

**ΠΡΟΤΑΣΗ ΓΙΑ STANDBY ROUTE REFLECTORS
ΣΤΟ ΠΡΩΤΟΚΟΛΛΟ BGP-4**

ΙΩΑΝΝΗΣ Γ. ΑΛΕΞΙΟΥ

μεταπτυχιακή εργασία

Επιβλέπων Καθηγητής: Καθ. Αναστάσιος Α. Οικονομίδης
Εξεταστές: Καθ. Βασίλειος Βίτσας

Δ.Π.Μ.Σ στα Πληροφοριακά Συστήματα

Πανεπιστήμιο Μακεδονίας
Θεσσαλονίκη

Φεβρουάριος 2009

Copyright © Ιωάννης Γ. Αλεξίου, 2009

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της μεταπτυχιακής εργασίας από το Δ.Π.Μ.Σ στα Πληροφοριακά Συστήματα του Πανεπιστημίου Μακεδονίας δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

ΠΕΡΙΛΗΨΗ

Το BGP-4 αποτελεί το πρωτόκολλο δρομολόγησης μεταξύ αυτόνομων συστημάτων (AS) για ανταλλαγή πληροφοριών δρομολόγησης. Σήμερα, το BGP ανταλλάσσει πολύ περισσότερα πράγματα από IPv4 διαδρομές, όπως πληροφορίες MPLS και φαντάζει αναντικατάστατο στην βιομηχανία των δικτύων υπολογιστών.

Το BGP υπάρχει σε δύο εκδοχές που συνεργάζονται στενά για την επίτευξη του τελικού αποτελέσματος. Το eBGP ανταλλάσσει μηνύματα μεταξύ δρομολογητών που ανήκουν σε διαφορετικά AS, ενώ το iBGP φροντίζει για την μετάδοση των πληροφοριών δρομολόγησης εντός AS.

Το iBGP, λόγω σχεδιασμού απαιτεί τη λογική συνδεσιμότητα των δρομολογητών σε πλήρες πλέγμα. Αυτό οφείλεται στον κανόνα του BGP που ορίζει ότι μια διαδρομή που έγινε γνωστή μέσω iBGP δεν πρέπει να διαδοθεί σε γειτονικό δρομολογητή του ίδιου AS για την αποφυγή δημιουργίας βρόχων δρομολόγησης.. Έτσι, ο αριθμός των λογικών συνδέσεων γίνεται απαγορευτικός για μεγάλα AS. Λόγω αυτού του περιορισμού, προτάθηκαν λύσεις για την μείωση των λογικών συνδέσεων μεταξύ των δρομολογητών ενός AS με δημοφιλέστερη την λύση του Route Reflection Clustering.

Βάσει του Route Reflection, οι δρομολογητές εντός του AS χωρίζονται σε ομάδες (clusters). Εντός κάθε ομάδος, υπάρχει ένας Route Reflector ή περισσότεροι για λόγους αξιοπιστίας, ενώ οι υπόλοιποι δρομολογητές του cluster χωρίζονται σε πελάτες (RR clients) ή μη (non clients). Αυτό που διαφοροποιεί τον Route Reflector από τους υπόλοιπους BGP δρομολογητές είναι πως έχει την δυνατότητα να διαδώσει μια διαδρομή που έγινε γνωστή μέσω iBGP σε γειτονικό δρομολογητή (RR client) του ίδιου AS. Έτσι, οι Route Reflectors παραβαίνοντας τον κανόνα του BGP και καταφέρνουν να διαδώσουν την πληροφορία δρομολόγησης στο AS, χωρίς να απαιτείται τοπολογία πλέγματος λογικών συνδέσεων. Η αποφυγή δημιουργίας βρόχων δρομολόγησης, γίνεται μέσω νέων μηχανισμών που εισάγει το Route Reflection. Αποτέλεσμα αποτελεί για n αριθμό δρομολογητών εντός ενός AS η μείωση των λογικών συνδέσεων σε (n-1), από $[n*(n-1)/2]$ για τοπολογία πλήρους πλέγματος.

Ωστόσο το Route Reflection δημιούργησε και κάποιες ατέλειες, όπως :

- αστάθεια επιλογής διαδρομής που οδηγεί σε ταλαντώσεις κάτω από συγκεκριμένες συνθήκες.
- πιθανότητα κάποιος RR client να μην αποκτήσει γνώση της βέλτιστης διαδρομής προς κάποιο AS.
- Ο Route Reflector αποτελεί κρίσιμο σημείο και πιθανή δυσλειτουργία του οδηγεί σε κατάρρευση ολόκληρου του cluster. Γενικότερα μια μεταβολή στην IGP τοπολογία μπορεί να οδηγήσει το AS σε απρόβλεπτη συμπεριφορά.
- Η χρήση Redundant Route Reflectors στην καλύτερη περίπτωση διπλασιάζει τον αριθμό λογικών συνδέσεων, που πολλές φορές γίνεται απαγορευτικός.

Πολλές επιστημονικές έρευνες έχουν βοηθήσει στην ελαχιστοποίηση των ελαττωμάτων του Route Reflection προτείνοντας βασικούς κανόνες σχεδιασμού, μεταβολή του iBGP σε βασικές παραμέτρους ή προσθέτοντας επιπλέον πρωτόκολλο.

Η λύση της προσθήκης στο BGP του πρωτοκόλλου Standby Route Reflection προτείνεται στο παρόν έγγραφο. Η λύση συνδυάζει τα πλεονεκτήματα του Route Reflection και του Redundant RR έχοντας μικρή απαίτηση σε μνήμη από τους RR client και προσφέροντας αξιοπιστία στο AS. Γνώση του πρωτοκόλλου πρέπει να έχουν μόνο οι Route Reflectors, οπότε οι clients μπορεί να είναι συμβατικοί δρομολογητές. Το Standby Route Reflection, υπόσχεται την καλύτερη δυνατή λειτουργία του AS έπειτα από σοβαρές αλλαγές της IGP τοπολογίας ή την δυσλειτουργία κάποιων δρομολογητών του AS. Στον αντίποδα, επιβαρύνει τους Route Reflectors με επιπλέον πρωτόκολλο που απαιτεί επιπλέον μνήμη, υπολογιστική ισχύ και εύρος ζώνης μεταξύ των λογικών συνδέσεων των Standby RR.

Το πρωτόκολλο BGP-4 καθώς και το Standby Route Reflection, έχουν υλοποιηθεί σε γλώσσα προγραμματισμού java (BGPConsole) ώστε να χρησιμοποιηθεί για πειραματικούς σκοπούς και λεπτομερή μελέτη του BGP routing table και των μηνυμάτων που ανταλλάσσονται, χωρίς ουσιαστικά να δρομολογεί πακέτα. Το πρόγραμμα μπορεί να εγκατασταθεί σε Η/Υ με λειτουργικό σύστημα windows ή linux . Βάση μιας σειράς πειραματικών μετρήσεων με το BGPConsole καθώς και πραγματικούς δρομολογητές εξάγουμε ασφαλή συμπεράσματα για τους παραπάνω ισχυρισμούς.

ΠΕΡΙΕΧΟΜΕΝΑ

1 Εισαγωγή	6
1.1 Χρησιμότητα του πρωτοκόλλου BGP.....	6
1.2 Βασικές έννοιες του πρωτοκόλλου BGP.....	7
1.2.1 Περίληψη λειτουργίας.....	7
1.2.2 Μορφή μηνυμάτων.....	9
1.2.3 Χαρακτηριστικά διαδρομών (Path Attributes).....	12
1.2.4 BGP Finite State Machine.....	14
1.2.5 BGP Path Selection.....	16
1.3 Εξωτερικό BGP (eBGP).....	16
1.4 Εσωτερικό BGP (iBGP).....	17
1.4.1 Περιγραφή του iBGP.....	17
1.4.2 Βελτίωση μέσω AS Confederations.....	18
1.4.3 Βελτίωση μέσω Route Reflection.....	19
2 Route Reflection	21
2.1 Περιγραφή λειτουργίας Route Reflection.....	21
2.2 Redundant Route Reflectors.....	25
3 Προβλήματα του Route Reflection	25
3.1 Αστάθεια τύπου-I.....	25
3.2 Επιλογή μη βέλτιστης διαδρομής.....	28
3.3 Συμπεριφορά σε περίπτωση δυσλειτουργίας.....	29
4 Επισκόπηση Επιστημονικής συνεισφοράς	32
4.1 Προτάσεις για σωστό σχεδιασμό με Route Reflectors.....	32
4.2 Προτάσεις για την επιλογή βέλτιστης διαδρομής.....	33
4.3 Προτάσεις για ομαλή λειτουργία σε περίπτωση βλάβης.....	35
5 Πρωτόκολλο Standby Route Reflection	38
5.1 Περιγραφή πρωτοκόλλου Standby Route Reflection.....	39
5.2 Ενσωμάτωση του Standby RR στο BGP.....	41
5.3 Τύποι μηνυμάτων.....	42
5.3.1 τύπος SUBTYPE_OPEN.....	42
5.3.2 τύπος SUBTYPE_UPDATE_ADD & WTD.....	43
5.4 Επίδραση στους timers του BGP.....	44
5.5 Διαδικασία επιλογής πελατών ανά RR.....	45
5.6 Διάδοση βέλτιστης διαδρομής.....	48
5.7 Αντίδραση σε περίπτωση δυσλειτουργίας.....	49
5.8 Πλεονεκτήματα-μειονεκτήματα.....	51
6 Υλοποίηση του πρωτοκόλλου BGP σε java (BGPConsole)	53
6.1 Στόχος του προγράμματος.....	53
6.2 Υλοποίηση σε java.....	53
6.2.1 Υποστηριζόμενα πρωτόκολλα.....	55
6.2.2 Σει κλάσεων.....	55
6.3 Εγκατάσταση προγράμματος.....	56
6.4 Μέθοδος παραμετροποίησης.....	57
6.5 Μέθοδος εισαγωγής IGP τοπολογίας.....	60
7 Εργαστηριακές μετρήσεις	61
7.1 Πειράματα.....	61
7.1.1 Redundant Route Reflection.....	61
7.1.2 Standby Route Reflection.....	61

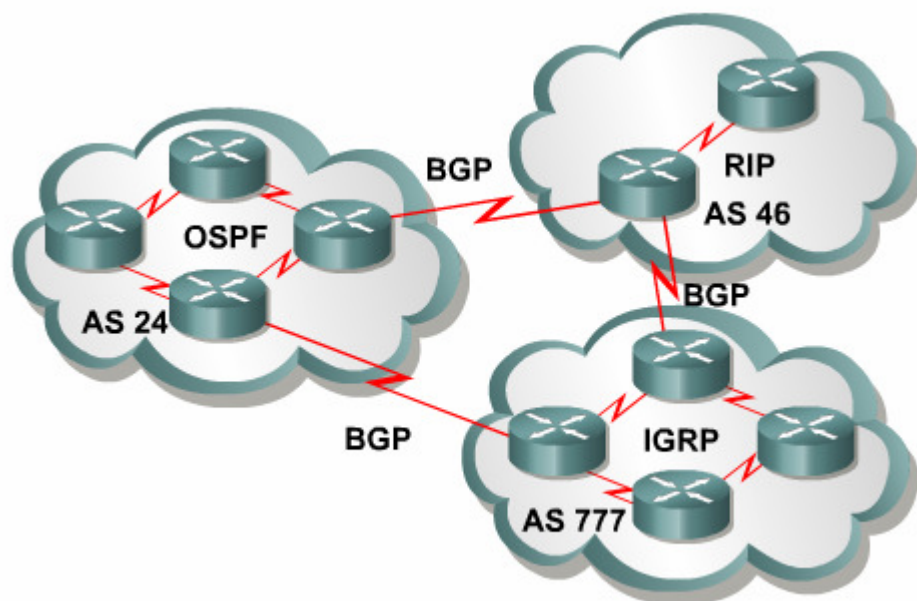
7.2 Σύγκριση πειραματικών μετρήσεων.....	62
7.3 Πείραμα συμβατότητας.....	67
8 Συμπεράσματα.....	71
Βιβλιογραφία.....	72

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Χρησιμότητα του πρωτοκόλλου BGP

Τα δίκτυα των παρόχων καθώς και αυτά των πελατών, συνήθως, χρησιμοποιούν πρωτόκολλα εσωτερικής δρομολόγησης (IGP) όπως είναι τα RIP, OSPF, Is-Is για την ανταλλαγή πληροφοριών δρομολόγησης, εντός των δικτύων τους. Συνήθως τα δίκτυα αυτά που βρίσκονται υπό κοινή διαχείριση, ονομάζονται αυτόνομα συστήματα (AS) Για να επιτευχθεί η διασύνδεση αυτών των δικτύων (AS), χρησιμοποιούμε πρωτόκολλο εξωτερικής δρομολόγησης (EGP). Στην κατηγορία αυτή (EGP), έχει επικρατήσει μέχρι σήμερα το πρωτόκολλο BGP-4 (Rekhter, Y & Li, T & Hares, S. 2003) , ώστε να πραγματοποιείται η ανταλλαγή πληροφοριών δρομολόγησης μεταξύ των παρόχων, των πελατών και του διαδικτύου (interdomain routing).



Εικόνα 1.1.1 : Διασύνδεση AS μέσω του πρωτοκόλλου BGP
(Πηγή Cisco Networking Academy v.1.0)

Η ιδιαιτερότητα του interdomain routing, έγκειται στην ανάγκη να ανταλλάσσονται οι πληροφορίες δρομολόγησης βάσει τεχνικών χαρακτηριστικών,

αλλά και βάσει πολιτικής. Έτσι, σε αντίθεση με τα IGP, το BGP δεν έχει πάντα ως στόχο την εύρεση και ανταλλαγή της τεχνικά καλύτερης διαδρομής.

Ειδικότερα, ένα πρωτόκολλο EGP καλείται να υποστηρίξει τα παρακάτω :

- Προσαρμοστικότητα : Τα EGP καλούνται να ανταλλάξουν τις πληροφορίες δρομολόγησης του διαδικτύου που ανέρχονται σε αριθμό μεγαλύτερο των 110,000 διαδρομών IPv4.
- Ασφάλεια : Καθώς οι δρομολογητές άλλων AS δεν μπορούν να θεωρηθούν αξιόπιστοι, πρέπει να υπάρχει μέθοδος ασφαλούς ανταλλαγής διαδρομών
- Υποστήριξη Πολιτικών : Πρέπει να υπάρχει η δυνατότητα φιλτραρίσματος των εισερχόμενων και εξερχόμενων διαδρομών , ώστε να μπορούν να εφαρμοστούν πολιτικές.

Σήμερα, το BGP-4 έχει μερικώς επιτύχει τα παραπάνω και έχει καθιερωθεί ως το πρωτόκολλο δρομολόγησης του διαδικτύου για IPv4 / IPv6 διευθύνσεις, καθώς και για την ανταλλαγή πληροφοριών δρομολόγησης σε δίκτυα MPLS VPN (VPNv4).

1.2 Βασικές έννοιες του πρωτοκόλλου BGP

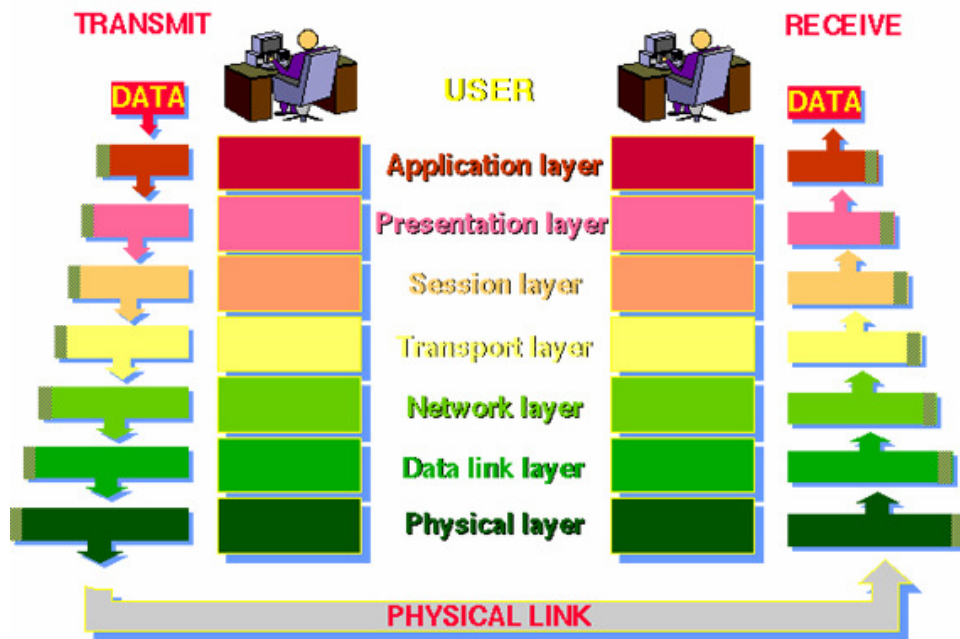
1.2.1 Περίληψη λειτουργίας

Το BGP-4 αποτελεί ένα distance vector πρωτόκολλο δρομολόγησης. Αυτό, σημαίνει ότι το BGP θα διαφημίσει σε άλλους δρομολογητές τα IP δίκτυα που γνωρίζει.

Εάν υπάρχουν διαθέσιμες δύο η περισσότερες διαδρομές προς ένα υποδίκτυο IP, τότε το BGP πρέπει να επιλέξει την καλύτερη δυνατή (Best Path). Όπως όλα τα distance vector πρωτόκολλα, έτσι και το BGP διαθέτει τους δικούς του μηχανισμούς για μέτρηση της απόστασης. Συγκεκριμένα, χρησιμοποιεί τα χαρακτηριστικά διαδρομής (Path Attributes) που χαρακτηρίζουν κάθε εφικτό υποδίκτυο IP.

Το BGP-4 εφαρμόζει αξιόπιστη μέθοδο για την ανταλλαγή μηνυμάτων μεταξύ των δρομολογητών, με αποτέλεσμα ο αποστολέας ενός μηνύματος να γνωρίζει εάν ο παραλήπτης το δέχτηκε. Έτσι δεν απαιτείται περιοδικά η ανταλλαγή όλης της πληροφορίας δρομολόγησης, αλλά μόνο οι μεταβολές. Για την επίτευξη των παραπάνω το BGP χρησιμοποιεί τα πρωτόκολλα TCP (port 179) και IP. Το BGP αποτελεί εφαρμογή και ανήκει στο Application layer του μοντέλου OSI και για να επιτύχει τα παραπάνω χρησιμοποιεί το πρωτόκολλο TCP (port 179) στο Transport layer καθώς και το IP σε Internet layer, βάσει του μοντέλου TCP/IP.

THE 7 LAYERS OF OSI



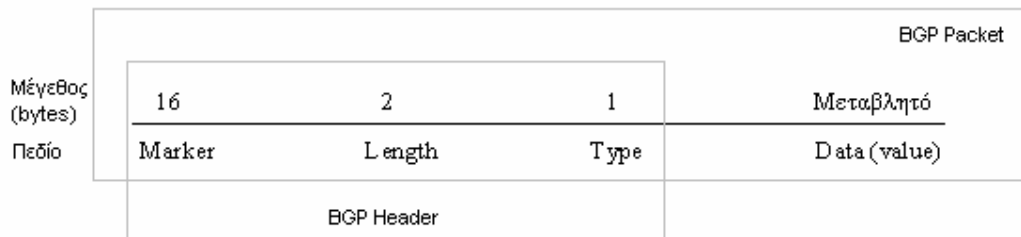
Εικόνα 1.2.1.1 : Τα επίπεδα του μοντέλου OSI
(Πηγή : <http://www.lensenet.com>)

Παρόλο που το TCP ως πρωτόκολλο γνωστοποιεί την ορθή παραλαβή από τον λήπτη του πακέτου εντούτοις, δεν σηματοδοτεί πιθανή αδυναμία επικοινωνίας μεταξύ των συνομιλητών. Βάση αυτού του γεγονότος, ένας δρομολογητής BGP που έχει μάθει κάποιες διαδρομές και τις διαφημίζει σε άλλους πρέπει να γνωρίζει το συντομότερο δυνατό την αποτυχία επικοινωνίας με την πηγή αυτών των διαδρομών ώστε να αποφευχθεί η άστοχη δρομολόγηση πακέτων. Για να επιβεβαιωθεί η συνδεσιμότητα σε επίπεδο TCP, το BGP χρησιμοποιεί μηνύματα επιβεβαίωσης που ονομάζονται Keepalive messages.

Το BGP δημιουργήθηκε με στόχο να μπορεί να διαχειριστεί μεγάλα ποσά IP διαδρομών σε μεγάλα και πολύπλοκα δίκτυα. Σε τέτοιες τοπολογίες, οι συνδέσεις μεταξύ των δρομολογητών (links) πολλές φορές αποτυγχάνουν προκαλώντας αλλαγή τοπολογίας. Σε αυτή την περίπτωση, το πρωτόκολλο απαιτεί την ανταλλαγή μεγάλου αριθμού μηνυμάτων και μεγάλης υπολογιστικής ισχύος με αποτέλεσμα να μπορεί να τεθεί σε κίνδυνο η δυνατότητα διαχείρισης μεγάλου αριθμού δικτύων. Για να αποφευχθεί μια τέτοια περίπτωση, το BGP έχει σχεδιαστεί με στόχο την υποστήριξη πολλών διαδρομών και όχι την γρήγορη αποκατάσταση από αστάθεια (convergence).

1.2.2 Μορφή μηνυμάτων

Το BGP-4 (Rekhter, Y & Li, T & Hares, S. 2003) βασίζεται στο πρωτόκολλο TCP και συγκεκριμένα στο Port 179 . Κάθε μήνυμα BGP βασίζεται στο τρίπτυχο τύπος – μέγεθος – τιμή (TLV) και περιλαμβάνει μια επικεφαλίδα που ονομάζεται BGP header. Το πρώτο πεδίο εύρους 16 Bytes είναι το marker και έχει χαρακτηριστική τιμή ώστε να αναγνωρίζεται ένα πακέτο BGP. Το δεύτερο πεδίο length είναι εύρους 2 bytes και ορίζει το συνολικό μέγεθος του πακέτου. Το τρίτο πεδίο type , έχει μέγεθος 1 byte και ορίζει τον τύπο του BGP μηνύματος.



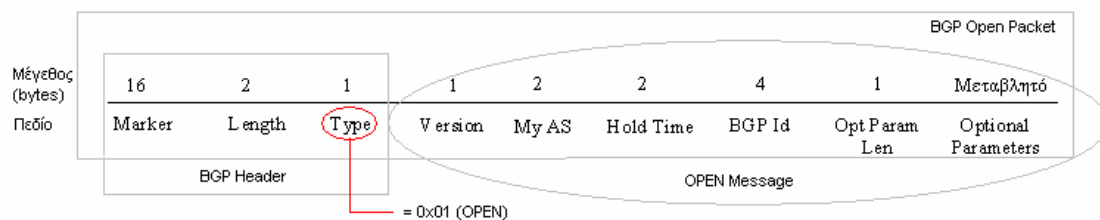
Εικόνα 1.2.2.1 : Το πακέτο BGP και τα πεδία του σε bytes

Οι τύποι μηνυμάτων στο BGP είναι τέσσερις :

- OPEN (type 1)
- UPDATE (type 2)
- NOTIFICATION (type 3)
- KEEPALIVE (type 4)

Κάθε BGP πακέτο ανήκει σε έναν και μόνο τύπο. Συνοπτικά, αναφέρουμε την λειτουργία του κάθε τύπου μηνύματος BGP.

Το OPEN χρησιμοποιείται κατά την έναρξη της συνομιλίας μεταξύ δύο δρομολογητών και έχει στόχο την ανταλλαγή παραμέτρων του πρωτοκόλλου.

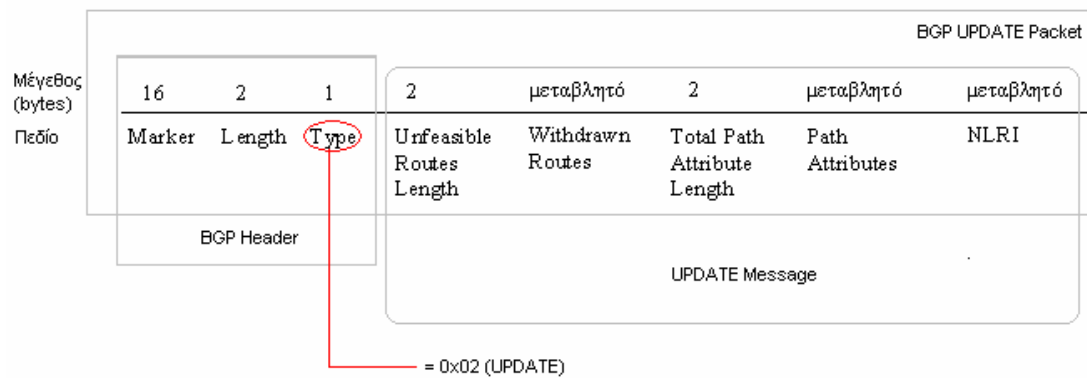


Εικόνα 1.2.2.2 : Το πακέτο BGP Open και τα πεδία του σε bytes

Το πρώτο πεδίο του μηνύματος OPEN είναι 1 byte και ορίζει την έκδοση του πρωτοκόλλου. Το δεύτερο πεδίο είναι 2 bytes και ορίζει τον αριθμό του αυτόνομου συστήματος AS που ανήκει ο δρομολογητής που δημιουργεί αυτό το μήνυμα. Το τρίτο πεδίο Hold Time ορίζει τον χρόνο σε δευτερόλεπτα (2 bytes), ο οποίος αν περάσει χωρίς να έχουμε δεχτεί μήνυμα από τον γειτονικό δρομολογητή, θα θεωρήσουμε μη έγκυρο το γειτονικό δρομολογητή. Επίσης, αυτομάτως, σε πολλές

υλοποιήσεις, ο χρόνος που αποστέλλονται τα μηνύματα τύπου keepalive ορίζεται στο $[0.33 * (\text{Hold Time})]$. Το τέταρτο πεδίο, BGP identifier, είναι 4 bytes και περιλαμβάνει έναν αριθμό χαρακτηριστικό του δρομολογητή. Αυτό πολλές φορές αποτελεί την Unicast IPv4 διεύθυνση που δέχεται μηνύματα BGP ο δρομολογητής. Το έκτο πεδίο Optional Parameters, περιλαμβάνει προαιρετικές παραμέτρους το μέγεθος των οποίων είναι μεταβλητό και ορίζεται στο πέμπτο πεδίο Optional Parameters Length μεγέθους 1 byte.

Το UPDATE χρησιμοποιείται όταν δύο δρομολογητές έχουν εδραιώσει μια συνομιλία BGP (FSM State : ESTABLISHED) και ανταλλάσσουν πληροφορίες δρομολόγησης.

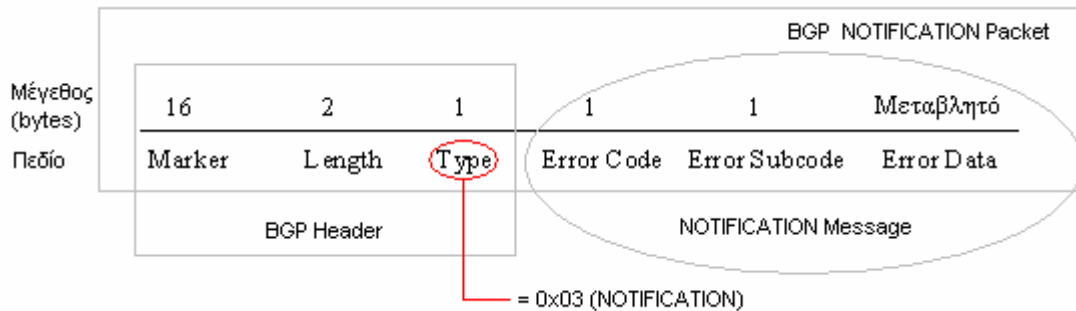


Εικόνα 1.2.2.3 : Το πακέτο BGP Update και τα πεδία του σε bytes

Το πρώτο πεδίο του μηνύματος Update είναι το Unfeasible Routes Length μεγέθους δύο Bytes και περιγράφει το μήκος σε Bytes του δεύτερου πεδίου. Το Withdrawn Routes περιλαμβάνει δυάδες (Length, Prefix) συνολικού μεγέθους πέντε bytes, εκ των οποίων ένα το length και τέσσερα το prefix. Οι διαδρομές αυτές διαφημίζονται στο γειτονικό δρομολογητή με στόχο την απόρριψη τους. Εν συνεχεία το τρίτο πεδίο, Total Path Attribute Length (2 bytes), ορίζει το μέγεθος σε bytes των χαρακτηριστικών που ορίζονται στο τέταρτο πεδίο, Path Attributes. Το πεδίο Path Attributes περιέχει χαρακτηριστικά τα οποία θα συζητηθούν παρακάτω (κεφ 1.2.3) και είναι κοινά για το πλήθος διαδρομών που διαφημίζονται σε ένα μήνυμα UPDATE. Το πέμπτο πεδίο Network Layer Reachability Information, περιλαμβάνει tuples (Length, Prefix), όπως ακριβώς και το Withdrawn Routes, ενώ το μέγεθος αυτού ορίζεται από το συνολικό μέγεθος του μηνύματος που ορίζει το BGP Header.

Ένα Update μήνυμα μπορεί να εισαγάγει νέες διαδρομές με κοινά χαρακτηριστικά και ταυτόχρονα να αποσύρει άλλες.

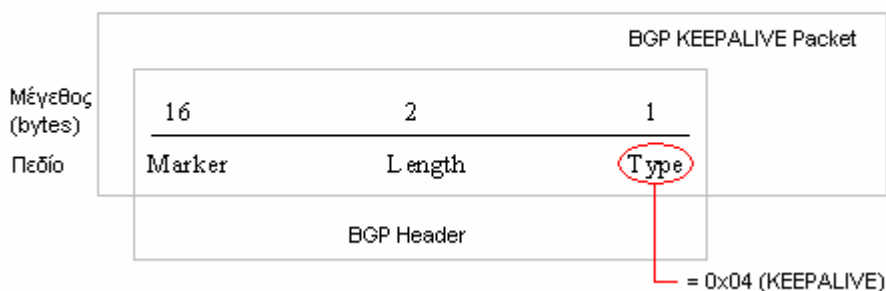
Το μήνυμα NOTIFICATION χρησιμοποιείται για τον εντοπισμό σφαλμάτων ή ασυμβατότητας μεταξύ της υλοποίησης BGP των δύο δρομολογητών που βρίσκονται σε σύνοδο. Μετά την αποστολή τέτοιου μηνύματος τερματίζεται η συνομιλία BGP.



Εικόνα 1.2.2.4 : Το πακέτο BGP Notification και τα πεδία του σε bytes

Το μήνυμα NOTIFICATION αποτελείται από τρία πεδία. Τα δύο πρώτα, ονομάζονται Error Code και Error Subcode με μέγεθος 1 byte έκαστο και αντιπροσωπεύουν τον κωδικό σφάλματος που είναι προσυμφωνημένος. Το τρίτο πεδίο μεταφέρει τα «ένοχα» δεδομένα. Έτσι, δίνεται η δυνατότητα να εξεταστεί λεπτομερώς ο λόγος τερματισμού της συνόδου.

Το μήνυμα τύπου KEEPALIVE, είναι το απλούστερο μήνυμα BGP. Ο λόγος ύπαρξης του είναι η επιβεβαίωση της καλής λειτουργίας της BGP συνόδου μεταξύ δύο δρομολογητών. Το TCP που χρησιμοποιείται, διαθέτει παρόμοια λειτουργία, η οποία όμως επιβεβαιώνει την καλή λειτουργία μόνο μέχρι και το Transport Layer του μοντέλου TCP/IP, οπότε δεν χρησιμοποιείται. Ο ρυθμός αποστολής μηνυμάτων keealive καθορίζεται ως το 33% του μικρότερου Hold Time μεταξύ των δύο δρομολογητών που δημιουργούν την σύνοδο BGP. Σε περίπτωση που η μια πλευρά BGP διαφημίζει Hold Time = 0, τότε δεν χρησιμοποιούνται keealive μηνύματα.



Εικόνα 1.2.2.5 : Το πακέτο BGP Keepalive και τα πεδία του σε bytes

Το μήνυμα KEEPALIVE αποτελεί απλώς ένα BGP Header με Type = 0x04, και δεν περιέχει επιπλέον πεδία ή δεδομένα. Συνεπώς έχει σταθερό μέγεθος 19 bytes.

1.2.3 Χαρακτηριστικά διαδρομών(Path Attributes)

Έχοντας καλύψει τους τύπους μηνυμάτων του BGP είναι επιβεβλημένο να αναφερθούμε στα χαρακτηριστικά διαδρομών (BGP Path Attributes) που μεταδίδονται μέσω μηνύματος τύπου Update. Τα χαρακτηριστικά χωρίζονται σε τέσσερις κατηγορίες :

- Γνωστά Υποχρεωτικά
- Γνωστά Διακριτικά
- Προαιρετικό Μεταβατικό
- Προαιρετικό μη Μεταβατικό

Τα γνωστά χαρακτηριστικά είναι αυτά που αναφέρονται στον ορισμό του πρωτοκόλλου BGP-4 (Rekhter, Y & Li, T & Hares, S. 2003) , ενώ τα προαιρετικά ορίζονται σε διάφορες προεκτάσεις του πρωτοκόλλου π.χ (Chandra, R & Scudder, J.(2002)) και μπορεί να αγνοηθούν από ένα δρομολογητή που δεν υλοποιεί τις προεκτάσεις του πρωτοκόλλου.

Τα υποχρεωτικά χαρακτηριστικά είναι αυτά που πρέπει να περιλαμβάνονται σε κάθε μήνυμα τύπου Update που διαφημίζει νέα IP υποδίκτυα. Ενώ τα διακριτικά χαρακτηριστικά μπορεί να περιλαμβάνονται ή όχι (Εικ 1.2.3.1).

Τα προαιρετικά χαρακτηριστικά χωρίζονται στα μεταβατικά και μη. Όταν ένας δρομολογητής δέχεται ένα μήνυμα Update που περιλαμβάνει κάποιο προαιρετικό χαρακτηριστικό, το οποίο συνοδεύει κάποιο IP υποδίκτυο (NLRI) και η διαδρομή αυτή επιλέγεται για διαφήμιση σε άλλους BGP συνομιλητές, τότε υπάρχουν δύο επιλογές. Εάν το χαρακτηριστικό είναι μαρκαρισμένο ως μεταβατικό, προωθείται στα εξερχόμενα Update μηνύματα, ενώ στην αντίθετη περίπτωση, σιωπηλά αφαιρείται το συγκεκριμένο χαρακτηριστικό και διαφημίζεται το IP υποδίκτυο χωρίς το χαρακτηριστικό αυτό.

Γνωστά χαρακτηριστικά	eBGP	iBGP
ORIGIN	+	+
AS PATH	+	+
NEXT HOP	+	+
MED	-	-
LOCAL PREF	-	+
ATOMIC AGGREGATE	-	-
AGGREGATOR	-	-

+ Υποχρεωτικό
- Διακριτικό

Εικόνα 1.2.3.1 : Τα γνωστά χαρακτηριστικά διαδρομής του BGP

Τα χαρακτηριστικά διαδρομών περιλαμβάνονται σε Update μηνύματα και χαρακτηρίζονται με τρίπτυχα (T-L-V). Κάθε χαρακτηριστικό μπορεί να περιλαμβάνεται το πολύ μία φορά ανά Update μήνυμα. Παρακάτω, γίνεται επεξήγηση των γνωστών χαρακτηριστικών καθώς και κάποιων περαιτέρω που θεωρούνται σημαντικά για την λειτουργία του πρωτοκόλλου BGP.

Το ORIGIN είναι χαρακτηριστικό το οποίο δημιουργεί ο δρομολογητής που διαδίδει την πληροφορία NLRI και παίρνει τρεις τιμές:

- IGP – το δίκτυο που διαφημίζεται ανήκει στο AS του δρομολογητή που ξεκινά την διάδοση
- EGP – το δίκτυο που διαφημίζεται έγινε γνωστό μέσω eBGP
- Incomplete – η προέλευση του δικτύου είναι άγνωστη.

Το AS_PATH, επιγραμματικά, αποτελεί τη διαδρομή αυτόνομων συστημάτων (AS) που πρέπει να διασχίσει ο δρομολογητής για να επικοινωνήσει με τα υποδίκτυα που διαφημίζονται στο μήνυμα Update (NLRI). Όταν ένας δρομολογητής στέλνει Update μέσω eBGP, δηλαδή σε άλλο AS, προσθέτει στη λίστα AS_Path το δικό του αριθμό AS.

Το NEXT_HOP είναι η unicast IP διεύθυνση που πρέπει να χρησιμοποιηθεί ως πρώτος προορισμός των πακέτων τα οποία προορίζονται για κάποιο IP υποδίκτυο που περιλαμβάνεται στη λίστα NLRI του μηνύματος Update. Στο iBGP που θα επεξηγήσουμε παρακάτω (κεφ.1.4), το Next Hop δεν πρέπει να μεταβάλλεται από τους δρομολογητές καθώς η πληροφορία διαδίδεται εντός του AS. Στο eBGP (κεφ.1.3) το Next Hop, προκαθορισμένα, παίρνει την τιμή της IP διεύθυνσης του συνοριακού δρομολογητή που δημιουργεί το Update. Η IP διεύθυνση που ορίζει το Next Hop πρέπει να είναι γνωστή μέσω κάποιου IGP πρωτοκόλλου δρομολόγησης ή μέσω στατικής διαδρομής, αλλιώς το Update απορρίπτεται.

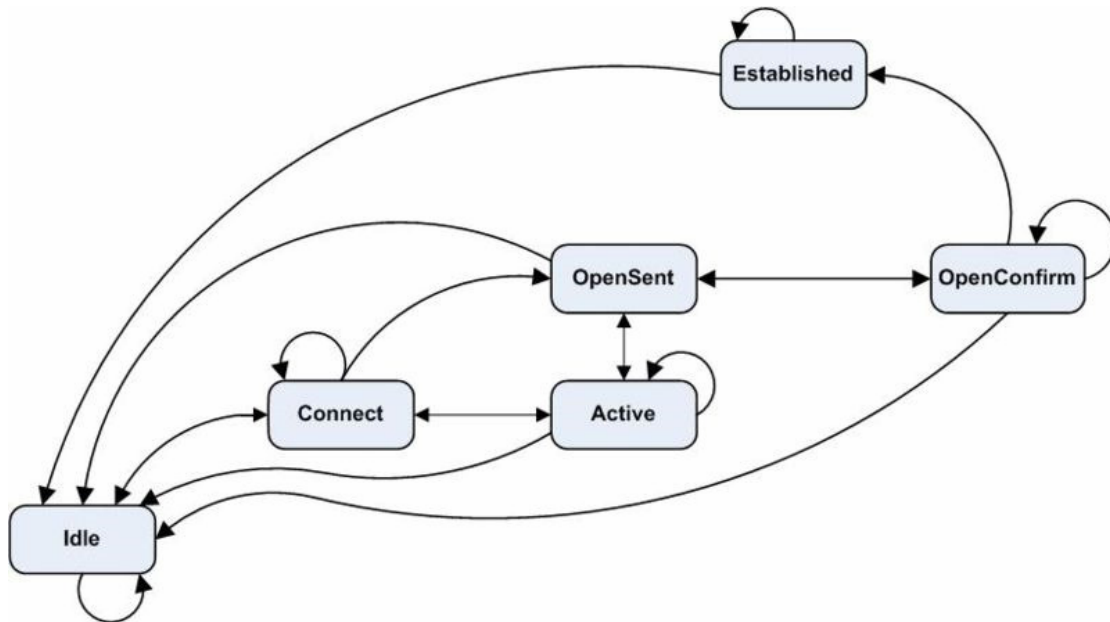
Το Multi Exit Discriminator (MED) συνιστά ένα 4-byte αριθμό ο οποίος αποτελεί πρόταση για το γειτονικό AS. Για παράδειγμα, όταν δύο AS συνορεύουν σε δύο ή περισσότερα σημεία, το AS που αποστέλλει BGP updates μέσω των δύο συνόρων, χρησιμοποιεί μικρότερο MED στο ένα από τα δύο για να δείξει την προτίμηση του να δεχτεί εισερχόμενη κίνηση από τον συνοριακό δρομολογητή με το υψηλό MED. Παρόλα αυτά, το AS προορισμού του Update μηνύματος μπορεί να αγνοήσει το MED.

Έτσι λοιπόν αντιθέτως με το MED, το Local Preference, αποτελεί πρόταση εντός του AS για επιλογή κάποιου συγκεκριμένου σημείου εξόδου προς κάποιο AS στην περίπτωση που υπάρχουν περισσότερα του ενός. Το Local Preference είναι χαρακτηριστικό που διαδίδεται αποκλειστικά εντός του AS μέσω iBGP.

Το Aggregator και Atomic Aggregate χρησιμοποιούνται για την συνάθροιση υποδικτύων που έχουν κοινά χαρακτηριστικά και συμβάλλει καθοριστικά στη δυνατότητα του πρωτοκόλλου να υπηρετήσει τις τεράστιες απαιτήσεις του διαδικτύου.

1.2.4 BGP Finite State Machine

Το BGP, ως εφαρμογή, μπορεί να βρεθεί σε πεπερασμένο αριθμό καταστάσεων (Finite State Machine). Οι έξι καταστάσεις που μπορεί να βρεθεί (Εικ.1.2.4.1) επεξηγούνται παρακάτω.



Εικόνα 1.2.4.1 : BGP Finite State Machine

(Πηγή : <http://upload.wikimedia.org/>)

Αρχικά το BGP βρίσκεται στην κατάσταση **Idle** κατά την οποία το BGP αρνείται όλες τις εισερχόμενες αιτήσεις για σύνδεση BGP. Όταν δημιουργηθεί το γεγονός έναρξης της εφαρμογής, αυτή προσπαθεί να συνδεθεί στον απομακρυσμένο BGP γείτονα, ενώ ταυτοχρόνως περιμένει για εισερχόμενες αιτήσεις πάντα σε επίπεδο TCP, μεταβαίνοντας σε κατάσταση **Connect**.

Εάν δεν πραγματοποιηθεί σύνδεση εντός ConnectRetry Timer διαστήματος ή ανιχνευτεί κάποιο σφάλμα, τότε το σύστημα παραμένει σε κατάσταση **Idle** και επανεκκινεί την διαδικασία μετά από χρόνο 60 sec.

Εν συνεχεία, επιχειρείται σύνδεση σε TCP επίπεδο και το BGP μεταβαίνει στην κατάσταση **Connect**. Επόμενο βήμα είναι η αποστολή μηνύματος OPEN στον BGP γείτονα και η μετάβαση σε κατάσταση **Opensent**. Αν ωστόσο η σύνδεση TCP αποτύχει, το BGP μεταβαίνει σε κατάσταση **Active**. Αν λήξει ο ConnectRetry Timer επανεκκινείται η διαδικασία και το σύστημα παραμένει την κατάσταση **Connect**. Σε περίπτωση σφάλματος, το BGP μεταβαίνει σε κατάσταση **Idle**.

Στην κατάσταση **Active**, το BGP προσπαθεί σε επίπεδο TCP να συνδεθεί με το γειτονικό δρομολογητή. Εάν αυτό επιτευχθεί, το BGP στέλνει OPEN στο γείτονα και επανεκκινεί τον timer ConnectRetry, ενώ θέτει τον hold timer σε μεγάλη τιμή (4min). Σε περίπτωση λήξης του timer ConnectRetry, μεταβαίνει στην κατάσταση **Connect**. Εάν διαπιστωθεί πως κάποιο γειτονικό σύστημα προσπαθεί να συνδεθεί με

μη αναμενόμενη IP διεύθυνση, τότε το BGP απορρίπτει την σύνδεση και παραμένει στην κατάσταση **Active**. Σε οποιαδήποτε άλλη περίπτωση η εφαρμογή επιστρέφει σε κατάσταση **Idle**.

Στην κατάσταση **OpenSent** η εφαρμογή περιμένει να παραλάβει μήνυμα τύπου OPEN από τον γειτονικό δρομολογητή. Εάν, ωστόσο, το εισερχόμενο μήνυμα δεν είναι τύπου OPEN ή τα πεδία του μηνύματος OPEN είναι εσφαλμένα, αποστέλλεται μήνυμα NOTIFICATION με τον ανάλογο κωδικό σφάλματος και η μηχανή επιστρέφει σε κατάσταση **Idle**. Επίσης, εάν παρέλθει διάστημα Hold Time και δεν έχει γίνει παραλαβή μηνύματος από τον γειτονικό δρομολογητή, αποστέλλεται μήνυμα NOTIFICATION και η μηχανή επιστρέφει σε κατάσταση **Idle**. Εάν ανιχνευτεί σφάλμα σε επίπεδο TCP, η μηχανή επιστρέφει σε κατάσταση **Active**. Στην περίπτωση που το OPEN μήνυμα που παραλάβαμε είναι σωστό, αποστέλλουμε μήνυμα τύπου Keepalive και ορίζουμε το hold Time ως το μικρότερο που περιλαμβάνεται στα μηνύματα OPEN που ανταλλάχθηκαν. Αναλόγως, ορίζεται Keepalive timer = 0.33% * (Hold Time) ως ο χρόνος αποστολής μηνυμάτων τύπου Keepalive στο συγκεκριμένο γείτονα. Εάν ο αριθμός αυτόνομου συστήματος AS είναι ίδιος στα μηνύματα OPEN που ανταλλάχθηκαν, τότε η σύνδεση ορίζεται ως εσωτερική iBGP (internal BGP), ενώ στην περίπτωση που διαφέρουν, η σύνδεση μαρκάρεται ως eBGP (external BGP). Τελικώς, μεταβαίνουμε σε κατάσταση **OpenConfirm**.

Στην κατάσταση **OpenConfirm**, το BGP περιμένει να παραλάβει μήνυμα τύπου KEEPALIVE ή NOTIFICATION. Εάν η εφαρμογή παραλάβει μήνυμα KEEPALIVE, τότε μεταβαίνει σε κατάσταση **Established**. Εάν παρέλθει διάστημα Hold Time και δεν έχει γίνει παραλαβή μηνύματος από τον γειτονικό δρομολογητή, αποστέλλεται μήνυμα NOTIFICATION και η μηχανή επιστρέφει σε κατάσταση **Idle**. Εάν το μήνυμα που παραλαμβάνει η εφαρμογή είναι τύπου NOTIFICATION ή υπάρξει σφάλμα η μηχανή πηγαίνει σε κατάσταση **Idle**.

Στην κατάσταση **Established** το BGP μπορεί να ανταλλάξει μηνύματα τύπου UPDATE, KEEPALIVE ή NOTIFICATION. Κάθε φορά που γίνεται παραλαβή μηνύματος UPDATE, KEEPALIVE γίνεται επανεκκίνηση του Hold Time. Σε περίπτωση παραλαβής μηνύματος NOTIFICATION το BGP επιστρέφει σε κατάσταση **Idle**. Εάν διαπιστωθεί σφάλμα σε μήνυμα UPDATE ή λήξει ο Hold Timer, το BGP αποστέλλει NOTIFICATION με το ανάλογο code / subcode και μεταβαίνει σε κατάσταση **Idle**. Εάν λήξει ο Keepalive timer, το BGP αποστέλλει μήνυμα τύπου KEEPALIVE.

Κάθε φορά που το BGP μεταβαίνει από τις καταστάσεις **OpenSent**, **OpenConfirm** και **Established** σε **Idle**, γίνεται επανεκκίνηση της διασύνδεσης TCP, καθώς και εκκαθάριση όλων των μεταβλητών και διαδικασιών που αφορούν την συγκεκριμένη σύνδεση.

Το BGP διατηρεί μια μηχανή FSM ανά σύνδεση, δηλαδή μπορεί να είναι σε κατάσταση Idle για μια σύνδεση και σε Established για μια άλλη. Ο αριθμός των συνολικών συνδέσεων αυξάνει την απαίτηση σε υπολογιστική ισχύ και μνήμη και κατά συνέπεια δημιουργεί προβλήματα που απασχολούν μέχρι και σήμερα την επιστημονική κοινότητα.

1.2.5 BGP Path Selection

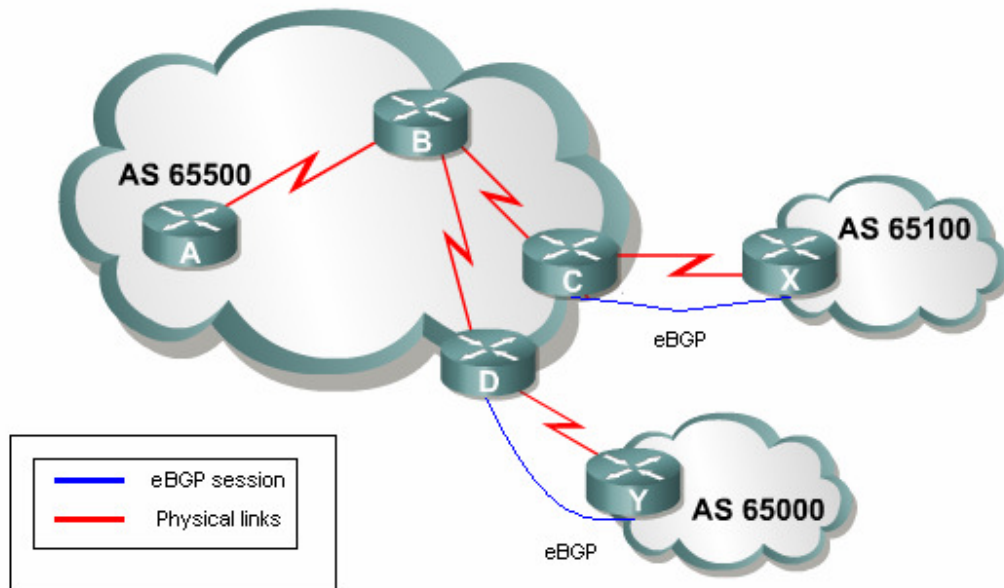
Όταν το BGP γνωρίζει πολλές διαδρομές για ένα συγκεκριμένο υποδίκτυο επιλέγει την καλύτερη και την τοποθετεί στον πίνακα δρομολόγησης. Προκειμένου να γίνει η επιλογή καλύτερης διαδρομής, ακολουθείται η εξής διαδικασία όπως περιγράφεται παρακάτω. Εάν σε κάποιο βήμα υπάρχει ισοπαλία προχωράμε στο επόμενο μέχρι να επιλεγεί μια διαδρομή.

- Εάν μία διαδρομή έχει χαρακτηριστικό NEXT_HOP μη προσβάσιμο τότε απορρίπτεται.
- Επιλέγεται η διαδρομή με το μεγαλύτερο αριθμητικά LOCAL_PREF.
- Επιλέγεται η διαδρομή που δημιουργήθηκε από το BGP που εκτελείται στο τοπικό δρομολογητή.
- Επιλέγεται η διαδρομή με την μικρότερη σε μήκος λίστα AS_PATH.
- Επιλέγεται η διαδρομή με το μικρότερο χαρακτηριστικό ORIGIN, όπου IGP < EGP < Incomplete(?).
- Επιλέγουμε τη διαδρομή με το μικρότερο αριθμητικά MED.
- Επιλέγεται η διαδρομή που έγινε γνωστή μέσω eBGP και όχι iBGP.
- Εάν οι διαδρομές είναι ακόμη ισόπαλες, προτιμούμε τη διαδρομή που έγινε γνωστή μέσω του πλησιέστερου IGP γείτονα βάσει των metrics που ορίζει το IGP που χρησιμοποιείται εντός του AS.
- Τέλος επιλέγεται η διαδρομή που διαφημίστηκε από το δρομολογητή με το μικρότερο BGP ID.

1.3 Εξωτερικό BGP (eBGP)

Το BGP ως πρωτόκολλο δρομολόγησης, διαθέτει δύο ειδών διασυνδέσεις μεταξύ δρομολογητών για ανταλλαγή πληροφοριών δηλαδή, το internal BGP και το external BGP (eBGP).

Το eBGP χρησιμοποιείται κατά την διασύνδεση δρομολογητών (border routers) που βρίσκονται σε διαφορετικά αυτόνομα συστήματα (AS). Ουσιαστικά, το eBGP χειρίζεται με διαφορετικό τρόπο τα χαρακτηριστικά διαδρομών. Για παράδειγμα, όταν διαφημίζεται ένα υποδίκτυο NLRI μέσω eBGP, ο αριθμός του δικού μας AS πρέπει να προστεθεί στην λίστα AS_PATH. Στόχος του eBGP είναι η διάδοση της πληροφορίας δρομολόγησης μεταξύ των αυτόνομων συστημάτων και όχι εντός αυτών.



Εικόνα 1.3 : BGP τοπολογία με εμφανή την σύνδεση eBGP
(Πηγή Cisco Networking Academy v.1.0)

1.4 Εσωτερικό BGP (iBGP)

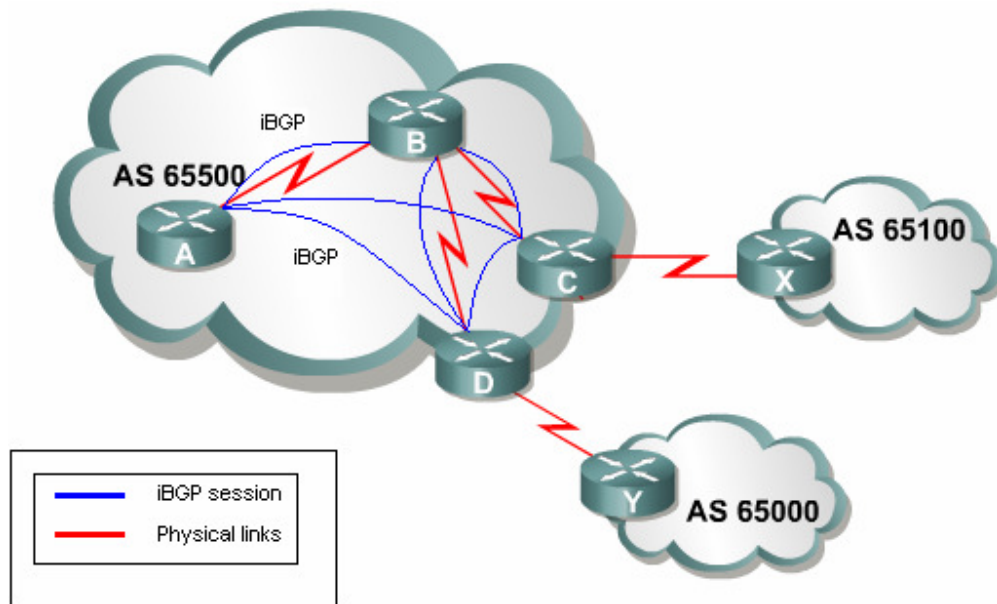
1.4.1 Περιγραφή του iBGP

Το iBGP, σε αντίθεση με το eBGP, είναι εκείνο το είδος BGP που εκτελεί ο δρομολογητής, όταν ανταλλάσει πληροφορίες δρομολόγησης με δρομολογητή του ιδίου AS. Για να μπορέσει να λειτουργήσει το iBGP, πρέπει να λειτουργεί κάποιο πρωτόκολλο IGP, αφού το iBGP δεν διαθέτει τέτοιους μηχανισμούς.

Σκοπός του iBGP είναι η διάχυση της πληροφορίας δρομολόγησης εντός του AS, καθώς και η διαφήμιση διαδρομών μέσω iBGP στους συνοριακούς δρομολογητές, ώστε να γίνει περαιτέρω η μετάδοση αυτών σε άλλα AS.

Το iBGP, εισάγει έναν βασικό κανόνα αποφυγής δημιουργίας βρόχων δρομολόγησης που απαγορεύει την διαφήμιση διαδρομών σε γείτονα iBGP όταν αυτές έγιναν γνωστές μέσω γείτονα iBGP. Ο κανόνας αυτός εγγυάται την ομαλή λειτουργία του πρωτοκόλλου. Ταυτόχρονα, όμως, εισάγει την ανάγκη για λογική διασύνδεση των δρομολογητών εντός ενός AS σε πλήρες πλέγμα. Έτσι, σε μεγάλα αυτόνομα συστήματα ο αριθμός συνδέσεων iBGP ,που πρέπει να διατηρεί κάθε δρομολογητής, γίνεται απαγορευτικός. Ενδεικτικά, αναφέρεται ότι για αυτόνομο σύστημα με n δρομολογητές απαιτούνται $[n*(n-1)/2]$ λογικές iBGP συνδέσεις. Η επιστημονική κοινότητα έδωσε λύσεις σε αυτό το πρόβλημα που ανέκυψε κατά την ανεξέλεγκτη επέκταση του διαδικτύου χωρίς, όμως, να καταφέρει να υποκαταστήσει την σταθερότητα της iBGP δικτύωσης πλήρους πλέγματος.

Στην συνέχεια παρουσιάζονται πολλές προτάσεις καθώς και τα πλεονεκτήματα και μειονεκτήματα τους.



Εικόνα 1.4.1 : BGP τοπολογία με εμφανή την σύνδεση iBGP σε πλήρες πλέγμα (Πηγή Cisco Networking Academy v.1.0)

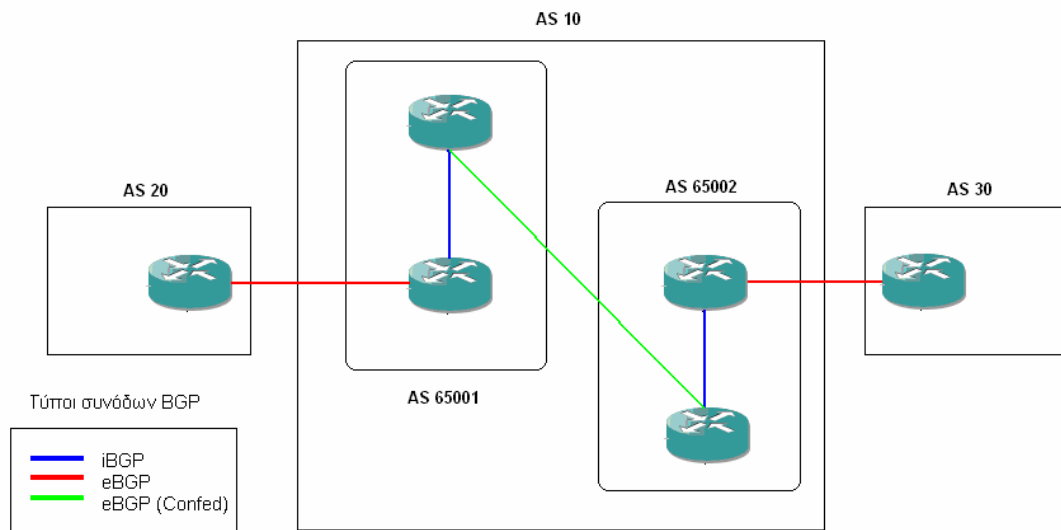
1.4.2 Βελτίωση μέσω AS Confederations

Λύση στο πρόβλημα που παρουσιάζεται στο (κεφ 1.4.1) αποτελούν τα AS Confederations (Traina, P & McPherson, D & Scudder, J. (2001)). Για να γίνει εφικτή η μείωση των λογικών συνδέσεων iBGP προτείνεται η εσωτερική διαίρεση ενός αυτόνομου συστήματος σε πολλά.

Κάνοντας χρήση των ιδιωτικών αριθμών AS συστημάτων (64532-65516), ορίζουμε αυθαίρετα ομάδες δρομολογητών σε κάποιο από αυτά και τα διασυνδέουμε σε ένα μεγαλύτερο που αποτελεί το πραγματικό AS. Η σηματοδότηση εντός των μικρών ιδιωτικών AS παραμένει iBGP. Τα μικρά ιδιωτικά AS επικοινωνούν με μια μεταλλαγμένη εκδοχή του eBGP που υποστηρίζει δύο νέα χαρακτηριστικά διαδρομής, το AS_CONFED_SET και το AS_CONFED_SEQUENCE, για την μεταφορά και διαχωρισμό των ιδιωτικών αριθμών AS από τα πραγματικά. Οι δρομολογητές που ανήκουν σε AS Confederation χρησιμοποιούν κλασικό eBGP στις συνομιλίες με δρομολογητές άλλων AS, αναιρώντας τις πληροφορίες ιδιωτικών AS που χρησιμοποιήθηκαν για να δημιουργηθεί η συμμαχία.

Για να αποφευχθούν οι βρόχοι δρομολόγησης, κάθε δρομολογητής που παραλαμβάνει μήνυμα Update, το οποίο που περιλαμβάνει το δικό του ιδιωτικό AS στο AS_PATH, αγνοεί το Update.

Λόγω της τεχνητής μείωσης των δρομολογητών εντός κάθε ιδιωτικού πλέον AS, επιτυγχάνεται δραματική μείωση των συνολικών BGP συνδέσεων εντός του πραγματικού AS.



Εικόνα 1.4.2 : BGP τοπολογία με χρήση AS Confederation

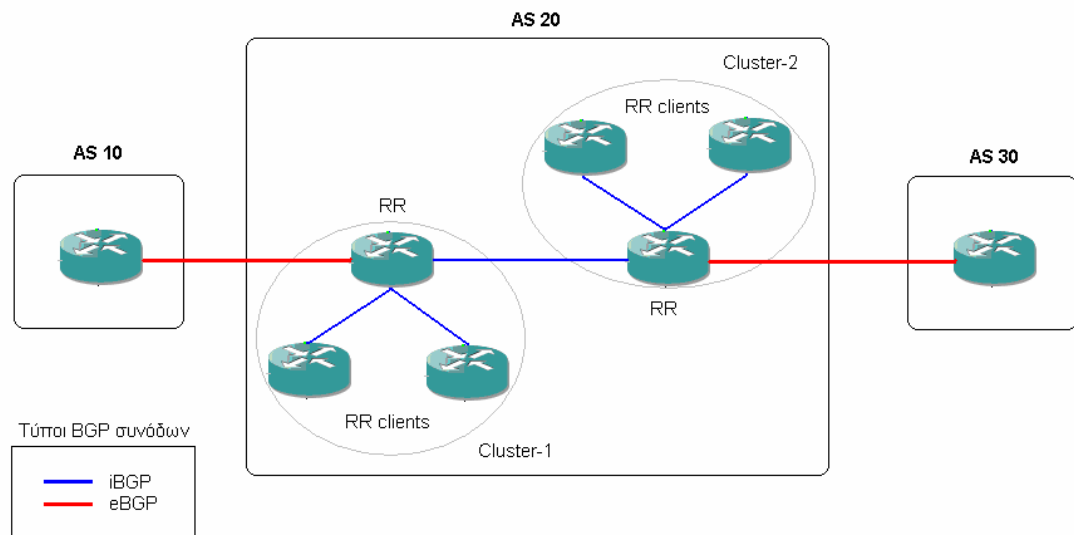
Στον αντίποδα, τα AS Confederations απαιτούν από το σύνολο των δρομολογητών που ανήκουν στο πραγματικό AS να γνωρίζουν τα νέα χαρακτηριστικά διαδρομής που εισήγαγε. Το γεγονός αυτό αποτέλεσε τον κύριο λόγο της ελάχιστης υλοποίησης της λύσης αυτής στον πραγματικό κόσμο. Επίσης, το γεγονός ότι μεταβάλλεται το μήκος της λίστας AS_PATH, λόγω της εισαγωγής των ιδιωτικών αριθμών AS, είναι πιθανό καθώς αυτή μεταδίδεται εντός του πραγματικού AS να δημιουργηθούν σφάλματα σε περίπτωση σύγκρισης AS_PATH από την διαδικασία επιλογής καλύτερης διαδρομής. Παρομοίως, σε περίπτωση σφάλματος κάποιου δρομολογητή μπορεί να υπάρξουν ανυπολόγιστες παρενέργειες.

1.4.3 Βελτίωση μέσω Route Reflection

Παρομοίως με προηγούμενες λύσεις (κεφ.1.4.2) στο πρόβλημα της απαίτησης για τοπολογία πλήρους πλέγματος (κεφ.1.4.1), παρουσιάστηκε η λύση του Route Reflection (Bates, T & Chen, E & Chandra, R.(2006)) και γρήγορα έγινε η δημοφιλέστερη υλοποίηση iBGP σε μεγάλα AS.

Το Route Reflection, στην ουσία αποτελεί παραβίαση των κανόνων split-horizon που επιτάσσει η αρχική ιδέα του BGP-4 (Rekhter, Y & Li, T & Hares, S. 2003)). Πιο συγκεκριμένα, το Route Reflection, χωρίζει το αυτόνομο σύστημα σε ομάδες που ονομάζονται clusters. Σε κάθε cluster υπάρχει τουλάχιστον ένας δρομολογητής που υλοποιεί την επέκταση του BGP, Route Reflection. Οι υπόλοιποι δρομολογητές μπορεί να έχουν άγνοια του Route Reflection και ονομάζονται Route

Reflector clients. Η ιδιαιτερότητα του Route Reflector είναι ότι μπορεί να διαφημίζει (reflect) διαδρομές που έμαθε μέσω iBGP σε δρομολογητές με τους οποίους επικοινωνεί μέσω iBGP. Έτσι, για να διαδοθούν οι γνωστές διαδρομές εντός cluster, αρκεί να υπάρχει μια διασύνδεση iBGP ανά δρομολογητή με τον Route Reflector. Ταυτόχρονα, οι Route Reflectors που ανήκουν σε διαφορετικά clusters, πρέπει να συνδεθούν με iBGP σε τοπολογία πλήρους πλέγματος.



Εικόνα 1.4.3 : BGP τοπολογία με χρήση Route Reflection

Από τα παραπάνω είναι εμφανής η δυνατότητα μείωσης των απαιτούμενων iBGP διασυνδέσεων εντός του AS, όταν γίνεται χρήση Route Reflection (Εικ 1.4.3).

ΚΕΦΑΛΑΙΟ 2

Route Reflection

2.1 Περιγραφή λειτουργίας Route Reflection

Το Route Reflection αποτελεί την σημαντικότερη εναλλακτική λύση στην τοπολογία πλήρους πλέγματος iBGP. Βασικό μειονέκτημα του AS Confederations (κεφ.1.4.2) που οδήγησε σε αυτήν την εξέλιξη είναι η απαίτηση όλοι οι δρομολογητές να γνωρίζουν τα νέα χαρακτηριστικά που εισάγει το AS Confederation. Το iBGP Route Reflection και τα μειονεκτήματά του αποτελούν λόγο της συγκεκριμένης διπλωματικής εργασίας

Για τους παραπάνω λόγους, επεξηγούνται διάφορες παράμετροι που σε προηγούμενα κεφάλαια (κεφ 1.4.3) δεν αναφέρθηκαν.

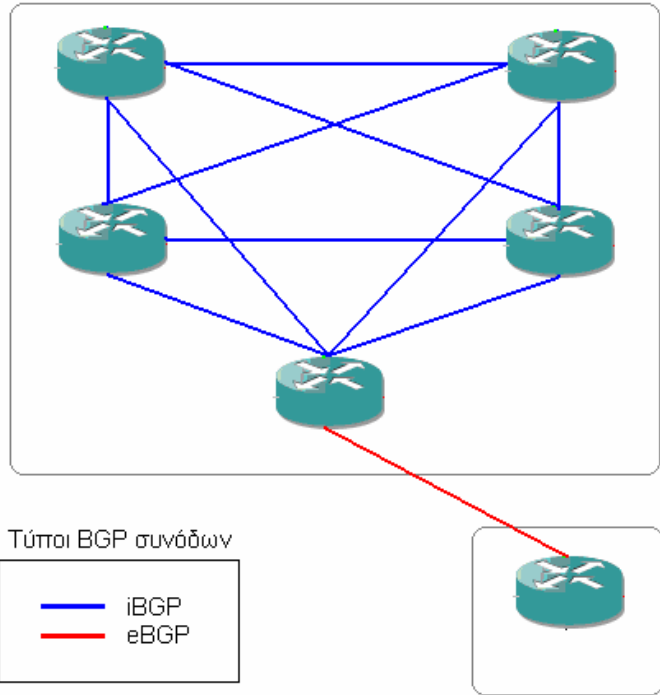
Εντός του αυτόνομου συστήματος, δεν υπάρχει απαίτηση όλοι οι δρομολογητές να ανήκουν σε κάποιο cluster, αλλά πολλοί μπορεί να βρίσκονται σε τοπολογία πλήρους πλέγματος μεταξύ τους και με τους RR (Route Reflectors).

Το Route Reflection εισάγει δύο νέα χαρακτηριστικά BGP που κυκλοφορούν μόνο εντός του cluster. Το **originator-id**, είναι αριθμητικό πεδίο 4-bytes που χαρακτηρίζει το cluster με μοναδικό αριθμό εντός του AS. Συνήθως το cluster-id ορίζεται μέσω παραμετροποίησης ως η IPv4 διεύθυνση του δρομολογητή που δρα ως RR εντός του Cluster. Το Cluster-list, αποτελεί μια λίστα από καταχωρήσεις originator-id (4-bytes). Σκοπός της λίστας αυτής είναι όταν ένας RR παραλάβει μια διαφήμιση (Update) και αναγνωρίσει το δικό του cluster-id να απορρίψει το μήνυμα, ώστε να μη δημιουργηθεί βρόχος δρομολόγησης.

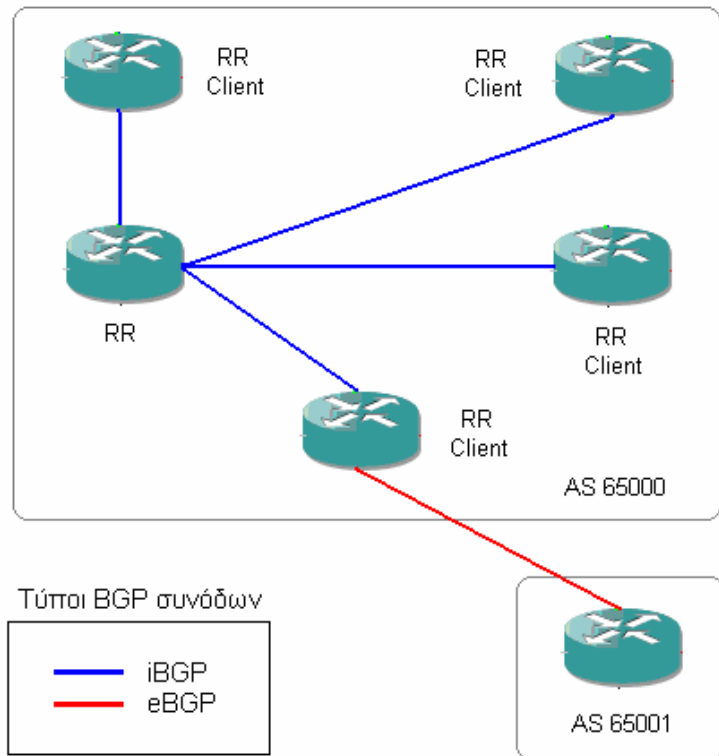
Κάνοντας χρήση του Route Reflection, μπορεί για ένα αυτόνομο σύστημα με n δρομολογητές να μειώσουμε τον αριθμό των απαιτούμενων iBGP συνδέσεων σε $[n-1]$, από $[n*(n-1)/2]$, αν υποθέσουμε ότι γίνεται χρήση ενός RR και ενός cluster εντός ολόκληρου του AS.

Στο παρακάτω παράδειγμα (Εικ.2.1.1) ορίζουμε $n=5$ για τους 5 δρομολογητές του AS 65000. Στην συγκεκριμένη τοπολογία BGP δρομολογητών γίνεται διασύνδεση πλήρους πλέγματος. Έτσι, βάσει του παραπάνω τύπου υπολογίζουμε $[5*(5-1)/2] = 10$ συνδέσεις iBGP στο AS. Έτσι επαληθεύεται και θεωρητικά η παρακάτω τοπολογία (Εικ.2.1.1).

Παρομοίως, (Εικ.2.1.2) χρησιμοποιούμε την ίδια τοπολογία, αλλά αυτή τη φορά επιλέγουμε να κάνουμε χρήση του Route Reflection. Έτσι βάσει του παραπάνω τύπου υπολογίζουμε $[5-1] = 4$ συνδέσεις iBGP στο AS. Η βελτίωση είναι εμφανής, καθώς λόγω της μείωσης των συνδέσεων iBGP μπορούμε να χρησιμοποιήσουμε δρομολογητές με μικρότερη υπολογιστική ισχύ και λιγότερη μνήμη. Επίσης, επιτυγχάνεται μείωση της χρήσης bandwidth από το BGP.



Εικόνα 2.1.1 : BGP τοπολογία με Full mess διασύνδεση



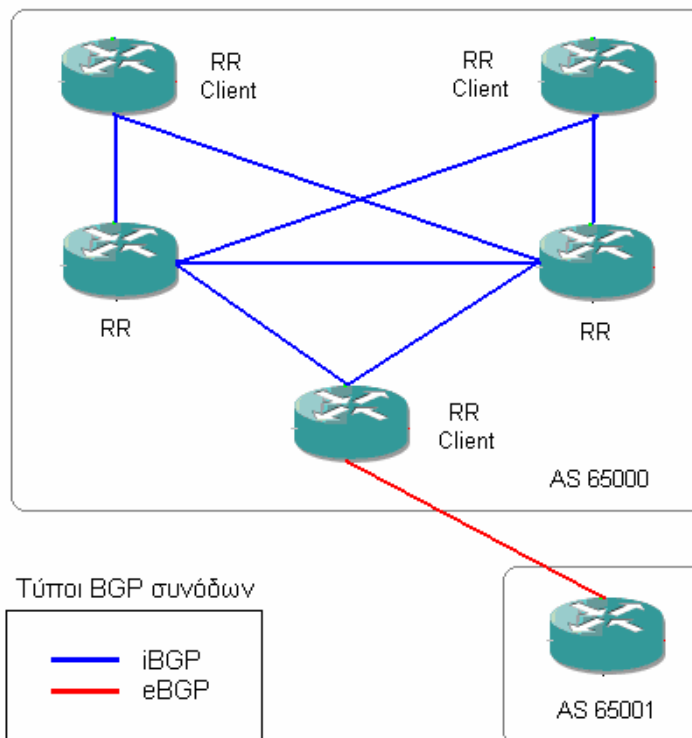
Εικόνα 2.1.2 : BGP τοπολογία με Route Reflector

2.2 Redundant Route Reflectors

Όπως αναφέρθηκε και προηγουμένως, κάθε cluster περιέχει έναν ή περισσότερους RR (Route Reflectors). Όσο αυξάνεται ο αριθμός των Route Reflector αυξάνεται και ο αριθμός των iBGP συνδέσεων που απαιτούνται, λόγω της ανάγκης οι RR-Clients (Route Reflector Clients) να είναι συνδεδεμένοι μέσω iBGP με όλους τους RR. Οπότε πρόσκαιρα γενάτε το ερώτημα γιατί να κάνουμε χρήση περισσότερων του ενός RR ανά cluster. Η απάντηση μονολεκτικά είναι η αξιοπιστία.

Όταν σχεδιάζουμε μια τοπολογία BGP για ένα αυτόνομο σύστημα πρέπει να λάβουμε υπόψη μας τους κινδύνους δυναμικής μεταβολής της τοπολογίας λόγω δυσλειτουργίας κάποιου δρομολογητή ή κάποιας φυσικής διασύνδεσης (link). Χρησιμοποιώντας ένα RR ανά cluster, υπάρχει κίνδυνος το σύνολο των δρομολογητών του cluster να χάσει την γνώση διαδρομών εκτός AS. Έτσι, για λόγους αξιοπιστίας, συνήθως γίνεται χρήση Redundant Route Reflectors. Στον αντίποδα, μετριάζονται τα οφέλη του «απλού» Route Reflection όσον αφορά των συνολικό αριθμό συνδέσεων BGP που πρέπει να διαχειρίζεται ένας δρομολογητής που εκτελεί την εφαρμογή BGP-4.

Ο σχεδιασμός τοπολογίας BGP με Route Reflection, πολλές φορές γίνεται σε επίπεδα. Δηλαδή οι δρομολογητές που ανήκουν σε ένα cluster ορίζονται βάση της απόστασης τους από τους συνοριακούς δρομολογητές. Επίσης, οι δρομολογητές που δρουν ως RR συνήθως έχουν μεγάλη υπολογιστική ισχύ, ενώ οι πελάτες τους μπορεί να είναι παλαιότερα μοντέλα με μικρές δυνατότητες.



Εικόνα 2.2.1 : BGP τοπολογία με Redundant Route Reflectors

Επεκτείνοντας το προηγούμενο παράδειγμα (Κεφ.2.1), στην ίδια τοπολογία ορίζουμε δύο Route Reflectors με ίδιο cluster-id, ώστε να λειτουργήσουν ως Redundant Route Reflectors. Ακολουθώντας τους κανόνες, πρέπει οι Route Reflectors να είναι συνδεμένη με πλήρες πλέγμα και κάθε Route Reflector Client να έχει μια τουλάχιστο iBGP σύνδεση με κάθε Route Reflector, καταλήγουμε στο σχεδιασμό των διασυνδέσεων iBGP (Εικ.2.2.1). Καταλήγουμε σε 7 απαιτούμενες συνδέσεις iBGP, δηλαδή, περισσότερες από το Route Reflection με ένα RR και λιγότερες από το πλήρες πλέγμα.

ΚΕΦΑΛΑΙΟ 3

Προβλήματα του Route Reflection

Στο κεφάλαιο αυτό, γίνεται ανάλυση προβλημάτων που προκύπτουν από την χρήση Route Reflectors.

3.1 Αστάθεια τύπου-I

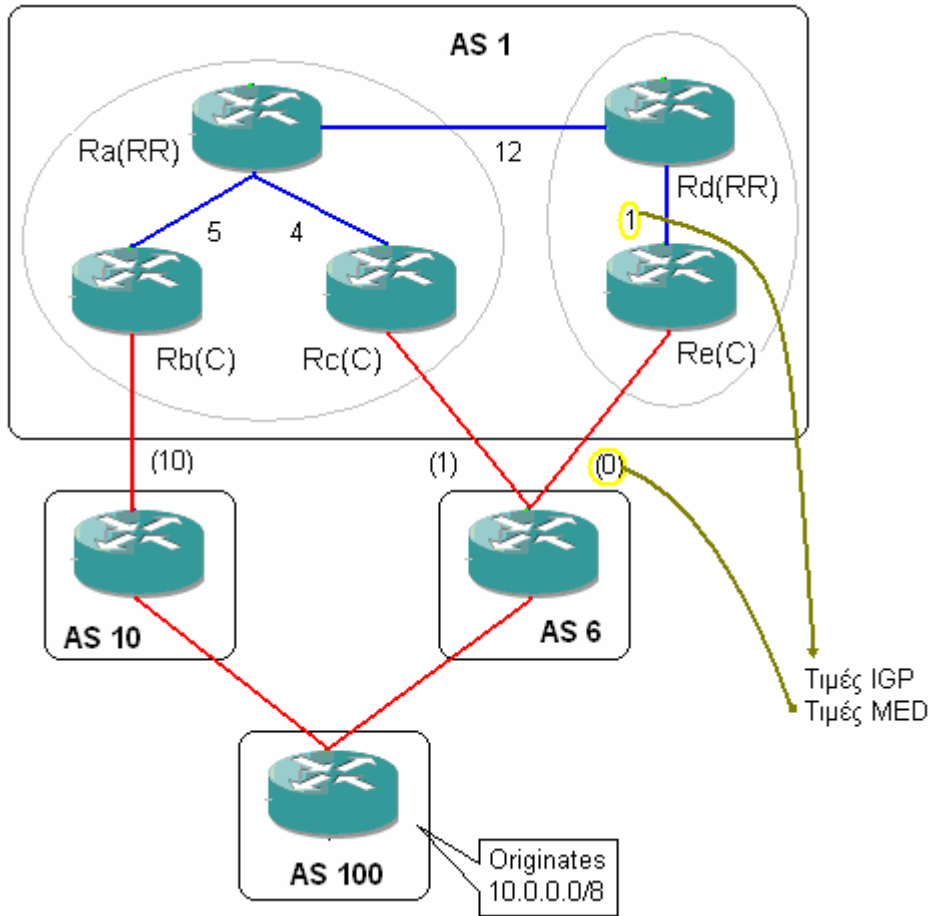
Ύστερα από την αρχική εφαρμογή της εναλλακτικής Route Reflection, αποδείχθηκε, πως κάτω από κάποιες συνθήκες υπάρχει περίπτωση οι δρομολογητές να αλλάζουν συνεχώς την επιλογή τους όσον αφορά την καλύτερη διαδρομή (McPherson, D & Gill, V & Walton, D & Renata, A. (2002)). Αυτό συνεπάγεται την αδυναμία πρόσβασης των συγκεκριμένων διαδρομών, καθώς και την υπερφόρτωση των δρομολογητών. Παρακάτω, αποδεικνύεται πως υπάρχει περίπτωση να εμφανιστούν τέτοια φαινόμενα σε τοπολογίες που συνυπάρχουν οι παρακάτω προϋποθέσεις :

- Route-Reflection ενός επιπέδου
- Το δίκτυο δέχεται το χαρακτηριστικό MED από δύο ή περισσότερα AS για ένα συγκεκριμένο υποδίκτυο και οι τιμές του MED είναι μοναδικές

Οι προϋποθέσεις αυτές συναντιούνται αρκετά συχνά όπως γίνεται εύκολα αντιληπτό. Για να ξεκινήσουμε την ανάλυση του προβλήματος υποθέτουμε την τοπολογία της εικόνας 3.1.1. Στην τοπολογία το AS 1 περιέχει δύο clusters, τα οποία περιέχουν από ένα RR (Ra και Rd), ενώ οι υπόλοιποι δρομολογητές του AS είναι Route Reflector Clients (C). Επίσης στο σχήμα φαίνονται τα metrics της IGP τοπολογίας καθώς και οι τιμές MED που συνοδεύουν το δίκτυο 10.0.0.0/8 (NLRI).Στον πίνακα 3.1.1, φαίνονται τα χαρακτηριστικά της διαδρομής 10.0.0.0/8. Υπενθυμίζεται ότι τα IGP metrics (Εικ.3.1.1) έχουν σημασία μόνο όταν γίνεται διάδοση μέσω iBGP.

Δρομολογητής	MED	AS_PATH
Rb	10	10 100
Rc	1	6 100
Re	0	6 100

Πίνακας 3.1.1 : Πίνακας χαρακτηριστικών διαδρομής 10.0.0.0/8



Εικόνα 3.1.1 : BGP τοπολογία με αστάθεια τύπου I

Στα παρακάτω βήματα η επιλεγμένη διαδρομή θα σημειώνεται με *.

- 1) Ο δρομολογητής Ra έχει τα παρακάτω περιεχόμενα (πιν.3.1.2) στον πίνακα δρομολόγησης του. Επιλεγμένη διαδρομή είναι αυτή που έγινε γνωστή μέσω του AS10.

AS_PATH	MED	Next Hop IGP Cost
6 100	1	4
* 10 100	10	5

Πίνακας 3.1.2 : Πίνακας BGP του Ra για την διαδρομή 10.0.0.0/8

Η διαδρομή «10 100» δεν θα έπρεπε να έχει επιλεγεί ως καλύτερη, λόγω μεγαλύτερου MED και Next Hop IGP metric. Στα επόμενα βήματα θα φανεί πως φτάνουμε σε αυτό το σημείο. Ο δρομολογητής Ra μέσω της διαδικασίας Best Path

selection αντιλαμβάνεται ότι χρησιμοποιεί την λάθος διαδρομή καθώς η «6 100» διαθέτει μικρότερο IGP metric. Συνεπώς, ο Ra επιλέγει την «6 100» και προωθεί μήνυμα Update στους γείτονες τους ανακοινώνοντας την «6 100» ως καλύτερη.

2) Ο Rd λαμβάνει το Update από τον Ra και αποκτά τον παρακάτω πίνακα BGP (πιν.3.1.3)

AS_PATH	MED	Next Hop IGP Cost
* 6 100	0	12
6 100	1	5

Πίνακας 3.1.3 : Πίνακας BGP του Rd για την διαδρομή 10.0.0.0/8

Ο Rd επιλέγει την «6 100,0,12» ως καλύτερη λόγω μικρότερου MED. Ο Rd στέλνει μήνυμα update στους γείτονες του περιλαμβάνοντας την καλύτερη για αυτόν διαδρομή προς το δίκτυο 10.0.0.0/8.

3) Ο Ra λαμβάνει το Update και έχει στον πίνακα BGP τα παρακάτω (πιν.3.1.4)

AS_PATH	MED	Next Hop IGP Cost
6 100	0	13
6 100	1	4
* 10 100	10	5

Πίνακας 3.1.4 : Πίνακας BGP του Ra για την διαδρομή 10.0.0.0/8

Η πρώτη διαδρομή «6 100,0,13» κερδίζει την δεύτερη «6 100,1,4» λόγω μικρότερου MED και εν συνεχεία η Τρίτη διαδρομή κερδίζει την πρώτη λόγω μικρότερου IGP metric προς το Next Hop. Ο Ra στέλνει μήνυμα Update για να ενημερώσει τους γείτονες του για την νέα καλύτερη διαδρομή που επέλεξε προς το δίκτυο 10.0.0.0/8

4) Ο Rd λαμβάνει το Update από τον Ra και αποκτά τον παρακάτω πίνακα BGP (πιν.3.1.5). Ο Rd επιλέγει την «10 100,10,6» ως καλύτερη λόγω μικρότερου IGP metric. Ο Rd στέλνει μήνυμα update στους γείτονες του περιλαμβάνοντας την καλύτερη για αυτόν διαδρομή προς το δίκτυο 10.0.0.0/8.

AS_PATH	MED	Next Hop IGP Cost
6 100	0	12
* 10 100	10	6

Πίνακας 3.1.5 : Πίνακας BGP του Rd για την διαδρομή 10.0.0.0/8

5) Ο Ra λαμβάνει το Update και έχει στον πίνακα BGP τα παρακάτω (πιν.3.1.6)

AS_PATH	MED	Next Hop IGP Cost
6 100	1	4
* 10 100	10	5

Πίνακας 3.1.6 : Πίνακας BGP του Ra για την διαδρομή 10.0.0.0/8

Ο Ra δέχθηκε μήνυμα Update/withdraw για την διαδρομή «6,100», και αλλάζει την απόφαση του για την καλύτερη διαδρομή επιλέγοντας την «10 100», παρόλο που η «6 100,1,4» είναι καλύτερη.

Σε αυτό το σημείο διαμορφώθηκε ένας πλήρης βρόχος και βρισκόμαστε και πάλι στο βήμα 1. Αυτό αποτελεί παράδειγμα αστάθειας τύπου I.

3.2 Επιλογή μη βέλτιστης διαδρομής

Το BGP, αρχικά σχεδιάστηκε για διασύνδεση πλήρους πλέγματος iBGP εντός του AS. Με την διασύνδεση πλήρους πλέγματος, όλοι οι δρομολογητές του AS γνωρίζουν όλες τις εναλλακτικές διαδρομές για κάθε υποδίκτυο. Με την σειρά του, κάθε δρομολογητής εκτελεί τον αλγόριθμο επιλογής βέλτιστης διαδρομής και την τοποθετεί στον πίνακα δρομολόγησης. Δυστυχώς, για μεγάλα αυτόνομα συστήματα, ο αριθμός των διασυνδέσεων iBGP, καθώς και η συνολική πληροφορία δρομολόγησης (Εναλλακτικές διαδρομές/δίκτυο) καθιστούν το πλήρες iBGP πλέγμα απαγορευτικό.

Έτσι, χρησιμοποιούμε Route Reflection. Στην περίπτωση αυτή, ο RR γνωρίζει όλες τις πιθανές διαδρομές ανά υποδίκτυο και τις τοποθετεί στον πίνακα εισερχόμενων διαδρομών BGP (adj-rib-in), εκτελεί τον αλγόριθμο επιλογής βέλτιστης διαδρομής και τοποθετεί τις καλύτερες από τη δική του σκοπιά στον πίνακα δρομολόγησης (loc-rib). Οι Route Reflector Clients, συνδέονται μέσω iBGP μόνο με τον Route Reflector. Ο Route Reflector τοποθετεί τις καλύτερες διαδρομές στον εξερχόμενο πίνακα (adj-rib-out) και στέλνει updates στους client.

Αποτέλεσμα των παραπάνω ενεργειών είναι οι Route Reflector clients να διατηρούν λίγες iBGP συνδέσεις σε σχέση με την τοπολογία πλήρους πλέγματος. Επίσης, καταναλώνουν μικρότερα ποσά μνήμης αφού εγκαθίστανται στον εισερχόμενο BGP πίνακα (adj-rib-in) μόνο μια διαδρομή ανά δίκτυο.

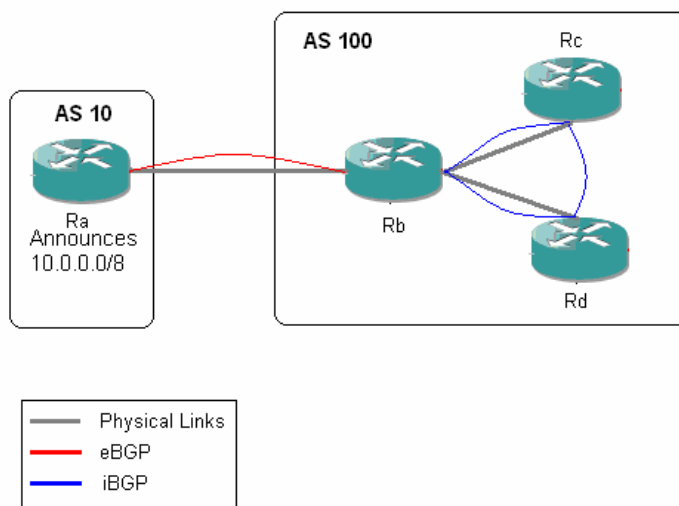
Έτσι, καταλήγουμε ότι σε ένα BGP δίκτυο και υπό συνθήκες τοπολογίας πλήρους πλέγματος ένας δρομολογητής πιθανώς να μάθαινε πολλές διαδρομές, ενώ σε τοπολογία Route Reflector και ταυτόχρονα έχοντας το ρόλο του client μαθαίνει μόνο μια. Συνεπώς, δημιουργούνται συνθήκες όπου πιθανώς κάποιοι δρομολογητές να μην κάνουν όμοιες επιλογές για κάποια δίκτυα, συγκριτικά με την τοπολογία πλήρους πλέγματος. Τελικώς, συμπεραίνουμε την πιθανότητα επιλογής μη βέλτιστης διαδρομής από πλευράς του Route Reflector Client.

3.3 Συμπεριφορά σε περίπτωση δυσλειτουργίας

Η συμπεριφορά ενός δικτύου σε περίπτωση βλάβης εξαρτάται από πολλές παραμέτρους και ειδικά από το IGP πρωτόκολλο και είναι δύσκολο να απομονωθεί η συμπεριφορά του BGP από τις λοιπές παραμέτρους του δικτύου, καθώς αυτό ουσιαστικά αποτελεί μια εφαρμογή.

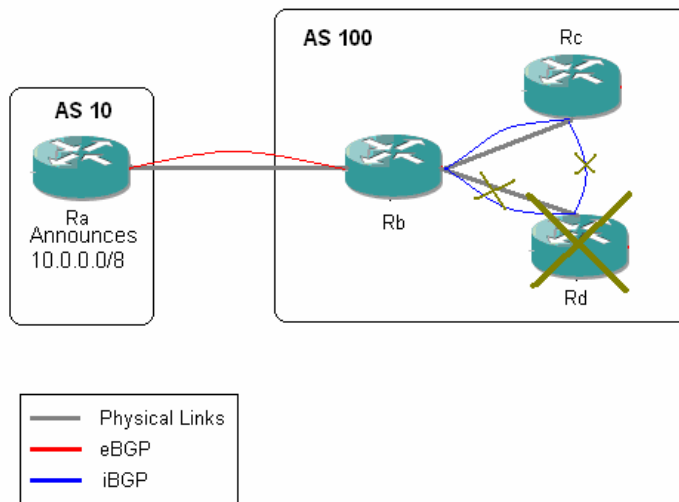
Η απώλεια ενός δρομολογητή ή μιας φυσικής σύνδεσης (Link) σε μια τοπολογία πλήρους πλέγματος έχει προβλέψιμα αποτελέσματα. Χρησιμοποιώντας απλό παράδειγμα θα αποδείξουμε ότι η συμπεριφορά του δικτύου σε περίπτωση βλάβης υποβαθμίζεται με τη χρήση Route Reflection.

Στην τοπολογία του παραδείγματος (εικ.3.3.1) υπάρχουν δύο AS. Το AS 10 διαφημίζει το δίκτυο 10.0.0.0/8 στο AS 100, το οποίο βρίσκεται σε τοπολογία πλήρους πλέγματος iBGP. Έτσι, η 10.0.0.0/8 γίνεται γνωστή από το σύνολο των δρομολογητών του AS 100.



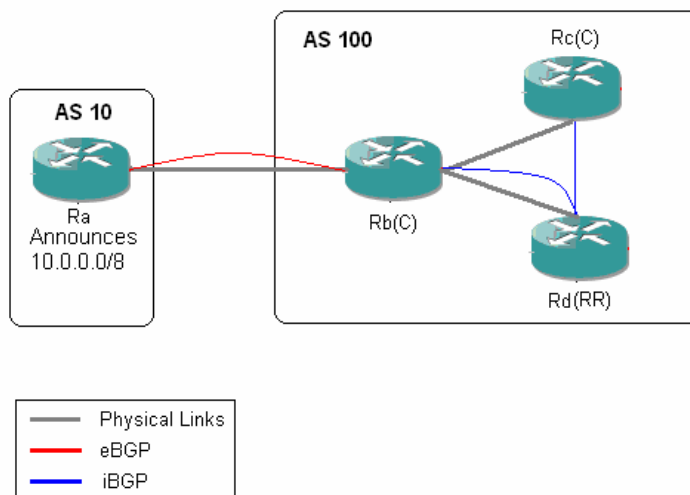
Εικόνα 3.3.1 : BGP τοπολογία με διασύνδεση full-mesh iBGP

Λόγω βλάβης, ο δρομολογητής Rd σταματάει την εκτέλεση BGP (εικ.3.3.2). Οι φυσικές διασυνδέσεις μένουν ανεπηρέαστες. Αποτέλεσμα της δυσλειτουργίας θα ήταν ο Rc να συνεχίσει να έχει γνώση της 10.0.0.0/8 και να λειτουργεί φυσιολογικά λόγω της φυσιολογικής λειτουργίας της διασύνδεσης με τον Rb.



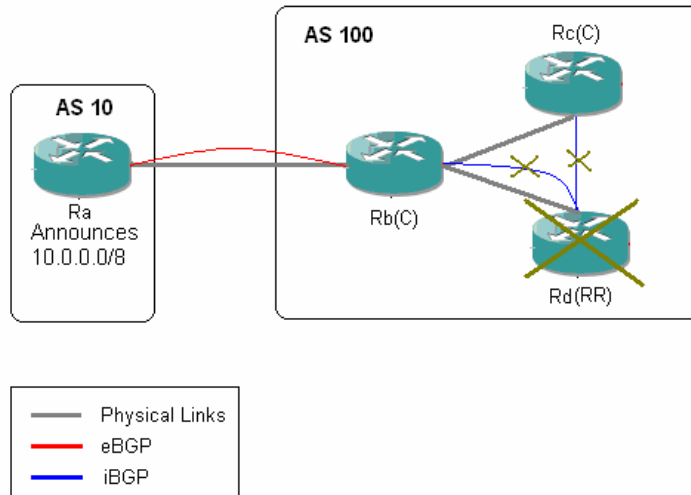
Εικόνα 3.3.2 : BGP τοπολογία με διασύνδεση full-mess iBGP (βλάβη Rd)

Στο ίδιο AS, για να μειώσουμε τον συνολικό αριθμό διασυνδέσεων iBGP από τρεις (εικ.3.3.1) σε δύο (εικ.3.3.3) εφαρμόζουμε το Route Reflection και θέτουμε τυχαία και λόγω συμμετρίας τον Rd ως Route Reflector από το πιθανό ζευγάρι (Rc,Rd) και τους υπόλοιπους δρομολογητές του AS 100 ως RR-clients. Η τοπολογία λειτουργεί ακριβώς όπως η παραπάνω (εικ.3.3.1).



Εικόνα 3.3.3 : BGP τοπολογία με Route Reflection

Λόγω βλάβης, ο δρομολογητής Rd σταματάει την εκτέλεση BGP (εικ.3.3.4). Οι φυσικές διασυνδέσεις μένουν ανεπηρέαστες. Αποτέλεσμα της δυσλειτουργίας θα ήταν ο Rc να μην έχει γνώση της 10.0.0.0/8 παρόλο που υπάρχει λειτουργικό φυσικό μονοπάτι προς το AS 10. Αυτό συμβαίνει διότι ο Rc δεν έχει γειτονικές σχέσεις BGP με τον Rb και προφανώς δεν ενημερώνεται μέσω αυτού του δρομολογητή για τις εκτός AS διαδρομές.



Εικόνα 3.3.4 : BGP τοπολογία με Route Reflection (βλάβη Rd)

Από το παραπάνω παράδειγμα, γίνεται εμφανής η ευαισθησία σε δυσλειτουργίες που παρουσιάζει το iBGP με Route Reflection, τουλάχιστο στη περίπτωση χρήσης ενός Route Reflector ανά cluster.

Έτσι, συνήθως για λόγους ασφαλείας του δικτύου χρησιμοποιούνται περισσότεροι του ενός RR ανά cluster. Επίσης, οι RR είναι συνδεδεμένοι σε πλήρες πλέγμα. Στην περίπτωση αυτή, υπάρχει αύξηση στις απαιτούμενες iBGP συνδέσεις και μετριάζονται τα πλεονεκτήματα του Route Reflection.

ΚΕΦΑΛΑΙΟ 4

Επισκόπηση Επιστημονικής συνεισφοράς

Στο προηγούμενο κεφάλαιο, έγινε επεξήγηση του Route Reflection, καθώς και των πλεονεκτημάτων και μειονεκτημάτων αυτού. Η λύση αυτή, έχει ήδη βρει εφαρμογή σε μεγάλο μέρος του διαδικτύου. Με κίνητρο την βελτίωση του Route Reflection, πολλές ομάδες επιστημόνων που ερευνούν το BGP-4, έχουν εστιάσει στην βελτίωση της τοπολογίας BGP με χρήση Route Reflection. Παρακάτω, γίνεται παρουσίαση των σημαντικότερων προτάσεων σχετικά με το Route Reflection.

4.1 Προτάσεις για σωστό σχεδιασμό με Route Reflectors

Ο μεγαλύτερος αριθμός επιστημονικών άρθρων και δημοσιεύσεων, βασίζεται στον σωστό σχεδιασμό BGP τοπολογίας με χρήση Route Reflection χωρίς να γίνεται μεταβολή του πρωτοκόλλου σε επίπεδο εφαρμογής. Στην αρχική υλοποίηση του Route Reflection, δεν γίνεται λόγος για την σωστή τοποθέτηση των RR, των clients, καθώς και τον αριθμό των clusters που πρέπει να χρησιμοποιήσουμε σε σχέση με το μέγεθος του AS. Αποτέλεσμα είναι να υπάρχουν ταλαντώσεις (McPherson, D & Gill, V & Walton, D & Renata, A. (2002)) και (Basu, A & Chih-Hao Luke, O & Rasala A & Shepherd, B & Wilfong, G (2002)) στην επιλογή διαδρομής που οδηγούν σε αύξηση της χρήσης της υπολογιστικής ισχύος και ανταλλαγή πολλών BGP μηνυμάτων.

Σε επιστημονικό περιοδικό, παρουσιάζεται επιστημονική μελέτη (Griffin, T & Wilfong, G. (2002)) σχετικά με τις συνθήκες που πρέπει να τηρούνται κατά τον σχεδιασμό τοπολογίας iBGP με Route Reflection. Στόχος του μαθηματικού μοντέλου είναι να αποφευχθούν οι ταλαντώσεις και η κακή επιλογή exit point από κάποιον route reflector client. Επίσης, γίνεται σαφές πως πιθανή βλάβη εντός του AS, μπορεί να παραβιάσει τα προαναφερόμενα και να κάνουν την εμφάνιση τους τα γνωστά προβλήματα του iBGP (κεφ.3).

Σε πρακτικά συνεδρίου παρουσιάζεται ο αλγόριθμος ORC, (Xiao, L & Wang, J & Nahrstedt, K. (2003)). Στόχος του αλγορίθμου είναι η τοποθέτηση των Route Reflector και των αντιστοίχων clients σε clusters, ώστε να αποφευχθούν αστάθειες τύπου I (παρ.3.1). Η μέθοδος βασίζεται στην τοποθέτηση των BGP Session βάσει του μικρότερου IGP metric. Ευελιξία του αλγορίθμου είναι η δυνατότητα να οριστούν στατικά κάποιοι Reflectors και κάποιοι clients στο διάγραμμα του AS, πριν αυτό δεχτεί την επεξεργασία του αλγορίθμου. Δεν υπάρχει σύσταση για ενσωμάτωση του αλγορίθμου στην εφαρμογή BGP-4 και προφανώς τα αποτελέσματα δεν μεταβάλλονται δυναμικά.

Στο άρθρο (Klockar, T & Carr-Motyckova, L. (2004)) προτείνεται μια μέθοδος αποθήκευσης παλαιών διαδρομών και σύγκρισης αυτών με τις νεότερες, ώστε να αποφευχθούν οι ταλαντώσεις, με αποτέλεσμα την αύξηση της σταθερότητας και της

ταχύτητας αποκατάστασης(convergence) Μειονέκτημα της μεθόδου αποτελεί η αυξημένη απαίτηση σε μνήμη από πλευράς RR client.

Σε επιστημονικό περιοδικό, παρουσιάζεται η ανάγκη για παράκαμψη του ανθρώπινου παράγοντα στην διαδικασία σχεδιασμού τοπολογίας iBGP (Van den Schrieck, V & Francois, P & Tandel, S & Bonaventure, O. (2006)). Οι εύστοχες παρατηρήσεις του άρθρου, κατά κύριο λόγο εστιάζουν στην πολυπλοκότητα του BGP γενικότερα. Δυστυχώς, δεν προτείνεται κάποια μέθοδος που να υλοποιεί τις απαιτήσεις από μια iBGP τοπολογία, όπως πολύ σωστά επισημαίνονται στο άρθρο.

Παρομοίως, σε επιστημονικό περιοδικό (Rawat, A & Shayman, M. (2006)) παρουσιάζεται μέθοδος σχεδιασμού τοπολογίας iBGP με Route Reflectors. Οι προϋποθέσεις που θέτει η μέθοδος εγγυώνται την εύρυθμη λειτουργία του πρωτοκόλλου, μη λαμβάνοντας υπόψη την συμπεριφορά του πρωτοκόλλου σε δυναμικές αλλαγές (π.χ αλλαγή της τοπολογίας λόγω βλάβης). Η μέθοδος δεν συνιστά μεταβολή ή προσθήκη στο πρωτόκολλο διότι βάσει των συγγραφέων κάτι τέτοιο θεωρείται ανέφικτο. Συνεπώς, η μέθοδος αφορά τον αρχικό σχεδιασμό ενός AS με χρήση iBGP Route Reflection.

Στο άρθρο (Buod, M & Meulle, M & Uhlig, S. (2006)), γίνεται μια αξιολόγηση της επιλογής βέλτιστης διαδρομής (exit point) με την χρήση route Reflector και τα αποτελέσματα συμβάλλουν στην ανάγκη για καλύτερο σχεδιασμό μιας τέτοιας τοπολογίας, χωρίς να προτείνεται κάποια αυτοματοποίηση.

Στο άρθρο των (Vutukuru, M & Valiant, P & Kopparty, S & Balakrishnan, H. (2006)) παρουσιάζεται ο αλγόριθμος BGP-Sep του MIT. Στόχος του είναι να δημιουργηθούν σωστές τοπολογίες με route reflector που να μην βασίζονται στην διαφήμιση πολλών πιθανών διαδρομών προς τους clients. Ο αλγόριθμος επιλέγει τους clients του κάθε reflector δημιουργώντας τα αντίστοιχα cluster βάσει διαγράμματος των IGP metrics. Μειονέκτημα του αλγορίθμου αποτελεί η offline χρήση του, δηλαδή αφορά μόνο σχεδίαση τοπολογίας με αποτέλεσμα να μην αντιδρά στην περίπτωση των δυναμικών αλλαγών των IGP metrics.

4.2 Προτάσεις για την επιλογή βέλτιστης διαδρομής

Το iBGP Route Reflection (Bates, T & Chen, E & Chandra, R.(2006)) έχει ως πρωταρχικό σκοπό την μείωση των iBGP συνδέσεων που απαιτούνται για την μετάδοση πληροφορίας δρομολόγησης εντός του AS. Εντός ενός AS οι Route Reflector είναι διασυνδεδεμένοι σε πλήρες πλέγμα και λαμβάνουν γνώση όλης της πληροφορίας δρομολόγησης, που περνά από το AS. Αντιθέτως, οι Route Reflector client, μαθαίνουν μόνο μια διαδρομή ανά δίκτυο προορισμού. Αυτό λαμβάνει χώρα διότι οι Route Reflector διατηρούν μια μόνο διαδρομή ανά IP προορισμό στον BGP πίνακα δρομολόγησης και ταυτόχρονα «ανακλούν» μια διαδρομή ανά δίκτυο προορισμού στους πελάτες τους.

Αποτέλεσμα των παραπάνω, είναι η αυξημένη πιθανότητα κάποιος RR client να χρησιμοποιεί διαφορετικές διαδρομές προς κάποια δίκτυα προορισμού, από την περίπτωση της τοπολογίας πλήρους πλέγματος. Έτσι, καταλήγουμε πως με την χρήση

Route Reflector, δεν γίνεται πάντα χρήση της βέλτιστης διαδρομής ή αξιοποίηση του καλύτερου σημείου εξόδου προς κάποιο AS. Οι επιστημονικές προτάσεις για την βελτίωση αυτής της συμπεριφοράς του Route Reflection, καλύπτουν και τα προβλήματα του (κεφ.4.1).

Στο άρθρο (Basu, A & Chih-Hao Luke, O & Rasala A & Shepherd, B & Wilfong, G (2002)) προτείνεται αντικατάσταση του iBGP πρωτοκόλλου ώστε να επιτρέπει αυτό σε έναν RR να διαφημίσει περισσότερες από μια διαδρομές σε client. Έτσι, ο RR client αποφασίζει για την καλύτερη διαδρομή μέσω του δικού του BGP decision process. Επομένως λόγω της πλήρους μετάδοσης της πληροφορίας δρομολόγησης, αποφεύγονται οι ταλαντώσεις και γίνεται χρήση της καλύτερης διαδρομής. Στα μειονεκτήματα τη μεθόδου είναι η ανταλλαγή περισσότερων BGP μηνυμάτων, η αυξημένη απαίτηση σε μνήμη και υπολογιστική ισχύ από πλευράς client και reflector.

Στο άρθρο (Walton, D & Renata, A & Scudder, J. (2002)), παρόμοια με (Basu, A & Chih-Hao Luke, O & Rasala A & Shepherd, B & Wilfong, G (2002)) προτείνεται μια μέθοδος που επιτρέπει σε έναν RR να διαφημίσει περισσότερες από μια διαδρομές σε client, ώστε να επιτρέψει στον client να αποφασίσει για την καλύτερη διαδρομή μέσω του δικού του BGP decision process. Έτσι, αποφεύγονται οι ταλαντώσεις και γίνεται χρήση της καλύτερης διαδρομής. Στα μειονεκτήματα τη μεθόδου είναι η ανταλλαγή περισσότερων BGP μηνυμάτων, η αυξημένη απαίτηση σε μνήμη και υπολογιστική ισχύ από πλευράς client και reflector, ακόμη και σε σχέση με την προηγούμενη πρόταση (Basu, A & Chih-Hao Luke, O & Rasala A & Shepherd, B & Wilfong, G (2002)).

Σε άρθρο που παρουσίασαν οι (Musunuri, R & Cobb, J (2003)) παρουσιάζεται ο αλγόριθμος SPD που ενσωματώνεται στο BGP πρωτόκολλο. Η λειτουργία του αφορά την επιλεκτική διαφήμιση διαδρομών στους client από RR, ενώ ταυτοχρόνως γίνεται κατάργηση της έννοιας cluster. Μειονέκτημα είναι η ανάγκη για ενσωμάτωση του πρωτοκόλλου σε όλους τους δρομολογητές που εκτελούν BGP εντός του AS, καθώς και η αυξημένη ζήτηση σε μνήμη, υπολογιστική ισχύς και κατανάλωση εύρους ζώνης. Ο SPD λύνει το πρόβλημα καλύτερης διαδρομής και αποφεύγει τις ανωμαλίες που οφείλονται στο route reflection. Σε άρθρο των ίδιων συγγραφέων (Musunuri, R & Cobb, J. (2004)) προτείνεται το πρωτόκολλο S-iBGP που αποτελεί την ενσωμάτωση του SPD στο BGP.

Στο άρθρο των (Bonaventure, O & Uhlig, S & Quoitin, B. (2004)) γίνεται λόγος, χωρίς να προτείνεται υλοποίηση για route reflectors οι οποίοι αποφασίζουν την καλύτερη διαδρομή εκ μέρους των client τους και μετά την διαφημίζουν σε αυτούς. Βάσει του θεωρητικού αυτού μοντέλου, οι RR απαιτούν πολύ μεγαλύτερη υπολογιστική ισχύ και μνήμη, ενώ οι clients όχι. Δίνεται λύση το θέμα της μη βέλτιστης επιλογής διαδρομής.

Βελτιωμένη εκδοχή του αλγορίθμου BGP-Sep (Vutukuru, M & Valiant, P & Kopparty, S & Balakrishnan, H. (2006)) υλοποιείται στο άρθρο Feng, Z & Xicheng, L & Peidong, Z & Jinjing Z. (2006). Ο αλγόριθμος μετονομάζεται σε BGPSep-S και εγγυάται επιπλέον του BGP-Sep επιλογή βέλτιστης διαδρομής. Μειονέκτημα του αλγορίθμου αποτελεί η σχεδιαστική του φύση, η οποία το κάνει ανήμπορο να

αντιδράσει σε δυναμικές μεταβολές της τοπολογίας. Έτσι, η συγκεκριμένη λύση δεν προσφέρει κάτι για ομαλή λειτουργία σε περίπτωση βλάβης.

4.3 Προτάσεις για ομαλή λειτουργία σε περίπτωση βλάβης

Η τεχνική Route Reflection, πολλαπλασιάζει τις επιπτώσεις από κάποια πιθανή δυσλειτουργία δρομολογητή ή διασύνδεσης (link) εντός του AS. Παρακάτω εξετάζουμε τις επιστημονικές προτάσεις σχετικά με την βέλτιστη συμπεριφορά μιας τοπολογίας βασισμένης σε route Reflection σε περίπτωση βλάβης.

Σε πρακτικά συνεδρίου, (Xiao, L & Wang, J & Nahrstedt, K. (2003)) προτείνεται μέθοδος για υπολογισμό του αριθμού των Route Reflectors, καθώς και την βέλτιστη κατανομή του ρόλου client και Route reflector. Το μαθηματικό μοντέλο, βασίζεται στη συχνότητα βλαβών που παρουσιάζουν συγκεκριμένοι δρομολογητές ή οι μεταξύ τους διασυνδέσεις. Σκοπός της μεθόδου είναι ο βέλτιστος σχεδιασμός ενός AS με χρήση Route Reflection, σε σχέση με την μείωση πιθανότητας εμφάνισης βλαβών. Η μέθοδος δεν ενσωματώνεται στην εφαρμογή BGP-4 και έτσι δεν τίθεται θέμα συμβατότητας. Εφαρμογές βασισμένες στη μέθοδο θα μπορούσαν να δέχονται αναφορές βλαβών από τους δρομολογητές του AS (π.χ μέσω SNMP) και να υπολογίζουν νέα προτεινομένη τοπολογία ύστερα από πέρας χρονικού διαστήματος ή λόγω της επέκτασης του AS.

Επίσης σε πρακτικά συνεδρίου, παρουσιάζεται ο αλγόριθμος C-DAP (Musunuri, R & Cobb, J. (2005)). Στόχος του αλγορίθμου είναι η αύξηση της σταθερότητας του πρωτοκόλλου BGP, περιλαμβάνοντας κάθε μορφή αυτού (eBGP ή iBGP). Ο αλγόριθμος βασίζεται στην δυναμική αξιολόγηση των διαδρομών βάσει της αξιοπιστίας τους σε βάθος χρόνου. Μειονέκτημα του αλγορίθμου, είναι η αναγκαία υλοποίηση του αλγορίθμου σε όλους τους δρομολογητές που εκτελούν BGP. Επίσης, είναι αρνητικό το γεγονός ότι δεν λαμβάνεται υπόψη κατά τον σχεδιασμό του αλγορίθμου το χαρακτηριστικό διαδρομών MED.

Στο άρθρο (Van den Schrieck, V & Bonaventure, O. (2005)) γίνεται πειραματική σύγκριση των πιθανών iBGP τοπολογιών. Συμπέρασμα του άρθρου είναι ότι η χρήση redundant Route Reflectors σε διάταξη που να αποδίδει κατά το δυνατόν διαφορετικές διαδρομές προς τους clients παρουσιάζει τα βέλτιστα αποτελέσματα.

Στον παρακάτω πίνακα (πιν.4.1), γίνεται σύγκριση των επιστημονικών άρθρων και προτάσεων που έχουν δημοσιευτεί σχετικά με το iBGP Route Reflection βάσει των κυρίων χαρακτηριστικών που προκαλούν προβληματισμό γύρω από τις αδυναμίες του πρωτοκόλλου.

ΤΙΤΛΟΣ ΑΡΘΡΟΥ*	Ανάγκη επιπλέον BGP μηνυμάτων	Μείωση iBGP συνδέσεων	Χρήση καλύτερης διαδρομής	Σχεδιασμός Τοπολογίας	Μείωση απαιτούμενης μνήμης	Αύξηση σταθερότητας
Reliability-aware IBGP Route Reflection Topology Design				++		+
Preventing Oscillations in Route Reflector-Based I-BGP			+		--	+
Route Oscillations in IBGP with Route Reflection			+		-	+
BGP Persistent Route Oscillation Condition				+		
BGP Persistent Route Oscillation Solution			+		--	+
Let BGP speakers configure their iBGP sessions on their own	-			+		
Checking for optimal egress points in iBGP routing				+		
A Complete Solution for iBGP Stability	--		+	+	-	+
Enforcing IBGP Convergence	--		+	+	-	+
Optimizing IBGP Route Reflection Network				++		
How to Construct a Correct and Scalable iBGP Configuration			+	+		
On the Correctness of IBGP Configuration				+		

Πίνακας 4.1 : Σύγκριση της σχετικής αρθογραφίας

* Αναφέρεται λόγω χώρου μόνο το όνομα του άρθρου περισσότερες λεπτομέρειες στη βιβλιογραφία

ΤΙΤΛΟΣ ΑΡΘΡΟΥ*	Ανάγκη επιπλέον BGP μηνυμάτων	Μείωση iBGP συνδέσεων	Χρήση καλύτερης διαδρομής	Σχεδιασμός Τοπολογίας	Μείωση απαιτούμενης μνήμης	Αύξηση σταθερότητας
Stable iBGP through Selective Path Dissemination	--		+			-
Comparison of IBGP Topologies						
The case for more versatile BGP Route Reflectors			+	+		
Standby Route Reflection (This Proposal)	-	+	+			+

Πίνακας 4.1 : (Συνέχεια)

ΚΕΦΑΛΑΙΟ 5

Πρωτόκολλο Standby Route Reflection

Η χρήση iBGP Route Reflection μαζί με τις λύσεις που έδωσε , δημιουργήσε συνοπτικά τα εξής προβλήματα :

- Είναι πιθανό μερικοί δρομολογητές που επιλέγουν την καλύτερη διαδρομή τους βάσει του MED να πέσουν σε ταλαντώσεις ή να υπάρξουν ανωμαλίες κατά την προώθηση των IP πακέτων. Για την περίπτωση αυτή έχουν προταθεί βελτιώσεις (κεφ.4.1).
- Λόγω μείωσης της πληροφορίας δρομολόγησης που φτάνει στους RR Clients είναι πιθανό να μη γίνει η βέλτιστη επιλογή διαδρομής από πλευράς RR Client. Για την περίπτωση αυτή έχουν επίσης προταθεί βελτιώσεις (κεφ.4.2).
- Σε περίπτωση δυσλειτουργίας κάποιας φυσικής σύνδεσης (link ή κάποιου δρομολογητή (Στην χειρότερη περίπτωση αυτού που δρα ως route reflector), υπάρχουν ανυπολόγιστες συνέπειες εντός του AS. Για την περίπτωση προστασίας από κατάρρευση του Route Reflector, υπάρχει η δυνατότητα χρήσης Redundant Route Reflector (Bates, T & Chen , E & Chandra, R.(2006)) , που όμως κατά ελάχιστο διπλασιάζει τον αριθμό των iBGP συνδέσεων, καταργώντας πολλά από τα πλεονεκτήματα του route reflection και κάνοντας το ασύμφορο για μικρές τοπολογίες.

Είναι προφανές πως τα παραπάνω αποτελούν μειονέκτημα της σημερινής υλοποίησης του πρωτοκόλλου δρομολόγησης iBGP. Τα άρθρα που αναφέρθηκαν σε προηγούμενο κεφάλαιο (κεφ.4), προτείνουν λύσεις οι οποίες καλύπτουν κάποια από τα υπάρχοντα ζητήματα, ενώ συνήθως θίγουν τα πλεονεκτήματα του route reflection σε ένα βαθμό.

Από την μελέτη της βιβλιογραφίας, γίνεται εύκολα αντιληπτό πως λίγα έχουν προταθεί ως εναλλακτική λύση στο Redundant Route Reflection που προτείνεται από τους (Bates, T & Chen , E & Chandra, R.(2006)). Επίσης, λίγα άρθρα (Buod, M & Uhlig, S & Muelle, M. (2008)), ασχολούνται με την πιθανότητα αλλαγής της τοπολογίας π.χ λόγω βλάβης, και αυτά προτείνουν σχεδιασμό με προϋπολογισμό πιθανής βλάβης προτείνοντας τοπολογία με τα διπλάσια session σε σχέση με το απλό Route Reflection.

Θα προσπαθήσουμε αξιοποιώντας τις προτάσεις που έχουν γίνει για την βελτίωση του πρωτοκόλλου iBGP, να προτείνουμε μια υλοποίηση που να προσφέρει:

- Μικρότερο αριθμό συνδέσεων iBGP σε σχέση με το full-mess iBGP.
- Προστασία από την δημιουργία βρόχων δρομολόγησης (routing loops) και ανακλάσεις πακέτων (deflections).
- Σταθερότητα, ακόμη και στην περίπτωση αποτυχίας πολλών δρομολογητών ή συνδέσεων, κατά το δυνατόν εφάμιλλη του full-mess iBGP.
- Επιλογή βέλτιστης διαδρομής από τους route reflector clients.

5.1 Περιγραφή πρωτοκόλλου Standby Route Reflection

Η λύση που προτείνουμε αφορά την επέκταση του πρωτοκόλλου BGP-4 (κεφ.5.2), ώστε να υποστηρίζει την λειτουργικότητα που περιγράφεται παρακάτω.

Η πρόταση μας αποτελεί βελτίωση στην περίπτωση χρήσης Redundant Route Reflectors. Λαμβάνοντας υπόψη ότι σε περιβάλλον παρόχου υπηρεσιών διαδικτύου, συνηθίζεται οι δρομολογητές του πελάτη (CPE) να είναι RR Client, θέτουμε ως στόχο την ελαχιστοποίηση της απαιτούμενης μνήμης και υπολογιστικής ισχύος από πλευράς RR Client. Επίσης, θεωρούμε απαράδεκτη οποιαδήποτε λύση απαιτεί επέκταση ή αντικατάσταση του πρωτοκόλλου BGP-4 στους RR Clients, διότι κάτι τέτοιο θα οδηγούσε σε υποχρεωτική αντικατάσταση ενεργού εξοπλισμού από πλευράς πελάτη.

Εντός ενός Cluster, σύμφωνα με το Redundant Route Reflection που έχει προταθεί (Bates, T & Chen , E & Chandra, R.(2006)), υπάρχουν τουλάχιστο δύο δρομολογητές που δρουν ως Route Reflectors, ενώ οι υπόλοιποι είναι RR Clients. Στην περίπτωση αυτή, κάθε RR Client διατηρεί δύο iBGP συνδέσεις, μία ανά RR. Προφανώς, είναι πολύ πιθανό ο RR Client να μαθαίνει τις ίδιες BGP διαδρομές από κάθε RR. Αποτέλεσμα είναι εκ πρώτης όψεως ο διπλασιασμός των BGP συνδέσεων, καθώς και της απαιτούμενης μνήμης. Αν μελετήσουμε σε βάθος το ζήτημα θα διαπιστώσουμε πως λόγω της ομοιότητας των διαδρομών που προέρχονται από τους RR, ο κάθε client αναγκάζεται να εκτελεί την BGP διαδικασία επιλογής διαδρομής για μεγάλο πλήθος διαδρομών και μάλιστα μέχρι το σημείο (tie-breaker) όπου γίνεται σύγκριση του BGP-ID των δύο RR. Έτσι, υπάρχει δραματική απαίτηση και σε υπολογιστική ισχύ.

Η πρόταση μας ορίζει τουλάχιστο δύο Standby RR ανά Cluster. Κατά την αρχική προσπάθεια του standby RR να δημιουργήσει γειτονικό δεσμό BGP με άλλο standby RR, οι δρομολογητές διαφημίζουν τις επιπλέον δυνατότητες τους, όπως αρχικά προτάθηκε από τους (Chandra, R & Scudder, J.(2002)). Εφόσον κάποιος δρομολογητής δεν υποστηρίζει τη δυνατότητα, διαμορφώνεται μια τυπική διασύνδεση iBGP. Φυσιολογικά, οι δύο δρομολογητές συμφωνήσουν στην υποστήριξη της συγκεκριμένης δυνατότητας.

Στο σημείο αυτό, οι δύο standby Route Reflectors έχουν διαμορφώσει σύνδεση iBGP (Established) και έχουν αποδεχτεί την υποστήριξη λειτουργικότητας Standby Route Reflection.

Η επιπλέον σηματοδότηση που πρόκειται να χρησιμοποιηθεί από το πρωτόκολλο standby Route Reflection, θα υλοποιηθεί πάνω σε νέο τύπο BGP μηνύματος, αριθμό στο οποίο θα πρέπει να απόδοση η αρμόδια αρχή (IANA). Για εργαστηριακούς σκοπούς χρησιμοποιείται αυθαίρετα ο BGP τύπος μηνύματος 0x07 (7), ο οποίος ύστερα από έρευνα δεν βρέθηκε να χρησιμοποιείται από κάποια γνωστή επέκταση του πρωτοκόλλου BGP-4.

Σε κάθε Cluster ένας δρομολογητής αποκτά τον ρόλο του master και είναι υπεύθυνος για τον καθορισμό διαφόρων παραμέτρων που επεξηγούνται εκτενώς παρακάτω. Κάθε standby RR, ανεξαρτήτως του IGP πρωτοκόλλου που χρησιμοποιείται στο AS γνωρίζει την απόσταση (metric) προς κάθε RR Client. Έτσι, κάθε standby RR δημιουργεί λίστα με δυάδες <BGP ID(IP),metric> που ανταλλάσσει με το γειτονικό standby RR, μέσω BGP μηνυμάτων τύπου 7. Κάθε RR συγκρίνει τη λίστα που παρέλαβε με τη δική του και αποφασίζει αν θα υποστηρίξει τη δεδομένη χρονική στιγμή τον client (active) ή όχι (passive). Σε κάθε περίπτωση όλοι οι client θα εξυπηρετηθούν από ένα standby RR. Η επιλογή βασίζεται στο IGP metric. Από πλευράς RR Client, υπάρχει η ρύθμιση ώστε να δημιουργηθούν iBGP συνδέσεις και με τους δύο standby RR. Οι standby RR αποδέχονται και δημιουργούν iBGP συνδέσεις με τους RR clients. Οι δύο standby RR αποδέχονται και επεξεργάζονται μηνύματα Update, ενώ ανταλλάσσουν Keepalive μηνύματα με τους RR client για την διατήρηση της σύνδεσης iBGP. Η διαφοροποίηση οφείλεται στο γεγονός πως μόνο ο active standby RR αποστέλλει μηνύματα τύπου update στους RR client.

Συνεπώς, το ζευγάρι των standby RR έχει επιλέξει τους RR client για τους οποίους είναι active η passive. Σε περίπτωση μεταβολής της IGP ή BGP τοπολογίας, γίνεται ανταλλαγή των μεταβολών σε δυάδες <BGP ID(IP),metric> και οι standby RR αποφασίζουν εκ νέου τον ρόλο active/passive ανά RR client. Σε περίπτωση απώλειας της iBGP σύνδεσης με το γειτονικό standby RR και ασχέτως του λόγου που προκάλεσε κάτι τέτοιο, ο standby Route Reflector, άμεσα λαμβάνει το ρόλο active για όλους τους client.

Λόγω της υψηλής σημασίας της iBGP σύνδεσης μεταξύ των δύο ή περισσότερων Standby RR, ορίζεται ένας εξ αυτών ως master και οι υπόλοιποι ως slave. Ο master, σε περίπτωση ισοπαλίας των IGP metric προς κάποιο client γίνεται active και οι υπόλοιποι standby RR passive. Επίσης, ο Master, παραβιάζοντας την κλασική μέθοδο επιλογής Hold Time και Keepalive timer του BGP-4, ορίζει προαιρετικά νέους μικρότερους BGP timers για την ειδική iBGP διασύνδεση των standby RR. Αυτό, γίνεται για την μείωση του απαιτούμενου χρόνου σταθεροποίησης του πρωτοκόλλου σε περίπτωση μεταβολών ή δυσλειτουργιών που λαμβάνουν χώρα εντός του AS.

Για να πετύχουμε την επιλογή καλύτερης διαδρομής ακόμη και σε αποτυχία αρκετών συνδέσεων ή δρομολογητών, πρέπει κάθε redundant route reflector να δρα ως route reflector για τους λοιπούς route reflector με την απαίτηση αυτοί να δέχονται το update παρόλο που βλέπουν το cluster-id τους εντός του cluster-list παραβιάζοντας τα μέχρι τώρα γνωστά. Για να μην προκαλέσει προβλήματα αυτό, στη διαδικασία επιλογής διαδρομής BGP και συγκεκριμένα στο βήμα που επιλέγεται διαδρομή που

έγινε γνωστή μέσω eBGP σε σχέση με iBGP προσθέτουμε την προτίμηση μιας iBGP διαδρομής από μια iBGP διαδρομή που προήλθε από reflection.

5.2 Ενσωμάτωση του Standby RR στο BGP

Το BGP-4, είναι πρωτόκολλο δρομολόγησης σχεδιασμένο ώστε να ανταλλάσει τρίπτυχα (Type-Length-Value) TLV μεταξύ των δρομολογητών. Έτσι, υπάρχει η ευελιξία να γίνει ανταλλαγή διαφόρων ειδών πληροφορίας μέσω του BGP. Έτσι, η επέκταση που προτείνουμε για το BGP μπορεί να υλοποιηθεί με ένα νέο τύπο μηνύματος πάνω σε υπάρχουσα εφαρμογή BGP. Πλεονέκτημα της λύσης που προτείνουμε είναι η ανάγκη να υλοποιείται η επέκταση του πρωτοκόλλου BGP-4 μόνο στους δρομολογητές που εκτελούν χρέη Redundant RR. Επίσης, οι δρομολογητές αυτοί μπορούν να αντικατασταθούν – αναβαθμιστούν σταδιακά, χωρίς να δημιουργούνται ασυμβατότητες.

Για να γίνει εφικτή η σταδιακή εξάπλωση της επέκτασης BGP-4 στο AS, έχουν δημιουργηθεί οι κατάλληλοι μηχανισμοί που επιτρέπουν σε αναβαθμισμένους δρομολογητές να συνεργαστούν με κλασικούς δρομολογητές.

Κατά την έναρξη δημιουργίας μιας σύνδεσης BGP, γίνεται ανταλλαγή μηνυμάτων τύπου OPEN. Κάθε δρομολογητής ελέγχει την ορθότητα του μηνύματος OPEN που έλαβε και εξετάζει τη λίστα με τις προαιρετικές ικανότητες που διαφημίζει ο γειτονικός δρομολογητής.

Σε περίπτωση ασυμφωνίας, κάθε δρομολογητής μπορεί να αντιδράσει με μήνυμα τύπου Notification και κωδικό σφάλματος που να χαρακτηρίζει την μη υποστηριζόμενη λειτουργία (Unsupported Capability). Σε αυτή την περίπτωση πρέπει οι δύο δρομολογητές να ξαναπροσπαθήσουν την δημιουργία σύνδεσης BGP χωρίς να περιλαμβάνουν στη λίστα δυνατοτήτων τις συγκεκριμένες δυνατότητες που προκάλεσαν την ασυμβατότητα. Έτσι οι δρομολογητές δημιουργούν σύνδεση BGP (Established). Για απόδειξη των παραπάνω, αναφερθείτε στο πείραμα συμβατότητας (κεφ.7.4).

Σε περίπτωση συμφωνίας (Optional Capability 0x100) των δύο δρομολογητών να χρησιμοποιήσουν την επέκταση πρωτοκόλλου Standby Route Reflection (0x07), οι δρομολογητές δημιουργούν σύνδεση iBGP και ανταλλάσσουν επιπλέον μηνύματα (τύπου 7) ώστε να καθοριστούν οι παράμετροι λειτουργίας και να ανταλλαχθούν πληροφορίες σχετικές με την επέκταση του πρωτοκόλλου.

5.3 Τύποι μηνυμάτων

Τα μηνύματα τύπου 7 που ανταλλάσσουν δύο δρομολογητές, οι οποίοι υποστηρίζουν την επέκταση standby route reflection, μεταφέρονται πάνω σε εδραιωμένη σύνδεση iBGP μεταξύ των δρομολογητών. Οι υπο-τύποι μηνύματος τύπου 7 που απαιτεί η επέκταση είναι τρεις:

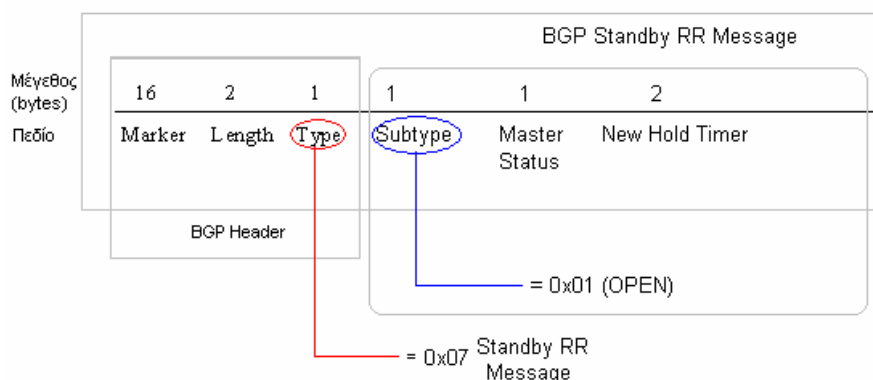
- Subtype OPEN (0x01) : Παρομοίως με το BGP μήνυμα OPEN, ο υποτύπος OPEN ανταλλάσσεται αμέσως μετά την εδραίωση της σύνδεσης BGP και περιλαμβάνει διάφορες παραμέτρους.
- Subtype UPDATE_ADD (0x02) : Χρησιμοποιείται για να δηλώσει την απόσταση metric από τους RR clients.
- Subtype UPDATE_WTD (0x03) : Χρησιμοποιείται για να δηλώσει την αδυναμία πρόσβασης σε κάποιους RR clients.

Κάθε μήνυμα τύπου 7 μπορεί να περιέχει έναν υποτύπο OPEN ή να περιέχει ταυτόχρονα υποτύπους 0x02 ή 0x03.

5.3.1 τύπος SUBTYPE_OPEN

Το μήνυμα υποτύπου OPEN (0x01) μηνύματος BGP, τύπου Standby Route Reflection (0x07) είναι συγκεκριμένων διαστάσεων (23 bytes) και ενθυλακώνεται φυσιολογικά σε πακέτο τύπου TCP/IP πριν τη μεταφορά του σε φυσικό επίπεδο.

Εφόσον το πεδίο Type του BGP Header αποκτήσει την τιμή 0x07, γίνεται κατανοητό πως το πακέτο αφορά BGP Standby RR's και δεν πρέπει να συναντηθεί σε σύνοδο BGP που δεν αφορά Redundant RR's. Το αμέσως επόμενο πεδίο παίρνει την τιμή του υποτύπου (Subtype) που στην περίπτωση του μηνύματος OPEN είναι 0x01.



Εικόνα 5.3.1 : BGP μήνυμα Standby RR υποτύπου OPEN

Μόλις ανιχνευτεί μήνυμα υποτύπου OPEN, αναμένονται δύο πεδία:

- **Master Status:** Αποτελεί πεδίο μεγέθους 1 byte και λαμβάνει τιμές 0x00 (Slave) ή 0x01 (Master). Μόνο ένας δρομολογητής μπορεί να τεθεί επικεφαλής (Master) σε συνομιλία standby Route Reflection. Ο master επιτάσσει νέους χρόνους Hold timer και Keepalive Interval timer στην iBGP σύνδεση μεταξύ των standby RR. Οι χρόνοι αυτοί κατά κανόνα, είναι αρκετά μικρότεροι από τους προκαθορισμένους του BGP και έχουν ως στόχο την έγκαιρη διάγνωση προβλήματος μεταξύ της διασύνδεσης δύο standby RR. Θα πρέπει να δοθεί ιδιαίτερη προσοχή ώστε οι χρόνοι αυτοί να ρυθμιστούν αρκετά μεγάλοι ώστε να αποφευχθεί υψηλή χρησιμοποίηση της υπολογιστικής ισχύος των δρομολογητών από την πρόκληση συχνών interrupts. Επίσης, ο Master δρομολογητής, αναλαμβάνει να εξυπηρετήσει (active) τους RR Clients για τους οποίους είναι όμοια τα κριτήρια μεταξύ των δύο standby RR δρομολογητών.
- **New Hold Timer :** αποτελεί πεδίο 2 bytes όπως και το Hold time του BGP OPEN μηνύματος και αφορά την δυνατότητα του δρομολογητή να προσαρμοστεί σε νέους χρόνους Hold timer και Keepalive Interval Timer.

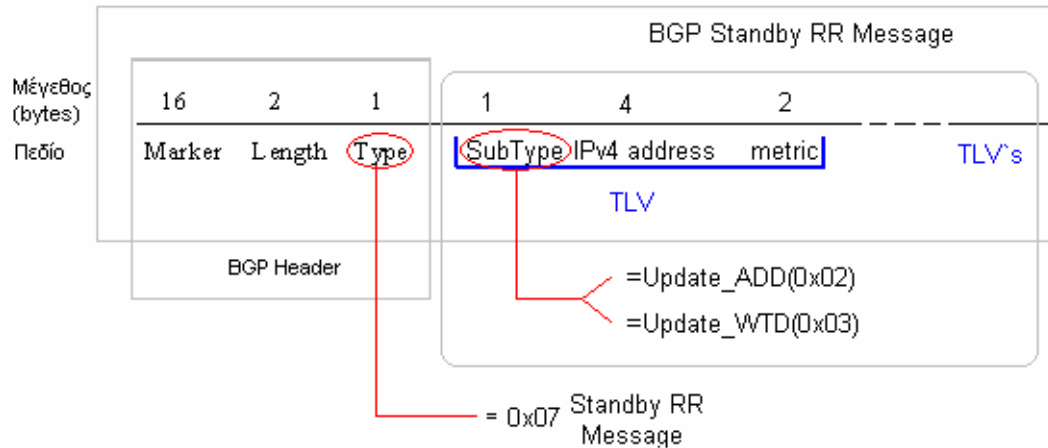
Μετά από την ανταλλαγή BGP standby μηνυμάτων υποτύπου OPEN, οι δύο δρομολογητές είναι έτοιμοι να ανταλλάξουν πληροφορίες που αφορούν την απόσταση (metric) των RR client. Στο σημείο αυτό, δεν αποστέλλονται BGP μηνύματα τύπου Update σε κανένα RR Client.

5.3.2 τύπος SUBTYPE_UPDATE_ADD & WTD

Η επέκταση του BGP-4 Standby Route Reflection, χρησιμοποιεί BGP πακέτα τύπου Standby RR (0x07), ώστε να ανταλλάξει τις απαραίτητες πληροφορίες για την σωστή εξυπηρέτηση των RR Clients(εικ.5.3.2). Τα μηνύματα αυτά αποτελούνται από τριάδες TLV σταθερού μεγέθους. Κάθε τριάδα περιέχει τα εξής πεδία :

- **Subtype :** Είναι πεδίο ενός byte.Λαμβάνει τιμές 0x02 (Update_ADD) ή 0x03 (Update_WTD) οι υπόλοιπες τιμές είναι για μελλοντική χρήση.
- **IPv4 Address :** Αποτελεί την unicast IPv4 διεύθυνση που αντιστοιχεί σε RR Client. Χρειάζεται 4-bytes.
- **Metric :** Αποτελεί το κόστος, βάσει του IGP πρωτοκόλλου, που χρειάζεται ο Standby RR για να φτάσει τον RR Client. Η παρούσα υλοποίηση χρησιμοποιεί 2-bytes. Σε μελλοντική υλοποίηση, θα μπορούσε να μετατραπεί σε δύο πεδία Metric Length (1-byte) για καθορισμό του μεγέθους του metric και metric(Variable). Κάτι τέτοιο, θα δημιουργούσε την ανάγκη για χρήση πεδίου Length(1-byte) που θα χαρακτήριζε το τώρα σταθερό, μέγεθος του TLV.

Στην ουσία η κάθε τριάδα χαρακτηρίζεται από το πρώτο πεδίο (SubType). Κάθε δρομολογητής αρχικά διαφημίζει τα IGP metrics προς τους RR Client μέσω υποτύπου (Update_ADD). Εφόσον γίνει πλήρη ανταλλαγή οι δρομολογητές διαφημίζουν τους RR Client για τους οποίους θεωρούν πως έχουν μεγαλύτερο κόστος σε σχέση με τον γείτονα Standby RR κάνοντας χρήση υποτύπου (Update_WTD).



Εικόνα 5.3.2 : BGP μήνυμα Standby RR υποτύπου UPDATE

Το συνολικό μήκος του BGP πακέτου, όπως ορίζεται στο BGP header, επιτρέπει τον υπολογισμό του συνολικού αριθμού TLV's που περιέχει το κάθε μήνυμα, καθώς αυτά είναι σταθερού μεγέθους (7-bytes).

5.4 Επίδραση στους timers του BGP

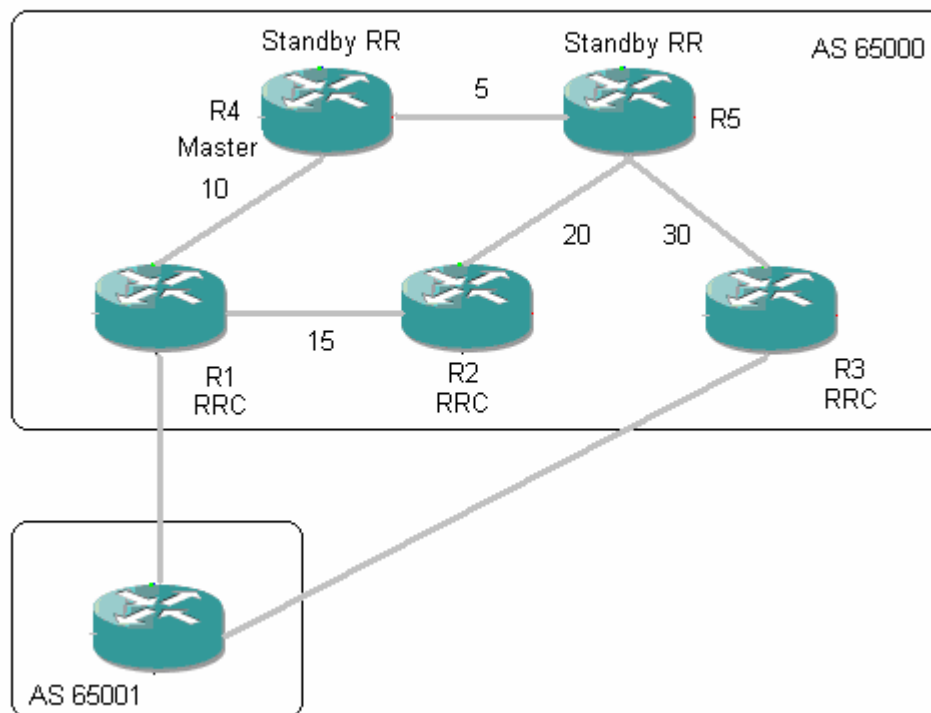
Το BGP προσπαθεί να διαμορφώνει συνδέσεις iBGP σε κατάσταση (FSM:Established) με τους γειτονικούς δρομολογητές. Μετά την ανταλλαγή μηνυμάτων OPEN (FSM:OpenConfirm) ορίζεται ο Hold timer ως ο μεγαλύτερος μεταξύ των δύο Hold Timer που ανταλλάχθηκαν μεταξύ των δρομολογητών με χρήση μηνύματος BGP OPEN. Όταν εδραιώνεται η σύνδεση (FSM:Established) ο Hold Timer έχει πάρει τιμή. Ο Hold Timer μετρά αντίστροφα και επανεκκινείται κάθε φορά που λαμβάνεται BGP μήνυμα από τον γειτονικό δρομολογητή. Αν για κάποιο λόγο ο Hold Timer λήξει, τότε η σύνδεση BGP θεωρείται προβληματική. Εν συνεχεία αποστέλλεται μήνυμα τύπου Notification (Hold Timer Expired) και κλείνει η σύνδεση (FSM:Idle). Ο Hold Timer συνήθως είναι τις τάξης μεγέθους των λεπτών, ενώ επιτρέπεται ελάχιστη τιμή 3 sec. Για την επιτυχή διατήρηση μιας σύνδεσης κάθε δρομολογητής στέλνει μηνύματα BGP Keepalive σε χρονικό διάστημα Keepalive Interval Timer. Το χρονικό διάστημα ορίζεται σαν το (1/3) του Hold Timer που έχει συμφωνηθεί.

Όταν δημιουργείται μια σύνδεση iBGP μεταξύ δύο standby Route Reflectors, κάθε πιθανή αποτυχία της μεταξύ τους φυσικής διασύνδεσης ή της iBGP διασύνδεσης είναι πιθανό να οδηγήσει σε απώλεια πακέτων εντός του AS. Οι standby Route Reflectors διαθέτουν τους μηχανισμούς για επίλυση του προβλήματος μέσω

δυναμικής μεταβολής της τοπολογίας. Παρόλα αυτά, ο standby Route Reflector πολλές φορές δεν μπορεί να εντοπίσει την αποτυχία της διασύνδεσης μέχρι την λήξη του Hold Timer. Έτσι, με την χρήση των μηνυμάτων BGP Standby Route reflection (κεφ.5.3.2), γίνεται μεταβολή των Hold Timer και Keepalive Interval Timer σε πολύ μικρότερες τιμές. Αποτέλεσμα είναι η γρήγορη αναγνώριση σφαλμάτων και μείωση του χρόνου ομαλοποίησης (convergence) του BGP.

5.5 Διαδικασία επιλογής πελατών ανά RR

Οι δύο δρομολογητές R4,R5 έχουν διαμορφώσει iBGP σύνδεση με τα χαρακτηριστικά του standby Route Reflection και έχουν ανταλλάξει μεταξύ τους δίπτυχα (IP,metric). Αποτέλεσμα αυτής της συναλλαγής, είναι κάθε standby RR να έχει διαμορφώσει δύο γραφικές παραστάσεις, μία με την απόσταση κάθε πιθανού RR Client από τον εαυτό του και μία με την απόσταση κάθε πιθανού RR Client από τον γειτονικό Standby RR.



Εικόνα 5.5.1 : BGP τοπολογία με Standby RR (Εμφανίζονται τα IGP metrics και οι φυσικές διασυνδέσεις)

Έτσι, το ζευγάρι Standby RR έχει δύο όμοιες γραφικές παραστάσεις προς τους πιθανούς RR Client (πιν.5.5.2 και πιν.5.5.3).

Πιθανός RR Client	Δική μου απόσταση	Απόσταση του Γείτονα (R5)
R1	10	15
R2	25	20
R3	35	30

Πίνακας 5.5.2 : Δεδομένα γραφικής παράστασης IGP στον R4

Πιθανός RR Client	Δική μου απόσταση	Απόσταση του Γείτονα (R4)
R1	15	10
R2	20	25
R3	30	35

Πίνακας 5.5.3 : Δεδομένα γραφικής παράστασης IGP στον R5

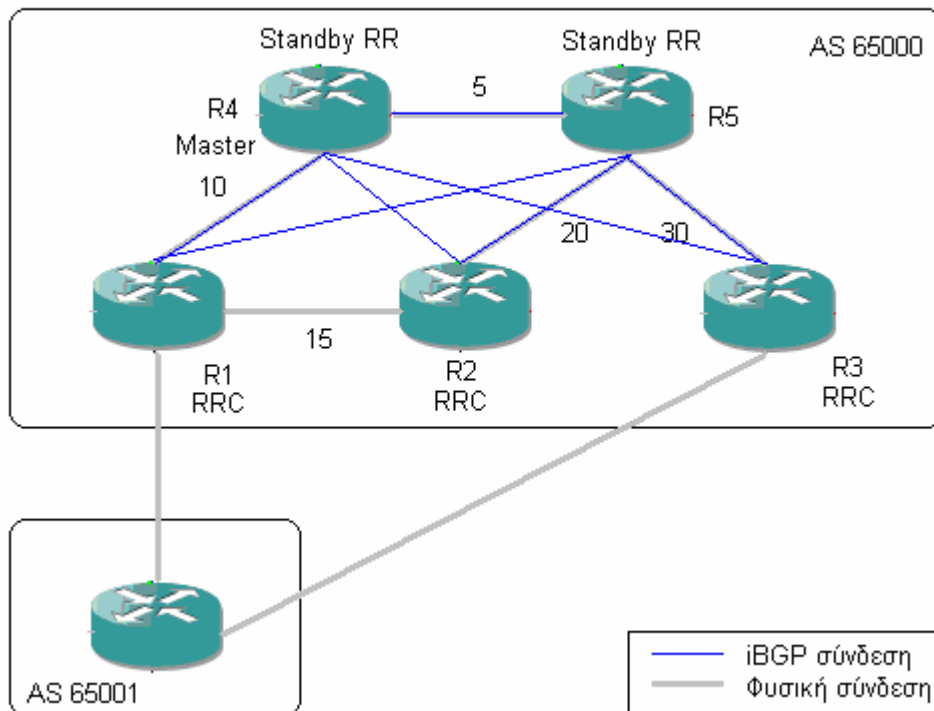
Στο σημείο αυτό, κάθε Standby Route Reflector έχει διαμορφώσει όλες τις απαραίτητες iBGP συνδέσεις (εικ.5.5.4). Ο αριθμός των iBGP συνδέσεων (7) είναι πάντα ίσος με τον αριθμό που θα απαιτούταν σε περίπτωση χρήσης τυπικών Redundant Route Reflectors και επίσης μικρότερος ή ίσος σε περίπτωση διασύνδεσης iBGP πλήρους πλέγματος (9).

Αμέσως μετά την διαμόρφωση των iBGP συνδέσεων με τους RR Client, οι Standby RR δέχονται μηνύματα Update και ανταλλάσσουν Keepalive με τους RR Client. Στο σημείο αυτό, όμως, δεν αποστέλλονται μηνύματα Update σε κανένα RR Client. Κάθε Standby RR, κάνοντας χρήση των γραφικών παραστάσεων (πιν.5.5.2 και πιν.5.5.3), δημιουργεί δύο λίστες, την Active και την Passive. Στην λίστα Active, τοποθετούνται οι RR Clients για τους οποίους ο Standby RR έχει μικρότερο IGP κόστος, ενώ στη λίστα Passive, τοποθετούνται οι RR Clients για τους οποίους ο Standby RR έχει ασύμφορο metric. Στην περίπτωση που οι δύο δρομολογητές έχουν ίσο κόστος προς κάποιους RR Client, ο Standby RR που αρχικά ορίστηκε Master, αναλαμβάνει να τον εισαγάγει στην λίστα Active. Κάθε RR Client, μπορεί να βρίσκεται μόνο σε μια λίστα Active ενός standby RR, ενώ μπορεί να βρίσκεται σε πολλές Passive λίστες λοιπών δρομολογητών Standby RR.

Έτσι, βάσει των παραπάνω για την τοπολογία (εικ.5.5.4) οι R4, R5 έχουν τις παρακάτω λίστες.

$$\text{Active}_{(R4)} = \{R1\} \text{ και } \text{Passive}_{(R4)} = \{R2,R3\}$$

$$\text{Active}_{(R5)} = \{R2,R3\} \text{ και } \text{Passive}_{(R5)} = \{R1\}$$

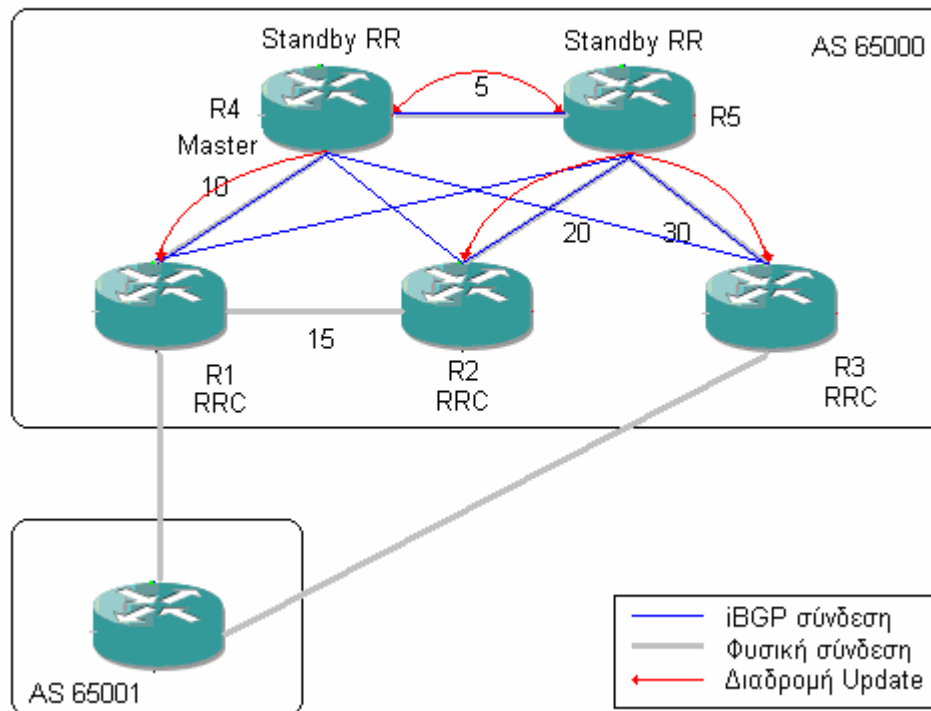


Εικόνα 5.5.4 : BGP τοπολογία με Standby RR (Εμφανίζονται τα IGP metrics και οι φυσικές διασυνδέσεις καθώς και οι iBGP συνδέσεις)

Οι Standby RR R4,R5 είναι έτοιμοι να εξυπηρετήσουν τους πελάτες RR Client. Τώρα, κάθε Standby RR αποστέλλει Update στους δρομολογητές που περιλαμβάνονται στη λίστα Active. Οι δρομολογητές που βρίσκονται στη λίστα Passive εξυπηρετούνται από άλλο Standby RR.

Με τη μέθοδο αυτή, δεν περιορίσαμε τον αριθμό συνδέσεων iBGP που απαιτούνται σε σχέση με τυπικό Redundant RR. Περιορίσαμε την περιττή πληροφορία και απαλλάξαμε τους RR Client από διπλοεγγραφές των ιδίων διαδρομών καθώς και την περιττή χρήση του αλγορίθμου επιλογής καλύτερης διαδρομής.

Στην τοπολογία που πραγματευόμαστε υπάρχουν τρεις ενεργές διαδρομές Update προς RR Client (εικ.5.5.5). Ενώ, σε τυπική τοπολογία Redundant RR θα υπήρχαν έξι ενεργές διαδρομές Update προς τους RR Client. Με το παράδειγμα αυτό γίνεται κατανοητή η μέθοδος του Standby Route Reflection και η διαφοροποίηση αυτής από το τυπικό Redundant Route Reflection. Ο αριθμός των ενεργών διαδρομών μηνυμάτων Update ισούται πάντα με τον αριθμό των iBGP συνδέσεων που θα είχαμε σε περίπτωση χρήσης Route Reflection με χρήση ενός RR.



Εικόνα 5.5.5 : BGP τοπολογία με Standby RR (Εμφανίζονται τα IGP metrics ,οι φυσικές διασυνδέσεις , οι iBGP συνδέσεις και οι Διαδρομή Update)

5.6 Διάδοση βέλτιστης διαδρομής

Πολύ σημαντικό ζήτημα σε όλες τις παραλλαγές του Route Reflection είναι η επιλογή μη βέλτιστης διαδρομής από τους RR Clients. Αυτό οφείλεται στην μειωμένη αντίληψη των RR Clients λόγω της ελλιπούς πληροφορίας που δέχονται.

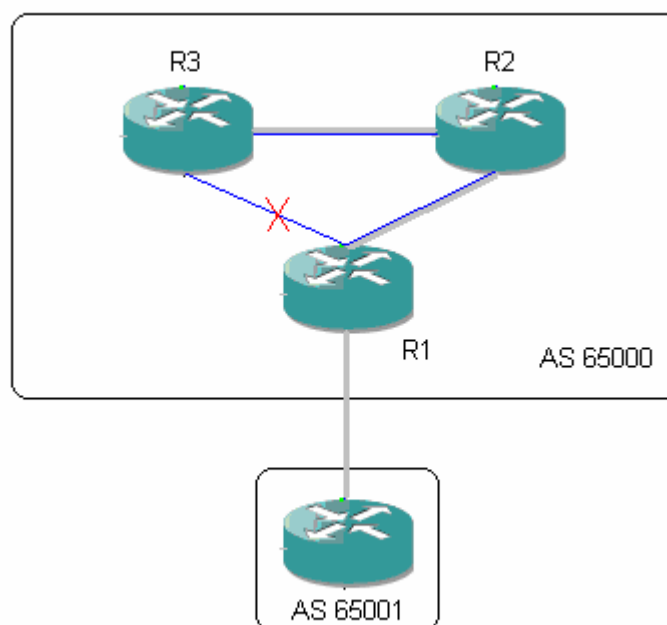
Προηγουμένως, (κεφ.4.2), αναφέρθηκαν προτάσεις για την επίλυση του θέματος. Η πρόταση μας για Standby Route Reflectors, δεν προσφέρει από μόνη της κάποια βελτίωση στο πρόβλημα, όταν γίνεται χρήση πολλών Clusters. Έτσι, είναι σημαντικό το πρωτόκολλο αυτό να ακολουθεί τις προτάσεις (κεφ.4.2) για διάδοση πολλών διαδρομών προς τους RR Client καθώς και τις σχεδιαστικές λύσεις για σωστή τοποθέτηση των RR εντός του AS βάσει των μαθηματικών μοντέλων που παρουσιάστηκαν προηγουμένως (κεφ.4.2).

5.7 Αντίδραση σε περίπτωση δυσλειτουργίας

Η πρόταση Standby Route Reflection, παρουσιάζει βέλτιστα αποτελέσματα σε σχέση με όλες τις γνωστές προτάσεις για βελτίωση της αξιοπιστίας του iBGP με χρήση Route Reflection.

Το τυπικό Redundant Route Reflection παρουσιάζει αντίστοιχη αξιοπιστία με το Standby Route Reflection. Ο βαθμός αξιοπιστίας που προσφέρουν και οι δύο λύσεις εξαρτάται άμεσα από τον αριθμό Route Reflector ανά cluster. Έτσι, και οι δύο λύσεις παρουσιάζουν εφάμιλλα αποτελέσματα με την τοπολογία πλήρους πλέγματος σε συνθήκες μεταβολής της τοπολογίας λόγω δυσλειτουργιών.

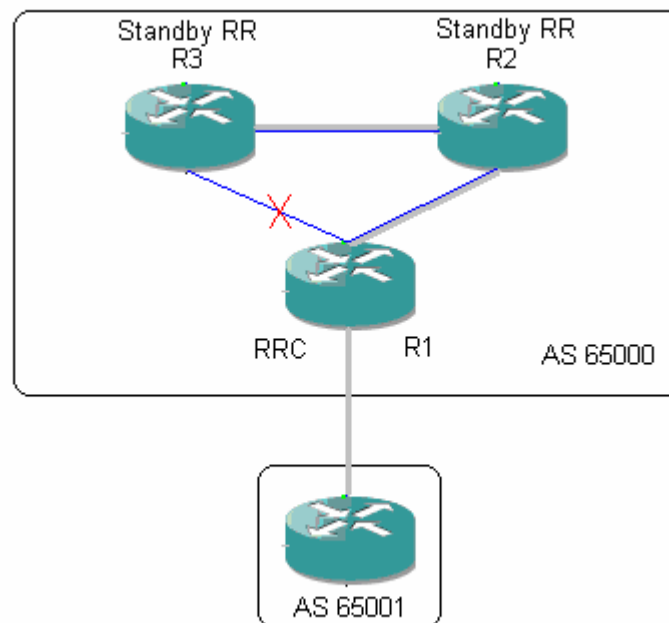
Πολλές φορές, παρόλο που υπάρχει διαθέσιμη η φυσική τοπολογία δημιουργούνται προβλήματα με τις iBGP συνδέσεις με αποτέλεσμα την απώλεια διαδρομών από τους πίνακες δρομολόγησης BGP. Σε τοπολογία πλήρους πλέγματος, (εικ.5.7.1), παρατηρείται δυσλειτουργία της iBGP σύνδεσης μεταξύ R1-R3. Ο R1 μαθαίνει μέσω eBGP την διαδρομή προς το δίκτυο 10.0.0.0/8 από το γειτονικό AS 65001. Ο R1 διαφημίζει μέσω iBGP την διαδρομή προς το 10.0.0.0/8 στον R2. Ο R2 τοποθετεί την διαδρομή στον BGP πίνακα δρομολόγησης και γνωρίζει πλέον διαδρομή προς το δίκτυο 10.0.0.0/8. Ο R2 δεν διαφημίζει τη διαδρομή 10.0.0.0/8 στον R3, λόγω των κανόνων split horizon που εισάγει το BGP-4. Αναλυτικότερα, ο R2 έμαθε, μέσω iBGP, την διαδρομή προς το δίκτυο 10.0.0.0/8 και δεν πρόκειται να την διαφημίσει μέσω iBGP. Επειδή ο R2 και ο R3 ανήκουν στο AS 65000, έχουν iBGP σύνδεση. Έτσι, ο R3 δεν έχει γνώση του δικτύου 10.0.0.0/8.



Εικόνα 5.7.1 : BGP τοπολογία πλήρους πλέγματος(Απώλεια iBGP σύνδεσης R1-R3)

Στο παραπάνω παράδειγμα αποδεικνύεται πως, παρόλο που υπάρχει το φυσικό μονοπάτι από τον R3 προς το δίκτυο 10.0.0.0/8, το BGP δεν καταφέρνει να διαδώσει την πληροφορία μέχρι τον R3, ακόμη και με τοπολογία πλήρους πλέγματος.

Στην ίδια τοπολογία, (εικ.5.7.2) οι R2 και R3 έχουν οριστεί ως Standby Route Reflectors και ο R1 ως Client. Λόγω της μη ύπαρξης φυσικής διασύνδεσης μεταξύ R1 – R3, είναι δεδομένο πως το IGP κόστος του R3 προς το R1 είναι μεγαλύτερο από το αντίστοιχο του R2 προς το R1. Έτσι, ο R2 βάζει στη λίστα Active τον R1 και ο R3 στην Passive. Οι R2 και R3 έχουν διαμορφώσει σύνδεση iBGP και υλοποιούν την επέκταση του πρωτοκόλλου BGP-4, Standby Route Reflection. Συνεπώς, ο R2 ανακλά τις iBGP διαδρομές στον R3 και αντιστρόφως, παρόλο που βρίσκονται στο ίδιο cluster. Υπό φυσιολογικές συνθήκες, οι δρομολογητές R2 και R3 μαθαίνουν για την 10.0.0.0/8 από τον R1 και από τον γειτονικό standby RR, R3 και R2 αντιστοίχως. Μέσω της διαφοροποιημένης μηχανής επιλογής καλύτερης διαδρομής, επιλέγεται ως καλύτερη η διαδρομή που έγινε γνωστή μέσω iBGP (R1) και όχι αυτή που έγινε γνωστή μέσω Reflection (R2,R3).



Εικόνα 5.7.2 : BGP τοπολογία Standby RR (Απώλεια iBGP σύνδεσης R1-R3)

Τώρα, όταν δημιουργείται βλάβη στην iBGP σύνδεση R1-R3, ο R3 γνωρίζει την 10.0.0.0/8 μέσω του R2 (Reflection) και τα πακέτα προωθούνται φυσιολογικά.

Το BGP-4 ορίζει τον κανόνα της μη διαφήμισης διαδρομής που έγινε γνωστή, μέσω iBGP, σε άλλο iBGP γείτονα για την αποφυγή δημιουργίας βρόχου δρομολόγησης. Το Standby Route reflection, όπως και το τυπικό Route Reflection, παραβαίνει αυτόν τον κανόνα και ανακλά της iBGP διαδρομές σε άλλους iBGP γείτονες. Για την προστασία από βρόχους δρομολόγησης γίνεται χρήση των cluster (κεφ.2). Με τη λογική των clusters, δεν έχει νόημα σε ένα ζευγάρι Redundant RR να γίνεται μεταξύ τους ανάκλαση διαδρομών (reflection). Προφανώς κάτι τέτοιο θα

οδηγούσε σε συνεχόμενη απόρριψη μηνυμάτων Update μεταξύ των R2,R3, γιατί έκαστος θα εντοπίζει το δικό τους Cluster-id στο cluster-list που περιέχεται στο μήνυμα Update.

Το Standby Route Reflection καταργεί τον κανόνα του Cluster. Έτσι, ένας δρομολογητής που εντοπίζει το δικό του cluster-id σε μήνυμα Update, δέχεται και επεξεργάζεται το μήνυμα. Στο σημείο αυτό, ο κίνδυνος εμφάνισης βρόχων δρομολόγησης γίνεται εμφανής.

Το BGP, κατά την διαδικασία επιλογής βέλτιστης διαδρομής, φτάνει σε βήμα που επιλέγει διαδρομή που έγινε γνωστή μέσω eBGP παρά μέσω iBGP. Στο Standby Route Reflection για να αποφύγουμε τους βρόχους δρομολόγησης, μεταβάλλουμε το βήμα αυτό. Έτσι, προτιμάται μια διαδρομή που έγινε γνωστή μέσω eBGP από μια διαδρομή που έγινε γνωστή μέσω iBGP. Επίσης, προτιμάται η διαδρομή που έγινε γνωστή μέσω iBGP από διαδρομή που έγινε γνωστή μέσω iBGP και περιέχει το δικό μας cluster-id.

eBGP > iBGP > iBGP (same cluster-id)

Έτσι, δεχόμαστε μεν τη διαδρομή, αλλά την υποβαθμίζουμε και εξαφανίζεται ο κίνδυνος εμφάνισης βρόχων δρομολόγησης. Τα πλεονεκτήματα γίνονται αντιληπτά με τη μελέτη του παραπάνω παραδείγματος.

Επομένως, με χρήση Standby Route Reflection είναι πιθανό να επιτύχουμε μεγαλύτερη αξιοπιστία από την αρχική υλοποίηση του BGP-4 με iBGP τοπολογία πλήρους πλέγματος.

5.8 Πλεονεκτήματα-μειονεκτήματα

Το Standby Route Reflection , αποτελεί μια προαιρετική επέκταση του πρωτοκόλλου BGP-4 και σχεδιάστηκε για να δώσει λύση σε προβλήματα που ανέκυψαν από τη χρήση του BGP-4 ως πρωτοκόλλου δρομολόγησης. Το Standby RR παρουσιάζει τα εξής πλεονεκτήματα κυρίως σε σχέση με το Redundant Route Reflection που αποτελεί εφάμιλλη πρόταση:

- Ο λόγος για τον οποίο προτείνουμε την υλοποίηση ενός τέτοιου πρωτοκόλλου είναι η ελαχιστοποίηση των διαδρομών που πρέπει να γνωρίζει ένας RR Client, με αποτέλεσμα την ελάχιστη δυνατή απαίτηση σε μνήμη και υπολογιστική ισχύ από πλευράς RR Client.
- Σε περίπτωση που η IGP τοπολογία είναι σταθερή, οι clients μαθαίνουν μέσω του BGP τον ίδιο αριθμό διαδρομών με το απλό route reflection, ενώ διατηρούν n*r sessions εντός ενός cluster, όπου n ο αριθμός των redundant route reflectors και r τα session που θα υπήρχαν στην περίπτωση του απλού route reflection.

- Γνώση πάντα της καλύτερης διαδρομής από πλευράς RR Client, αφού πρώτα έχουν ακολουθηθεί σωστές τακτικές σχεδιασμού του AS (κεφ.4.2) σε συνδυασμό με την πρόταση για Standby RR.
- Διατήρηση βέλτιστης λειτουργίας ύστερα από κατάρρευση ή αλλαγή μεγάλου μέρους της IGP τοπολογίας σε σχέση με οποιαδήποτε εναλλακτική λύση.
- Οι RR Client δεν χρειάζεται να έχουν γνώση της επέκτασης του πρωτοκόλλου.
- Μείωση των μηνυμάτων BGP μεταξύ Standby RR και RR Client

Δυστυχώς, όπως και κάθε άλλη εναλλακτική πρόταση (κεφ.4) το Standby Route Reflection παρουσιάζει και ορισμένα μειονεκτήματα :

- Επιπρόσθετο πρωτόκολλο, ανταλλαγή επιπλέον μηνυμάτων BGP μεταξύ των Standby RR
- Επιπλέον απαίτηση σε υπολογιστική ισχύς από του Standby Route Reflectors.
- Επιπλέον απαίτηση σε μνήμη από τους Standby Route Reflectors..
- Ο RR Client μπορεί προσωρινά να μην έχει γνώση κάποιων διαδρομών, σε περίπτωση δυσλειτουργίας, για χρονικό διάστημα που μεσολαβεί μέχρις ότου να αναλάβει ο εναλλακτικός δρομολογητής Standby Route Reflector.

ΚΕΦΑΛΑΙΟ 6

Υλοποίηση του πρωτοκόλλου BGP σε java (BGPConsole)

Κατά τον θεωρητικό σχεδιασμό του της επέκτασης του πρωτοκόλλου BGP-4, Standby Route Reflection, προέκυψε η ανάγκη για υλοποίηση της επέκτασης. Έτσι, δημιουργήθηκε εφαρμογή που υλοποιεί τις βασικές λειτουργίες του πρωτοκόλλου BGP-4, καθώς και του Route Reflection. Η εφαρμογή έχει γραφτεί σε Java, ώστε να είναι ανεξάρτητη από το λειτουργικό σύστημα που εκτελείται.

6.1 Στόχος του προγράμματος

Στόχος της εφαρμογής είναι η υλοποίηση της επέκτασης του BGP-4, Standby route Reflection. Για να φτάσουμε στο σημείο αυτό, πρέπει να υλοποιηθεί μέρος του BGP-4, του τυπικού Route Reflection και ,εν συνεχεία, το Standby Route Reflection. Η εφαρμογή είναι σχεδιασμένη για ερευνητικούς σκοπούς και συγκεκριμένα για την εργαστηριακή αξιολόγηση της επέκτασης του BGP-4, Standby Route Reflection.

Η εφαρμογή BGPConsole , διαθέτει ισχυρές μεθόδους λεπτομερούς παρακολούθησης των μηνυμάτων BGP που ανταλλάσσει ο H/Y με δρομολογητές. Το BGPConsole, δημιουργεί συνδέσεις BGP-4 με τυπικούς δρομολογητές που εκτελούν BGP, καθώς και με H/Y που εκτελούν την εφαρμογή BGP-4. Το BGPConsole, υποστηρίζει την ανταλλαγή πληροφοριών IPv4. Λόγω του εργαστηριακού σκοπού του προγράμματος, δεν κρίθηκε απαραίτητο η εφαρμογή να προωθεί πακέτα. Έτσι, οι διαδρομές IPv4 που τοποθετεί η εφαρμογή στον τοπικό BGP πίνακα δρομολόγησης, δεν τοποθετούνται στον αντίστοιχο πίνακα δρομολόγησης του λειτουργικού συστήματος.

6.2 Υλοποίηση σε java

Το BGPConsole αποτελεί μια ομάδα κλάσεων Java και συγκεκριμένα, βασίζεται στην ομάδα κλάσεων Java-BGP, που γράφτηκε από τον Paul Jimenez,1999 (<http://www.place.org/~pj/software>). Το java-BGP είχε τη δυνατότητα δημιουργίας και διατήρησης συνδέσεων BGP-4. Το BGPConsole (εικ.6.1), επεκτείνει τις διαθέσιμες δυνατότητες ώστε να μπορούν να εξαχθούν συμπεράσματα για το Standby Route Reflection. Αναλυτικότερα, Το BGPConsole έχει τις παρακάτω δυνατότητες :

- Δημιουργία πολλών ταυτόχρονων BGP-4 συνδέσεων και διατήρηση αυτών μέσω keepalive μηνυμάτων.
- Αποδοχή μηνυμάτων τύπου Update και τοποθέτηση της έγκυρης πληροφορίας δρομολόγησης στον εισερχόμενο πίνακα BGP (adj-rib-in).

- Περιοδική και οδηγημένη από γεγονότα διαδικασία επιλογή καλύτερης διαδρομής και τοποθέτησης αυτής στον πίνακα δρομολόγησης BGP (loc-rib).
- Δημιουργία τοπικών BGP διαδρομών και διάδοση αυτών.
- Τοποθέτηση των προς διάδοση διαδρομών στον αντίστοιχο εξερχόμενο πίνακα BGP (adj-rib-out).
- Αποστολή μηνυμάτων Update.
- Παρακολούθηση συνδέσεων BGP.
- Δυνατότητες debugging του BGP, των διαφόρων γεγονότων των πακέτων που ανταλλάσσονται, καθώς και της διαδικασίας επιλογής καλύτερης διαδρομής.
- Πλήρης υλοποίηση Route Reflection, βάσει τυπικών προδιαγραφών (Bates, T & Chen , E & Chandra, R.(2006)).
- Πλήρης υλοποίηση Standby Route Reflection, όπως παρουσιάζεται στο παρόν.
- Εισαγωγή θεωρητικών IGP metrics από αρχείο ή σε πραγματικό χρόνο.
- Εισαγωγή παραμετροποίησης από αρχείο ή σε πραγματικό χρόνο.
- Δυνατότητα μεταβολής BGP timers.

```

C:\WINDOWS\system32\cmd.exe - java BGPConsole
Microsoft Windows XP [Έκδοση 5.1.26001
(C) Πνευματικά δικαιώματα 1985-2001 Microsoft Corporation
C:\Documents and Settings\Administrator>bin
C:\Documents and Settings\Administrator>cd\Program Files\Java\jdk1.6.0_06\bin
C:\Program Files\Java\jdk1.6.0_06\bin>java BGPConsole
Now listening on port 179...
BGP Console is written by Alexiou Ioannis ,University Of Macedonia, 2008
mail to : galexioy@gmail.com
and is based on BGPListener set of java classes which was originally written
by Paul Jimenez,1999
RFC support :
- rfc4271: BGP-4
- rfc4456: BGP Route Reflection <modified>
- rfc2918: Capabilities Advertisement with BGP-4
- rfc3345: Border Gateway Protocol (BGP) Persistent Route Oscillation Condition
-----
>

```

Εικόνα 6.1 : Εκτέλεση του προγράμματος BGPConsole

Από τα παραπάνω, είναι εμφανές πως έχει δοθεί ιδιαίτερη προσοχή στην υποστήριξη βασικών λειτουργιών του πρωτοκόλλου BGP-4, ώστε να γίνει δυνατή η εξαγωγή συμπερασμάτων με ασφάλεια. Μελλοντικά, θα ήταν δυνατή η μετατροπή του BGPConsole, ώστε να προωθεί πακέτα, να υποστηρίζει aggregation, φίλτρα, IPv6 και MPLS/VPNv4 διαδρομές.

6.2.1 Υποστηριζόμενα πρωτόκολλα

Η ομάδα κλάσεων BGPConsole, παρέχει μερική υποστήριξη για τα παρακάτω πρωτόκολλα :

- BGP-4 (Rekhter, Y & Li, T & Hares, S.(2003)). Στην παρούσα υλοποίηση δεν υποστηρίζεται aggregation.
- BGP Route Reflection (Bates, T & Chen , E & Chandra, R.(2006)). Πλήρης υλοποίηση.
- Capabilities Advertisement with BGP-4 (Chandra, R & Scudder, J.(2002)). Πλήρης υλοποίηση.
- Standby Route Reflection. Πλήρης υλοποίηση.

6.2.2 Σετ κλάσεων

Η ομάδα κλάσεων περιλαμβάνει τις παρακάτω κλάσεις :

- BGPConsole.class : Υπεύθυνη για τις διαδικασίες διασύνδεσης με το χρήστη και τα αρχεία παραμετροποίησης.
- BGPEngine.class : Δημιουργεί τον τοπικό πίνακα δρομολόγησης BGP. Επίσης, υλοποιεί την μηχανή επιλογής καλύτερης διαδρομής και τις αντίστοιχες διαδικασίες για την επιλογή των διαδρομών που πρέπει να διαδοθούν σε γειτονικούς δρομολογητές.
- BGPHeader.class : Δημιουργεί, ως αντικείμενο κεφαλίδα BGP ή αποκωδικοποιεί εισερχόμενη κεφαλίδα BGP. Η χρήση του αντικειμένου είναι υποχρεωτική πριν ή κατά τη λήψη ή αποστολή μηνύματος BGP.
- BGPListener.class : Νήμα που δημιουργείται ανά σύνδεση BGP και υλοποιεί το FSM , τους timers του BGP και είναι υπεύθυνη για τον χειρισμό γεγονότων που αφορούν το BGP.
- BGPNotificationPacket.class : Δημιουργεί ως αντικείμενο πακέτο τύπου Notification ή επεξεργάζεται εισερχόμενο πακέτο τύπου Notification και δημιουργεί event ανάλογα με τον κωδικό σφάλματος BGP (μερική υλοποίηση των υποστηριζόμενων κωδικών).
- BGPOpenPacket.class : Δημιουργεί ως αντικείμενο πακέτο τύπου OPEN ή αναλύει εισερχόμενο πακέτο τύπου OPEN.

- `BGPOpenParameter.class` : Δημιουργεί ως αντικείμενο τις παραμέτρους που περιλαμβάνονται σε μήνυμα OPEN ή αποκωδικοποιεί εισερχόμενες παραμέτρους. Η χρήση του αντικειμένου είναι υποχρεωτική πριν τη χρήση του αντικειμένου `BGPOpenPacket`.
- `BGPPathAttribute.class` : αποκωδικοποιεί τα χαρακτηριστικά διαδρομών BGP που περιέχονται σε αντικείμενο τύπου `BGPUpdatePacket` ή κωδικοποιεί τα εξερχόμενα χαρακτηριστικά διαδρομών ώστε να περιληφθούν σε πακέτο BGP Update. Η χρήση του αντικειμένου είναι υποχρεωτική κατά την αποστολή ή λήψη μηνύματος update.
- `BGPStatus.class` : Δημιουργείται ένα αντικείμενο αυτού του τύπου ύστερα από επιτυχή σύνδεση BGP (`BGPListener`) και διατηρούνται στατιστικά στοιχεία εισερχόμενος πίνακας BGP, ο εξερχόμενος πίνακας BGP για κάθε γειτονικό δρομολογητή.
- `BGPtype7.class` : Κωδικοποιεί/αποκωδικοποιεί τα πακέτα BGP τύπου 0x07 που ανταποκρίνονται στο Standby Route Reflection.
- `BGPUpdatePacket.class` : Δημιουργεί αντικείμενο για την κωδικοποίηση/αποκωδικοποίηση πακέτων τύπου Update. Γίνεται υποχρεωτική χρήση του υπεύθυνου αντικειμένου για τον χειρισμό των χαρακτηριστικών διαδρομής.
- `DataPacket.java` : Υλοποιεί γενικό αντικείμενο αποστολής/λήψης πακέτων TCP/IP.
- `FileHandler.class` : Κλάσης που διαβάζει αρχεία παραμετροποίησης ή τοπολογίας IGP και περνά τα δεδομένα στο `BGPConsole` για περαιτέρω επεξεργασία.
- `Hex.class` : Κλάση υπεύθυνη για μετατροπή δεδομένων Hex.
- `IGPEngine.class` : Κλάση που διατηρεί πίνακα δρομολόγησης IGP. Πιθανές μεταβολές δημιουργούν event που προκαλούν εκτέλεση κώδικα στο `BGPListener`.
- `InetNetwork.class` : Κλάση για μετατροπή διευθύνσεων IPv4 σε τύπους δεδομένων.
- `Util.class` : Κλάση που περιλαμβάνει ρουτίνες υποστήριξης και είναι απαραίτητη από το σύνολο της ομάδας κλάσεων `BGPConsole`.

6.3 Εγκατάσταση προγράμματος

Η εγκατάσταση του προγράμματος `BGPConsole` είναι αρκετά απλή. Ανεξαρτήτως λειτουργικού συστήματος, εγκαθιστούμε τη Java στον Η/Υ. Εν συνεχεία τοποθετούμε τα αρχεία *.java που υπάρχουν στο CD Εγκατάστασης σε διαδρομή στο δίσκο που έχει τη δυνατότητα εύρεσης της JAVA (π.χ `PATH=.`). Κάνουμε `compile` τα αρχεία με την εντολή : **javac** και τέλος εκτελούμε το πρόγραμμα με την εντολή : **java BGPConsole**. Αν επιθυμούμε την εκτέλεση αρχείου παραμετροποίησης, εκτελούμε με : **java BGPConsole -conf**. Η διαδικασία `compile` έχει δοκιμαστεί με Java έκδοσης 1.6.0 και εξάγει τις κλάσεις επιτυχώς και χωρίς σφάλματα.

6.4 Μέθοδος παραμετροποίησης

Η εφαρμογή παραμετροποιείται σύμφωνα με τις ανάγκες της κάθε πειραματικής τοπολογίας. Οι εντολές που παρουσιάζονται παρακάτω εισάγονται με όμοιο τρόπο σε πραγματικό χρόνο ή μέσω του αρχείου παραμετροποίησης. Το αρχείο παραμετροποίησης είναι απλό αρχείο κειμένου και πρέπει να έχει το όνομα αρχείο `bgp.conf`, ώστε να αναγνωστεί επιτυχώς από το πρόγραμμα. Για εκτέλεση αρχείου παραμετροποίησης, κατά την έναρξη της εφαρμογής, εισάγουμε το `argument -conf`.

Οι εντολές που αφορούν τη συμπεριφορά του τοπικού δρομολογητή (local) πρέπει να εισάγονται, πριν εισαχθούν άλλες εντολές. Επίσης, οι εντολές που επιχειρούν γειτονικές συνδέσεις (neighbor) πρέπει να εισάγονται τελευταίες.

Παρακάτω, παρουσιάζονται οι διαθέσιμες εντολές(πιν.6.4.1). Δεξιά της εντολής, ορίζεται η παράμετρος που δέχεται, καθώς και ο τύπος δεδομένων.

ΕΝΤΟΛΗ	ΠΕΡΙΓΡΑΦΗ
local AS <integer>	Ορισμός τοπικού αριθμού AS 1-65535.
local bind <a.b.c.d>	Διεύθυνση IPv4 που θα χρησιμοποιήσει ο δρομολογητής ως BGP-ID.
local holdtime <time>	Ορισμός του τοπικού Hold timer (sec). Χρησιμοποιείται στα εξερχόμενα μηνύματα OPEN.
local send-opt-cap	Γίνεται χρήση της προαιρετικής ικανότητας 0x100 στα μηνύματα OPEN ώστε να υποστηριχθεί σύνδεση Standby Route Reflector.
local redundant-RR <master/slave>	Ορίζουμε αν ο τοπικός δρομολογητής θα είναι master/slave σε περίπτωση σύνδεσης Standby RR.
local redundant-RR holdtime <time>	Ορίζουμε το νέο επιθυμητό Hold Timer σε περίπτωση σύνδεσης Standby RR .
local redundant-RR	Ορίζουμε ότι ο δρομολογητής πρέπει να εκτελέσει την επέκταση πρωτοκόλλου Standby RR.
local	Εμφανίζει την τρέχουσα κατάσταση των μεταβλητών τύπου local BGP παραμέτρων
announce <a.b.c.d/e>	Χρησιμοποιείται για να διαφημίσει ο τοπικός δρομολογητής κάποιο δίκτυο.
neighbor <a.b.c.d>	Ορισμός γειτονικών δρομολογητών. Απαραίτητη εντολή για δημιουργία σύνδεσης BGP.
status	Εμφανίζει τις διαθέσιμες συνδέσεις BGP με τα αντίστοιχα στατιστικά (Uptime, updates κτλ)

ΕΝΤΟΛΗ	ΠΕΡΙΓΡΑΦΗ
status <connection>	Εμφανίζει τα στατιστικά μιας συγκεκριμένης σύνδεσης BGP.
show adj-rib-in	Εμφανίζει τα περιεχόμενα του εισερχόμενου πίνακα BGP (adj-rib-in) για κάθε γειτονική σύνδεση BGP.
show adj-rib-out	Εμφανίζει τα περιεχόμενα του εξερχόμενου πίνακα BGP (adj-rib-out) για κάθε γειτονική σύνδεση BGP.
show loc-rib	Εμφανίζει τα περιεχόμενα του πίνακα δρομολόγησης BGP (loc-rib). Δεν περιλαμβάνονται τα χαρακτηριστικά διαδρομής
show loc-rib detail	Εμφανίζει τα περιεχόμενα του πίνακα δρομολόγησης BGP (loc-rib) και περιλαμβάνει τα χαρακτηριστικά διαδρομής για κάθε διαδρομή.
show igp-rib	Εμφανίζει τον εικονικό πίνακα δρομολόγησης IGP.
show route-reflector	Εμφανίζει την κατάσταση του τοπικού δρομολογητή σχετικά με το Standby Route Reflection.
route-reflect <a.b.c.d>	Χρησιμοποιείται μετά από εντολή neighbor και ορίζει πως ο γειτονικός δρομολογητής BGP είναι RR Client.
redundant-RR <a.b.c.d>	Χρησιμοποιείται μετά από εντολή neighbor και ορίζει πως πρέπει να επιχειρηθεί σύνδεση Standby RR με τον γειτονικό δρομολογητή BGP.
import igp	Χρησιμοποιείται για να εισάγουμε εικονική IGP τοπολογία από το αρχείο igp.rib.
quit	Έξοδος από το πρόγραμμα.
help or ?	Εμφάνιση βοήθειας.
debug <type> on off	Ενεργοποιεί/ απενεργοποιεί το debugging για συγκεκριμένα σημεία του BGP που ορίζονται μέσω της παραμέτρου type. Αν δεν δώσουμε on off, τότε εμφανίζεται η κατάσταση του debugging.

Πίνακας 6.4.1 : Εντολές προγράμματος BGPConsole.

Παρακάτω (πιν.6.4.2), παρουσιάζονται οι παράμετροι <type> που μπορεί να πάρει η εντολή debug, καθώς και το αποτέλεσμα της κάθε μιάς.

ΤΙΜΕΣ <type>	ΑΠΟΤΕΛΕΣΜΑ DEBUGGING
connect	Προσπάθεια σύνδεσης / ConnectRetryTimer.
Redundant RR	Χαρακτηριστικά / μηνύματα πρωτοκόλλου Standby Route Reflection.
BGP Engine	Διαδικασία επιλογής διαδρομής / τοποθέτηση προς διάδοση διαδρομών στο adj-rfb-out .
header	Αποκωδικοποίηση κεφαλίδας BGP
listener	BGP FSM / μηνύματα BGP / keepalive / holdtimer.
open	Ανάλυση μηνυμάτων Open.
attributes	Αποκωδικοποίηση χαρακτηριστικών διαδρομής.
status	Λήψη / αποστολή μηνυμάτων BGP.
updates	Ανάλυση μηνυμάτων Update.
IGP	Παρακολούθηση εικονικής IGP Μηχανής.
?	Εμφανίζει αυτές τις επιλογές.

Πίνακας 6.4.2 : Παράμετροι της εντολής debug <type>.

Παρακάτω παρουσιάζονται παραδειγματικά τα περιεχόμενα ενός αρχείου bgp.conf.

FILE : bgp.conf

```

local AS 65000
local bind 192.168.1.2
local holdtime 90
local send-opt-cap
local redundant-RR holdtime 30
local redundant-RR master
debug redundant_RR on
debug BGP Engine on
debug listener on
debug open on
debug IGP on
import igp
redundant-RR 192.168.1.5
neighbor 192.168.1.5
route-reflect 192.168.1.254

```

```
neighbor 192.168.1.254
route-reflect 192.168.1.100
neighbor 192.168.1.100
EOF
```

6.5 Μέθοδος εισαγωγής IGP τοπολογίας

Το BGPConsole λειτουργεί ανεξαρτήτως πλατφόρμας λογισμικού και για το λόγο αυτό δεν συνδέεται με τον αντίστοιχο πίνακα δρομολόγησης του λειτουργικού.

Έτσι, υπάρχει η ανάγκη για εικονικό IGP πίνακα δρομολόγησης, ώστε να διεξαχθούν πειραματικές ρυθμίσεις. Στην συγκεκριμένη περίπτωση, μας ενδιαφέρει το κόστος (metric) που απαιτείται για επικοινωνία με γειτονικούς δρομολογητές BGP.

Στο αρχείο `igr.rib` ορίζουμε τα metrics ανά γειτονικό BGP δρομολογητή και όχι τα metrics προς κάθε υποδίκτυο. Το `igr.rib` είναι αρχείο κειμένου και περιέχει σε κάθε σειρά την unicast IPv4 διεύθυνση γειτονικού δρομολογητή και το κόστος (metric) προς αυτή, χωρισμένη με κόμμα. Για παράδειγμα, παρακάτω παρουσιάζεται ένα αρχείο `igr.rib`

FILE : igr.rib

```
192.168.1.5,10
192.168.1.100,30
192.168.1.254,30
EOF
```

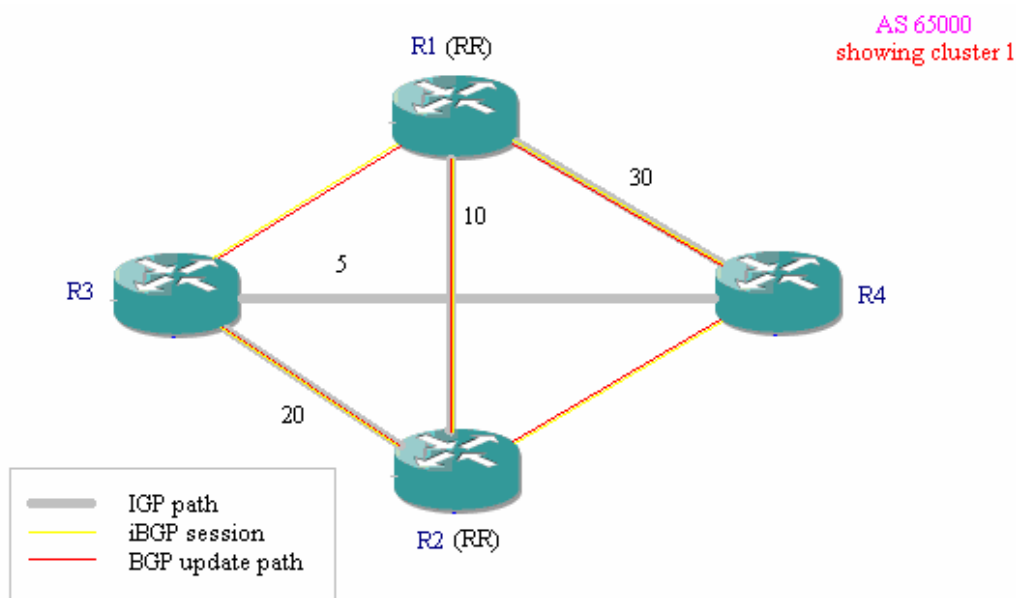
ΚΕΦΑΛΑΙΟ 7

Εργαστηριακές μετρήσεις

7.1 Πειράματα

7.1.1 Redundant Route Reflection

Στην παρακάτω τοπολογία παρουσιάζεται το AS65000 και συγκεκριμένα το cluster-1, το οποίο υλοποιείται με Redundant Route Reflector. Οι δρομολογητές R1,R2 είναι υλοποιημένοι σε java (BGP Console) και υποστηρίζουν πλήρως το BGP-4 και το τυπικό Route Reflection.

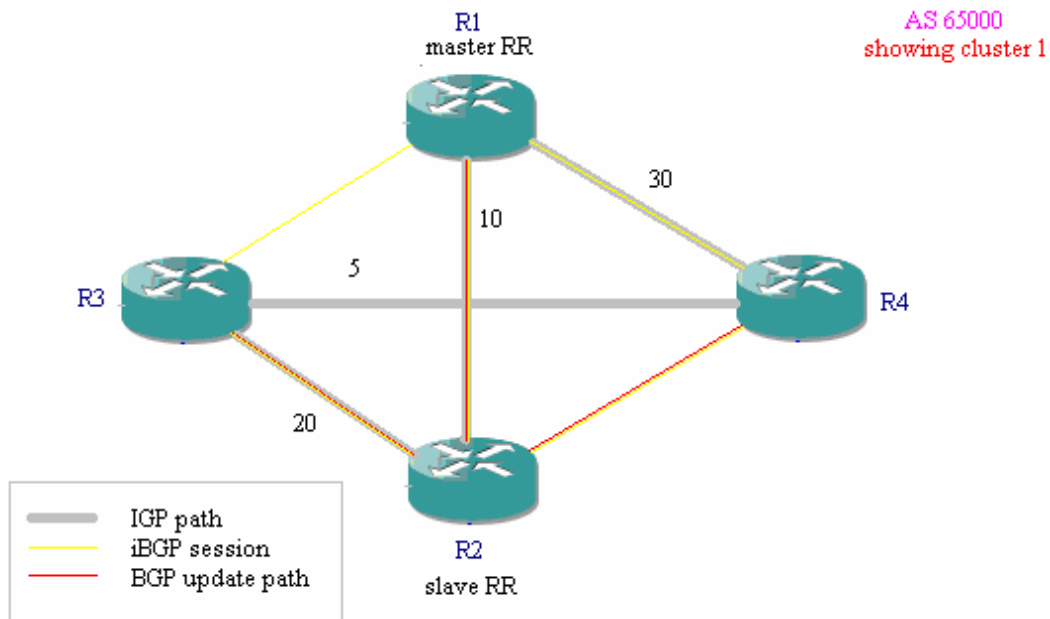


Εικόνα 7.1.1 : Πειραματική τοπολογία BGP με Redundant RR's

Στο πείραμα αυτό, οι δρομολογητές R3,R4 είναι route reflector client. Ο R4 διαφημίζει αυτοδημιούργητες διαδρομές, οι οποίες φτάνουν στους R1, R2, οι οποίοι με τη σειρά τους διαφημίζουν την καλύτερη διαδρομή στον R3.

7.1.2 Standby Route Reflection

Στην παρακάτω τοπολογία παρουσιάζεται το AS65000 και συγκεκριμένα το cluster-1 και υλοποιείται με Standby Route reflector. Οι δρομολογητές R1, R2 είναι υλοποιημένοι σε java (BGP Console) και υποστηρίζουν πλήρως το BGP-4 , το τυπικό Route Reflection καθώς και υλοποιούν την επέκταση Standby Route Reflection για βελτίωση του πρωτοκόλλου iBGP.



Εικόνα 7.1.2 : Πειραματική τοπολογία BGP με Standby RR's

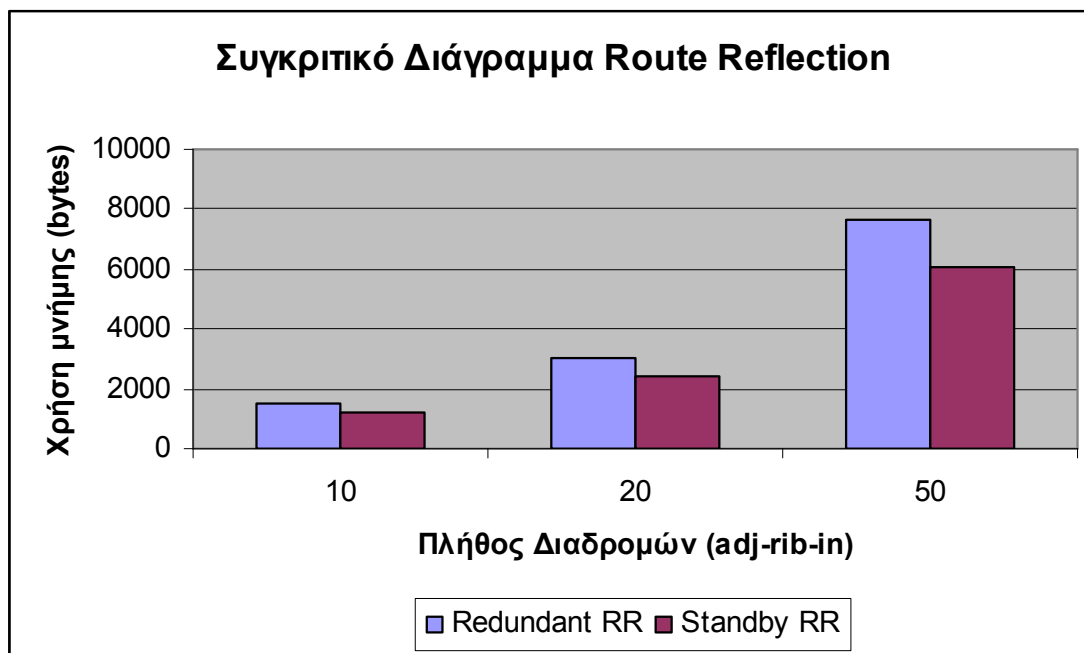
Στο πείραμα αυτό, οι δρομολογητές R3, R4 είναι RR Client. Ο R4 διαφημίζει αυτοδημιούργητες διαδρομές, οι οποίες φτάνουν στους R1, R2. Σε αντίθεση με το προηγούμενο πείραμα, μόνο ο R2 (λόγω μικρότερου IGP metric) θα εξυπηρετήσει τον R3, διαφημίζοντας σε αυτόν τις διαδρομές που έμαθε από τον R4.

7.2 Σύγκριση πειραματικών μετρήσεων

Λόγω της απλότητας του πειράματος, δηλαδή ένα exit point ανά διαδρομή, παρατηρούμε πως στα αποτελέσματα το ποσοστό μνήμης που κερδίζουμε είναι ανεξάρτητο από το πλήθος των διαδρομών που διαφημίζουμε. Επίσης, παρατηρούμε (πιν.7.2.1) και (εικ.7.2.2) πως το κέρδος είναι μικρότερο του θεωρητικά αναμενόμενου 50%, αφού οι δύο RR διαφημίζουν όμοιες διαδρομές στον RR Client. Αυτό οφείλεται στη συμπίεση δεδομένων που εφαρμόζει ο συγκεκριμένος δρομολογητής στο Adj-Rib-In καθώς και στην απλότητα του πειράματος (Οι διαδρομές έχουν παρόμοια χαρακτηριστικά [Path Attributes]).

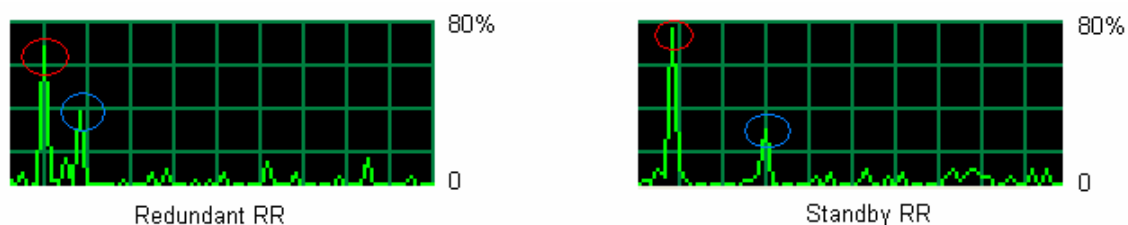
Ο R4 διαφημίζει διαδρομές – Ο R3 είναι Cisco 1600R				
	Redundant RR	Standby RR		
Πλήθος διαδρομών	Χρήση μνήμης Adj-Rib-In (bytes)	Χρήση μνήμης Adj-Rib-In (bytes)	Κέρδος μνήμης (bytes)	Ποσοστό κέρδους
10	1530	1210	320	20,9%
20	3060	2420	640	20,9%
50	7650	6050	1600	20,9%

Πίνακας 7.2.1 : Συγκριτικά αποτελέσματα πειράματος



Εικόνα 7.2.1 : Συγκριτικά αποτελέσματα πειράματος

Στα παρακάτω διαγράμματα (Εικ.7.2.2) εμφανίζεται η χρήση CPU στον Redundant RR και στον Standby RR, όταν αυτοί υλοποιούνται με BGPConsole σε όμοιο Η/Υ στα πειράματα 7.1.1 & 7.1.2 με 50 Routes.

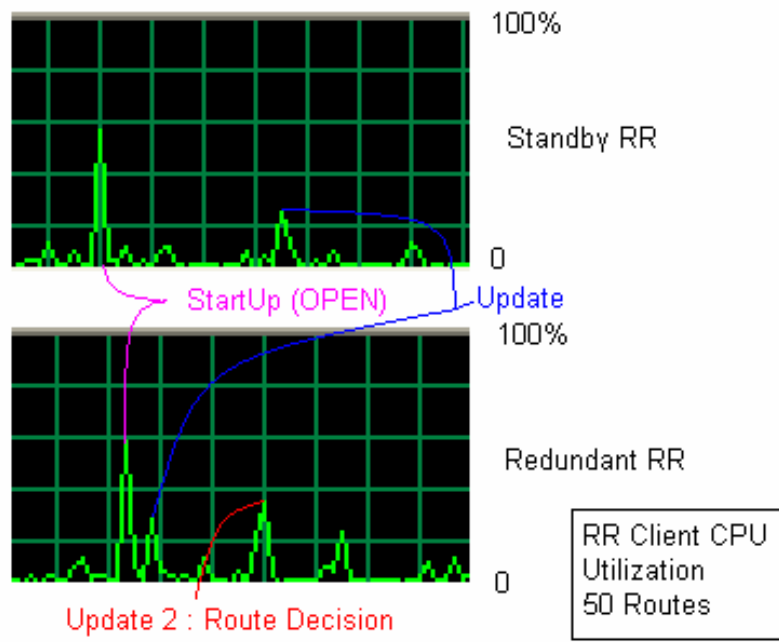


Εικόνα 7.2.2 : Συγκριτικό χρήσης CPU στον Reflector

Είναι εμφανές πως στην αρχή της διαδικασίας ο Standby RR καταναλώνει μεγαλύτερο ποσοστό της διαθέσιμης υπολογιστικής ισχύος, προφανώς, λόγω της επιπλέον πολυπλοκότητας του πρωτοκόλλου. Επίσης, εν συνεχεία, κατά την διαδικασία αποστολής update μηνυμάτων, ο Standby RR καταναλώνει μικρότερο ποσοστό της διαθέσιμης υπολογιστικής ισχύος, επειδή εξυπηρετεί λιγότερους RR Client σε σχέση με το Redundant RR.

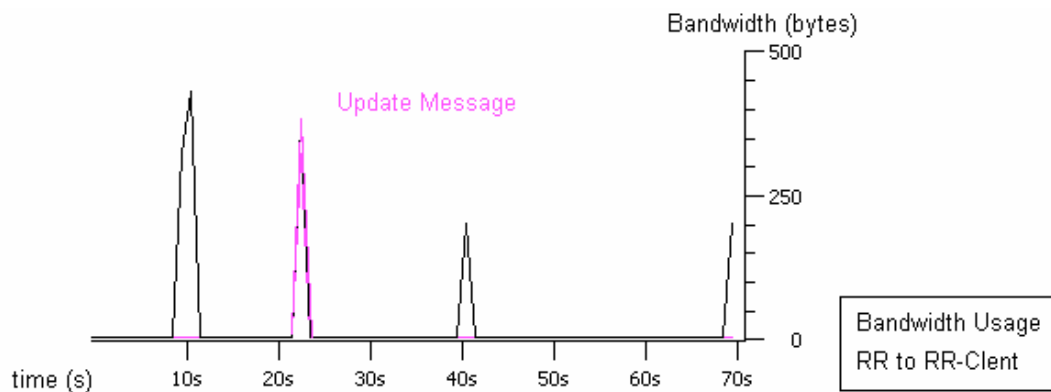
Από πλευράς Client, παρατηρούμε μείωση της χρήσης CPU (Εικ.7.2.3) στην περίπτωση χρήσης standby Route Reflector. Αυτό συμβαίνει, γιατί ο client μαθαίνει 50 διαδρομές για 50 δίκτυα προορισμού και δεν καλεί τον αλγόριθμο επιλογής διαδρομής. Επίσης, επεξεργάζεται λιγότερα μηνύματα τύπου Update. Στην περίπτωση Redundant Route Reflector, παρατηρούμε πως η χρήση CPU είναι ίδια μέχρι το σημείο που ο Client έχει δεχτεί το πρώτο update. Μόλις, όμως, ο Client δεχτεί το

δεύτερο update από τον άλλο reflector, καλεί την διαδικασία επιλογής διαδρομής για 50 διαδρομές και παρουσιάζεται χρήση της CPU κοντά στο 40%.

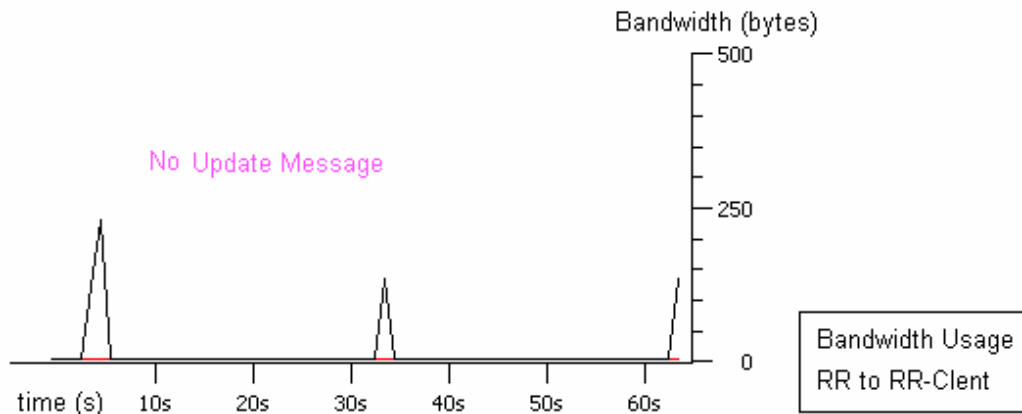


Εικόνα 7.2.2 : Συγκριτικό χρήσης CPU στον RR Client

Στα δύο πειράματα 7.1.1 & 7.1.2 παρατηρείται όμοια χρησιμοποίηση Bandwidth μεταξύ RR και Client , μόνο όταν ο RR είναι active (Εικ.7.2.4). Στην περίπτωση που ο Standby RR είναι σε κατάσταση Passive για τον συγκεκριμένο Client , τότε η χρησιμοποίηση του Bandwidth ελαττώνεται αισθητά, αφού δεν γίνεται αποστολή μηνυμάτων Update προς τον Client (Εικ.7.2.5).



Εικόνα 7.2.4 : Χρήση Bandwidth μεταξύ Active RR / Redundant RR και client



Εικόνα 7.2.5 : Χρήση Bandwidth μεταξύ Passive RR και client (απουσία Update)

Στον αντίποδα, στα δύο πειράματα 7.1.1 & 7.1.2 παρατηρείται μεγαλύτερη χρησιμοποίηση του Bandwidth μεταξύ δύο Standby RR, σε σχέση με τους Redundant RR(Εικ.7.2.6).

Στην περίπτωση των redundant RR, οι δύο δρομολογητές δεν έχουν κάποια ειδική σηματοδότηση μεταξύ τους. Επίσης, οι redundant RR δεν κάνουν αμοιβαία ανάκλαση διαδρομών με αποτέλεσμα να υστερούν και σε περίπτωση δυσλειτουργίας. Στο συγκεκριμένο σημείο, πρέπει να αναφέρουμε πως η μέτρηση που αφορά τους Redundant RR είναι ανεξάρτητη του αριθμού των διαδρομών, λόγω της μη ανάκλασης διαδρομών.

Στην περίπτωση των Standby RR, η χρησιμοποίηση της μεταξύ τους διασύνδεσης, έχει άμεση σχέση με τον αριθμό των διαδρομών που γνωρίζει κάθε Reflector. Αυτό συμβαίνει, γιατί κάθε standby RR ανακλά διαδρομές στο γείτονα Standby RR και προκύπτουν οφέλη σχετικά με την αξιοπιστία (Κεφ.5.7). Επίσης, αυξάνεται η κατανάλωση εύρους ζώνης λόγω των μηνυμάτων BGP τύπου 7 που διασχίζουν το Link και έχουν ως στόχο την ανταλλαγή των απαραίτητων δεδομένων για τη λειτουργία του πρωτοκόλλου.

Έπειτα και από τις εργαστηριακές προτάσεις που παρουσιάζονται, γίνεται εύκολα αντιληπτό πως η μεθοδολογία που παρουσιάζουμε έχει να προσφέρει πλεονεκτήματα που συνοψίζονται παρακάτω :

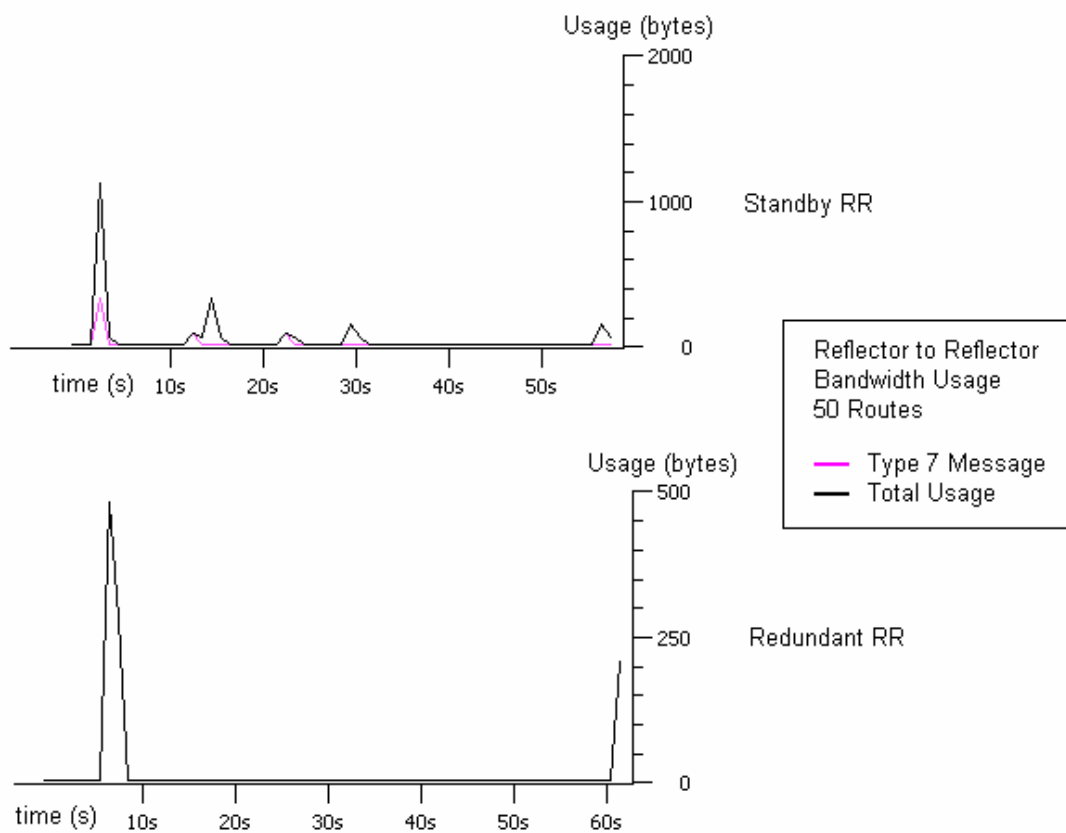
- 1) Μείωση απαίτησης σε μνήμη των clients που οφείλεται στην ελαχιστοποίηση της διαφήμισης διαδρομών μέσω Standby route reflectors
- 2) Μείωση της υπολογιστικής ισχύος από πλευράς route reflector client, αφού λαμβάνοντας μόνο μια διαδρομή ανά προορισμό, ο δρομολογητής δεν χρειάζεται να τρέξει τον αλγόριθμο επιλογής καλύτερης διαδρομής.
- 3) Προσαρμογή σε δυναμικές αλλαγές της IGP τοπολογίας.
- 4) Οι RR client δεν χρειάζεται να έχουν γνώση του πρωτοκόλλου.

5) Μικρότερη χρήση εύρους ζώνης μεταξύ Standby RR και RR Client, λόγω των λιγότερων διαδρομών που φτάνουν τους RR Clients.

Στον αντίποδα, για να πετύχουμε τα παραπάνω, επιβαρύνουμε :

1) Τους δρομολογητές που δρουν ως Standby Route Reflectors με περισσότερα δεδομένα και περισσότερους υπολογισμούς.

2) Το ίδιο το δίκτυο με την ροή περισσότερων δεδομένων μεταξύ των Standby RR.



Εικόνα 7.2.6 : Χρήση Bandwidth μεταξύ Reflectors (Συγκριτικό)

7.3 Πείραμα συμβατότητας

Παρακάτω, παρουσιάζεται η έξοδος του τερματικού (BGPConsole) σε μια προσπάθεια σύνδεσης με δρομολογητή (εικ.5.2.1) που δεν υποστηρίζει τη δυνατότητα standby RR (0x07).



Εικόνα 7.3 : iBGP σύνδεση δύο δρομολογητών

BGPConsole Output

```
C:\Program Files\Java\jdk1.6.0_06\bin>java BGPConsole -conf
Now listening on port 179...
BGP Console is written by Alexiou Ioannis ,University Of Macedonia, 2008
mail to : galexioy@gmail.com
and is based on BGPListener set of java classes which was originally written
by Paul Jimenez,1999
RFC support :
- rfc4271: BGP-4
- rfc4456: BGP Route Reflection (modified)
- rfc2918: Capabilities Advertisement with BGP-4
- rfc3345: Border Gateway Protocol (BGP) Persistent Route Oscillation Condition
-----
>EXECUTING : local AS 65000
[local] AS Number set to : 65000
>EXECUTING : local bind 192.168.1.2
>EXECUTING : local holdtime 90
[local] holdtimer set to : 90
>EXECUTING : local send-opt-cap
>EXECUTING : local redundant-RR holdtime 30
[local] RR holdtimer set to : 30
>EXECUTING : local redundant-RR master
>EXECUTING : debug redundant_RR on
>EXECUTING : debug BGPEngine on
>EXECUTING : debug listener on
>EXECUTING : debug open on
>EXECUTING : debug IGP on
>EXECUTING : import igp
[IGP] add route to : 192.168.1.254
[IGP] add route to : 192.168.1.5
[IGP] add route to : 192.168.1.100
>EXECUTING : redundant-RR 192.168.1.100
>EXECUTING : neighbor 192.168.1.100
>[TCP] Connected to : 192.168.1.100
[FSM] Passive
[FSM] Active
[BGPOpen] Sending OPEN: version: 4 AS: 65000 holdtime: 90 BGPID:192.168.1.2/32
```

```
[BGPOpen] Sent OPEN: version: 4 AS: 65000 holdtime: 90 BGPID: 192.168.1.2/32
[29/12/2008 1:38 'HH] [FSM] Open Sent From : 192.168.1.2/32
[BGPOpen]: Waiting for header
[FSM] Open Confirm. Received Notification --> Proceeding without Opt Capabilities...
clearing session thread...
Routes processed : 0
Routing table contents :
```

```
>[RRStartuptimer] Failed to peer with Redundant RR --> Acting as simple RR
Routes processed : 0
Routing table contents :
```

```
show adj-rib-in
Adjacency Information Base-in :
```

```
-----
Neighbor id : 192.168.1.100/32
-----
```

```
Total : 0 memory used : 0 bytes
>[TCP] Connected to : 192.168.1.100
[FSM] Passive
[FSM] Active
[BGPOpen] Sending OPEN: version: 4 AS: 65000 holdtime: 90 BGPID: 192.168.1.2/32
[BGPOpen] Sent OPEN: version: 4 AS: 65000 holdtime: 90 BGPID: 192.168.1.2/32
[29/12/2008 1:39 'HH] [FSM] Open Sent From : 192.168.1.2/32
[BGPOpen]: Waiting for header
[BGPOpen]: Waiting for body
[BGPOpen]: received body
Optional Capability : T 1
Optional Capability : T 128
Optional Capability : T 2
[BGPOpen] Optional parameters length : 16
[FSM] received OPEN: version: 4 AS: 65000 holdtime: 0 BGPID: 192.168.1.100/32
Session type : iBGP
[29/12/2008 1:39 'HH] Keepalive sent From : 192.168.1.100/32
[FSM] OpenConfirm
[29/12/2008 1:39 'HH] [FSM] Open Confirm : From : 192.168.1.100/32
[FSM] Established
quit[29/12/2008 1:39 'HH] Update 0 : From : 192.168.1.100/32
[29/12/2008 1:39 'HH] withdraw: 0 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] add: 10 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 1.1.1.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 10.10.10.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 9.9.9.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 8.8.8.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 7.7.7.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 6.6.6.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 5.5.5.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 4.4.4.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 3.3.3.0/24 From : 192.168.1.100/32
[29/12/2008 1:39 'HH] Announce: 2.2.2.0/24 From : 192.168.1.100/32
NLRI: [1.1.1.0/24, 10.10.10.0/24, 9.9.9.0/24, 8.8.8.0/24, 7.7.7.0/24, 6.6.6.0/24
, 5.5.5.0/24, 4.4.4.0/24, 3.3.3.0/24, 2.2.2.0/24]
BGP Engine : Route Decision : 5.5.5.0/24
Routes processed : 10
Routing table contents :
10.10.10.0/24 , 3.3.3.0/24 , 1.1.1.0/24 , 7.7.7.0/24 , 5.5.5.0/24 , 9.9.9.0/24 ,
2.2.2.0/24 , 6.6.6.0/24 , 4.4.4.0/24 , 8.8.8.0/24 ,
```

```
adj-rib-out calculation for neighbor : 192.168.1.100/32
loc-rib Pass : 1

adj-rib-out table contents :

>
```

BGPConsole configuration file : bgp.conf

```
local AS 65000
local bind 192.168.1.2
local holdtime 90
local send-opt-cap
local redundant-RR holdtime 30
local redundant-RR master
debug redundant_RR on
debug listener on
debug open on
import igp
redundant-RR 192.168.1.100
neighbor 192.168.1.100
```

BGPConsole configuration file : igp.rib

```
192.168.1.5,10
192.168.1.100,30
192.168.1.254,30
```

Cisco 1601R configuration

```
Cisco#sh run
Building configuration...

Current configuration:
!
version 12.0
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname Cisco
!
enable password cisco
!
ip subnet-zero
!
!
!
interface Ethernet0
ip address 192.168.1.100 255.255.255.0
no ip directed-broadcast
!
interface Serial0
no ip address
no ip directed-broadcast
shutdown
!
interface BRI0
no ip address
no ip directed-broadcast
```

```
shutdown
!  
router bgp 65000  
no synchronization  
bgp router-id 192.168.1.254  
bgp cluster-id 3305177345  
timers bgp 30 100  
neighbor 192.168.1.2 remote-as 65000  
neighbor 192.168.1.5 remote-as 65000  
!  
ip classless  
!  
!  
line con 0  
transport input none  
line vty 0 4  
password cisco  
login  
!  
end  
  
Cisco#
```

ΚΕΦΑΛΑΙΟ 8

Συμπεράσματα

Σκοπός της παρούσας εργασίας είναι η βελτίωση του πρωτοκόλλου BGP-4. Στο κείμενο αυτό, γίνεται εκτενής αναφορά στο πρωτόκολλο BGP-4 και τον τρόπο λειτουργίας του. Εν συνεχεία, γίνεται ανάλυση των προβλημάτων και των μειονεκτικών σημείων του BGP-4. Βάσει της ανάλυσης αυτής παρουσιάζεται η λύση iBGP Route Reflection που επέτρεψε στο BGP-4 να εξυπηρετήσει επιτυχώς το μεγαλύτερο δίκτυο του κόσμου (Internet). Η λύση αυτή, δημιούργησε και νέα προβλήματα, τα οποία επεξεργάστηκε η επιστημονική κοινότητα και έδωσε λύσεις και προτάσεις για την βελτίωση του iBGP route Reflection.

Παρόλα αυτά, ελάχιστες προτάσεις, παρουσιάζουν αξιόλογο έργο σχετικά με την ταυτόχρονη μείωση των iBGP συνδέσεων και την διατήρηση της αξιοπιστίας του δικτύου σε ικανοποιητικό επίπεδο.

Λαμβάνοντας υπόψη τα παραπάνω αποφασίστηκε ο σχεδιασμός της επέκτασης του πρωτοκόλλου BGP-4, Standby Route Reflection. Η πρόταση που παρουσιάζεται στο παρόν έγγραφο :

- Παρέχει τα πλεονεκτήματα του Redundant iBGP Route Reflection
- Επιπρόσθετα αυξάνει την αξιοπιστία του δικτύου
- Παρουσιάζει δυναμική προσαρμογή σε πιθανές αλλαγές της IGP τοπολογίας.
- Έχει μειωμένες απαιτήσεις σε μνήμη ,υπολογιστική ισχύ και χρήση εύρους ζώνης από τους RR Client

Ακολούθως, παρουσιάζεται η εφαρμογή Java BGPConsole, η οποία υλοποιεί δρομολογητή BGP σε H/Y. Σκοπός της εφαρμογής είναι η υλοποίηση του BGP-4, καθώς και της επέκτασης Standby Route Reflection, ώστε να γίνει η πειραματική επαλήθευση των παραπάνω.

Εν κατακλείδι, παρουσιάζονται πειραματικές μετρήσεις και επαληθεύουν τα πλεονεκτήματα – μειονεκτήματα του Standby Route Reflection.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Rekhter, Y & Li, T & Hares, S.(2003) *A Border Gateway Protocol 4 (BGP-4)*. Ανάκτηση 29/11/2008 από World Wide Web: http://www.rfc-editor.org/cgi-bin/rfcdoctype.pl?loc=RFC&letsgo=4271&type=http&file_format=txt
- Bates, T & Chen , E & Chandra, R.(2006) *BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)*. Ανάκτηση 29/11/2008 από World Wide Web: http://www.rfc-editor.org/cgi-bin/rfcdoctype.pl?loc=RFC&letsgo=4456&type=http&file_format=txt
- Traina, P & McPherson, D & Scudder, J. (2001) *Autonomous System Confederations for BGP*. Ανάκτηση 29/11/2008 από World Wide Web: http://www.rfc-editor.org/cgi-bin/rfcdoctype.pl?loc=RFC&letsgo=3065&type=http&file_format=txt
- Chandra, R & Scudder, J.(2002) *Capabilities Advertisement with BGP-4*. Ανάκτηση 29/11/2008 από World Wide Web: http://www.rfc-editor.org/cgi-bin/rfcdoctype.pl?loc=RFC&letsgo=3392&type=http&file_format=txt
- Xiao, L & Wang, J & Nahrstedt, K. (2003). Reliability-aware IBGP route reflection topology design. In: *Proceedings 11th IEEE International Conference on Network Protocols*, Atlanta,Georgia, USA
- Klockar, T & Carr-Motyckova, L. (2004). Preventing oscillations in route reflector-based I-BGP. In: *International Conference on Computer Communications and Networks, 2004. ICCCN 2004. Proceedings. 13th* , Chicago, Illinois, USA
- Basu, A & Chih-Hao Luke,O & Rasala A & Shepherd ,B & Wilfong, G (2002): Route oscillations in I-BGP with route reflection. In: *Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, Pittsburgh, Pennsylvania, USA
- McPherson, D & Gill, V & Walton, D & Renata, A. (2002) *Border Gateway Protocol (BGP) Persistent Route Oscillation Condition*. Ανάκτηση 29/11/2008 από World Wide Web: <http://www.ietf.org/rfc/rfc3345.txt>
- Walton, D & Renata, A & Scudder, J. (2002) *BGP Persistent Route Oscillation Solution*. Ανάκτηση 29/11/2008 από World Wide Web: <http://www.potaroo.net/ietf/all-ids/draft-walton-bgp-route-oscillation-stop-00.txt>
- Van den Schrieck, V & Francois, P & Tandel, S & Bonaventure, O. (2006): Let BGP Speakers Configure their iBGP Sessions on Their Own. In: *WIRED 2006 workshop* , Atlanta,Georgia, USA
- Buod, M & Meulle, M & Uhlig, S. (2006) *Checking for optimal egress points in iBGP routing*. Ανάκτηση 29/11/2008 από World Wide Web: <http://www.nas.ewi.tudelft.nl/people/Steve/papers/drcn07.pdf>

- Xiao, L & Wang, J & Nahrstedt, K. (2003). Optimizing IBGP route reflection network. In: *ICC '03. IEEE International Conference on Communications*, Atlanta, Georgia, USA
- Vutukuru, M & Valiant, P & Kopparty, S & Balakrishnan, H. (2006): How to Construct a Correct and Scalable iBGP Configuration. In: *INFOCOM 2006, 25th IEEE International Conference on Computer Communications*, Barcelona, Spain
- Griffin, T & Wilfong, G. (2002) “On the correctness of IBGP configuration”, *Communications of the ACM*,32(4),17-29
- Musunuri, R & Cobb, J (2003) *Stable iBGP through Selective Path Dissemination*, ACTA Press
- Van den Schrieck, V & Bonaventure, O. (2005) Comparison of IBGP topologies. In: *Proceedings of the 2005 ACM conference on Emerging network experiment and technology*, Toulouse, France
- Bonaventure, O & Uhlig, S & Quoitin, B. (2004) : *The case for more versatile BGP Route Reflectors*. Ανάρτηση 29/11/2008 από World Wide Web: <http://tools.ietf.org/html/draft-bonaventure-bgp-route-reflectors-00>
- Musunuri, R & Cobb, J. (2004): A Complete Solution for iBGP Stability. In: *Conference on Communications, 2004 IEEE International* , Paris, France.
- Musunuri, R & Cobb, J. (2004): Enforcing iBGP Convergence. In: *Conference on Communications, 2004 IEEE International* , Paris, France.
- Feng, Z & Xicheng, L & Peidong, Z & Jinjing Z. (2006) “BGPSep_S: An Algorithm for Constructing iBGP Configurations with complete visibility” , *LNCS Volume 4308/2006*, Berlin, Springer, 379-384
- Musunuri, R & Cobb, J. (2005) “Comprehensive Solution for Anomaly-Free BGP” , *LNCS Volume 3751/2005*, Berlin, Springer,130-141
- Rohit, D. (1999) “A Comparison of scaling Techniques for BGP”, *Communications of the ACM*,29(3),44-46
- Rawat, A & Shayman, M. (2006) “Preventing persistent oscillations and loops in IBGP configuration with route reflection”, *Communications of the ACM*,50(18),3642-3665.
- Buod, M & Uhlig, S & Muelle, M. (2008) “Designing optimal iBGP route-reflection topologies”, *NETWORKING 2008 Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet* , Berlin, Springer, 542-553