομότιμος < Πελάτης

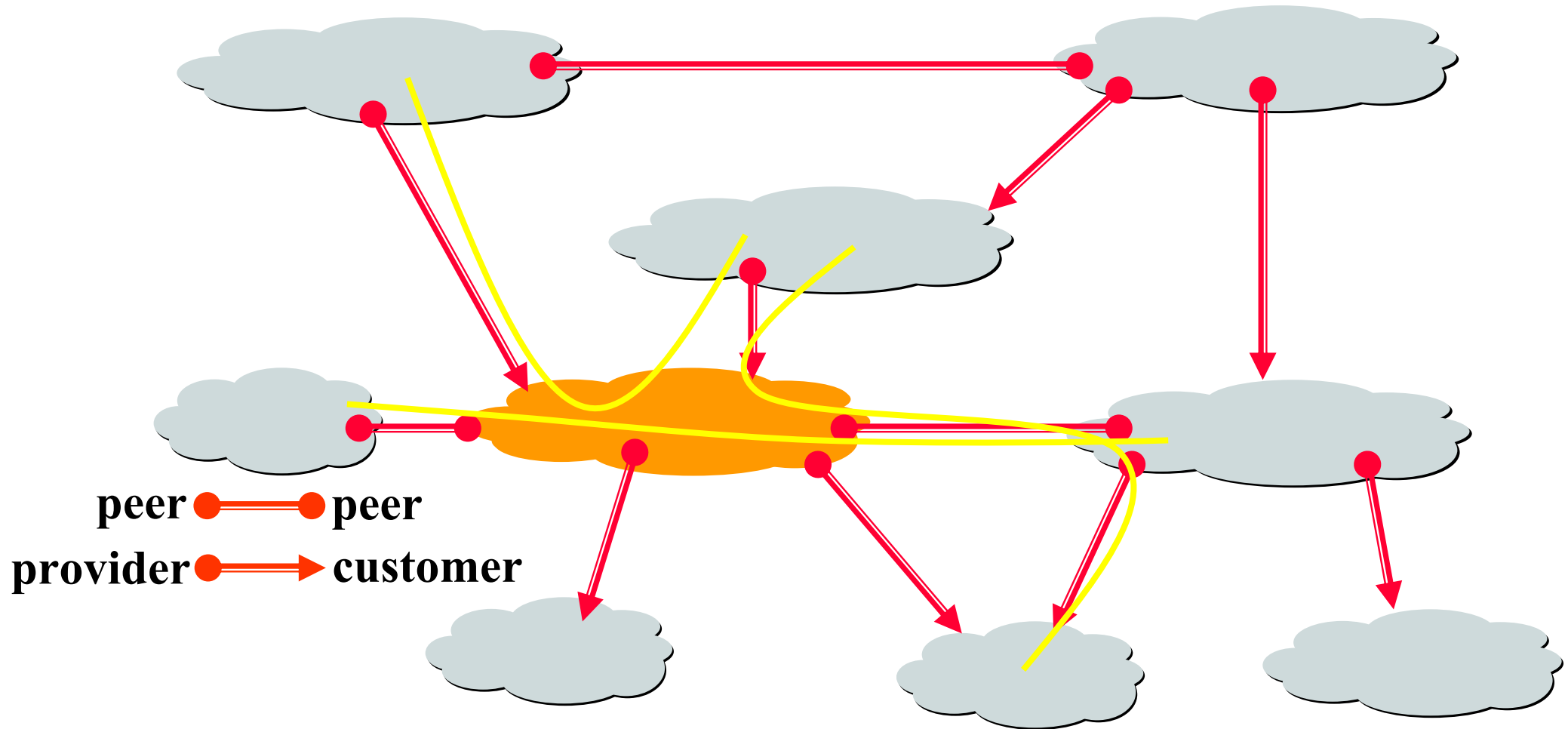
Εισαγωγή διαδρομών



Φίλτρα διαδρομών



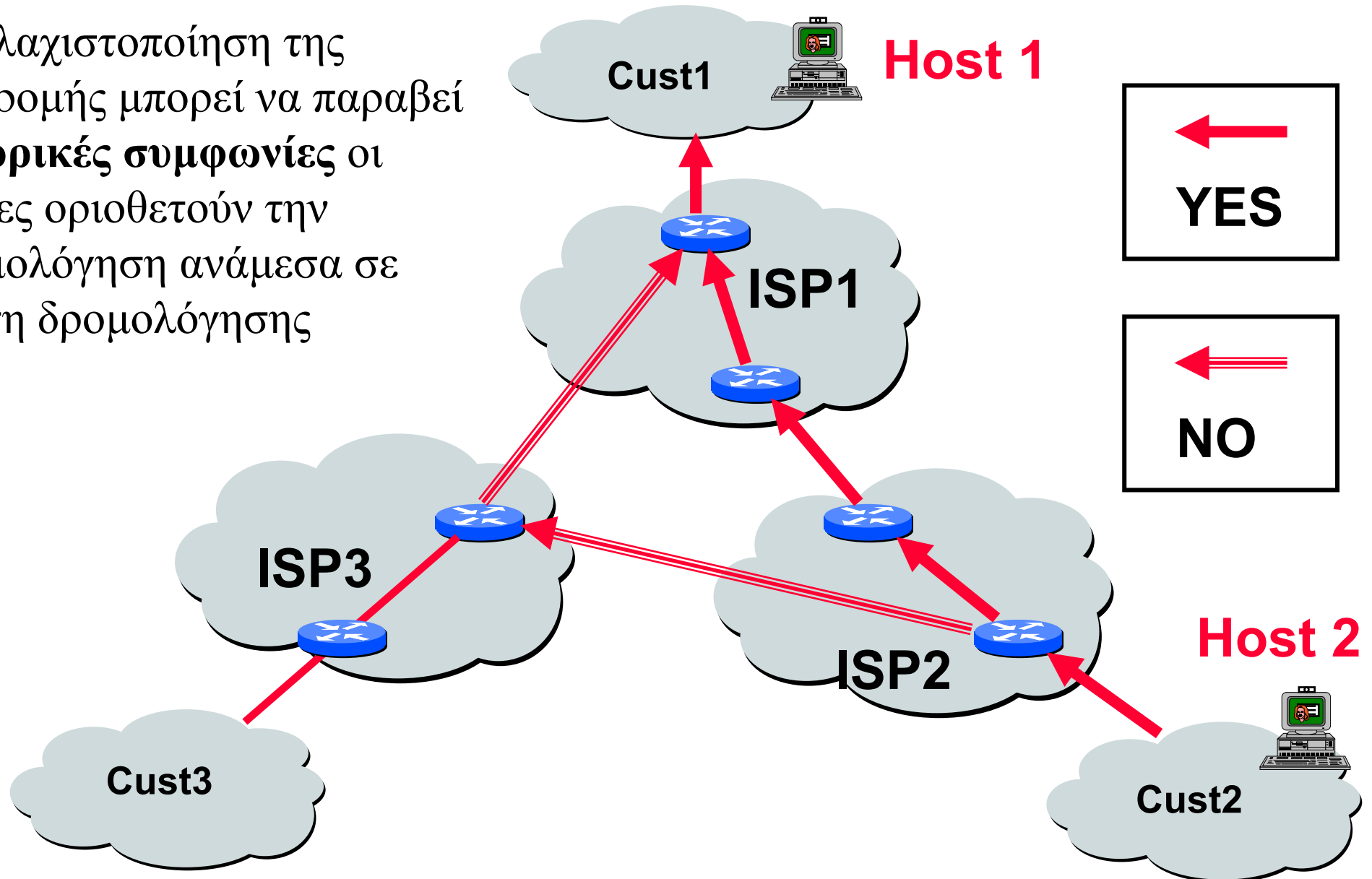
Πραγματοποίηση των σχέσεων



- Κόβοντας-φιλτράροντας τις παραπάνω διαδρομές αποτρέπουμε σε κίνηση από ένα πάροχο σε άλλον να περάσει διαμέσου εμάς
- Αποτρέπουμε επίσης κίνηση από ομότιμο μας (και πελάτες του) να περάσει σε άλλο ομότιμο μας
- Αποτρέπουμε ομότιμο να περάσει κίνηση από δικό μας πάροχο, όπως και στον πάροχο μας να περάσει κίνηση σε ομότιμο μας (και πελάτη του)

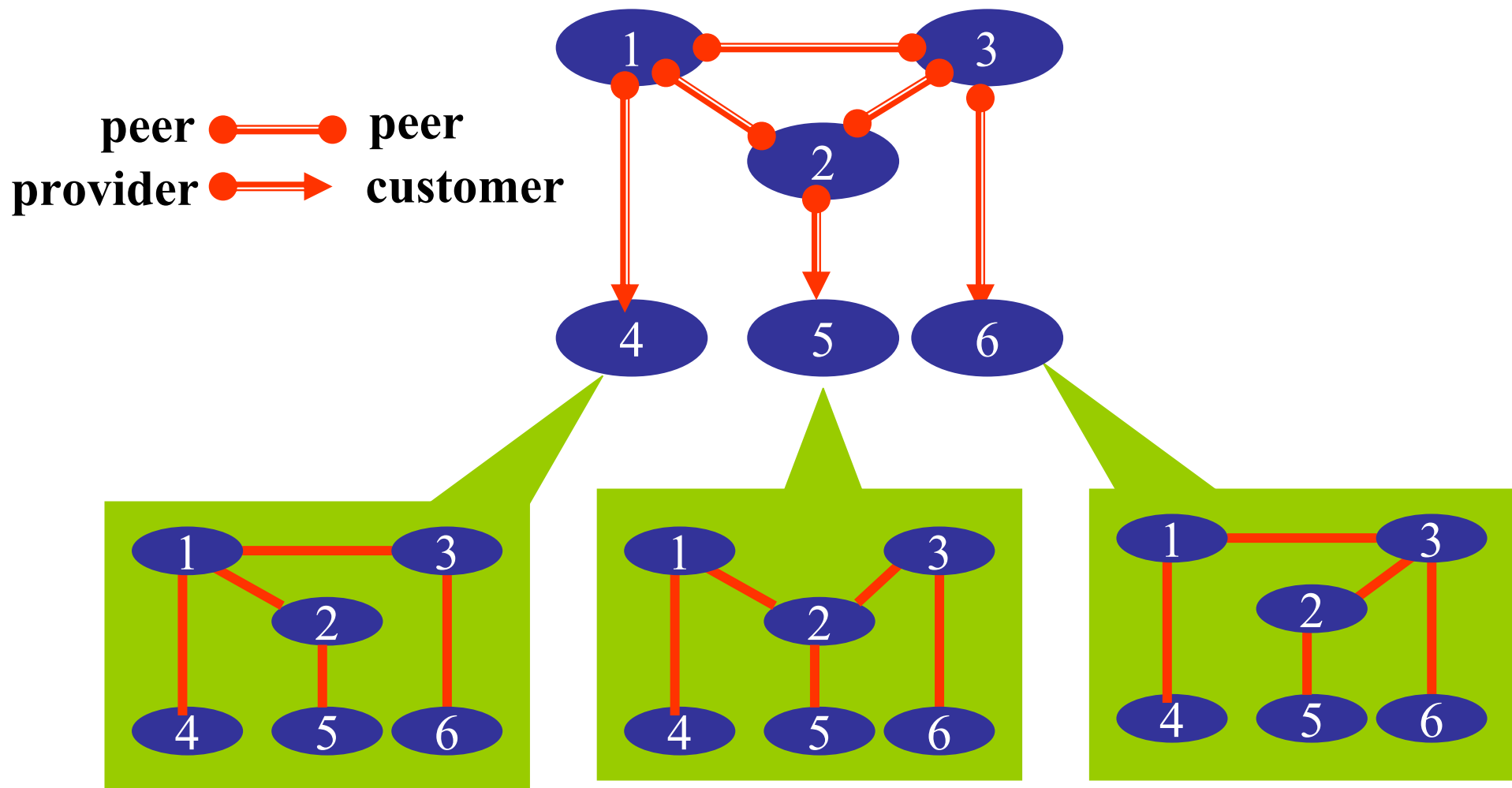
Δρομολόγηση βασισμένη στην τακτική ή στην απόσταση

- Η ελαχιστοποίηση της διαδρομής μπορεί να παραβεί **εμπορικές συμφωνίες** οι οποίες οριοθετούν την δρομολόγηση ανάμεσα σε κράτη δρομολόγησης



Οι γράφοι των αυτόνομων συστημάτων εξαρτώνται από την οπτική μας γωνία

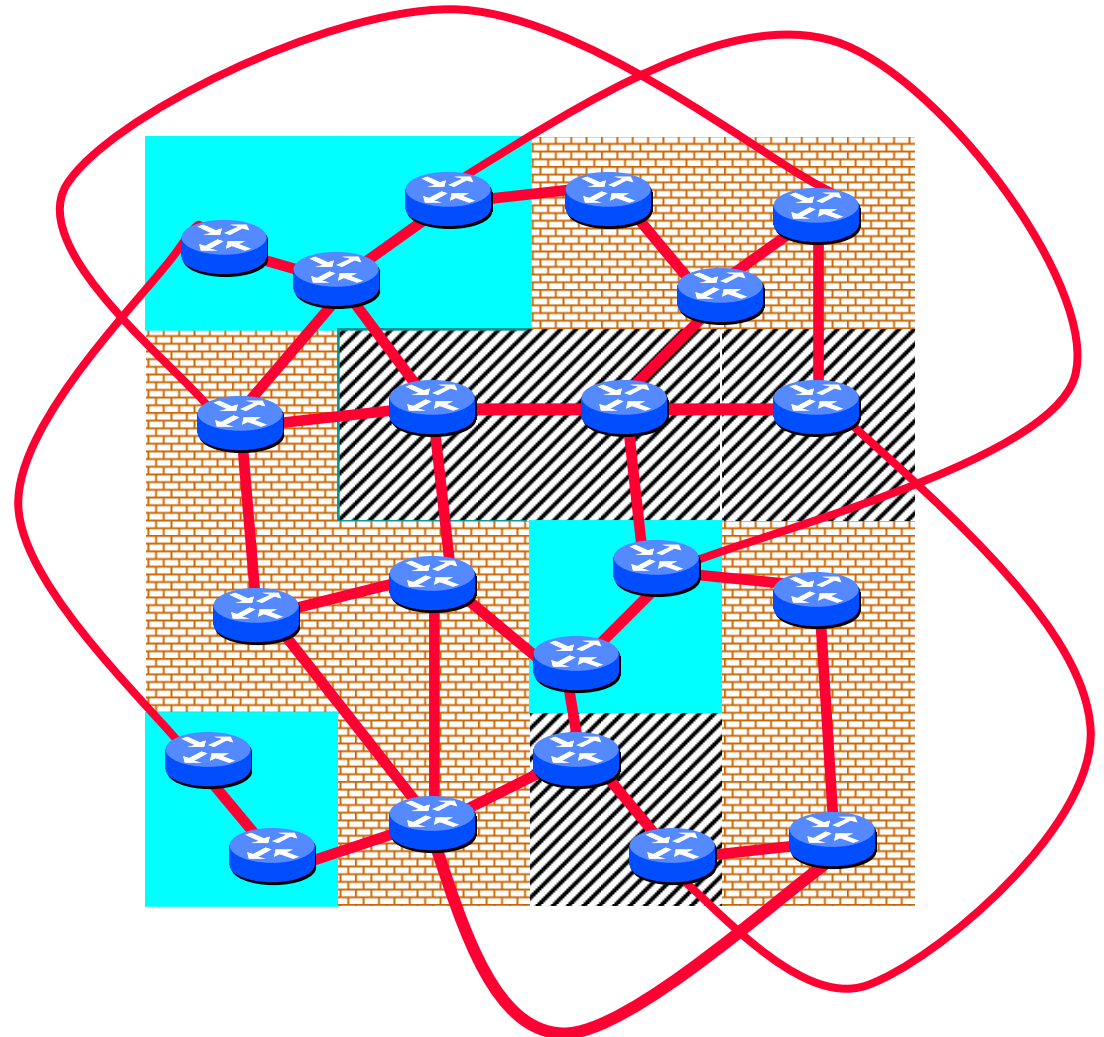
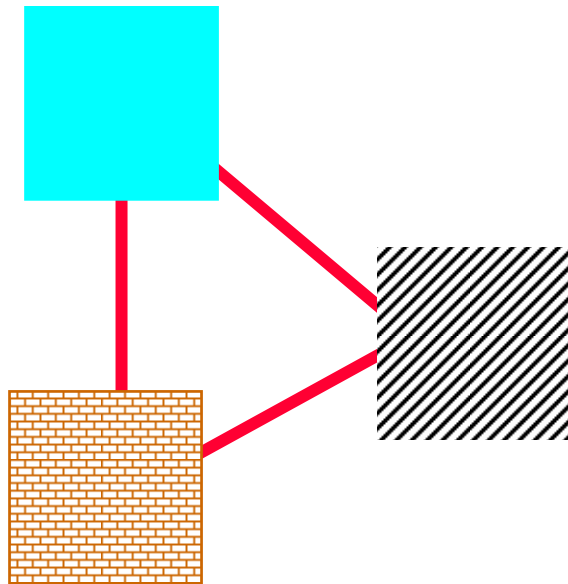
- Δηλαδή δεν έχουμε αντίληψη της πραγματικής **τοπολογίας**



Οι γράφοι των AS δεν δείχνουν την τοπολογία

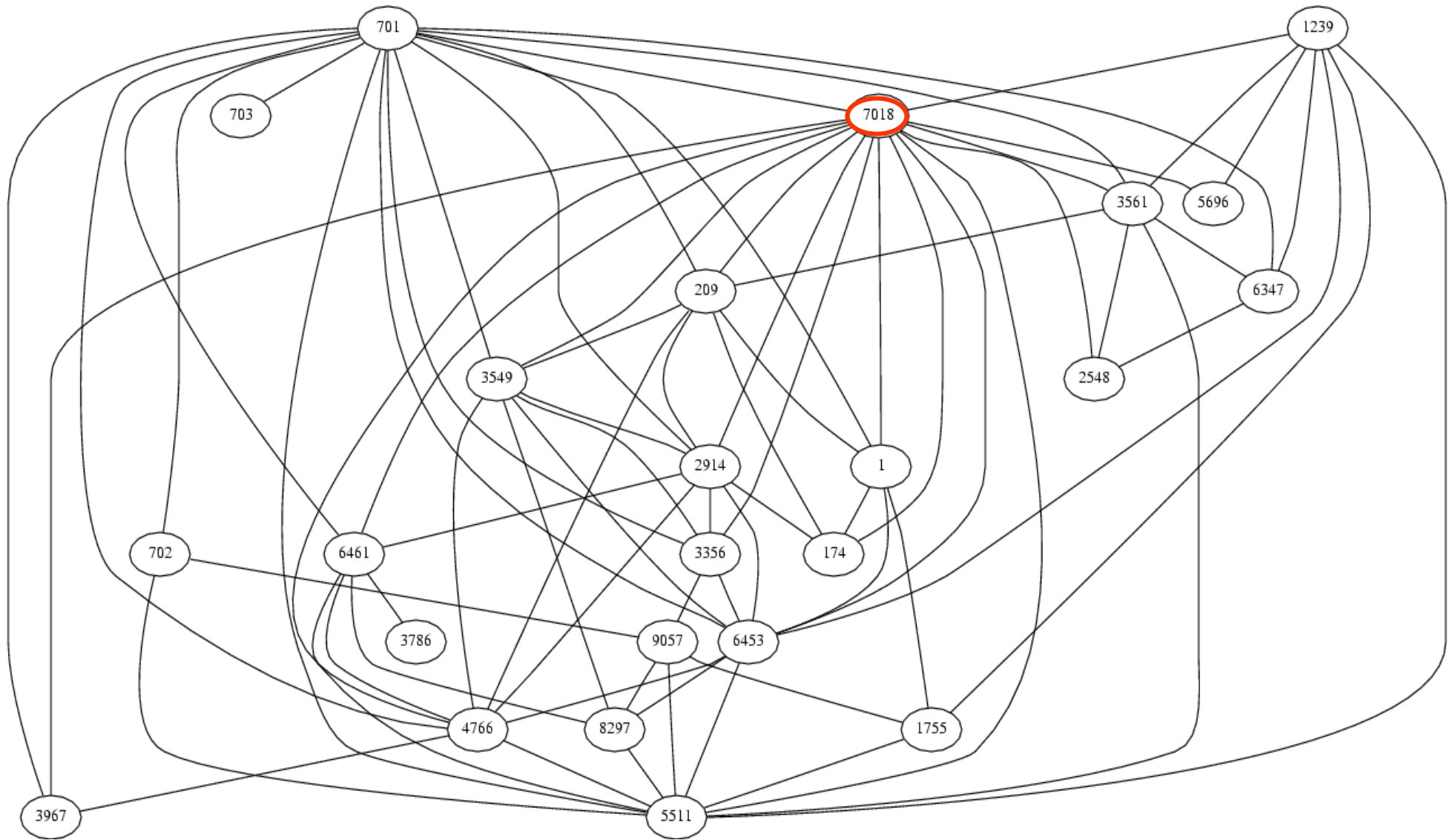
- Τα EGP πρωτόκολλα όπως το BGP αποκρύβουν την πραγματική τοπολογία

Όπως φαίνεται από το BGP



Στην πραγματικότητα...

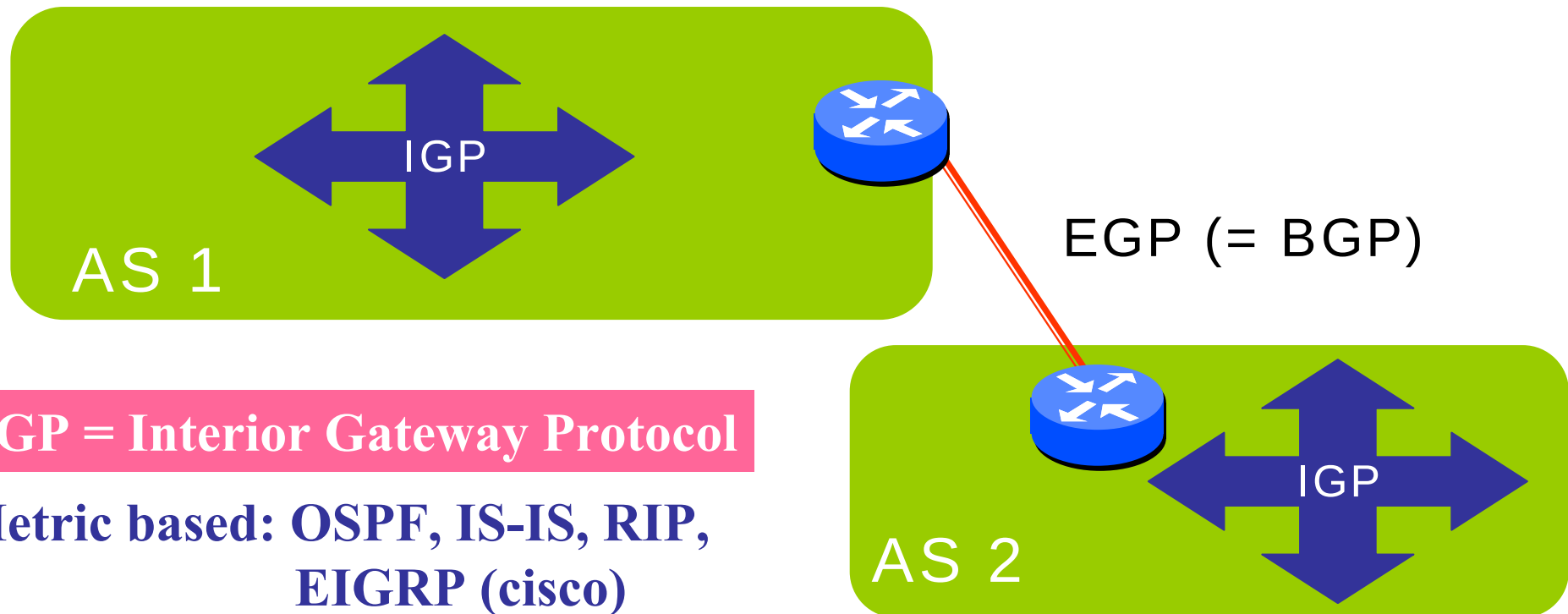
Οι γράφοι των αυτόνομων συστημάτων μπορεί να είναι πολύπλοκοι



Κατηγορίες Πρωτοκόλλων Δρομολόγησης

Υπάρχουν δύο κατηγορίες πρωτοκόλλων:

- **IGP**, Interior Gateway Protocol
- **EGP**, Exterior Gateway Protocol



IGP = Interior Gateway Protocol

**Metric based: OSPF, IS-IS, RIP,
EIGRP (cisco)**

EGP = Exterior Gateway Protocol

Policy based: BGP

IGP, Interior Gateway Protocol

- Συνδέει τους δρομολογητές σε ένα **αυτόνομο σύστημα**
- Μεταφέρει πληροφορία σχετικά με τα **προθέματα** που χρησιμοποιούνται στο **AS**
- Παραδείγματα είναι το **OSPF, ISIS, EIGRP**
- Ο σκοπός της σχεδίασης είναι η **ελαχιστοποίηση** του αριθμού των προθεμάτων ώστε να είναι πιο εύκολη η **κλιμάκωση** και η **σύγκλιση** του δικτύου
- Αυτόματη **ανακάλυψη** των γειτόνων
- Γενικά οι IGP δρομολογητές είναι **εμπιστέψιμοι** κάτι που είναι λογικό αφού ανήκουν όλοι στην ίδια αρχή διαχείρισης
- Τα προθέματα διαδίδονται σε **όλους** τους IGP δρομολογητές

EGP, Exterior Gateway Protocol

- Συνδέει όλα τα **AS**
- Αναφέρεται στις διαδικασίες δρομολόγησης μεταξύ δρομολογητών που βρίσκονται στα άκρα διαφορετικών AS συνεδμεμένων μεταξύ τους
- Τα EGP είναι **ανεξάρτητα** από την τοπολογία των εσωτερικών δικτύων
- Θέτει **διαχειριστικά όρια**
- Το πρώτο πρωτόκολλο αυτού του τύπου ήταν το **EGP** το οποίο έχει σοβαρές **αδυναμίες**, η σοβαρότερη είναι ότι πρόκειται για ένα πρωτόκολλο ευρεσης της προσιτότητας παρά δρομολόγησης
- Το τρέχων EGP είναι το **BGP**
- Δημιουργήθηκε για χρήση στο **διαδίκτυο**
- Αντίθετα με το EGP σχεδιάστηκε να ανιχνεύει **βρόχους** στο δίκτυο
- Μπορούμε να το δούμε σαν το EGP της **επόμενης γενιάς**
- Το BGP **αντικαθιστά** αργά αλλά σταθερά άλλα EGP πρωτόκολλα στο δίκτυο

EGP, Exterior Gateway Protocol

Ένα EGP πρωτόκολλο βοηθά τη:

- **Κλιμάκωση** ενός μεγάλου δικτύου
 - Ιεράρχιση
 - Περιορίζει την έκταση βλάβης
- **Πολιτική** δρομολόγησης
 - Έλεγχος του που διαδίδονται τα προθέματα
 - Συνένωση διαφορετικών οργανισμών
 - Σύνδεση περισσότερων IGP's

Τύποι Πρωτοκόλλων Δρομολόγησης

Κατάστασης συνδέσμων (Link state)

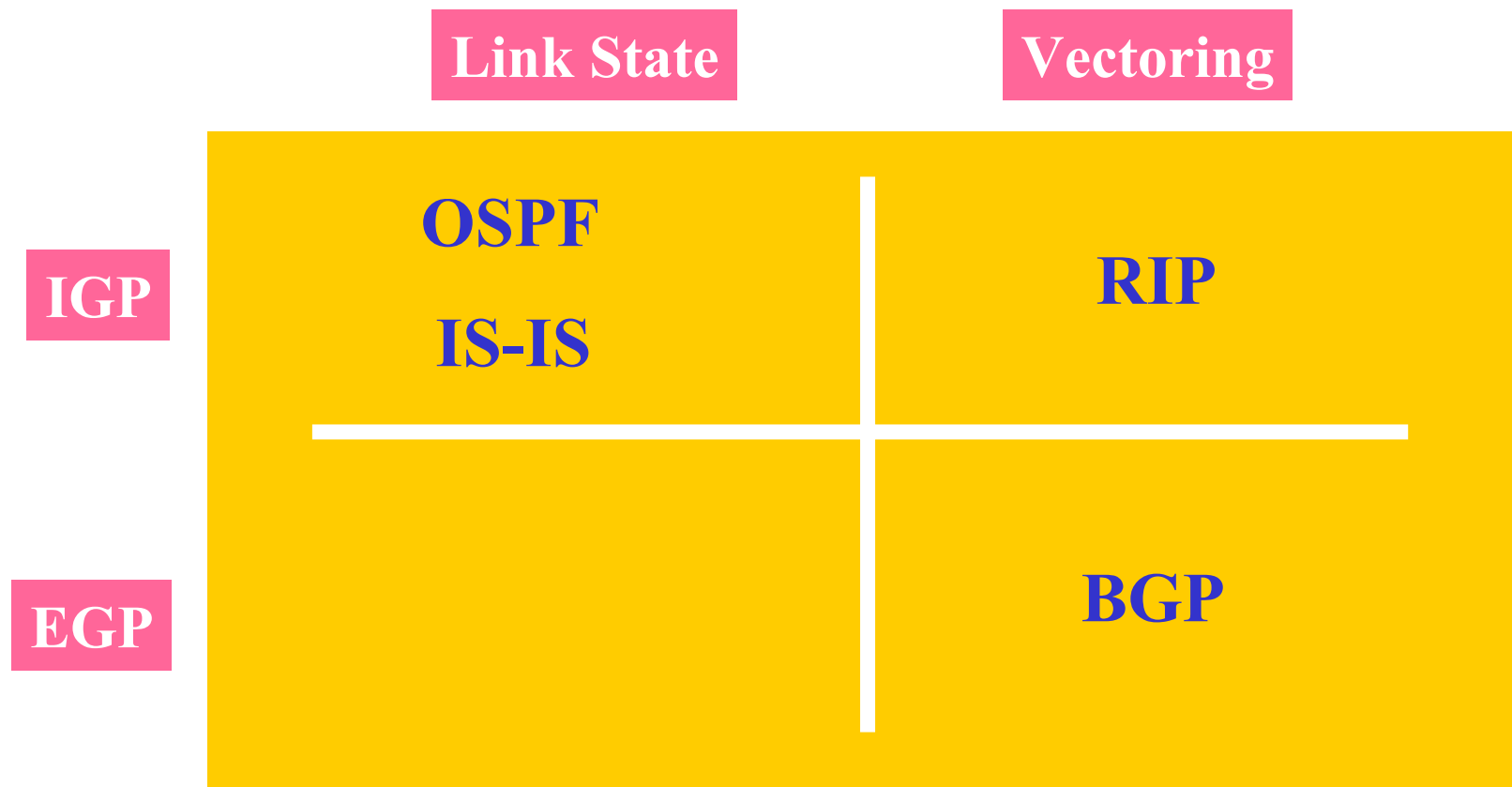
- Η πληροφορία τοπολογίας διαχέεται εντός του κράτους δρομολόγησης
- Οι βέλτιστες διαδρομές από άκρο σε άκρο **υπολογίζονται τοπικά** σε κάθε δρομολογητή
- Οι βέλτιστες διαδρομές καθορίζουν και το **επόμενο βήμα** (next hop)
- Βασίζονται στην **ελαχιστοποίηση** κάποιας παραμέτρου απόστασης
- Λειτουργούν μόνο αν η τακτική δρομολόγησης (policy) είναι ομοιόμορφη
- Παραδείγματα είναι τα **OSPF, IS-IS**

Τύποι Πρωτοκόλλων Δρομολόγησης

Διανυσματικά (Vectoring)

- Κάθε δρομολογητής γνωρίζει ελάχιστα για την **τοπολογία** δικτύου
- Μόνο τα βέλτιστα **επόμενα βήματα** επιλέγονται για κάθε δρομολογητή για κάθε προορισμό
- Οι βέλτιστες από άκρο σε άκρο **διαδρομές** παράγονται σαν **σύνθεση** όλων των επιλογών για το επόμενο βήμα
- Δεν απαιτεί κάποιο παράμετρο απόστασης
- Δεν απαιτεί **ομοιόμορφη τακτική** δρομολόγησης σε όλους τους δρομολογητές
- Παραδείγματα είναι τα **RIP, BGP**

Τα τέσσερα πιο σημαντικά πρωτόκολλα



Εισαγωγή στο BGP

BGP, Border Gateway Protocol, βασικά χαρακτηριστικά

- Τύπου **EGP** πρωτόκολλο δρομολόγησης ανάμεσα σε **αυτόνομα συστήματα** (interautonomous system routing protocol)
- Πρωτόκολλο **τακτικής** (Policy Based)
- Είναι το **defacto EGP** πρωτόκολλο στο **διαδίκτυο** (και **μοναδικό**)
- Σχετικά **απλό** σε λειτουργία
- **Πολύπλοκο** σε διάρθρωση
- Απαιτεί προσεκτική **σχεδίαση** αφού όλος ο κόσμος μπορεί να δει και να επηρεαστεί από τυχόν **λάθη** μας
- **Στιβαρό**

BGP, Border Gateway Protocol, βασικά χαρακτηριστικά

- Επιτρέπει εύκολη **κλιμάκωση** (περισσότερες από 100,000 διαδρομές στους πίνακες δρομολόγησης του BGP στο διαδίκτυο)
- Είναι το μόνο πρωτόκολλο που μπορεί να χρησιμοποιηθεί **ανάμεσα** σε **AS** αφού έχει **εγγενή** υποστήριξη σε τακτική δρομολόγησης
- Υποστηρίζει **classless** δρομολόγηση
- Τρέχει σε περισσότερους από **100K** δρομολογητές
- Η **δυναμική** και οι **αρχές κλιμάκωσης** του δεν έχουν γίνει ακόμα κατανοητές

BGP, ιστορία

BGP-1 [RFC 1105]

- 1989
- Αντικαταστάτης του EGP (1984, RFC 904)

BGP-2 [RFC 1163]

- 1990

BGP-3 [RFC 1267]

- 1991

BGP-4 [RFC 1771]

- 1995
- Κανείς δεν χρησιμοποιεί προηγούμενες εκδόσεις πλέον
- Συνεχής εξέλιξη: www.ietf.org/internet-drafts/draft-ietf-idr-bgp4-23.txt
- Πρόσθετες ιδιότητες ορίζονται με καινούργια RFC
- Υποστήριξη για Classless Interdomain Routing (CIDR)

Ποιο είναι το βασικό πρόβλημα που επιλύει το BGP;

- Το διαδίκτυο αποτελείται από πολύ **μεγάλο αριθμό** δικτύων
- Πολλά από αυτά έχουν κάποιου είδους **αστάθεια**, γνωστή ή άγνωστη στους διαχειριστές τους
- Συνέχεια στοιχεία του δικτύου αλλάζουν **λειτουργική κατάσταση**
- Κάτω από αυτές τις συνθήκες η **ευστάθεια** του πρωτοκόλλου γίνεται πρωταρχικός στόχος
- Το **BGP** σχεδιάστηκε έχοντας αυτό σαν **πρωταρχικό στόχο**
- Επίσης βασικός στόχος της σχεδίασης του ήταν να επιτρέπει στους πάροχους να παίρνουν αποφάσεις δρομολόγησης με βάση την **τακτική** που **επιθυμούσαν** ή ήταν **υποχρεωμένοι** να έχουν

Πως θα βρούμε τη
συντομότερη διαδρομή



RIP, OSPF, IS-IS

Πως θα βρούμε τη
σταθερότερη διαδρομή



BGP

BGP, λειτουργία

- Μαθαίνει πολλαπλές διαδρομές μέσω **εσωτερικών** και **εξωτερικών** BGP ομοτίμων
- Επιλέγει την **καλύτερη** διαδρομή και την εγκαθιστά στον πίνακα **προώθησης**
- Το BGP **δεν** παρέχει δρομολόγηση “**shortest path**”
- Η **βέλτιστη** διαδρομή στέλνεται στους γείτονες
- Πολιτική** δρομολόγησης εφαρμόζεται επηρεάζοντας την **επιλογή** του βέλτιστου μονοπατιού

BGP, τρόποι λειτουργίας

- Παρότι σχεδιάστηκε σαν πρωτόκολλο δρομολόγησης ανάμεσα σε **διαφορετικά AS**, μπορεί να χρησιμοποιηθεί **μέσα** και **ανάμεσα** στα AS
- Το πρωτόκολλο μένει το ίδιο, αλλάζουν απλώς οι **κανόνες** εγκατάστασης των διαδρομών

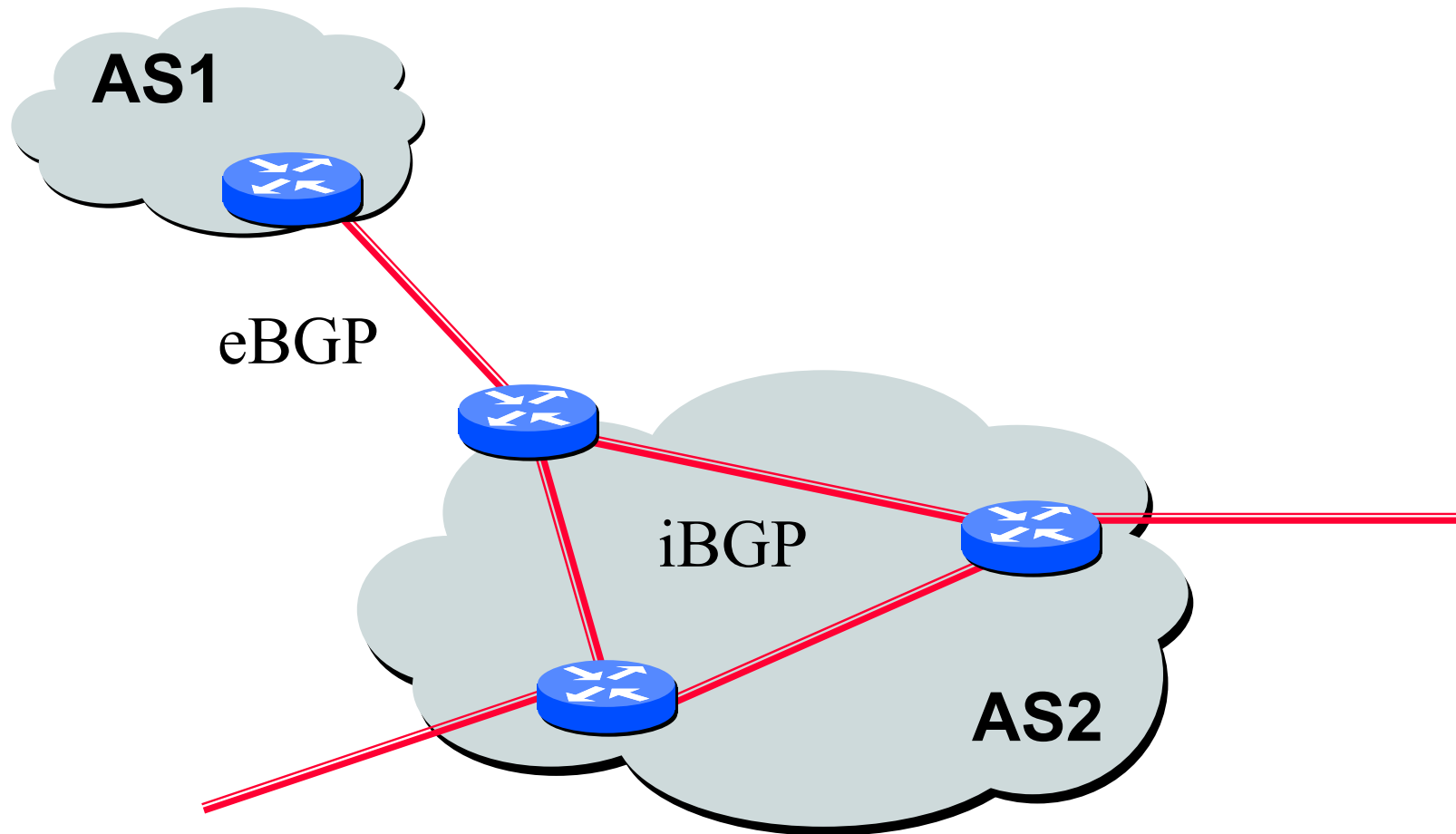
EBGP, External BGP

- Όταν το BGP χρησιμοποιείται **ανάμεσα** σε αυτόνομα συστήματα
- Υλοποιεί την εκάστοτε **τακτική** δρομολόγησης

IBGP, Internal BGP

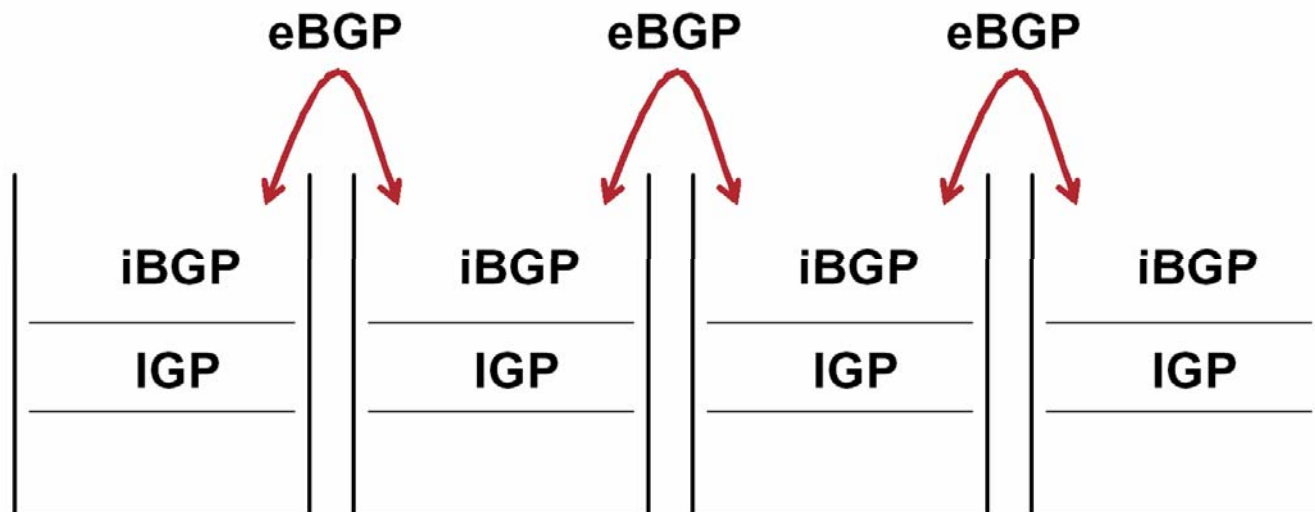
- Αν ένας πάροχος χρησιμοποιεί BGP για την ανταλλαγή διαδρομών **εντός** του AS

IBGP, EBGP

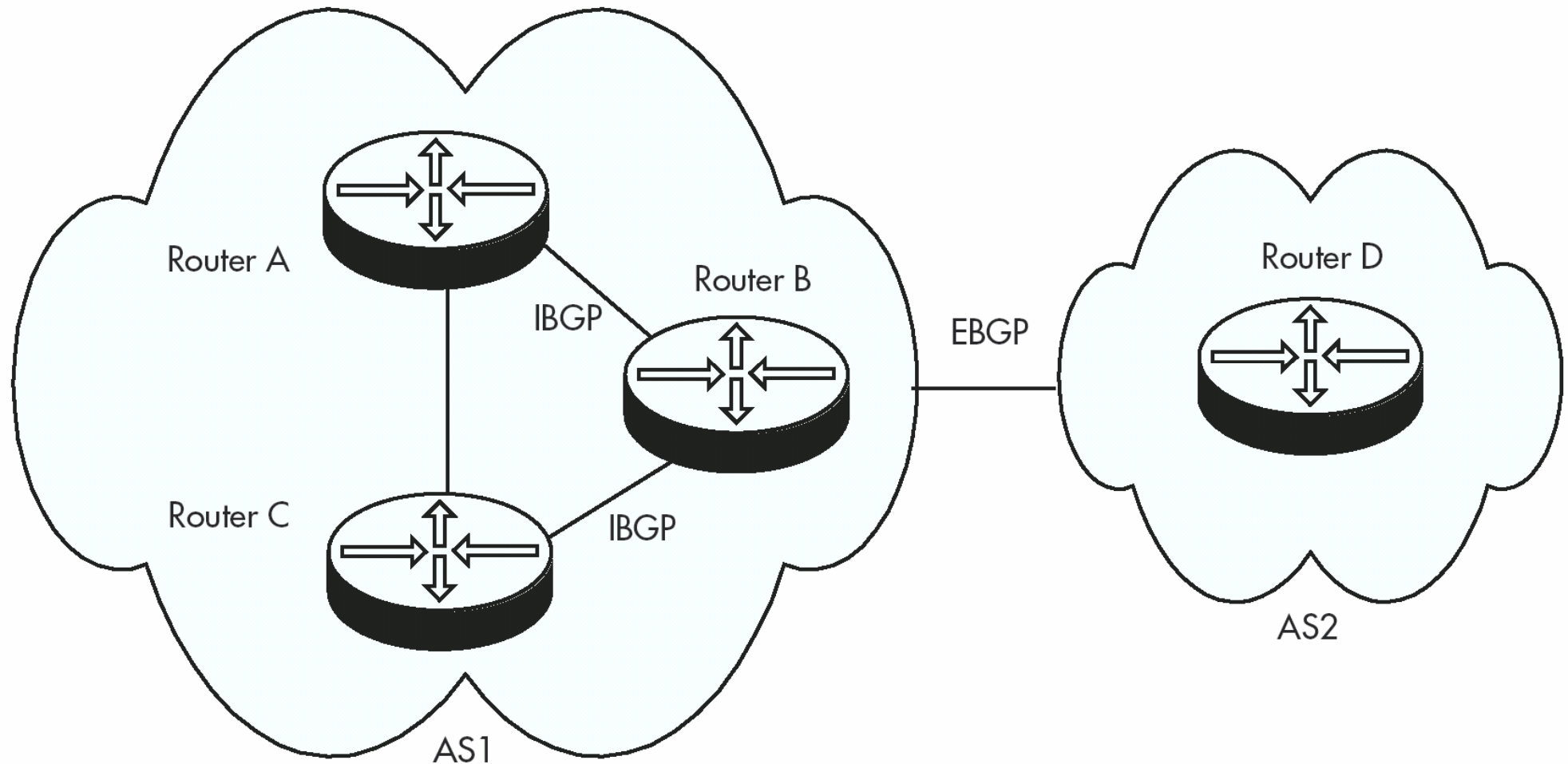


IBGP, EBGP

- Το **EBGP** αναλαμβάνει τη δρομολόγησης ανάμεσα στα AS
- Είναι απαραίτητο τα δίκτυα **εντός** του AS να μπορεί να προσεγγιστούν πριν στείλουμε την πληροφορία προς άλλα AS
- Τούτο επιτυγχάνεται με συνδυασμό του **IBGP** εντός του AS και **αναδιανομής** της BGP πληροφορίας προς τα IGP πρωτόκολλα που τρέχουν στο AS

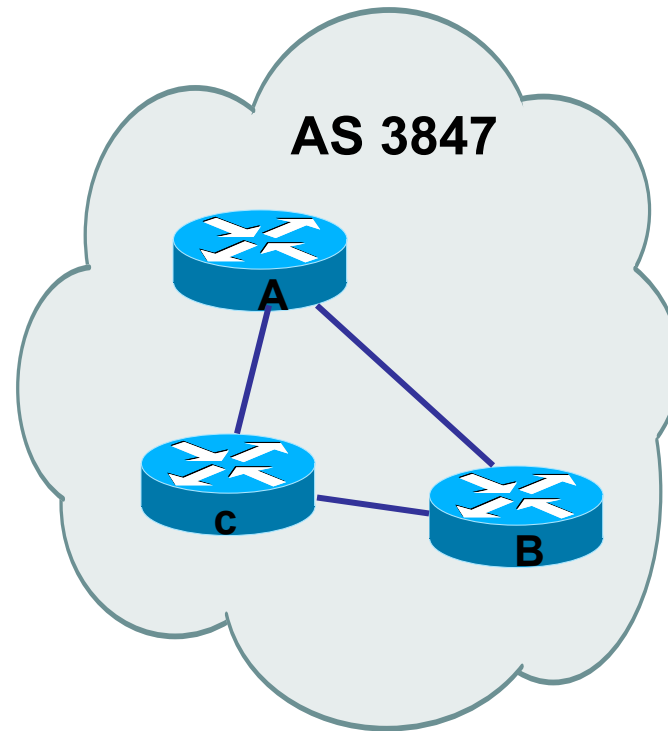


Stub AS



- Ένα αυτόνομο σύστημα με **ένα** BGP δρομολογητή και **μία EBGP** σύνδεση αναφέρεται σαν **στέλεχος** (stub AS)

IBGP, Internal BGP



IBGP, Internal BGP

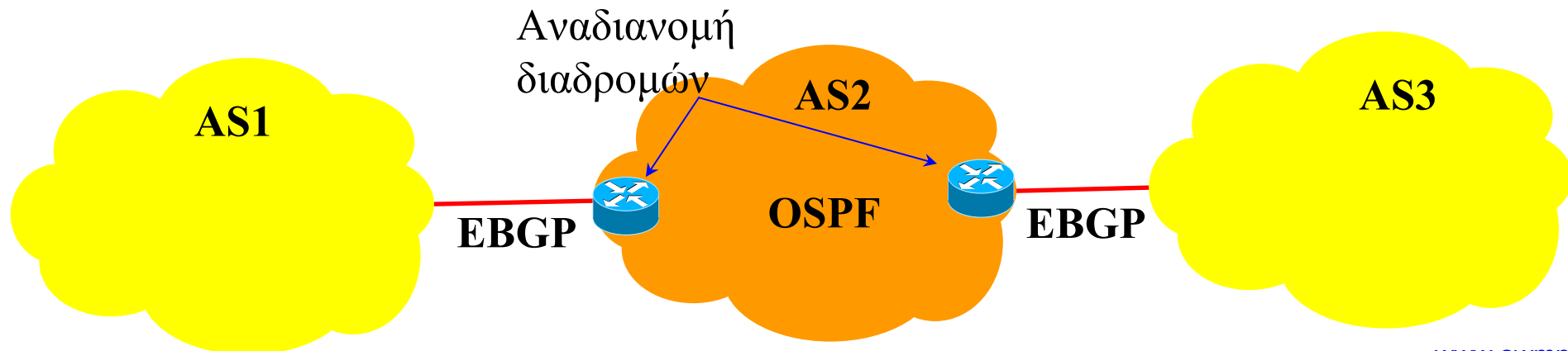
- Οι **BGP** ομότιμοι βρίσκονται στο **ίδιο AS**
- Δεν απαιτείται να είναι **κατευθείαν** συνεδεδεμένοι, ένα **IGP** μπορεί να αναλάβει την **IP σύνδεση** μεταξύ τους
- Το IBGP χρησιμοποιείται (είναι βολικό) όταν ένα AS χρησιμοποιείται σαν σύστημα **διέλευσης** (transit) προς άλλα AS
- Επιτρέπει την επιλογή του καλύτερου σημείου εξόδου από το AS
- Το IBGP **δεν είναι ένα IGP**, μπορεί να χρησιμοποιεί IGP για τη σύνδεση μεταξύ των IBGP ομοτίμων
- Οι BGP δρομολογητές εντός του ίδιου AS επικοινωνούν μεταξύ τους για να διατηρούν μια **κοινή αντίληψη** του AS και για να αποφασίσουν ποιος από όλους θα εξυπηρετεί σαν το **σημείο σύνδεσης** προς και από άλλα εξωτερικά AS
- Κάποια AS περνάνε διαφανώς την κίνηση διαμέσου τους. Έτσι θα πρέπει το BGP να συνεργάζεται με οποιοδήποτε **IGP** πρωτόκολλο χρησιμοποιείται σε αυτά τα AS

Γιατί IBGP και όχι κάποιο IGP;

- Το ίδιο μπορεί να επιτευχθεί με **αναδιανομή** του EBGP σε κάποιο IGP και μετά **πάλι αναδιανομή** σε άλλο AS

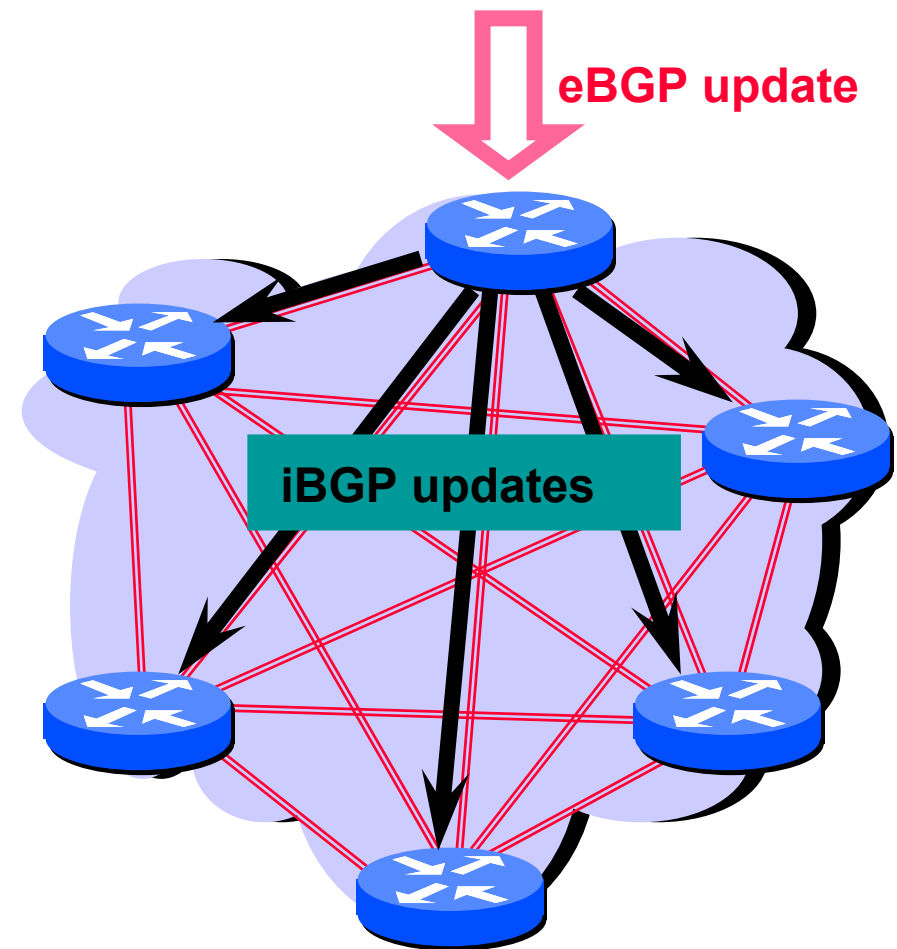
Όμως με το IBGP:

- Μεγαλύτερη **ευελιξία** και έχει **αποτελεσματικούς** τρόπους για την ανταλλαγή της πληροφορίας δρομολόγησης εντός του AS
- Για παράδειγμα με το IBGP μπορούμε να **ελέγξουμε** ποιο θα είναι το καλύτερο σημείο εξόδου από το AS (local preference)
- Εισάγωντας εξωτερικές διαδρομές εντός του IGP, χάνουμε την ικανότητα **κλιμάκωσης**, όπως και την πληροφορία σχετικά με την BGP **τακτική** δρομολόγησης



IBGP, όλοι με όλους

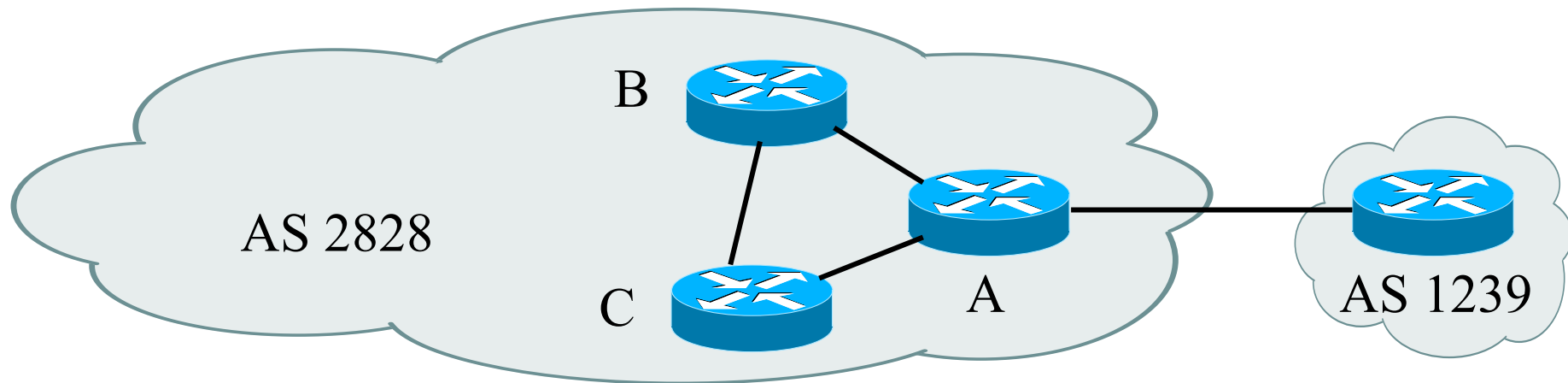
- Ένας IBGP ομότιμος μπορεί μόνο να στείλει διαδρομές που **εισάγει** (τα δίκτυα με τα οποία είναι κατευθείαν συνεδεμένοι) ή διαδρομές που μαθαίνει **κατευθείαν** από **εξωτερικούς** ομότιμους
- Έτσι δεν μπορεί να αναγγείλει διαδρομές που έμαθε από **IBGP** ομότιμο σε άλλο **IBGP** ομότιμο
- Άρα οι IBGP ομότιμοι πρέπει να είναι συνεδεμένοι **όλοι με όλους** σε τοπολογία **πλέγματος** (mesh)



IBGP, όλοι με όλους

Παράδειγμα

- Το AS1239 αναγγέλει τη διαδρομή 10.0.0.0/8 στο AS2828
- Ο A θα στείλει τη διαδρομή αυτή στους B, C
- Ο B όμως δεν θα την διαδώσει στον C και ούτε ο C στον B, διότι αυτές έγιναν γνωστές από έναν **IBGP** ομότιμο

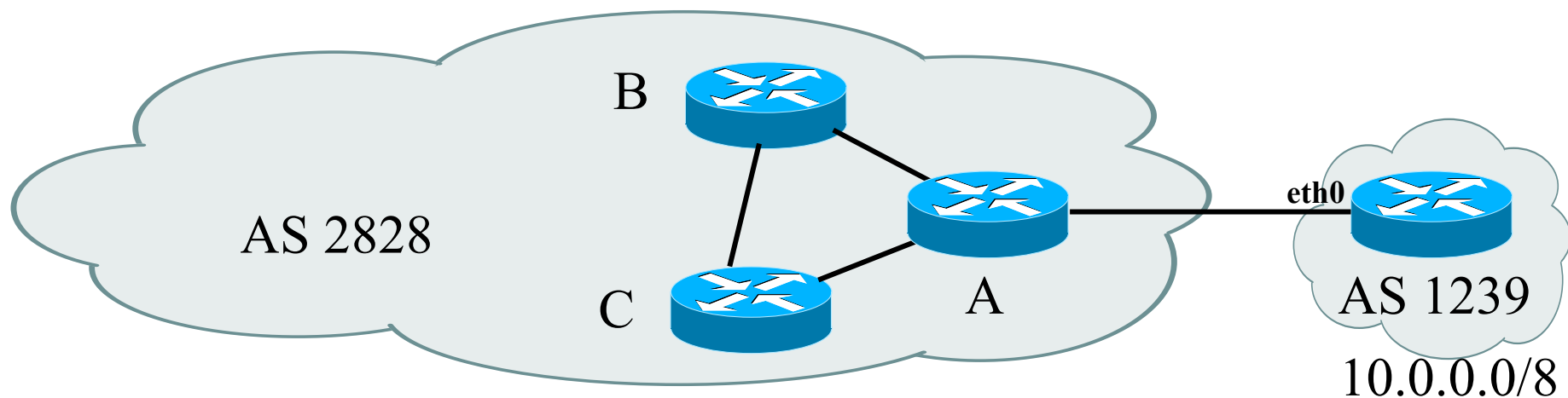


IBGP, λειτουργία

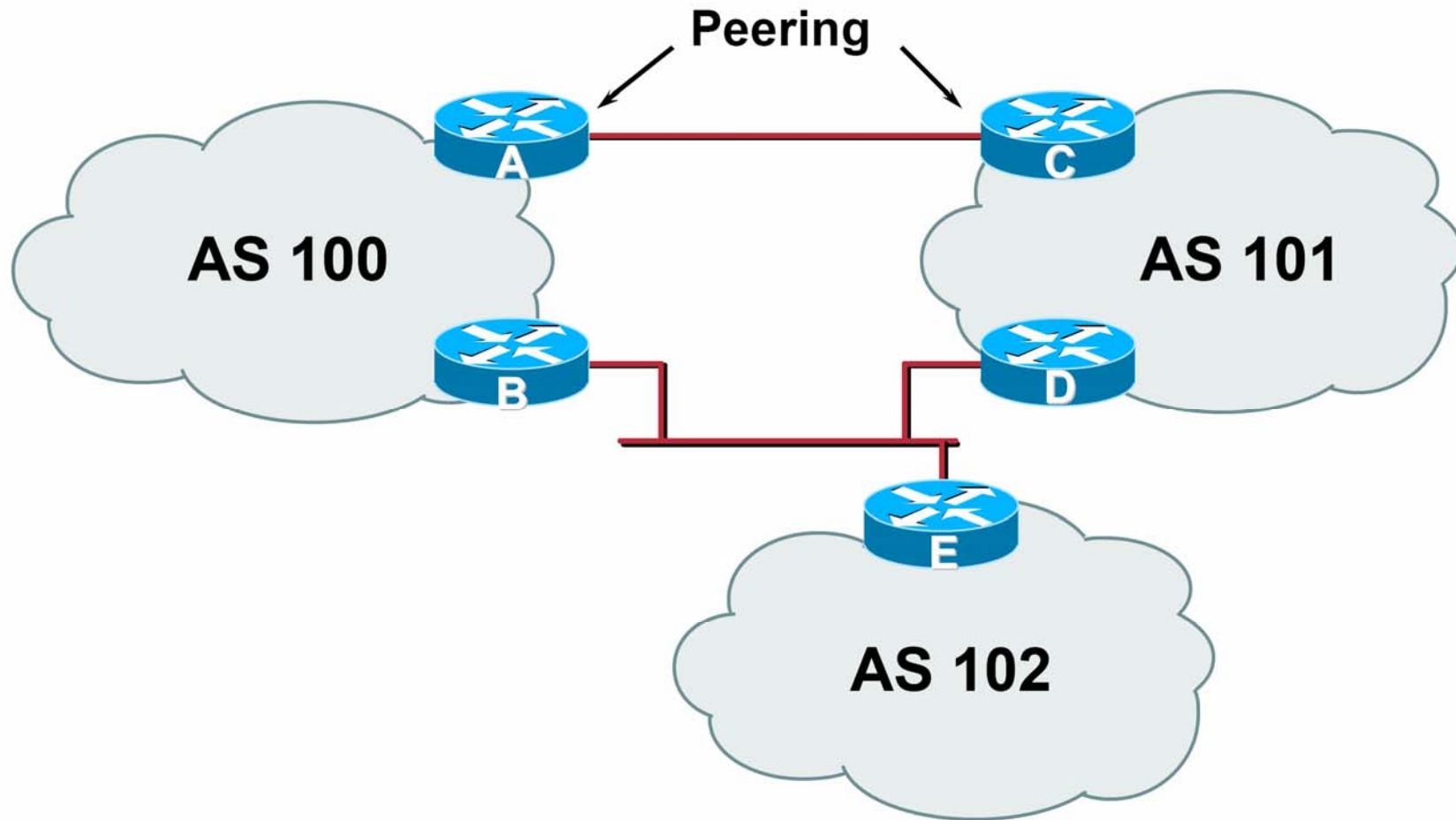
- Είναι δυνατό να αναγγελθούν πάνω από **IBGP** διαδρομές που έγιναν γνωστές με **αναδιανομή** από άλλα πρωτόκολλα **OSPF, RIP**
- Είναι δυνατό να αναγγελθούν διαδρομές που έγιναν γνωστές από **IBGP** σε μια **EBGP** σύνδεση
- Αποστολή **όλων των ιδιοτήτων** των διαδρομών
- Οι ιδιότητες **AS_path, NextHop** μεταδίδονται **διαφανώς**

IBGP, NextHop

- Το NextHop για το 10.0.0.0/8 θα είναι η διεπαφή eth0 στο AS1239, ακόμα και για τους B, C
- Στο IBGP το NextHop δεν είναι κατευθείαν συνεδεδμένο στο δρομολογητή
- Έτσι χρειάζεται **αναδρομική** έρευνα (recursive lookup) στον πίνακα δρομολόγησης
- Έτσι για προορισμό στο 10.0.0.0/8 ο B και ο C βρίσκουν μια εγγραφή που δείχνει το NextHop
- Στη συνέχεια πρέπει να βρουν μια εγγραφή για προορισμό το NextHop

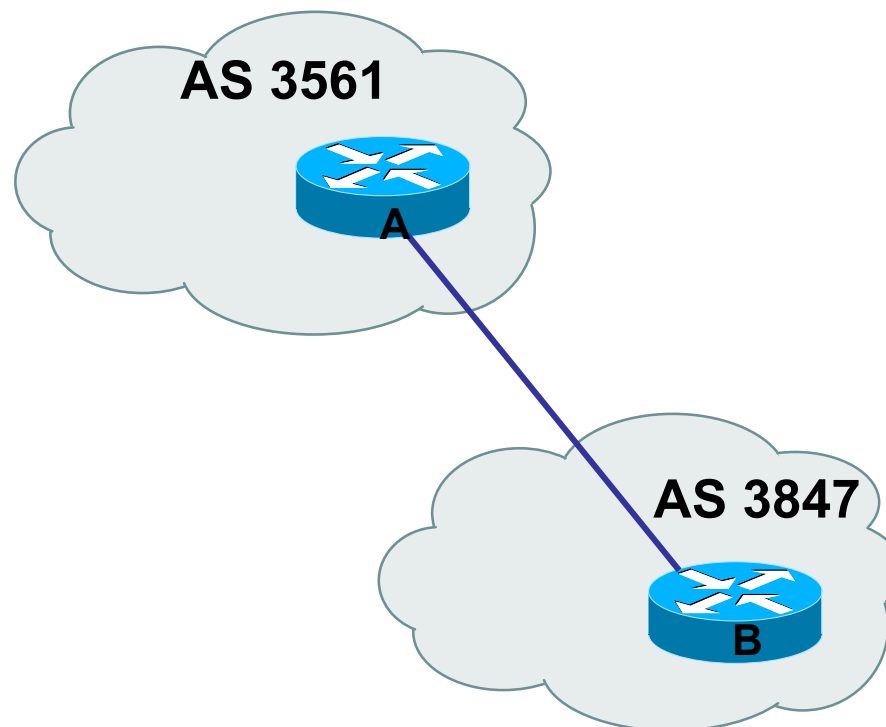


EBGP, External BGP



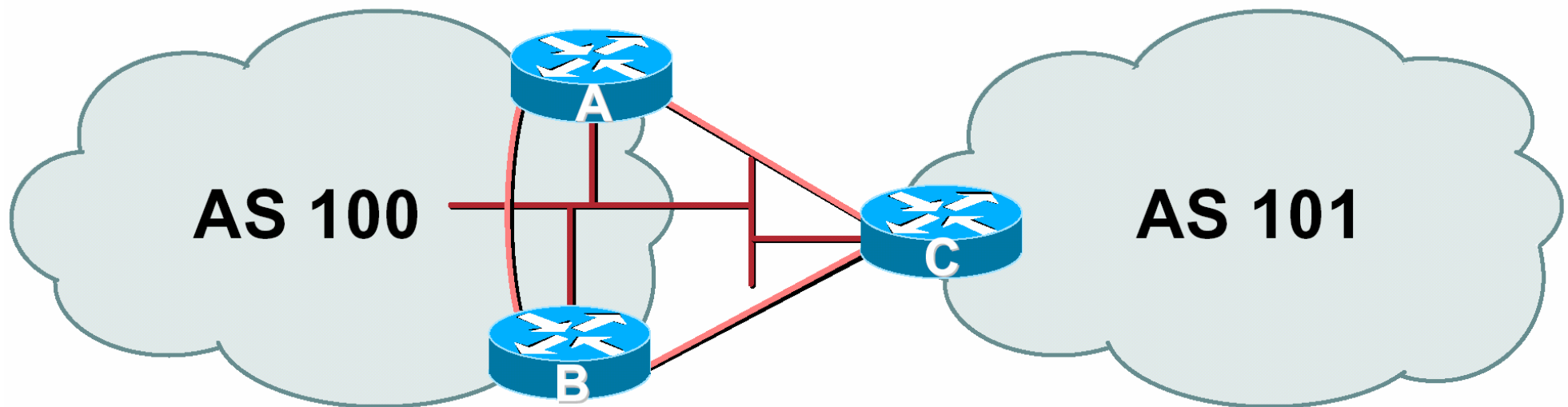
EBGP, External BGP

- Οι ομότιμοι βρίσκονται σε **διαφορετικά AS**
- Οι ομότιμοι πρέπει να είναι **κατευθείαν συνεδεδμένοι** δηλαδή πρέπει να βρίσκονται στο ίδιο **φυσικό δίκτυο** (εξαίρεση το Multihop)
- **Ποτέ δεν πρέπει να τρέχει κάποιο IGP** ανάμεσα σε ομότιμους EBGP



EBGP, External BGP

- Αποστολή της **καλύτερης** διαδρομής για κάθε προορισμό
- Αποστολή όλων των **σημαντικών** ιδιοτήτων, εκτος του **local preference**
- Προσθέτει το ASN του ομιλητή στην **AS_path** λίστα
- Συνήθως ξαναγράφει την ιδιότητα **NextHop**



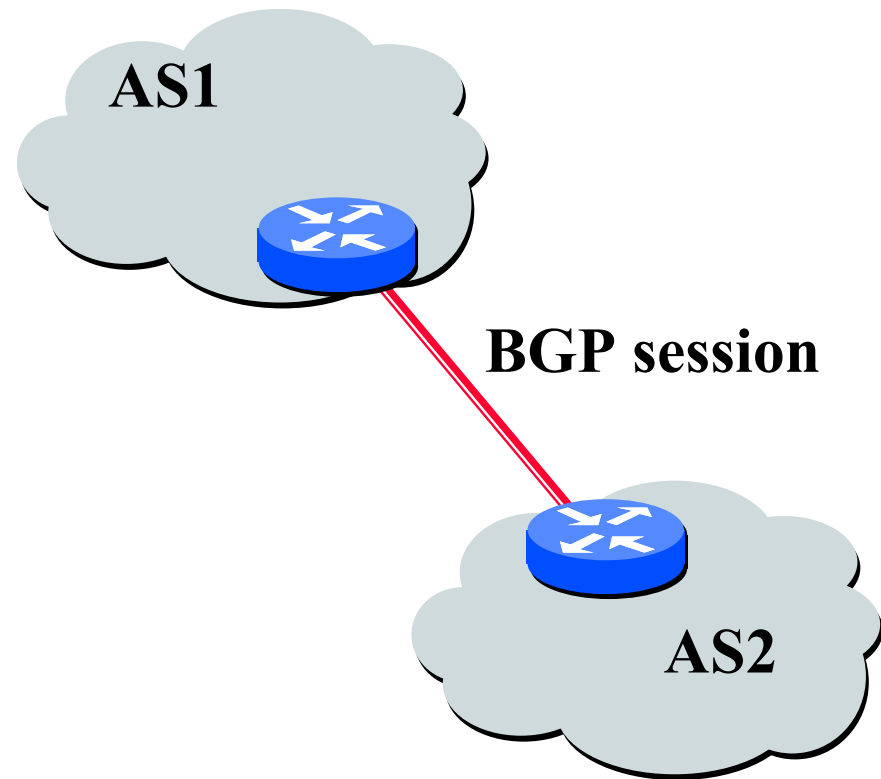
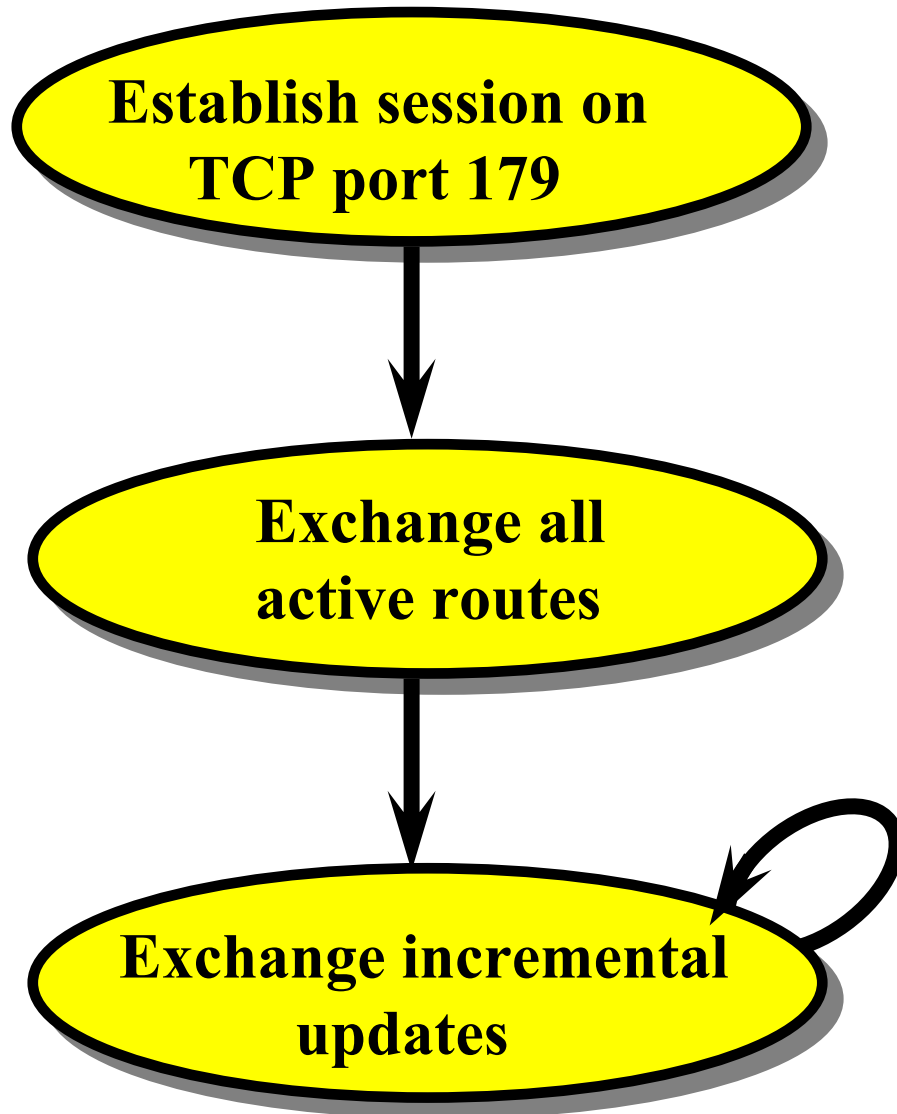
IBGP, EBGP και ιδιότητες

Attribute	EBGP	IBGP
ORIGIN	Mandatory	Mandatory
AS_PATH	Mandatory	Mandatory
NEXT_HOP	Mandatory	Mandatory
MULTI_EXIT_DISC	Discretionary	Discretionary
LOCAL_PREF	Not Allowed	Required
ATOMIC_AGGREGATE	Discretionary	Discretionary
AGGREGATOR	Discretionary	Discretionary
COMMUNITY	Discretionary	Discretionary



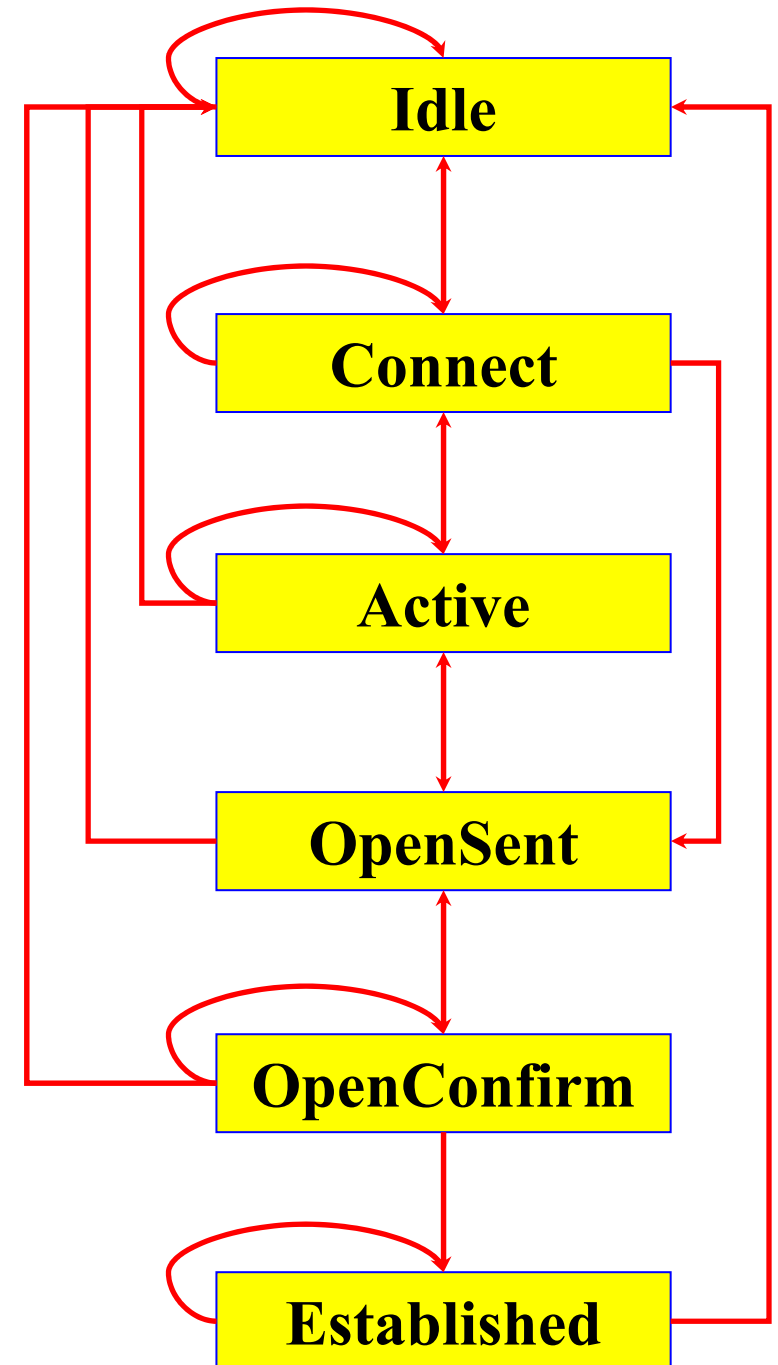
Λειτουργία

BGP, Λειτουργία

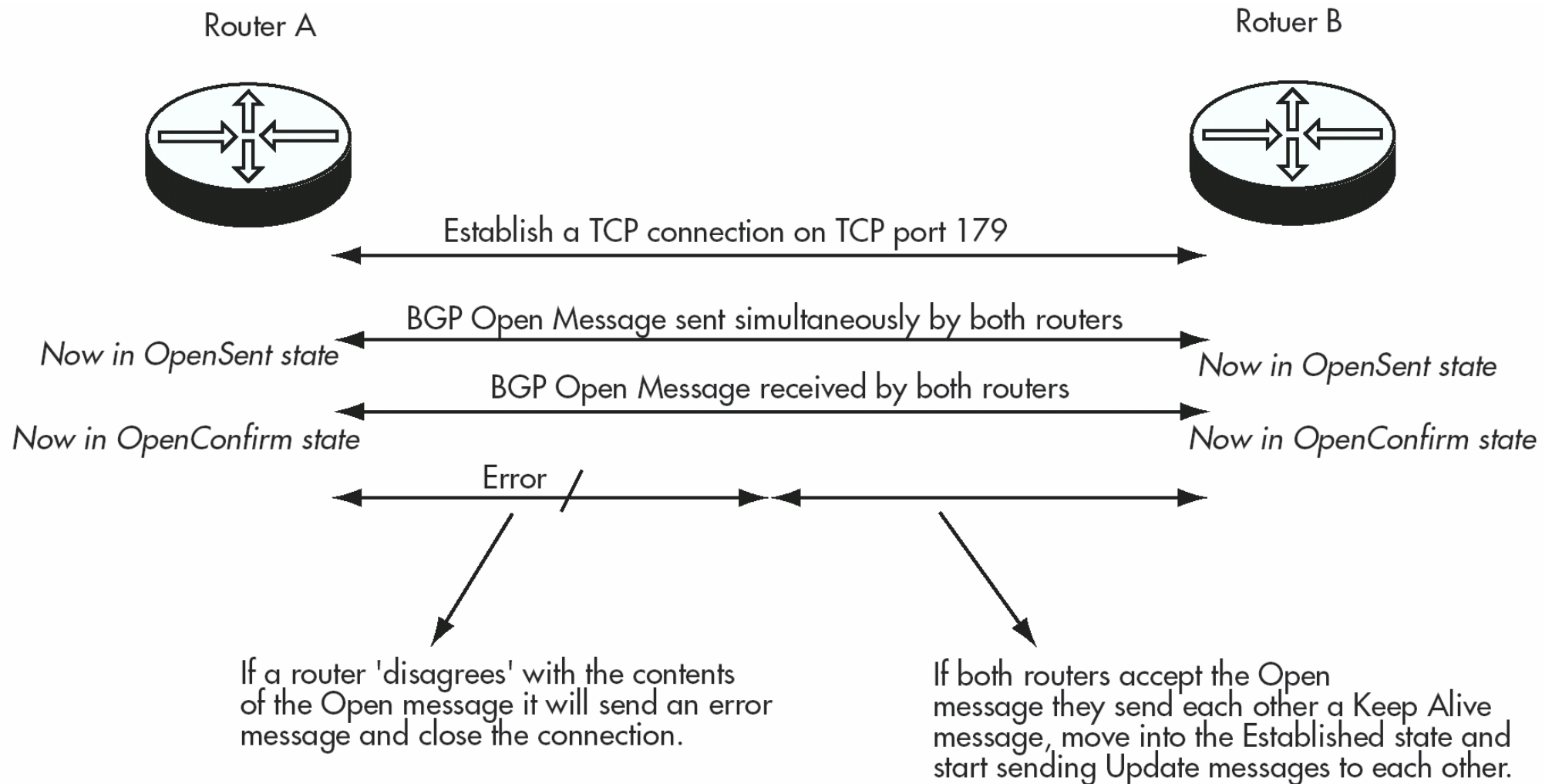


Μηχανή καταστάσεων (State machine)

- Ορίζεται μία **μηχανή καταστάσεων** σχετικά με τη ίδρυση, διατήρηση και κατάργηση των **BGP συνόδων**
- Αυτή είναι ένα μοντέλο μιας μηχανής που ορίζονται καταστάσεις στις οποίες μπορεί να βρεθεί η μηχανή
- Γεγονότα όπως **μετρητές χρόνου** ή **εξωτερικοί είσοδοι** προκαλούν **μεταβάσεις** ανάμεσα στις καταστάσεις



και ανταλλάσσονται μηνύματα...



Μετρητές χρόνου

- Ορίζονται σε δευτερόλεπτα
- Οι υλοποιήσεις του BGP επιτρέπουν την αλλαγή των παραμέτρων αυτών

ConnectRetry

- Πόσο συχνά γίνεται προσπάθεια για TCP σύνδεση
- Προτεινόμενη τιμή τα 120'
- Χρησιμοποιείται κατά την αρχικοποίηση της BGP συνόδου

Hold Time

- Πόσος χρόνος μπορεί να περάσει χωρίς λήψη μηνύματος Update ή KeepAlive
- Μετά την πάροδο του χρόνου ο ομότιμος θεωρείται εκτός λειτουργίας (εξ' ορισμού 180', προτεινόμενη τα 90')

KeepAlive

- Παράγονται μηνύματα KeepAlive κάθε 1/3 του χρόνου Hold Time (εξ' ορισμού 60', προτεινόμενη τα 30')

MinASOriginationInterval

- Πόσο συχνά στέλνονται οι εσωτερικές αλλαγές του AS, προτεινόμενη τα 15'

MinRouteAdvertisementInterval

- Πόσο συχνά στέλνονται οι αλλαγές της ίδιας διαδρομής, προτεινόμενη τιμή τα 30'

- Οι δύο τελευταίοι μετρητές είναι εσωτερικοί σε κάθε BGP υλοποίηση και αποσκοπούν να μειώσουν τη διάδοση **γρήγορων** μεταβολών οι οποίες προκαλούν **αστάθεια** στο πρωτόκολλο

Μηχανή καταστάσεων (State machine)

- Ξεκινάμε από την κατάσταση **Idle**
- Δημιουργείται ανάμεσα στους BGP γείτονες μια **TCP σύνδεση** στην θύρα **179** (πρέπει να υπάρχει IP συνδεσιμότητα προφανώς είτε λόγω κατευθείαν φυσικής σύνδεσης είτε μέσω κάποιου IGP πρωτοκόλλου)
- Αυτό εξασφαλίζει την **αξιόπιστη επικοινωνία**
- Περισσότερα του ενός BGP μηνύματα μπορούν να σταλούν στο ίδιο TCP τμήμα
- Οι συνομιλούντες δρομολογητές **στέλνουν** μηνύματα **Open** ο ένας στο άλλον δίνοντας την ταυτότητα τους, ώστε να **αρχικοποιήσουν** και **επιβεβαιώσουν** τις **παραμέτρους σύνδεσης** και μεταβαίνουν στην κατάσταση **OpenSend**

Μηχανή καταστάσεων (State machine)

- Όταν κάποιος λάβει το **Open** μήνυμα μεταβαίνει στην κατάσταση **OpenConfirm**
- Ακολουθεί **επεξεργασία** των περιεχομένων του **Open** μηνύματος και αν κάποιος από τους δύο **διαφωνήσει** στέλνει πίσω ένα **Notification** μήνυμα εξηγώντας τους λόγους και η σύνοδος τερματίζεται - κατάσταση **Idle**
- Αν δεχτεί τη σύνδεση στέλνει πίσω ένα **KeepAlive** μήνυμα και μεταβαίνει στην κατάσταση **Established**
- Σε αυτή την κατάσταση λέμε ότι η **BGP σύνοδος** είναι **ανοικτή** και οι δύο ομότιμοι μπορούν να ανταλλάσσουν μηνύματα **Update**, **KeepAlive**, **Notification**
- Οι δύο συνομιλούντες δρομολογητές ονομάζονται τότε **ομότιμοι** (peers) ή και γείτονες
- Οι ομότιμοι πρέπει να **οριστούν** από την αρχή και αυτό είναι θετικό

Μηχανή καταστάσεων (State machine)

- Αν ένα **λάθος** συμβεί ενώ η σύννοδος είναι ανοικτή ο δρομολογητής στέλνει ένα **Notification** μήνυμα **διευκρινίζοντας** το λάθος και **τερματίζει** την **TCP** σύνδεση και μεταβαίνει στην **Idle** κατάσταση
- Το ίδιο θα συμβεί αν διακοπεί η φυσική σύνδεση οπότε τα **KeepAlive** μηνύματα ή/και η TCP σύννοδος θα **εκπνεύσουν**
- Σε αυτή την κατάσταση ο δρομολογητής σταματά να χρησιμοποιεί **όλη** την πληροφορία δρομολόγησης που έμαθε από τον ομότιμο του
 1. Όλες οι εγγραφές που έγιναν γνωστές από αυτόν σημαδεύονται σαν μην έγκυρες
 2. Αποσύρονται οι συγκεκριμένες εγγραφές από όλους τους ομότιμους
 3. Διαγράφονται από τον πίνακα

Καταστάσεις δρομολογητή

Idle

- Αρχική κατάσταση
- Απόρριψη όλων των εισερχομένων BGP συνδέσεων και μη ανάθεση πόρων στους γείτονες
- Ένα λάθος ή μία clear εντολή γυρνάει τη μηχανή σε αυτή την κατάσταση

RS# **bgp show summary**

Local router ID is 192.168.0.1, Local AS number 65030

BGP Route Entries 0, Unique AS Paths 2

Unique Communities 0, Unique Extended Communities 0

Neighbor	V	AS	MsgRcvd	MsgSent	Up/Down	Prefixes Rcvd/Sent
10.0.0.2	0	65015	0	0	Idle	

Connect

- Αναμονή για ολοκλήρωση της TCP σύνδεσης
- Σε περίπτωση επιτυχούς περάτωσης μεταβαίνει στην κατάσταση OpenSend διαφορετικά στην Active

Καταστάσεις δρομολογητή

Active

- Αρχικοποίηση TCP σύνδεσης
- Στην περίπτωση που έχουμε άλλα συμβάντα, ελευθερώνονται όλοι οι πόροι που έχουν σχέση με τη σύνδεση και η κατάσταση γίνεται Idle
- Αν δημιουργηθεί TCP σύνοδος, στέλνεται Open μήνυμα και μεταβαίνει στην κατάσταση OpenSend
- Αν ο μετρητής ConnectRetry εκπνεύσει, αρχικοποιείται η TCP σύνδεση και γίνεται μετάβαση στην κατάσταση Open
- Ταλάντωση ανάμεσα στις καταστάσεις Connect – Active σημαίνει μάλλον πρόβλημα στην **IP συνδεσιμότητα** ανάμεσα στους ομότιμους

RS# **bgp show summary**

Local router ID is 192.168.0.1, Local AS number 65030

BGP Route Entries 0, Unique AS Paths 2

Unique Communities 0, Unique Extended Communities 0

Neighbor	V	AS	MsgRcvd	MsgSent	Up/Down	Prefixes Rcvd/Sent
10.0.0.2	0	65015	0	0	Active	

Καταστάσεις δρομολογητή

OpenSend

- Υπάρχει TCP σύνδεση ανάμεσα στους ομοτίμους
- Έχει σταλεί μήνυμα Open
- Αναμονή για λήψη μηνύματος Open
- Σε περίπτωση λήψης και έγκρισης των περιεχομένων του μεταβαίνει στην OpenConfirm και επιβεβαιώνει με μήνυμα KeepAlive, διαφορετικά στέλνει μήνυμα Notification και μεταβαίνει στην Idle

OpenConfirm

- Έχει στείλει και λάβει μήνυμα Open και έχει στείλει KeepAlive
- Αναμονή για μήνυμα KeepAlive ή Notification
- Αν λάβει KeepAlive ματαβαίνει στην κατάσταση Established, διαφορετικά στην Idle

Ανίχνευση Διπλών TCP συνδέσεων

- Υπάρχει η **πιθανότητα** να δημιουργηθούν δύο TCP σύνοδοι ανάμεσα σε δύο ομότιμους
- Προβλέπεται ο μηχανισμός **Connection Collision Detection** ο οποίος κλείνει τη μία TCP σύνδεση
- Το κριτήριο είναι ότι κρατείται η σύνδεση η οποία αρχικοποιήθηκε από τον δρομολογητή με το **υψηλότερο Rid**

Διαπραγμάτευση έκδοσης του BGP

- Οι BGP ομότιμοι κάνουν προσπάθεια να δημιουργήσουν μια BGP σύνοδο, ξεκινώντας από την **υψηλότερη** έκδοση του πρωτοκόλλου
- Αν η προσπάθεια αποτύχει με Error Code= OPEN Message Error, Error Subcode=Unsupported Version Number, ξαναπροσπαθούν, γνωρίζοντας τις κοινές εκδόσεις που υποστηρίζουν και οι δύο, επιλέγοντας την υψηλότερη
- Πλέον σπανίως η **BGPv4** δεν είναι αυτή που χρησιμοποιείται

Καταστάσεις δρομολογητή

Established

- Το BGP βρίσκεται τώρα σε λειτουργική κατάσταση
- Μπορεί να ανταλλάξει μηνύματα Update, Notification, KeepAlive με τους ομότιμους
- Τα KeepAlive στέλνονται στο 1/3 του HoldTime, αν λάθη συμβούν στέλνονται Notification μηνύματα και γίνεται μετάβαση σε Idle
- **Οποιαδήποτε άλλη κατάσταση σημαίνει ότι οι δρομολογητές δεν είναι ομότιμοι και άρα δεν μπορούν να ανταλλάξουν ενημερώσεις**

RS# **bgp show summary**

Local router ID is 192.168.0.1, Local AS number 65030

BGP Route Entries 3, Unique AS Paths 3

Unique Communities 0, Unique Extended Communities 0

Neighbor	V	AS	MsgRcvd	MsgSent	Up/Down	Prefixes Rcvd/Sent
10.0.0.2	4	65015	5	5	0d0h2m0s	3/0

BGP, Λειτουργία

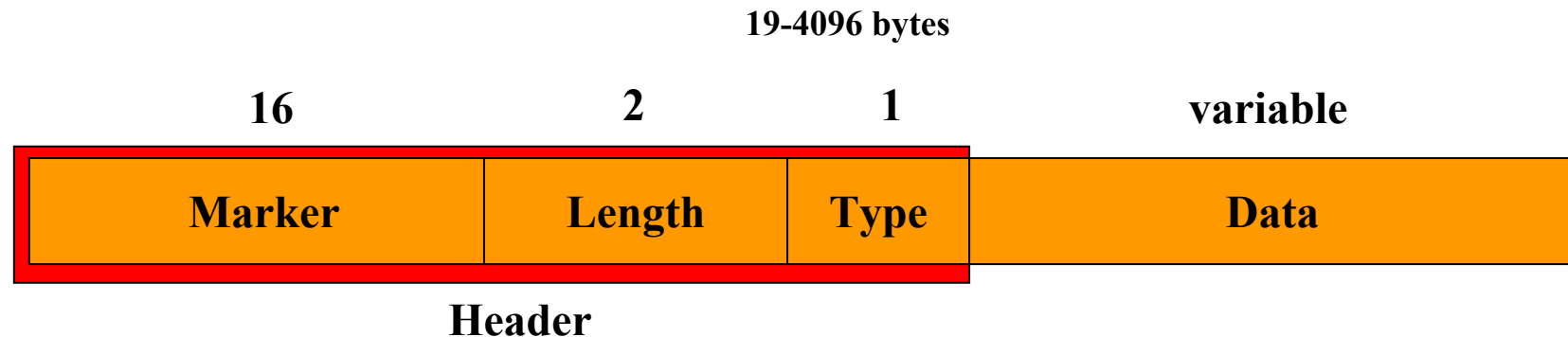
- Αρχικά οι δρομολογητές ανταλλάσσουν **ολόκληρο** τον BGP πίνακα δρομολόγησης που έχει ο καθέ ένας, ενημερώνει δηλαδή τον άλλον για τα δίκτυα που **μπορούν** να **προσεγγιστούν** μέσω αυτού
- **Σταδιακές ενημερώσεις** στέλνονται σε κάθε **αλλαγή** του πίνακα δρομολόγησης
- Αντίθετα με άλλα πρωτόκολλα δεν γίνεται **περιοδική αποστολή** όλου του πίνακα
- Οι BGP δρομολογητές κρατάνε την **πιο ενημερωμένη έκδοση** του πίνακα δρομολόγησης του κάθε ομότιμου
- Κρατάνε έναν **αριθμό έκδοσης του πίνακα** και ο οποίος πρέπει να είναι ο ίδιος για όλους τους ομότιμους
- Η έκδοση αλλάζει οποτεδήποτε γίνεται **ενημέρωση** του πίνακα
- Ο **ρυθμός** και ο **τρόπος αύξησης** του είναι ένδειξη για τη **σταθερότητα** του δικτύου
- Σε ένα πολύ **σταθερό** δίκτυο πολύ λίγη κίνηση εκτός των **KeepAlive** μηνυμάτων πρέπει να παρατηρείται

BGP, Λειτουργία

- Παρότι κρατούνται στον πίνακα δρομολόγησης του πρωτοκόλλου **όλες οι πιθανές διαδρομές προς ένα προορισμό**, μόνο η **βέλτιστη και μόνο αυτή** προωθείται στον **πίνακα προώθησης** και διαφημίζεται προς τους άλλους δρομολογητές
- Υπάρχει **μόνο μία βέλτιστη διαδρομή για ένα προορισμό** και αυτή μπορεί να μην είναι αυτή που εγκαθίσταται στην RIB/FIB
- Υπολογίζονται οι βέλτιστες διαδρομές **σε σχέση με τις ενημερώσεις** και **όχι τη συνολική κατάσταση** όπως συμβαίνει στο OSPF ή το IS-IS με τους spf υπολογισμούς
- Ανταλλάσσεται πληροφορία σχετικά με την **προσιτότητα** δικτύων, όπου κυρίως είναι μια ένδειξη των **AS** που πρέπει να ακολουθηθούν για κάποιο προορισμό
- Αυτή η πληροφορία θα χρησιμοποιηθεί για τη δημιουργία ενός **γράφου** με όλα τα AS η οποία δεν έχει **βρόχους** και όπου μπορεί να εφαρμοστεί κάποια **τακτική** (policy) ώστε να επηρεάσει η δρομολόγηση

Μηνύματα

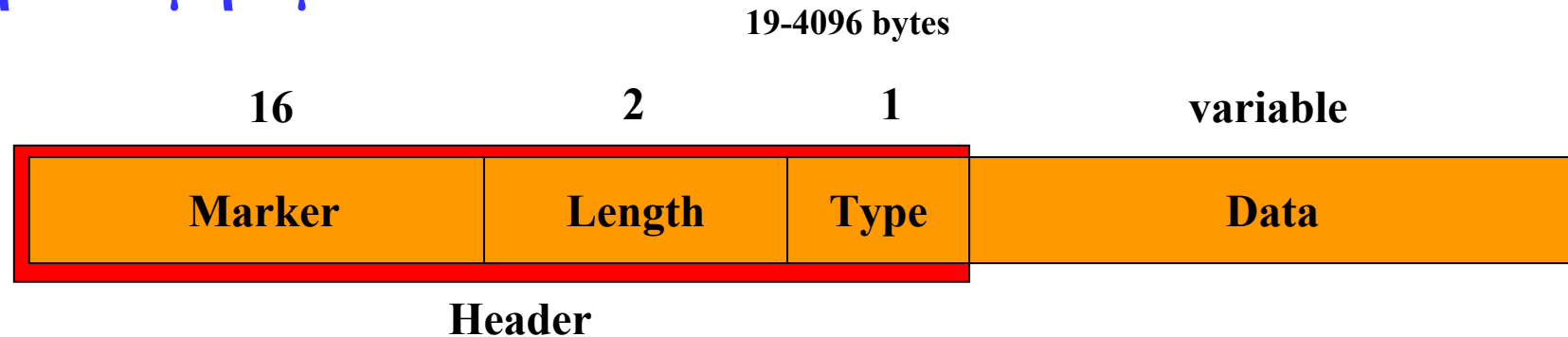
Μορφή BGP μηνυμάτων



Marker

- Περιέχει μια τιμή την οποία ο δέκτης του μηνύματος μπορεί να **προβλέψει**
- Αν το μήνυμα είναι τύπου Open ή αν το Open δεν μεταφέρει πληροφορία αυθεντικοποίησης έχει την τιμή **FF**
- Διαφορετικά έχει μια τιμή που μπορεί να υπολογιστεί συναρτήσει της αυθεντικοποίησης που ορίζεται
- Μπορεί να χρησιμεύει για την ανίχνευση απώλειας συγχρονισμού ανάμεσα στους ομοτίμους όπως και στο συγχρονισμό όταν περισσότερα από ένα μηνύματα περιέχονται στο ίδιο TCP segment αλλά και για την αυθεντικοποίηση των εισερχομένων μηνυμάτων

Μορφή BGP μηνυμάτων



Length

- Το συνολικό μήκος του μηνύματος σε byte. Έτσι μπορούμε να βρούμε το επόμενο μήνυμα σε ένα TCP πακέτο που περιέχει πολλά μηνύματα

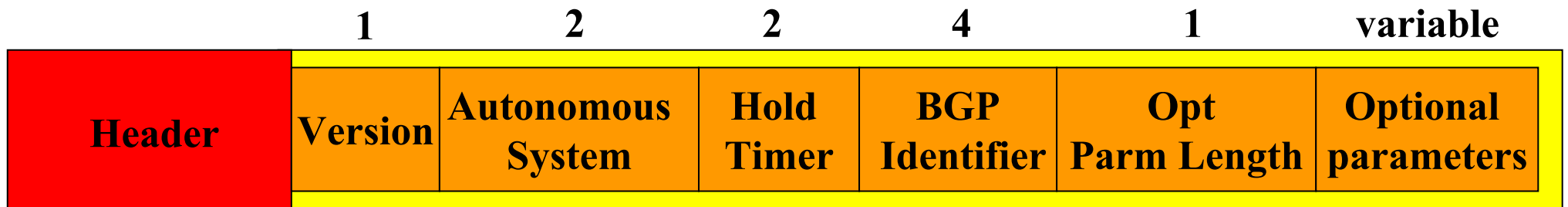
Type

- Ο τύπος του μηνύματος
- Ορίζονται στην **RFC1163** τέσσερις τύποι μηνυμάτων
 - 1-Open** Ίδρυση μια ομότιμης συνόδου
 - 2-Update** Αναγγελία ή απόρριψη διαδρομών
 - 3-Notification** Κατάργηση ομότιμης συνόδου
 - 4-KeepAlive** Χειραψία σε τακτά διαστήματα

Open μήνυμα

- **Ανοίγει** μια BGP σύνοδο μεταξύ ομοτίμων
- Επιτρέπει στον κάθε ομότιμο να **προσδιοριστεί** στον άλλον και να **συμφωνήσει-διαπραγματευτεί** για διάφορες **παραμέτρους** όπως οι μετρητές χρόνου
- Η επιβεβαίωση τους γίνεται με **Keepalive** μήνυμα και είναι απαραίτητη για να αρχίσει η ανταλλαγή των ενημερώσεων

Open μήνυμα



Open μήνυμα

Version

- Η έκδοση του BGP πρωτοκόλλου, οι δύο συνομιλητές πρέπει να τρέχουν την ίδια έκδοση του πρωτοκόλλου, εξ' ορισμού η **BGP-4**

Autonomous System

- Ο αριθμός του AS στο οποίο ανήκει ο αποστολέας

Hold time

- Ο ανώτερος χρόνος σε δευτερόλεπτα που ο αποστολέας προτείνει ότι μπορεί να περάσει χωρίς τη λήψη μηνυμάτων Update ή KeepAlive, εξ' ορισμού τα **180'**, ελάχιστος τα **3'**, μπορεί να πάρει και την τιμή **0**
- Μετά την πάροδο του διαστήματος, ο ομότιμος θεωρείται εκτός λειτουργίας
- Με τη λήψη του Open μηνύματος ο δέκτης θέτει σαν HoldTime την μικρότερη από τις δύο τιμές

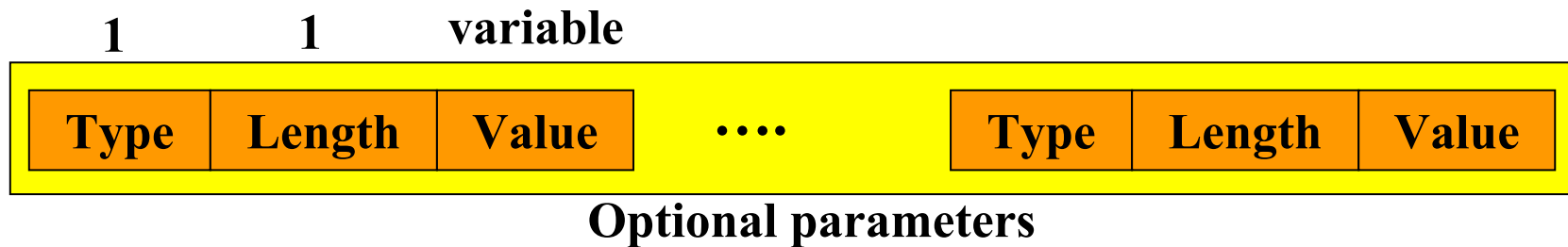
BGP Identifier

- Ταυτότητα του αποστολέα BGP δρομολογητή, συνήθως είναι ίδια με την **RID**, Router Identifier

Optional Parameter Length

- Είναι 0 αν δεν υπάρχουν προαιρετικές παράμετροι

Open μήνυμα



Optional Parameters

- Λίστα με προαιρετικές παραμέτρους, η κάθε μία περιγράφεται από μία τριπλέτα [**type**-τύπος της παραμέτρου-1byte, **length**-μήκος του πεδίου value-1byte, **value**]

Authentication code (RFC1771)

- Τύπος 1
- Δεν έχει υλοποιηθεί από τους κατασκευαστές
- Το πεδίο Value περιέχει κωδικό αυθεντικοποίησης 1byte, που δείχνει τον μηχανισμό αυθεντικοποίησης που χρησιμοποιείται ακολουθούμενο από δεδομένα αυθεντικοποίησης μεταβλητού μήκους

Update μήνυμα

- Μεταφέρουν **πληροφορία δρομολόγησης** μεταξύ των ομοτίμων
- Μπορούν να **ανακοινώσουν** ή να **απορρίψουν** πληροφορία δρομολόγησης
- Η **ανακοίνωση** είναι μια **λίστα με προθέματα** και μία **σειρά από ιδιότητες** για αυτά τα προθέματα
- Ένα μήνυμα Update μπορεί να **ανακοινώσει** διαδρομές που να έχουν πολλές ιδιότητες, να **αποσύρει** πολλές διαδρομές ή και τα δύο
- Το **πρόθεμα** είναι το τμήμα της IP διεύθυνσης που δηλώνει το δίκτυο ακολουθούμενο από / και έναν αριθμό 0 ως 32
- Κάθε **πρόθεμα** αναπαριστά ένα **δίκτυο** που μπορεί να προσεγγιστεί μέσω της IP διεύθυνσης που δίνεται στην ιδιότητα **NextHop**
- Η πληροφορία που περιέχεται χρησιμοποιείται για την κατασκευή ενός **γράφου** που περιγράφει τη σχέση ανάμεσα στα διάφορα AS



Ανακοίνωση = Πρόθεμα(τα) + κοινές ιδιότητες

Update μήνυμα, πεδία

	2	variable	2	variable	variable
Header	Withdrawn Routes Length	Withdrawn Routes	Total Path Attribute Length	Path Attribute	Network Layer Reachability Information

Withdrawn Routes Length

- Μήκος του πεδίου με τις **απορριφθέντες** διαδρομές, 0 σημαίνει ότι δεν υπάρχουν

Withdrawn Routes

- 2-tuple προθέματα **διαδρομών** που **απορρίπτονται** , σε μορφή <length, prefix>

Total Path Attribute Length

- Συνολικό μήκος των **ιδιοτήτων** των διαδρομών, 0 σημαίνει ότι δεν υπάρχουν NLRI

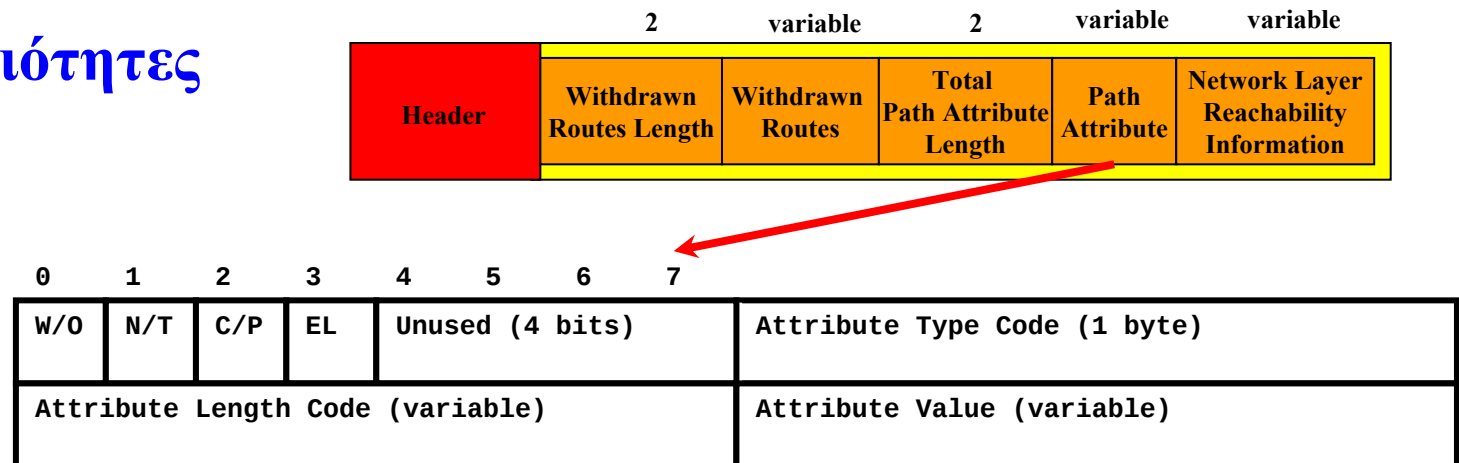
Path Attributes

- Ιδιότητες** των διαδρομών, παρουσιάζονται σαν μεταβλητού μήκους **TLV**, type-length-value

NLRI, Network Layer Reachability Information

- 2-tuple προθέματα <length, prefix>
- Περιλαμβάνουν μία διεύθυνση και μία μάσκα (10.26.122.0/24)

Update μήνυμα, ιδιότητες



- Τα πρώτα bits περιγράφουν την **κατηγορία** που ανήκει η ιδιότητα:

Bit 0/Optional: Well-known ή Optional

Bit 1/ Transitive : Non-transitive ή Transitive, για τις well-known είναι 1

Bit 2/Partial: Complete ή Partial

Complete: Η ιδιότητα διέσχισε ολόκληρη τη διαδρομή

Partial: Κάποιος δρομολογητής ενδιάμεσα δεν υλοποιεί την ιδιότητα και πιθανώς να έχει χαθεί κάποια πληροφορία

Για τις well-known και για τις optional non-transitive είναι 0

Bit 3/ Extended Length : Δηλώνει το μήκος της ιδιότητας, 1 byte ή 2 bytes

Bits 4 – 7: είναι 0 και δεν χρησιμοποιούνται

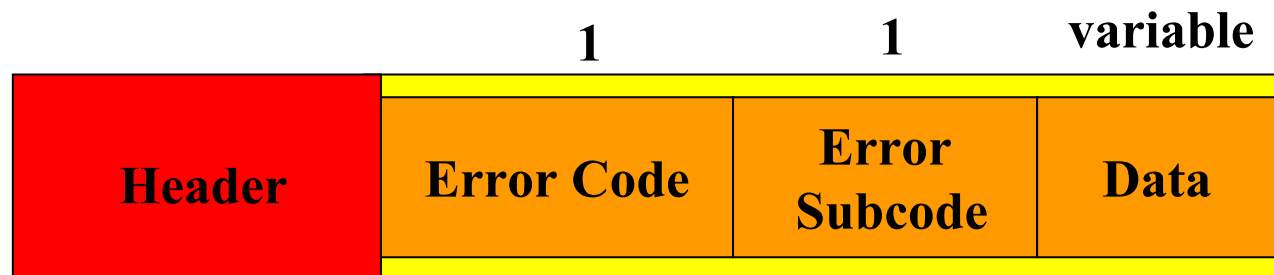
- **Attribute Type Code:** Κωδικός της ιδιότητας

- **Attribute Length :** Είναι 1 ή 2bytes ανάλογα με το ExtendedLength Bit. Δηλώνει το μήκος της ιδιότητας

- **Attribute Value:** Η τιμή της ιδιότητας

Notification μήνυμα

- Χρησιμοποιούνται για να **κλείσουν** μια ενεργή σύνοδο και να πληροφορήσουν τους συνεδεμένους δρομολογητές για το **λόγο**
- Στέλνονται όταν μια κατάσταση **λάθους** ανιχνευθεί ή και όταν μία σύνοδος τερματιστεί από το **διαχειριστή** με την κατάλληλη εντολή
- Η **TCP** σύνδεση τερματίζεται **αμέσως** μετά την **αποστολή** ή **λήψη** του μηνύματος
- Να παρατηρήσουμε ότι μια BGP σύνοδος είναι πολύ **ευαίσθητη** σε λάθη που μπορεί να συμβούν



Notification μήνυμα

Error Code

1-Message Header Error

- Πρόβλημα στην **επικεφαλίδα** του μηνύματος, όπω μη αποδεκτό μήκος μηνύματος, AS αριθμός, πεδίο marker, πεδίο type

2-OPEN Message Error

- Πρόβλημα με το **open** μήνυμα, όπως μη υποστηριζόμενη έκδοση πρωτοκόλλου, λάθος στη λίστα με τις ιδιότητες ή μη αποδεκτό next-hop

3-UPDATE Message Error

- Πρόβλημα με το Update μήνυμα

4-Hold Timer Expired

- Ο μετρητής HoldTime εξέπνευσε

5-Finite State Machine Error

- Οποιοδήποτε λάθος στη μηχανή καταστάσεων του BGP

6-Cease

- Ένας ομότιμος μπορεί να επιλέξει, χωρίς να υπάρχουν λάθη, να κλείσει την BGP σύνοδο (για παράδειγμα από τη διαχείριση)

Notification μήνυμα

Error subcode

- Περισσότερες λεπτομέρειες για το σφάλμα
- Κάθε Error Code μπορεί να έχει ένα ή περισσότερα σχετικά Error Subcodes
- Αν δεν έχει οριστεί αυτό το πεδίο, χρησιμοποιείται το 0

Error data

- Προαιρετικά δεδομένα σχετικά με το σφάλμα
- Χρησιμοποιείται για τη **διάγνωση** της αιτίας για το Notification μήνυμα
- Τα περιεχόμενα εξαρτώνται από τα πεδία Error Code και Error Subcode

Message Header Error

- 1 - Connection Not Synchronized
- 2 - Bad Message Length
- 3 - Bad Message Type

OPEN Message Error

- 1 - Unsupported Version Number
- 2 - Bad Peer AS
- 3 - Bad BGP Identifier
- 4 - Unsupported Optional Parameter
- 5 - Authentication Failure
- 6 - Unacceptable Hold Time

UPDATE Message Error

- 1 - Malformed Attribute List
- 2 - Unrecognized Well-known Attribute
- 3 - Missing Well-known Attribute
- 4 - Attribute Flags Error
- 5 - Attribute Length Error
- 6 - Invalid ORIGIN Attribute
- 7 - AS Routing Loop
- 8 - Invalid NEXT_HOP Attribute
- 9 - Optional Attribute Error
- 10 - Invalid Network Field
- 11 - Malformed AS_PATH

KeepAlive μήνυμα

- Στέλνονται για να αποτρέψουν την **εκπνοή** του αντίστοιχου μετρητή χρόνου (**hold-time**)
- Στέλνονται αρκετά συχνά από 1 ως 21845'
- Στέλνονται στο **1/3** του χρόνου Hold Time, εξ'ορισμού τιμή τα **60'**
- Αν η παράμετρος **HoldTime** έχει τεθεί σε 0, τότε μηνύματα KeepAlive **δεν στέλνονται**



Header

Route-Refresh μήνυμα

- Δεν ορίζεται στην RFC1771, αλλά σαν **δυνατότητα** στην **RFC2918**
- Έχει **ευρέως** αναπτυχθεί
- Χρησιμοποιείται για να ζητήσει την **επανεκπομπή ολόκληρης** της πληροφορίας δρομολόγησης ενός ομότιμου, **χωρίς** να χρειαστεί **επανεκκίνηση** της BGP συνόδου
- Έτσι αλλαγές στην **τακτική** δρομολόγησης μπορούν να εφαρμοστούν, χωρίς hard reset του πρωτοκόλλου που προκαλεί **μεγάλη αναστάτωση** σε όλο το δίκτυο

AFI, Address Family Identifier

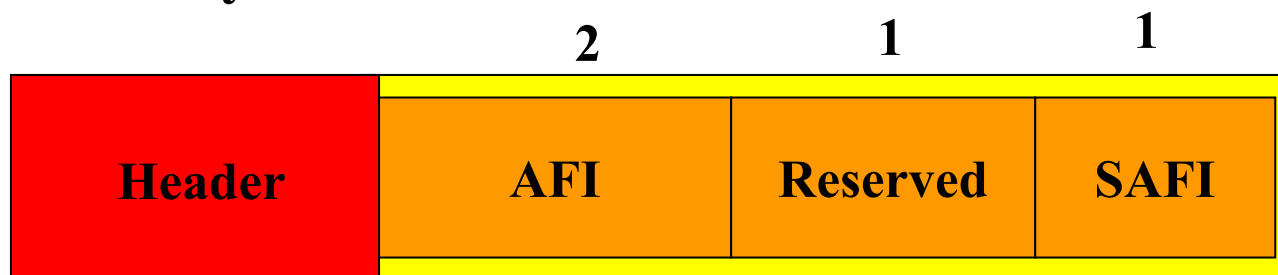
- IPv4 ή IPv6

SAFI, Subsequent Address Family Identifier

- Π.χ unicast ή multicast

Reserved

- Πάντα 0



Ιδιότητες Διαδρομών

BGP, Ιδιότητες διαδρομών (attributes)

- Κάθε διαδρομή προς ένα δίκτυο που ανακοινώνεται κουβαλά μαζί της κάποιες παραμέτρους που ονομάζονται **ιδιότητες** (attributes)
- Μπορούμε να δούμε τη διαδρομή σαν **αντικείμενο** ή σαν ένα κουτί με ιδιότητες ή σαν μικρή **βάση δεδομένων**
- Η χρήση τους βοηθά στην εύκολη **κλιμάκωση** του δικτύου, στον ορισμό **πολιτικών** δρομολόγησης και στην διατήρηση ενός **σταθερού περιβάλλοντος** δρομολόγησης
- Οι ιδιότητες αυτές χρησιμοποιούνται στην επιλογή της **καλύτερης διαδρομής προς έναν προορισμό**
- Η **κατανόηση** τους είναι απολύτως απαραίτητη για τη σχεδίαση ενός στιβαρού δικτύου

BGP, Ιδιότητες διαδρομών (attributes)

- Οι ιδιότητες ανήκουν στις ακόλουθες **κατηγορίες**:

Προαιρετική μεταβατική

- Δεν είναι υποχρεωτική η αναγνώριση και υλοποίηση της, αλλά πρέπει να αναμεταδίδεται σε άλλους ακόμα και αν δεν την καταλαβαίνουμε

Προαιρετική μη μεταβατική

- Δεν είναι υποχρεωτική η αναγνώριση και υλοποίηση της και δεν μεταδίδεται σε άλλους ομότιμους αν δεν είναι αντιληπτή

Γνωστή υποχρεωτική

- Πρέπει να είναι αναγνωρίσιμη από όλους τους BGP ομιλητές και πρέπει να περιλαμβάνεται σε όλες τις ενημερώσεις

Γνωστή διακριτική

- Πρέπει να αναγνωρίζεται-υλοποιείται από όλους τους BGP ομιλητές αλλά δεν είναι υποχρεωτικό να περιλαμβάνεται σε όλες τις ενημερώσεις

BGP, Ιδιότητες διαδρομών

	Value	Code	Reference	
Υποχρεωτικές	1	ORIGIN	[RFC1771]	Μεταβατικές
	2	AS_PATH	[RFC1771]	
	3	NEXT_HOP	[RFC1771]	
Προαιρετικές	4	MULTI_EXIT_DISC	[RFC1771]	Μη Μεταβατικές
	5	LOCAL_PREF	[RFC1771]	
	6	ATOMIC_AGGREGATE	[RFC1771]	
	7	AGGREGATOR	[RFC1771]	
	8	COMMUNITY	[RFC1997]	
	9	ORIGINATOR_ID	[RFC2796]	
	10	CLUSTER_LIST	[RFC2796]	
	11	DPA	[Chen]	
	12	ADVERTISER	[RFC1863]	
	13	RCID_PATH / CLUSTER_ID	[RFC1863]	
	14	MP_REACH_NLRI	[RFC2283]	
	15	MP_UNREACH_NLRI	[RFC2283]	
	16	EXTENDED COMMUNITIES	[Rosen]	
	...			
	255	reserved for development		

Καταγωγή (Origin)

- **Γνωστή υποχρεωτική** ιδιότητα
- Είναι **ιστορική** ιδιότητα, αλλά χαμηλής σημασίας, από τις **τελευταίες** ιδιότητες στον αλγόριθμο επιλογής της καλύτερης διαδρομής
- Δηλώνει τον **τρόπο** με τον οποίο το **BGP** **πρωτοέμαθε** για τη συγκεκριμένη διαδρομή, δηλ. πως γεννήθηκε η διαδρομή στο αρχικό AS
- **Προτεραιότητα** έχουν οι διαδρομές **IGP > EGP > Unknown**
- Οι υλοποιήσεις του BGP προτιμούν τις **IGP** από τις **EGP** διαδρομές διότι οι EGP αποτυγχάνουν στην περίπτωση βρόχων
- Μπορούμε να αλλάξουμε την τιμή της αλλά δε συνηθίζεται

Καταγωγή (Origin)

- Είπαμε ότι το BGP μεταφέρει τις διαδρομές ανάμεσα στους ομότιμους, αλλά από πού προήλθαν-γεννήθηκαν αυτές **εξ αρχής**;
- Κατευθείαν** συνεδεμένες διεπαφές
- Στατικές** διαδρομές
- Εσωτερικά ή εξωτερικά δυναμικά πρωτόκολλα με **αναδιανομή** στο BGP

Καταγωγή (Origin)

- Έχει τρεις πιθανές τιμές:

0-IGP

- Η διαδρομή είναι **εσωτερική** στο **AS** που την **παρήγαγε** δηλ. το δίκτυο είναι μέρος του AS, μαθεύτηκε δηλ. από IGP πρωτόκολλο
- Αυτό συμβαίνει όταν ο δρομολογητής έχει εισάγει τη διαδρομή στο BGP, μέσω της εντολής διάρθρωσης **network** ή μέσω αναδιανομής από ένα δυναμικό εσωτερικό ή εξωτερικό πρωτόκολλο
- Η ένδειξη στον πίνακα του BGP είναι τότε ένα “i”

1-EGP

- Η διαδρομή έγινε γνωστή μέσω ενός **EGP** πρωτοκόλλου
- Η ένδειξη στον πίνακα του BGP είναι τότε ένα “e”

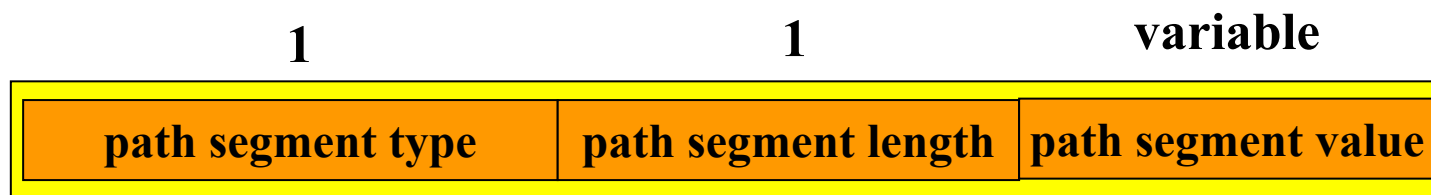
2-Incomplete

- Η καταγωγή της διαδρομής **δεν είναι γνωστή** ή έχει μαθευτεί με κάποιον άλλον τρόπο
- Αυτό συμβαίνει όταν μία **στατική διαδρομή** αναδιανέμεται στο BGP
- Η ένδειξη στον πίνακα του BGP είναι τότε ένα “?”

Μονοπάτι αυτόνομων συστημάτων (AS_path)

- Γνωστή υποχρεωτική ιδιότητα
- Είναι μία **λίστα με τα AS** τα οποία **διέσχισε** μια ανακοίνωση διαδρομής
- Καθώς η ανακοίνωση διασχίζει τα AS, κάθε AS προσθετει **το δικό του**
- Παρέχει ένα μηχανισμό **ανίχνευσης βρόχου**
- Η διαδρομή με τη **μικρότερη** λίστα προτιμάται
- Κάθε φορά που η ανακοίνωση διασχίζει δύο AS, ο EBGP Ομότιμος προσθέτει το δικό του ASN στη λίστα
- Έτσι το AS που ανακοίνωσε τη διαδρομή βρίσκεται **αριστερά** στη λίστα, ένα AS από το οποίο διάβηκε βρίσκεται στη **μέση** και το AS που γέννησε τη διαδρομή βρίσκεται στο **τέλος** της AS_path λίστας

AS_path



- Κάθε AS παριστάνεται από μία **τριπλέτα**

path segment type

- 1 AS_SET: Μη διατεταγμένη σειρά AS τα οποία διέσχισε μια ανακοίνωση
- 2 AS_SEQUENCE: Διατεταγμένη σειρά AS τα οποία διέσχισε μια ανακοίνωση

path segment length

- Αριθμός των AS στο πεδίο path segment value

path segment value

- Ένας ή περισσότεροι AS αριθμοί, ο κάθε ένας με μήκος 2 bytes

AS_path

Αν ο ομότιμος **διαδίδει** μια διαδρομή που έμαθε από άλλο BGP ομιλητή:

- Αν ο BGP ομότιμος ανακοινώσει τη διαδρομή σε άλλο BGP που βρίσκεται στο **ίδιο AS**, δεν αλλάζει το AS_path
- Αν ο BGP ομότιμος ανακοινώσει τη διαδρομή σε άλλο BGP που βρίσκεται σε **γειτονικό AS**, ενημερώνει το AS_path

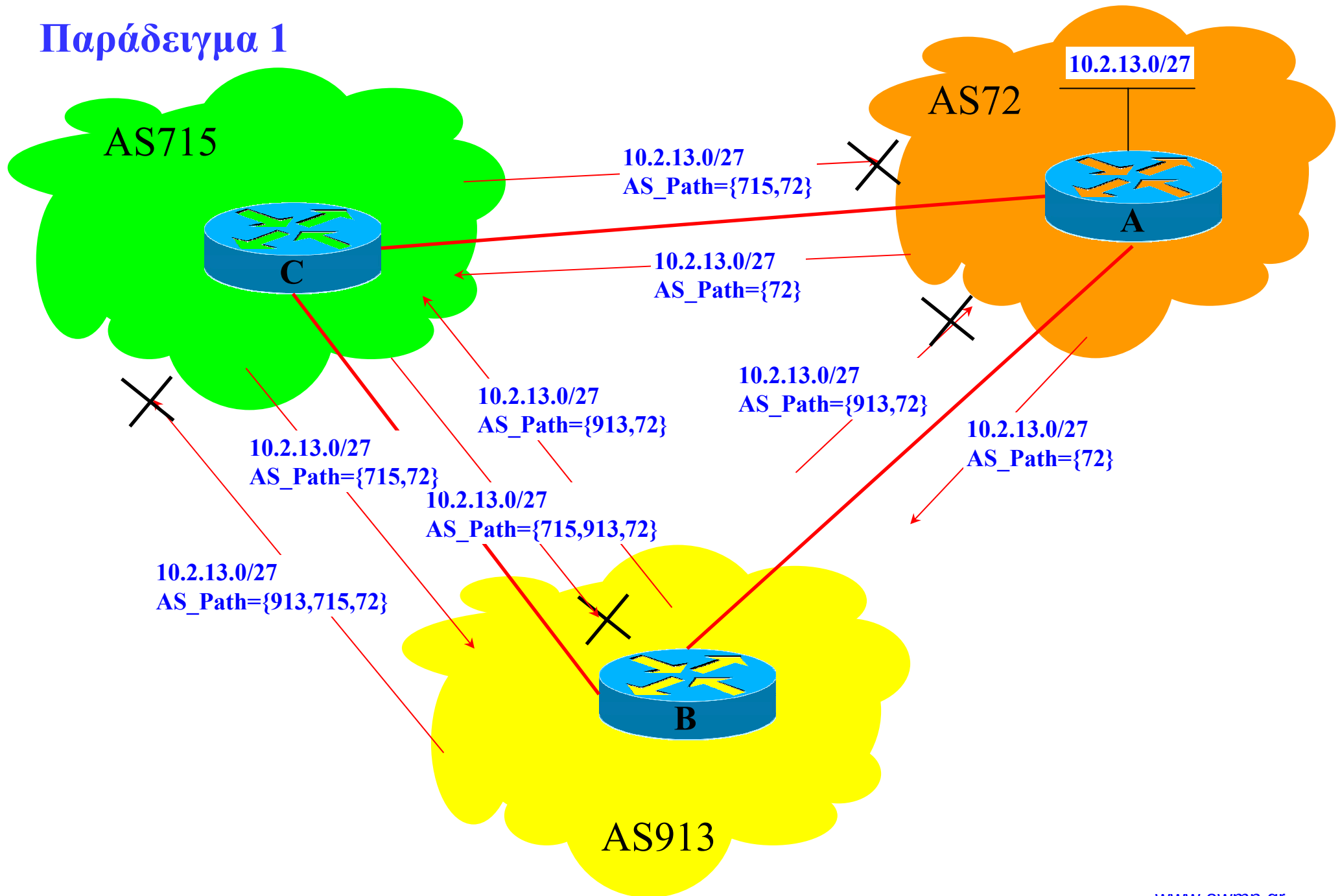
Αν το πρώτο path segment είναι τύπου AS_SEQUENCE, θα τοποθετήσει το ASN του στο τέλος (αριστερά)

Αν το πρώτο path segment είναι τύπου AS_SET, θα τοποθετήσει ένα νέο path segment με το ASN του, τύπου AS_SEQUENCE

Αν ο BGP ομιλητής **γεννεί** τη διαδρομή:

- Αν στέλνει την ενημέρωση σε ομότιμο σε **διαφορετικό AS**, εισάγει το δικό του ASN στη λίστα
- Αν στέλνει την ενημέρωση σε ομότιμο στο **ίδιο AS** περιλαμβάνει μια κενή AS_path (length=0) στις ενημερώσεις που στέλνει

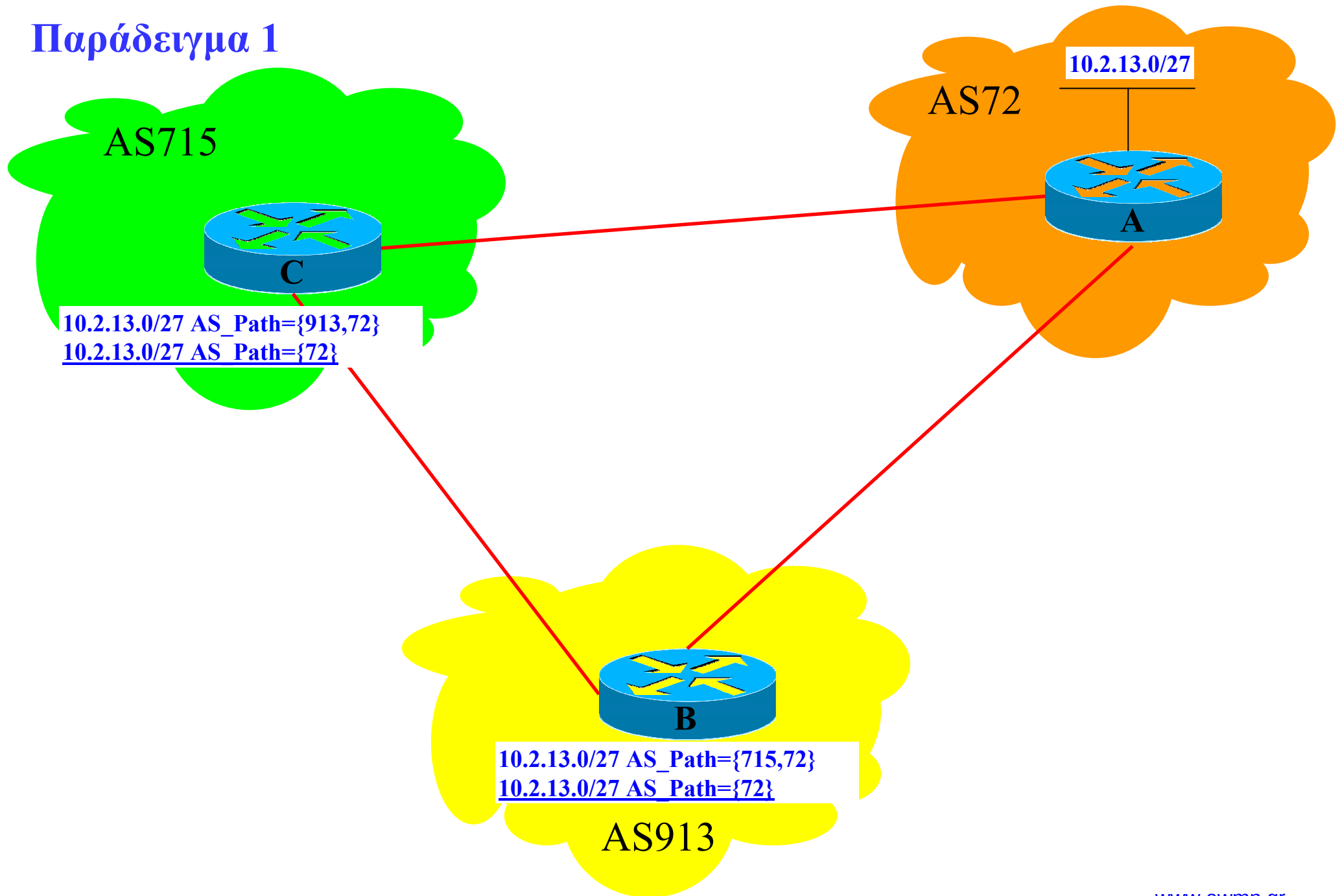
Παράδειγμα 1



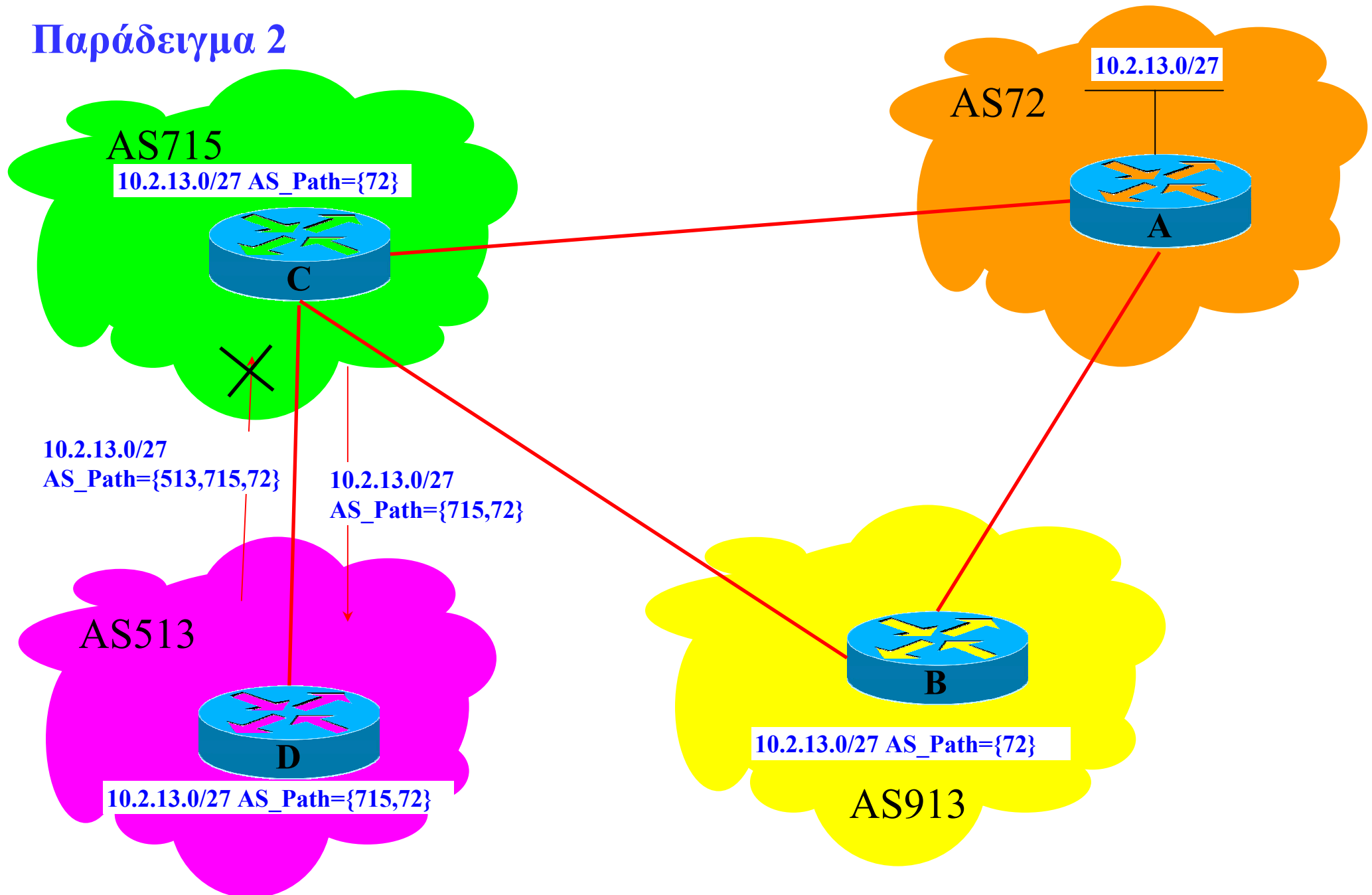
Παράδειγμα 1

- Το AS72 παράγει τη διαδρομή προς το 10.2.13.0/27 και τη διαφημίζει προς τα AS913 και AS715, με το πεδίο AS_path ίσο με {72}
- Η AS913 διαφημίζει προς το AS72 με AS_path ίσο με {913,72} και η AS715 προς το AS72 με το πεδίο AS_path ίσο με {715,72}
- Η AS72 θα απορρίψει και τις δύο διαδρομές αφού περιέχουν τους δικούς της αριθμούς
- Αυτός είναι ο μηχανισμός με τον οποίο το BGP μπορεί να ανιχνεύει **βρόχους διαδρομών**
- Η AS913 διαφημίζει προς το AS715 με AS_path ίσο με {913,72} και η AS715 προς το AS913 με AS-path ίσο με {715,72}
- Οι δύο τελευταίες διαδρομές δεν θα εγκατασταθούν στον πίνακα δρομολόγησης του πυρήνα αφού οι AS913, AS715 έχουν μάθει διαδρομή προς το 10.2.13.0/27 με ένα **μικρότερο μήκος λίστας AS** αριθμών

Παράδειγμα 1



Παράδειγμα 2



Παράδειγμα 2

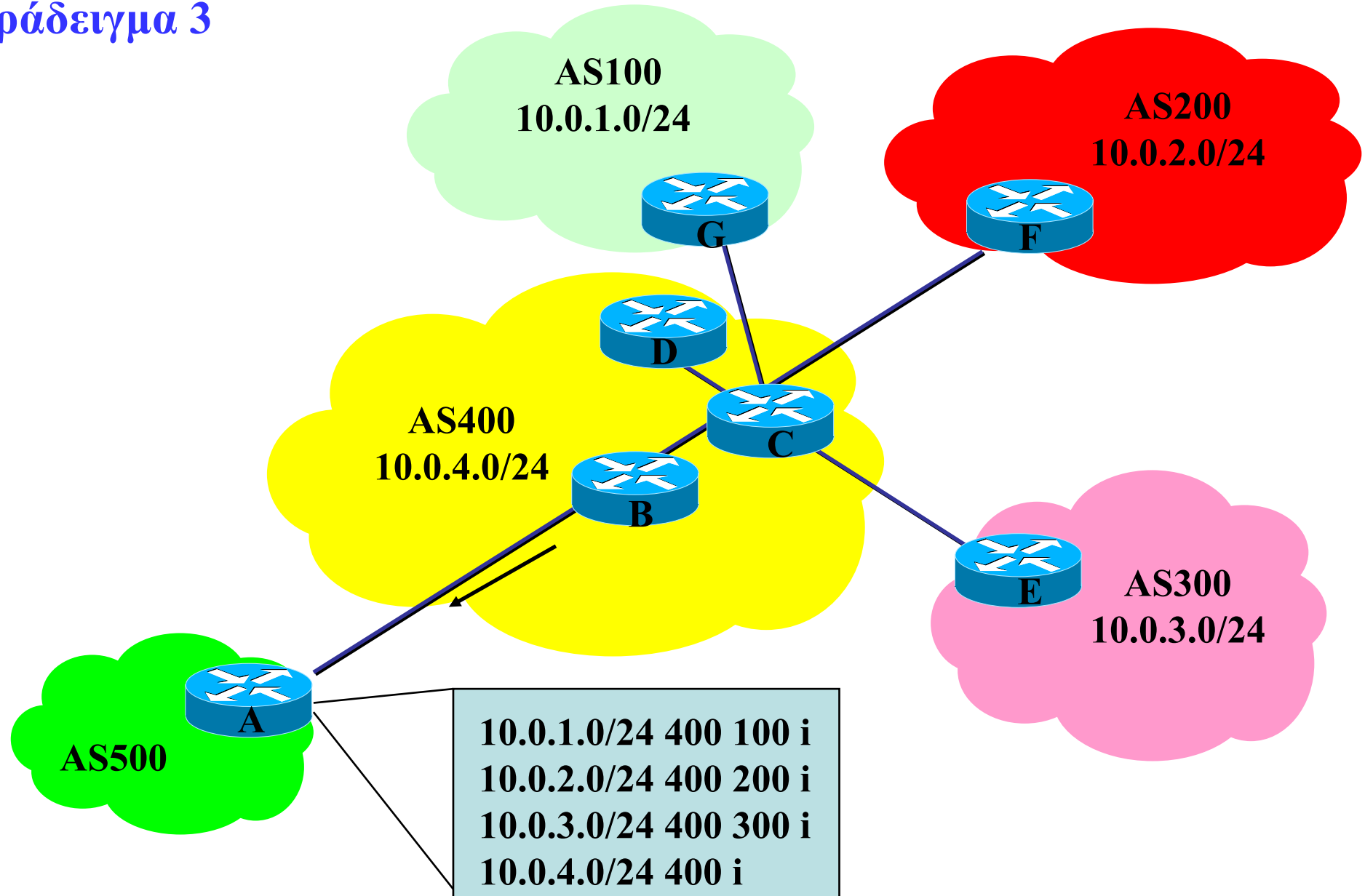
- Έστω τώρα ότι το AS513 συνδέεται στο AS715
- Το AS715 διατηρεί στον πίνακα δρομολόγησης του BGP και τις δύο διαδρομές

10.2.13.0/27 AS_Path={913,72}

10.2.13.0/27 AS_Path={72}

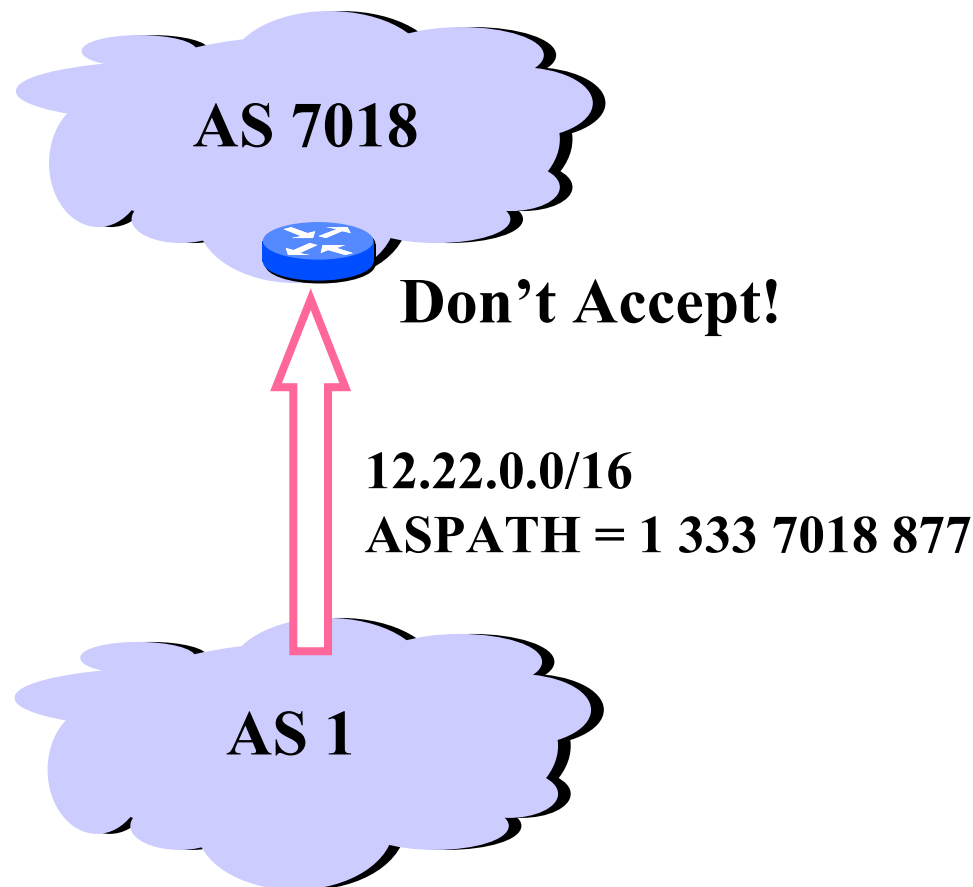
- Στον πίνακα προώθησης του όμως έχει μόνο την
10.2.13.0/27 AS_Path={72} και αυτή θα προωθήσει προς το AS513

Παράδειγμα 3

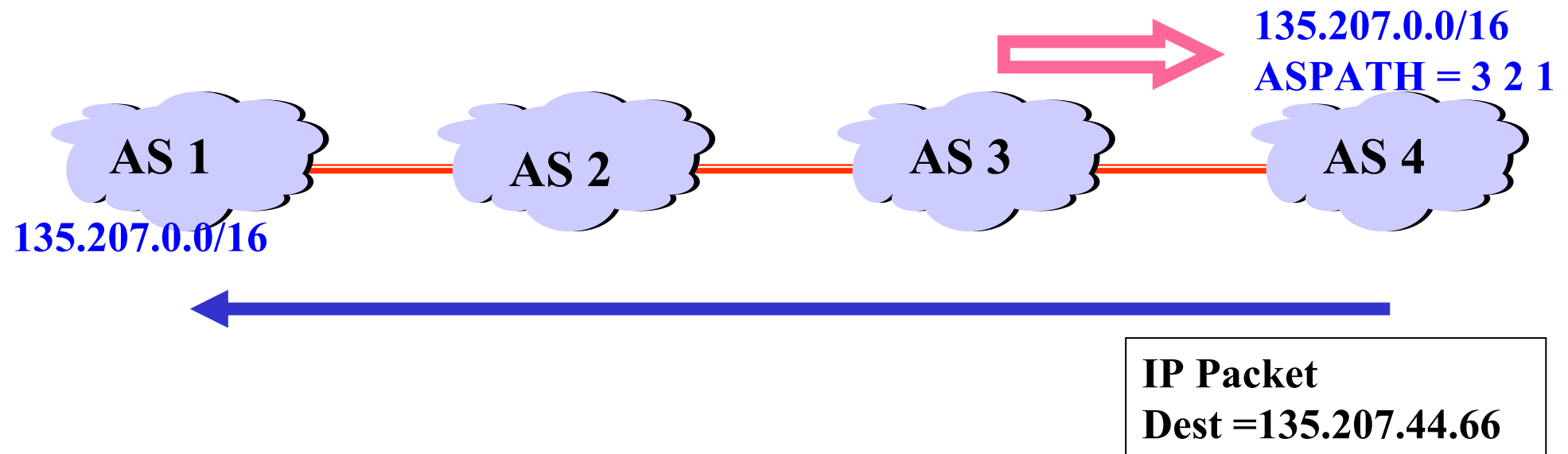


Ανίχνευση βρόχου

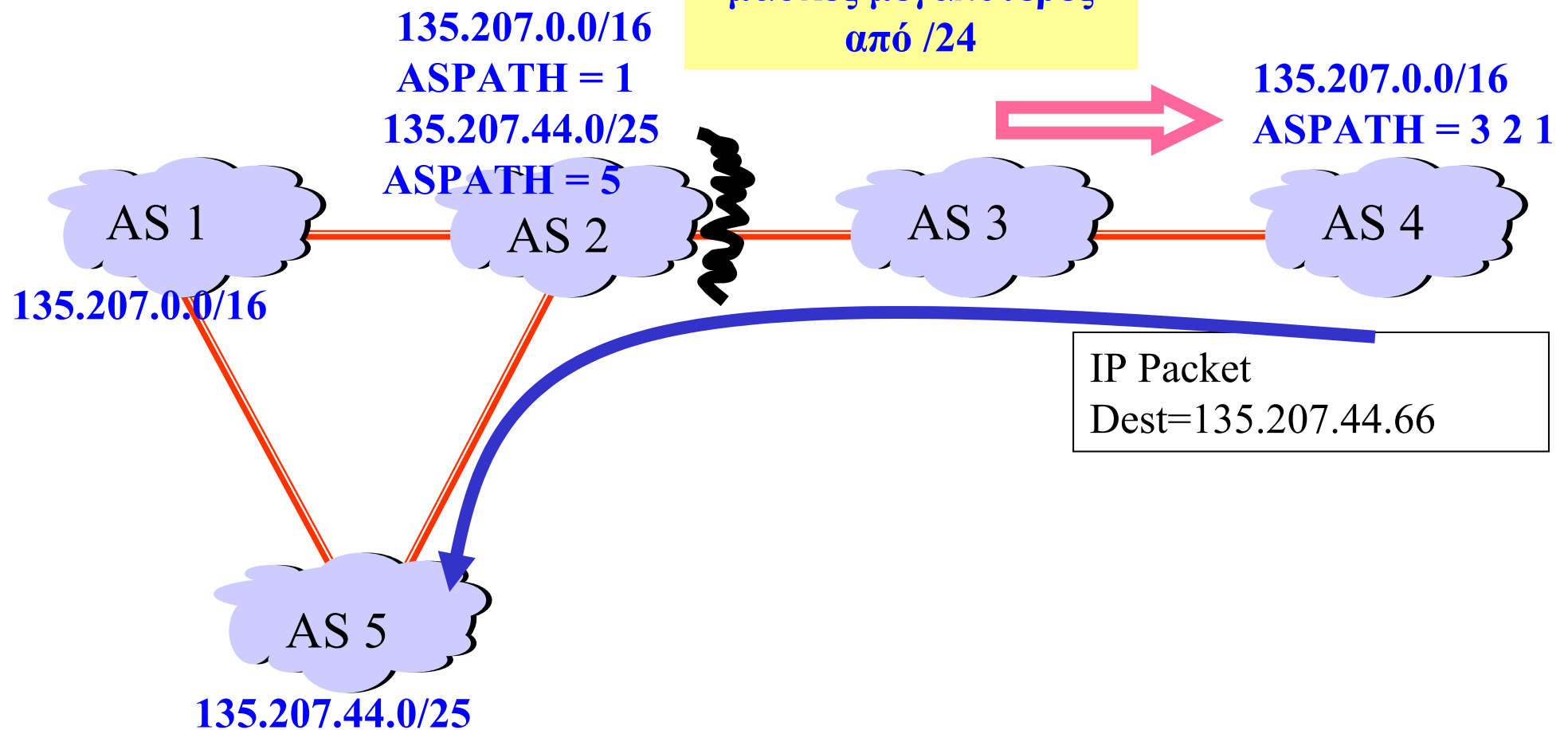
- Δρομολογητής BGP στο AS XXX δεν θα κάνει ποτέ δεκτή διαδρομή με ιδιότητα AS_path που να περιέχει το XXX



Η κίνηση συχνά ακολουθεί το AS_Path

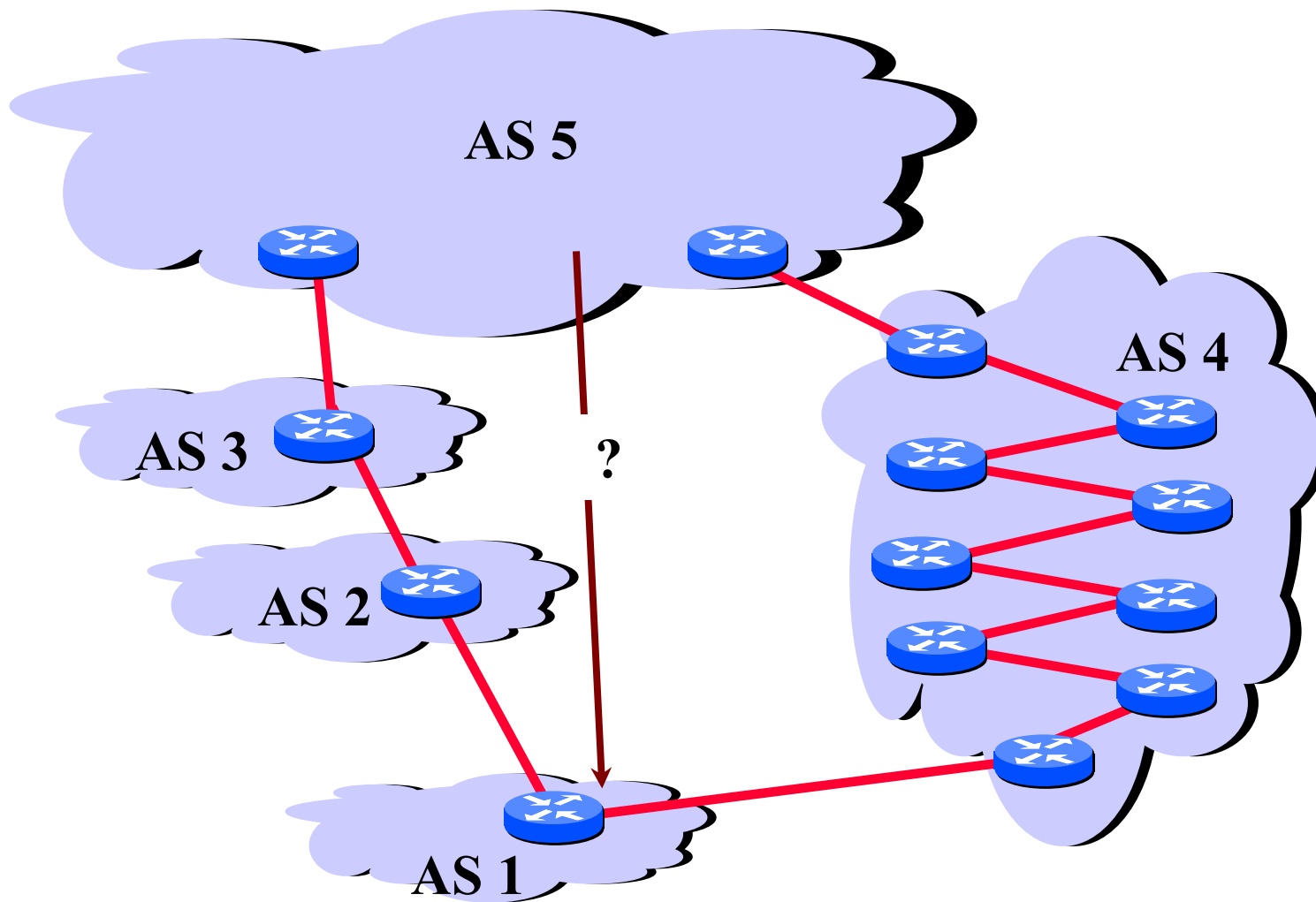


Αλλά μπορεί και όχι



- Το AS4 βλέπει ότι το πακέτο με προορισμό ο δίκτυο 135.207.0.0/16 θα ακολουθήσει τη διαδρομή 3 2 1, ενώ στην πραγματικότητα ακολουθεί την 3 2 5

Το BGP δεν βρίσκει την κοντινότερη διαδρομή



Το BGP δεν βρίσκει την κοντινότερη διαδρομή

Παράδειγμα

- Το BGP αποκρύπτει την πραγματική φυσική τοπολογία και έτσι προτείνει την διαδρομή μέσω των AS $\{4,1\}$ και όχι μέσω των $\{3,2,1\}$

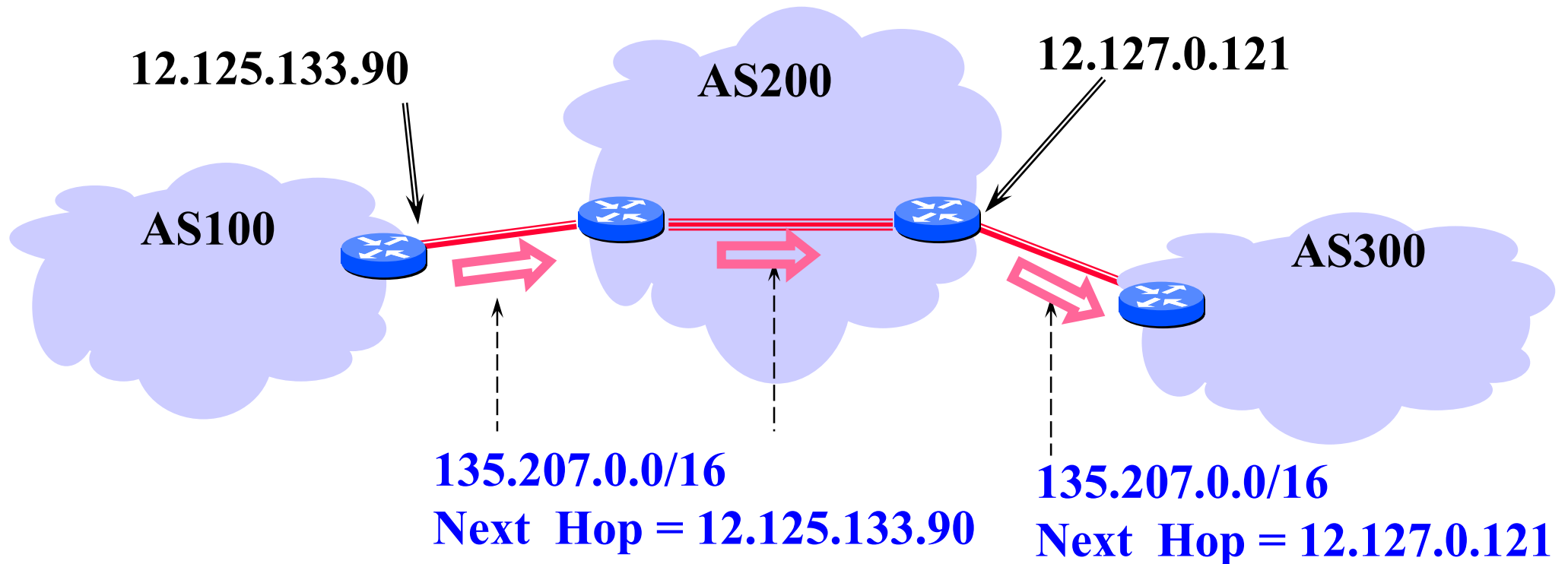
Γιατί;

- Αν το πρωτόκολλο λάβει υπόψη και την εσωτερική τοπολογία θα λάβει πιο σοφές αποφάσεις, αλλά θα αυξήσει δραματικά την **αστάθεια** και την **ποσότητα** της πληροφορίας που πρέπει να διαχειριστεί και αφετέρου δεν θα είναι εύκολη η **κλιμάκωση** του
- Αν ορίζαμε **ένα δρομολογητή ανά AS** αυτό το πρόβλημα δεν θα υπήρχε!

Next-Hop

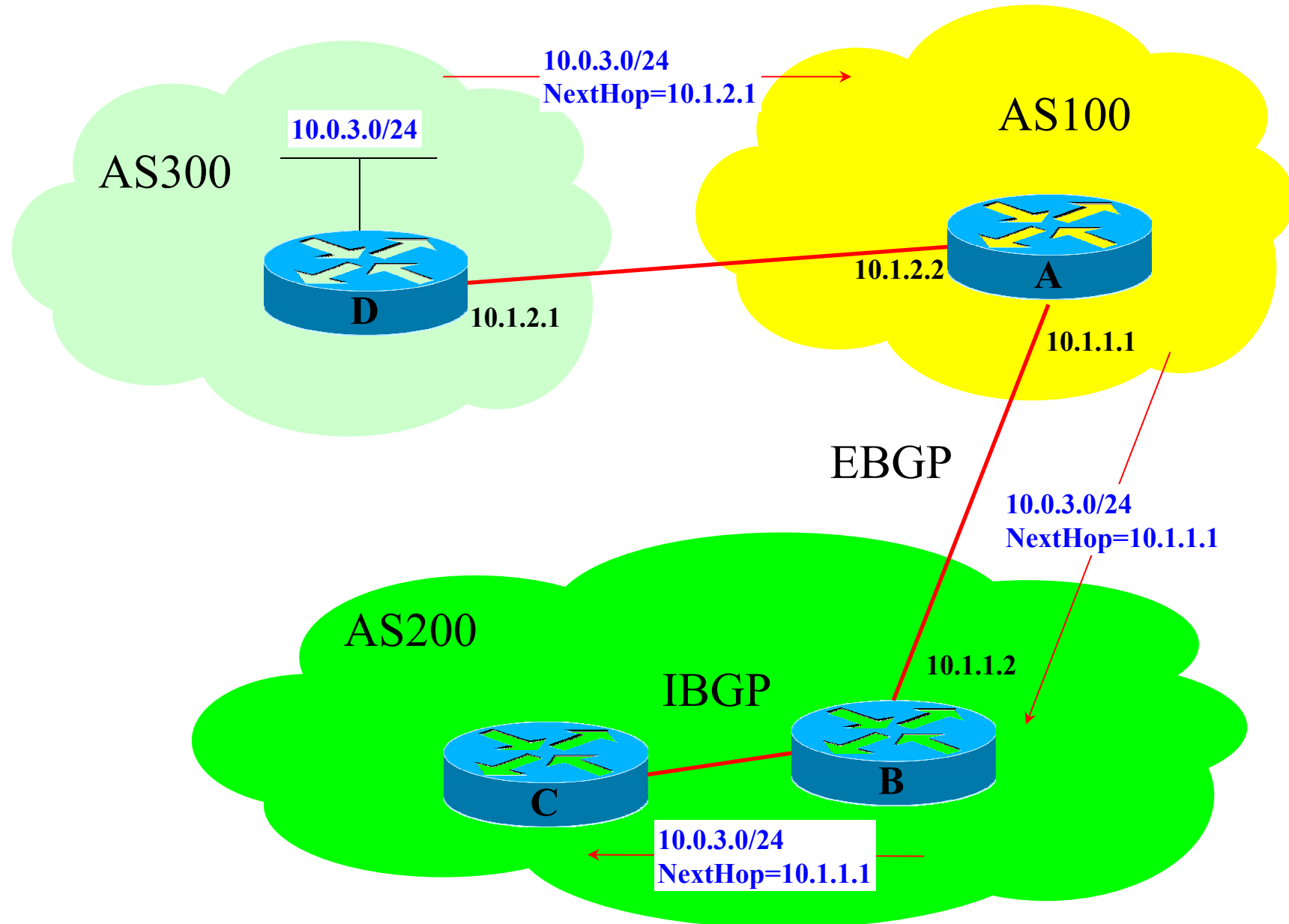
- **Γνωστή υποχρεωτική**
- Είναι η **IP διεύθυνση** του **επόμενου κόμβου** στον οποίο θα πρέπει να στείλουμε τα πακέτα ώστε να φτάσουν κοντύτερα σε κάποιο **προορισμό**
- Για ομότιμους **EBGP**, είναι η **IP διεύθυνση** της **σύνδεσης** του ομοτίμου (η διεύθυνση που ορίστηκε με την εντολή neighbor)
- Για ομότιμους **IBGP** είναι η **IP διεύθυνση** του ομοτίμου, συνήθως χρησιμοποιούμε τη διεύθυνση βρόχου
- Σε **IBGP**, η **EBGP NextHop** μεταφέρεται **διαφανώς** στο **τοπικό AS**
- Απόσυνδέει το **BGP** από την πραγματική **φυσική τοπολογία**
- Επιτρέπει στο **IGP** να πάρει **έξυπνες αποφάσεις προώθησης**
- Ο δρομολογητής πρέπει να κάνει αναδρομική αναζήτηση ώστε να βρει το **BGP Next_Hop** στον πίνακα δρομολόγησης
- Κάποιο **IGP** ή στατική δρομολόγησης ή κατευθείαν συνεδεσμένη διεπαφή μπορεί να παρέχει αυτή τη διαδρομή
- Αν το **NextHop** δεν είναι **προσεγγίσιμο**, η **BGP διαδρομή** δεν λαμβάνεται υπόψη

Παράδειγμα 1



- Κάθε φορά που μία ανακοίνωση διαδρομής διασχίζει το όριο ανάμεσα σε δύο AS η ιδιότητα next hop αλλάζει στην IP διεύθυνση του **ακραίου** δρομολογητή που ανακοίνωσε τη διαδρομή

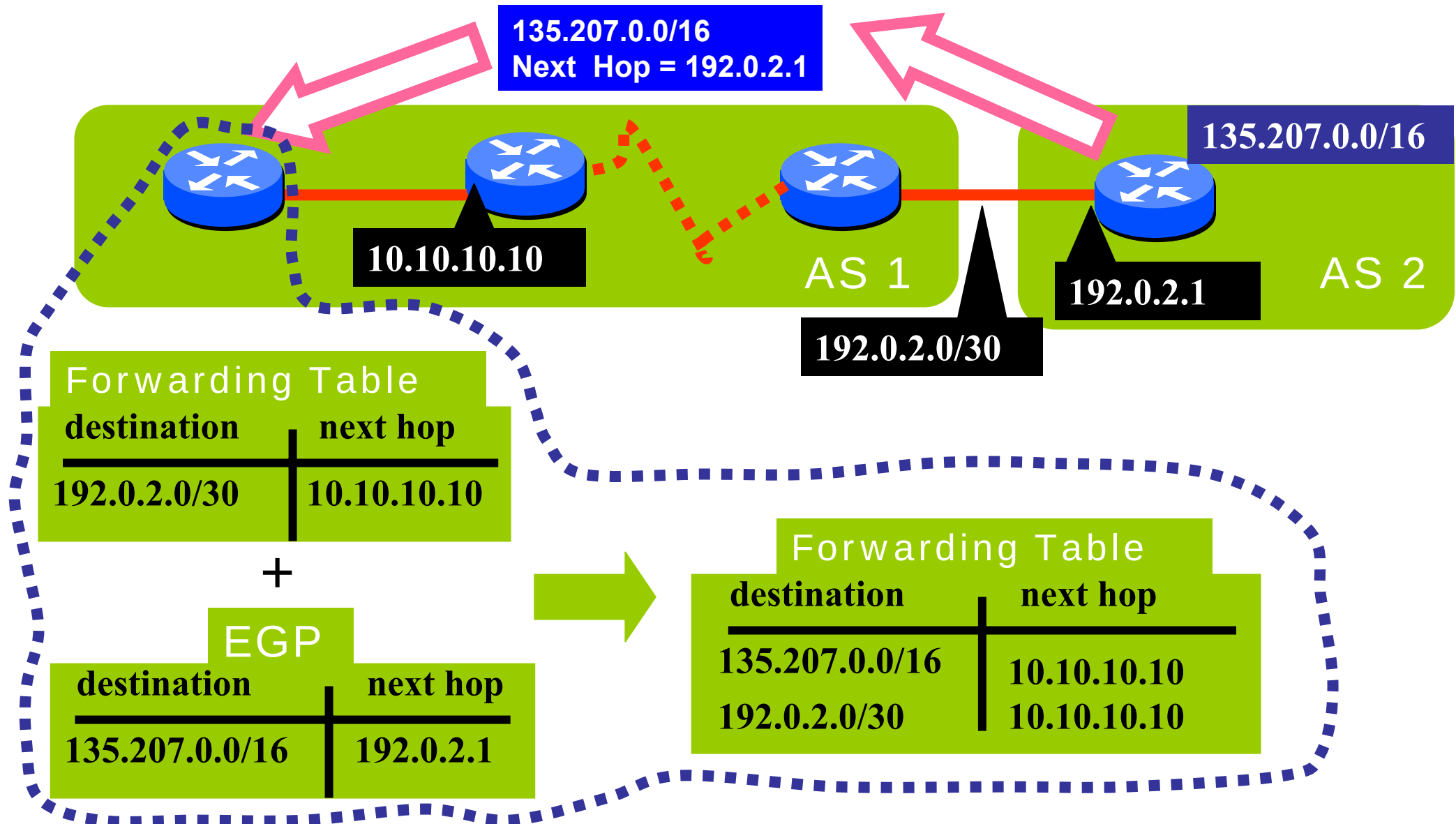
Παράδειγμα 2



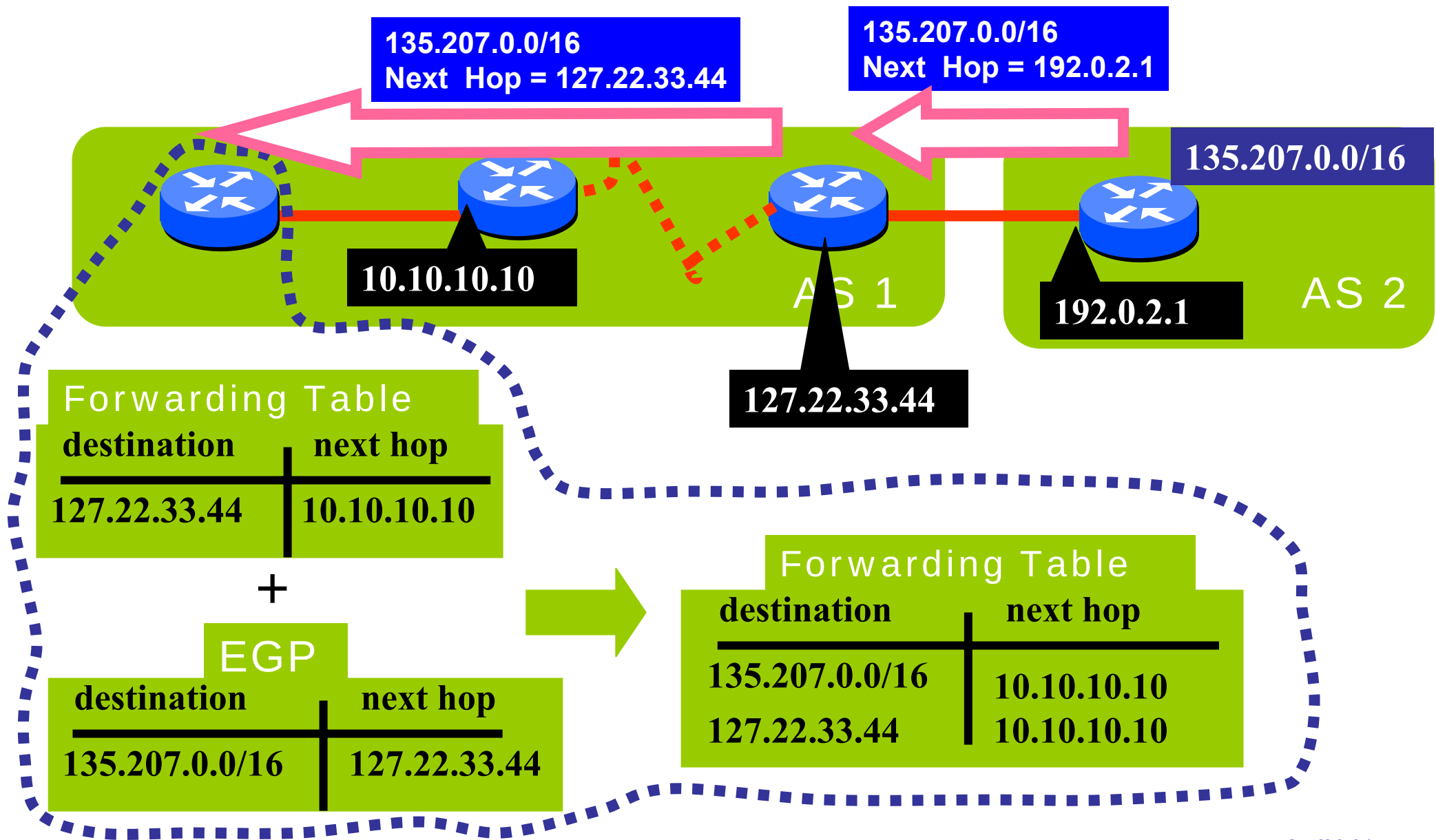
Παράδειγμα 2

- Ο D διαφημίζει το δίκτυο 10.0.3.0/24 με τιμή NextHop = 10.1.2.1
- Ο A διαφημίζει το δίκτυο 10.0.3.0/24 που έμαθε από τον D, αλλάζοντας το NextHop=10.1.1.1
- Όταν λοιπόν **διασχίζει** ένα AS η τιμή **αλλάζει** ανάλογα
- Όταν ο B διαδώσει αυτή τη διαδρομή στο AS του, η πληροφορία NextHop **κρατείται** (στο IBGP δεν αλλάζει η τιμή)
- Αν ο B ή ο C δεν έχουν πληροφορία δρομολόγησης για το NextHop, τότε η διαδρομή θα απορριφθεί
- Έτσι είναι σημαντικό να έχουμε ένα IGP στο AS ώστε να διαδώσει την πληροφορία για το NextHop
- Ο C για να προωθήσει πακέτο στο 10.0.3.0/24 πρέπει να το στείλει στο NextHop=10.1.1.1
- Για να το στείλει στο NextHop=10.1.1.1 πρέπει να συμβουλευτεί τον πίνακα δρομολόγησης του για μια αντίστοιχη εγγραφή
- Αυτό ονομάζεται **αναδρομική αναζήτηση**

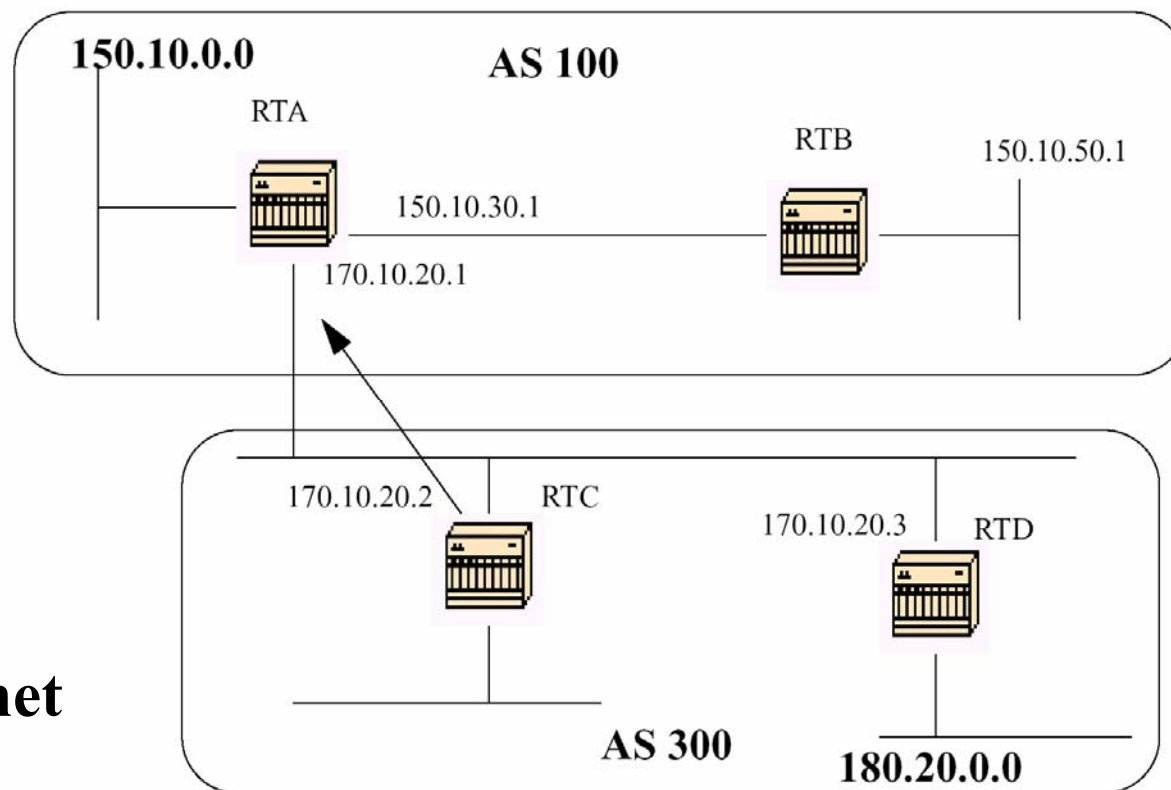
Η συνδεσιμότητα απαιτεί EGP + IGP



Συχνά η NextHop αλλάζει στην διεύθυνση βρόχου



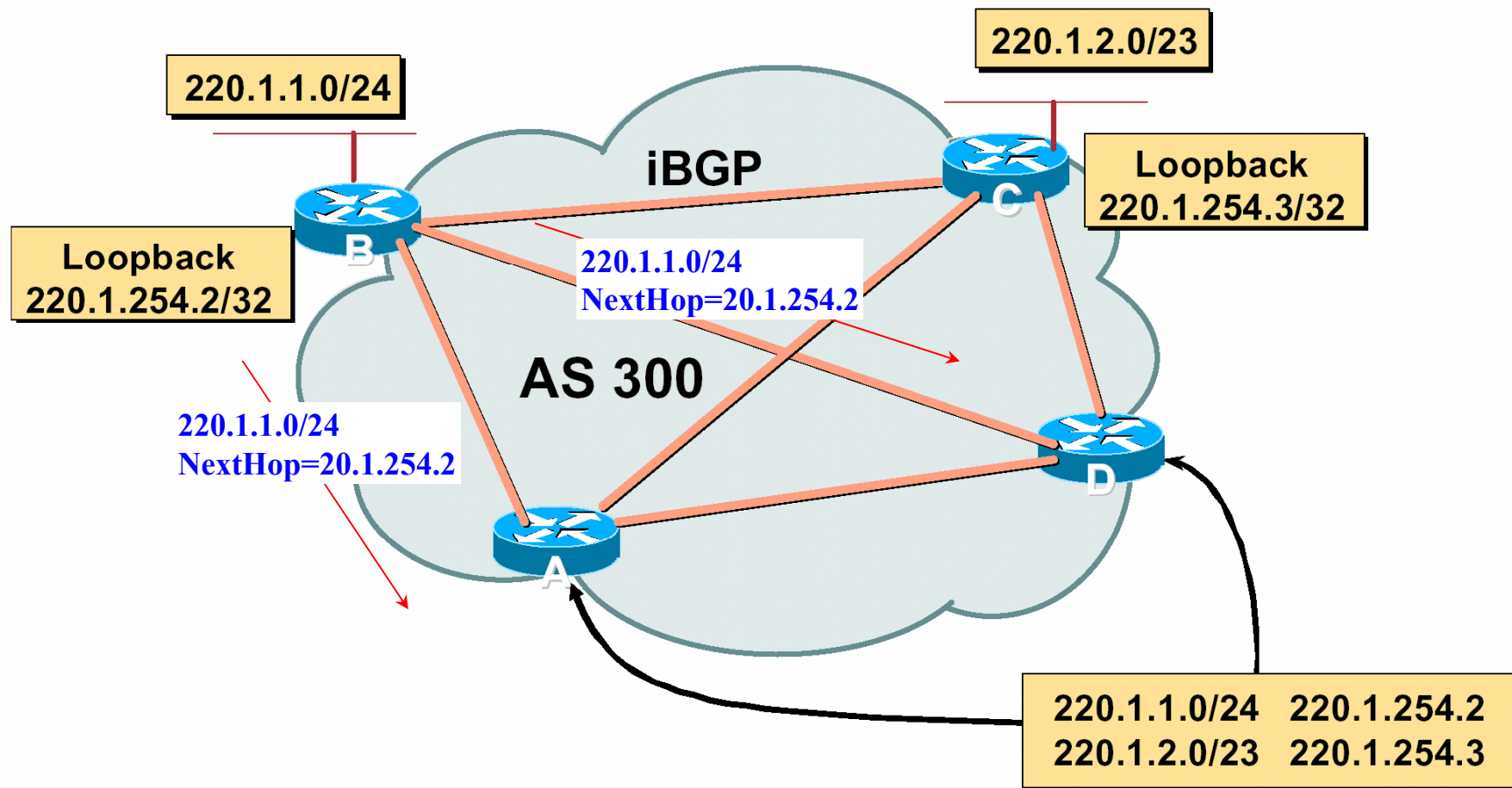
Next-Hop, Multiaccess networks



- Παράδειγμα είναι ένα **ethernet**
- Ο RTC, RTD τρέχουν OSPF
- Οι RTC, RTA τρέχουν BGP
- Ο RTC προσεγγίζει το δίκτυο 180.20.0.0 μέσω του NextHop 170.10.20.3
- Όταν ενημερώσει ο RTC τον RTA για το 180.20.0.0 θα του στείλει NextHop 170.10.20.3 **και όχι** την .2
- Αυτό διότι οι RTC, RTD βρίσκονται σε ένα δίκτυο πολλαπλής πρόσβασης και **δεν έχει νόημα** η δημιουργία ενός παραπάνω βήματος

Next Hop, IBGP

- Συνήθως είναι η **διεύθυνση βρόχου** του δρομολογητή, αλλιώς οποιαδήποτε φυσική διεύθυνση του δρομολογητή



Multi-Exit Discriminator (MED) ή metric

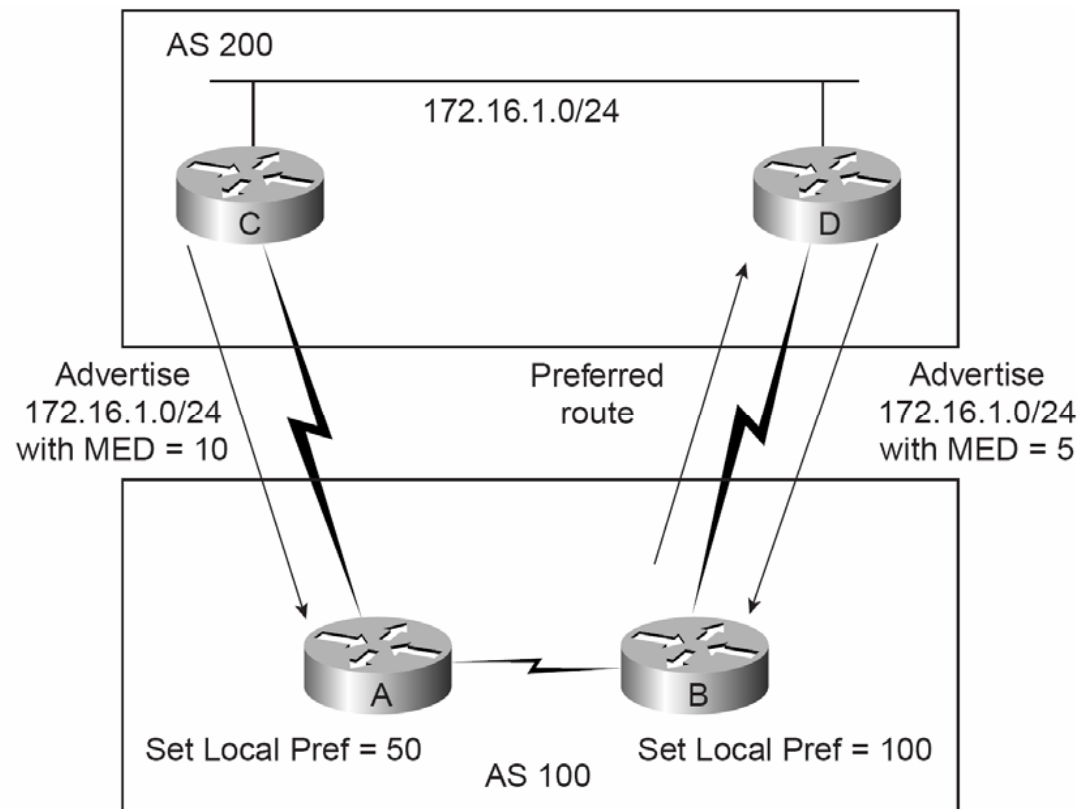
- Προαιρετική μη μεταβατική
- Χρησιμοποιείται σαν μία **πρόταση** (hint) σε ένα **εξωτερικό AS** σχετικά με την προτιμώμενη διαδρομή προς το AS το οποίο διαφήμισε την μετρική (του **προτείνουμε σημείο εισόδου** προς τα εμάς)
- Επηρεάζει την **εισερχόμενη** κίνηση
- Ο όρος πρόταση υπονοεί ότι το AS το οποίο έλαβε την MED, μπορεί να χρησιμοποιήσει **άλλες ιδιότητες** για την επιλογή των διαδρομών
- Δυναμικός τρόπος να επηρεάσουμε τις αποφάσεις των άλλων AS στην περίπτωση που έχουμε πολλές **πύλες εισόδου** στο AS
- Η μετρική ανταλλάσσεται **ανάμεσα στα AS**
- Λαμβάνεται από ένα AS, χρησιμοποιείται για τη **λήψη** απόφασης εντός του AS. Όταν η ενημέρωση περνάει σε ένα τρίτο AS η μετρική **μηδενίζεται**
- Εξ' ορισμού τιμή το **0**
- Ένα AS θα **θέσει** αυτή την τιμή και ένα άλλο θα τη χρησιμοποιήσει για να **αποφασίσει** για τη **βέλτιστη** διαδρομή

Multi-Exit Discriminator (MED) ή metric

- Είναι **συγκρίσιμες** τιμές αν **προέρχονται** από το **ίδιο AS**, έχει **τοπική** σημασία στο **AS**
- Υπάρχει εντολή με την οποία μπορούμε να ενεργοποιήσουμε τη **σύγκριση** των MED που προέρχονται από **διαφορετικά AS**
- Πρέπει αυτή να χρησιμοποιείται με **προσοχή** και φειδώ
- Οι τιμές έχουν εύρος **0 – 65535** (4 bytes)
- Προτιμώνται οι **χαμηλότερες** τιμές
- Τιμή **-1**, σημαίνει **κενό** πεδίο και αντιμετωπίζεται σαν η **υψηλότερη** προτίμηση
- Οι **IGP μετρικές** μπορεί να μετατραπούν σε **MED**
- Χρησιμοποιείται **συχνά**
- Σε βαρύτητα στην επιλογή των διαδρομών έρχεται **μετά** την **AS_path** και άρα επηρεάζει διαδομές με το **ίδιο AS_path**
- Οι NSPs διαφημίζουν τις MED μεταξύ τους ώστε να γνωρίζουν ποιο **POP**, Point of Presence είναι **κοντινότερο**

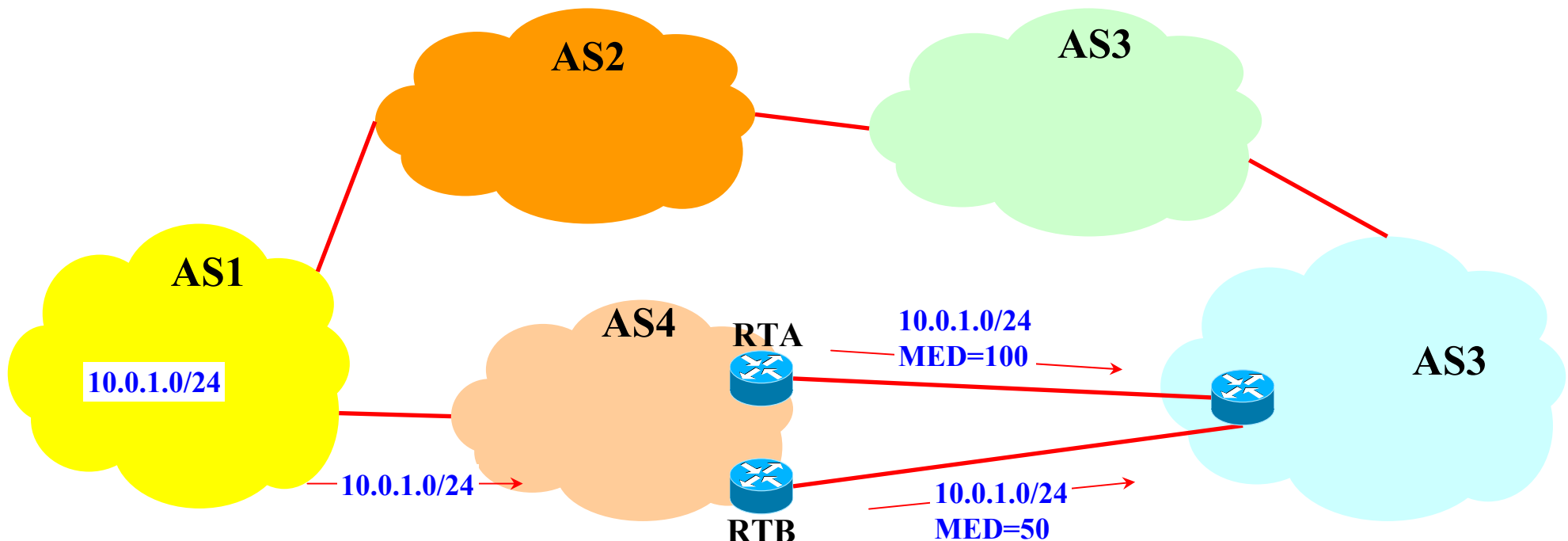
Παράδειγμα 1

- Ο RTC διαφημίζει την διαδρομή 172.16.1.0 με μετρική 10 , ενώ ο RTD τη διαφημίζει με μετρική 5
- Η χαμηλότερη θα προτιμηθεί, οπότε το AS 100 θα επιλέξει τη διαδρομή προς τον RTD για το υποδίκτυο 172.16.1.0
- Οι τιμές των MED διαφημίζονται σε όλο το **τοπικό AS**



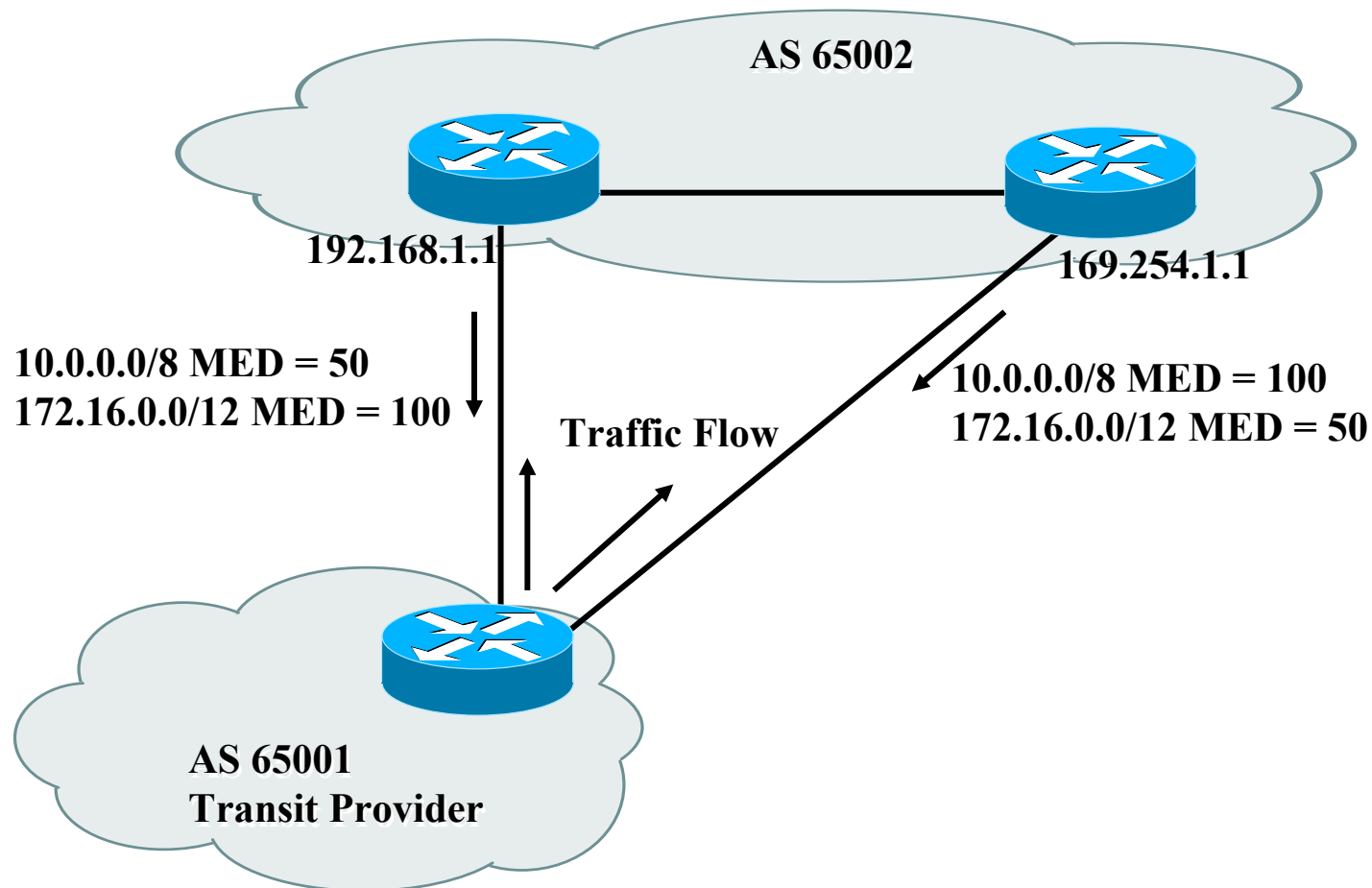
Παράδειγμα 2

- Το AS4 διαφημίζει στο AS3 το δίκτυο 10.0.1.0/24 με δύο διαφορετικά MED, 100 από τον RTA και 50 από τον RTB
- Ο AS3 αν λάβει υπόψη του θα έχει δύο διαδρομές και οι δύο με AS_Path={4,1} αλλά με διαφορετικά MED
- Έτσι θα επιλέξει για προορισμούς στο 10.0.1.0/24 να στείλει την κίνηση στον RTB



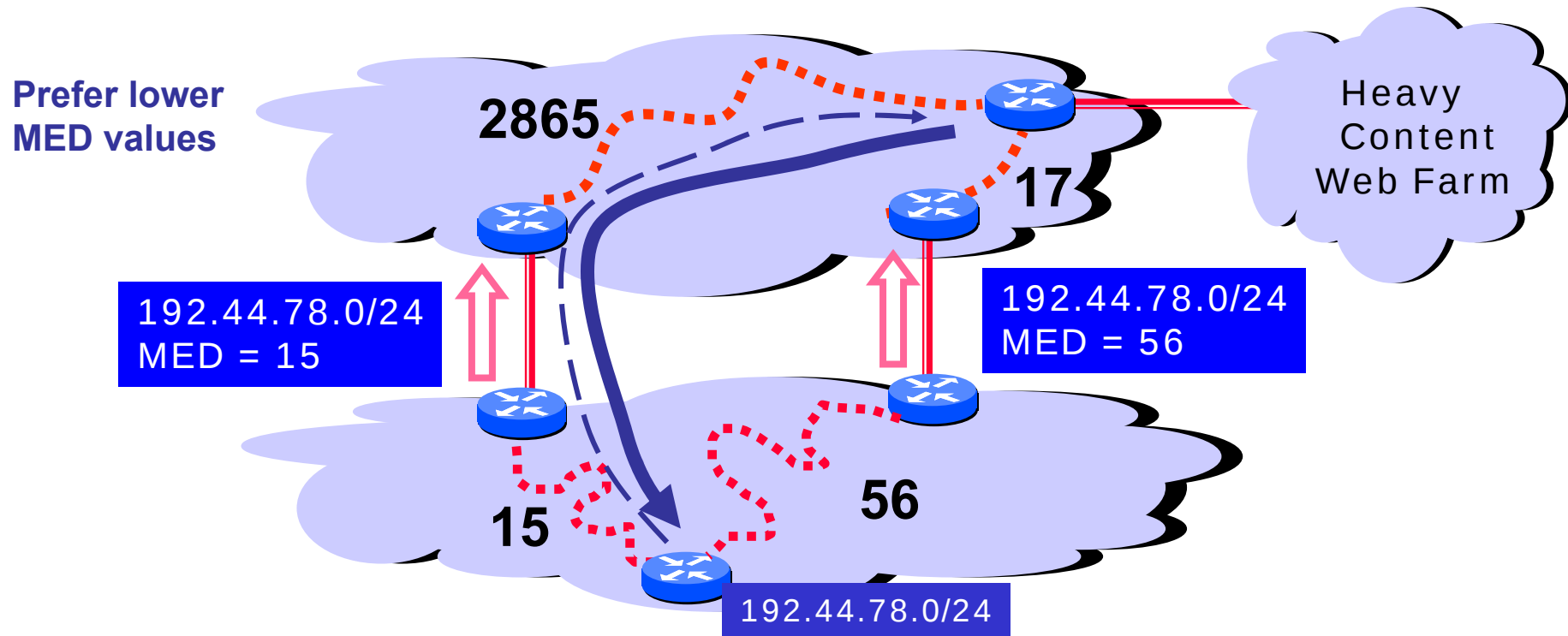
Διαμοιρασμός φόρτου, εφεδρεία

- Ορίζοντας διαφορετικές τιμές ανά πρόθεμα, μία χαμηλή και μία υψηλή, μπορούμε να πετύχουμε **διαμοιρασμό του φόρτου**, αλλά και **εφεδρεία**



IGP και MED

- Οι **IGP** μετρικές μπορεί να χρησιμοποιηθούν σαν **MED** ώστε το IGP δυναμικά να κατευθύνει την εισερχόμενη κίνηση στον **καλύτερο** δρομολογητή άκρης, μέσω BGP

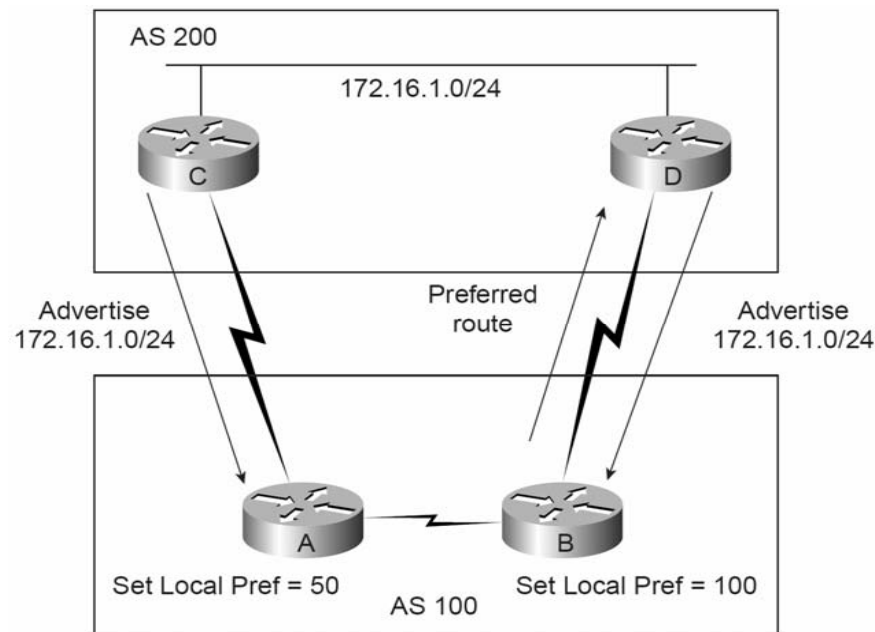


Τοπική προτίμηση (Local Preference)

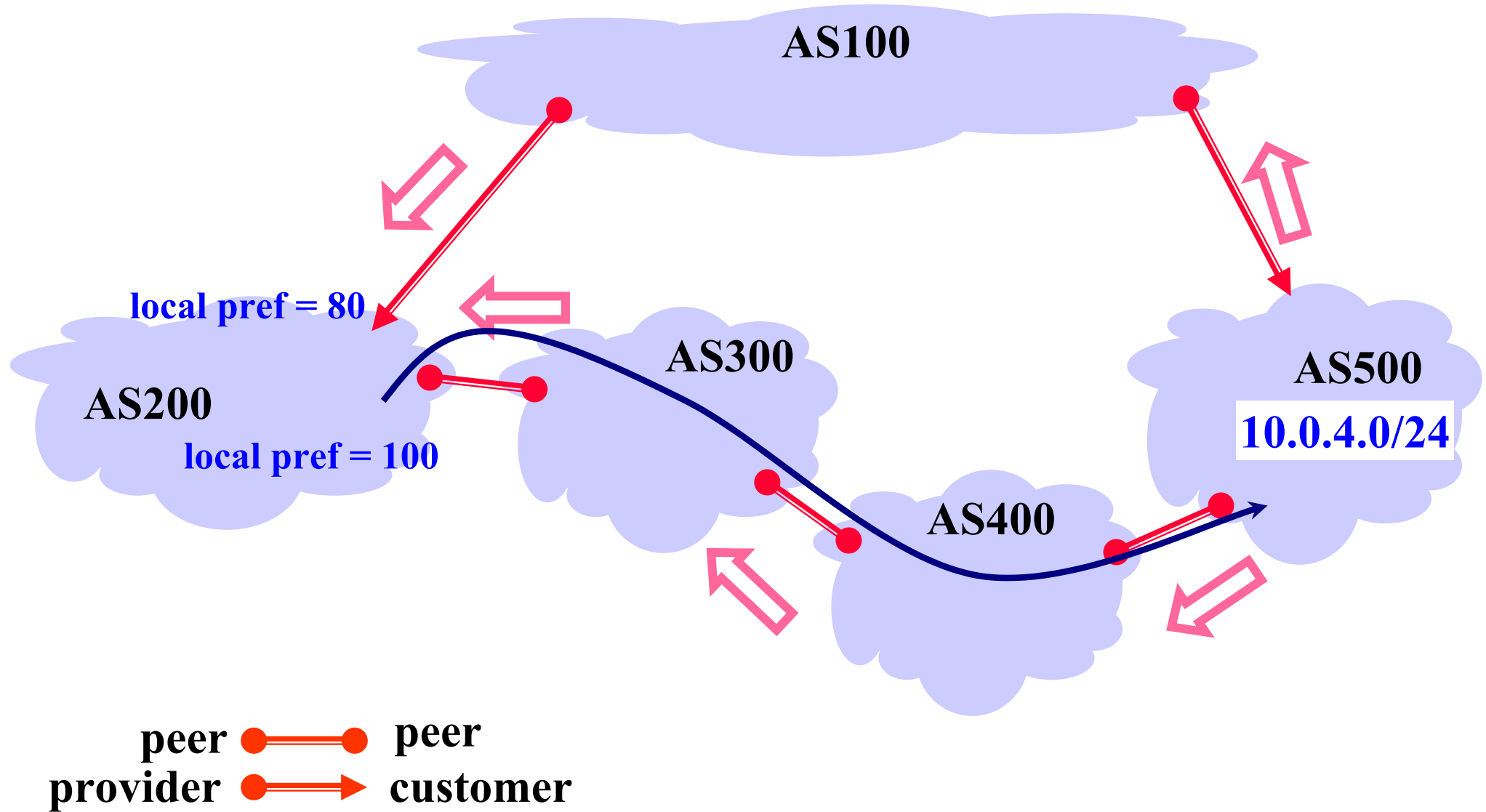
- Γνωστή διακριτική
- Διαδίδεται από άκρο σε άκρο στο **τοπικό AS** (μεταβατική)
- Έχει **τοπική σημασία** σε ένα AS
- Χρησιμοποιείται μόνο στο **IBGP**
- Οι **EBGP** ομότιμοι **ποτέ** δεν στέλνουν την τιμή αυτή, ενώ την αγνοούν αν την λάβουν
- Η μετρική που χρησιμοποιείται στο **IBGP** μόνο ώστε ο κάθε ένας να ξέρει ποια διαδρομή μέσα στο AS πρέπει να χρησιμοποιηθεί ώστε να προσεγγιστεί ένα ανακοινωθέν πρόθεμα
- Χρησιμοποιείται για να δώσουμε προτίμηση για ένα **σημείο εξόδου** από ένα AS για **συγκεκριμένο προορισμό**
- Εξ' ορισμού τιμή **100**, επιτρεπτές τιμές **0 – 65535** (4 bytes)
- Η διαδρομή με την **υψηλότερη τιμή** κερδίζει
- **Συχνά** χρησιμοποιούμενη
- Υπερνικά την AS_path έχοντας μεγαλύτερη **προτεραιότητα**

Παράδειγμα 1

- Το AS 100 λαμβάνει δύο διαφημίσεις για το δίκτυο 172.16.1.0 από το AS200
- Στον RTA η τοπική προτίμηση της διαδρομής τίθεται στο 50, ενώ στον RTB στο 100
- Οι RTA, RTB θα ανταλλάξουν τις διαδρομές αυτές μεταξύ τους
- Αυτή μέσω του RTB θα επιλεγεί να είναι το **σημείο εξόδου** από το **AS100** προς το δίκτυο **172.16.1.0**



Παράδειγμα 2

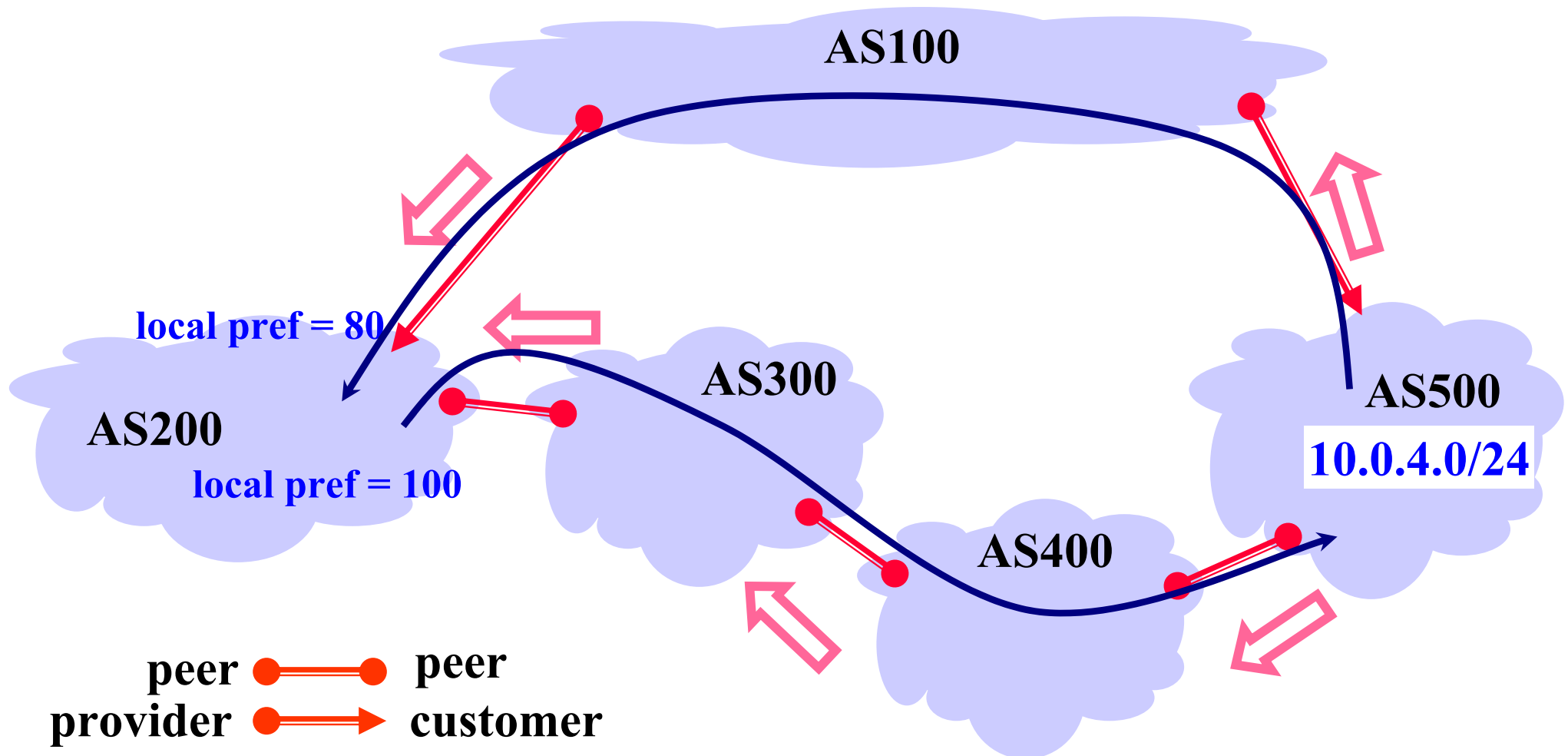


Παράδειγμα 2

- Από το AS200 προς το AS500 υπάρχουν δύο διαδρομές
- Αυτή μέσω των AS300, AS400 θα **προτιμηθεί** διότι έχει υψηλότερη **τοπική προτίμηση**
- Στο παράδειγμα το AS100 θα μπορούσε να είναι ένας **κοινός πάροχος** τον οποίο δεν θέλουμε να επιβαρύνουμε από τη στιγμή που έχουμε ομότιμες συνδέσεις ανάμεσα στα AS200, AS500
- Μπορεί να αντιπροσωπεύει επίσης ένα **δίκτυο κορμού** το οποίο δεν θέλουμε να επιβαρύνουμε παρότι είναι η πιο σύντομη διαδρομή, αφού υπάρχει εναλλακτική περιφερειακή σύνδεση

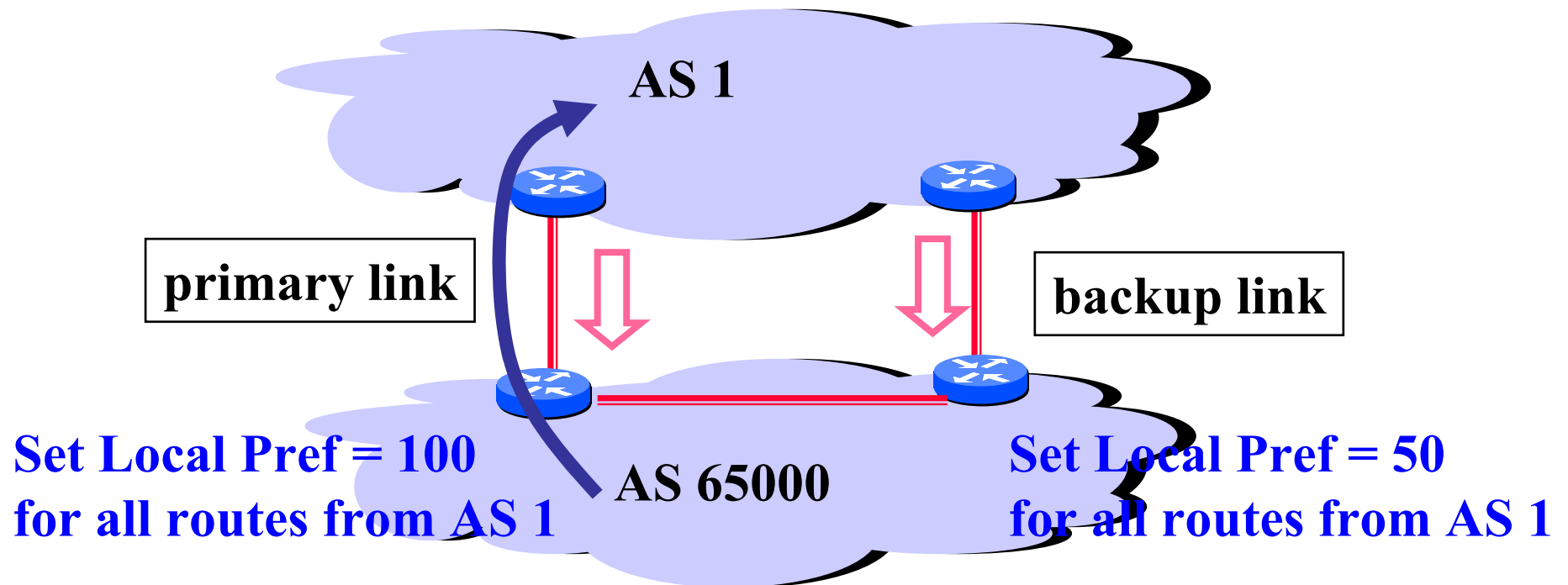
Παράδειγμα 2

- Παρατηρήστε ότι η κίνηση από το AS500 στο AS200 αν δεν έχουν οριστεί αντίστοιχα τιμές τοπικής προτίμησης θα ακολουθήσει τη διαδρομή μέσω του **AS100** λόγω της μικρότερης λίστας **AS_path**



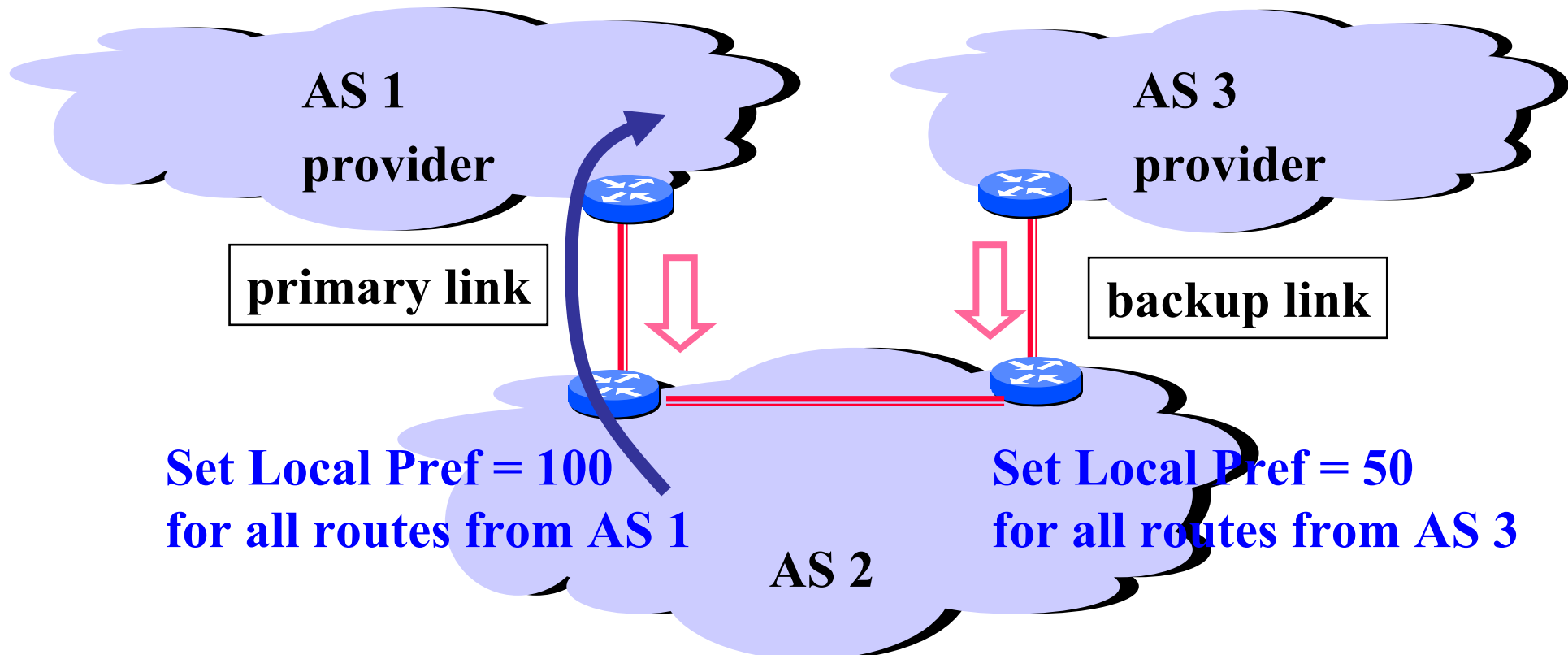
Υλοποίηση εφεδρικών συνδέσεων

- Αναγκάζουμε την **εξερχόμενη** από το AS65000 κίνηση να ακολουθήσει την πρωτεύουσα σύνδεση, εκτός αν η σύνδεση είναι εκτός λειτουργίας



Υλοποίηση εφεδρικών συνδέσεων Multihomed Backups

- Αναγκάζουμε την **εξερχόμενη** από το AS2 κίνηση να ακολουθήσει την πρωτεύουσα σύνδεση, προς το AS1, εκτός αν η σύνδεση είναι εκτός, οπότε χρησιμοποιείται η εφεδρεία προς το AS3



Atomic-Aggregate

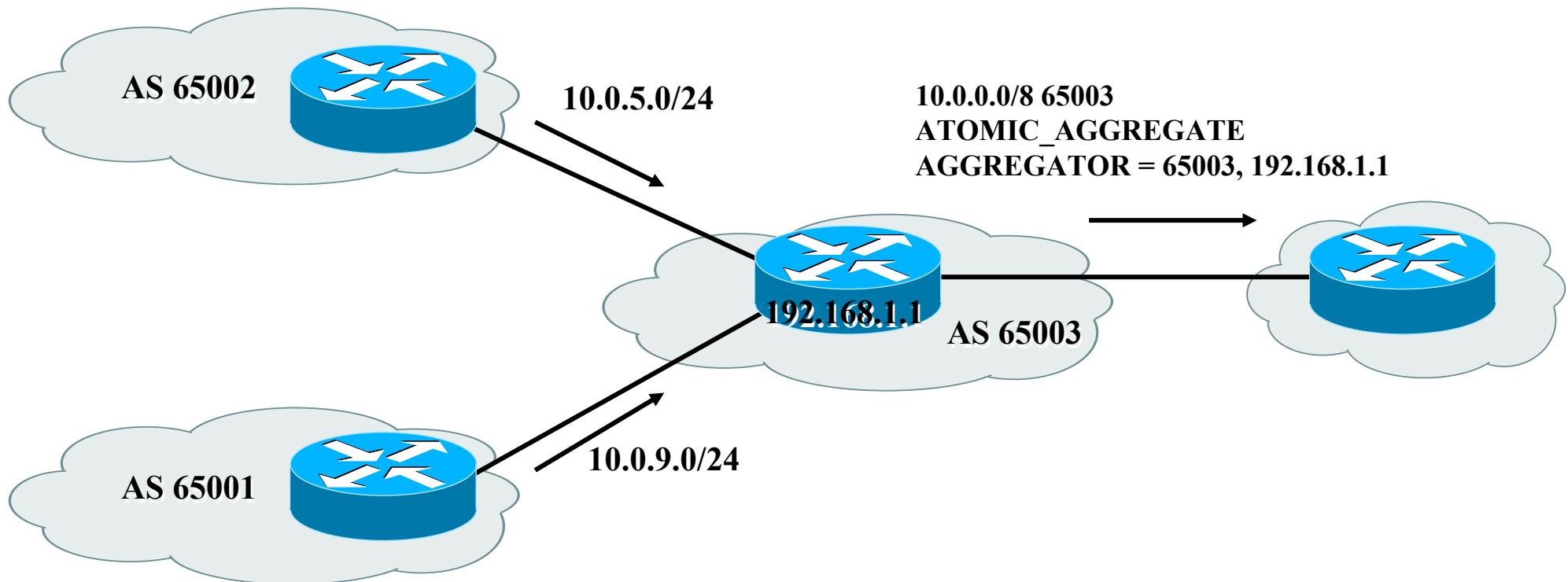
- Γνωστή διακριτική
- Έχει μήκος 0
- Επιτρέπει στους ομότιμους να ενημερώνει ο ένας τον άλλον για αποφάσεις που πήρανε όσον αφορά **επικαλυπτόμενες** διαδρομές
- Αν ο A λάβει επικαλυπτόμενες διαδρομές, και επιλέξει **την λιγότερο συγκεκριμένη**, τότε **θέτει την Atomic-Aggregate**, όταν ανακοινώσει τη διαδρομή αυτή σε ομότιμους
- Όταν ένας γείτονας λάβει ένα πρόθεμα με την ιδιότητα αυτή ορισμένη, **δεν** θα πρέπει να **απομακρύνει** την ιδιότητα αυτή όταν διαδώσει τη διαδρομή σε άλλους ομότιμους
- Επίσης δεν θα πρέπει να ανακοινώσει ένα πιο συγκεκριμένο πρόθεμα όταν ανακοινώνει τη διαδρομή σε άλλους ομότιμους, αφού έτσι θα χαθούν προορισμοί
- Ο δέκτης διαδρομής με αυτή την ιδιότητα πρέπει να γνωρίζει ότι η **πραγματική διαδρομή** προς τον προορισμό, μπορεί να είναι **διαφορετική** από αυτή που αναφέρεται στη λίστα **AS_path**

Aggregator

- Προαιρετική μεταβατική
- Μπορεί να οριστεί σε ένα **συναθροισμένο** (aggregated) πρόθεμα ώστε να προσδιορίσει το AS και τον δρομολογητή που έκανε τη συνάθροιση
- Είναι η **IP διεύθυνση του δρομολογητή** και το **AS** του που παρήγαγε την συναθροισμένη διαδρομή
- Χρήσιμο πεδίο για **debugging**
- Δεν επηρεάζει την **επιλογή** της καλύτερης διαδρομής
 - Έχει μήκος 6 bytes

Atomic-Aggregate, Aggregator, παράδειγμα

- Ο 192.168.1.1 συναθροίζει τις δύο διαδρομές για τα προθέματα 10.0.5.0/24 και 10.0.9.0/24 σε μία ανακοίνωση μόνο, την 10.0.0.0/8
- Ταυτόχρονα θέτει την Atomic_Aggregate και ορίζει στο πεδίο Aggregator τα στοιχεία του
- Αυτοί που λάβουν την ανακοίνωση δεν επιτρέπεται να κάνουν de-aggregate αφού έτσι θα χαθούν προορισμοί

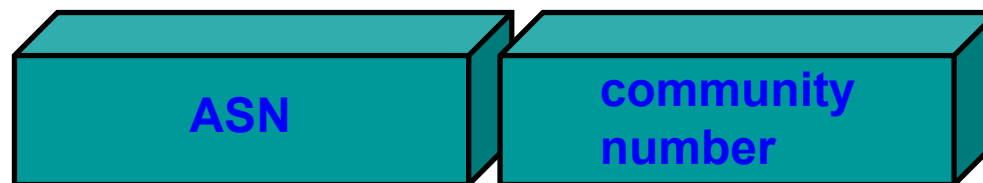


Community

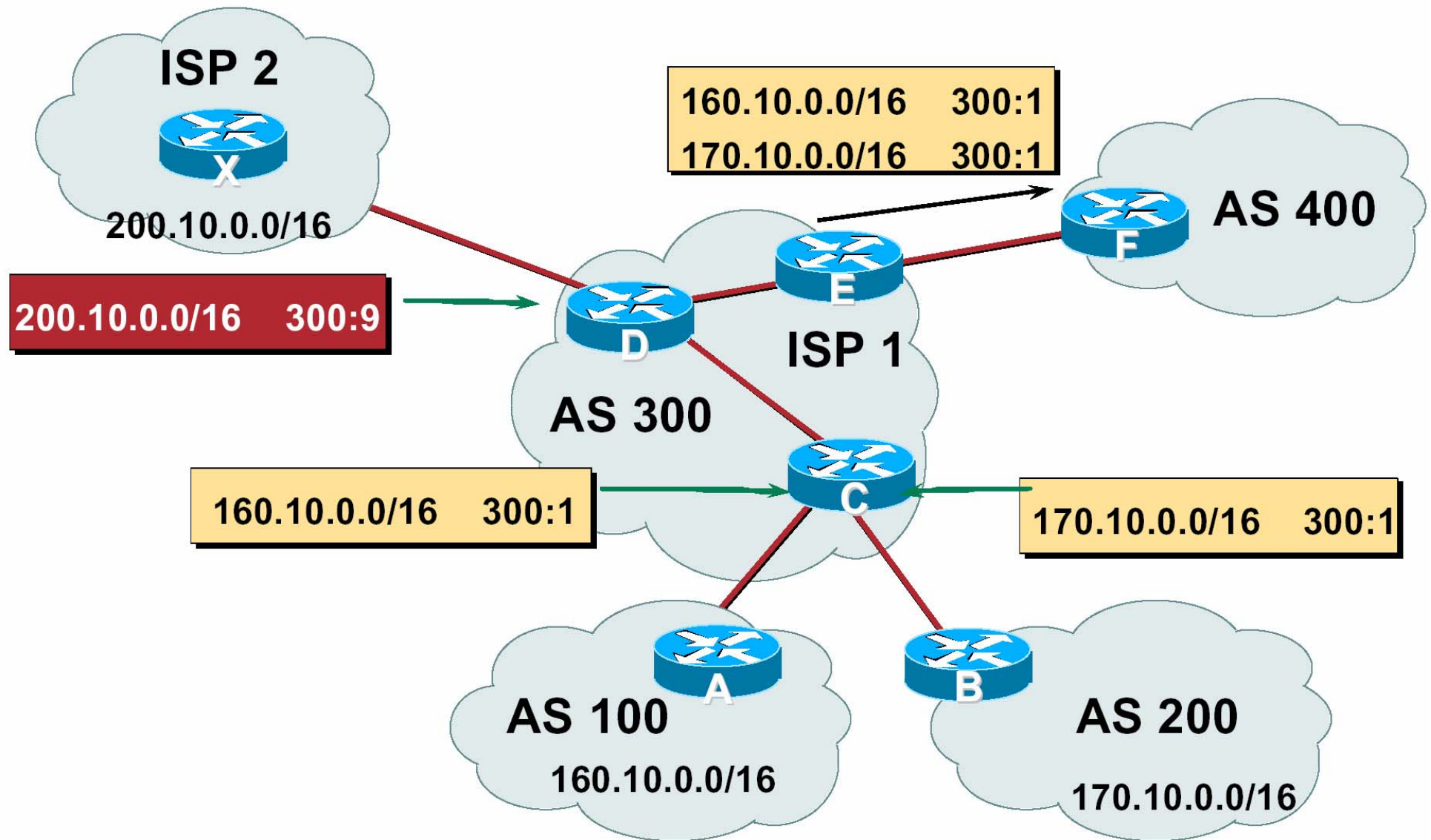
- **Προαιρετική μεταβατική**
- Χρησιμοποιείται ώστε το BGP να **ελέγχει** ποια πληροφορία δρομολόγησης **δέχεται, προτιμά ή διανείμει** σε άλλους ομότιμους
- Ένας ομιλητής που δέχεται μια διαδρομή η οποία δεν έχει την ιδιότητα αυτή μπορεί να την **ορίσει** όταν τη διαδώσει στους ομότιμους του
- Ένας ομιλητής που λαμβάνει μια διαδρομή που έχει την ιδιότητα αυτή μπορεί να την **τροποποιήσει** ανάλογα με την τακτική του
- Παρέχει ένα τρόπο να **ομαδοποιούμε τους προορισμούς** σε ομάδες που ονομάζουμε **κοινότητες** (communities)
- Σε αυτές **κοινές αποφάσεις δρομολόγησης** (όπως αποδοχή, προτίμηση, αναδιανομή) μπορούν να εφαρμοστούν
- Κάθε προορισμός μπορεί να είναι μέρος σε **πολλές κοινότητες**
- Είναι πολύ χρήσιμο για την εφαρμογή **τακτικής** (policy)

Community

- RFC1997, RFC1998
- Ακέραιος μήκους **32bit**, που αποτελείται από δύο ακεραίους των 16bit (0–65535) με μορφή <local-ASN>:xx
- Μέθοδος για χρωματισμό** (ομαδοποίηση - λογική ταξινόμηση) των προορισμό και εφαρμογή αποφάσεων δρομολόγησης με τη route-map (αποδοχή, προτίμηση, αναδιανομή)
- Σηματοδοσία** μέσα και ανάμεσα στα AS
- Ισχυρό** εργαλείο
- Δεν έχει προκαθορισμένη** σημασία
- Η ιδιότητα community στις ανακοινώσεις είναι μια **λίστα** από community τιμές

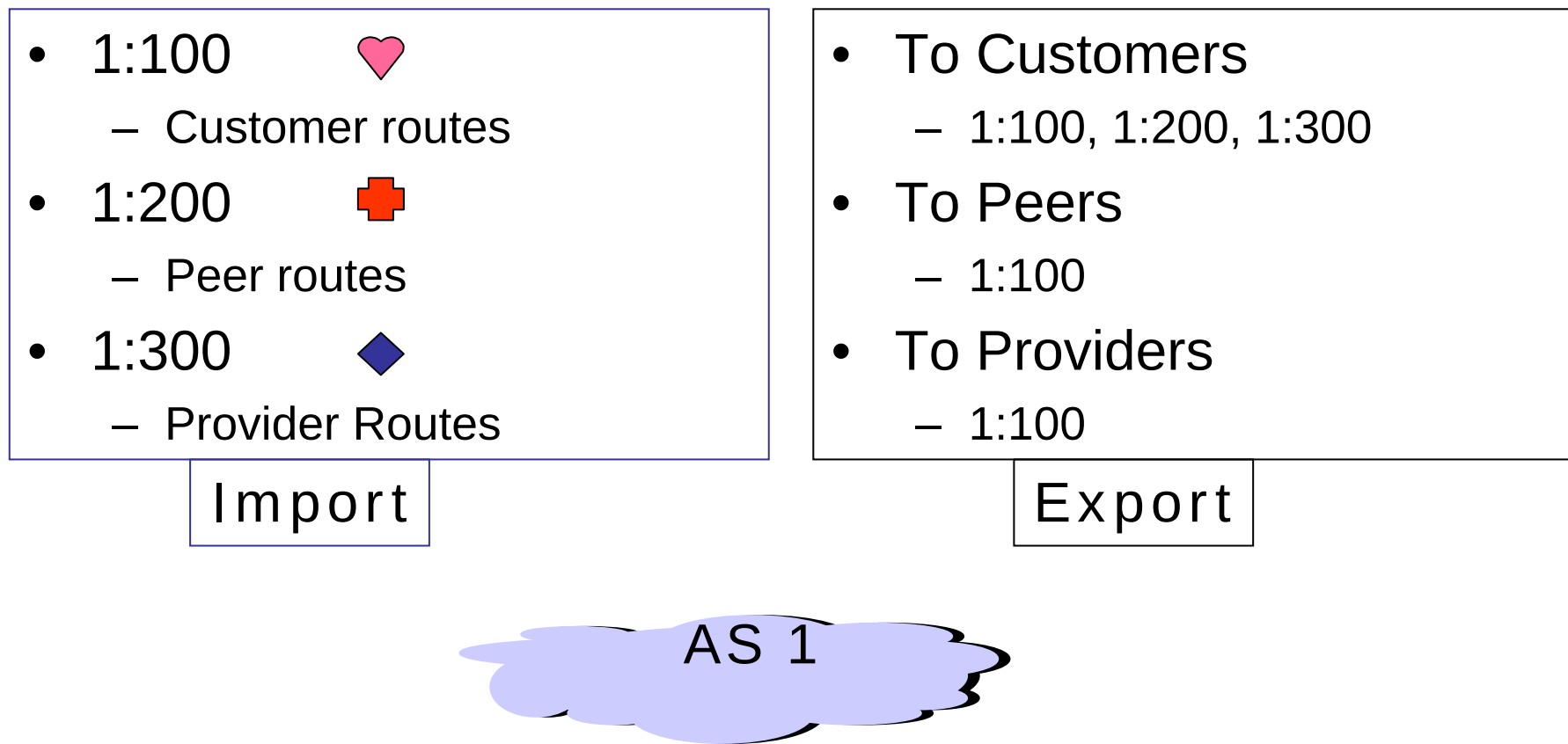


Παράδειγμα

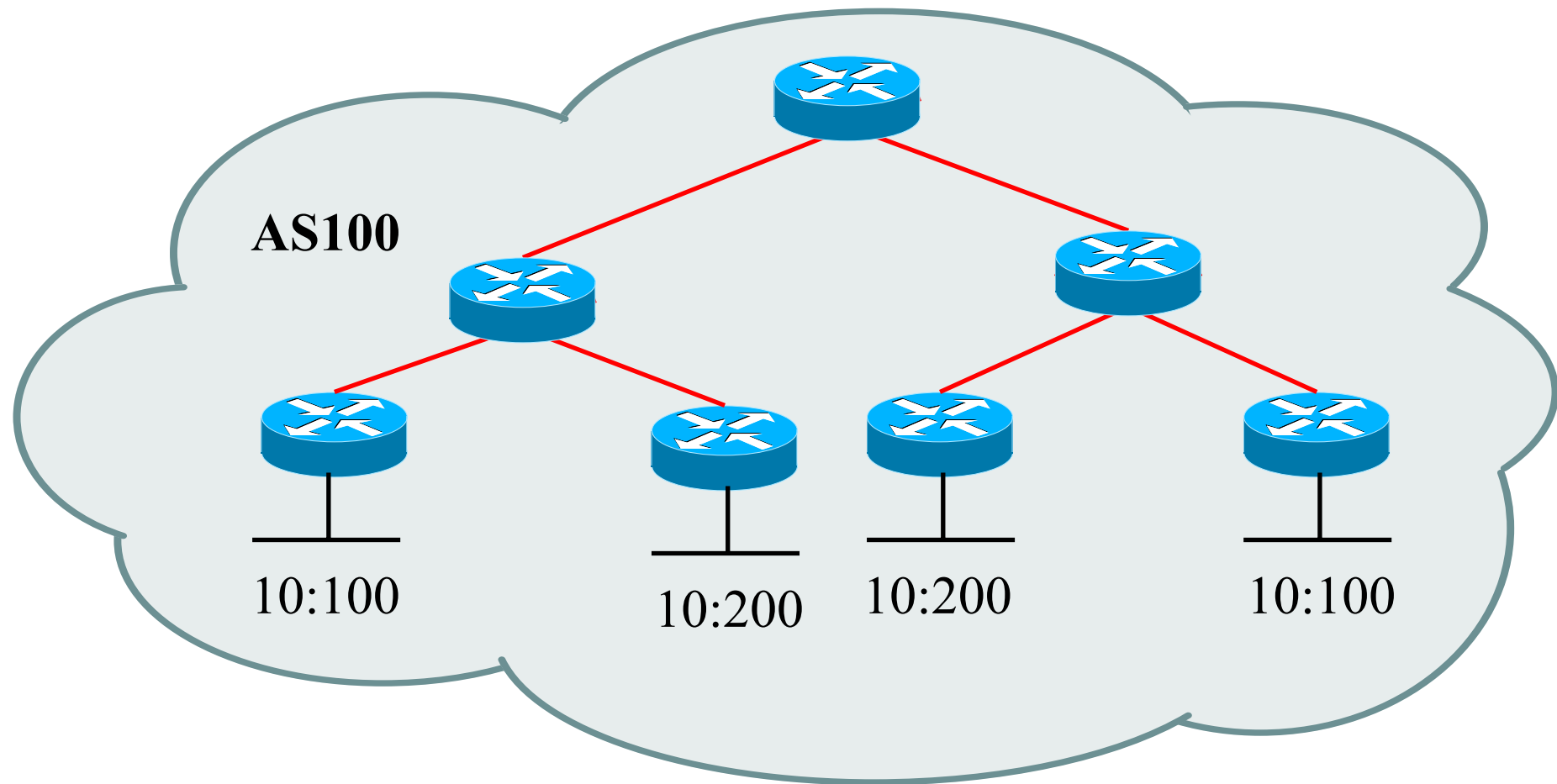


Δημιουργία σχέσεων

- **Χρωματίζουμε** τις διαδρομές ανάλογα με το από που μας έρχονται
Αναλόγως το χρωματισμό προωθούμε τις διαδρομές



Χρωματισμός πελατών



Προκαθορισμένες κοινότητες

- Οι τιμές 0:0 ως 0:65535 και 65535:0 ως 65535:65535 είναι δεσμευμένες
- Κάποιες **προκαθορισμένες** κοινότητες ορίζονται από αυτό το χώρο και είναι κατανοητές από όλους τους δρομολογητές

no-export (65535:65281 ή xFFFFFFFF01)

- Μην διαφημίσετε αυτή τη διαδρομή σε ομότιμους EBGP

no-advertise (65535:65282 ή 0xFFFFFFFF02)

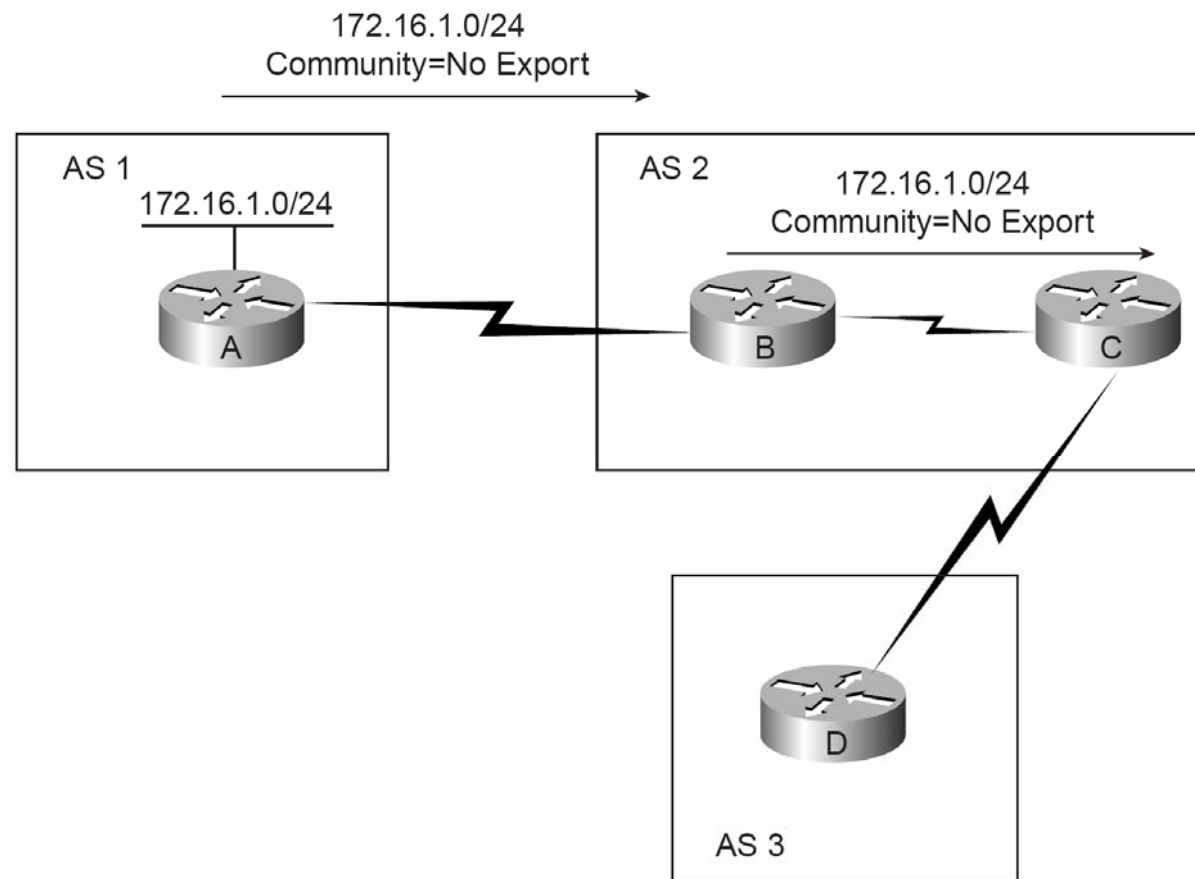
- Μην διαφημίσετε αυτή τη διαδρομή σε οποιοδήποτε ομότιμο
- Μη διαφημίσεις έξω από το τοπικό AS (χρησιμοποιείται μόνο με τα confederations)

Internet

- Διαφήμισε αυτή τη διαδρομή στην κοινότητα του Internet (όλοι οι δρομολογητές σε ένα δίκτυο ανήκουν σε αυτή)

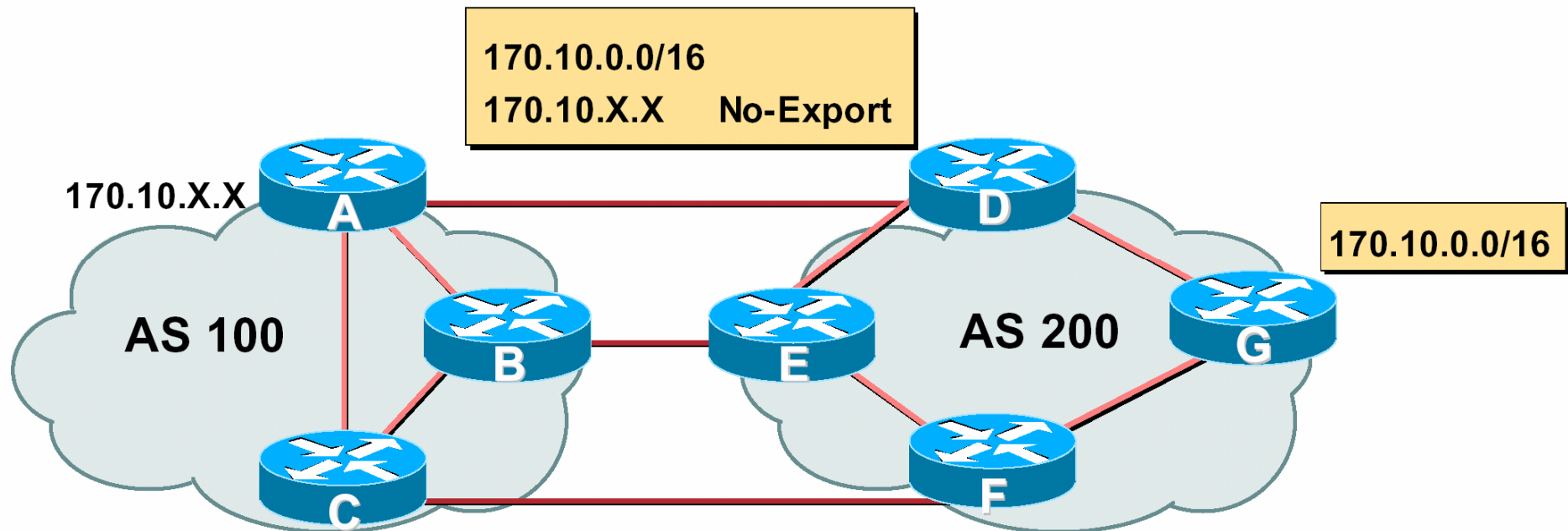
no-export community

- Το AS1 διαφημίζει το 172.16.1.0 στο AS2 με την ιδιότητα community = **no-export**
- Το AS2 θα διαδώσει τη διαδρομή στο AS2 αλλά δεν θα στείλει τη διαδρομή προς το AS3



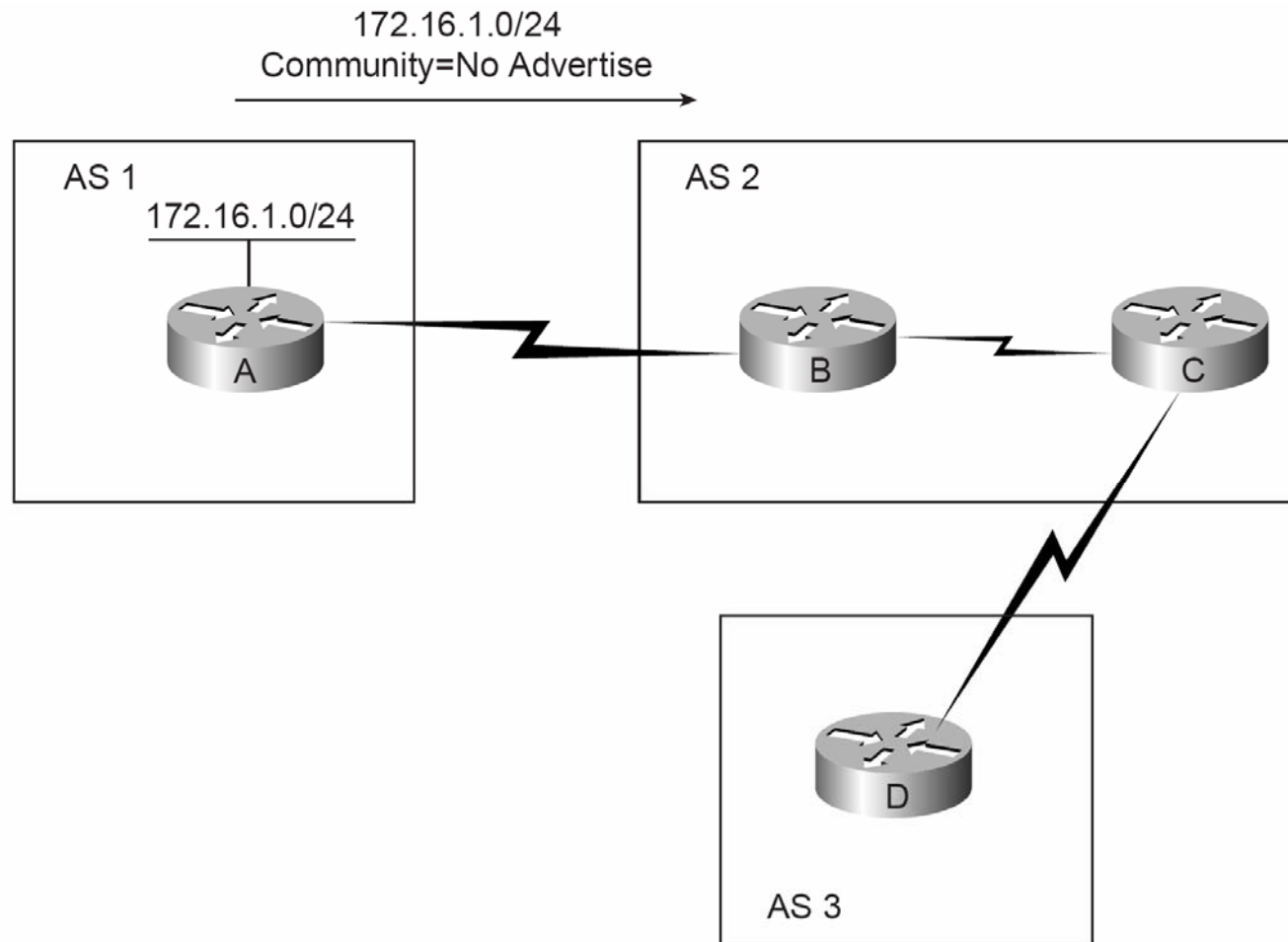
no-export community

- Το AS100 διαφημίζει το 170.10.0.0/16
- Επίσης διαφημίζει τα υποπροθέματα 170.10.X.Y με community=no-export
- Ο σκοπός είναι η βελτίωση του διαμοιρασμού του φορτίου, αφήνοντας να διαρρεύσουν τα υποπροθέματα
- Ο RTG δεν ανακοινώνει τα προθέματα με την ιδιότητα no-export community



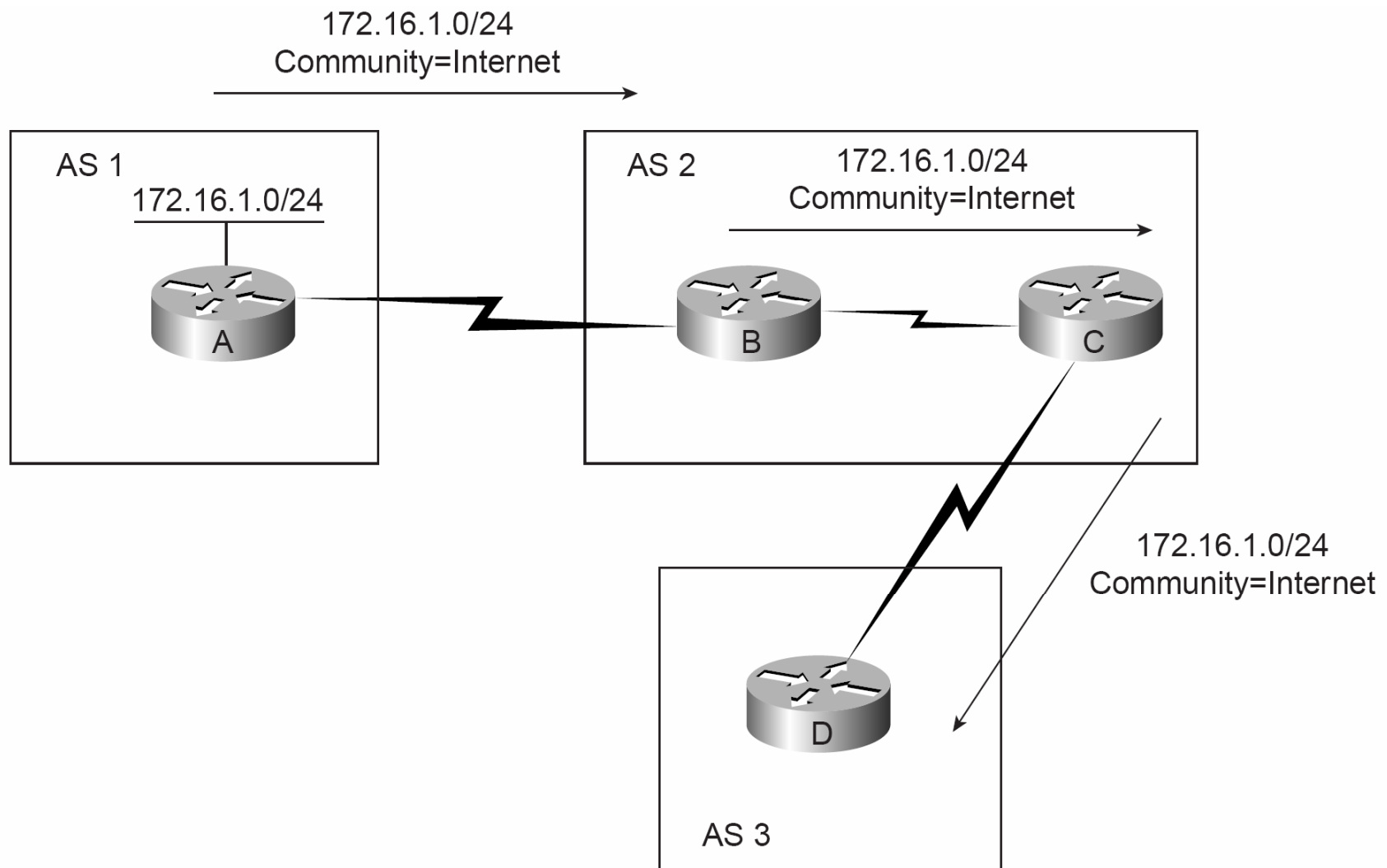
no-advertise community

- Το AS1 διαφημίζει το 172.16.1.0 στο AS2
- Ο RTB δεν θα διαφημίσει τη διαδρομή σε κανένα άλλο δρομολογητή



internet community

- Εδώ δεν υπάρχουν περιορισμοί στη διαφήμιση της διαδρομής



Extended Communities

- Προαιρετική μεταβατική
- Ίδιος τρόπος λειτουργία με την ιδιότητα Community
- Επιτρέπουν πιο **ευρεία** κλίμακα τιμών και έχουν πιο καθορισμένη δομή
- Έχει **8 bytes** μέγεθος και ορίζεται σαν 2 byte τύπος + 6 byte τιμή
- Κάθε byte στον **τύπο** ορίζει τη **δομή** και τη **σημασία** της **τιμής**
- Δεν έχουν χρησιμοποιηθεί ευρέως
- Χρησιμοποιούνται συνήθως μόνο σε συνδιασμό με **MPLS VPNs**

Extended Communities

Αρκετοί τύποι extended communities έχουν οριστεί:

Route Target Community (extended type 0x02)

- Ορίζει ένα προορισμό για ένα πρόθεμα, το οποίο περνάει διαφανώς από τα AS σύνορα
- Ορίζεται ως “target:AS:value” ή “target:IP address:value”

Route Origin Community (extended type 0x03)

- Ταυτοποιεί το δημιουργό ενός προθέματος, περνάει διαφανώς από τα AS σύνορα
- Ορίζεται ως “origin:AS:value” ή “origin:IP address:value”

Link Bandwidth Community (extended type 0x04)

- Ορίζει μια μετρική για μία σύνδεση ανάμεσα σε IBGP ή EBGP ομότιμους, περνάει διαφανώς από τα AS σύνορα

Βάρος (Weight)

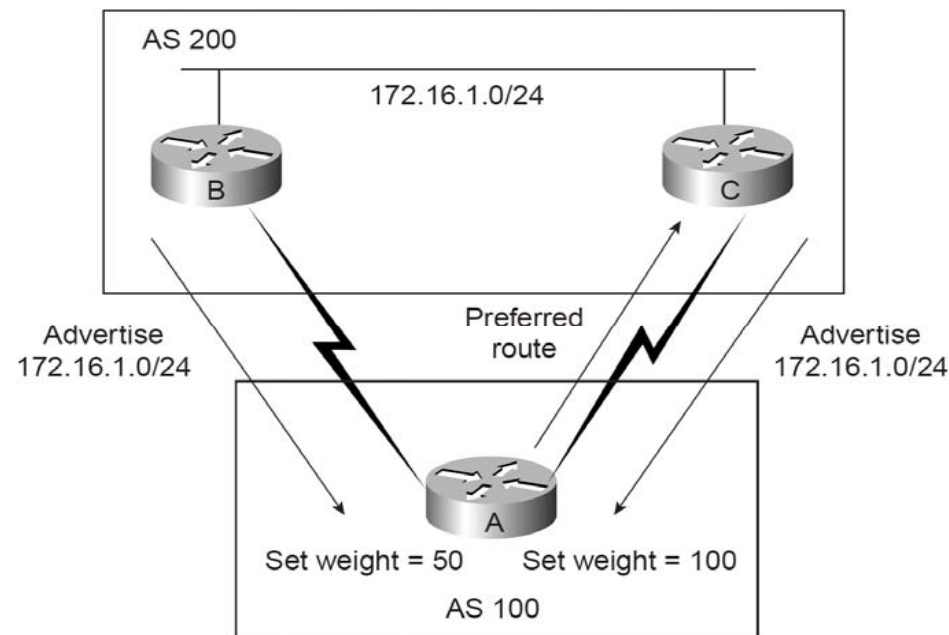
- Ορισμένη από την **Cisco**, **δεν ανήκει σε κάποια προδιαγραφή**
- Έχει **τοπική** σημασία στο δρομολογητή, **δεν** είναι στην πραγματικότητα ιδιότητα
- Τιμές 0 - 65535
- Η εξ' ορισμού είναι η **32768** για τα δίκτυα που είναι **τοπικά** και **0** για όλα τα **άλλα**
- Η **υψηλότερη** τιμή προτιμάται
- Χρησιμοποιείται **σπάνια** και είναι **ισχυρότερη** από όλες τις άλλες ιδιότητες
- Δεν ανακοινώνεται σε άλλους δρομολογητές **ούτε** αυτούς που είναι στο **ίδιο AS**
- Έχει περισσότερο το νόημα **πρόχειρης** διάρθρωσης

Βάρος (Weight)

- Επιτρέπει τον έλεγχο παρομοία με την ιδιότητα της **τοπικής προτίμησης**
- Η διαφορά είναι ότι **δεν διαφημίζεται** στους γειτονικούς δρομολογητές
- Αν υπάρχουν περισσότερες διαδρομές προς τον ίδιο προορισμό αυτή με το **μεγαλύτερο βάρος** προτιμάται
- Μπορούμε να πούμε ότι είναι ο **βαθμός προτίμησης** που δίνει ο **δρομολογητής** σε κάθε μία από τις **εξόδους** του

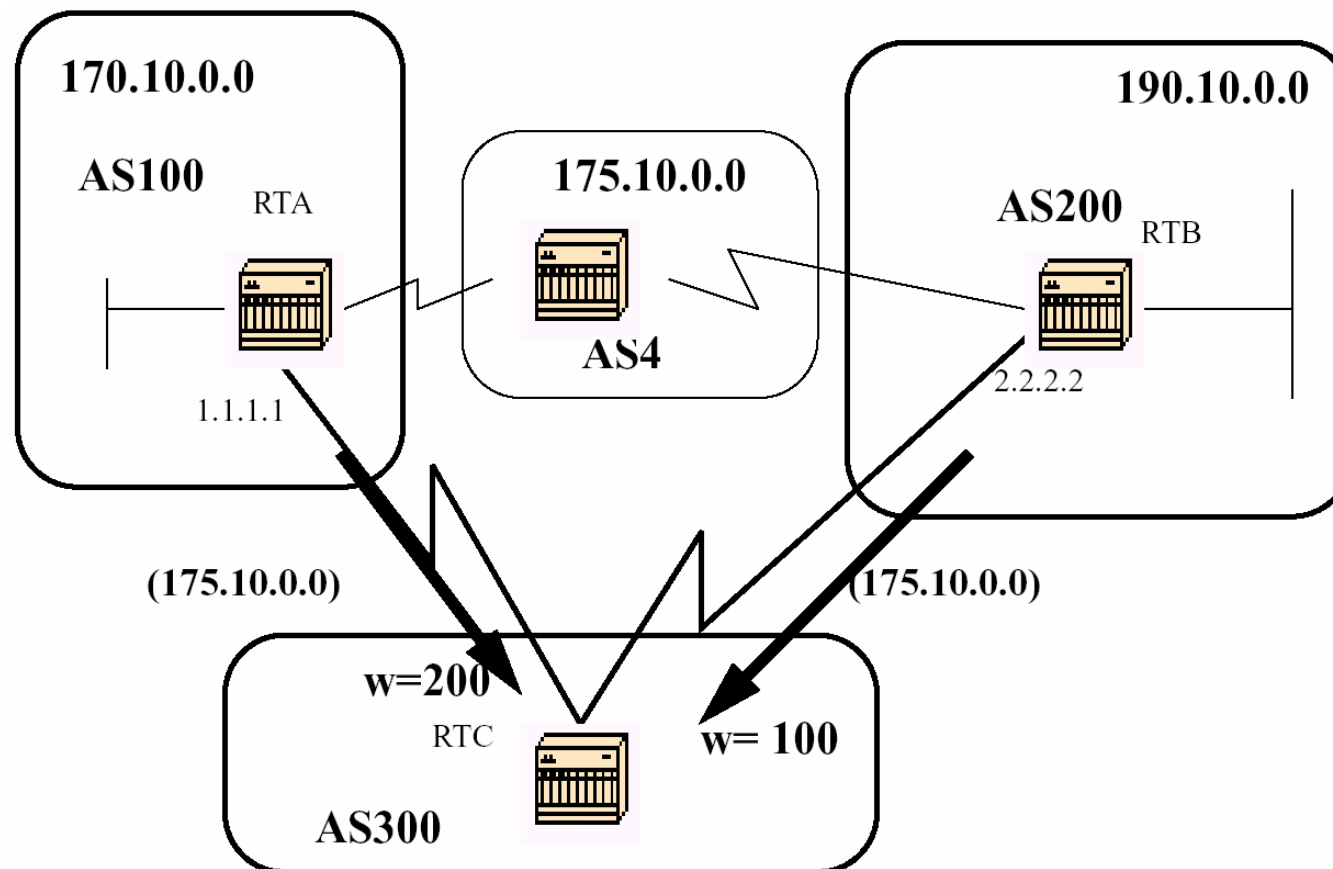
Παράδειγμα 1

- Ο RTA λαμβάνει από μία διαφήμιση για το δίκτυο 172.16.1.0 από τους RTB, RTC
- Το βάρος για τις διαδρομές που μαθεύονται από τον RTB είναι 50, ενώ για αυτές από τον RTC είναι 100
- Και οι δύο διαδρομές παραμένουν στον πίνακα δρομολόγησης του BGP, αλλά αυτή με το **μεγαλύτερο βάρος** εγκαθίσταται στον πίνακα δρομολόγησης του πυρήνα και ανακοινώνεται παραπέρα



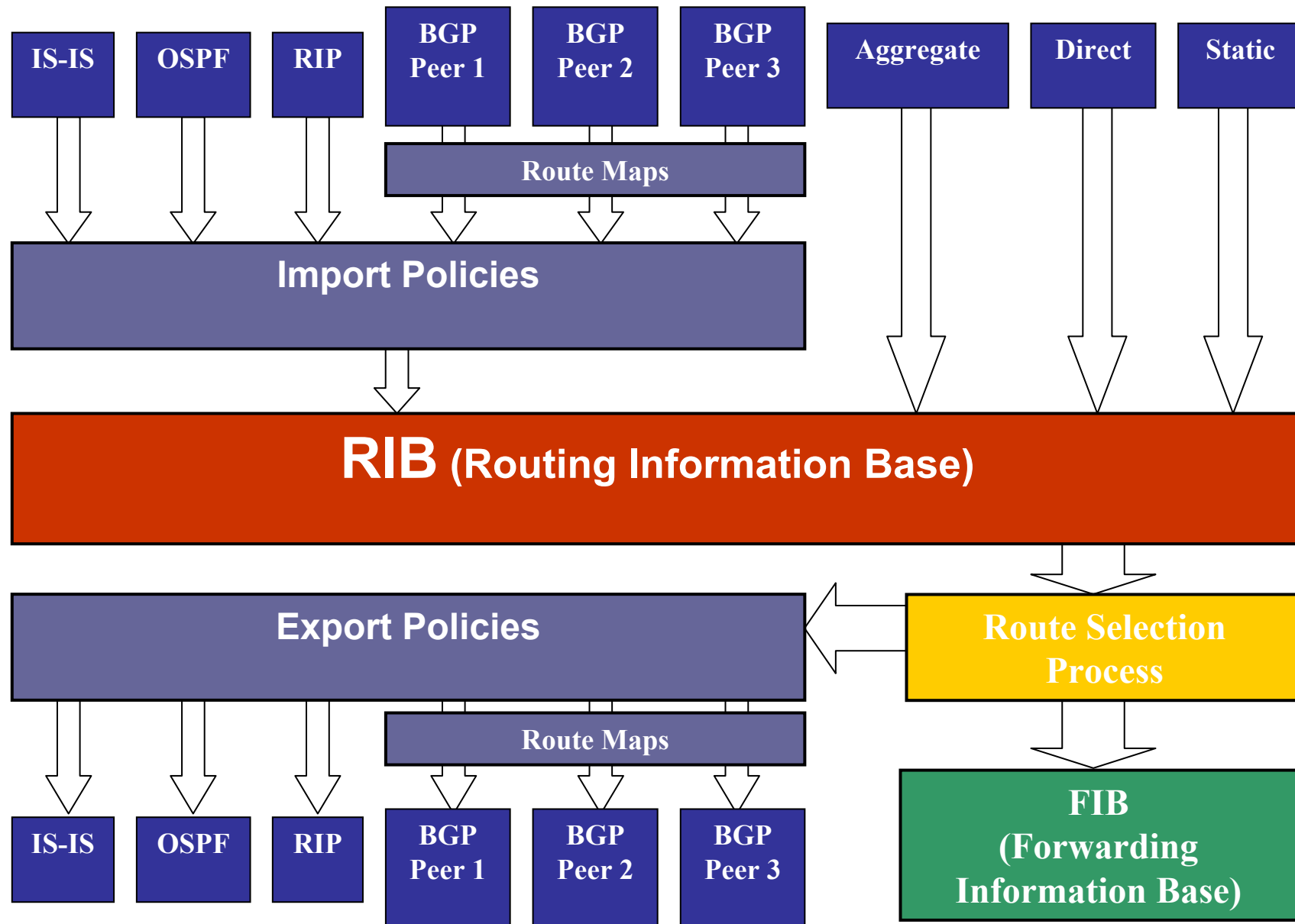
Παράδειγμα 2

- Ο RTC έχει λάβει δύο διαδρομές προς το δίκτυο 175.10.0.0
- Επειδή έχουμε ορίσει το βάρος για τις διαδρομές που λαμβάνει ο RTC από τον RTA σε 200, ενώ από τον RTB σε 100, θα γίνει αποστολή προς τον RTA



Επιλογή διαδρομής

RIB - FIB



RIB - FIB

- Οι διαδρομές γίνονται γνωστές από **διαφορετικές** πηγές, κατευθείαν συνδεδεμένα δίκτυα, στατικές διαδρομές, συναθροισμένες διαδρομές, IGP, EGP
- Η **RIB, Routing Information Base** κρατεί όλες τις διαδρομές από όλα τα πρωτόκολλα
- Οι εγγραφές μπορούν να ελεγχθούν-φιλτραριστούν
- Η **FIB, Forwarding Information Base** κρατεί τη καλύτερη διαδρομή για κάποιο προορισμό και χρησιμοποιείται για την προώθηση της κίνησης βασισμένο στο καλύτερο ταίριασμα με τον προορισμό
- Επιλογή διαδρομής** είναι η διαδικασία με την οποία επιλέγεται μια διαδρομή από όλες τις διαθέσιμες της RIB προς εγκατάσταση στην FIB
- Οι διαδρομές της FIB μπορούν να **εξαχθούν** σε άλλα πρωτόκολλα αφού φιλτραριστούν

RIBs στο BGP

- Η RFC 1771 ορίζει τρεις RIB για το BGP

Adj-RIBs-In

- Όλες οι διαδρομές που ελήφθησαν μέσω Update μηνυμάτων
- Αυτές οι διαδρομές είναι διαθέσιμες στον αλγόριθμο επιλογής
- **bgp show peer-host x.x.x.x all-received-routes**

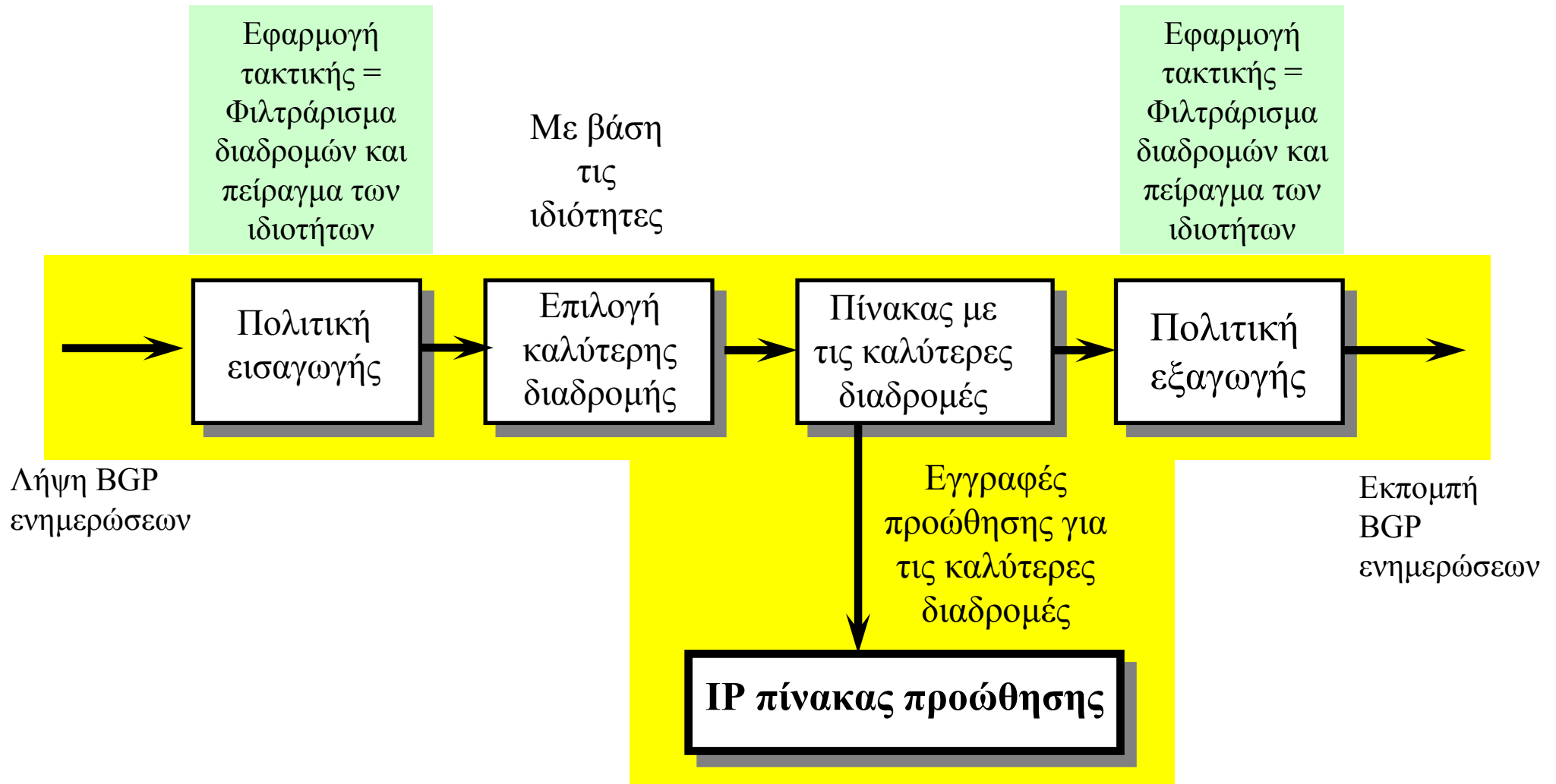
Loc-RIB

- Όλες οι έγκυρες BGP διαδρομές, περιέχει την πληροφορία δρομολόγησης της **Adj-RIBs-In** μετά την εφαρμογή της τακτικής δρομολόγησης
- **bgp show routes all**

Adj-RIBs-Out

- Οι διαδρομές που επέλεξε το πρωτόκολλο για να ανακοινώσει στους ομότιμους
- **bgp show peer-host x.x.x.x advertised routes**

Επεξεργασία διαδρομής



Επεξεργασία διαδρομής

- Για κάθε εισερχόμενη διαδρομή γίνεται έλεγχος για την **εγκυρότητα** της, περνάει από ένα **φίλτρο** και αν το περάσει τοποθετείται στον **πίνακα δρομολόγησης του BGP**
- Αν η καινούργια αναφέρεται σε ένα προορισμό για τον οποίο υπάρχουν **και άλλες διαδρομές**, τότε γίνεται ένας υπολογισμός ώστε να βρεθεί η **καλύτερη** από όλες
- Αν η καλύτερη αλλάξει τότε πρέπει να ενημερωθούν οι **RIB** και **FIB**
- Παράλληλα ενημερώνονται και οι **ομότιμοι**, αφού πρώτα περάσει η ενημέρωση από ένα **φίλτρο** εξόδου

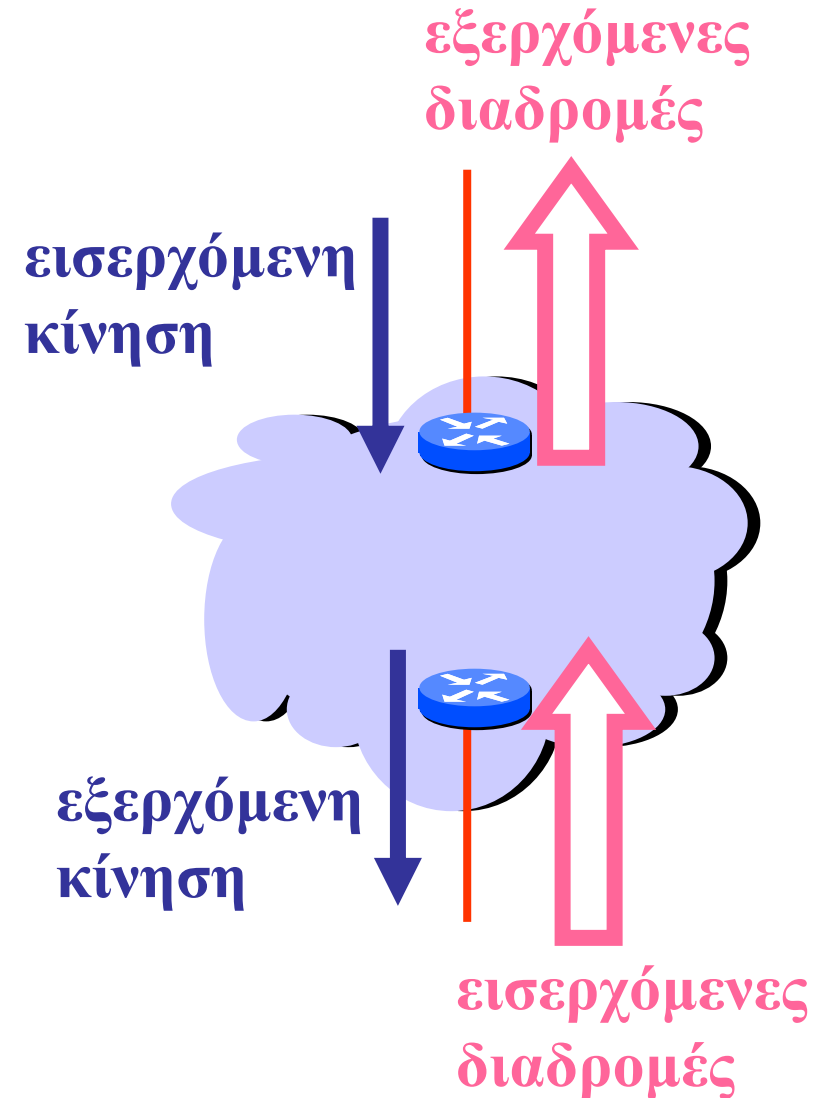
Πολιτική εισαγωγής - εξαγωγής

Εισερχόμενη κίνηση

- **Φιλτράρισμα** στις εξερχόμενες διαδρομές
- **Πείραγμα** των ιδιοτήτων των εξερχόμενων διαδρομών με την ελπίδα να **επηρεάσουμε** τις επιλογές βέλτιστων διαδρομών των γειτόνων

Εξερχόμενη κίνηση

- **Φιλτράρισμα** εισερχόμενων διαδρομών
- **Πείραγμα** των ιδιοτήτων των εισερχομένων διαδρομών ώστε να **επηρεάσουμε** την επιλογή της βέλτιστης διαδρομής
- Γενικά μπορούμε να έχουμε περισσότερο έλεγχο στην εξερχόμενη κίνηση



Λήψη Update μηνύματος

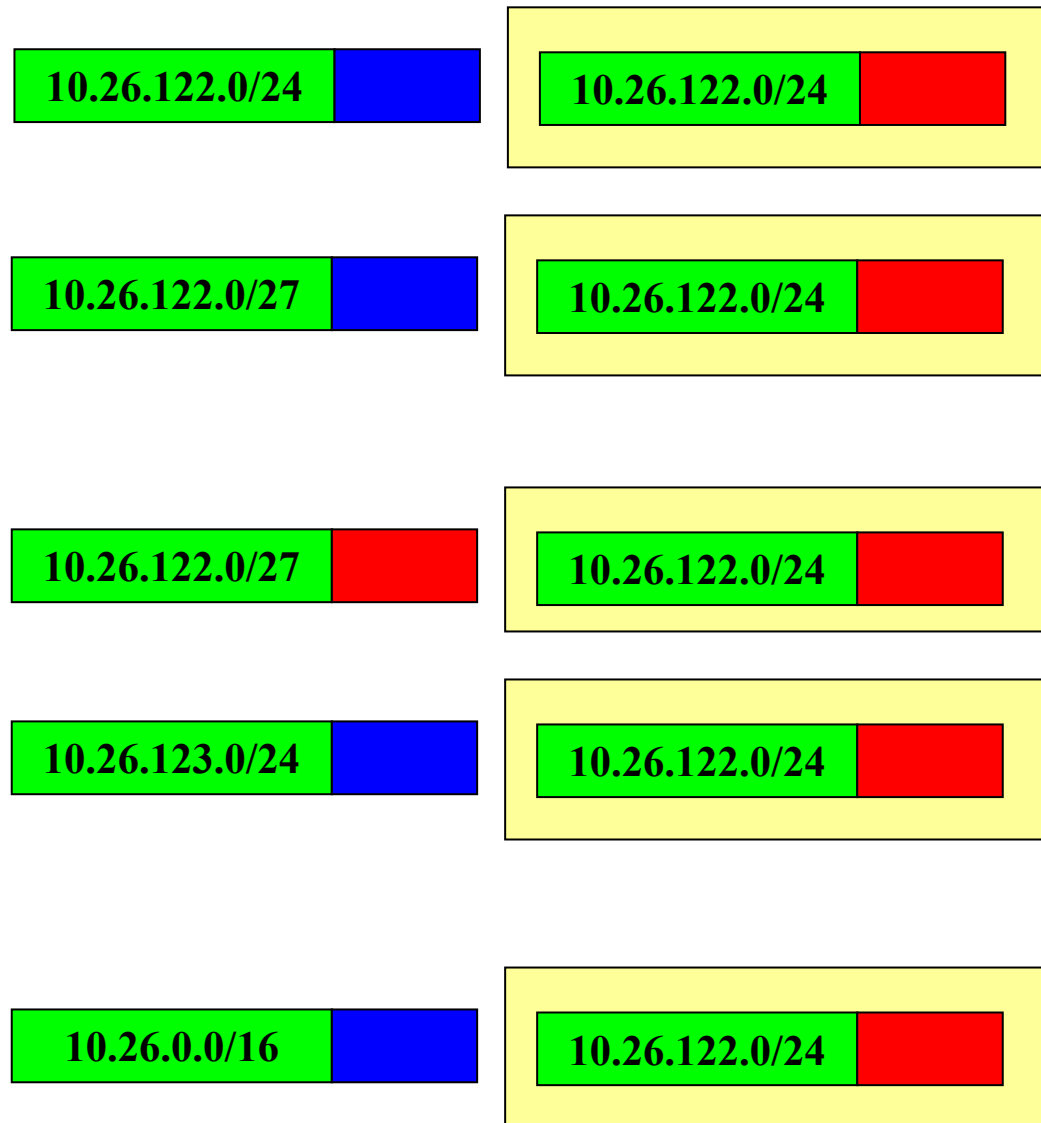
- Έλεγχος για την **ορθότητα** του
- Αν μία ιδιότητα είναι **μη μεταβατική** και **δεν** είναι **αναγνωρίσιμη**, τότε αγνοείται
- Αν μία **προαιρετική μεταβατική** ιδιότητα **δεν** είναι **αναγνωρίσιμη**, τότε το **Partial bit** γίνεται 1 και η ιδιότητα κρατείται για **μετάδοση** στους ομότιμους
- Αν μια **προαιρετική** ιδιότητα **αναγνωρίζεται** και έχει έγκυρη τιμή τότε γίνεται **επεξεργασία** της και αν χρειάζεται **ενημερώνεται** ώστε να **διαδωθεί** πιθανώς προς τους ομότιμους
- Αν η ενημέρωση περιέχει διαδρομές προς **απόσυρση**, αυτές θα πρέπει να **αφαιρεθούν** από τη βάση **Adj-RIB- In**
- Στη συνέχεια θα πρέπει να εκτελεστεί μια **διαδικασία επιλογής** αφού πλέον η διαδρομή **δεν** είναι διαθέσιμη

Λήψη νέας διαδρομής

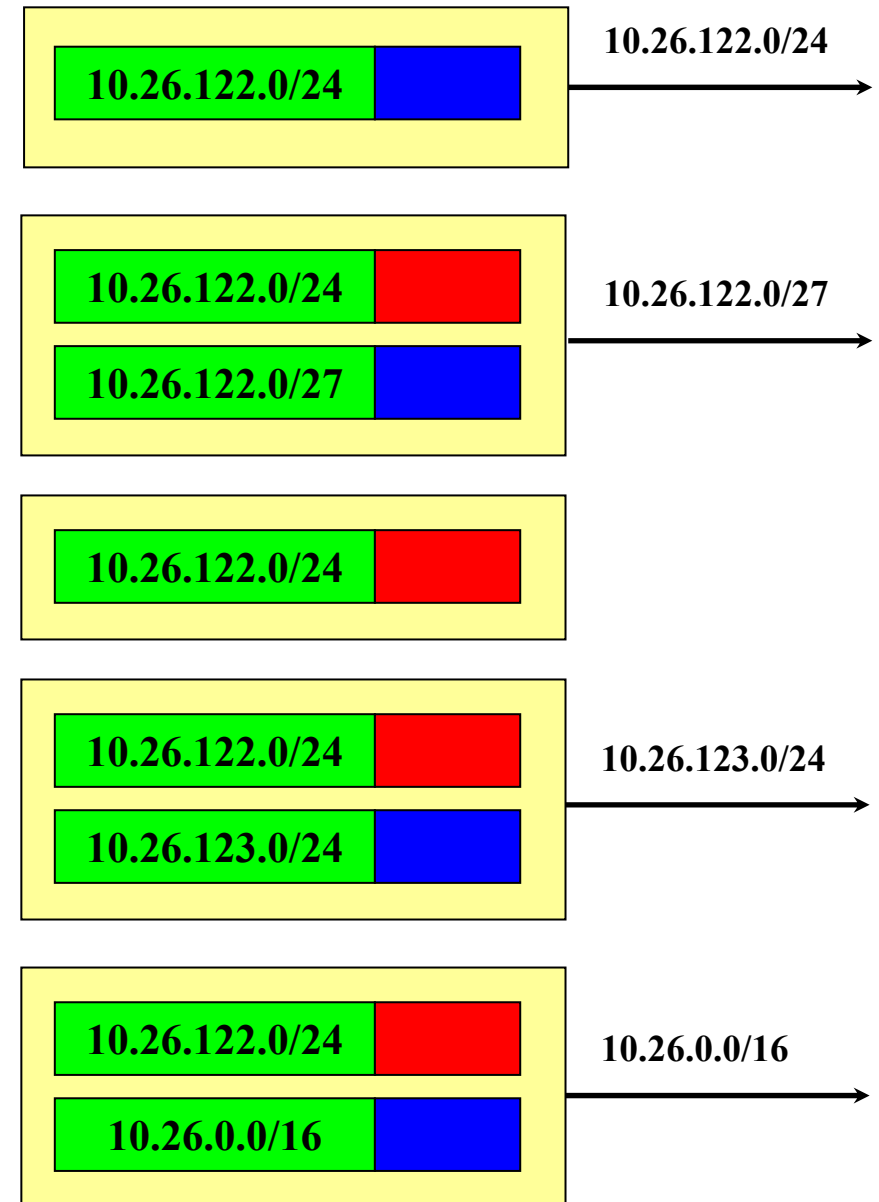
- Αν το **πρόθεμα** είναι **ίδιο** με κάποιας διαδρομή που ήδη υπάρχει στην Adj-RIB-In, τότε η νέα διαδρομή πρέπει να **αντικαταστήσει** την παλιά και η διαδικασία επιλογής πρέπει να εκτελεστεί
- Αν η νέα διαδρομή έχει **πρόθεμα** που **επικαλύπτεται** με μιας που υπάρχει στην Adj-RIB-In, εισάγεται και εκτελείται η διαδικασία επιλογής
- Αν η νέα διαδρομή έχει ίδιες **ιδιότητες** με μια προηγούμενη που υπάρχει στην Adj-RIB-In, αλλά είναι πιο **συγκεκριμένη**, τότε δεν γίνεται τίποτα
- Αν η νέα διαδρομή έχει **πρόθεμα** που δεν υπάρχει σε διαδρομή της Adj-RIB-In, τότε η νέα διαδρομή τοποθετείται στην Adj-RIB-In και εκτελείται η διαδικασία επιλογής
- Αν η νέα διαδρομή έχει πρόθεμα που **επικαλύπτεται** με μιας που υπάρχει στην Adj-RIB-In, αλλά είναι **λιγότερο συγκεκριμένη**, τότε εκτελείται η διαδικασία επιλογής για τους προορισμούς που περιγράφονται από την καινούργια διαδρομή μόνο

Λήψη νέας διαδρομής

Adj-RIB-In (πριν)

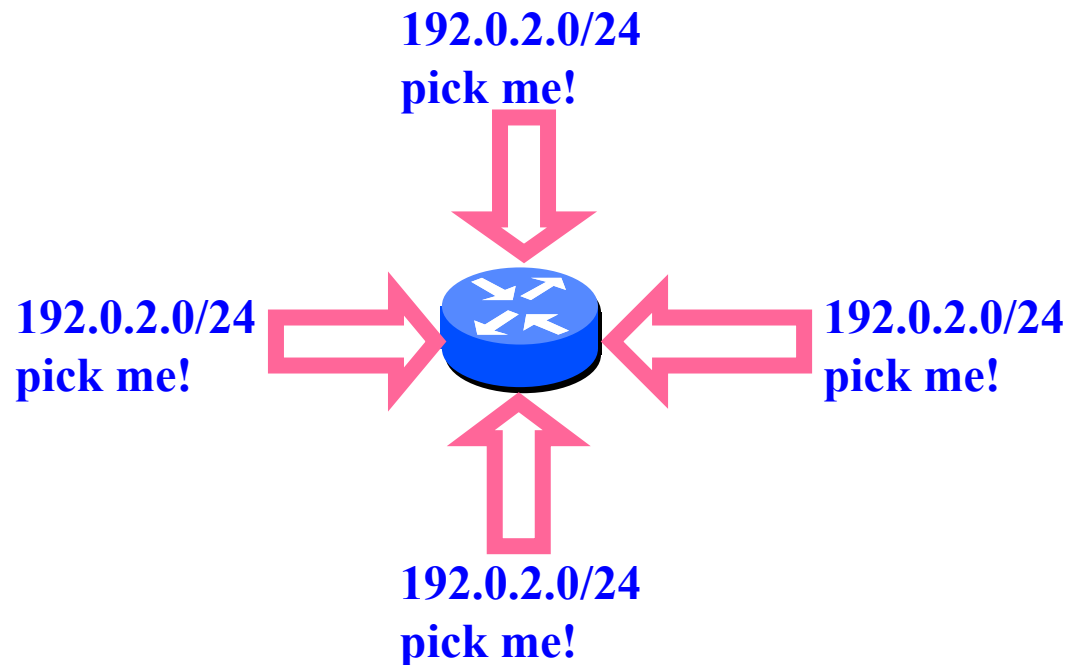


Adj-RIB-In (μετά)



Επιλογή καλύτερης διαδρομής

- Το πρωτόκολλο μπορεί να λάβει περισσότερες της μιας **ανακοινώσεις** για κάποιον προορισμό-πρόθεμα
- Το BGP επιλέγει **μόνο μία** διαδρομή ως την καλύτερη (μπορεί και να τις απορρίψει **όλες**)
- Αυτή θα εισαχθεί στον **πίνακα δρομολόγησης του πυρήνα** και θα **διαδωθεί** προς τους γειτονικούς δρομολογητές
- Τα κριτήρια επιλογής βασίζονται στις **ιδιότητες**



Έγκυρες διαδρομές

- Η σύγκριση για την επιλογή της διαδρομής, γίνεται μόνο ανάμεσα σε **έγκυρες** διαδρομές, οι άλλες αγνοούνται
- Στη συνέχεια εφαρμόζεται ένας **αλγόριθμος επιλογής** ανάμεσα στις έγκυρες, που περιλαμβάνει έναν αριθμό κριτηρίων
- Σε αυτόν ξεκινάμε με το πιο σημαντικό κριτήριο επιλογής. Αν με βάση αυτό έχουμε μία μόνο ή καμία διαδρομή τότε δεν συνεχίζουμε στο επόμενο, διαφορετικά συγκρίνουμε με βάση το επόμενο κριτήριο και ούτω κάθε'ξης

Μη έγκυρες διαδρομές

Συγχρονισμός

- Αν δεν είναι σε κατάσταση **συγχρονισμού** μη λάβετε υπόψη IBGP μονοπάτι (Cisco IOS)
- Είναι ενεργοποιημένο εξ' ορισμού σε Cisco αλλά απενεργοποιημένο από 12.2(8)T και μετά
- Πρέπει να υπάρχει **εγγραφή** για το πρόθεμα στον πίνακα δρομολόγησης, ώστε μια **IBGP** διαδρομή να θεωρηθεί έγκυρη
- Αν η διαδρομή έχει γίνει γνωστή από ένα OSPF γείτονα, πρέπει το RID του να ταιριάζει με το BGP RID του IBGP ομότιμου
- Συχνά απενεργοποιούμε την δυνατότητα αυτή, με την εντολή **no synchronization**

Μη έγκυρες διαδρομές

NextHop μη προσβάσιμο

- Μία διαδρομή στην οποία το **NextHop** δεν είναι **προσβάσιμο** (δεν υπάρχει διαδρομή στον πίνακα για αυτό) δεν λαμβάνεται υπόψη
- Πιθανώς να χρειαστεί **αναδρομική** αναζήτηση στον πίνακα δρομολόγησης προκειμένου να βρεθεί εγγραφή (συνήθως από IGP) που να ταιριάζει στο NextHop

Τοπικό ASN στην AS_path

- Διαδρομές από ένα EBGP ομότιμο, αν το **τοπικό ASN** εμφανίζεται στη λίστα AS_path (ανίχνευση βρόχου)

Μη έγκυρες διαδρομές

Φίλτρο εισερχομένων

- Διαδρομές οι οποίες φιλτράρονται σύμφωνα με την **τακτική** δρομολόγησης που έχουμε ορίσει δεν θα εμφανιστούν στον πίνακα του BGP
- Στην περίπτωση που ορίσουμε τη λειτουργία **soft-reconfiguration inbound** για ένα ομότιμο οι διαδρομές **φυλάσσονται** στην μνήμη ακόμα και αν απορρίπτονται από το φίλτρο εισόδου
- Αυτές τις διαδρομές έχουν σημαδευτεί σαν **received-only**

First_AS

- Αν ενεργοποιήσουμε την επιλογή **enabled bgp enforce-first-as** και το Update μηνυμα δεν περιέχει το **ASN** του γείτονα σαν τον **πρώτο** ASN στη λίστα, ο δρομολογητής στέλνει μήνυμα **Notification** και κλείνει την σύνοδο

Administrative Distance

- Τα περισσότερα πρωτόκολλα έχουν αλγορίθμους και μετρικές οι οποίες είναι **ασύμβατες** μεταξύ τους
- Όταν υπάρχουν περισσότερες από μία διαδρομές για τον ίδιο προορισμό που προέρχονται από **διαφορετικά πρωτόκολλα** δίνουμε **προτεραιότητα** (βαθμό αξιοπιστίας) στο κάθε πρωτόκολλο με την παράμετρο

Administrative Distance

- Είναι το **πρώτο κριτήριο** επιλογής διαδρομής
- Έχει **τοπική σημασία** στον κάθε δρομολογητή
- Η **μικρότερη τιμή** ορίζει μεγαλύτερη αξιοπιστία
- Για παράδειγμα αν έχουμε δύο διαδρομές για τον ίδιο προορισμό, μία που να προέρχεται από OSPF και μία από BGP θα προτιμηθεί αυτή από το BGP
- Αν χαθεί αυτή από το BGP τότε θα χρησιμοποιηθεί η άλλη

Administrative Distance

- Οι **στατικές** διαδρομές έχουν πάντα **1**
- Αν η τιμή για ένα πρωτόκολλο είναι **255**, ο δρομολογητής δεν το **εμπιστεύεται καθόλου**
- Όταν έχουμε **αναδιανομή** διαδρομών μερικές φορές χρειάζεται να την τροποποιήσουμε ώστε να προτιμηθεί το ένα πρωτόκολλο
- Τροποποιούμε την απόσταση με την εντολή **distance**
- Η τροποποίηση χρησιμοποιείται συχνά κατά το **μεταβατικό στάδιο** από ένα πρωτόκολλο σε άλλο, με το **τελευταίο** να έχει **μεγαλύτερη** τιμή
- Η τροποποίηση μπορεί να δημιουργήσει **βρόχους** και **μαύρες τρύπες** οπότε πρέπει να γίνεται με **προσοχή**
- Δεν υπάρχουν **γενικοί κανόνες**, κάθε δίκτυο έχει τις δικές του απαιτήσεις
- Μια εφαρμογή μπορεί να είναι όταν χρησιμοποιούμε **στατικές** διαδρομές σαν **εφεδρεία** μιας IGP διαδρομής, όπου ορίζουμε την εφεδρεία με ένα μεγάλο διαχειριστικό κόστος
- Παράδειγμα `ip route 10.0.0.0 255.0.0.0 Dialer 1 250`

Administrative Distance

Route Source	Default Distance Values
Connected interface	0
Static route*	1
Enhanced Interior Gateway Routing Protocol (EIGRP) summary route	5
External Border Gateway Protocol (BGP)	20
Internal EIGRP	90
IGRP	100
OSPF	110
Intermediate System-to-Intermediate System (IS-IS)	115
Routing Information Protocol (RIP)	120
Exterior Gateway Protocol (EGP)	140
On Demand Routing (ODR)	160
External EIGRP	170
Internal BGP	200
Unknown**	255

Κριτήρια επιλογής διαδρομής

1. Weight (Βάρος) - BGP

- Διαδρομή με το **μεγαλύτερο βάρος** (τοπική σημασία στον **δρομολογητή**, ιδιοταγής ιδιότητα από την Cisco)

2. Local_Preference - BGP

- Μεγαλύτερη **τοπική προτίμηση** (ενιαία στο ίδιο **AS**)
- Διαδρομή με κενή αυτή την ιδιότητα θεωρείται ότι έχει την εξ ‘ ορισμού τιμή (100 ή αυτή που έχει οριστεί με την εντολή **bgp default local-preference**)
- Η τιμή της υπολογίζεται τοπικά για διαδρομές που έγιναν γνωστές μέσω EBGP ή από IGP ή στατικές διαδρομές
- Η τιμή της γίνεται γνωστή από άλλο δρομολογητή για διαδρομές που γίνανε γνωστές από έναν **IBGP** ομότιμο (μέσω Update μηνύματος που περιέχει την ιδιότητα αυτή και υποδυκνείει το βαθμό προτίμησης)

Κριτήρια επιλογής διαδρομής

3. Τοπική διαδρομή

- Διαδρομή η οποία δημιουργήθηκε **τοπικά** από BGP το οποίο τρέχει σε αυτό τον δρομολογητή (**locally** originated route)
- Τέτοιες διαδρομές έχουν δημιουργηθεί με τις εντολές **network** ή **aggregate** ή με **αναδιανομή** από ένα IGP
- Οι τοπικές διαδρομές που παρήχθησαν από τις εντολές **network** ή **redistribute** προτιμώνται από αυτές που δημιουργήθηκαν με την **aggregate-address**

Κριτήρια επιλογής διαδρομής

4. AS_path - BGP

- Διαδρομή με το μικρότερο μέγεθος λίστα **AS_path**
- Αν έχουμε ορίσει την **bgp bestpath as-path ignore**, αυτό το βήμα δεν πραγματοποιείται
- Η AS_SET υπολογίζεται σαν 1
- Οι AS_CONFED_SEQUENCE και AS_CONFED_SET δεν περιλαμβάνονται στο μήκος της AS_PATH λίστας

5. Origin - BGP

- Αν οι διαδρομές έχουν το ίδιο μέγεθος του AS_path προτίμησε αυτή μικρότερο **origin type** (IGP < EGP < incomplete)

Κριτήρια επιλογής διαδρομής

6. MED – BGP

- Προτίμησε αυτή με τη **μικρότερη** τιμή MED
- Η σύγκριση γίνεται **μόνο** αν το πρώτο (γειτονικό) ASN στη λίστα είναι το ίδιο
- Αν υπάρχουν ύπο-AS (confederation), αγνοούνται (AS_CONFED_SEQUENCE)

bgp always-compare-med (bgp set compare-med always)

- Οι MED συγκρίνονται για **όλες** τις διαδρομές
- Αυτή η επιλογή πρέπει να ενεργοποιείται παντού και όχι μεμονωμένα, αλλιώς βρόχοι μπορεί να δημιουργηθούν
- Διαδορετικά ο εξορισμού τρόπος είναι να συγκρίνονται μόνο αν **προέρχονται από το ίδιο AS**

bgp bestpath med-confed

- Οι MED συγκρίνονται μόνο για τις διαδρομές που περιέχουν μόνο μία λίστα AS_CONFED_SEQUENCE, δηλαδή διαδρομές που γεννήθηκαν από το τοπικό confederation
- Στις διαδρομές που λαμβάνονται με MED=4,294,967,295 θα τροποποιηθεί ο MED σε 4,294,967,294 πριν εισαχθεί στον πίνακα Στις διαδρομές που λαμβάνονται με κενό MED, ανατίθεται MED=0, εκτός αν ενεργοποιήσουμε την **bgp bestpath missing-as-worst** οπότε ανατίθεται MED=4,294,967,294

Κριτήρια επιλογής διαδρομής

7. (IGP)>EBGP>IBGP

- Προτίμησε μια **EBGP** από μία **IBGP** διαδρομή (αν έχει επιλεχθεί bestpath τότε πήγαινε στο βήμα 9 multipath)
- Οι διαδρομές που περιέχουν AS_CONFED_SEQUENCE και AS_CONFED_SET είναι τοπικά στο confederation και άρα αντιμετωπίζονται σαν τοπικές διαδρομές. Δεν υπάρχει διάκριση σε Internal/External Confederation
- Αν υπάρχει και IGP διαδρομή προτίμησε την και από τις EBGP, IBGP
- Το IGP θεωρείται πιο εμπιστέψιμο από το EGP διότι δεν περιέχει εξωτερικές διαδρομές στο δίκτυο

8. IGP cost/metric – BGP

- Αν όλες οι διαδρομές έχουν γίνει γνωστές μέσω IBGP
- Προτίμησε τη διαδρομή με τη μικρότερη IGP μετρική προς το BGP NextHop

Κριτήρια επιλογής διαδρομής

9. Multipath

- Εκτελείται ακόμα και αν έχει βρεθεί ήδη η καλύτερη διαδρομή
- Έλεγχος αν περισσότερες διαδρομές πρέπει να εγκατασταθούν για την δυνατότητα Multipath
- Είναι εξ' ορισμού ενεργοποιημένο
- Υποστηρίζει μέχρι 4 ίσου κόστους διαδρομές για κάθε πρωτόκολλο

Κριτήρια επιλογής διαδρομής

10. Received Time

- Άν όλες οι διαδρομές είναι εξωτερικές, προτίμησε αυτή που ελήφθηκε πρώτα
- Αυτό ώστε να εκλαχιστοποιηθούν οι ταλαντώσεις των διαδρομών
- Αυτό το βήμα δεν εκτελείτε αν
 - **bgp best path compare-routerid**
 - Το RID είναι το ίδιο, και άρα οι διαδρομές ελήφθησαν από τον ίδιο δρομολογητή
 - Δεν υπάρχει τώρα καλύτερη διαδρομή (όταν π.χ. Ο ομότιμος χάνεται)

Κριτήρια επιλογής διαδρομής

11. RID - BGP

- Αν όλες έχουν γίνει γνωστές με EBGP
- Προτίμησε τη διαδρομή που προέρχεται από το δρομολογητή με την πιο μικρή RID (είναι η μεγαλύτερη IP του ή η διεύθυνση βρόχου αν υπάρχει, ή μπορεί να οριστεί χειροκίνητα - **bgp router-id**)
- Αν η διαδρομή περιέχει ιδιότητες σχετικές με RR, η ιδιότητα Originator ID χρησιμοποιείται όταν εκτελείται αυτό το βήμα

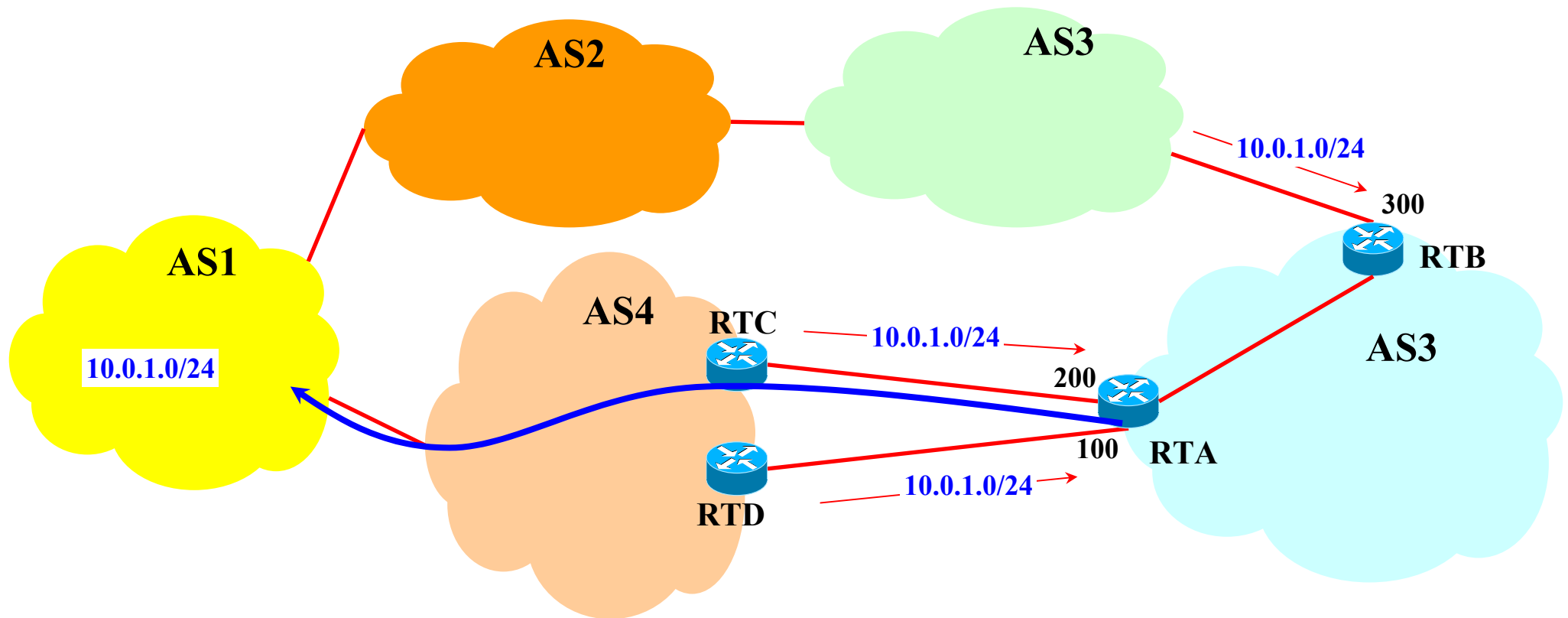
12. Cluster-List

- Αν το RID είναι το ίδιο προτίμησε αυτό με το μικρότερο μήκος της **Cluster-List** (έχει νόημα μόνο όταν έχει οριστεί RR, ο πελάτης πρέπει να είναι ενήμερος για τις ιδιότητες του Route Reflector)

13. Neighbor IP

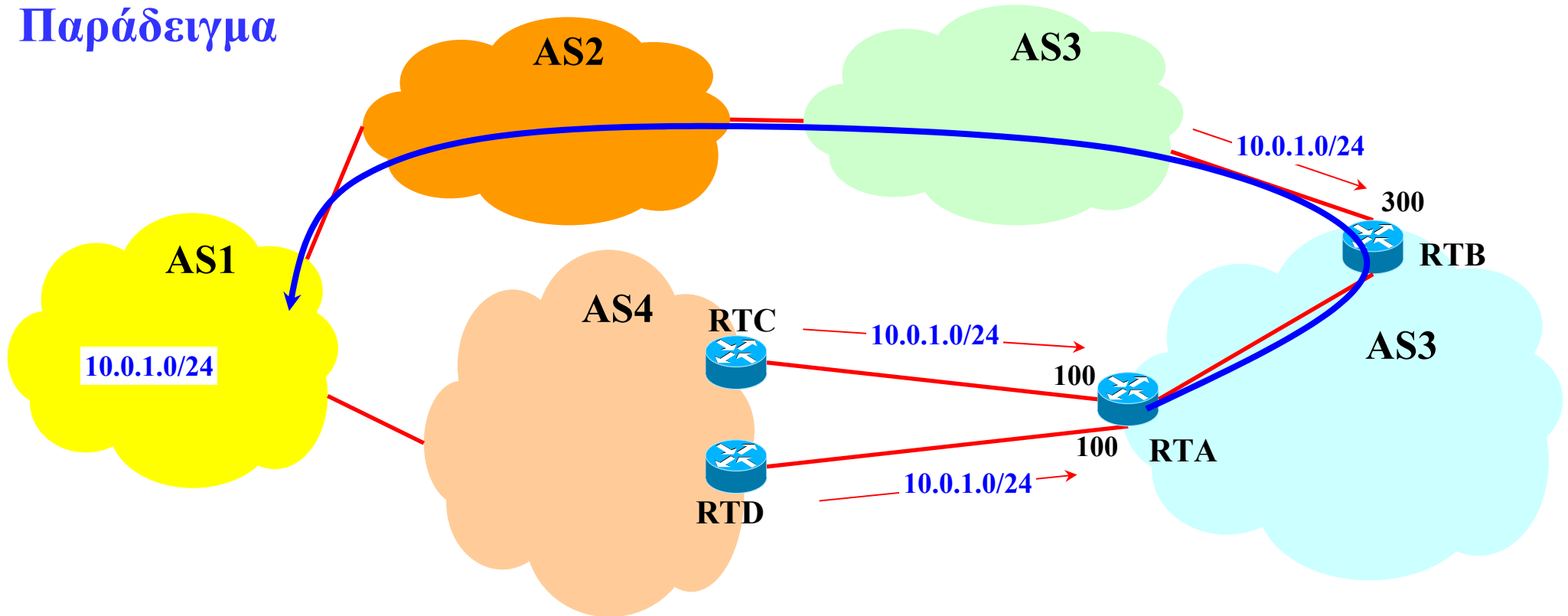
- Προτίμησε τη διαδρομή που προέρχεται από την μικρότερη IP διεύθυνση γείτονα (αυτή που ορίζεται με την εντολή neighbor)

Παράδειγμα



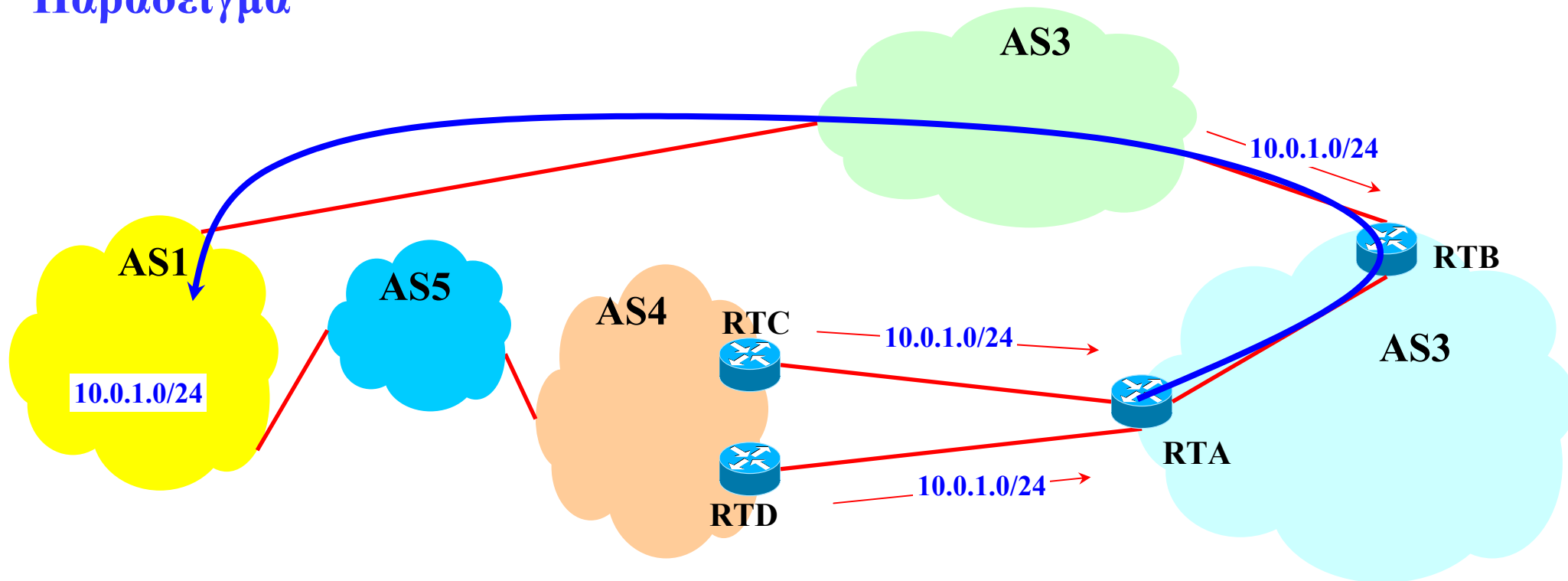
- Ορίζουμε στον RTA **βάρη 100 και 200** για τους προορισμούς 10.0.1.0/24
- Επίσης στον RTB ορίζουμε βάρος 300 για τους ίδιους προορισμούς
- Ο RTA θα πάρει τρεις διαδρομές για το 10.0.1.0/24 δύο από EBGP ομότιμους και μία από IBGP ομότιμο
- Θα επιλέξει τη διαδρομή μέσω του RTC

Παράδειγμα

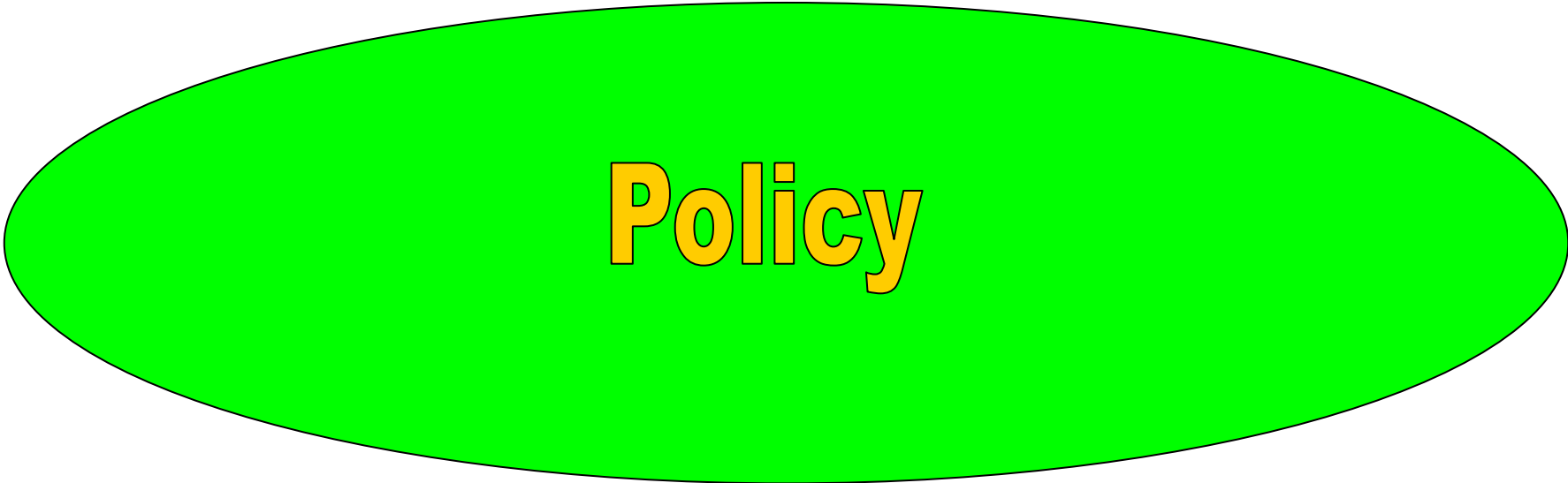


- Έστω ότι τα βάρη είναι ίδια στον RTA και για τις δυο εξόδους
- Ορίζουμε στον RTA **τοπική προτίμηση 100** και στις δύο εξόδους για το πρόθεμα 10.0.1.0/24
- Επίσης στον RTB ορίζουμε τοπική προτίμηση **300** για το ίδιο πρόθεμα
- Ο RTA θα πάρει τρεις διαδρομές για το 10.0.1.0/24 δύο από EBGP ομότιμους και μία από IBGP ομότιμο
- Θα επιλέξει τη διαδρομή μέσω του RTB

Παράδειγμα



- Έστω ότι τα βάρη είναι ίδια στον RTA και για τις δυο εξόδους
- Επίσης έστω ότι η τοπική προτίμηση είναι η ίδια και για τα τρία σημεία εξόδου από το AS3
- Ο RTA θα λάβει τρεις διαδρομές για το 10.0.1.0/24, δύο από το AS4 με $AS_path=\{1,5,4\}$ και μία μέσω του RTB με $AS_path=\{1,3\}$, το IBGP μεταφέρει διαφανώς το AS_path
- Θα επιλέξει τη διαδρομή μέσω του RTB αφού έχει μικρότερη λίστα AS_path



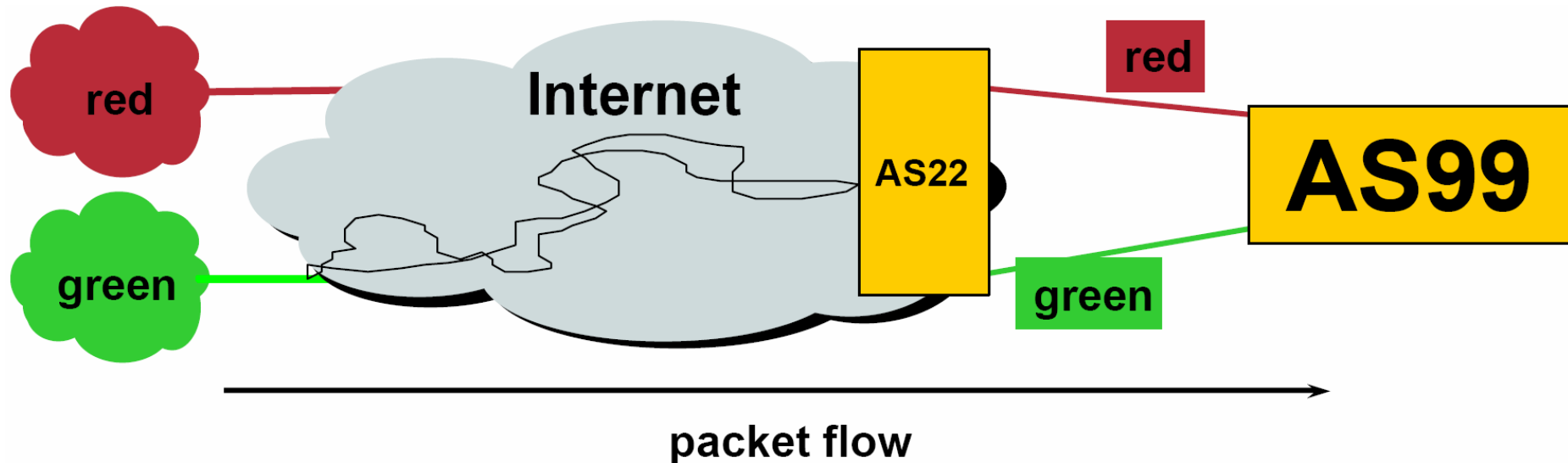
Policy

Γιατί Πολιτική δρομολόγησης;

- Χρησιμοποιείται για σιγουρέψει ότι ένας πάροχος μόνο **χρησιμοποιεί** και **διαφημίζει** πληροφορία δρομολόγησης που επιτρέπεται σύμφωνα με τις συμφωνίες που έχει με τα δίκτυα με τα οποία είναι συνεδεδεμένος
- **Πλήρης έλεγχος** σχετικά με τη **διαδρομή** που παίρνει η κίνηση προς τον προορισμό (είτε παράγεται είτε διασχίζει το δίκτυο)
- Μπορεί να βεβαιωθούμε ότι κίνηση η οποία δε προβλέπεται **δε διασχίζει** το δίκτυο μας
- Μπορεί να βεβαιώσει ότι η δική μας κίνηση ταξιδεύει προς τον προορισμό μόνο μέσω των δικτύων με τα οποία έχουμε **συμφωνία**

Παράδειγμα

- Η AS99 μπορεί να χρησιμοποιεί τη κόκκινη σύνδεση για την κίνηση προς το κόκκινο AS και την πράσινη για την υπόλοιπη κίνηση
- Για την επίτευξη αυτής της πολιτικής πρέπει το AS99 να δέχεται διαδρομές που παράγονται από το κόκκινο AS, από την κόκκινη σύνδεση και να δέχεται όλες τις άλλες διαδρομές από την πράσινη σύνδεση
- Για το σκοπό αυτό πρέπει όμως να συνεργαστεί το AS22 ώστε να προωθήσει την κίνηση από το πράσινο AS στην πράσινη σύνδεση



Πως εφαρμόζεται η Πολιτική δρομολόγησης;

- **Φιλτράρισμα** ανάλογα με **AS_path**, **Community**, **πρόθεμα**
- **Αποδοχή/απόρριψη** διαδρομών
- Ορισμός **ιδιοτήτων** ώστε να επηρεάσουμε τη επιλογή διαδρομής

Φιλτράρισμα διαδρομών

- Το φιλτράρισμα των διαδρομών είναι η **βάση** για την εφαρμογή μιας τακτικής δρομολόγησης
- Γίνεται με βάση τις ενημερώσεις από ή προς ένα συγκεκριμένο **ομότιμο**
- Σε **εισερχόμενες – εξερχόμενες** ανακοινώσεις
- Γίνεται με βάση το

Πρόθεμα

AS_path

Community

- Κάθε μέθοδος φιλτραρίσματος μπορεί να χρησιμοποιηθεί για το **ίδιο αποτέλεσμα**, η επιλογή εξαρτάται από την **περίπτωση**

- Κάθε μέθοδος φιλτραρίσματος έχει διαφορετικό **βαθμό τραχύτητας**

Prefix_filter < AS_path < community filtering

- **Routemap filtering** είναι ένας τρόπος να εφαρμόσουμε **AS_path** και **community** φιλτράρισμα . Επίσης παρέχει έναν τρόπο για **ταίριασμα, τροποποίηση** της λαμβανόμενης ή διαφημιζόμενης πληροφορίας πριν να χρησιμοποιηθεί ή πριν να προωθηθεί

Φίλτρο με βάση το πρόθεμα

•Φίλτρο distribute-list με Standard Access List

neighbor { ip-address| peer-group-name} **distribute-list** access-list-number {in | out}

access-list <access-list-number> {deny|permit} protocol source mask

•Φίλτρο distribute-list με Extended Access List

neighbor { ip-address| peer-group-name} **distribute-list** access-list-number {in | out}

access-list <access-list-number> {deny|permit} protocol source source-wildcard mask mask-wildcard

•Φίλτρο ip prefix-list

neighbor 172.16.1.2 **prefix-list** cisco in

ip prefix-list cisco seq 10 **permit** 0.0.0.0/0 **le** 19

Φίλτρο με βάση το AS_path

neighbor { ip-address| peer-group-name} **filter-list** access-list-number {in|out}

ip as-path access-list access-list-number {permit|deny} as-regular-expression1

Φίλτρο με βάση το Community

ip community-list community-list-number {permit|deny} community-number

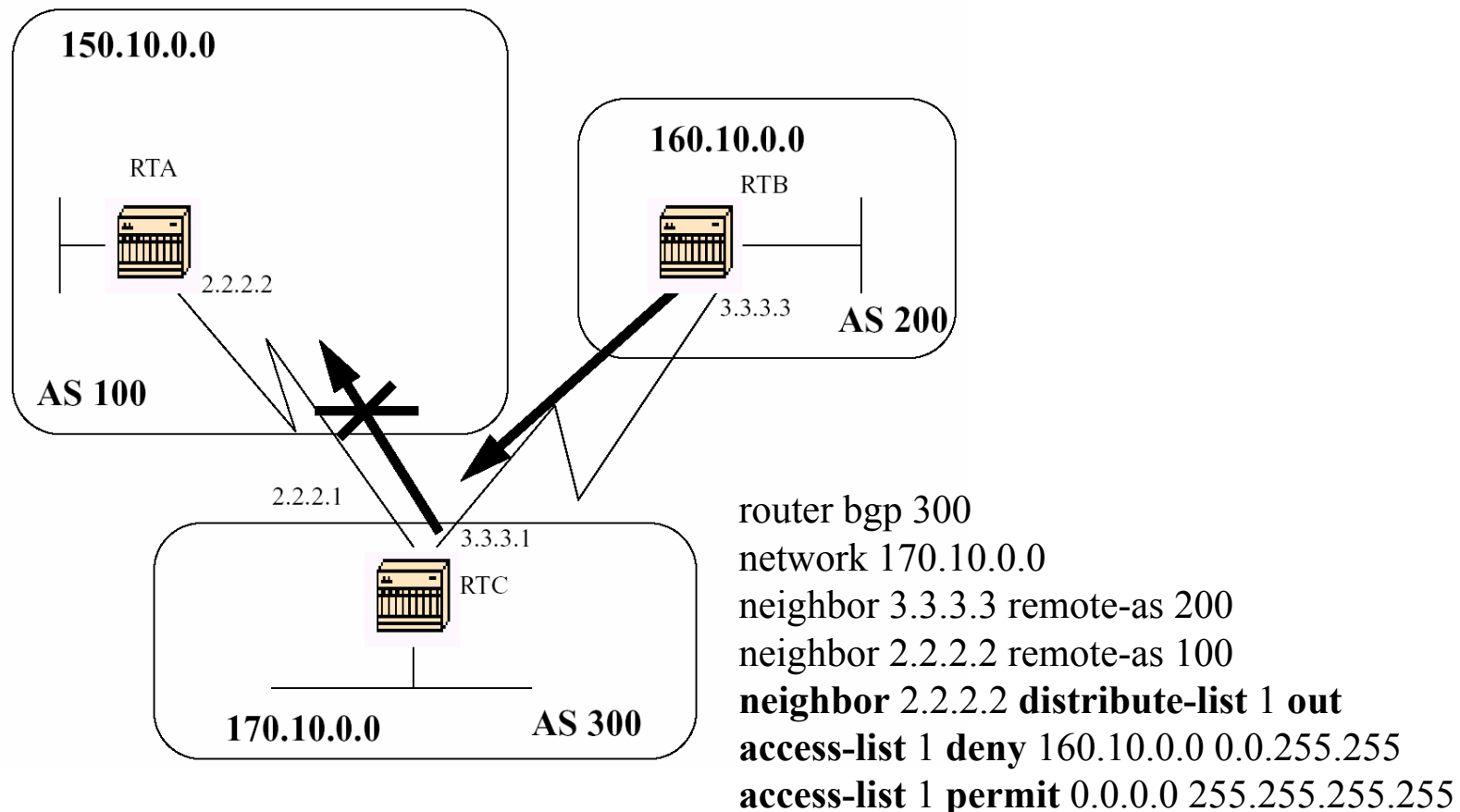
Φίλτρο **distribute-list** με **Standard Access List**

neighbor { ip-address| peer-group-name} **distribute-list** access-list-number {**in** | **out**}

access-list <access-list-number> {**deny**|**permit**} protocol source mask

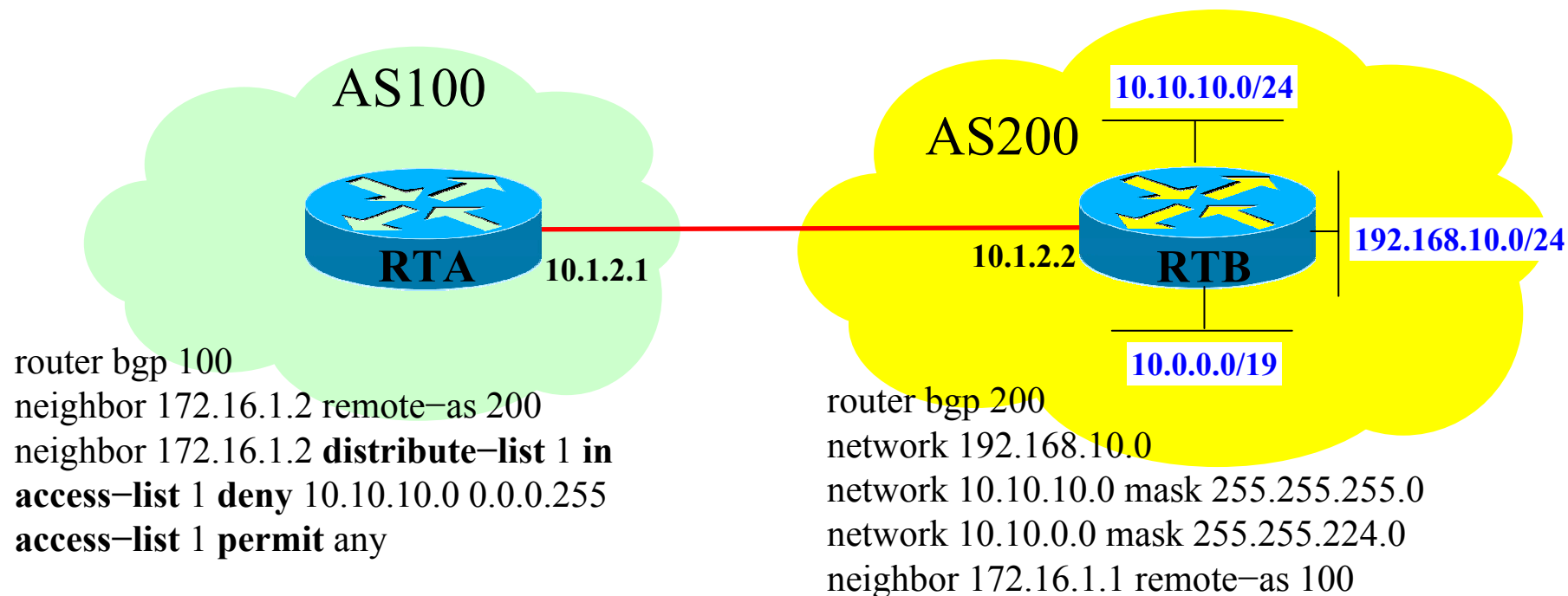
Παράδειγμα 1

- Ο RTB παράγει το πρόθεμα 160.10.0.0
- Θέλουμε ο RTC να φιλτράρει αυτές τις ανακοινώσεις και να μη διαδωθούν στο AS100 (αυτός θα τις χρησιμοποιεί)



Παράδειγμα 2

- Θέλουμε να φιλτραριστούν τα προθέματα 10.10.10.0/24



show ip bgp

BGP table version is 3, local router ID is 172.16.1.1

Status codes: s suppressed, d damped, h history, * valid, > best, i – internal

Origin codes: i – IGP, e – EGP, ? – incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path
*>10.10.0.0/19	172.16.1.2	0		0	200 i
*>192.168.10.0/24	172.16.1.2	0		0	200 i

Φίλτρο distribute-list με Extended Access List

- Μια απλή λίστα **access-list 1 permit 10.26.122.0 0.0.0.255** αφήνει τα προθέματα 10.26.122.0/24, 10.26.122/25, 10.26.122/26 κ.τ.λ αφού έχει την μοναδική απαίτηση τα 24 πρώτα bits να ταιριάζουν
- Αν θέλουμε να φιλτράρουμε υπερδίκτυα πρέπει να χρησιμοποιήσουμε την extended access list
- Η εντολή είναι:
access-list <access-list-number> {deny|permit} protocol source source-wildcard mask mask-wildcard

Παράδειγμα

access-list 1 permit 10.26.122.0 0.0.0.255 255.255.255.0 0.0.0.0

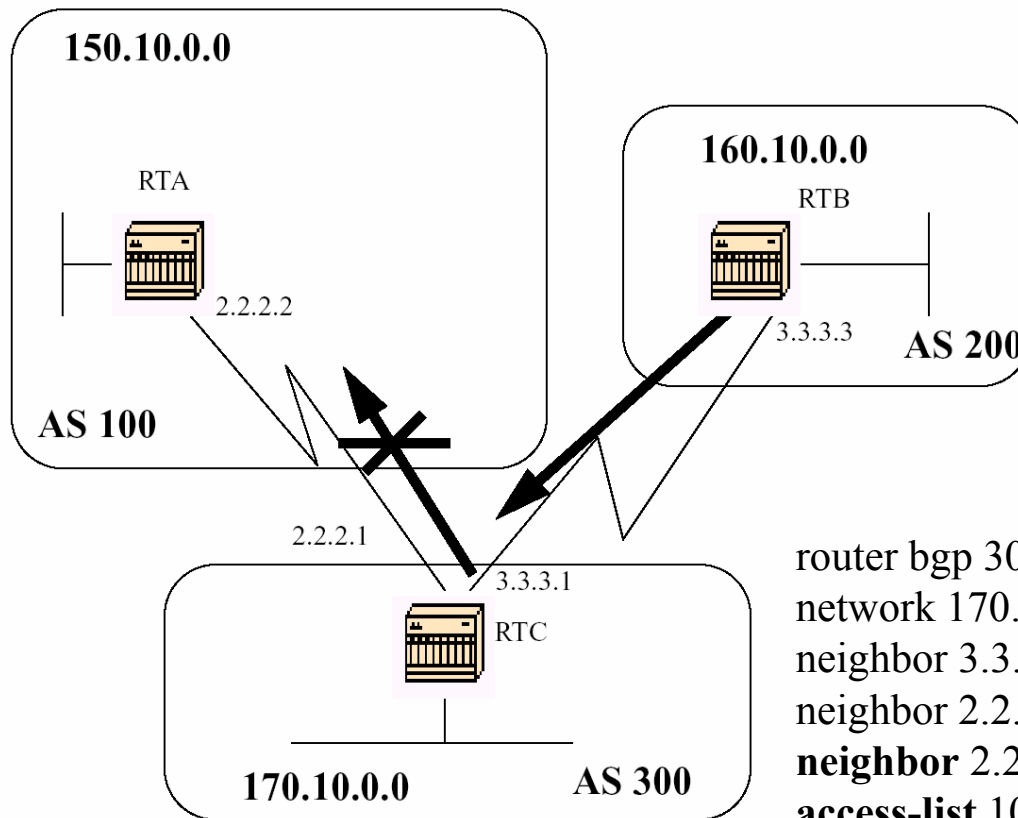
- Θα αφήσει μόνο το 10.26.122.0/24

access-list 1 permit 10.26.122.0 0.0.0.255 255.255.255.224 0.0.0.0

- Θα αφήσει μόνο τα 10.26.122.0/27

Παράδειγμα 1

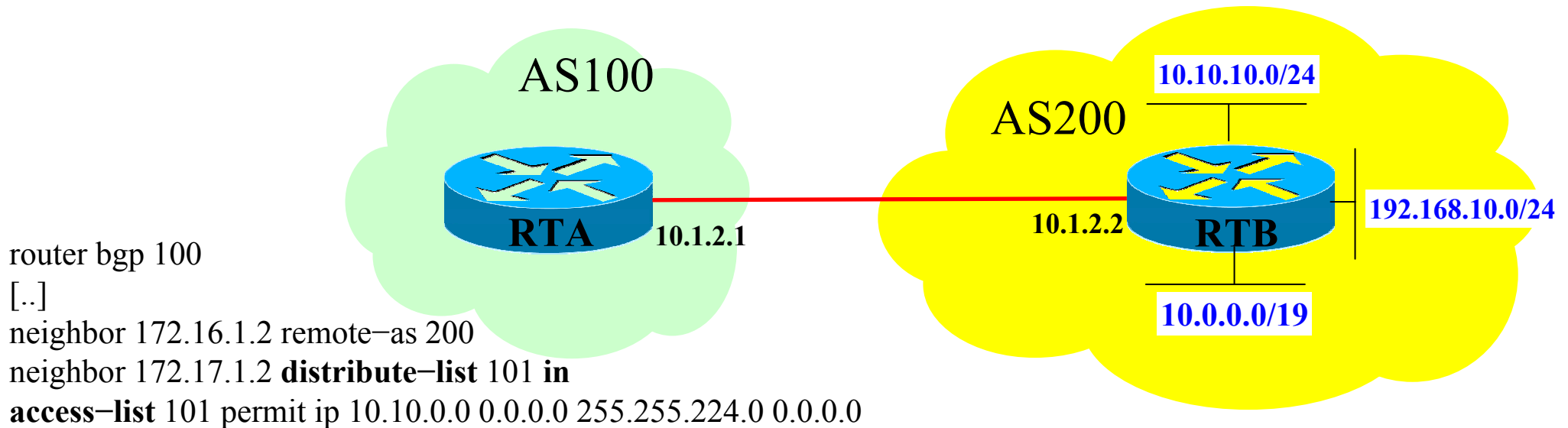
- Ο RTB έχει διαφορετικά υποδίκτυα της μορφής 160.10.X.X
- Θέλουμε να ανακοινώσουμε μόνο 160.0.0.0/8



```
router bgp 300
network 170.10.0.0
neighbor 3.3.3.3 remote-as 200
neighbor 2.2.2.2 remote-as 100
neighbor 2.2.2.2 distribute-list 1 out
access-list 101 permit ip 160.0.0.0 0.255.255.255 255.0.0.0 0.0.0.0
```

Παράδειγμα 2

- Το AS200 ανακοινώνει διάφορα προθέματα 10.10.1.0/24 ως και 10.10.31.0/24 και το συνάθροισμα τους 10.10.0.0/19
- Θέλουμε ο RTA να δεχεται μόνο το 10.10.0.0/19



Router100# **show ip bgp**

BGP table version is 2, local router ID is 172.16.1.1

Status codes: s suppressed, d damped, h history, * valid, > best, i – internal

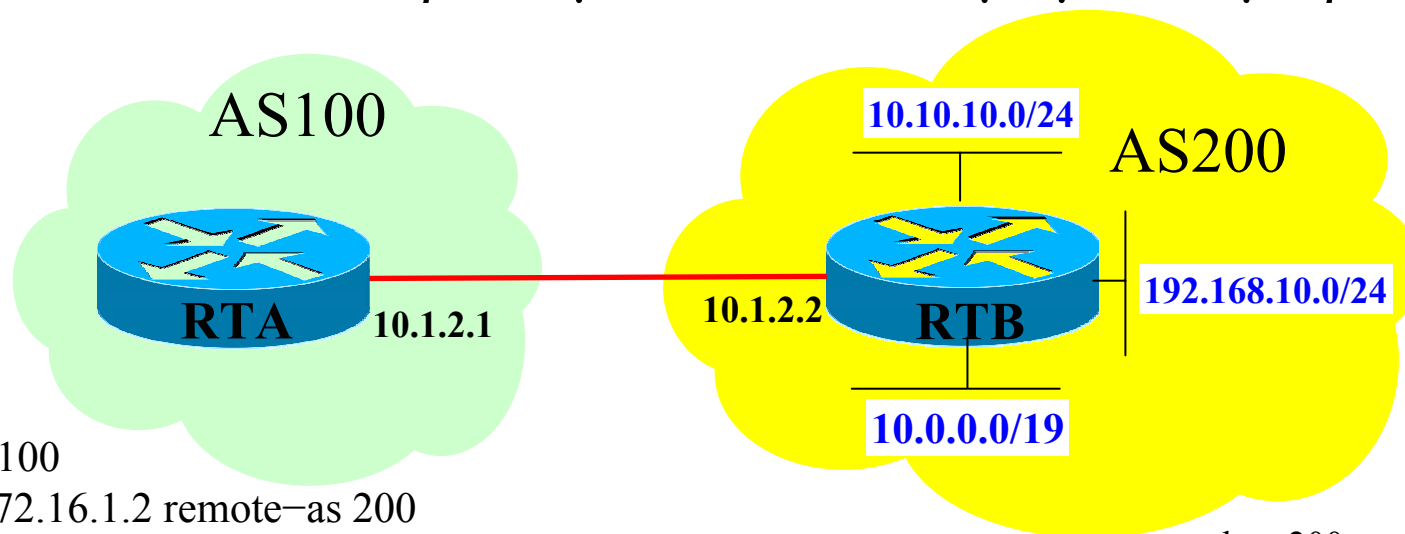
Origin codes: i – IGP, e – EGP, ? – incomplete

Network Next Hop Metric LocPrf Weight Path

*> 10.10.0.0/19 172.16.1.2 0 0 200 i

Φίλτρο ip prefix-list

- Θέλουμε ο RTA να επιτρέπει μόνο τα δίκτυα με μάσκα μικρότερη ή ίση με /19



```
router bgp 100
neighbor 172.16.1.2 remote-as 200
neighbor 172.16.1.2 prefix-list cisco in
ip prefix-list cisco seq 10 permit 0.0.0.0/0 le 19
```

```
router bgp 200
network 192.168.10.0
network 10.10.10.0 mask 255.255.255.0
network 10.10.0.0 mask 255.255.224.0
neighbor 172.16.1.1 remote-as 100
```

show ip bgp

BGP table version is 2, local router ID is 172.16.1.1

Status codes: s suppressed, d damped, h history, * valid, > best, i – internal

Origin codes: i – IGP, e – EGP, ? – incomplete

Network Next Hop Metric LocPrf Weight Path

```
*> 10.10.0.0/19 172.16.1.2 0 0 200 i
```

Φίλτρο ip prefix-list

- Φιλτράρει τις διαδρομές με βάση το **πρόθεμα**
- Εισερχόμενο-εξερχόμενο φίλτρο
- Θέλουμε να φιλτραριστούν όλα τα εισερχόμενα προθέματα 218.10.0.0/16 και να επιτραπούν σαν εξερχόμενα μόνο τα 215.7.0.0/16

```
router bgp 200
neighbor 220.200.1.1 remote-as 210
neighbor 220.200.1.1 prefix-list PEER-IN in
neighbor 220.200.1.1 prefix-list PEER-OUT out
!
ip prefix-list PEER-IN deny 218.10.0.0/16
ip prefix-list PEER-IN permit 0.0.0.0/0 le 32
ip prefix-list PEER-OUT permit 215.7.0.0/16
```

Φίλτρο AS_path

- Σημαίνει το φιλτράρισμα των διαδρομών με βάση την ιδιότητα **AS_path**
- Χρησιμοποιούνται **κανονικές εκφράσεις**
- Εισερχόμενο-εξερχόμενο φίλτρο
- Θα προσπαθήσει να ταιριάξει την AS_path ιδιότητα μιας ενημέρωσης με μια λίστα εγγραφών

ip as-path access-list access-list-number {**permit|deny**} as-regular-expression
neighbor { ip-address| peer-group-name} **filter-list** access-list-number {**in|out**}

show ip bgp regexp < regular expression>

- Μπορούμε να δούμε όλα τα μονοπάτια που ταιριάζουν στη έκφραση

NautilusBgpd> show ip bgp regexp ^72

BGP table version is 0, local router ID is 10.26.122.225

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal

Origin codes: i - IGP, e - EGP, ? - incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.2.12.0/24	10.2.13.150			0	72 240 ?
*> 10.2.13.0/24	10.2.13.150	0		0	72 i
*> 10.80.183.0/24	10.2.13.150			0	72 1351 888 405 i
*> 10.80.190.0/24	10.2.13.150			0	72 1351 888 ?

Εκφράσεις (Regular Expressions – regexp – regex)

- Μία παράσταση που περιγράφει **πρότυπα** συμβολοσειρών
- Χρησιμοποιούνται για τον προσδιορισμό ενός **συνόλου κανόνων** που θα πρέπει να πληρεί ένα σύνολο συμβολοσειρών ώστε να ταιριάζει σε μία αναζήτηση
- Στο BGP χρησιμοποιούνται για την **αναζήτηση** λιστών **AS_path** ώστε να διαπιστωθεί αν ταιριάζουν σε συγκεκριμένο πρότυπο
- Μπορούν να χρησιμοποιηθούν σε **CLI** εντολές ή σε **τμήματα εντολών** ώστε να χτιστούν πολύπλοκες **τακτικές** δρομολόγησης

Εκφράσεις

^	Ταίριαγμα της αρχής της AS_path λίστας
\$	Ταίριαγμα του τέλους της AS_path λίστας
<κενό>	Χωρίζει AS αριθμούς στην λίστα
<AS number>	Ταίριαγμα του AS αριθμού
*	Ταίριαγμα καμίας ή περισσότερων επαναλήψεων του προηγούμενου υποδείγματος
+	Ταίριαγμα μίας ή περισσότερων επαναλήψεων του προηγούμενου υποδείγματος
.	Ταίριαγμα οποιοδήποτε μονού χαρακτήρα ακόμα και του κενού
?	Ταίριαγμα καμίας ή μίας επαναλήψεων του προηγούμενου υποδείγματος
_	Ταίριαγμα των , { } , αρχή και τέλος ενός string, ή ένα κενό
[]	Ορίζει μία κλίμακα τιμών για ένα μονό χαρακτήρα
-	Χωρίζει τα άκρα μιας κλίμακας τιμών
\character	Ταιριάζει τον χαρακτήρα

Παραδείγματα

.*	όλες οι AS_path λίστες
.+	όλες οι λίστες εκτός της κενής
^\$	κενή λίστα, άρα ταιριάζει σε διαδρομές που γενήθηκαν εντός του AS
_100\$	λίστα που τελειώνει σε 100
100	λίστα που περιλαμβάνει το AS=100
^100	λίστες που αρχίζουν με το AS=100
^100_	έχει ληφθεί από το 100
^100\$	λίστα με μόνο το 100
^100.*	λίστα που αρχίζει με το 100
100	διαμέσω του 100
_200_100_	διαμέσω του 100 και μετά του 200
(100)+	τουλάχιστον ένα 100
\\(100\\)	διαμέσω του 100 (confederation AS)
^100 200\$	100 200

Παραδείγματα

<code>^[0-9]+\$</code>	λίστα με έναν αριθμό
<code>^[0-9]+_[0-9]+\$</code>	λίστα με δύο αριθμούς
<code>^[0-9]*_[0-9]+\$</code>	λίστα με έναν ή δύο αριθμούς
<code>^[0-9]*_[0-9]*\$</code>	λίστα με έναν ή δύο αριθμούς, αλλά και το μηδέν
<code>^[0-9]+_[0-9]+_[0-9]+\$</code>	λίστα με τρεις αριθμούς
<code>_(100 200)_</code>	οτιδήποτε έχει διέλθει από τα AS100 ή AS200
<code>_100(_.+_)200\$</code>	οτιδήποτε γενήθηκε στο 200 και περνά μέσω του 100

Παραδείγματα

bgp show regexp <regexp>

bgp show regexp “.”

•Όλες οι διαδρομές ένα μονό αριθμό στην AS_path λίστα

bgp show regexp “65050 .”

•Όλες οι διαδρομές που μας ανακοινώθηκαν από το AS65050

bgp show regexp “.* 65100”

•Όλες οι διαδρομές που γεννήθηκαν από το AS65100

bgp show regexp “.*”

•Όλες οι διαδρομές με οποιοδήποτε AS_path

bgp show regexp “. .+”

•Όλες οι διαδρομές με τουλάχιστον δύο AS αριθμούς

bgp show regexp “.?”

•Όλες οι διαδρομές με όχι λιγότερο από ένα και όχι περισσότερα από δύο AS αριθμούς

bgp show regexp “”

•Όλες οι διαδρομές που έχουν γεννηθεί τοπικά

bgp show regexp “.* 65499 .”

•Όλες οι διαδρομές που έχουν διασχίσει το AS65499

bgp show regexp “65499|65000”

•Όλες οι διαδρομές με μόνο το 65499 ή το 65000 στη λίστα

•bgp show regexp “.* [^65000 65001]”

•Όλες οι διαδρομές με έναν τουλάχιστον AS αριθμό και δεν τελειώνουν σε 65000 65001

bgp show regexp “.* [^64512-65535]* .”

•Όλες οι διαδρομές που περιέχουν μόνο δημόσιους AS αριθμούς (εξαρούνται οι διαδρομές με ιδιωτικούς αριθμούς)

bgp show regexp “65000 65100 [^65499] .”

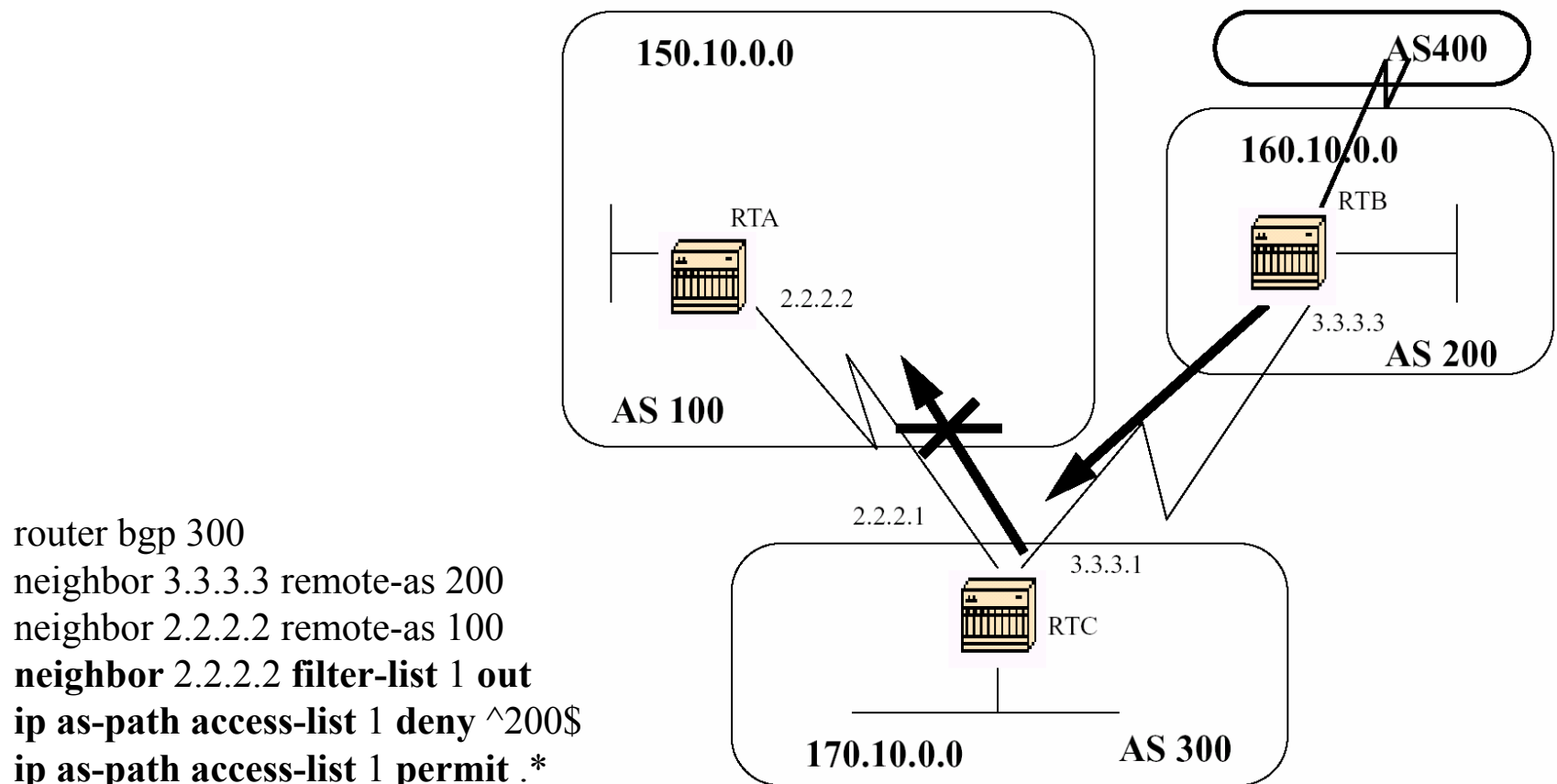
•Όλες οι διαδρομές που περιέχουν το 65000 65100 και δεν περιέχουν το AS 65499 σαν τον επόμενο αριθμό

bgp show regexp “.* 65499 [^65020] .”

•Όλες οι διαδρομές με το AS 65499 με εξαίρεση αυτά που περιέχουν το 65499 65020

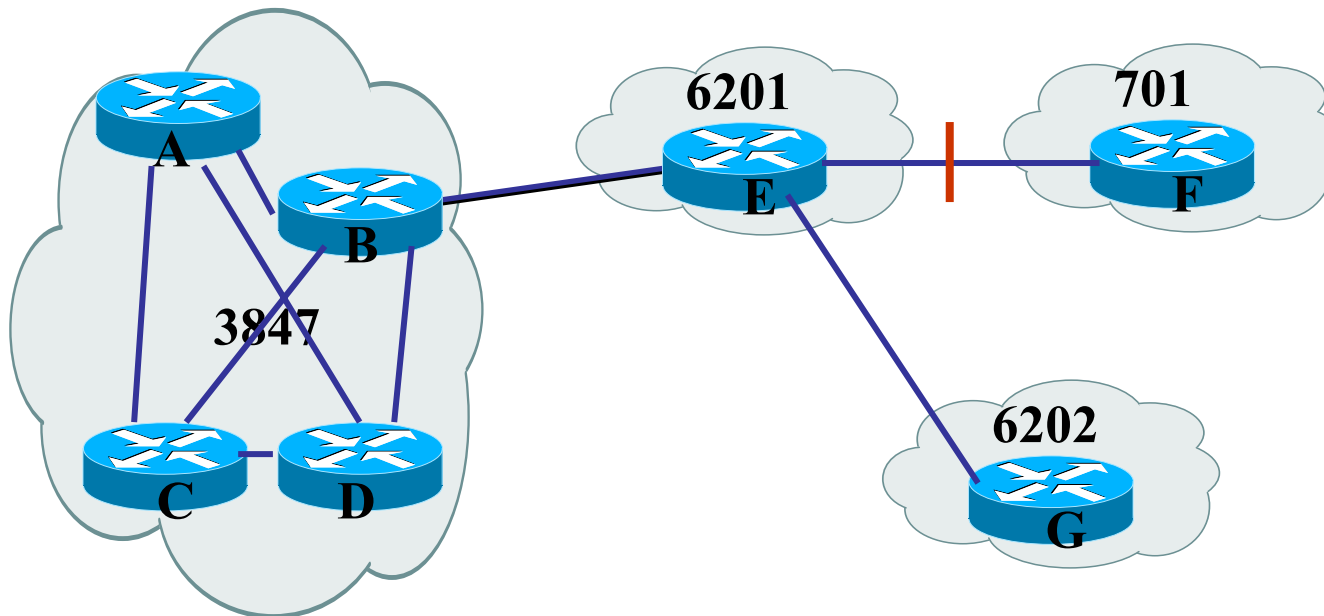
Παράδειγμα 1

- Θέλουμε ο RTC να απορρίπτει όλες τις ενημερώσεις με AS_path λίστα που να ξεκινά και τελειώνει με 200 (δηλαδή όλες που παράγονται από το AS200) και να δέχεται όλες τις υπόλοιπες



Παράδειγμα 2

- Ο Β θέλουμε να δέχεται διαδρομές από τα AS6201 & 6202 και μόνο



```
ip as-path access-list 10 permit ^6201$  
ip as-path access-list 10 permit ^6201_6202$  
ip as-path access-list 10 deny .*
```

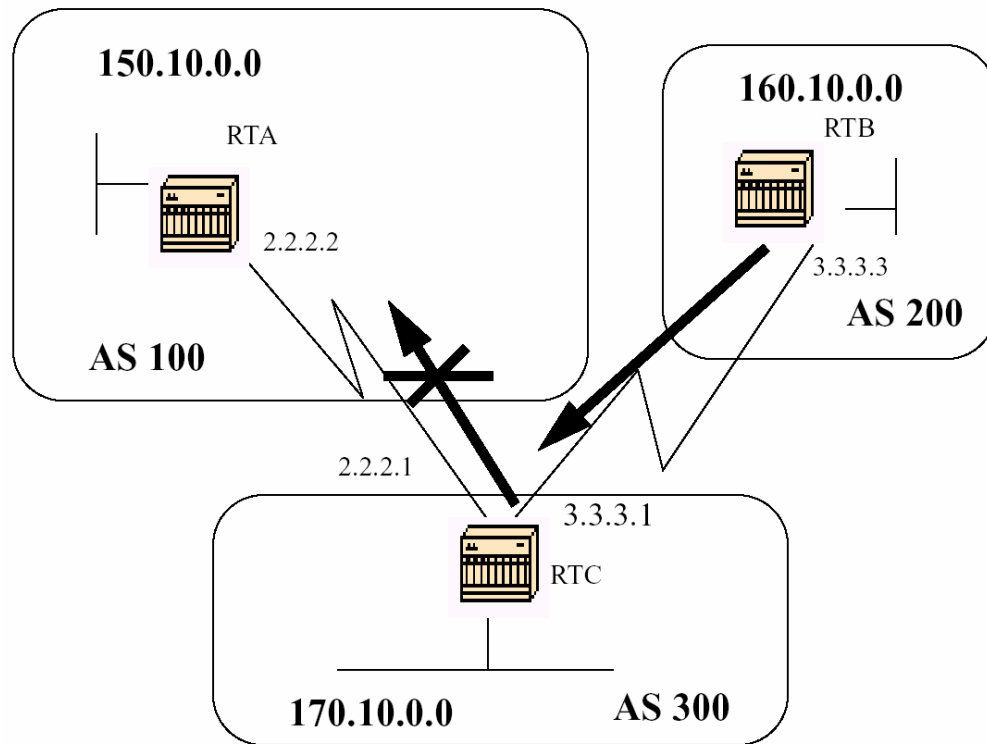
Παράδειγμα 3

- Θέλουμε να επιτραπούν μόνο οι ενημερώσεις που έχουν στη ν AS_path μόνο το AS=200 ή το AS=150

```
router bgp 100
neighbor 220.200.1.1 remote-as 210
neighbor 220.200.1.1 filter-list 5 out
neighbor 220.200.1.1 filter-list 6 in
!
ip as-path access-list 5 permit ^200$
ip as-path access-list 6 permit ^150$
```

Φίλτρο Community

- Κοινότητα είναι μία ομάδα προορισμών που μοιράζονται κάποια κοινή ιδιότητα
- Θέλουμε ο RTB να ορίσει μια κοινότητα για τις διαδρομές που ανακοινώνει, ώστε ο RTC να μην τις διαδώσει παραπέρα



RTB#

```
router bgp 200
```

```
network 160.10.0.0
```

```
neighbor 3.3.3.1 remote-as 300
```

```
neighbor 3.3.3.1 send-community
```

```
neighbor 3.3.3.1 route-map setcommunity out
```

```
route-map setcommunity
```

```
match ip address 1
```

```
set community no-export
```

```
access-list 1 permit 0.0.0.0 255.255.255.255
```

- Η κοινότητα no-export έχει το νόημα ότι οι διαδρομές δεν ανακοινώνονται σε άλλο AS

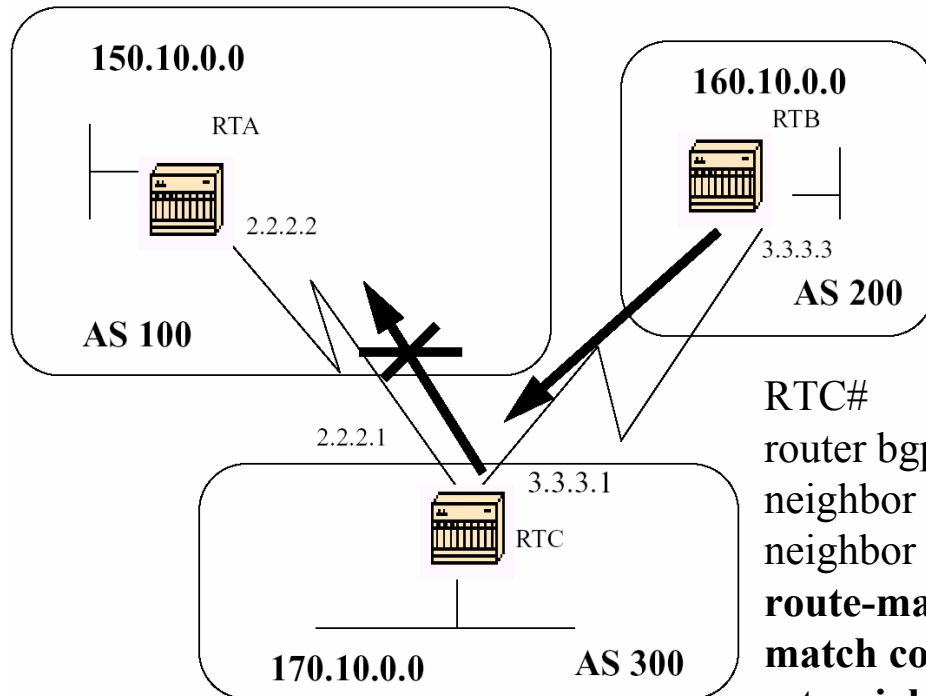
Παράδειγμα 1

- Θέλουμε ο RTB να ορίσει για τις διαδρομές που ανακοινώνει community=100 200 additive (η 100 200 θα προστεθεί σε τυχόν άλλες υπάρχουσες τιμές)

ip community-list community-list-number {**permit|deny**} community-number

- Οι τιμές αυτές μπορούν να χρησιμοποιηθούν για να τεθούν διάφορες ιδιότητες όπως το βάρος, ανάλογα με την τιμή της κοινότητας
- Κάθε διαδρομή που περιέχει community=100 θα πάρει βάρος 20
- Κάθε διαδρομή που περιέχει μόνο την community=200 θα πάρει βάρος 10
- Οι υπόλοιπες δεν θα απορριφθούν

Παράδειγμα 1



RTB#

```
router bgp 200
network 160.10.0.0
neighbor 3.3.3.1 remote-as 300
neighbor 3.3.3.1 send-community
neighbor 3.3.3.1 route-map setcommunity out
route-map setcommunity
match ip address 2
set community 100 200 additive
access-list 2 permit 0.0.0.0 255.255.255.255
```

RTC#

```
router bgp 300
neighbor 3.3.3.3 remote-as 200
neighbor 3.3.3.3 route-map check-community in
route-map check-community permit 10
match community 1
set weight 20
route-map check-community permit 20
match community 2 exact
set weight 10
route-map check-community permit 30
match community 3
ip community-list 1 permit 100
ip community-list 2 permit 200
ip community-list 3 permit internet
```

Route Maps

- Είναι μία μέθοδος **ελέγχου** και **τροποποίησης** της πληροφορίας δρομολόγησης
- Ο τρόπος που λειτουργεί είναι ότι ορίζουμε **συνθήκες** για **αναδιανομή** των διαδρομών από το ένα πρωτόκολλο στο άλλο ή όταν **εισάγονται** ή **εξάγονται** στο/από το BGP
- Χρησιμοποιούνται επίσης για λειτουργίες όπως το **traffic shaping** και το **Policy routing**
- Χρησιμοποιούνται πάρα πολύ

- Είναι σαν ένα πρόγραμμα
- Έχει αριθμούς γραμμής
- Κάθε γραμμή είναι μια ξεχωριστή **συνθήκη/ενέργεια**

Αυτή είναι

εάν *match* τότε κάνε *expression* και *exit*

διαφορετικά

εάν *match* τότε κάνε *expression* και *exit*

διαφορετικά *etc*

Ορισμός

route-map <name> permit|deny <sequence> [match options] [set options]

name Το όνομα του route-map. Πολλές περιπτώσεις στιγμές του ίδιου route map μπορούν να οριστούν

sequence Ένδειξη της θέσης που έχει η περίπτωση σε μια λίστα με το ίδιο όνομα

Match options Ορίζονται κάποια κριτήρια

Set options Ενέργειες που γίνονται αν ικανοποιούνται τα παραπάνω κριτήρια

Permit

- Αν το πρόθεμα που εξετάζεται περάσει την match τότε οι εντολές set εκτελούνται και η route-map τερματίζει
- Αν δεν περάσει την match η επόμενη route-map με το ίδιο όνομα και μεγαλύτερο αριθμό σειράς εκτελείται
- Αν εκτελεστούν όλες οι route-map χωρίς το πρόθεμα να περάσει καμία match, το πρόθεμα απορρίπτεται

Deny

- Αν το πρόθεμα περάσει την match τότε η διαδρομή απορρίπτεται και η route-map τερματίζει
- Αν το πρόθεμα δεν περάσει την match η επόμενη εκτελείται
- Αν εκτελεστούν όλες οι route-map χωρίς το πρόθεμα να περάσει καμία match, το πρόθεμα απορρίπτεται

Route Maps

neighbor ip-address **route-map** route-map-name

- Η εντολή **neighbor** μπορεί να χρησιμοποιηθεί με τα **route maps** για να φιλτράρει ή να αλλάξει-θέσει παραμέτρους σε εισερχόμενες – εξερχόμενες ενημερώσεις
- Όταν εφαρμόζεται σε **εισερχόμενες** ενημερώσεις με την εντολή **neighbor**, με ταίριασμα με βάση την **IP διεύθυνση** δεν έχει αποτέλεσμα

Εντολές

match

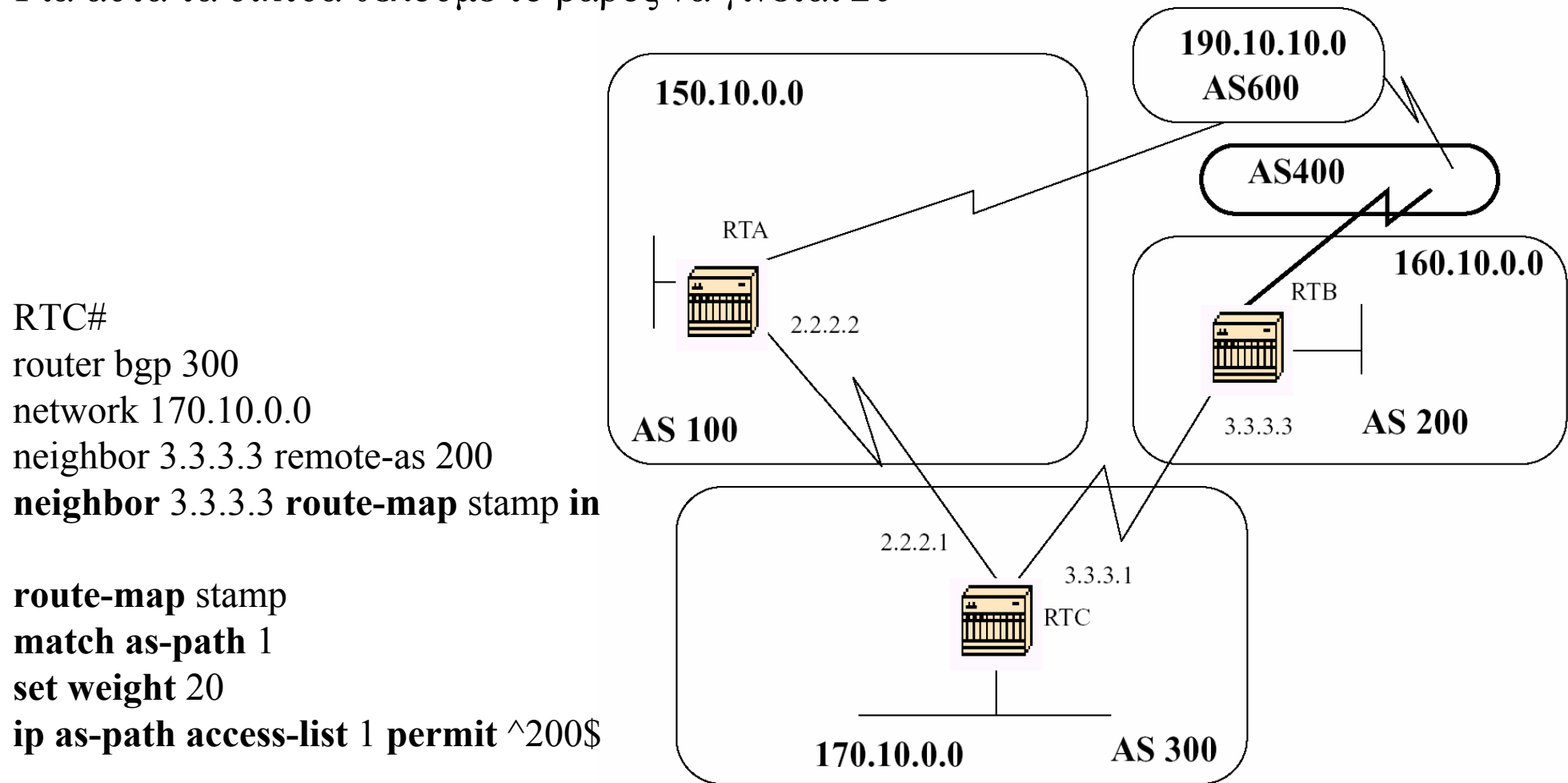
match as-path
match community
match clns
match interface
match ip address
match ip next-hop
match ip route-source
match metric
match route-type
match tag

set

set as-path
set automatic-tag
set community
set clns
set interface
set default interface
set ip next-hop
set ip default next-hop
set ip precedence
set tos
set level
set local-preference
set metric
set metric-type
set next-hop
set origin
set tag
set weight

Παράδειγμα 1

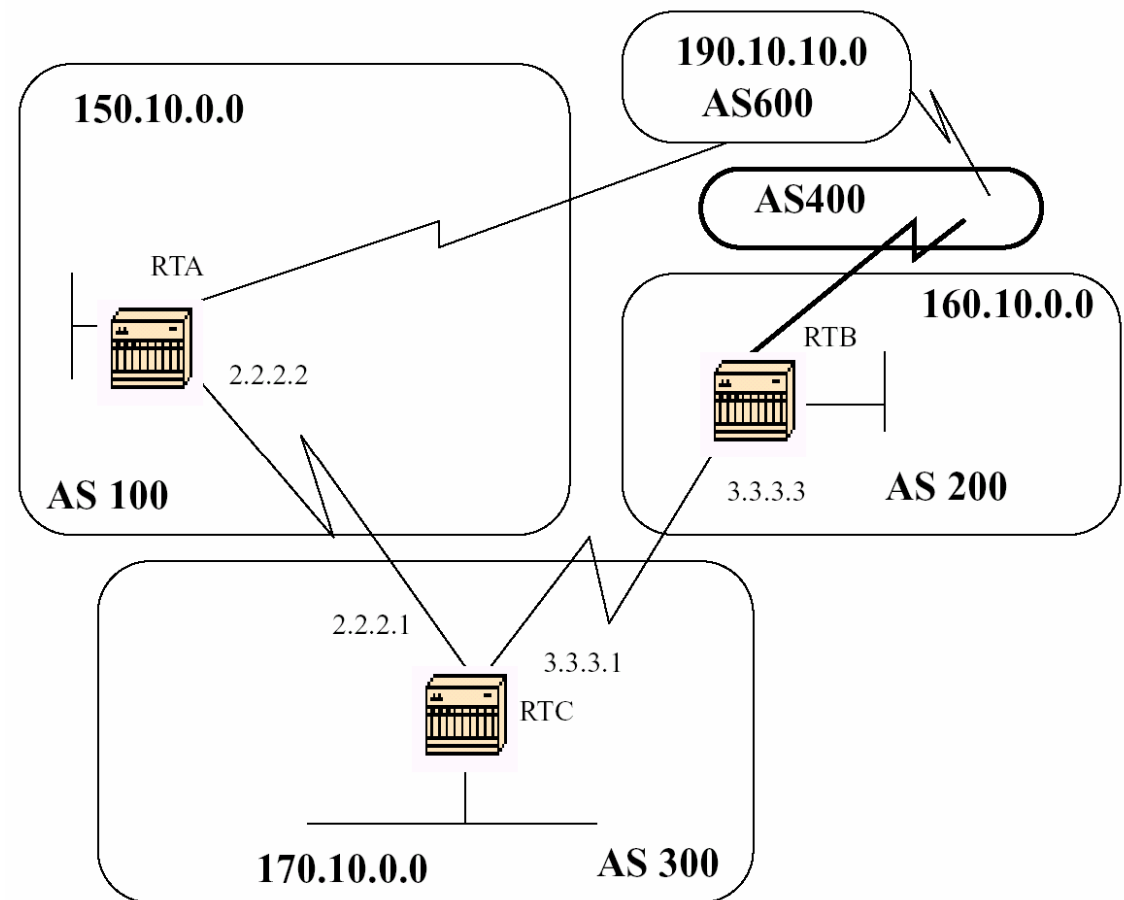
- Θέλουμε ο RTC να δέχεται ενημερώσεις από το AS200 μόνο για δίκτυα που είναι τοπικά εκεί (άρα ενημερώσεις με AS_path={200})
- Για αυτά τα δίκτυα θέλουμε το βάρος να γίνεται 20



Παράδειγμα 2

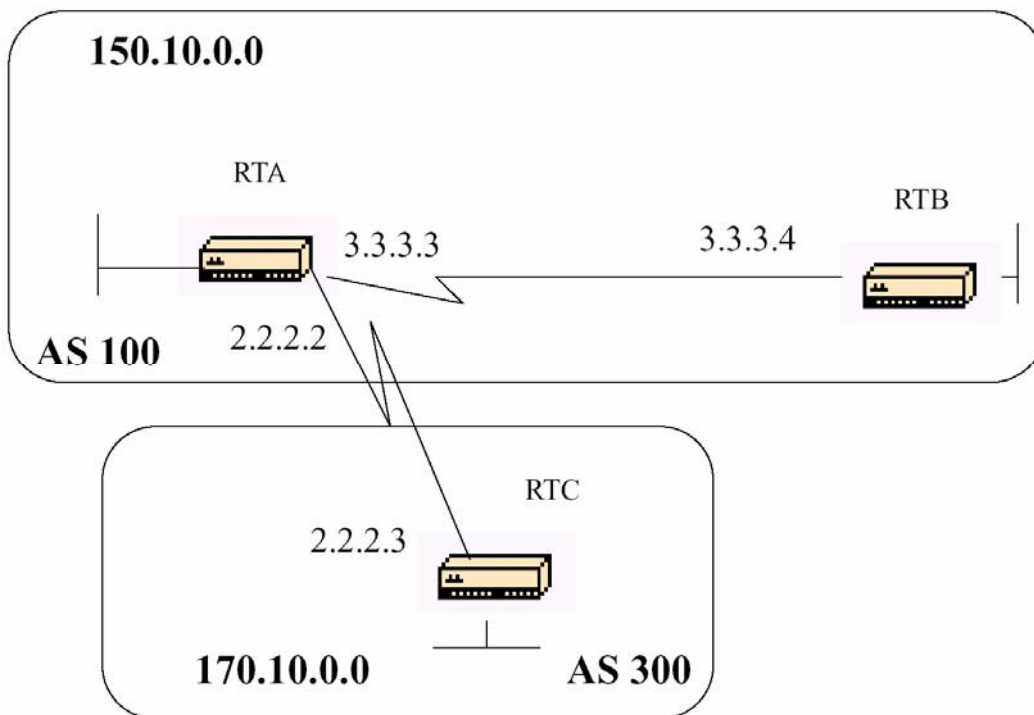
- Θέλουμε ο RTC να δέχεται ενημερώσεις από το AS200 μόνο για δίκτυα που είναι τοπικά εκεί (άρα ενημερώσεις με AS_path={200}) και μάλιστα με βάρος 20
- Ενημερώσεις που γεννήθηκαν στο AS400 να απορρίπτονται
- Όλες οι υπόλοιπες να έχουν βάρος 10

```
RTC#  
router bgp 300  
network 170.10.0.0  
neighbor 3.3.3.3 remote-as 200  
neighbor 3.3.3.3 route-map stamp in  
  
route-map stamp permit 10  
match as-path 1  
set weight 20  
route-map stamp permit 20  
match as-path 2  
set weight 10  
ip as-path access-list 1 permit ^200$  
ip as-path access-list 2 permit ^200 600 .*
```



Παράδειγμα 3

- Οι RTA, RTB τρέχουν rip
- Οι RTA, RTC τρέχουν BGP
- Ο RTA αναδιανέμει το BGP στο rip
- Θέλουμε ο RTA να αναδιανέμει στον RTB τις διαδρομές σχετικά με το 170.10.0.0/16 με metric=2 και τις υπόλοιπες με metric=5



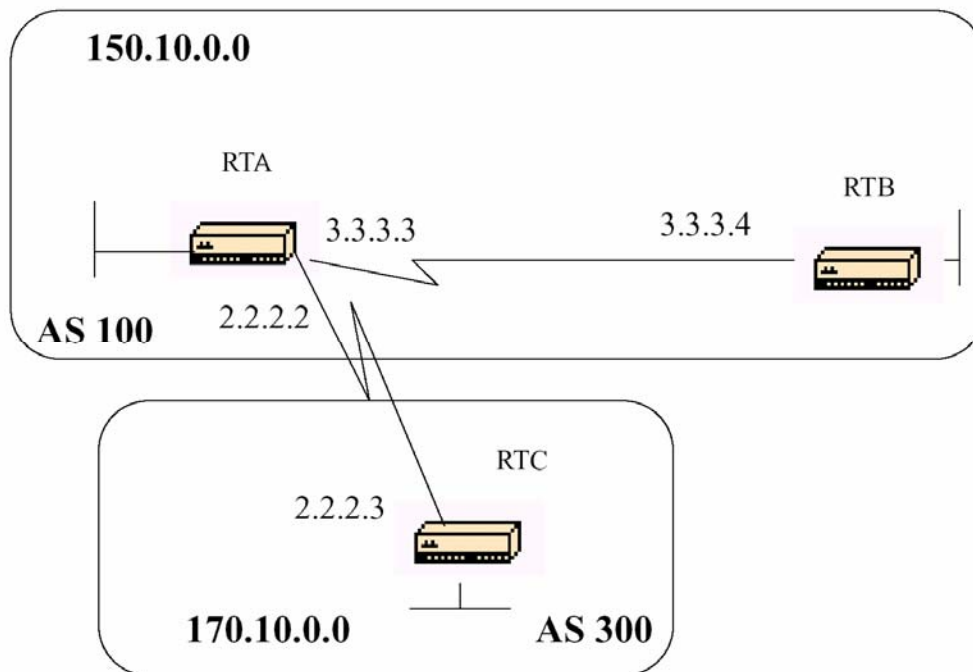
```
RTA#  
router rip  
network 3.0.0.0  
network 2.0.0.0  
network 150.10.0.0  
passive-interface Serial0  
redistribute bgp 100 route-map SETMETRIC
```

```
router bgp 100  
neighbor 2.2.2.3 remote-as 300  
network 150.10.0.0
```

```
route-map SETMETRIC permit 10  
match ip-address 1  
set metric 2  
route-map SETMETRIC permit 20  
set metric 5  
access-list 1 permit 170.10.0.0 0.0.255.255
```

Παράδειγμα 4

- Δεν θέλουμε το AS100 να δέχεται ενημερώσεις για το 170.10.0.0/16
- Οι route maps δεν μπορούν να εφαρμοστούν για τις εισερχόμενες ανακοινώσεις, όταν η match βασίζεται στην IP διεύθυνση, οπότε:



```
RTC#  
router bgp 300  
network 170.10.0.0  
neighbor 2.2.2.2 remote-as 100  
neighbor 2.2.2.2 route-map STOPUPDATES out
```

```
route-map STOPUPDATES permit 10  
match ip address 1  
access-list 1 deny 170.10.0.0 0.0.255.255  
access-list 1 permit 0.0.0.0 255.255.255.255
```

Παράδειγμα 5

- Θέλουμε οι εισερχόμενες διαδρομές για προθέματα 10.0.0.0/8 να πέρνουν τοπική προτίμηση 120 και οι εισερχόμενες για προθέματα 20.0.0.0/8 να πέρνουν τοπική προτίμηση 80

```
router bgp 100
neighbor 1.1.1.1 route-map infilter in
!
route-map infilter permit 10
match ip address prefix-list HIGH-PREF
set local-preference 120
!
route-map infilter permit 20
match ip address prefix-list LOW-PREF
set local-preference 80
!
route-map infilter permit 30
!
ip prefix-list HIGH-PREF permit 10.0.0.0/8
ip prefix-list LOW-PREF permit 20.0.0.0/8
```

Παράδειγμα 6

- Θέλουμε οι διαδρομές που παράγονται από το AS=150 να παίρνουν τοπική προτίμηση 80, ενώ αυτές που διασχίζουν το AS=210 να πέρνουν τοποική προτίμηση 200

```
router bgp 100
neighbor 220.200.1.2 route-map filter-on-as-path in
!
route-map filter-on-as-path permit 10
match as-path 1
set local-preference 80
!
route-map filter-on-as-path permit 20
match as-path 2
set local-preference 200
!
route-map filter-on-as-path permit 30
!
ip as-path access-list 1 permit _150$
ip as-path access-list 2 permit _210_
```

Παράδειγμα 7

•Θέλουμε για όλα τα προθέματα 172.168.0.0/16 που είναι μεγαλύτερα ή ίσα από /17 να οριστεί community=no-export, οπότε όταν αυτά ληφθούν από τον ομότιμο δεν θα ανακοινωθούν σε άλλα AS

```
router bgp 100
neighbor 220.200.1.1 remote-as 200
neighbor 220.200.1.1 send-community
neighbor 220.200.1.1 route-map set-community out
!
route-map set-community permit 10
match ip address prefix-list NO-ANNOUNCE
set community no-export
!
route-map set-community permit 20
!
ip prefix-list NO-ANNOUNCE permit 172.168.0.0/16 ge 17
```

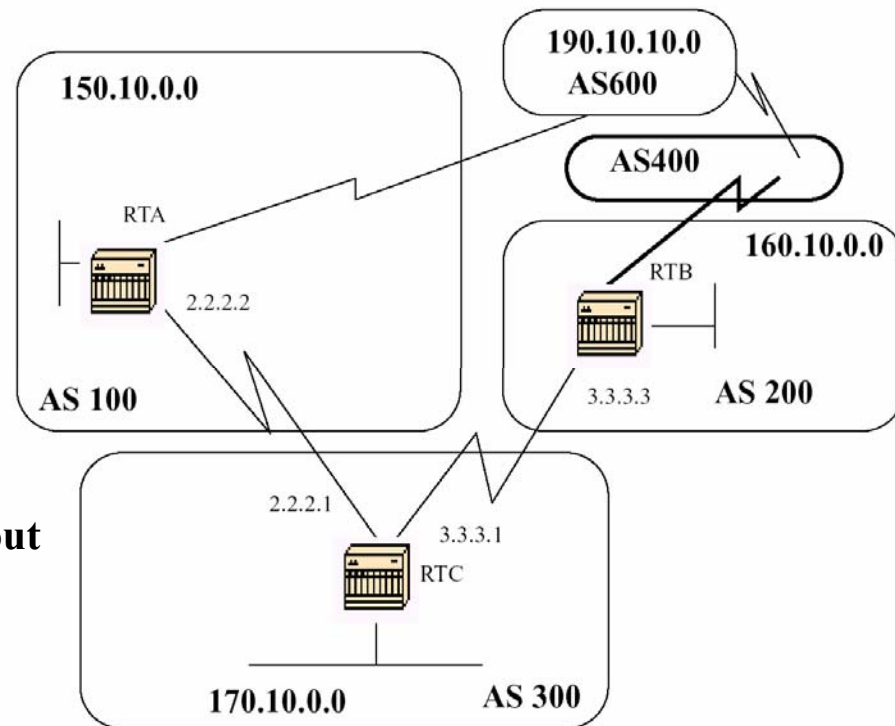
Route-Map , set as-path prepend

- Μπορούμε να τροποποιήσουμε την **AS_path** λίστα με σκοπό να **επηρεάσουμε** την επιλογή των διαδρομών με την εντολή **set as-path prepend** <as-path#> <as-path#> ...
- Μπορούμε να επαναλάβουμε τον δικό μας ASN στη λίστα ώστε να την κάνουμε αρκετά μεγάλη
- Προσοχή χρειάζεται να μην τοποθετήσουμε ξένο ASN διότι τότε θα δημιουργηθούν βρόχοι και κατάργηση BGP συνόδων
- Αν ο απέναντι έχει ορίσει τιμές weight και local preference τότε αφού αυτές έχουν μεγαλύτερη ισχύ θα υπερισχύσουν

Παράδειγμα

- Θέλουμε να προσπαθήσουμε να επηρεάσουμε τη διαδρομή που θα επιλέξει το AS600 προς το AS300, ώστε να προτιμήσει αυτή μέσω του AS100

```
router bgp 300
network 215.7.0.0
neighbor 2.2.2.2 remote-as 100
neighbor 2.2.2.2 route-map SETPATH out
!
route-map SETPATH permit 10
set as-path prepend 300 300
```



- Το AS600 θα λάβει δύο διαδρομές για το 170.10.0.0, μία με $AS_path=\{100, 300, 300, 300\}$ και μία με $AS_path=\{400, 200, 300\}$

Capabilities

BGP Capabilities, RFC2842

- Οι παράμετροι για τις **δυνατότητες** που υποστηρίζονται μεταδίδονται μέσω του **Open** μηνύματος
- Αν κάποιες είναι άγνωστες ή μη υποστηριζόμενες από τον παραλήπτη, τότε ένα **Notification** μήνυμα θα αποσταλεί και η διαδικασία **διαπραγμάτευσης** θα επαναληφθεί

Κωδικοί:

- 0 - 63 έχουν ανατεθεί από την IANA από IETF consensus
- 64 - 127 έχουν ανατεθεί από την IANA “first come first served”
- 128 - 255 είναι vendor specific

BGP Capabilities, RFC2842

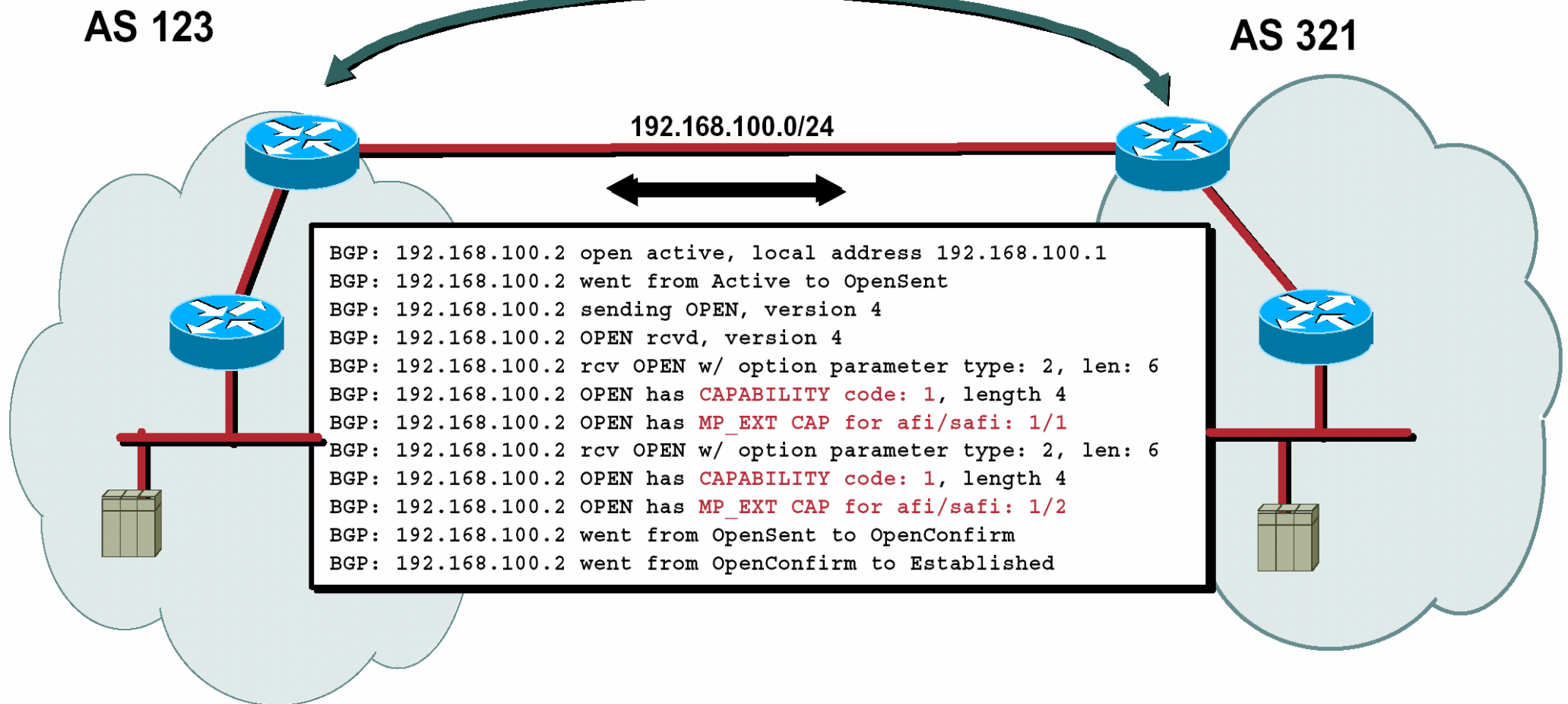
•Οι τρέχουσες **δυνατότητες** είναι:

- 0 Δεσμευμένη τιμή [RFC3392]
- 1 Multiprotocol Extensions for BGP-4 [RFC2858]
- 2 Route Refresh Capability for BGP-4 [RFC2918]
- 3 Cooperative Route Filtering Capability []
- 4 Multiple routes to a destination capability [RFC3107]
- 64 Graceful Restart Capability []
- 65 Support for 4 octet ASNs []
- 66 Support for Dynamic Capability

<http://www.iana.org/assignments/capability-codes>

Διαπραγμάτευση δυνατοτήτων

BGP session for unicast and multicast NLRI



Ευστάθεια-Δυναμική

Δυναμική του BGP

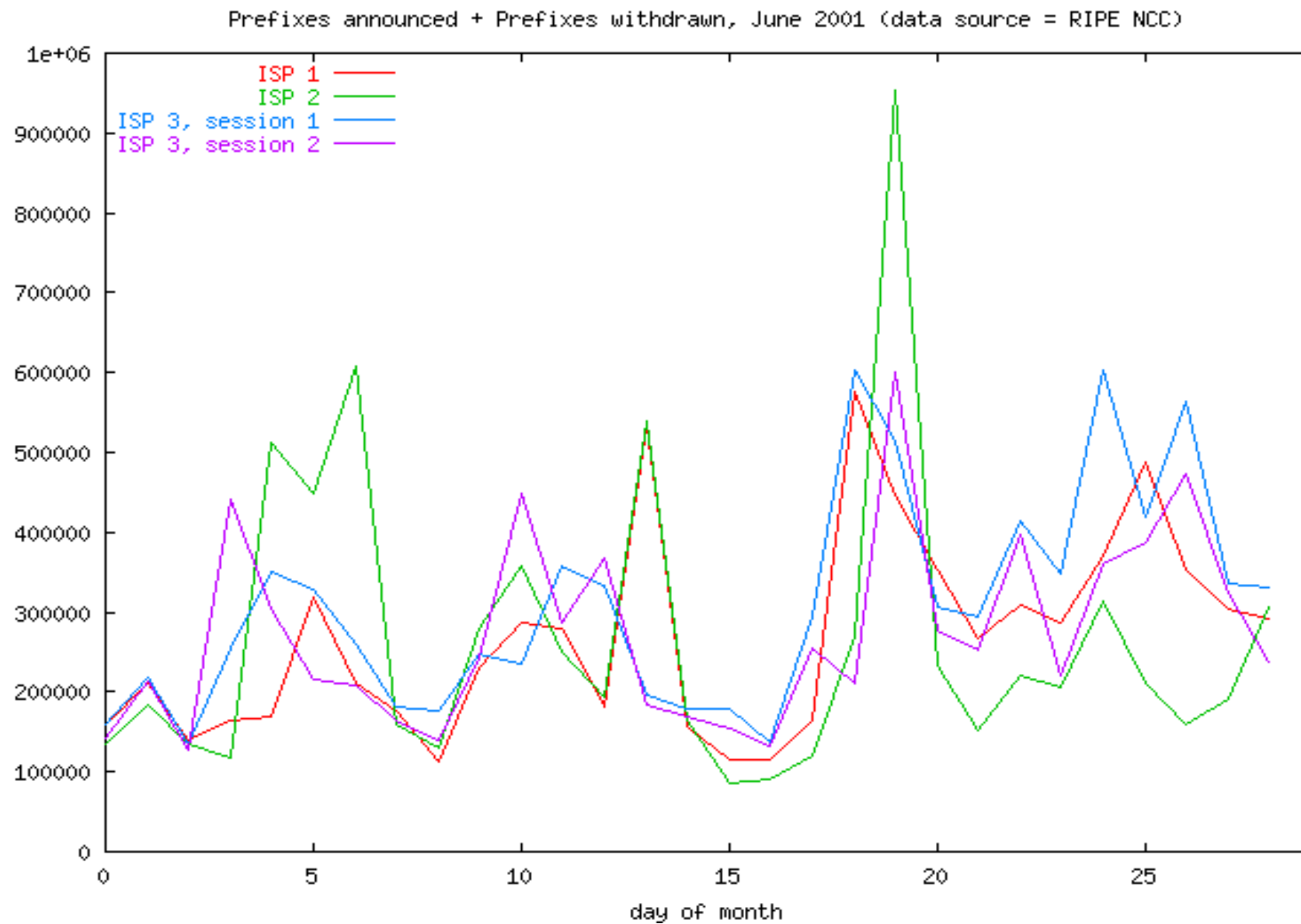
Περιλαμβάνει τα κεφάλαια:

1. Πλήθος ενημερώσεων στο δίκτυο
2. Πόσο γρήγορα γίνεται η ενημέρωση των διαδρομών, δηλαδή η **σύγκλιση**

Ο στόχος είναι:

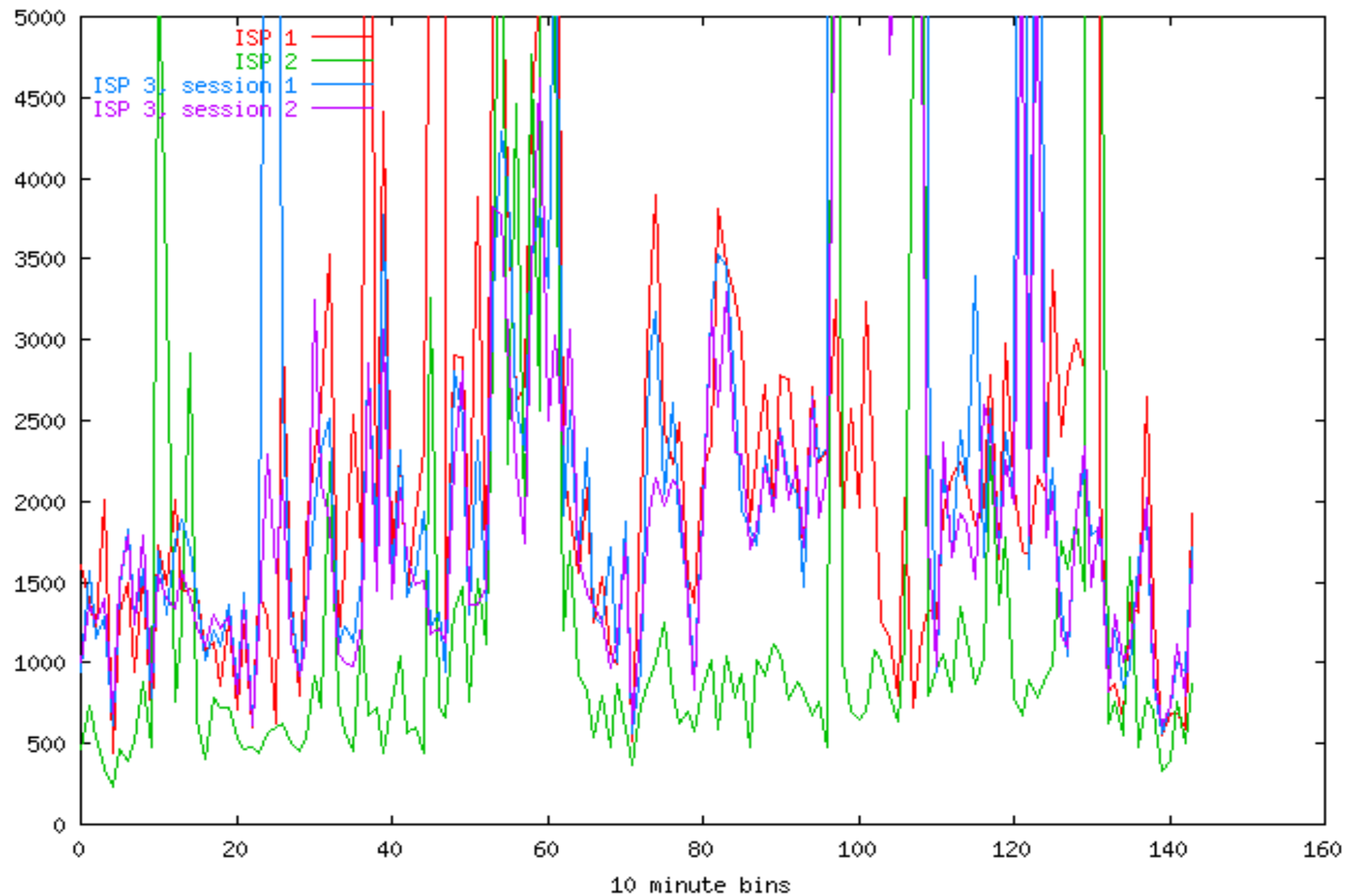
- Γρήγορη σύγκλιση
- Ελάχιστος αριθμός ενημερώσεων

Αριθμός καθημερινών ενημερώσεων



Αριθμός ενημερώσεων

Prefixes announced + Prefixes withdrawn, June 25, 2001 (data source = RIPE NCC)



Αναμενόμενοι λόγοι για ενημερώσεις

Τεράστιος αριθμός ενημερώσεων διότι:

- Δίκτυα ανεβοκατεβαίνουν
- Δρομολογητές επανεκκινούν
- Βλάβες υλικού, προβληματικές κάρτες δικτύου, κομμένες οπτικές ίνες, πλημύρες και καταστροφές

Κακοί λόγοι για ενημερώσεις

- Λάθος διάρθρωση
- Ο μηχανισμός Route flap dampening δεν χρησιμοποιείται συχνά
- Το BGP εξερευνά πολλές εναλλακτικές διαδρομές
- Λάθη στο λογισμικό υλοποίησης των πρωκόλλων δρομολόγησης
- Επανεκκινήσεις στο BGP λόγω συμφόρησης ή λόγω έλλειψης διαλειτουργικότητας
- Οι BGP σύνοδοι είναι εύθραυστοι. Μία κακοσχηματισμένη ενημέρωση είναι αρκετή να επανεκκινήσει μια BGP σύνοδο και να προκαλέσει αλλαγή κατάστασης σε 100K διαδρομές
- Αυτό είναι επακόλουθο της προσέγγισης να στέλνονται ενημερώσεις μόνο των αλλαγών

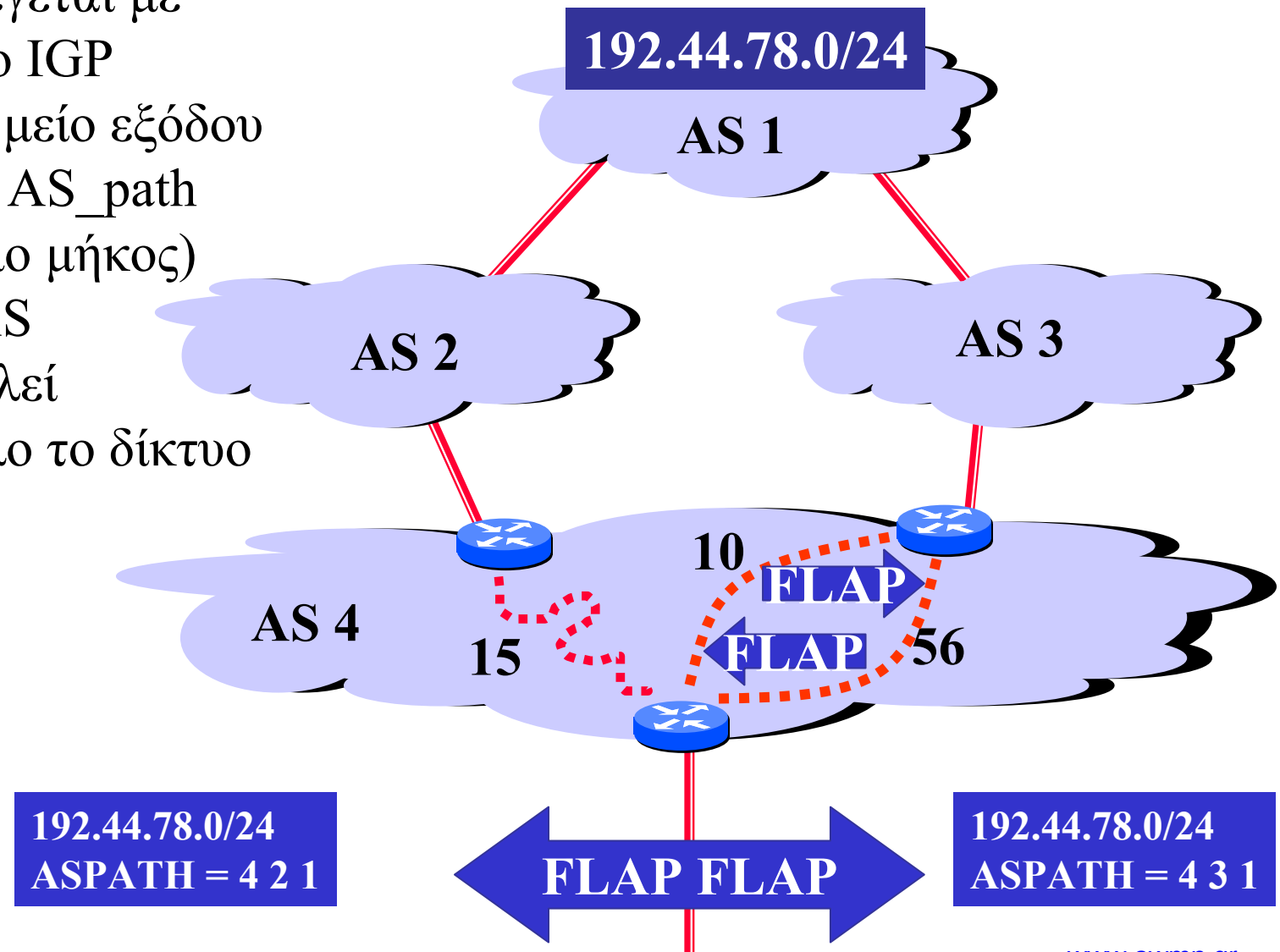
Κακοί λόγοι για ενημερώσεις

- Η αστάθεια του IGP μπορεί να εξαχθεί με τη χρήση των MEDs
- Κατώτερες του βέλτιστου επιλογές υλοποίησης από τους κατασκευαστές
- Μυστικοί αλγόριθμοι δρομολόγησης οι οποίοι προσπαθούν να επιτύχουν διάφορα μαγικά ..
- Παράξενες αλληλοεπιδράσεις των διαφόρων τακτικών (ταλάντωση του MED, κακές επινοήσεις, ..)
- Νάνοι, φαντάσματα, στοιχειά και νεράϊδες
-Τελικά άπειροι λόγοι....

Η αστάθεια του IGP μπορεί να εξαχθεί

Παράδειγμα

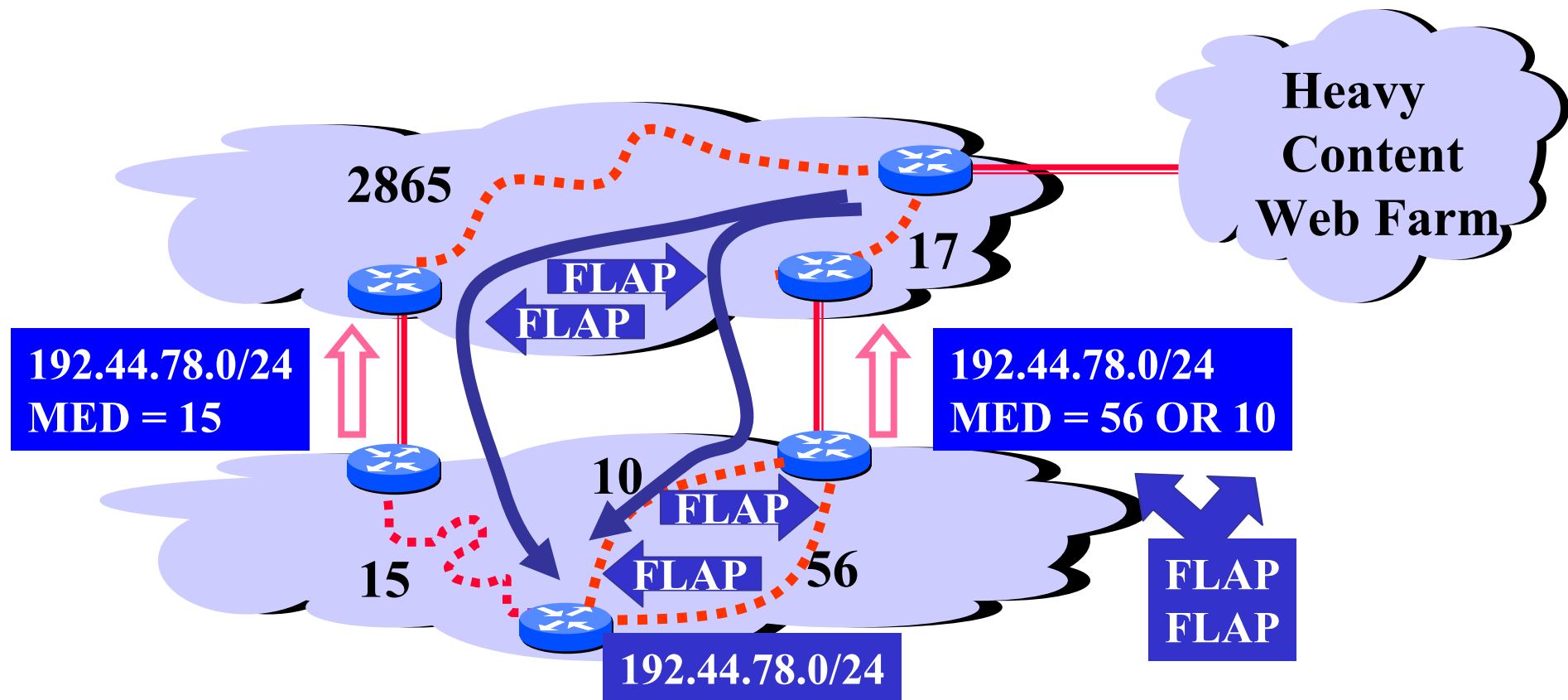
- Η διαδρομή επιλέγεται με βάση το μικρότερο IGP κόστος προς το σημείο εξόδου από το AS (οι δύο AS_path λίστες έχουν το ίδιο μήκος)
- Εσωτερική στο AS ταλάντωση προκαλεί ταλαντώσεις σε όλο το δίκτυο



Η αστάθεια του IGP μπορεί να εξαχθεί μέσω των MED

Παράδειγμα

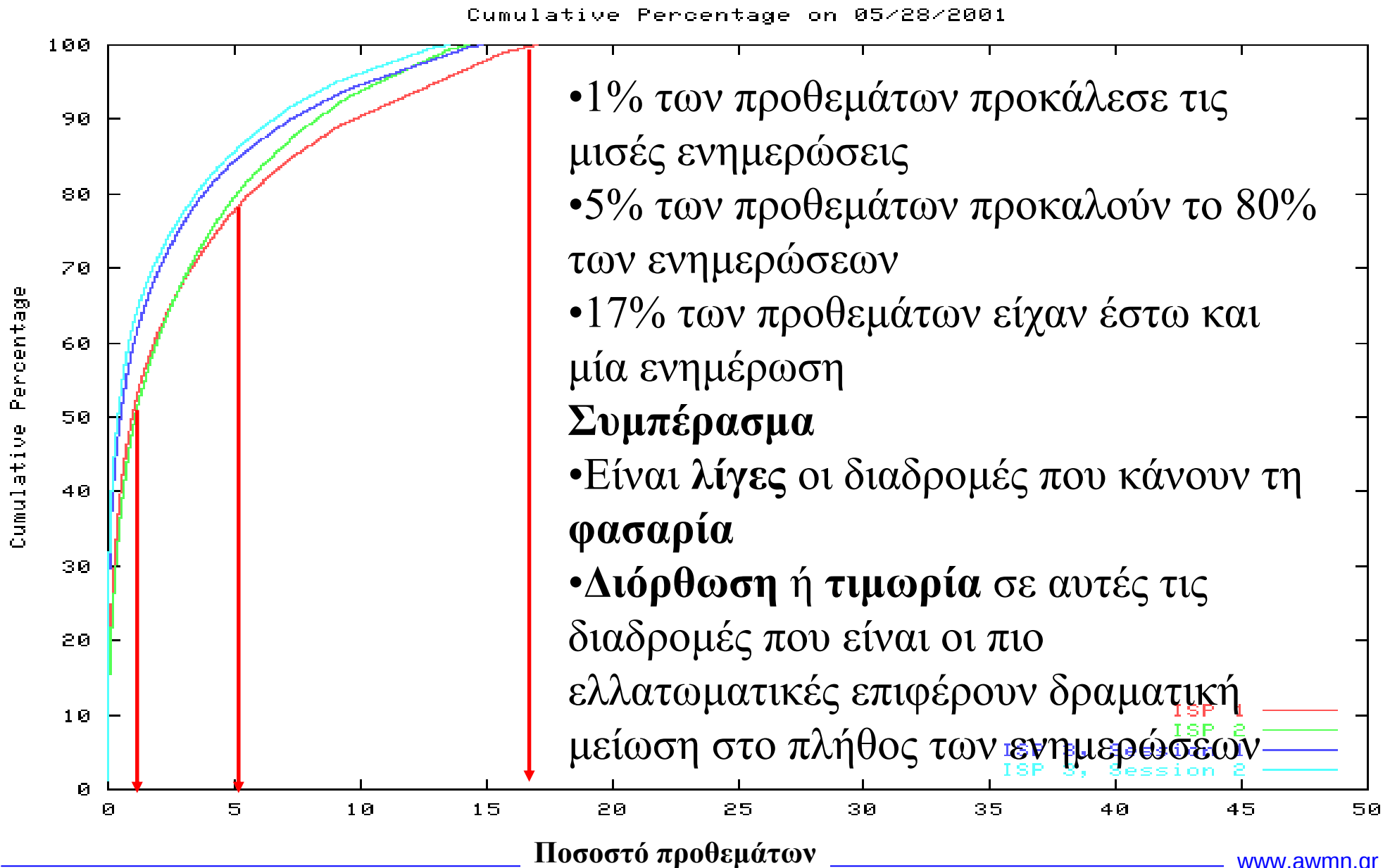
- Γράφουμε σαν MED την τιμή του IGP κόστους
- Έτσι όμως τυχόν εσωτερική ταλάντωση μπορεί να εξαχθεί σε όλο το δίκτυο



Γιατί πειράζει να έχουμε πολλές ενημερώσεις;

- Οι ταλαντώσεις αυτές συμβαίνουν συνέχεια σε ένα ευρύ δίκτυο όπως το internet
- Αυτό προκαλεί όμως μεγάλη σπατάλη σε **CPU**, **μνήμη** και **εύρος** και βεβαίως μη αξιόπιστη μετάδοση
- Ταλαντώσεις σε **EBGP** προκαλούν μία γενική αναστάτωση στους πίνακες δρομολόγησης γιατί οι μεταβολές διαδίδονται στο διαδίκτυο και κάθε δρομολογητής πρέπει να επεξεργαστεί τις **αλλαγές**
- Ταλαντώσεις σε **IBGP** προκαλεί ακανόνιστη ροή της κίνησης και προβλήματα προσέγγισης στο **τοπικό AS** και μπορεί να επηρεάσει την **ευστάθεια** του **EBGP** αν οι IBGP διαδρομές ανακοινώνονται σε EBGP ομότιμους

Πόσες διαδρομές δεν είναι σταθερές;



Μηχανισμοί για μείωση των ενημερώσεων – βελτίωσης της ευστάθειας

- Τρόπος αποστολής ενημερώσεων
- Συνάθροιση διαδρομών (aggregation) - CIDR
- Περιορισμός του ρυθμού με τον οποίο στέλνονται οι ενημερώσεις
- Χρήση διευθύνσεων βρόχου για τις IBGP συνόδους
- Ομάδες ομοτίμων (Peer groups)
- Αλληλοεπίδραση BGP/IGP
- Επαναδιάρθρωση
 - Ανανέωση διαδρομών (Route Refresh)
 - Ήπια αναδιάρθρωση (BGP soft reconfiguration)
- Απόσβεση ταλαντώσεων διαδρομών (Route Flap Dampening)

Τρόπος αποστολής ενημερώσεων

- Ένας ομότιμος που θέλει να αποσύρει μία διαδρομή και να στείλει μια ενημέρωση για ένα πιο συγκεκριμένο πρόθεμα, τα συνδιάζει στο **ίδιο μήνυμα ενημέρωσης**

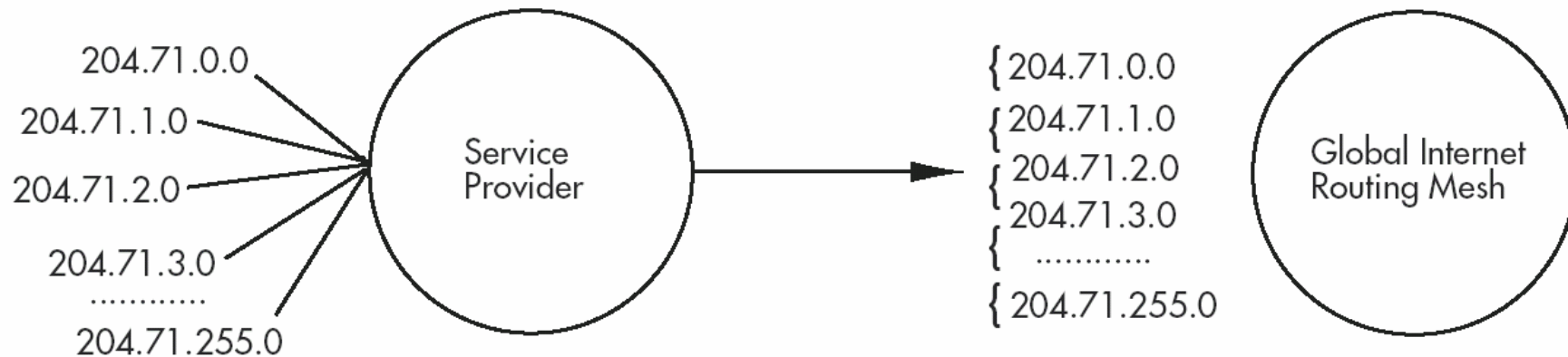
CIDR, Classless Interdomain Routing

- Έκτος από τις BGP ιδιότητες αξιοποιούμε και το CIDR ώστε να **μειωθεί** το μέγεθος των **πινάκων** δρομολόγησης
- Με το CIDR μπορούμε να κάνουμε συνάθροιση **πολλών classfull** δικτύων σε ένα **classless** παρέχοντας δρομολόγηση σε πολλούς προορισμούς με **μία μόνο ανακοίνωση**
- Στην περίπτωση που υπάρχει **υπερκάλυψη** των προθεμάτων το CIDR επιλέγει τη διαδρομή με το **καλύτερο** ταίριασμα

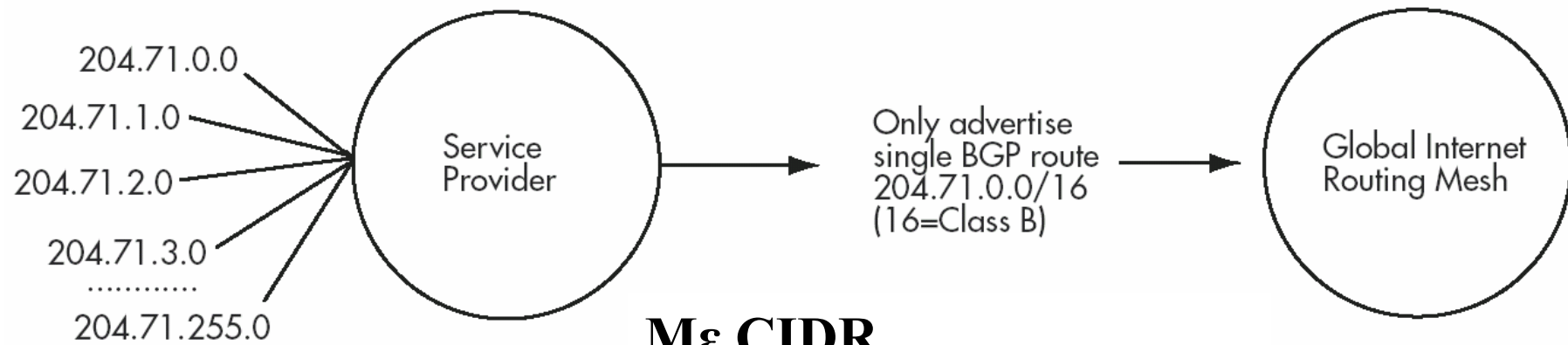
Παράδειγμα 1

- Ένας ISP κατέχει το τμήμα IP διευθύνσεων 195.10.x.x το οποίο προέρχεται από το χώρο διευθύνσεων για δίκτυα τάξης C
- Αυτό το τμήμα από 256 δίκτυα τάξης C, από το 195.10.0.x ως το 195.10.255.x
- Έστω ότι το κάθε δίκτυο τάξης C ανατίθεται από τον ISP στους πελάτες του
- Χωρίς CIDR, ο ISP θα ανακοίνωνε 256 δίκτυα τάξης C προς τους BGP ομότιμους
- Με CIDR, BGP μπορεί να ανακοινώσει ένα δίκτυο μόνο το 195.10.x.x το οποίο έχει το ίδιο μέγεθος με ένα δίκτυο τάξης B
- Έτσι θα υπάρχει μια σημαντική ελάττωση στον πίνακα δρομολόγησης του BGP

Παράδειγμα 1



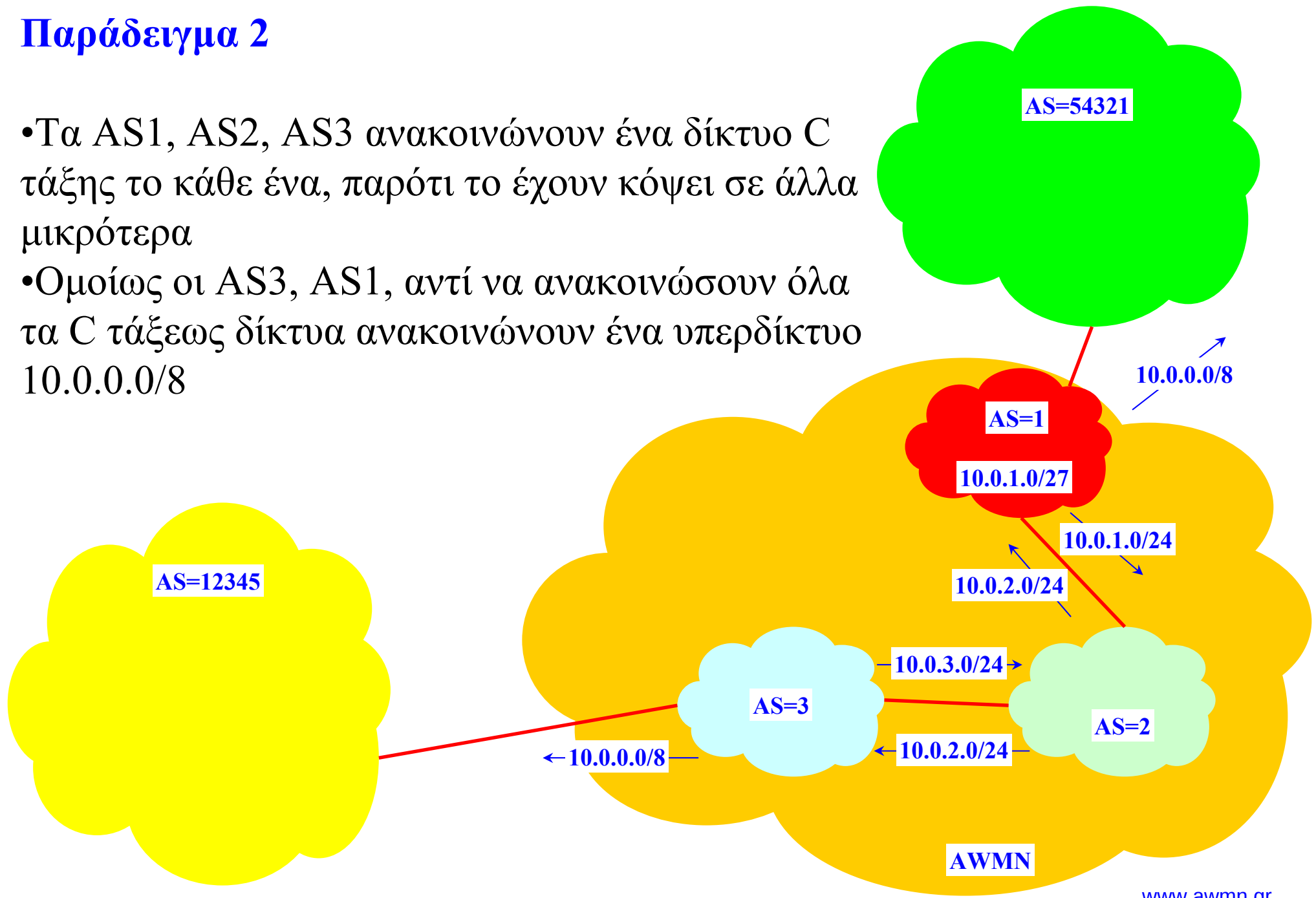
Χωρίς CIDR



Με CIDR

Παράδειγμα 2

- Τα AS1, AS2, AS3 ανακοινώνουν ένα δίκτυο C τάξης το κάθε ένα, παρότι το έχουν κόψει σε άλλα μικρότερα
- Ομοίως οι AS3, AS1, αντί να ανακοινώσουν όλα τα C τάξεως δίκτυα ανακοινώνουν ένα υπερδίκτυο 10.0.0.0/8



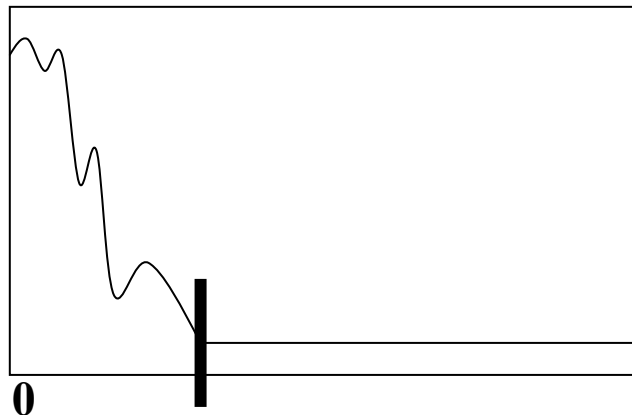
Περιορισμός του ρυθμού με τον οποίο στέλνονται οι ενημερώσεις

- Αποστολή ομάδας ενημερώσεων κάθε **inRouteAdvertisementInterval** seconds (+/- random fuzz)
- Για να μην έχουμε αιχμές σε συγκεκριμένες στιγμές το διάστημα αυτό αυξομειώνεται κατά μικρό τυχαίο χρονικό διάστημα (**jitter**)
- Τυπική τιμή τα **30'**
- Ο δρομολογητής μπορεί να αλλάξει γνώμη για τις βέλτιστες διαδρομές **όσες φορές θέλει** μέσα σε αυτό το διάστημα, χωρίς να ενημερώσει τους γείτονες
- Αποτελεσματικός τρόπος **απόσβεσης των ταλαντώσεων** που παρουσιάζονται σε πρωτόκολλα **τύπου vectoring**
- Ο ελάχιστος αυτός χρόνος αναφέρεται σε **ανακοινώσεις** για τον **ίδιο προορισμό** από τον ίδιο BGP ομιλητή
- Ο περιορισμός ρυθμού δεν ισχύει για **IBGP** ενημερώσεις, αφού εκεί θέλουμε ταχεία σύγκλιση
- Επίσης δεν ισχύει για την **απόσυρση** διαδρομών, αφού αυτό θα δημιουργούσε μαύρες τρύπες για μεγάλα χρονικά διαστήματα

Γιατί χρειάζεται περιορισμός του ρυθμού ενημέρωσης;

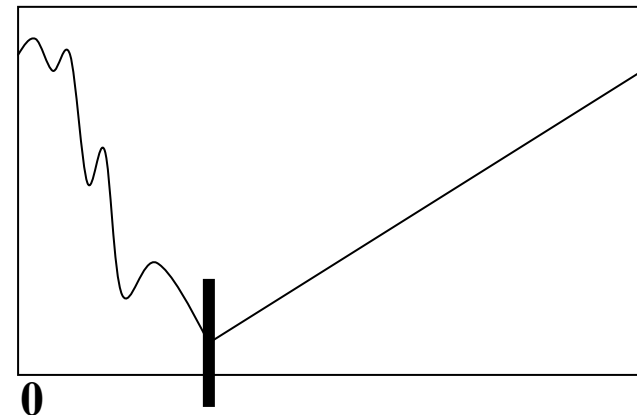
- Θέτοντας ένα **όριο** στο ρυθμό καταφέρνουμε να καταπνίξουμε μερικές ταλαντώσεις που έτσι και αλλιώς συμβαίνουν στα vectoring πρωτόκολλα
- Η τιμή των **30'** επιλέχθηκε χωρίς κανένα λόγο
- Υπάρχει μία τιμή που **βελτιστοποιεί** το χρόνο σύγκλισης για δεδομένο δίκτυο
- Όσο καλύτερο το δίκτυο τόσο πιο μεγαλύτερος θα είναι ο βέλτιστος ρυθμός

Updates
to convergence



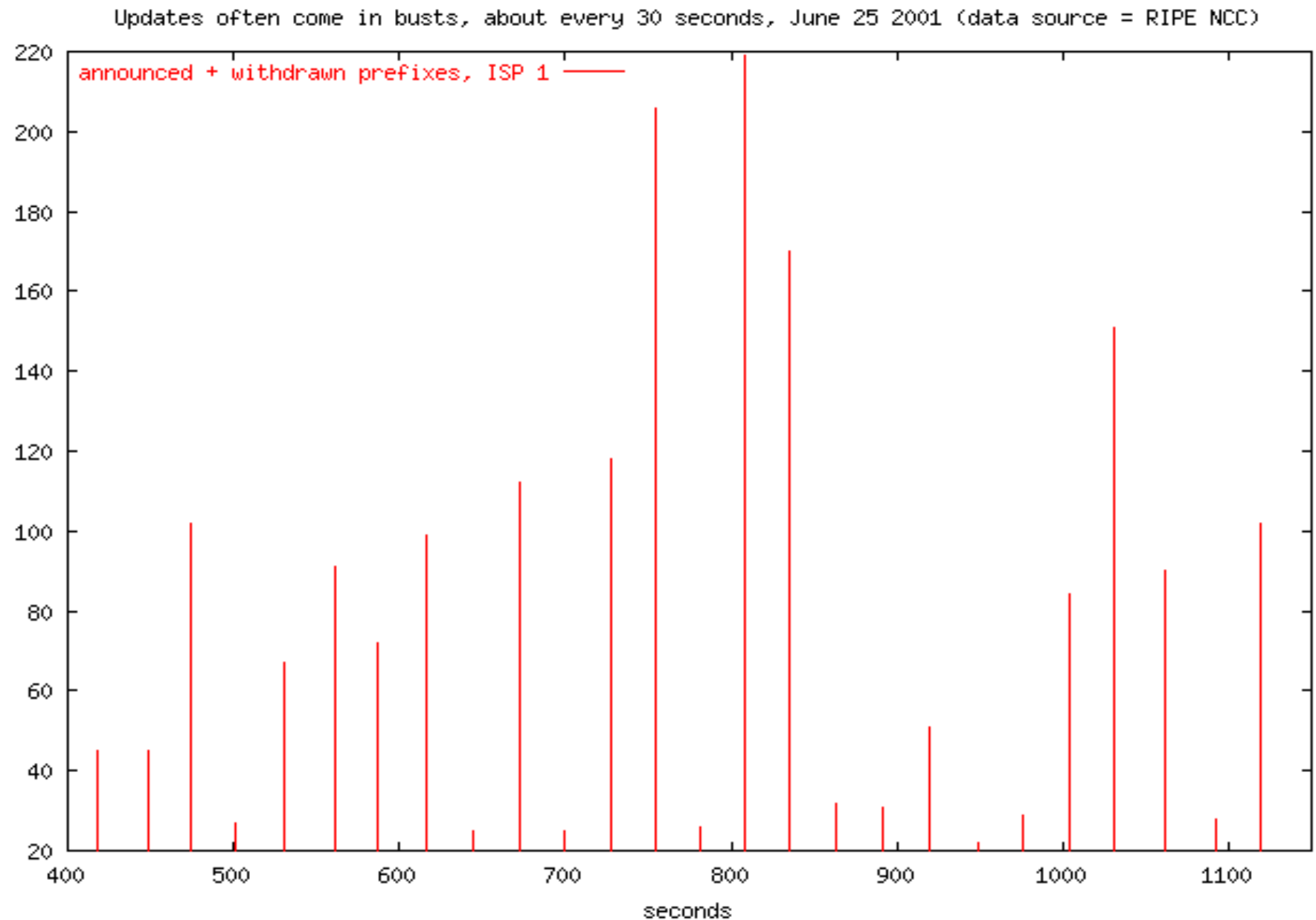
MinRouteAdvertisementInterval

Time
to convergence



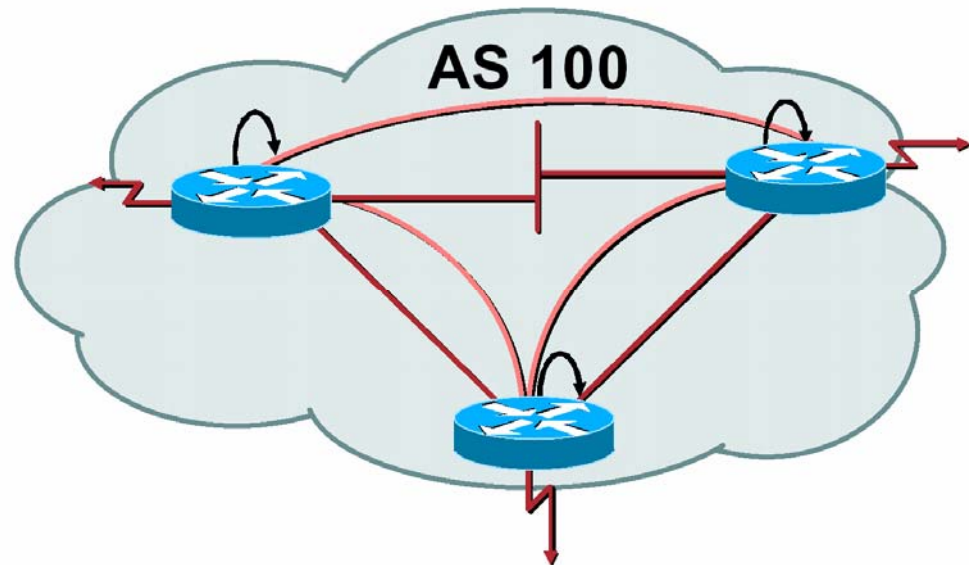
MinRouteAdvertisementInterval

Ριπές ενημέρωσης ανα 30'



Χρήση διευθύνσεων βρόχου για τις IBGP συνόδους

- Χρησιμοποιούμε σαν διεύθυνση στον **ορισμό** μιας ομότιμης σύνδεσης IBGP την **διεύθυνση βρόχου** του δρομολογητή, αντί για τη **φυσική** διεύθυνση μιας διεπαφής του
 - Η λογική αυτή διεπαφή βρίσκεται πάντοτε σε **λειτουργία** και έτσι η BGP σύνοδος να μην εξαρτάται από τη λειτουργική κατάσταση μιας μόνο **φυσικής** διεπαφής
 - Επίσης η χρήση της διεύθυνσης βρόχου είναι **πλεονεκτική** και στην περίπτωση **διαμοιρασμού** του φορτίου
-
- Στο EBGP τον περισσότερο χρόνο οι ομότιμοι δρομολογητές είναι **κατευθείαν** **συνεδεμένοι** οπότε **δεν** χρησιμοποιούμε τη διεύθυνση βρόχου



Αλληλοεπίδραση BGP/IGP

- Το BGP μεταφέρει ολόκληρο τον πίνακα δρομολόγησης του **διαδικτύου**
- Το IGP μεταφέρει την πληροφορία για τις **εσωτερικές διαδρομές** και για το NextHop
- Οι διαδρομές **ποτέ δεν αναδιανέμονται από το ένα στο άλλο**
- Αντίθετα γίνεται **αναδρομική αναζήτηση** των διαδρομών

Ομάδες ομοτίμων (Peer groups)

- Οι IBGP ομότιμοι λαμβάνουν τις **ίδιες** ενημερώσεις
- Σε δίκτυα με πολλούς IBGP ομότιμους, υπάρχει μεγάλη **καθυστέρηση** για σύγκλιση
- Χρειάζεται περισσότερη **επεξεργαστική ισχύς** για τους υπολογισμούς

Ιδέα!!

- Ομαδοποιούμε τους ομότιμους με την **ίδια τακτική** όσον αφορά τις **εξερχόμενες** διαδρομές
- Οι ενημερώσεις δημιουργούνται μόνο **μια φορά** ανά ομάδα
- Η **διάρθρωση** γίνεται πιο εύκολα, λιγότερο επισφαλής σε λάθη και πιο ευανάγνωστη
- Λιγότερη επεξεργαστική ισχύς και απαιτούμενη μνήμη
- Πιο γρήγορη σύγκλιση
- Τα μέλη μιας ομάδας μπορεί να έχουν **διαφορετική τακτική** για τις **εισερχόμενες** διαδρομές
- Συνίσταται ομαδοποίηση ανάμεσα σε δρομολογητές με κοινή εξερχόμενη τακτική δρομολόγησης
- Μπορεί να εφαρμοστεί και ανάμεσα σε EBGP γείτονες

Ομάδες ομοτίμων, Απαιτήσεις

- Όλα τα μέλη της ομάδας πρέπει να έχουν την **ίδια** τακτική δρομολόγησης όσον αφορά τις **εξερχόμενες** ανακοινώσεις (distribute-list, filter-list, route-map), εκτός από την default-originate, η οποία να ρυθμίζεται ανά ομότιμο
- Μπορούμε να ορίσουμε την τακτική δρομολόγησης για τις **εισερχόμενες** ενημερώσεις, **ανά ομότιμο**

Ομάδες ομοτίμων, Περιορισμοί

- Αν ομαδοποιούμε ομότιμους που είναι **clients** σε έναν **RR**, όλοι οι clients πρέπει να είναι σε τοπολογία **πλέγματος**
- Αν έχουμε μία ομάδα **EBGP** ομοτίμων, δεν μπορούμε να έχουμε **transit** ανάμεσα στα μέλη της ομάδας
- Όλα τα EBGP μέλη της ομάδας πρέπει να είναι στο **ίδιο υποδίκτυο** ώστε να αποφύγουμε ανακοινώσεις μη σύνδεσης με το NextHop
- Από τα IOS 11.1(18)CC, 11.3(4), 12.0 οι περιορισμοί αυτοί άρονται

Συνήθειες Ομάδες ομοτίμων

- Μία συχνά χρησιμοποιούμενη λίστα ομάδων ομοτίμων είναι :

Normal iBGP

- Κανονικοί IBGP ομότιμοι

iBGP client

- Reflection ομότιμοι ενός RR

eBGP full-routes

- Ομότιμοι που λαμβάνουν όλες τις διαδρομές

eBGP customer-routes

- Ομότιμοι που λαμβάνουν μόνο τις διαδρομές από κατευθείαν πελάτες του ISP

eBGP default-routes

- Ομότιμοι που λαμβάνουν την default διαδρομή και πιθανώς μερικές ακόμα

Επαναδιάρθρωση

- Μετά από κάθε αλλαγή στην **τακτική δρομολόγησης** (policy) χρειάζεται επανεκκίνηση των BGP συνόδων με τους ομότιμους προκειμένου να ενεργοποιηθούν οι αλλαγές
- Ο λόγος είναι ότι οι διαδρομές που απορρίπτονται δεν φυλλάσσονται στον δρομολογητή αφού αυτό θα κατανάλωνε CPU και μνήμη
- Μια τέτοια επανεκκίνηση όμως θα διάκοπτε σοβαρά την **συνδεσιμότητα** σε όλα τα δίκτυα και θα προκαλούσε μεγάλο αριθμό **ενημερώσεων** σε όλο το δίκτυο
- Γενικά η επανεκκίνηση των **BGP συνόδων** πρέπει να είναι το τελευταίο μέτρο, αφού προκαλεί μεγάλη **αναστάτωση** στο δίκτυο

Route Refresh RFC 2918

- Μία λύση στο πρόβλημα είναι η **δυνατότητα route refresh**
- Επιτρέπει αλλαγές στην **τακτική δρομολόγησης** χωρίς να διακόπτεται η **κίνηση**
- **Δεν** χρειάζεται **διάρθρωση**
- **Δεν** απαιτεί πρόσθετη **μνήμη**
- Ο ομότιμος που θέλει να επανεκκινήσει την BGP σύνοδο στέλνει ένα μήνυμα Route Refresh
- Όταν ο ομότιμος λάβει το μήνυμα στέλνει όλες τις ανακοινώσεις από την αρχή
- Γίνεται **αυτοματα διαπραγμάτευση** για τη δυνατότητα χρήσης της κατά την δημιουργία μιας σχέσης ομοτίμων
- Απαιτεί οι ομότιμοι να υποστηρίζουν την **RFC2918, Route Refresh Capability**

Route Refresh RFC 2918

- Με την **show ip bgp neighbor** μπορούμε να δούμε αν ο ομότιμος υποστηρίζει τη δυνατότητα

clear ip bgp x.x.x.x in

- Δίνουμε εντολή στον ομότιμο να ξαναστείλει όλες τις BGP ανακοινώσεις

clear ip bgp x.x.x.x out

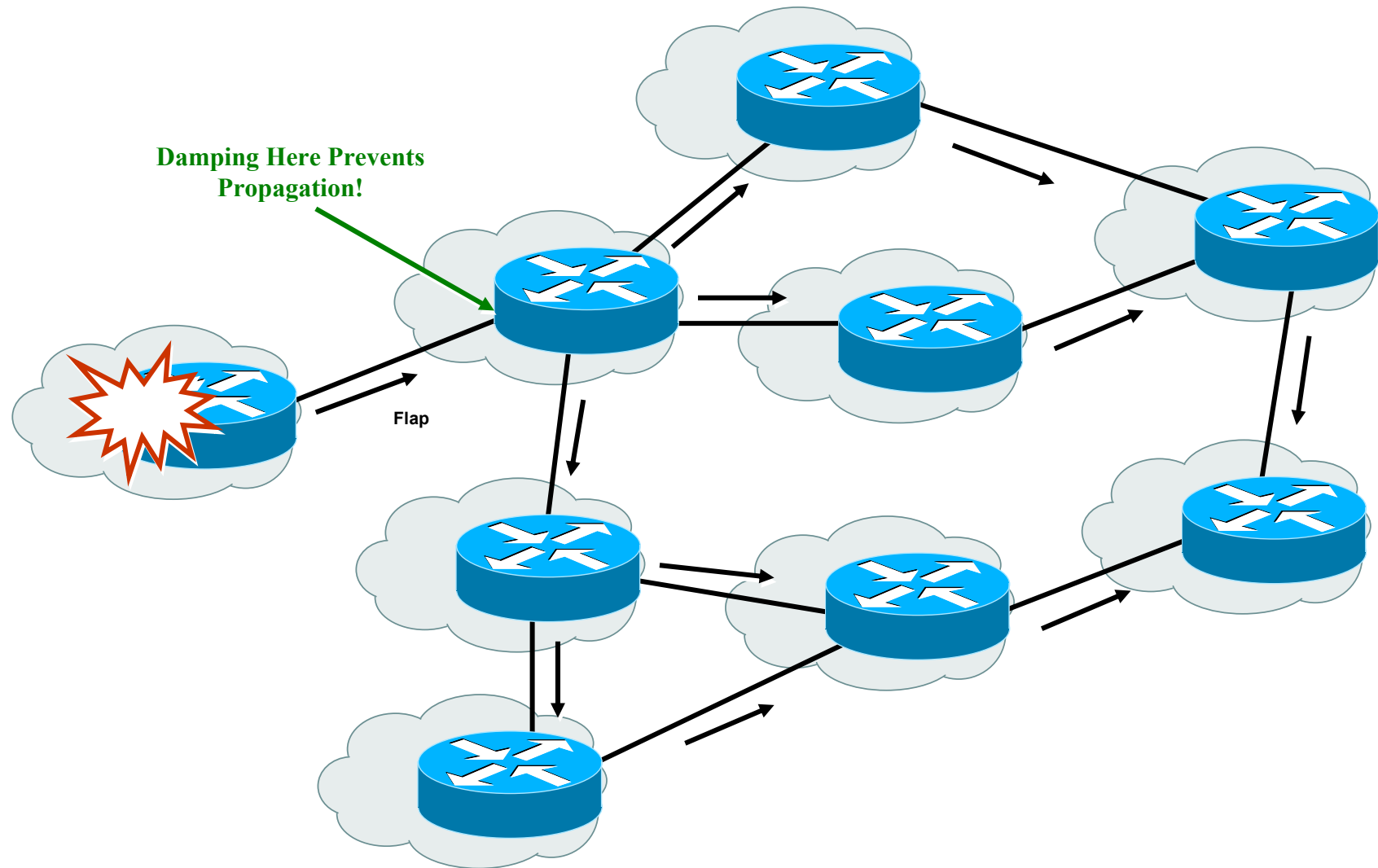
- Ξαναστέλνουμε στον ομότιμο όλες τις BGP ανακοινώσεις

.

Soft Reconfiguration

- Άλλη λειτουργία που επιτρέπει τη δυναμική αναδιάρθρωση είναι η **Soft Reconfiguration** που υποστηρίζεται από το IOS
- Όταν οι ομότιμοι δεν υποστηρίζουν την δυνατότητα RFC 2918 Route Refresh
- Με την ενεργοποίηση του SR ο δρομολογητής κρατεί στην **μνήμη** του εκτός από τα προθέματα και τις ιδιότητες τους **μετά** της εφαρμογή της τακτικής δρομολόγησης επίσης και τα προθέματα και τις ιδιότητες τους, **πριν** την εφαρμογή της τακτικής δρομολόγησης
- Έτσι **καινούργια τακτική δρομολόγησης** μπορεί να εφαρμοστεί χωρίς να απαιτείται **επανεκκίνηση των BGP συνόδων**
- Ορίζεται **ανά ομότιμο** και μόνο για τις **εισερχόμενες διαδρομές**
- Απαιτεί **περισσότερη μνήμη** για να κρατά τα προθέματα που απορρίφθηκαν ή αυτά στα οποία αλλάχτηκαν οι ιδιότητες
- Μας δίνει το πλεονέκτημα να γνωρίζουμε τα προθέματα τα οποία μας έχουν σταλεί, **εξαρχής**, πριν την εφαρμογή της τακτικής δρομολόγησης

Ταλάντωση διαδρομής, Route Flap

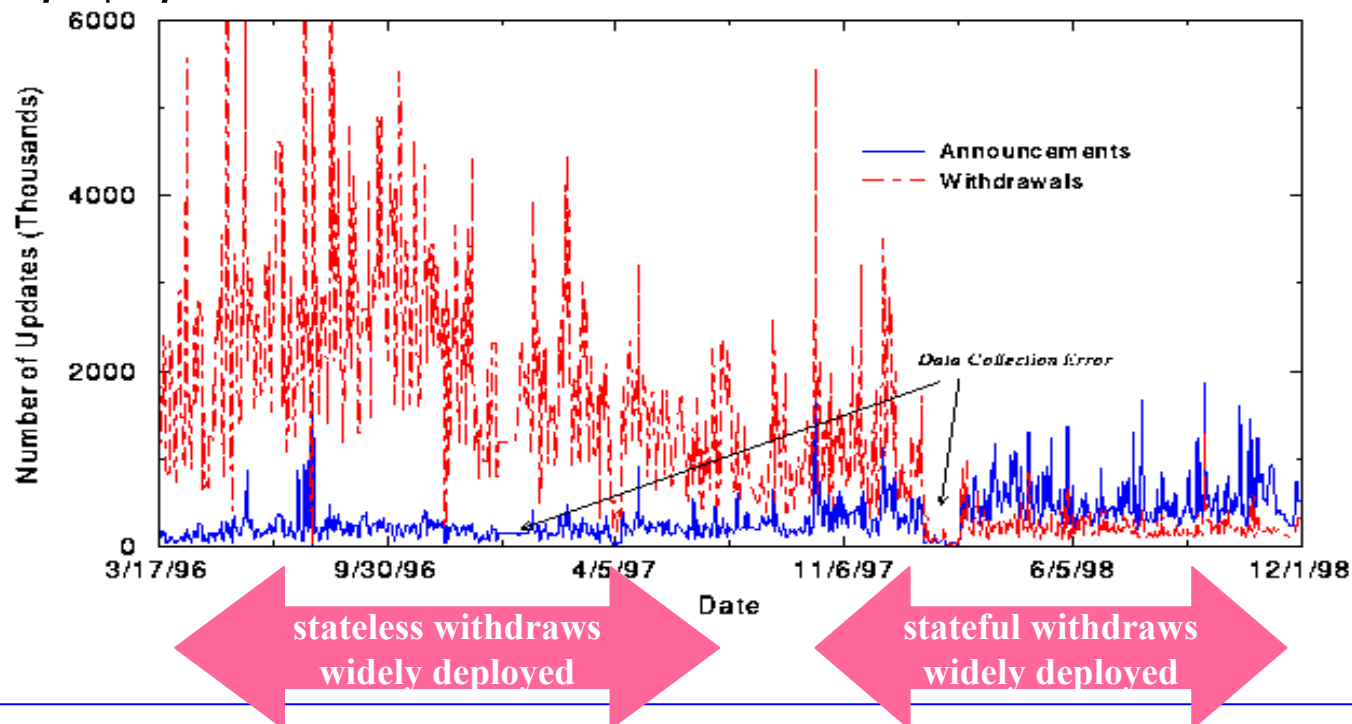


Ταλάντωση διαδρομής, Route Flap

- Ονομάζουμε το φαινόμενο όπου η διαδρομή **ανεβοκατεβαίνει** ή **αλλάζει** συνέχεια κάποια **ιδιότητα** της
- Έτσι η διαδρομή ανακοινώνεται και μετά αποσύρεται ή το αντίστροφο
- Αυτή είναι μία φυσιολογική κατάσταση, αυτός είναι και ο βασικός λόγος που έχουμε **δυναμικά** πρωτόκολλα
- Όταν όμως γίνεται με μεγάλο **ρυθμό** συνήθως δείχνει κάποιο πρόβλημα όπως κάποια σύνδεση που ανεβοκατεβαίνει ή κάποια σοβαρά προβλήματα ανάμεσα σε ομότιμους

Απόσβεση ταλαντώσεων διαδρομών, Route Flap Damping, RFC 2439

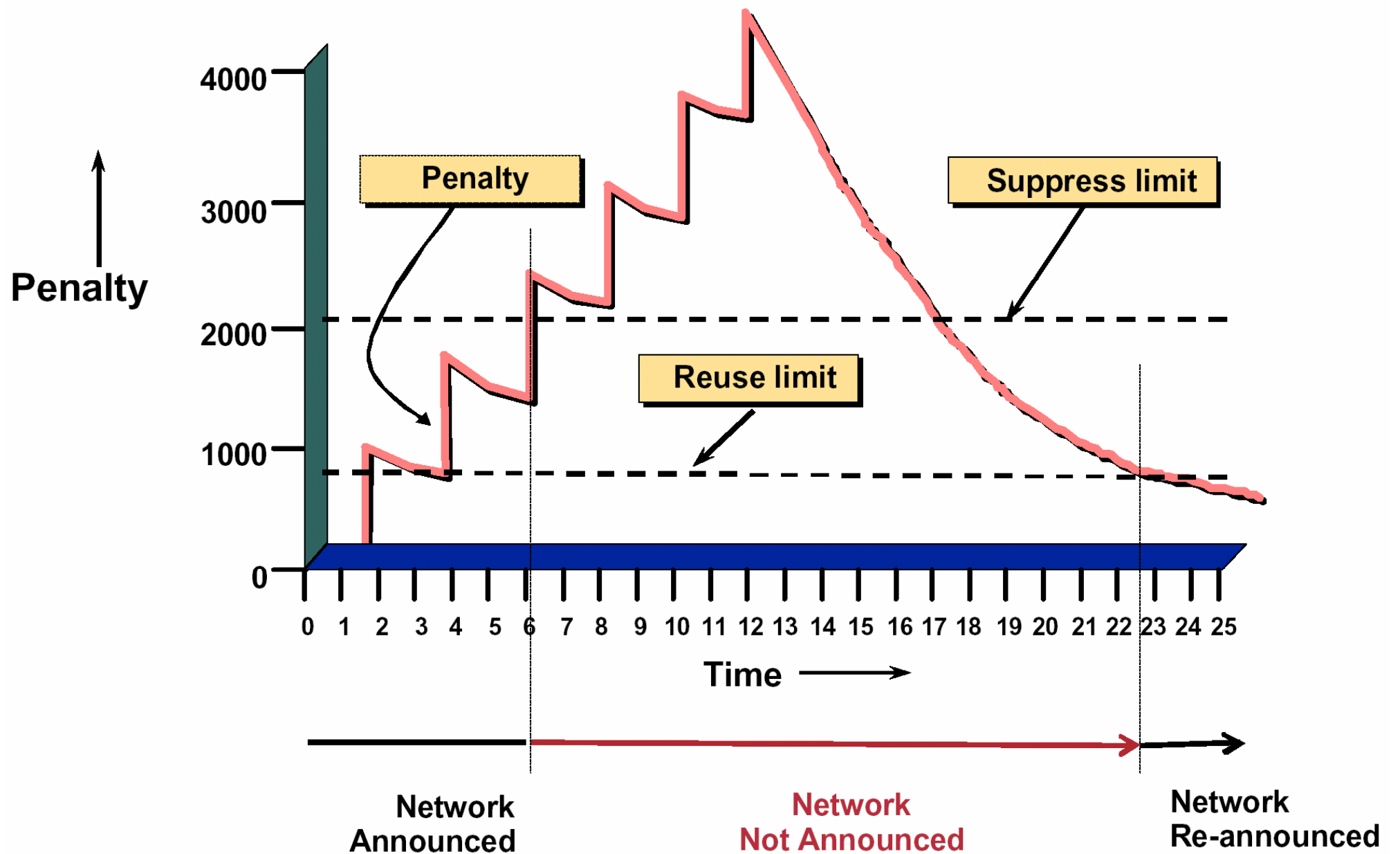
- **Κατασταλτικό** μέτρο αντιμετώπισης της **διάδοσης** των **ταλαντώσεων** των διαδρομών στο δίκτυο, καταστέλοντας τις ανακοινώσεις
- Ο στόχος είναι η **ελαχιστοποίηση** της **αστάθειας** που προκαλείται από τις ταλαντώσεις των διαδρομών και η σταθεροποίηση του δικτύου
- Η ιδέα είναι ότι η **παρελθοντική** συμπεριφορά μας οδηγεί στην πρόγνωση της μελλοντικής άρα αρκεί να τιμωρήσουμε αυτούς που έχουν επιδείξει **κακή συμπεριφορά**



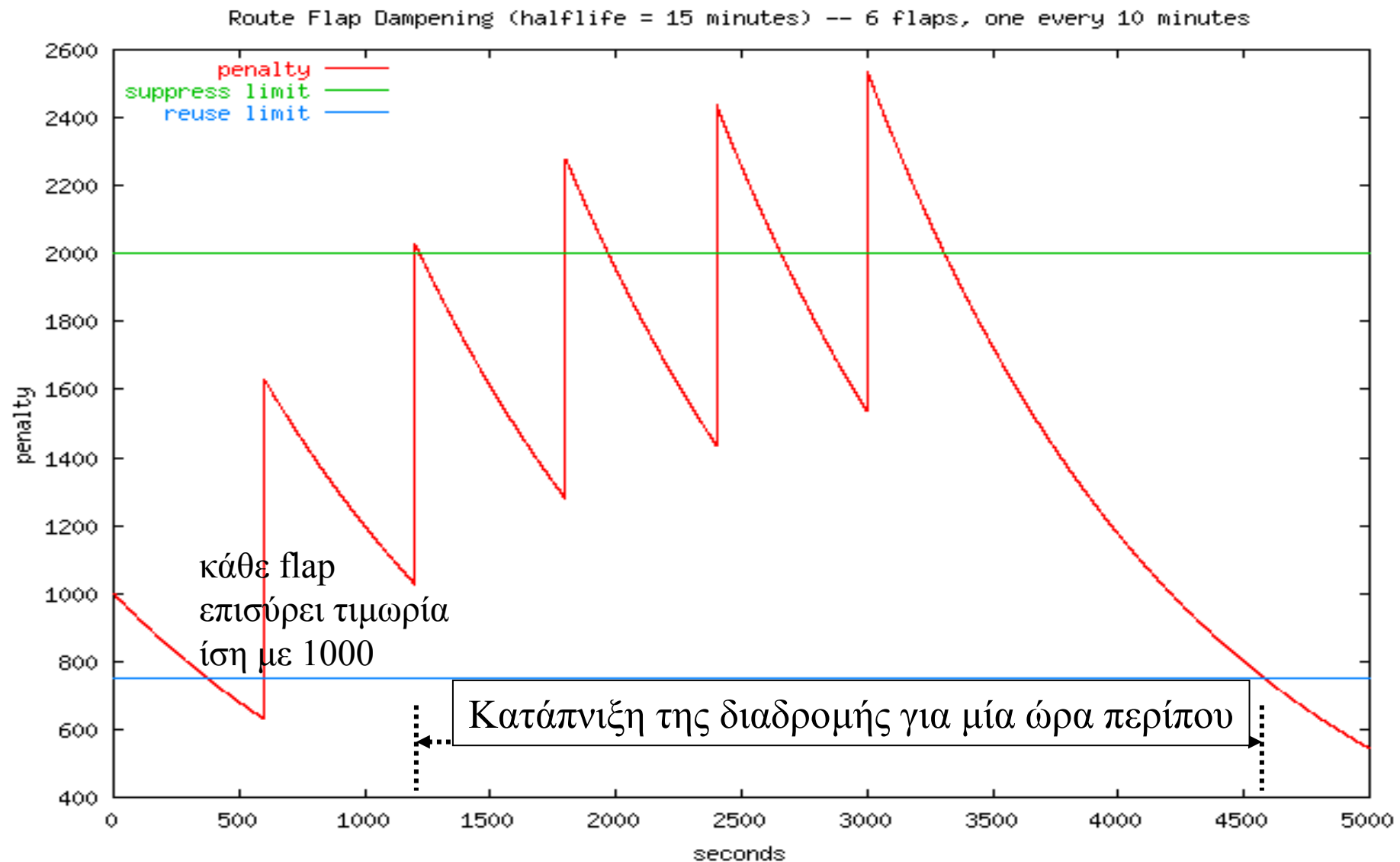
Route Flap Damping, RFC 2439

- Όταν ενεργοποιήσουμε τον μηχανισμό κρατείται μία μετρική ανά **ομότιμο** ανά **διαδρομή** η οποία ονομάζεται **τιμωρία (penalty)** και είναι βαθμός αστάθειας της διαδρομής
- Οι διαδρομές **τιμωρούνται** για κάθε αλλαγή
 - Κάθε αλλαγή κατάστασης (απόσυρση διαδρομής) επισύρει τιμωρία ίση με 1000
 - Κάθε αλλαγή σε μια ιδιότητα παίρνει τιμωρία ίση με 500
- Αν η τιμωρία ξεπεράσει το **όριο καταπίεσης**, τότε η διαδρομή καταπνίγεται, δηλαδή δεν γίνεται διαφήμιση της διαδρομής στους BGP ομότιμους
- Όταν η διαδρομή δεν αλλάζει η τιμωρία μειώνεται με **εκθετικό** ρυθμό (Ο **χρόνος ημίσειας ζωής** καθορίζει το ρυθμός μείωσης)
- Όταν η τιμωρία μειωθεί κάτω από το **όριο επαναχρησιμοποίησης**, τότε ανακοινώνεται πάλι στους BGP ομότιμους
- Η πληροφορία σχετικά με την τιμωρία **σβήνεται** χαριστικά όταν φτάσει το μισό του ορίου επαναχρησιμοποίησης

Route Flap Damping, RFC 2439



Route Flap Damping, RFC 2439



Route Flap Damping, RFC 2439

- Χρησιμοποιείται μόνο ανάμεσα σε **EBGP** ομότιμους (εφαρμόζεται μόνο για τις **εισερχόμενες EBGP** ενημερώσεις) ώστε οι ταλαντώσεις να καταπιεστούν όσο το δυνατό **πιο κοντά στην πηγή της αστάθειας**
- Οι διαδρομές εξωτερικές του AS, που έχουν γίνει γνωστές με **IBGP** δεν θα καταπνιχθούν, έτσι ώστε να αποφύγουμε την περίπτωση να έχουν οι IBGP ομότιμοι μεγαλύτερη τιμωρία για εξωτερικές στο AS διαδρομές
- Η τιμωρία μειώνεται σε βήμα 5" και οι διαδρομές επανέρχονται σε βήμα 10"

Route Flap Damping, RFC 2439

- Μπορεί να μειώσει με δραματικό τρόπο των **αριθμό των ενημερώσεων**
- Καταπιέζει τις διαδρομές που **ταλαντώνουν**
- Διαφήμιση μόνο **σταθερών** διαδρομών
- Απαιτεί επιπρόσθετους **πόρους** στον δρομολογητή
- Κάποιος βαθμός ταλάντωσης πρέπει να γίνεται αποδεκτός αφού αντικατοπτρίζει κανονικές μεταβολές στο δίκτυο
- Για παράδειγμα μια μεταβολή σε μερικά λεπτά δεν είναι απαραίτητα ένα πρόβλημα, αλλά πέντε μεταβολές σε μερικά λεπτά μπορεί να είναι
- Πρέπει να μην θυσιάζεται ο **χρόνος σύγκλισης** για διαδρομές που συμπεριφέρονται καλά

Παράμετροι λειτουργίας

bgp dampening [<half-life> <reuse-value> <suppresspenalty> <maximum suppress time>]

half-life-time

- Ο χρόνος ημίσειας ζωής (τυπικά 15'), 1-45'

reuse-value

- Το όριο επαναχρησιμοποίησης (τυπικά 750), 1-20000
- Όταν η τιμωρία είναι πάνω από αυτό το όριο τότε η διαδρομή ανακοινώνεται πάλι

suppress-value

Το όριο καταπίεσης (τυπικά 2000), 1-20000

Όταν η τιμωρία είναι πάνω από αυτό το όριο τότε η διαδρομή καταπιέζεται

max-suppress-time

- Μέγιστος χρόνος κατάπνιξης (τυπικά 60' 4 φορές ο half-life), 1-255
- Εξασφαλίζει ότι μία διαδρομή που ταλαντώνει πολύ για μικρό χρονικό διάστημα, αλλά μετά σταθεροποιείται θα καταπιεστεί μόνο το πολύ 60'

Παράμετροι λειτουργίας

- Οι παράμετροι πρέπει να επιλέγονται με **προσοχή**
- Ο **μέγιστος χρόνος καταπίεσης** και ο χρόνος **ημίσειας ζωής** πρέπει να επιτρέπουν η **τιμωρία** να είναι μεγαλύτερη από το **όριο καταπίεσης**
- Η **μέγιστη τιμή** της ποινής είναι:

$$\text{max-penalty} = \text{reuse-limit} \times 2^{\left(\frac{\text{max-suppress-time}}{\text{half-life}} \right)}$$

- Το **όριο καταπίεσης** πρέπει να είναι μικρότερο από τη **μέγιστη ποινή**
διαφορετικά δεν θα υπάρχει απόσβεση των ταλαντώσεων

Παράμετροι λειτουργίας

Παράδειγμα

bgp dampening [<half-life> <reuse-value> <suppresspenalty> <maximum suppress time>]

bgp dampening 30 750 3000 60

•Όριο επαναχρησιμοποίησης 750 σημαίνει ότι η μέγιστη πιθανή τιμωρία είναι 3000

•Έτσι τα προθέματα δεν θα καταπιεστούν αφού η τιμωρία δεν μπορεί να υπερβεί το όριο καταπίεσης

Παράδειγμα

bgp dampening [<half-life> <reuse-value> <suppresspenalty> <maximum suppress time>]

bgp dampening 30 2000 3000 60

•Όριο επαναχρησιμοποίησης 2000 σημαίνει ότι η μέγιστη πιθανή ποινή είναι 8000

•Το όριο καταπίεσης εύκολα προσεγγίζεται

Παράμετροι λειτουργίας, εξ' ορισμού τιμές

- Οι παράμετροι πρέπει να επιλεγούν με **προσοχή**
- Οι Cisco/Juniper τιμωρούν όλες τις αλλαγές διαδρομών
- Οι εξ' ορισμού ρυθμίσεις και αυτές που προτείνονται από τη σύσταση **RIPE-229** είναι πολύ **επιθετικές** – αυστηρές

flap 3 times in 15 minutes and you are damped

- Ρύθμιση παραμέτρων**
- Μείωση του κατωφλίου καταπίεσης
- Μείωση της τιμωρίας σε κάθε αλλαγή, ιδιαίτερα για τις αλλαγές στην τιμή των ιδιοτήτων
- Πιθανώς **απενεργοποίηση** του μηχανισμού

Parameter	Cisco	Juniper
Withdrawal penalty	1000	1000
Re-advertisement penalty	0	1000
Attributes change penalty	500	500
Suppress threshold	2000	3000
Half-life (min)	15	15
Reuse threshold	750	750
Max suppress time (min)	60	60

Ρύθμιση της λειτουργίας

- Η **RIPE** έχει δημοσιεύσει συστάσεις για τις κατάλληλες παραμέτρους
- Συστήνει ότι η διαδικασία θα πρέπει να αρχίσει μετά από 4 συνεχόμενες ταλαντώσεις

Μήκος προθέματος

- Το κόστος είναι πολύ πιο σημαντικό από την κατάπνιξη ενός κοντού προθέματος/μεγάλου δικτύου

</21 MaxSuppressTime=30 minutes

/22 /23 MaxSuppressTime=45 minutes

/24 MaxSuppressTime=60 minutes

Μερικά σημαντικά προθέματα δεν καταπνίγονται ποτέ, όπως οι DNS root servers ([golden networks](#))

Τύπος ομοτίμου

- Μικροί χρόνοι για διαδρομές **πελατών**, μεγαλύτερες για άλλους, διότι είναι οι πελάτες μας

- Ο κάθε τύπος δικτύου θέλει τις δικές του τιμές, κάποιες φορές έχουμε **καλύτερο δίκτυο χωρίς την ενεργοποίηση της λειτουργίας**

Καταστάσεις διαδρομής

sh ip bgp

- Ο πίνακας δρομολόγησης του BGP

Κωδικοί κατάστασης εγγαφών

s suppressed

d damped

h history

* valid

> Best

I μαθεύτηκε από IBGP ομότιμο

i IGP

e EGP

? incomplete

Καταστάσεις διαδρομής

History (h)

- Το πρόθεμα έχει **αποσυρθεί** και είναι τώρα **μη προσεγγίσιμο**, έχει οριστεί τιμωρία

Damped (d)

- Το πρόθεμα είναι **προσεγγίσιμο** και **χρησιμοποιείται** και **ανακοινώνεται** αλλά έχει τιμωρία **κάτω** από όριο καταπίεσης

Suppressed (s)

- Το πρόθεμα είναι **προσεγγίσιμο** αλλά έχει τιμωρία **μεγαλύτερη** από το όριο καταπίεσης και άρα **δεν ανακοινώνεται** και **δεν εγκαθίσταται** στην **FIB**

Παράδειγμα 1

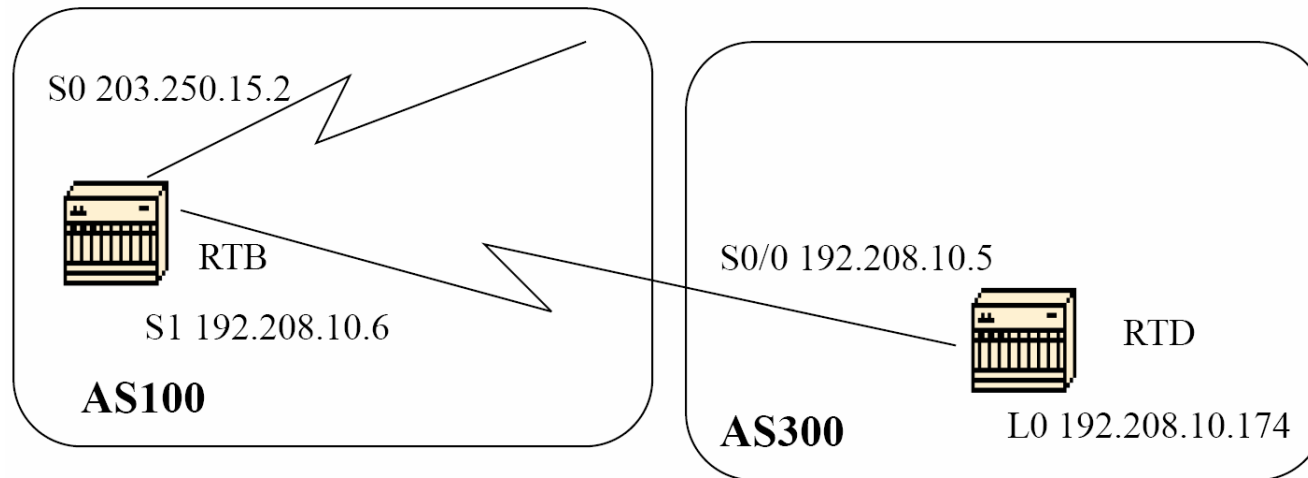
Time 8/10	A/W	ASPath	Penalty
18:00:15	A	9177 3320 1 2914 3130 3927	500
18:00:41	A	9177 6730 5400 2914 3130 3927	990
18:01:41	A	9177 3320 2914 3130 3927	1445
18:03:06	A	9177 3320 1239 2914 3130 3927	1853
18:03:35	W		2812
18:04:03	A	9177 6730 5400 2914 3130 3927	2752
18:04:31	W		3694

Above suppress threshold

Time 8/20	A/W	ASPath	Penalty
01:00:16	A	3257 1299 2914 3130 3927	500
01:00:47	A	3257 1299 2914 3130 3927	988
01:00:50	W		1985
01:00:50	A	3257 1299 4200 2914 3130 3927	1985
01:01:13	A	3257 1299 701 2914 3130 3927	2451
01:02:05	W		3354

Above suppress threshold

Παράδειγμα 2



```
hostname RTB
interface Serial0
ip address 203.250.15.2 255.255.255.252
interface Serial1
ip address 192.208.10.6 255.255.255.252
router bgp 100
bgp dampening
network 203.250.15.0
neighbor 192.208.10.5 remote-as 300
```

```
hostname RTD
interface Loopback0
ip address 192.208.10.174 255.255.255.192
interface Serial0/0
ip address 192.208.10.5 255.255.255.252
router bgp 300
network 192.208.10.0
neighbor 192.208.10.6 remote-as 100
```

Παράδειγμα 2

Σταθερή διαδρομή

- Δύο εγγραφές μία από κατευθείαν συνεδεδεμένο δίκτυο και μία που διαφημίζεται από το AS300

sh ip bgp

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 192.208.10.0	192.208.10.5	0		0	300 i
*> 203.250.15.0	0.0.0.0	0		32768	i

Παράδειγμα 2

Κομμένη διαδρομή, αλλά κάτω από το όριο καταπίεσης

• Η διαδρομή μπαίνει σε κατάσταση **history**, δηλ. δεν έχουμε διαδρομή αλλά η πληροφορία για το route flapping κρατείται

sh ip bgp

Network	Next Hop	Metric	LocPrf	Weight	Path
h 192.208.10.0	192.208.10.5	0		0	300 i
*> 203.250.15.0	0.0.0.0	0		32768	i

sh ip bgp 192.208.10.0

BGP routing table entry for 192.208.10.0 255.255.255.0, version 25

Paths: (1 available, no best path)

300 (history entry)

192.208.10.5 from 192.208.10.5 (192.208.10.174)

Origin IGP, metric 0, external

Dampinfo: penalty 910, flapped 1 times in 0:02:03

The route has been given a penalty for flapping but the penalty is still below the “suppress limit” (default is 2000). The route is not yet suppressed.

Παράδειγμα 2

Διαδρομή πάνω από το όριο καταπίεσης

sh ip bgp

Network	Next Hop	Metric	LocPrf	Weight	Path
*d 192.208.10.0	192.208.10.5	0		0	300 i
*> 203.250.15.0	0.0.0.0	0		32768	i

sh ip bgp 192.208.10.0

BGP routing table entry for 192.208.10.0 255.255.255.0, version 32

Paths: (1 available, no best path)

300, (suppressed due to dampening)

192.208.10.5 from 192.208.10.5 (192.208.10.174)

Origin IGP, metric 0, valid, external

Dampinfo: penalty 2615, flapped 3 times in 0:05:18 , reuse in 0:27:00

- Η διαδρομή ανεβοκατέβηκε 3 φορές, προκαλώντας 7 ενημερώσεις, αλλαγές της έκδοσης του πίνακα
- Η τιμωρία έφτασε τα 2000 άρα η διαδρομή **καταπιέζεται**
- Η διαδρομή θα χρησιμοποιηθεί όταν η τιμωρία φτάσει το όριο επαναχρησιμοποίησης, 750
- Η πληροφορία για το dampening θα σβήσει όταν η τιμωρία γίνει μικρότερη από $750/2=375$

Παράδειγμα 3

***d** 10.14.141.128/29 10.21.123.230 **129 05:27:31 00:55:34** 121 941 2801 ?

- Η διαδρομή έχει ανεβοκατέβει **129** φορές σε χρόνο **5:27:31**
- Η διαδρομή είναι damped και σε **55:34** θα ξαναεμφανιστεί στον πίνακα δροπμολόγησης και θα ανακοινωθεί στους ομότιμους, με την προθπόθεση ότι δεν θα ανεβοκατεύει πάλι

h 10.45.165.0/26 10.21.123.230 **85 05:28:55** 121 941 191 ?

- Η διαδρομή αφού ανεβοκατέβηκε **85** φορές σε **5:28** δεν εμφανίστηκε ξανά οπότε παραμένει σαν history route

Εντολές για τα στατιστικά

show ip bgp flap-statistics

- displays flap statistics for all the paths

show ip bgp-flap-statistics regexp <regexp>

- displays flap statistics for all paths that match the regexp

show ip bgp flap-statistics filter-list <list>

- (displays flap statistics for all paths that pass the filter)

show ip bgp flap-statistics A.B.C.D m.m.m.m

- (displays flap statistics for a single entry)

show ip bgp flap-statistics A.B.C.D m.m.m.m longer-prefixes

- (displays flap statistics for more specific entries)

show ip bgp neighbor [dampened-routes] | [flap-statistics]

- (displays flap statistics for all paths from a neighbor)

clear ip bgp flap-statistics

- clears flap statistics for all routes

clear ip bgp flap-statistics regexp <regexp>

- (clears flap statistics for all the paths that match the regexp)

clear ip bgp flap-statistics filter-list <list>

- (clears flap statistics for all the paths that pass the filter)

clear ip bgp flap-statistics A.B.C.D m.m.m.m

- (clears flap statistics for a single entry)

clear ip bgp A.B.C.D flap-statistics

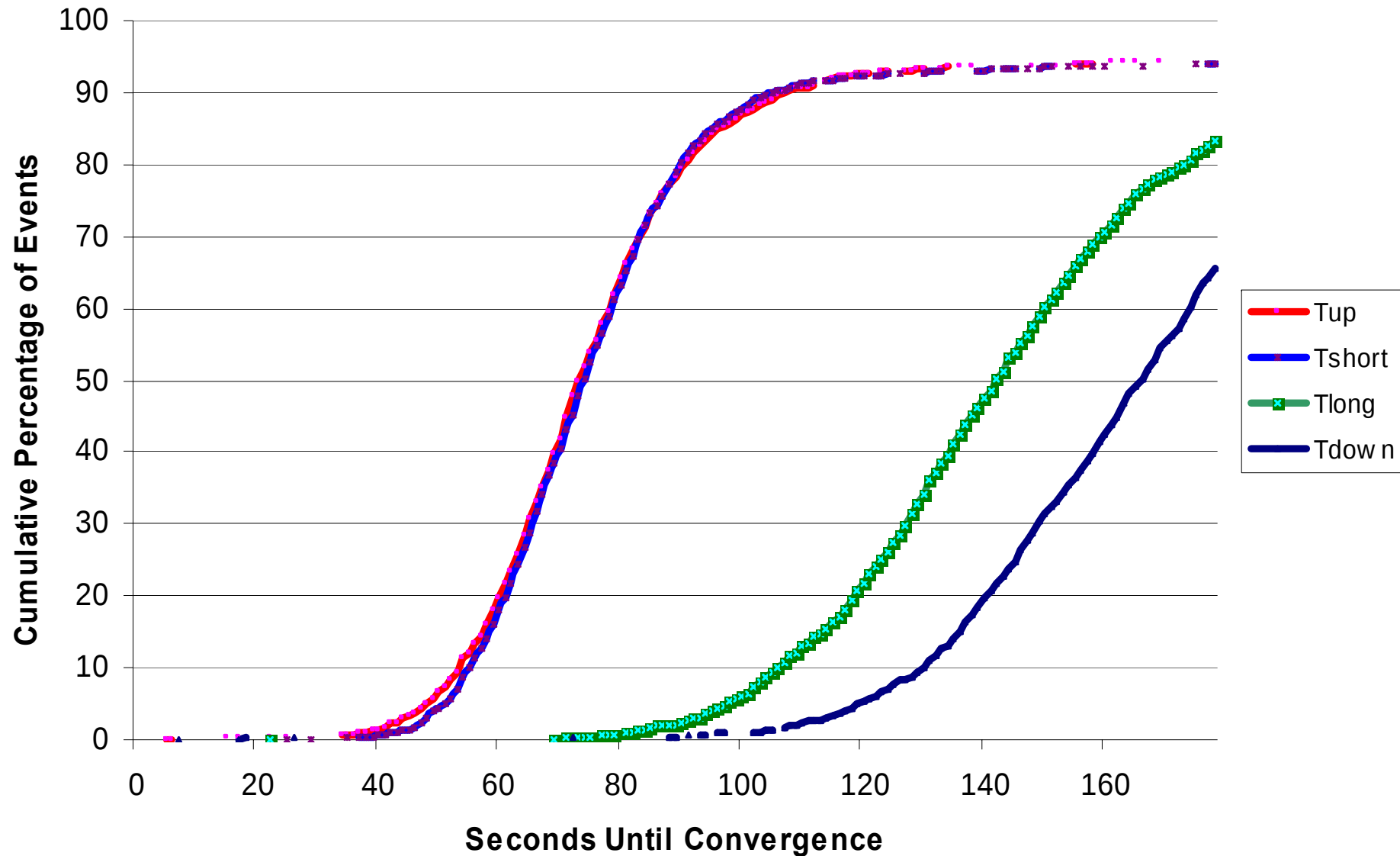
- (clears flap statistics for all paths from a neighbor)

BGP Beacons (Φάροι)

- **Ανακοινώσεις** και **αποσύρσεις** διαδρομών ανά **σταθερά** διαστήματα δύο ωρών
- Αποσκοπούν σε δοκιμές ευστάθειας του δικτύου
- Η πρώτη ενημέρωση 3:00AM GMT
- Η πρώτη απόσυρση 1:00AM GMT

Prefix	Source AS	Start date	Upstream Providers	Beacon Host
198.133.206.0/24	3927	8/10/02	AS2914, AS1	Randy Bush
192.135.183.0/24	5637	9/4/02	AS3701, AS2914	David Meyer
203.10.63.0/24	1221	9/25/02	AS1221	Geoff Huston
198.32.7.0/24	3944	10/24/02	AS2914, AS8001	Andrew Partan

Χρόνος σύγκλισης (πόσο παίρνει για την προσαρμογή σε αλλαγές;)



Γιατί έχουμε καθυστερημένη σύγκλιση;

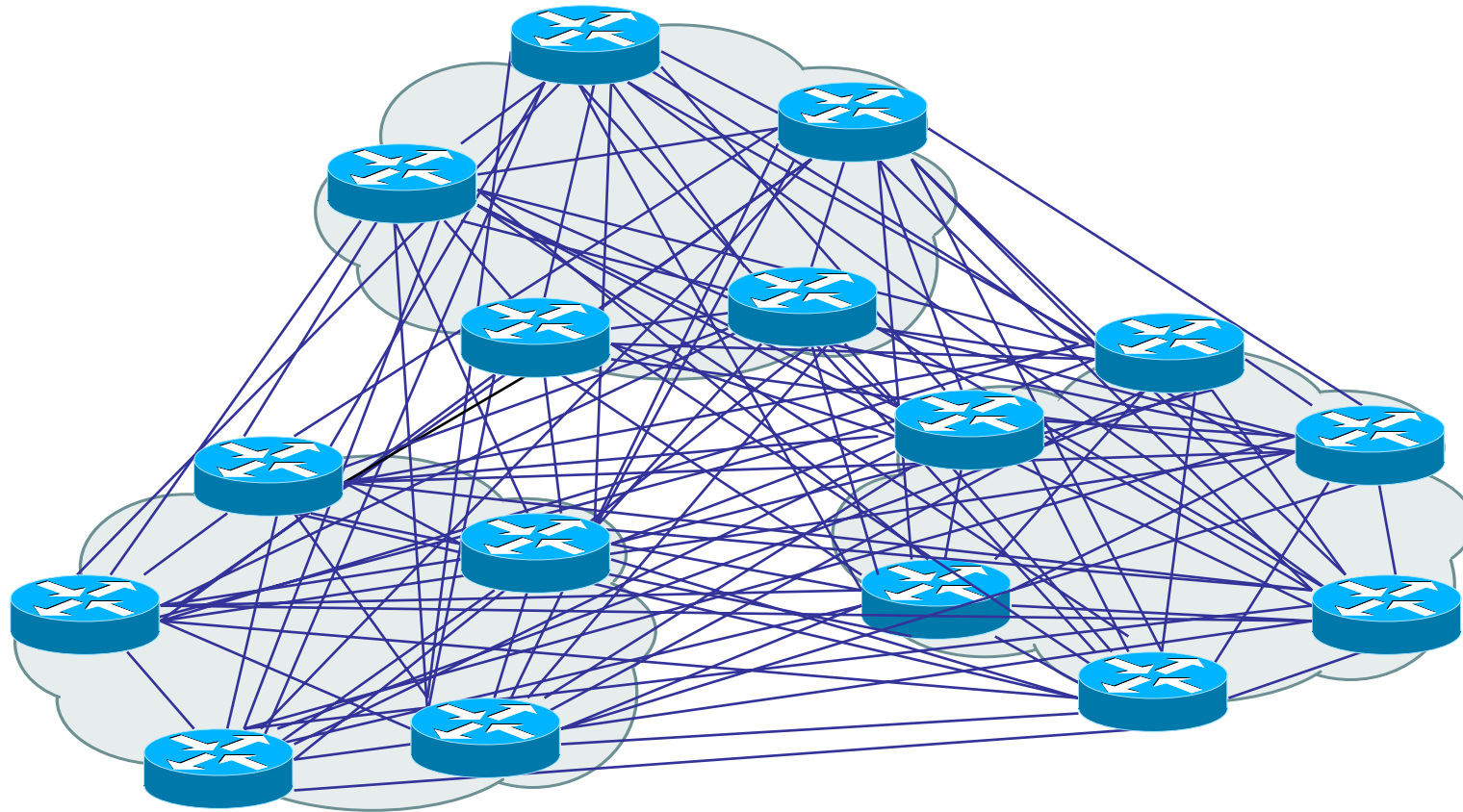
- Ο περιορισμός του **ρυθμού ενημέρωσης** καθυστερεί τη διάδοση των ενημερώσεων
- Καθυστέρηση στη **διάδοση** των διαδρομών
- Χρόνος επεξεργασίας στο δρομολογητή
- Το πρωτόκολλο μπορεί να εξερευνά πολλές **εναλλακτικές** διαδρομές πριν να παραιτηθεί ή πριν την άφιξη μιας νέας διαδρομής (δεν υπάρχει **συνολική γνώση του δικτύου** στα πρωτόκολλα τύπου vectoring)

Τελευταίο αλλά και πιο σημαντικό ...

- Είναι γνωστό ότι τα **distance vector** πρωτόκολλα παρουσιάζουν **αργή σύγκλιση**
- Η μικρή ταχύτητα σύγκλισης είναι **αιτία αστάθειας**
- Το BGP παράγει μεγάλο ποσό ασταθούς πληροφορίας δρομολόγησης, μέχρι τη στιγμή που θα συγκλίνει
- Η απόσβεση της ταλάντωσης των διαδρομών είναι ένα **κατασταλτικό** μέτρο, δεν αποτελεί λύση του προβλήματος
- Είναι ένα μέτρο που λαμβάνουμε εναντίον κάποιων που μας δημιουργούν πρόβλημα ή είναι ένα μέτρο που λαμβάνουν άλλοι εναντίον μας γιατί δημιουργούμε πρόβλημα
- Η λύση θα ήταν να σταματήσουν/σταματήσουμε να δημιουργούμε το πρόβλημα
- Λύσεις είναι η εφαρμογή **σταθερών** συνδέσεων, η σωστή **συνάθροιση** διαδρομών, η εφαρμογή τακτικής δρομολόγησης, κ.α.
- **Τελευταίο** μέτρο και εφόσον δεν λυθεί εντελώς το πρόβλημα, είναι η **κατάπνιξη** των διαδρομών

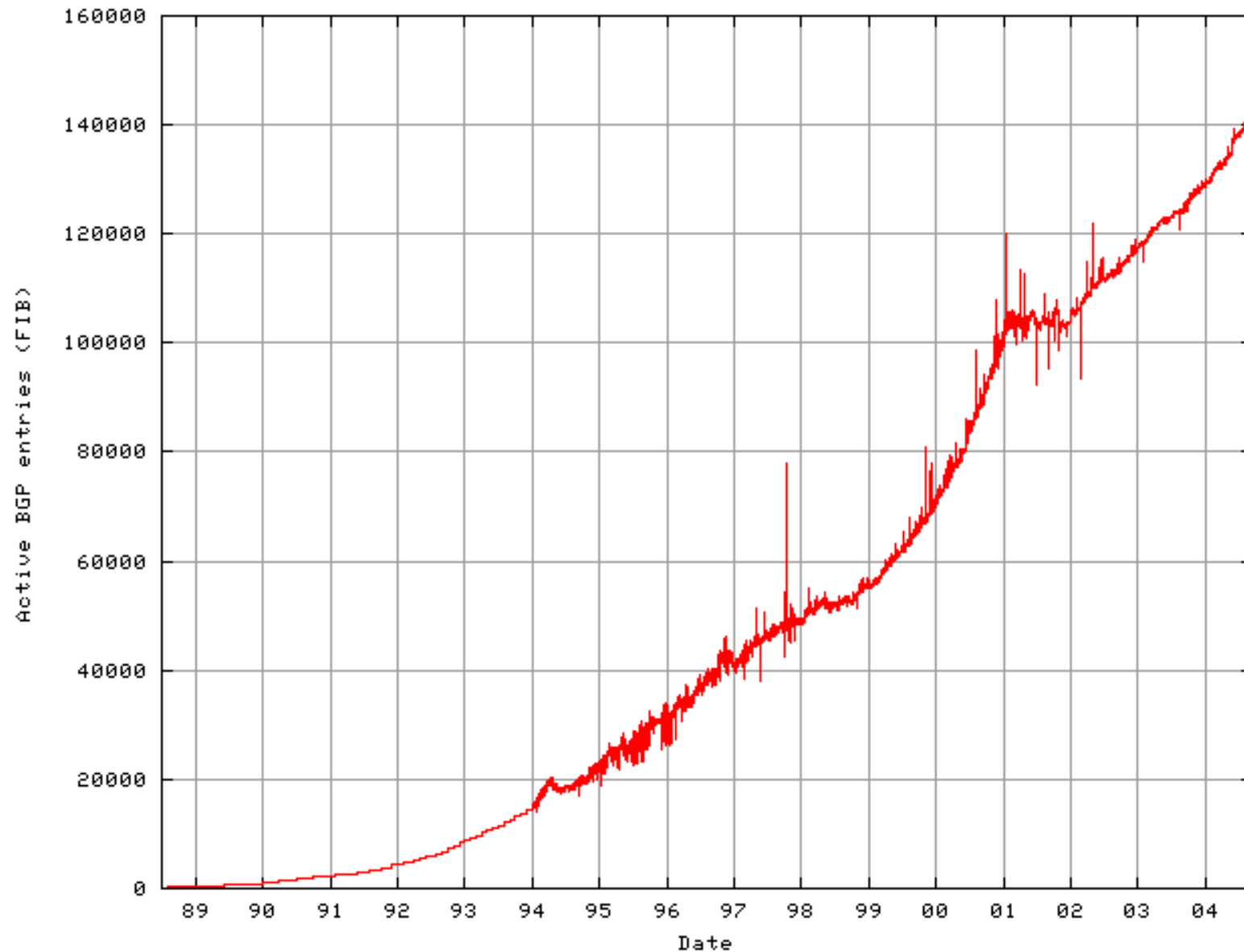
Κλιμάκωση

Κλιμάκωση δικτύου



•Τι κάνουμε όταν **μεγαλώνει** το δίκτυο;

Όσο κλιμακώνεται ένα δίκτυο, τόσο μεγαλώνουν οι πίνακες δρομολόγησης



- CPU
- Μνήμη
- Αναβαθμίσεις
- Υπερωρίες
- Έξοδα..

Οι μεγάλοι πίνακες δρομολόγησης δημιουργούν προβλήματα

- Στους πίνακες κρατούνται οι **καλύτερες** διαδρομές αλλά και οι **εναλλακτικές**
- Το μέγεθος μπορεί να είναι μεγάλο για δρομολογητές με πολλές **εναλλακτικές** διαδρομές, όπως οι RR (Route Reflectors)
- Προκαλείται αύξηση στο φόρτο της **CPU**, ιδιαίτερα κατά την επανεκκίνηση της συνόδου
- Μπορεί να λυθεί με περισσότερη **επεξεργαστική ισχύ**, αλλά αυτό συνεπάγεται μεγαλύτερο κόστος για **αναβαθμίσεις**

Η τοπολογία mesh του IBGP δεν κλιμακώνεται

2 → 1

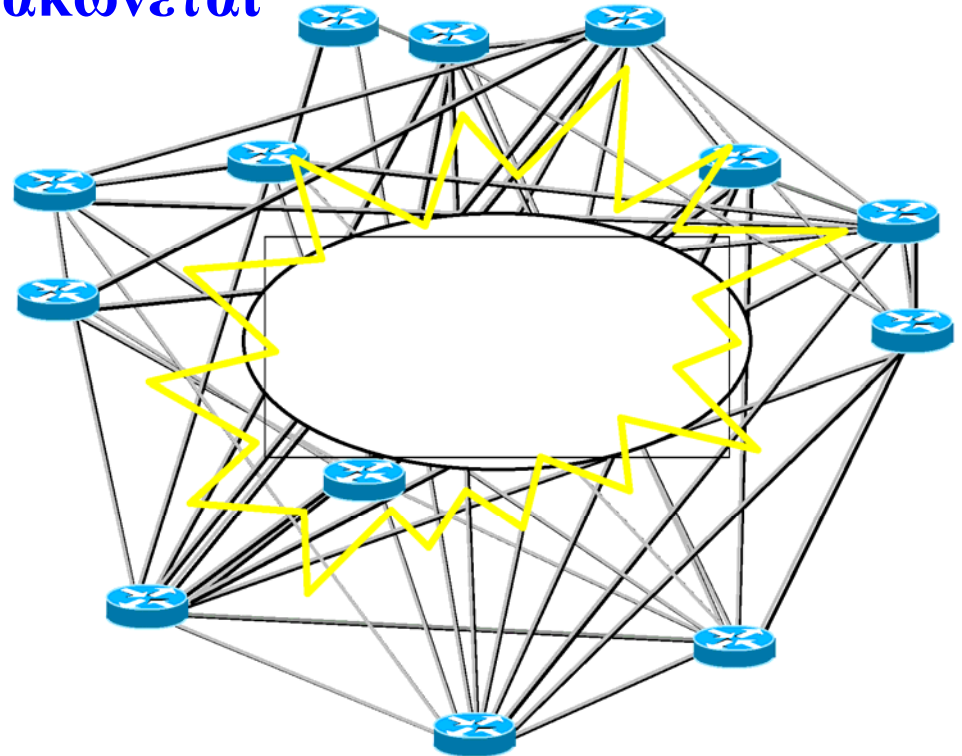
3 → 3

4 → 6

10 → 45

20 → 190

1000 → 500000



- Στο IBGP ο κάθε ομότιμος δεν μπορεί να βάλει το δικό του ASN στη AS_path λίστα
- Άρα τυχόν ανακοίνωση διαδρομών IBGP σε IBGP ομότιμο μπορεί να δημιουργήσει βρόχο και έτσι οι **IBGP** ομότιμοι δεν αναμεταδίδουν τις ανανεώσεις που έλαβαν από IBGP ομότιμο
- Άρα πρέπει να υπάρχει συνδεσιμότητα **όλων με όλους (πλέγμα)**
- $N(N-1)/2$ συνδέσεις ανάμεσα σε ομότιμους

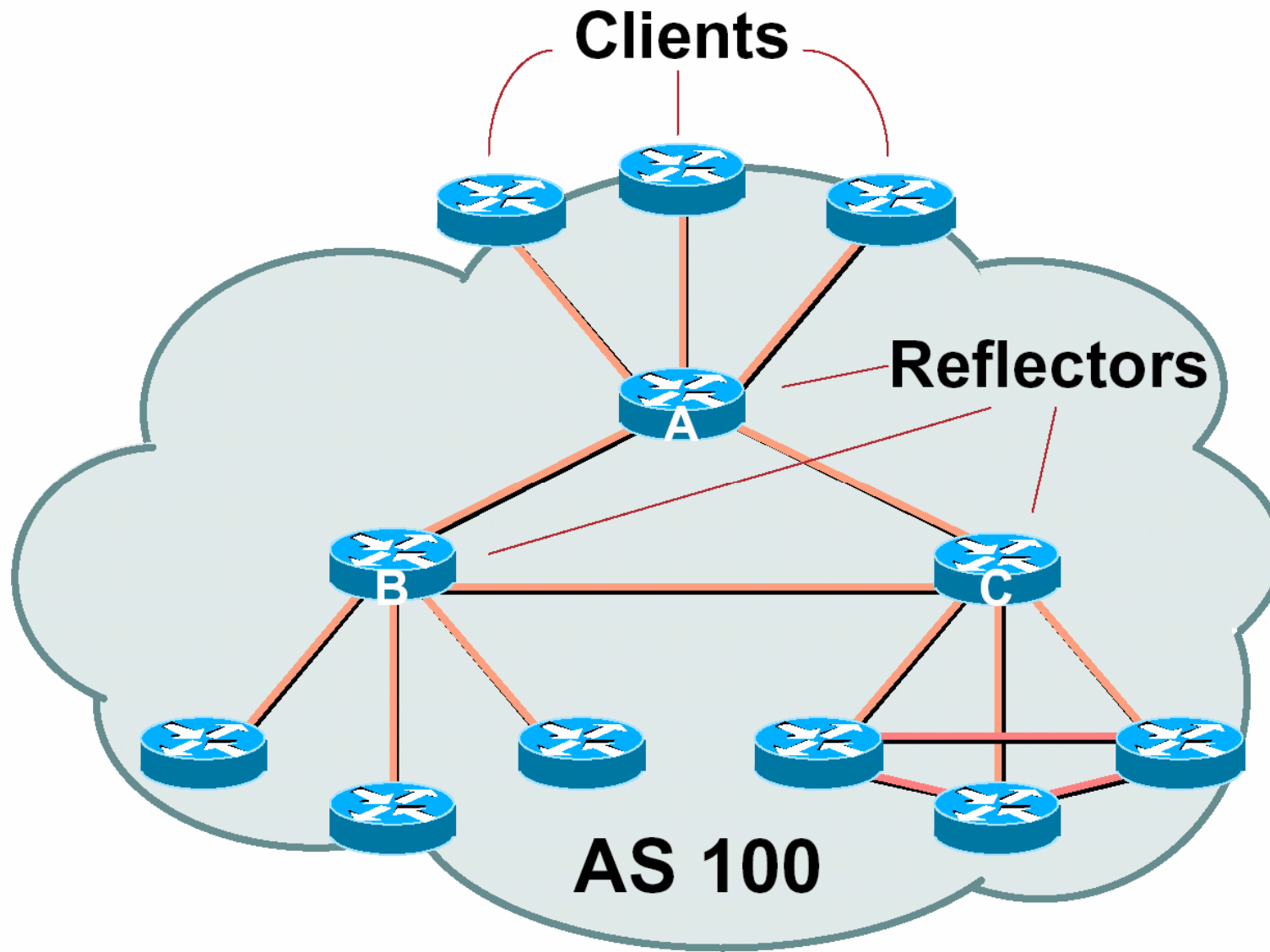
Η τοπολογία mesh του IBGP δεν κλιμακώνεται

- N δρομολογητές σημαίνει $N(N-1)/2$ ομότιμες **συνόδοι**
- Σε κάθε δρομολογητή **N-1** IBGP σύνοδοι πρέπει να διαρθρωθούν
- Η προσθήκη ενός απλού IBGP κόμβου απαιτεί **αλλαγές σε όλους**
- Το **μέγεθος** του πίνακα δρομολόγησης μπορεί να είναι N φορές μεγαλύτερος από τον αριθμό των βέλτιστων διαδρομών
- Κάθε δρομολογητής πρέπει να ακούει τον **θόρυβο** από τις ενημερώσεις του κάθε γείτονα
- Μεγάλος επεξεργαστικός **φόρτος** και κατανάλωση σε **μνήμη**

Λύσεις

- Μεγαλύτεροι δρομολογητές
- Κατάτμηση του AS σε μικρότερα
- **Route Reflectors**, απλότητα
- **Confederations**, περισσότερη πολυπλοκότητα, περισσότερα πλεονεκτήματα

Ανακλαστήρες διαδρομών (Route Reflectors), τοπολογία



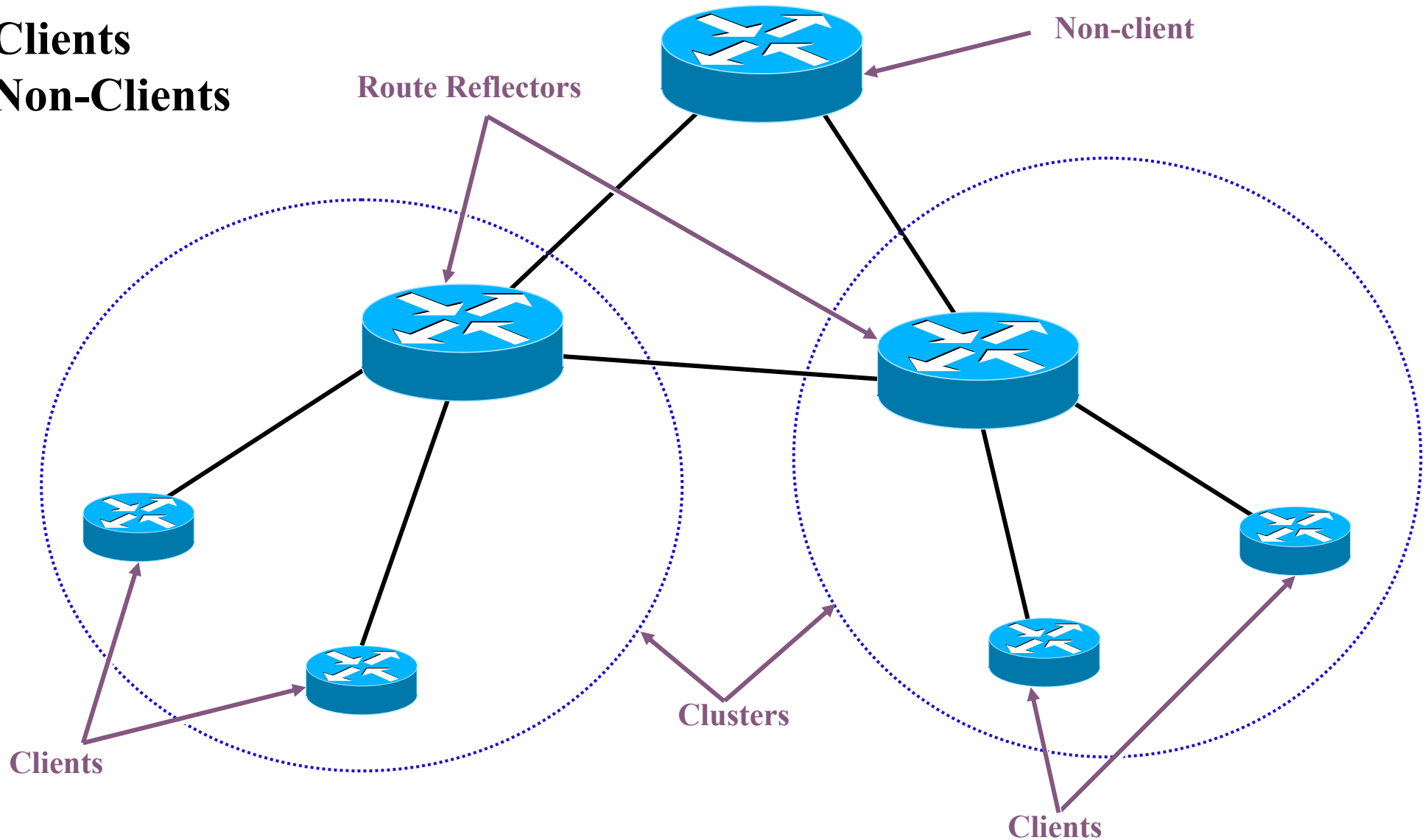
Route Reflectors, τοπολογία

- Περιγράφεται στην **RFC2796**
- Εισάγει μια **ιεραρχία**
- Χωρίζουμε το AS σε ομάδες δρομολογητών που ονομάζονται **clusters**
- Ορίζουμε έναν τουλάχιστον ή περισσότερους (δεν προτείνεται) **Route Reflectors** ανά **cluster**
- Οι RR πρέπει να τοποθετηθούν ακολουθώντας πάντα την **φυσική** τοπολογία (αυτό θα εγγυηθεί ότι η προώθηση των πακέτων δεν θα επηρεαστεί)
- Οι υπόλοιποι δρομολογητές στο cluster ονομάζονται **clients** και συνδέονται με τον **RR**
- Οι RR συνδέονται μεταξύ τους σε τοπολογία **πλέγματος (meshed)**
- Οι πελάτες σε ένα cluster μπορεί να είναι σε τοπολογία πλέγματος, μπορεί και όχι

- Η διάρθρωση είναι **εύκολη**
- Η **μετάβαση** σε τοπολογία με RR είναι εύκολη

Route Reflectors, τοπολογία

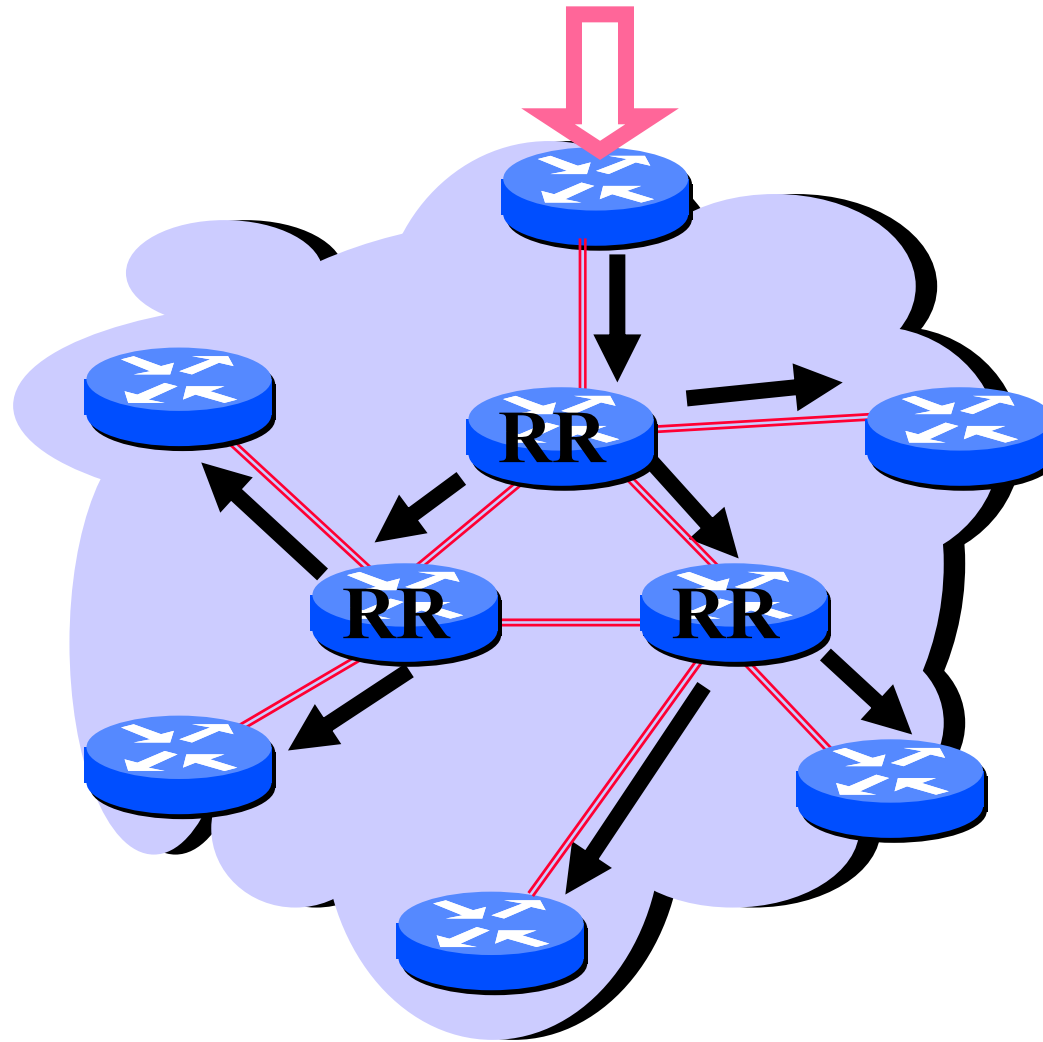
- Route Reflectors
- Clients
- Non-Clients



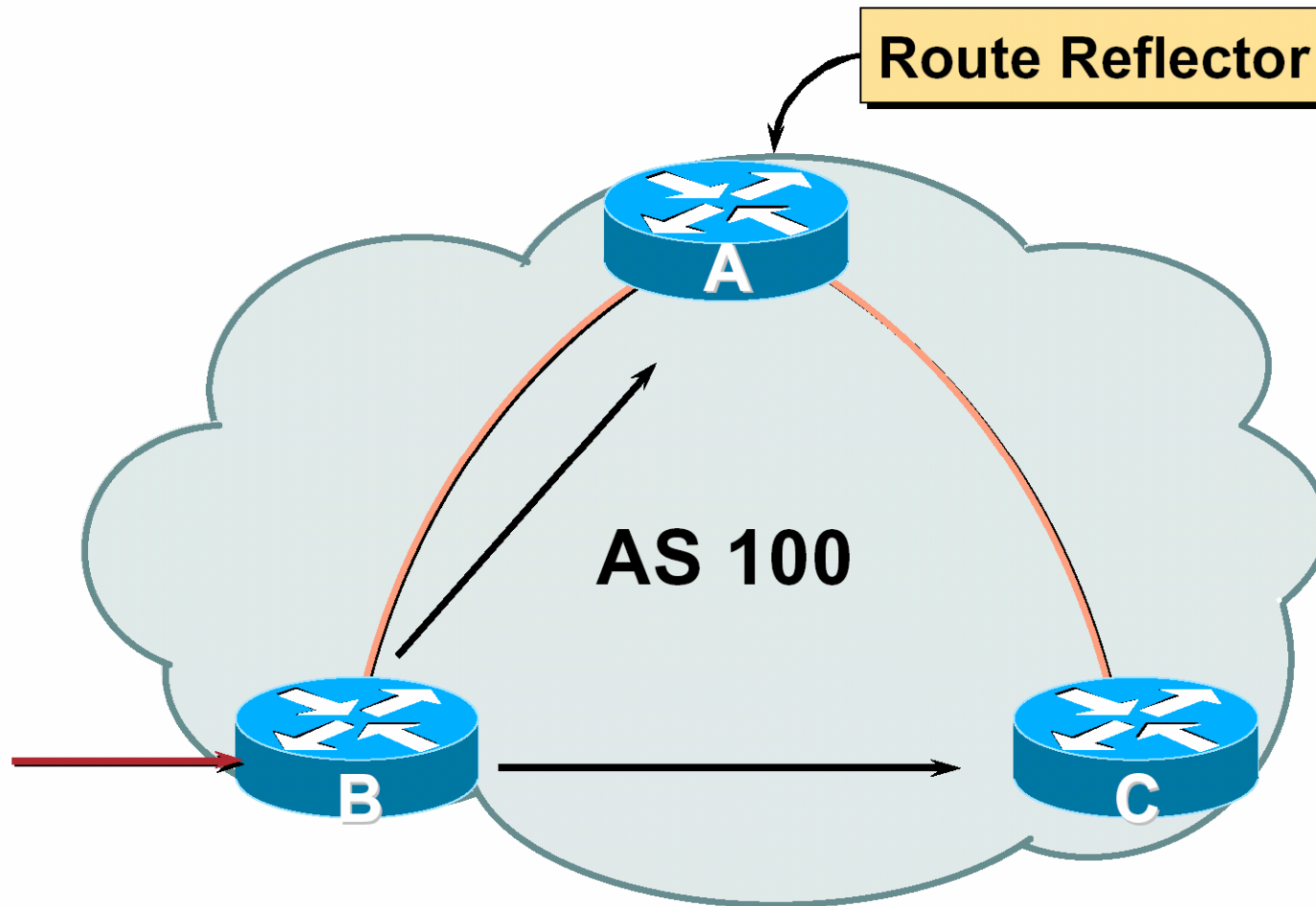
Route Reflectors, λειτουργία

- Ο κάθε **RR** λαμβάνει διαδρομές από **πελάτες** και **μη-πελάτες**
- Επιλέγει και ανακοινώνει μόνο την **καλύτερη** διαδρομή
- Αν αυτή προέρχεται από **πελάτη**, την ανακλά σε άλλους **πελάτες** και **μη-πελάτες**
- Αν αυτή προέρχεται από **μη-πελάτη** την ανακλά μόνο σε **πελάτες**
- Η λειτουργία του BGP αλλάζει μόνο για τους **RR**, οι clients δεν αντιλαμβάνονται καμία διαφορά και **δεν γνωρίζουν** ότι είναι ομότιμοι με έναν RR

Route Reflectors, πορεία μιας ενημέρωσης από EBGP



Route Reflectors, πορεία μιας ενημέρωσης από EBGP



Πρόσθετες Ιδιότητες

- Οι ιδιότητες **Originator_ID** και **Cluster_List** χρειάζονται για την αποφυγή βρόχων

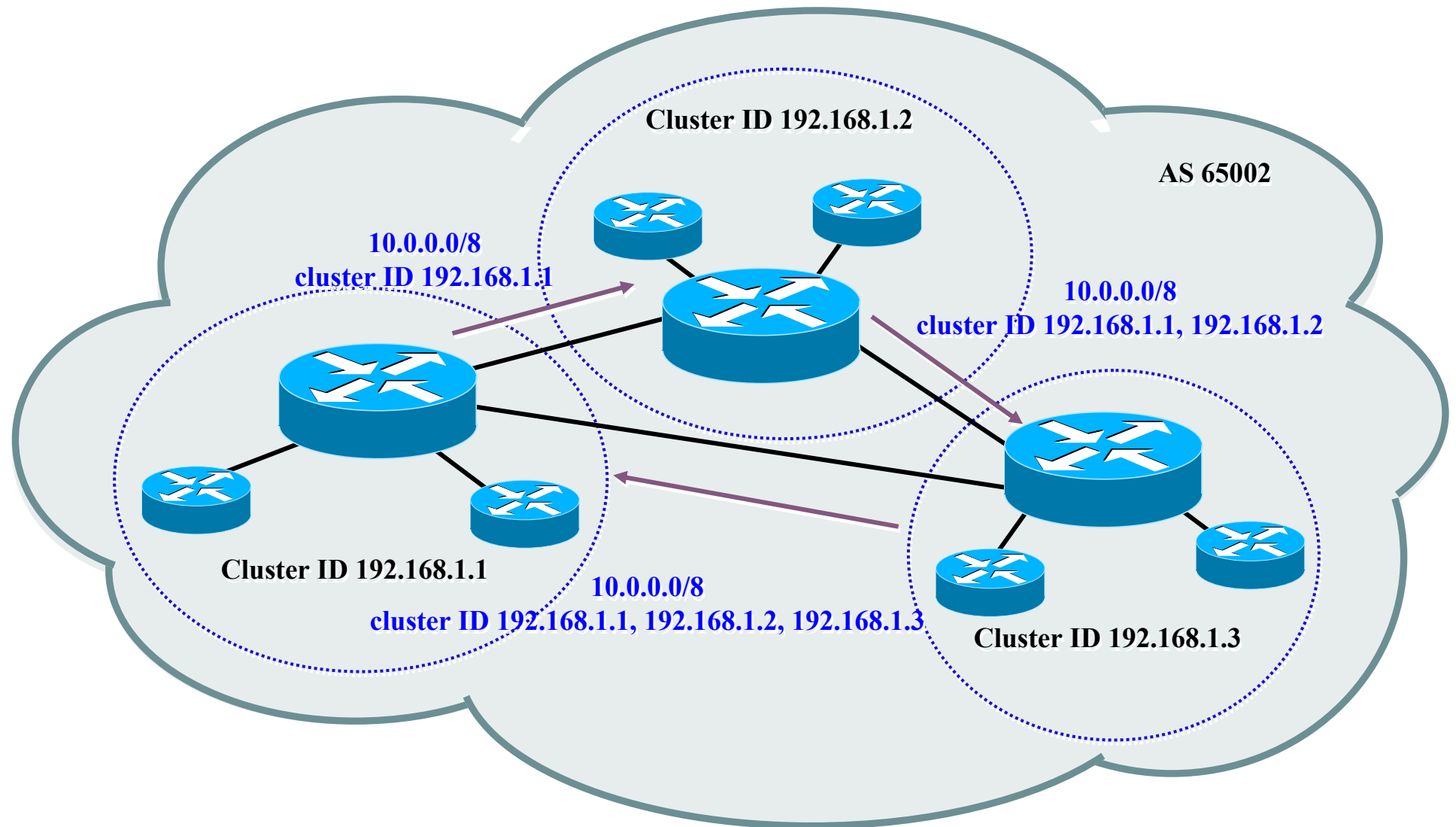
Originator_ID attribute

- Μεταφέρει την **RID** του δρομολογητή που παρήγαγε τη διαδρομή στο τοπικό AS

Cluster_list (RFC 2796)

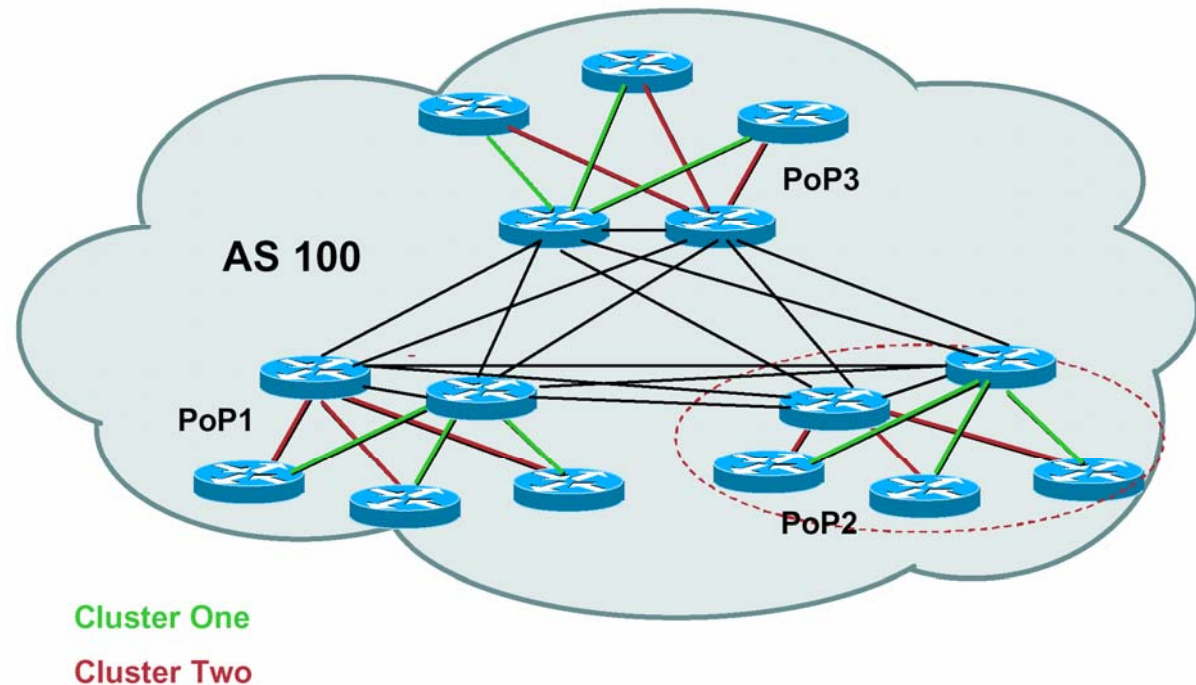
- Προαιρετική μη μεταβατική
- Είναι μια λίστα με τα **clusters** τα οποία διέσχισε μια ενημέρωση
- Ο RR τοποθετεί το **ClusterID** (οι RR σε ένα cluster έχουν το ίδιο **cluster-id**) στη λίστα όταν ανακοινώνει τη διαδρομή **εκτός του cluster**
- Αν ο RR ανιχνεύσει το δικό του **ClusterID** σε μια ανακοίνωση, δεν την κάνει αποδεκτή, αφού αυτό θα προκαλούσε ένα **βρόχο** δρομολόγησης
- Οι clients **δεν τροποποιούν** το πεδίο αυτό

Αποφυγή βρόχου



Εφεδρεία

- Ένας δρομολογητής μπορεί να είναι πελάτης RR που βρίσκονται σε **διαφορετικά cluster**
- Είναι συνήθης πρακτική η **αλληλοκάλυψη** δύο clusters ώστε να επιτευχθεί **εφεδρεία**



Τυπικό ISP δίκτυο

- PoP έχει δύο κεντρικούς RR δρομολογητές
- Δύο **επικαλυπτόμενα** clusters
- Οι κεντρικοί έχουν υλοποιημένο fully meshed iBGP

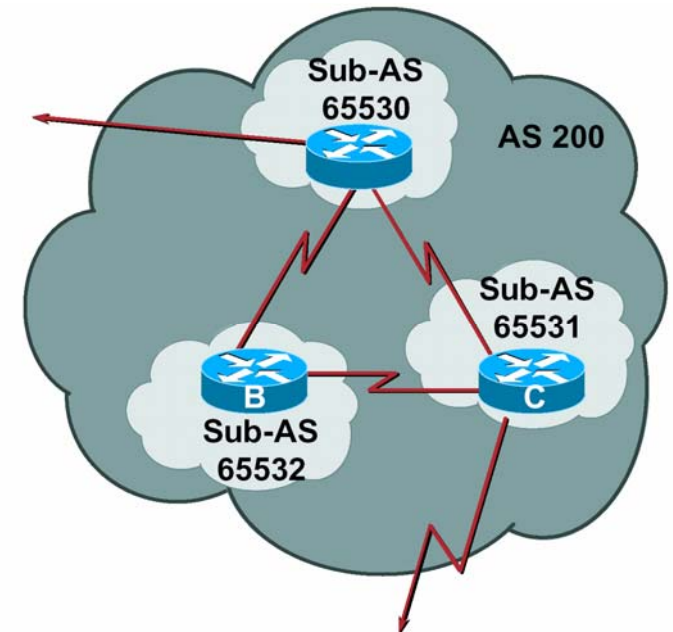
Περαιτέρω κλιμάκωση, κανόνες υλοποίησης

- Αν το πλέγμα γίνεται πολύ μεγάλο δημιουργείτε **επιπλέον ιεραρχία**
- Σπάστε το δίκτυο σε **τμήματα**
- Διαρθρώστε ένα ζευγάρι **cluster** τη φορά
- Εξαλείψτε τις εφεδρικές **iBGP** συνόδους
- Τοποθετήστε **ένα το πολύ RR ανά cluster**
- Υλοποίηση **μικρών** τμημάτων του δικτύου, ένα κάθε φορά
- **IGP** χρειάζεται για να μεταφέρει την πληροφορία για το **next hop** και τα **local routes**
- Πολλές φορές ταυτίζουμε μια **IGP area** με ένα **cluster** ώστε να είναι πιο αποδοτική η σχεδίαση

Συνομοσπονδίες (Confederations), τοπολογία

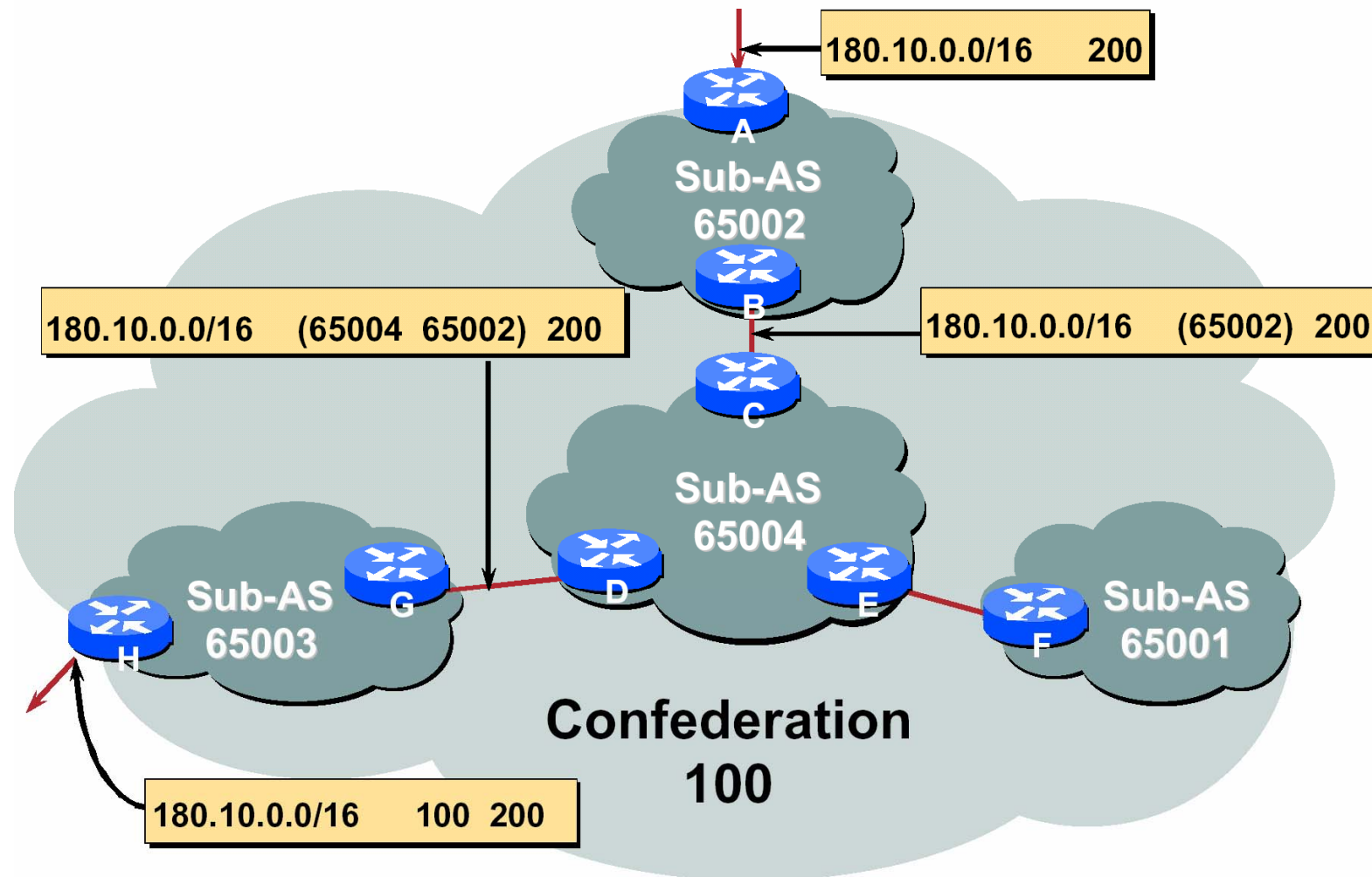
- **RFC3065**

- Τακτική Διαίρει και Βασίλευε
- Χωρίζουμε το AS σε άλλα μικρότερα **ύπο-AS**
- Συνήθως το ύπο-AS θα αναφέρεται σε ένα **POP** (Point of Presence)
- Όλοι οι ομότιμοι στο ίδιο **ύπο-AS** είναι **fully-meshed**
- Κάθε ύπο-AS έχει ένα αριθμό από την κλίμακα των **ιδιωτικών ASN** (64512-65534) ο οποίο ονομάζεται **Confederation Identifier**
- Ο αριθμός αυτός αφαιρείται από την **AS_path** λίστα στο **σύνορο** του ύπο-AS



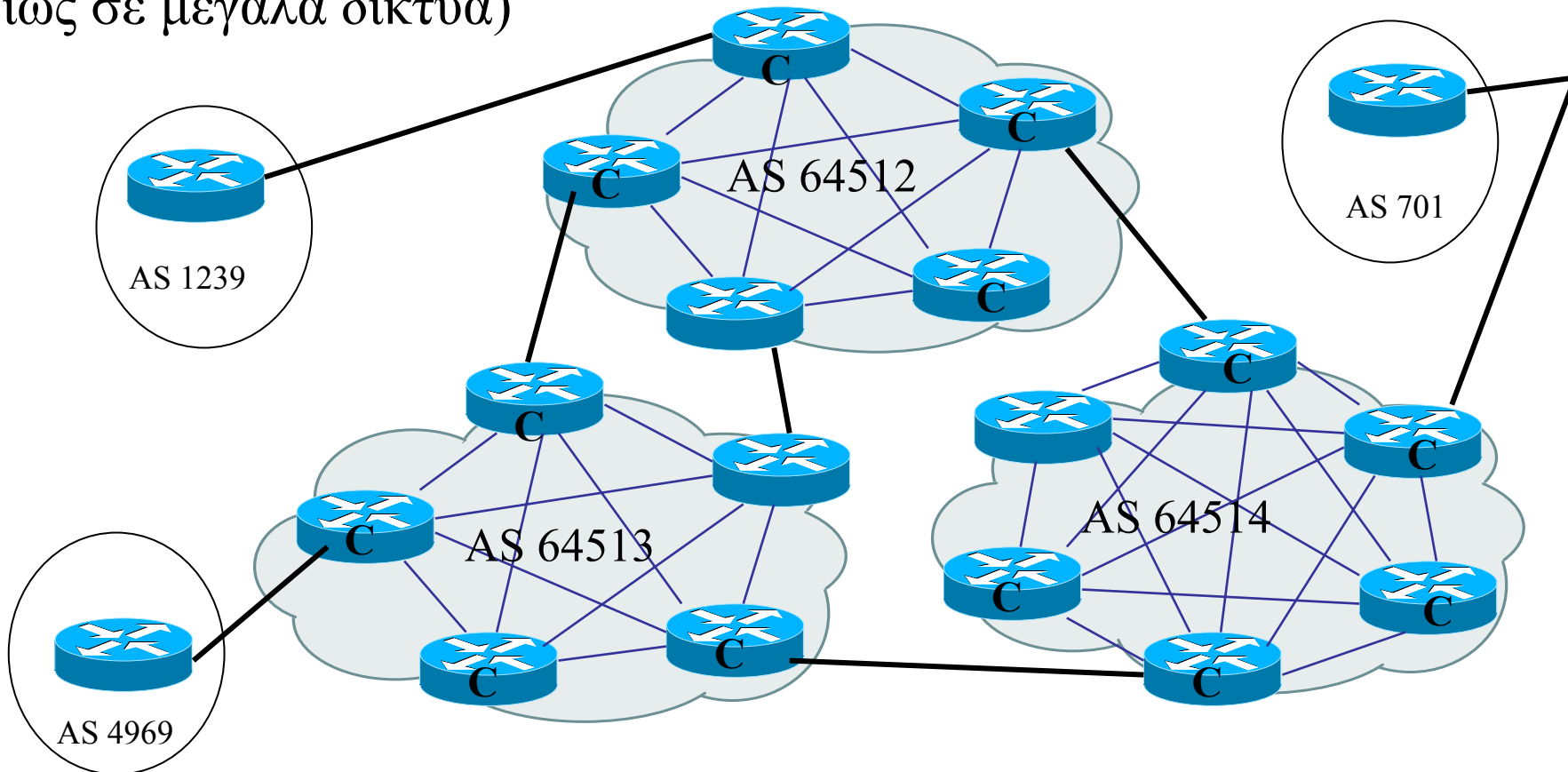
Confederations, τοπολογία

- Το AS φαίνεται στον έξω κόσμο σαν ένα **μονό AS**



Confederations, τοπολογία

- Οι IBGP ομότιμοι σε κάθε ύπο-AS είναι σε τοπολογία **fully meshed**
- Ο συνολικός αριθμός γειτόνων μειώνεται περιορίζοντας την απαίτηση για τοπολογία πλέγματος μόνο στους ομότιμους στο ίδιο ύπο-AS
- Μπορούμε να χρησιμοποιήσουμε και **Route-Reflector** μέσα στο ύπο-AS (ιδίως σε μεγάλα δίκτυα)



Confederations, λειτουργία

- **Παρόμοια** λειτουργία με το κανονικό BGP
- Ομότιμος στο ίδιο **ύπο-AS** ανακοινώνει τις διαδρομές **μόνο** σε **εξωτερικούς** ομότιμους
- Οι **εξωτερικοί** ομότιμοι ανακοινώνουν σε **όλους** τους ομότιμους, μέσα στο ύπο-AS, αλλά και ομότιμους σε άλλα ύπο-AS

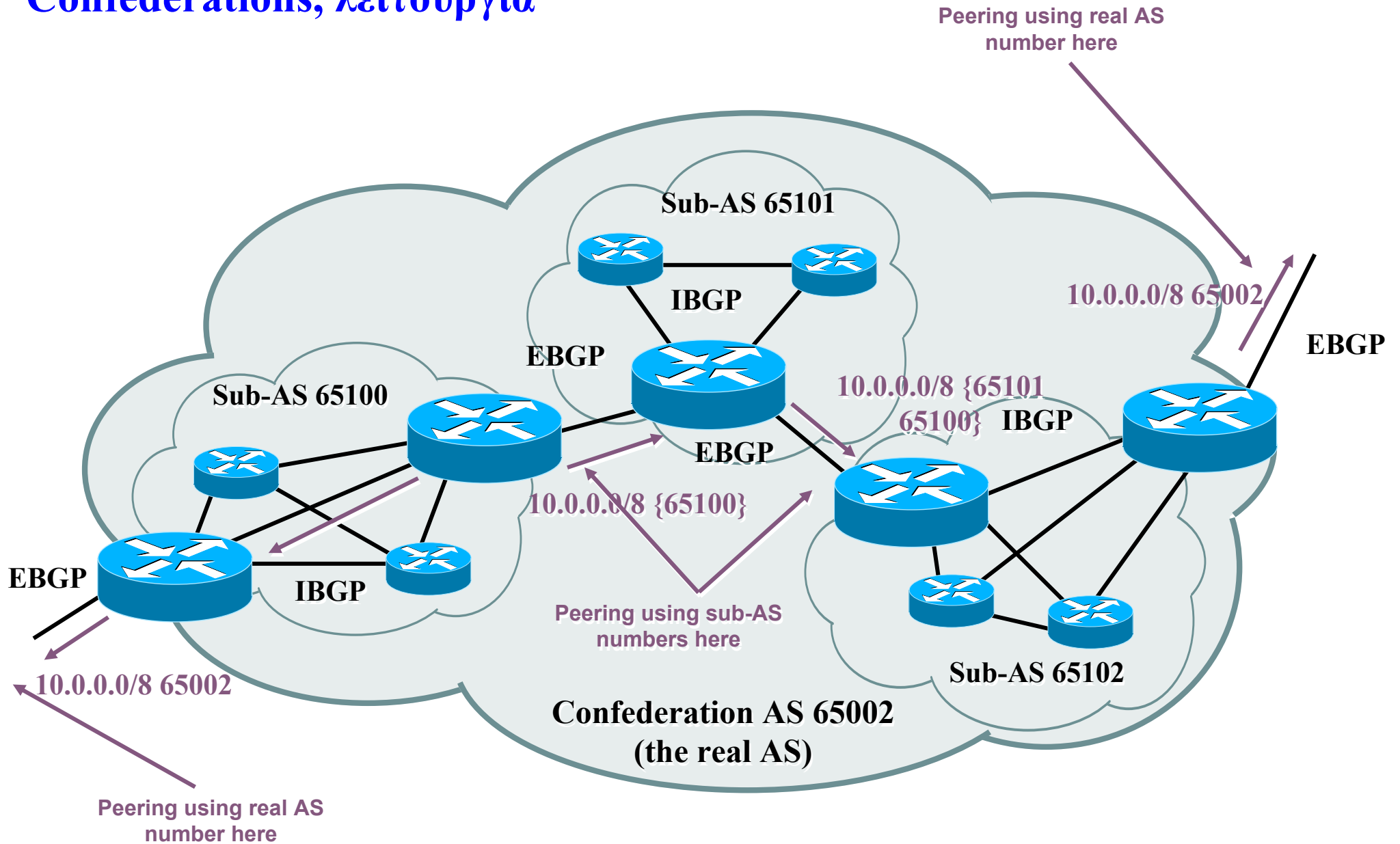
Παράδειγμα

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.0.0	141.153.14.3	0	100	0	(65531) 1 i
*> 141.153.0.0	141.153.30.2	0	100	0	(65530) i
*> 144.10.0.0	141.153.12.1	0	100	0	(65530) i
*> 199.10.10.0	141.153.29.2	0	100	0	(65530) 1 i

Confederations, λειτουργία

- **Confederation EBGP** = **EBGP** ανάμεσα στα ύπο-ASes, αλλά κάποια IBGP πληροφορία κρατείται, είναι δηλαδή ένα **τροποποιημένο EBGP**
- Διατηρείται η τιμή του **NextHop**, **Local_Preference** και **Med**
- Η σύγκριση ανάμεσα σε τιμές **MED** οι οποίες προέρχονται από διαφορετικά ύπο-AS επιτρέπεται (με την υπόθεση βέβαια ότι έχουν οριστεί με κοινή πολιτική)
- Οι MED μπορούν να χρησιμοποιηθούν σαν **IGP μετρικές** ανάμεσα στα ύπο-AS
- Ένα **IGP** χρησιμοποιείται συνήθως για τις **εσωτερικές** διαδρομές
- Μπορούμε να ταιριάξουμε την **IGP ιεραρχία** με τα ύπο-AS ή μπορούμε να έχουμε διαφορετικά IGP στην ίδια ύπο-AS

Confederations, λειτουργία



Confederations, χαρακτηριστικά

- Μπορούμε να εφαρμόσουμε **πολιτική δρομολόγησης** ανά **ύπο-AS**
- Η υλοποίηση είναι **γρήγορη**
- Το κάθε ύπο-AS διατηρεί την **αυτονομία** του
- Η **μετάβαση** από απλό IBGP σε confederation **δεν** είναι τόσο **απλή**, αφού κάθε δρομολογητής στο δίκτυο πρέπει να διαρθρωθεί ώστε να γνωρίζει για το confederation AS και τα ύπο-AS

Πρόσθετες ιδιότητες

AS_Path

- Γνωστή υποχρεωτική
- Ορίζεται στην **RFC 1771**, και επεκτείνεται στην **RFC 3065** προκειμένου να υποστηρίξει τα confederations
- Δύο καινούργιοι **τύποι** ορίζονται για να δείχνουν ότι υπάρχει ένα ύπο-AS στη διαδρομή
- Οι τύποι αυτοί είναι παρόμοιοι με τους AS_SET, AS_SEQUENCE που ορίζονται στην RFC 1771, και χρησιμοποιούνται μόνο όταν έχουμε confederations
- Βοηθά στην αποφυγή **βρόχων** ανάμεσα στα ύπο-AS

AS_CONFED_SET (path segment type 3)

- Ένα σύνολο μη διατεταγμένων ASN, που χρησιμοποιείται για τη συνάθροιση διαδρομών με διαφορετικά AS_path
- Ξεχωρίζει με χρήση των χαρακτήρων “<>”, π.χ. <65001 65002> 65003

AS_CONFED_SEQUENCE (path segment type 4)

- Ένα σύνολο διατεταγμένων ύπο-AS διαδρομών
- Ξεχωρίζει με τη χρήση των χαρακτήρων “{}”, π.χ {65001 65002} 65003

Community

- Προαιρετική μεταβατική
- Ορίζεται στην **RFC 1991**, και επεκτείνεται στην **RFC 3065** προκειμένου να υποστηρίξει τις confederations

Route Reflectors ή Confederations

	Route Reflectors	Confederation
Προσέγγιση	Ιεραρχική προσέγγιση	Διαίρει και Βασίλευε
Πολλαπλά επίπεδα ιεραρχίας	√	√
Κλιμάκωση	+	+++
Πολυπλοκότητα υλοποίησης	++	-
Ευκολία Επέκτασης	+	+
Policy	Ανά cluster	Ανά sub-AS
Αυτονομία	Δεν υπάρχει αυτονομία αναμεσα στα cluster	Αυτονομία ανάμεσα στα sub-AS

- Οι περισσότεροι πάροχοι προτιμούν τους **RR** εξ' αρχής

Πόση κλιμάκωση ακόμα;

- Υπάρχει η άποψη ότι τα υπάρχοντα πρωτόκολλα για δρομολόγηση γρήγορα πλησιάζουν στα **όρια** τους
- Προσθήκες** στο BGP δεν μπορούν να λύσουν θεμελιώδεις αδυναμίες κλιμάκωσης, απλά δίνουν παρατάσεις στη ζωή του πρωτοκόλλου
- Κανείς** δεν ξέρει πόση κλιμάκωση μπορεί να αντέξει ακόμα...

Κατασκευαστές

Κατασκευαστές

Cisco (Crisco)

- Κυριαρχεί στην αγορά όσον αφορά τους δρομολογητές κορμού
- Όλοι τρέχουν cisco

Δεν υποστηρίζουν BGP

- 2501 - 68030, 25mhz, 16mb ram max
- 4000/4000M/4500, 16-32mb max

Υποστηρίζουν BGP

- 2600 - risc, 64mb max - < 3k
- 3620/3640 - risc, 64/128mb max 3-10k
- 4500M/4700 - risc, 64/128mb max 5-10k
- 7206 - risc, 128mb max 12-30k
- 7000 - motorola, 64mb max 4-30k
- 7500 - risc, dist switch, 128mb max 15-70k
- GSR/BFR - “anything you want” 30-120k

Κατασκευαστές

Bay Networks (Bad Networks)

- Ο αμέσως επόμενος κατασκευαστής με μεγάλη διαφορά
- Ιδιοταγής υλοποίηση του BGP
- Αργή ενσωμάτωση των νέων δυνατοτήτων
- Μειονέκτημα το GUI, command-line interface δεν υπάρχει ακόμα

•**Livingston** (Lucington)

- Η υλοποίηση τους δεν έχει όλες τις δυνατότητες, αλλά είναι αποτελεσματική

Ascend (ASS-END)

- Τρέχει gated και είναι χάλια, θέλει επανεκκίνηση κάθε αλλαγή

Allied Tellesys

- Δρομολογητές με καλή τιμή - απόδοση
- Πολύ καλή υποστήριξη – τεκμηρίωση
- Υψηλό κόστος για την ενεργοποίηση BGP λογισμικού

PC router w/ unix

- Κάποιο **λογισμικό** όπως το gated τρέχει σε ένα Unix κουτί, δίνοντας δυνατότητα για ένα πλήθος από πρωτόκολλα
- Έχει **μηχανικά** μέρη – σκληρός δίσκος ο οποίος παρουσιάζει αυξημένη πιθανότητα βλάβης
- Θόρυβος, κατανάλωση**
- Συνήθως έχει φτωχή υποστήριξη και τεκμηρίωση
- Καλή **επίδοση**, καλύτερη από H/W δρομολογητές χαμηλής-μεσαίας σκάλας
- Δυνατότητα **επέκτασης**, υποστήριξης **νέων** πρωτοκόλλων
- Ελάχιστο **κόστος**
- Πολλοί πάροχοι προτιμούν τη λύση αυτή

PC router w/ unix

Zebra (<http://www.zebra.org>)

- GNU μορφή ενός εμπορικού πακέτου
- Κάποιες δυνατότητες-πρωτόκολλα υπάρχουν μόνο στο εμπορικό πακέτο

Quagga (<http://www.quagga.net>)

- Κλώνος της Zebra ακολουθεί την ίδια αρχιτεκτονική
- Περισσότερες δυνατότητες
- Αναπτύσσεται από μία κοινότητα
- Διαθέσιμα πακέτα για RedHat & Fedora Core, Debian, Gentoo
- Το CLI της είναι παρόμοιο με του Cisco's IOS
- Χρησιμοποιείται από πολλούς ISPs
- Μπορεί να ξεπεράσει σε επίδοση άλλους δρομολογητές

PC router w/ unix

Quagga (<http://www.quagga.net>)

- Αποτελείται από τους ακόλουθους δαίμονες:

Zebra : Μεσολαβεί ανάμεσα στους πίνακες των πρωτοκόλλων και τον πίνακα του πυρήνα

OSPFd: OSPF v2

OSPF6d: OSPF v3 (for IPv6)

RIPd: RIP v1 and v2

RIPngd: RIP v3 (for IPv6)

BGPd: BGP v4+

VTYsh: Γραμμή εντολών

- Τεκμηρίωση : <http://www.quagga.net/docs.php>

- Πολλές προχωρημένες δυνατότητες επιπρόσθετα με τη βασική υποστήριξη

Βιβλιογραφία

Δρομολόγηση

- Internet Routing Architectures. Bassam Halabi. Second edition Cisco Press, 2000
 - BGP4: Inter-domain Routing in the Internet. John W. Stewart, III. Addison-Wesley, 1999
 - Routing in the Internet. Christian Huitema. 2000
 - Parkhurst, William R., and David R. Jackson. Practical BGP for Internet Routing. Cisco Press
 - ISP Survival Guide: Strategies for Running a Competitive ISP. Geoff Huston. Wiley, 1999
 - Interconnection, Peering and Settlements. Geoff Huston. The Internet Protocol Journal. March and June 1999
- BGP Protocol Technology Page,
http://www.cisco.com/en/US/tech/tk365/tk80/tech_protocol_family_home.html
- BGP4 Case Studies, <http://www.cisco.com/warp/public/459/bgp-toc.html>

Interdomain Routing

- Cooperative Association for Internet Data Analysis (CAIDA)
 - <http://www.caida.org/>
 - Tools and analyses promoting the engineering and maintenance of a robust, scalable global Internet infrastructure
- Internet Performance Measurement and Analysis (IPMA)
 - <http://www.merit.edu/ipma/>
 - Studies the performance of networks and networking protocols in local and wide-area networks
- National Laboratory for Applied Network Research (NLNR)
 - <http://www.nlanr.net/>
 - Analysis, tools, visualization.
- IRTF Routing Research Group (IRTF-RR)
 - <http://puck.nether.net/irtf-rr/>
- <http://www.cambridge.intel-research.net/~tgriffin/interdomain/>
- http://www.cambridge.intel-research.net/~tgriffin/talks_tutorials/CL_2031024.ppt

Θεωρία του BGP

- Persistent Route Oscillations in Inter-Domain Routing. Kannan Varadhan, Ramesh Govindan, and Deborah Estrin. Computer Networks, Jan. 2000. (Also USC Tech Report, Feb. 1996)
 - Shows that BGP is not guaranteed to converge
- An Architecture for Stable, Analyzable Internet Routing. Ramesh Govindan, Cengiz Alaettinoglu, George Eddy, David Kessens, Satish Kumar, and WeeSan Lee. IEEE Network Magazine, Jan-Feb 1999.
 - Use RPSL to specify policies. Store them in registries. Use registry for configuration generation and analysis.
- An Analysis of BGP Convergence Properties. Timothy G. Griffin, Gordon Wilfong. SIGCOMM 1999
 - Model BGP, shows static analysis of divergence in policies is NP complete
- Policy Disputes in Path Vector Protocols. Timothy G. Griffin, F. Bruce Shepherd, Gordon Wilfong. ICNP 1999
 - Define Stable Paths Problem and develop sufficient condition for “sanity”
- A Safe Path Vector Protocol. Timothy G. Griffin, Gordon Wilfong. INFOCOM 2001
 - Dynamic solution for SPVP based on histories
- Stable Internet Routing without Global Coordination. Lixin Gao, Jennifer Rexford. SIGMETRICS 2000
 - Show that if certain guidelines are followed, then all is well.
- Inherently safe backup routing with BGP. Lixin Gao, Timothy G. Griffin, Jennifer Rexford. INFOCOM 2001
 - Use SPP to study complex backup policies

Ευστάθεια και σύγκλιση του BGP

- The Impact of Internet Policy and Topology on Delayed Routing Convergence. Craig Labovitz, Abha Ahuja, Roger Wattenhofer, Srinivasan Venkatachary. INFOCOM 2001
- An Experimental Study of BGP Convergence. Craig Labovitz, Abha Ahuja, Abhijit Abose, Farnam Jahanian. SIGCOMM 2000
- Origins of Internet Routing Instability. C. Labovitz, R. Malan, F. Jahanian. INFOCOM 1999
- Internet Routing Instability. Craig Labovitz, G. Robert Malan and Farnam Jahanian. SIGCOMM 1997
- Persistent Route Oscillations in Inter-Domain Routing. Kannan Varadhan, Ramesh Govindan, and Deborah Estrin. Computer Networks, Jan. 2000. (Also USC Tech Report, Feb. 1996)
- An Analysis of the MED oscillation Problem. Griffin and Wilfong. ICNP 2002

Ευστάθεια και σύγκλιση του BGP

- RIPE-229: “RIPE Routing-WG Recommendations for Coordinated Route-flap Damping Parameters”, 2001
- RFC 2439: “BGP Route Flap Damping”, 1998
- Randy Bush, Tim Griffin, Z. Mao, “Route flap Damping: harmful?” RIPE 43, 2002
- Tim Griffin, “What is the sound of one route flapping” Dartmouth talk slides, June 2002
- C. Labovitz, G. R. Malan, F. Jahanian, “Internet Routing Instability”, TON 1998
- C. Labovitz, R. Malan, F. Jahanian, “Origins of Internet Routing Instability”, Infocom 1999
- C. Labovitz, A. Ahuja, F. Jahanian, “Experimental Study of Internet Stability and Wide-Area Network Failures”, FTCS 1999
- C. Labovitz, A. Ahuja, A. Bose, F. Jahanian, “Delayed Internet Routing Convergence” Sigcomm 2000
- C. Labovitz, A. Ahuja, R. Wattenhofer, S. Venkatachary, “The Impact of Internet Policy and Topology on Delayed Routing Convergence”, Infocom 2001
- Z. Mao, R. Govindan, G. Varghese, R. Katz “Route Flap Damping Exacerbates Internet Routing Convergence” Sigcomm 2002

Γενικά για BGP

- Recommendations for ISPs : <http://www.ripe.net/docs/ripe-229.html>
- An Analysis of BGP Convergence Properties. Timothy G. Griffin, Gordon Wilfong. SIGCOMM 1999
- Policy Disputes in Path Vector Protocols. Timothy G. Griffin, F. Bruce Shepherd, Gordon Wilfong. ICNP 1999
- A Safe Path Vector Protocol. Timothy G. Griffin, Gordon Wilfong. INFOCOM 2001
- Stable Internet Routing without Global Coordination. Lixin Gao, Jennifer Rexford. SIGMETRICS 2000
- Inherently safe backup routing with BGP. Lixin Gao, Timothy G. Griffin, Jennifer Rexford. INFOCOM 2001
- On the Correctness of IBGP Configurations. Griffin and Wilfong. SIGCOMM 2002.

Internet Route Registries

- Internet Route Registry
 - <http://www.irr.net/>
- Routing Policy Specification Language (RPSL)
 - RFC 2622 Routing Policy Specification Language (RPSL)
 - RFC 2650 Using RPSL in Practice
- Internet Route Registry Daemon (IRRd)
 - <http://www.ird.net/>
- RAToolSet
 - <http://www.isi.edu/ra/RAToolSet/>

Addressing και ASN RFCs

- RFC 1380** IESG Deliberations on Routing and Addressing (1992)
- RFC 1517** Applicability Statement for the Implementation of Classless Inter- Domain Routing (CIDR) (1993)
- RFC 1518** An Architecture for IP Address Allocation with CIDR (1993)
- RFC 1519** Classless Inter-Domain Routing (CIDR) (1993)
- RFC 1467** Status of CIDR Deployment in the Intrenet (1983)
- RFC 1520** Exchanging Routing Information Across Provider Boundaries in the CIDR Environment (1993)
- RFC 1817** CIDR and Classful routing (1995)
- RFC 1918** Address Allocation for Private Internets (1996)
- RFC 2008** Implications of Various Address Allocation Policies for Internet Routing (1996)
- RFC 2050** Internet Registry IP Allocation Guidelines (1996)
- RFC 2260** Scalable Support for Multi-homed Multi-provider Connectivity (1998)
- RFC 2519** A Framework for Inter-Domain Route Aggregation (1999)
- RFC 1930** Guidelines for creation, selection, and registration of an Autonomous System (AS)
- RFC 2270** Using a Dedicated AS for Sites Homed to a Single Provider

Επιλεγμένα BGP RFCs

IETF, Interdomain Routing : <http://www.ietf.org/html.charters/idr-charter.html>

RFC 1771 A Border Gateway Protocol 4 (BGP-4) - draft-ietf-idr-bgp4-12.txt

RFC 1772 Application of the Border Gateway Protocol in the Internet

RFC 1773 Experience with the BGP-4 protocol

RFC 1774 BGP-4 Protocol Analysis

RFC 2796 BGP Route Reflection An alternative to full mesh IBGP

RFC 3065 Autonomous System Confederations for BGP

RFC 1997 BGP Communities Attribute

RFC 1998 An Application of the BGP Community Attribute in Multi-home Routing

RFC 2439 Route Flap Dampening

Μερικά πρόσφατα Internet Drafts

- Dynamic Capability for BGP-4
- Application of Multiprotocol BGP-4 to IPv4 Multicast Routing
- Graceful Restart mechanism for BGP
- Cooperative Route Filtering Capability for BGP-4
- Address Prefix Based Outbound Route Filter for BGP-4
- Aspath Based Outbound Route Filter for BGP-4
- Architectural Requirements for Inter-Domain Routing in the Internet
- BGP support for four-octet AS number space
- Autonomous System Number Substitution on Egress
- BGP Extended Communities Attribute
- Controlling the redistribution of BGP routes
- BGP Persistent Route Oscillation Condition
- Benchmarking Methodology for Basic BGP Convergence
- Terminology for Benchmarking External Routing Convergence Measurements