



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Χαρτογράφηση του Διαδικτύου με μετρήσεις βασισμένες στο
CAIDA**

Βασίλειος Δ. Ροζής

Επιβλέπων: Δημήτριος Βαρουτάς, Επίκουρος Καθηγητής

ΑΘΗΝΑ

ΝΟΕΜΒΡΙΟΣ 2016

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Χαρτογράφηση του Διαδικτύου με μετρήσεις βασισμένες στο CAIDA

Βασίλειος Δ. Ροζής

A.M.: 1115200900289

ΕΠΙΒΛΕΠΩΝ: **Δημήτριος Βαρουτάς, Επίκουρος Καθηγητής**

ΠΕΡΙΛΗΨΗ

Βασικό αντικείμενο, της πτυχιακής αυτής εργασίας είναι όπως μαρτυρά και ο τίτλος της η χαρτογράφηση του Διαδικτύου με μετρήσεις βασισμένες στον CAIDA. Το Διαδίκτυο είναι ένα πολύπλοκο δίκτυο το οποίο μέρα με τη μέρα μεγαλώνει και αυξάνεται η πολυπλοκότητά του. Καθίσταται ως εκ τούτου απαραίτητη για την ευρωστία και την ασφάλεια του Διαδικτύου μια πλήρης και αναλυτική χαρτογράφηση του. Από την αναπαράσταση του Διαδικτύου μέσα από χάρτες δίδεται μια πλήρη εικόνα της συνδεσιμότητας. Ο CAIDA με το πρόγραμμα της Μακροσκοπικής Τοπολογίας του Διαδικτύου που στηρίζεται σε μεγάλο πλήθος εργαλείων λογισμικού, με συγκεκριμένη λειτουργία το καθένα τους και με το μεγάλο δίκτυο συσκευών και υποδομών παρακολούθησης που διαθέτει, αποτελεί τη μεγαλύτερης κλίμακας προσπάθεια της επιστημονικής κοινότητας για τη χαρτογράφηση του Διαδικτύου.

Η πτυχιακή αυτή εργασία βασίστηκε στην ανάλυση των αποτελεσμάτων που προέκυψαν από την εκτέλεση των εργαλείων του CAIDA με είσοδο τους τα δεδομένα που έχει συλλέξει ο CAIDA καθώς και στα αποτελέσματα που έχουν παραχθεί μέσω των μελετών που έχει εκπονήσει το ερευνητικό προσωπικό του CAIDA.

Αρχικά γίνεται μια περιγραφή των οντοτήτων και πρωτοκόλλων η γνώση των οποίων απαιτείται στη προσπάθεια κατανόησης και χαρτογράφησης στη συνέχεια του Διαδικτύου. Ακολούθως γίνεται μια περιγραφή του CAIDA και της υποδομής μετρήσεων που έχει αναπτύξει του Archipelago. Έπειτα γίνεται ανάλυση των τοπολογικών ερευνών του CAIDA. Γίνεται ανάλυση της επίλυσης ψευδωνύμων δηλαδή της διαδικασίας με την οποία γίνεται η αναγνώριση των διευθύνσεων IP που ανήκουν στον ίδιο δρομολογητή, που αποτελεί ένα πολύ σημαντικό βήμα για τη δημιουργία τοπολογικών χαρτών του Διαδικτύου. Χαρτογραφούνται οι θέσεις μεγάλου πλήθους δρομολογητών και των συσκευών παρακολούθησης του Archipelago. Μελετάτε η AS core οπτικοποίηση του CAIDA που αντιπροσωπεύει ένα μακροσκοπικό στιγμιότυπο δειγμάτων IPv4 και IPv6 τοπολογίας Διαδικτύου που έχει ληφθεί κατά τη διάρκεια ενός έτους και τέλος γίνεται μία συγκριτική ανάλυση των Διαδικτυακών τοπολογιών. Η ακριβής και αναλυτική χαρτογράφηση του Διαδικτύου δεν είναι μια εύκολη υπόθεση. Διαφορετικές μεθόδους μέτρησης της τοπολογίας δίνουν διαφορετικά αποτελέσματα, συνεπώς απαιτείται μεγάλος αριθμός διαφορετικών πηγών δεδομένων μέτρησης και μελέτη διαφορετικών ειδών προκυπτόμενων γραφημάτων ώστε να εξαχθούν έγκυρα αποτελέσματα και να δημιουργηθεί ένας έγκυρος χάρτης του Διαδικτύου.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Χαρτογράφηση Διαδικτύου

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: αυτόνομα συστήματα, CAIDA, χαρτογράφηση, τοπολογία, σύνολα δεδομένων

ABSTRACT

Main object of this diploma thesis is as evidenced by the title of this, the mapping of the Internet with measurements based on CAIDA. The Internet is a complex network which day by day grows and increases its complexity. It becomes therefore necessary for the health and safety of the Internet a complete and detailed mapping. From the representation of the Internet through maps given a full picture of the connectivity. The CAIDA the program of visual topology of the Internet which includes a large number of software tools, with specific function in each of them and with the large devices network monitoring infrastructure available, is the larger effort of the scientific community to map the Internet.

The thesis work is based on an analysis of the results obtained from the performance of his CAIDA tools with input the data collected by the CAIDA and the results are obtained through studies drawn up by the research staff of CAIDA.

First is a description of the entities and protocols, knowledge of which is required in understanding and mapping then the Internet. Next is a description of CAIDA and Archipelago, CAIDA's measurement infrastructure. After an analysis of topological surveys of CAIDA. An analysis of alias resolution that process which is the recognition of IP addresses belonging to the same router, which is a very important step in creating topological maps of the Internet takes place. Mapped the positions of a large number of routers and the Archipelago monitors. Studying the AS core visualization, CAIDA's AS Core visualization represents a macroscopic snapshot of IPv4 and IPv6 Internet topology samples captured over the course of a year and finally made a comparative analysis of Internet topologies. Accurate and detailed mapping of the Internet is not an easy task. Different methods of measuring topology give different results, thus necessitating a large number of different sources of measurement data and study different types of graphs that arise to draw valid results and to provide a valid Internet map.

SUBJECT AREA: Internet Mapping

KEYWORDS: autonomous systems, CAIDA, mapping, topology, datasets

ΕΥΧΑΡΙΣΤΙΕΣ

Η πτυχιακή αυτή εργασία αποτελεί την κορύφωση των σπουδών μου στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ. Ως εκ τούτου οφείλω να ευχαριστήσω όλους αυτούς που διαδραμάτισαν σημαντικό ρόλο στις σπουδές μου στο τμήμα. Θα ήθελα λοιπόν καταρχήν να εκφράσω τις ευχαριστίες μου στον Επίκουρο Καθηγητή κ. Δημήτριο Βαρουτά για την εμπιστοσύνη που μου έδειξε με την ανάθεση μίας ιδιαίτερα ενδιαφέρουσας πτυχιακής εργασίας που συνέβαλε ουσιαστικά στη βελτίωση των γνώσεών μου και για τις γνώσεις που μου παρείχε κατά τη διάρκεια των σπουδών μου.

Ακόμα θα ήθελα να ευχαριστήσω το σύνολο του προσωπικού του τμήματος Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ για τις γνώσεις που μου μετέδωσαν μέσα από ένα πλαίσιο διαλόγου και συνεργασίας και όχι μέσα από στείρα και μηχανιστική παράδοση μαθήματος.

Τέλος θα ήθελα να απευθύνω τις ευχαριστίες μου στην οικογένειά μου, για τη στήριξη και την εμπιστοσύνη που μου έδειξε όλα αυτά τα χρόνια των σπουδών μου, φροντίζοντας για τη καλύτερη δυνατή μόρφωσή μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ	14
1. ΕΙΣΑΓΩΓΗ	15
1.1 Γενικές Αρχές	15
1.1.1 Δρομολογητές.....	15
1.1.2 Αυτόνομα Συστήματα	16
1.1.3 Πρωτόκολλο BGP.....	17
2. ΜΑΚΡΟΣΚΟΠΙΚΗ ΤΟΠΟΛΟΓΙΑ (CAIDA)	20
2.1 Επισκόπηση των Δραστηριοτήτων του CAIDA	20
2.1.1 Ερευνητικά Προγράμματα Χαρτογράφησης του Διαδικτύου	20
2.1.2 Επιστημονικά Προγράμματα για την Ασφάλεια του Διαδικτύου	21
2.1.3 Επιστημονικό Πρόγραμμα στο Τομέα των Οικονομικών	22
2.1.4 Έρευνες για το Μέλλον του Διαδικτύου.....	22
2.2 Ο CAIDA σε Αριθμούς	23
2.3 Δεδομένα του CAIDA	27
2.3.1 Στατιστικά της Συλλογής των Δεδομένων	27
2.3.2 Στατιστικά Κατανομής των Δεδομένων	28
2.4 Archipelago (Ark)	31
2.4.1 Είδη Μετρήσεων.....	32
2.4.1.1 Φιλοξενούμενες Μετρήσεις	32
2.4.1.2 Διαρκείς Μετρήσεις	33
2.4.1.3 Ειδικές Μετρήσεις.....	35
3. ΤΟΠΟΛΟΓΙΚΗ ΕΡΕΥΝΑ	38
3.1 Επίλυση Ψευδωνύμων (Alias Resolution)	38
3.1.1 Εργαλεία Επίλυσης Ψευδωνύμων	38
3.2 Χαρτογράφηση με Χρήση του plot-latlong	42
3.3 Χαρτογράφηση του Archipelago	48
3.4 AS Core: Οπτικοποιώντας την IPv4 και IPv6 Τοπολογία Διαδικτύου σε Μακροσκοπική Κλίμακα	

3.4.1	Εργαλείο Otter	50
3.4.2	Πηγή Δεδομένων	51
3.4.3	Μεθοδολογία Κατάταξης AS.....	52
3.4.3.1	Ποσοτικοποίηση των Επιπτώσεων των Αλλαγών στους Υπολογισμούς του IPv4 AS βαθμού 54	
3.4.3.2	Επιδράσεις στο IPv6.....	56
3.4.3.3	Σύνοψη των Επιδράσεων των δύο Αλλαγών	58
3.4.4	Μεταβολή του AS core στο Πέρασμα των Χρόνων.....	59
3.4.4.1	Μεταβολή σε Σχέση με το 2013.....	60
3.4.4.2	IPv4 και IPv6 AS core το 2005	63
3.5	Συγκριτική Ανάλυση των Διαδικτυακών Τοπολογιών	64
3.5.1	Πηγές Δεδομένων της Μελέτης	64
3.5.1.1	Δεδομένα του traceroute.....	65
3.5.1.2	Δεδομένα BGP για Τοπολογίες AS Επιπέδου.....	66
3.5.1.3	Δεδομένα IRR	67
3.5.2	Τοπολογικές Μετρήσεις.....	68
3.5.3	Κατασκευή Γραφημάτων τοπολογίας Διαδικτύου από τα Δεδομένα.....	69
3.5.3.1	IP-Επιπέδου Γραφήματα.....	69
3.5.3.2	Γραφήματα Επιπέδου Δρομολογητή	70
3.5.3.3	Σύγκριση των IP και Δρομολογητή Επιπέδου Γραφημάτων Βασισμένη σε Πληροφορίες Αμεσης Παρατήρησης.....	75
3.5.3.4	Γραφήματα AS Επιπέδου	79
3.5.4	Η Ανάλυση των Γραφημάτων Τοπολογίας του Διαδικτύου.....	81
3.5.4.1	Χαρακτηριστικά των IP και Δρομολογητή Επιπέδου Γραφημάτων.....	81
3.5.4.2	Χαρακτηριστικά των AS Επιπέδου Γραφημάτων.....	84
4.	ΣΥΜΠΕΡΑΣΜΑΤΑ	95
	ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ	96
	ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ	99

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1: Κατανομή ανά πηγή χρηματοδότησης.....	24
Σχήμα 2: Λειτουργικά έξοδα	25
Σχήμα 3: Έξοδα ανα πηγή χρηματοδότησης.....	26
Σχήμα 4: Έξοδα ανα περιοχή προγράμματος.....	27
Σχήμα 5: Αριθμός των χρηστών που κατεβάζουν δημόσια δεδομένα	29
Σχήμα 6: Πλήθος των δημόσιων δεδομένων που έχουν ληφθεί	29
Σχήμα 7: Ληφθείσες/εγκεκριμένες αιτήσεις για περιορισμένης πρόσβασης πακέτα δεδομένων	30
Σχήμα 8: Πλήθος των περιορισμένης πρόσβασης δεδομένων που έχουν ληφθεί.....	30
Σχήμα 9: Αριθμός των δημοσιευμένων εργασιών	31
Σχήμα 10: Συντεταγμένες του AS στην AS core οπτικοποίηση	52
Σχήμα 11: Σύγκριση του γενικού βαθμού, του βαθμού διέλευσης και του βαθμού εξόδου των ASes.....	53
Σχήμα 12: Συμπεραίνοντας AS συνδεσιμότητα σε όλο το IX χώρο διευθύνσεων.....	54
Σχήμα 13: Υπολογισμός διαγράμματος κουτιού	54
Σχήμα 14: Αλλαγή του AS βαθμού κατά τη μετάβαση από το βαθμό εξόδου στο βαθμό διέλευσης για IPv4 ASes	55
Σχήμα 15: Ποσοστιαία διαφορά βαθμού διέλευσης	55
Σχήμα 16: Βαθμός διέλευσης, βαθμός εξόδου και βαθμός διέλευσης συνεπαγόμενος σε όλο το IX χώρο διευθύνσεων των 10 κορυφαίων ASes.....	56
Σχήμα 17: Ποσοστιαία διαφορά μεταξύ του βαθμού εξόδου και του βαθμού διέλευσης των ASes σε IPv6	57
Σχήμα 18: Ποσοστιαία διαφορά βαθμού διέλευσης σε IPv6	57
Σχήμα 19: Βαθμός διέλευσης, βαθμός εξόδου και βαθμός διέλευσης συνεπαγόμενος σε όλο το IX χώρο διευθύνσεων των 10 κορυφαίων ASes σε IPv6.....	58
Σχήμα 20: Επίδραση της μετάβασης από το βαθμό εξόδου στο βαθμό διέλευσης σε IPv4.....	59
Σχήμα 21: Επίδραση της μετάβασης από το βαθμό εξόδου στο βαθμό διέλευσης σε IPv6.....	59
Σχήμα 22: Μεταβολή του βαθμού διέλευσης συνεπαγόμενου σε όλο το IX χώρο διευθύνσεων σε ένα χρόνο σε IPv4.....	61

Σχήμα 23: Μεταβολή του βαθμού διέλευσης συνεπαγόμενου σε όλο το IX χώρο διευθύνσεων σε ένα χρόνο σε IPv6.....	61
Σχήμα 24: Μεταβολή του βαθμού των κορυφαίων ASes σε IPv4 σε ένα χρόνο	62
Σχήμα 25: Μεταβολή του βαθμού των κορυφαίων ASes σε IPv6 σε ένα χρόνο	63
Σχήμα 26: Στατιστικά στοιχεία των καταχωρήσεων στη RIPE NCC Whois βάση δεδομένων	68
Σχήμα 27:Επιλεγμένα παραδείγματα της αρχικής επεξεργασίας του karar των IP μονοπατιών σε τμήματα	73
Σχήμα 28: Επιλεγμένα βήματα της επεξεργασίας του karar των IP τμημάτων σε συνδέσμους.	73
Σχήμα 29: Βαθμοί κόμβου που προκύπτουν από τις τρεις μεθόδους για τη μέτρηση των μη ανταποκρινόμενων hops	74
Σχήμα 30: Βαθμού κόμβου κατανομές που προκύπτουν από διαφορετικές μεθόδους που ασχολούνται με τα μη ανταποκρινόμενα hops.....	75
Σχήμα 31: Αντιστοίχιση των πραγματικών δρομολογητών σε ένα δεδομένο αριθμό συναγόμενων δρομολογητών	77
Σχήμα 32: Αντιστοίχιση των συναγόμενων δρομολογητών σε ένα δεδομένο αριθμό πραγματικών δρομολογητών.....	77
Σχήμα 33: Το CCDF των βαθμών κόμβου για κάθε μέθοδο επεξεργασίας και πηγή δεδομένων.....	79
Σχήμα 34: CCDF του βαθμού κόμβου των IP και δρομολογητή επιπέδου	82
Σχήμα 35: Μέσος βαθμός γείτονα των IP και δρομολογητή επιπέδου.....	83
Σχήμα 36: Τοπική ομαδοποίηση των IP και δρομολογητή επιπέδου.....	83
Σχήμα 37: CCDF του βαθμού κόμβου των βασισμένων στο Ark AS	85
Σχήμα 38: Μέσος βαθμός γείτονα των βασισμένων στο Ark AS.....	85
Σχήμα 39: Τοπική ομαδοποίηση των βασισμένων στο Ark AS.....	86
Σχήμα 40: CCDF του βαθμού κόμβου των βασισμένων στο traceroute AS	86
Σχήμα 41: Μέσος βαθμός γείτονα των βασισμένων στο traceroute AS.....	87
Σχήμα 42: Τοπική ομαδοποίηση των βασισμένων στο traceroute AS.....	87
Σχήμα 43: Συμβολή του βαθμού των γειτόνων στο μεγαλύτερο μέσο βαθμό γείτονα στο iPlane AS γράφημα έναντι του Ark ITDK ASro	88

Σχήμα 44: Τοπική ομαδοποίηση υπολογισμένη για υπογραφήματα σχηματισμένα απο κόμβους με βαθμούς εντός 25% του ενός από του άλλου	89
Σχήμα 45: CCDF του βαθμού κόμβου των βασισμένων στο BGP AS	90
Σχήμα 46: Μέσος βαθμός γείτονα των βασισμένων στο BGP AS	90
Σχήμα 47: Τοπική ομαδοποίηση των βασισμένων στο BGP AS	91
Σχήμα 48: CCDF του βαθμού κόμβου του συνόλου των AS επιπέδου.....	92
Σχήμα 49: Μέσος βαθμός γείτονα του συνόλου των AS επιπέδου	92
Σχήμα 50: Τοπική ομαδοποίηση του συνόλου των AS επιπέδου	93
Σχήμα 51: Τοπική ομαδοποίηση για υπογραφήματα απο κόμβους με βαθμούς εντός 25% του ενός από του άλλου.....	94

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Δημιουργία μέτρησης στο Vela	36
Εικόνα 2: Αποτελέσματα μέτρησης με χρήση της μεθόδου Ping	36
Εικόνα 3: Γεωγραφικός χάρτης των traceroute μονοπατιών.....	37
Εικόνα 4: Αναφορά του χάρτη της Ελλάδας στο αρχείο marinfo.....	44
Εικόνα 5: Εκτέλεση του plot-latlong με είσοδο το midar-iff.nodes2010.geo	44
Εικόνα 6: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το midar.iff.nodes2010.geo	44
Εικόνα 7: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το midar.iff.nodes2010.geo	45
Εικόνα 8: Εκτέλεση του plot-latlong με είσοδο το kapar-midar-iff.nodes2010.geo	45
Εικόνα 9: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το kapar-midar.iff.nodes2010.geo	45
Εικόνα 10: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το kapar-midar.iff.nodes2010.geo	46
Εικόνα 11: Εκτέλεση του plot-latlong με είσοδο το midar-iff.nodes2013.geo	46
Εικόνα 12: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το midar.iff.nodes2013.geo	46
Εικόνα 13: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το midar.iff.nodes2013.geo	47
Εικόνα 14: Εκτέλεση του plot-latlong με είσοδο το kapar-midar-iff.nodes2013.geo	47
Εικόνα 15: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το kapar-midar.iff.nodes2013.geo	47
Εικόνα 16: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το kapar-midar.iff.nodes2013.geo	48
Εικόνα 17: Χάρτης συστημάτων παρακολούθησης του Archipelago	49
Εικόνα 18: Κύριο παράθυρο του otter	51
Εικόνα 19: Απεικόνιση 30.000 κόμβων από το otter	51
Εικόνα 20: Μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου το 2014	60
Εικόνα 21: Μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου για το 2013.....	60
Εικόνα 22: Μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου το 2005	63

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1:Ακριβές ποσά χρηματοδότησης.....	25
Πίνακας 2: Λειτουργικά έξοδα(ποσά).....	26
Πίνακας 3: Έξοδα ανά πηγή(ποσά).....	26
Πίνακας 4: Έξοδα ανα περιοχή προγράμματος(ποσά).....	27
Πίνακας 5: Συστήματα παρακολούθησης του Archipelago στην Ελλάδα.....	49
Πίνακας 6: Σύνολα δεδομένων που χρησιμοποιούνται στη μελέτη.....	64
Πίνακας 7: Βασικά στατιστικά στοιχεία των IP και δρομολογητή τοπολογίας γραφημάτων	82
Πίνακας 8: Βασικά στατιστικά των AS γραφημάτων	84

ΠΡΟΛΟΓΟΣ

Η παρούσα πτυχιακή εργασία με τίτλο «Χαρτογράφηση του Διαδικτύου με μετρήσεις βασισμένες στο Caida» εκπονήθηκε στα πλαίσια της ολοκλήρωσης των προϋποθέσεων για τη λήψη του πτυχίου μου από το τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ με επιβλέπων καθηγητή τον Επίκουρο Καθηγητή κ. Δημήτριο Βαρουτά.

Σκοπός μου κατά την εκπόνηση της πτυχιακής αυτής εργασίας, δεν ήταν μόνο η ακριβής και όσο το δυνατόν πληρέστερη ανάλυση του θέματος. Έγινε προσπάθεια ώστε το περιεχόμενο της εργασίας να είναι σαφές και κατανοητό, για το λόγο αυτό έγινε χρήση μεγάλου αριθμού σχημάτων, παραδειγμάτων, εικόνων και πινάκων και ευκολονόητος σχολιασμός των αποτελεσμάτων που προέκυπταν μετά από τη κατάλληλη επεξεργασία των συνόλων δεδομένων του CAIDA κατά την ανάλυση του θέματος.

Η πτυχιακή αυτή εργασία βασίστηκε στην ανάλυση των αποτελεσμάτων που προέκυψαν από την εκτέλεση των εργαλείων του CAIDA με είσοδο τους τα δεδομένα που έχει συλλέξει ο CAIDA καθώς και στα αποτελέσματα που έχουν παραχθεί μέσω των μελετών που έχει εκπονήσει το ερευνητικό προσωπικό του CAIDA.

Μέσα από αυτή την εργασία διεύρυνα σημαντικά τις γνώσεις μου πάνω στην οργάνωση και τη συνδεσιμότητα στο Διαδίκτυο και στο τομέα των δικτύων εν γένει. Ένα τομέα τον οποίο βρίσκω ιδιαίτερα ενδιαφέρον και με τον οποίο θα ήθελα να ασχοληθώ και στο μέλλον.

Θα ήθελα να εκφράσω τις ευχαριστίες μου στον επιβλέποντα καθηγητή μου κ. Δημήτριο Βαρουτά, για τη δυνατότητα που μου έδωσε να ασχοληθώ με ένα ιδιαίτερα ενδιαφέρον θέμα και μέσω αυτού να βελτιώσω τις γνώσεις μου, όπως επίσης και για τις γνώσεις τις οποίες μου μεταλαμπάδευσε, μέσα από τη παρακολούθηση των μαθημάτων που διδάσκει.

Επίσης θα ήθελα να ευχαριστήσω όλο το προσωπικό του τμήματος Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ για τις γνώσεις που μου παρείχαν κατά τη διάρκεια των σπουδών μου μέσα από ένα γόνιμο κλίμα διαλόγου και συνεργασίας καθώς και για τη διαρκή προσπάθεια που καταβάλλουν για τη βελτίωση των σπουδών στο τμήμα.

Τέλος θα ήθελα να ευχαριστήσω την οικογένειά μου για τη στήριξη που μου παρείχαν κατά τη διάρκεια των σπουδών μου και την εμπιστοσύνη που μου έδειχναν.

1. ΕΙΣΑΓΩΓΗ

Το Διαδίκτυο είναι ένα πολύπλοκο δίκτυο που αποτελείται μεταξύ άλλων από συσκευές δρομολόγησης (δρομολογητές) οι οποίες είναι οργανωμένες σε διοικητικές οντότητες, τα αυτόνομα συστήματα. Κάθε αυτόνομο σύστημα έχει τις δικές του πολιτικές δρομολόγησης και κριτήρια σχεδιασμού, ανάλογα με τους τεχνολογικούς, γεωγραφικούς και οικονομικούς περιορισμούς και στοχεύει στη μεγιστοποίηση της απόδοσης από την άποψη του εύρους ζώνης και της καθυστέρησης. Κάθε αυτόνομο σύστημα δεν γνωρίζει τη κατάσταση των υπολοίπων.

Το Διαδίκτυο είναι αχανές και συνεχώς μεγαλώνει με πολύ μεγάλο ρυθμό. Στο Διαδίκτυο είναι συνδεδεμένες πολύ κρίσιμες υποδομές σε ολόκληρο το κόσμο και διακινούνται πολύ σημαντικές πληροφορίες. Απαιτείται μια προσπάθεια για τη δημιουργία μίας εικόνας της συνδεσιμότητας στο Διαδίκτυο. Μια πλήρης και αναλυτική χαρτογράφηση του Διαδικτύου αποτελεί εχέγγυο για την ασφάλεια του καθώς μπορεί να το θωρακίσει έναντι κακόβουλων παρεμβάσεων. Μέσα από την ύπαρξη ενός χάρτη του Διαδικτύου γίνεται εφικτός ο εντοπισμός της πηγής της κακόβουλης επίθεσης και οι συνακόλουθες αλλαγές στην συνδεσιμότητα.

Μια πλήρης και αναλυτική χαρτογράφηση του Διαδικτύου απαιτεί την ύπαρξη μεγάλου όγκου δεδομένων που να έχει ληφθεί από ένα μεγάλο δίκτυο συσκευών παρακολούθησης του Διαδικτύου διασκορπισμένων σε ολόκληρο το πλανήτη και τα οποία είναι πολύτιμα για την ανάπτυξη εμπειρικών μοντέλων, που θα προσομοιώνουν τη λειτουργία του Διαδικτύου. Καθώς επίσης και κατάλληλων εργαλείων για την ανάλυση και την οπτικοποίηση των δεδομένων που λήφθηκαν από τα συστήματα παρακολούθησης. Ο CAIDA με το πρόγραμμα της Μακροσκοπικής Τοπολογίας του Διαδικτύου που στηρίζεται σε πολλά εργαλεία λογισμικού, καθένα από τα οποία επιτελεί συγκεκριμένες λειτουργίες και με το μεγάλο δίκτυο συσκευών και υποδομών παρακολούθησης που διαθέτει. Αποτελεί τη μεγαλύτερης κλίμακας προσπάθεια της επιστημονικής κοινότητας για τη χαρτογράφηση του Διαδικτύου. Τα εργαλεία αυτά του CAIDA και τα δεδομένα που εκείνος έχει συλλέξει και προσφέρει χρησιμοποιούνται στα πλαίσια αυτής της εργασίας για τη μελέτη και τη χαρτογράφηση του Διαδικτύου.

Η μελέτη της διαδικασίας της χαρτογράφησης του Διαδικτύου θα ήταν όμως σε μεγάλο βαθμό δύσκολη χωρίς τη κατανόηση ορισμένων δομικών στοιχείων και πρωτοκόλλων στη προσπάθεια χαρτογράφησης του Διαδικτύου που αποτελούν βασικότατα στοιχεία αυτής της μελέτης. Για το λόγο αυτό αρχικά αποσαφηνίζονται, μέσα από την παρουσίαση τους που λαμβάνει χώρα στο πρώτο κεφάλαιο, οι όροι δρομολογητής, αυτόνομο σύστημα, πρωτόκολλα δρομολόγησης, πρωτόκολλο δρομολόγησης BGP.

1.1 Γενικές Αρχές

1.1.1 Δρομολογητές

Το δημόσιο Διαδίκτυο είναι ένα παγκόσμιο δίκτυο υπολογιστών, δηλαδή ένα δίκτυο που διασυνδέει εκατομμύρια υπολογιστικές συσκευές σε όλο το κόσμο. Οι δρομολογητές αποτελούν ένα από τα σημαντικότερα συστατικά της δομής του διαδικτύου. Οι δρομολογητές

ανήκουν στο επίπεδο 3 του μοντέλου OSI, το επίπεδο δικτύου (network layer) και λειτουργούν ως σημεία ένωσης δύο ή περισσότερων δικτύων, δικτύων που μπορεί να διαφέρουν, παρόλα αυτά οι δρομολογητές αποκρύπτουν τις τεχνικές διαφορές που υπάρχουν στα επιμέρους δίκτυα και εγκαθιδρύουν επικοινωνία ανάμεσά τους.

Ο κύριος ρόλος των δρομολογητών είναι η προώθηση πακέτων από εισερχόμενες ζεύξεις σε εξερχόμενες ζεύξεις [1][2]. Όταν ο αποστολέας και ο παραλήπτης ανήκουν σε διαφορετικά δίκτυα, μεσολαβεί η προώθηση των πακέτων από τους αντίστοιχους ενδιάμεσους δρομολογητές. Στο δημόσιο Διαδίκτυο υπάρχουν πολλά εναλλακτικά μονοπάτια. Οι ενδιάμεσοι δρομολογητές που θα προωθήσουν το μήνυμα πρέπει να επιλέξουν τη καλύτερη διαδρομή. Κάθε δρομολογητής χρησιμοποιεί ένα ή περισσότερα πρωτόκολλα δρομολόγησης. Με βάση αυτά τα πρωτόκολλα ο δρομολογητής καθορίζει ποιος ή ποιοι δρομολογητές είναι οι καταλληλότεροι κάθε χρονική στιγμή και δρομολογεί τα πακέτα δεδομένων προς αυτούς. Η διαδικασία της δρομολόγησης (routing) βασίζεται στους πίνακες δρομολόγησης (routing tables). Κάθε δρομολογητής κάνει χρήση ενός πίνακα δρομολόγησης στον οποίο αντιστοιχίζεται κάθε τελικός προορισμός με έναν επόμενο δρομολογητή. Ο πίνακας δρομολόγησης είναι προσαρμοσμένος στις εκάστοτε αλλαγές του δικτύου καθώς περιοδικά γίνεται ανανέωσή του. Ώς εκ τούτου με την άφιξη ενός πακέτου, ο δρομολογητής το προωθεί αφού πρώτα αντιστοιχίσει το παραλήπτη με τη κατάλληλη εγγραφή του πίνακα δρομολόγησης. Οι δρομολογητές χρησιμοποιούν πρωτόκολλα δυναμικής δρομολόγησης (dynamic routing protocols) για τη δημιουργία και τη συντήρηση των πινάκων δρομολόγησης. Τα πρωτόκολλα δυναμικής δρομολόγησης χωρίζονται σε δύο κατηγορίες. Τα πρωτόκολλα Κατάστασης Ζεύξης (Link – State) και τα πρωτόκολλα Διανύσματος Απόστασης (Distance – Vector).

Πρωτόκολλα Κατάστασης Ζεύξης αποτελούν το πρωτόκολλο IS – IS και το πρωτόκολλο ανοικτό – πρώτα η βραχύτερη διαδρομή (Open Shortest Path First) OSPF. Στα πρωτόκολλα Κατάστασης Ζεύξης κάθε δρομολογητής έχει καταχωρημένο ένα πλήρη τοπολογικό χάρτη(δηλαδή ένα γράφο) του δικτύου. Ο δρομολογητής στη συνέχεια εκτελεί τοπικά τον αλγόριθμο ελάχιστου κόστους του Dijkstra για να καθορίσει ένα δέντρο βραχύτερης διαδρομής προς όλα τα υποδίκτυα, με το συγκεκριμένο δρομολογητή ως ριζικό κόμβο και να αποφασίσει ως εκ τούτου τη βέλτιστη διαδρομή.

Στα πρωτόκολλα Διανύσματος Απόστασης ανήκουν το πρωτόκολλο Δρομολόγησης Πληροφοριών (Routing Information Protocol) RIP και το Εσωτερικής Πύλης πρωτόκολλο δρομολόγησης (Interior Gateway Routing Protocol) IGRP. [3] Στα πρωτόκολλα Διανύσματος Απόστασης κάθε δρομολογητής καταχωρεί στο πίνακα δρομολόγησης του όλες τις γειτονικές συνδέσεις που μπορεί να πραγματοποιήσει, το τελικό προορισμό της κάθε γειτονικής σύνδεσης και το κόστος της σύνδεσης. Κατόπιν κάθε δρομολογητής επικοινωνεί με τους γειτονικούς του δρομολογητές στο σύνολό τους περιοδικά και τους αποστέλλει τις πληροφορίες αυτές. Με τη λήψη κάποιας γειτονικής σύνδεσης σε ένα προορισμό με χαμηλότερο κόστος με χρήση του αλγορίθμου Bellman – Ford γίνεται αντικατάσταση των εγγραφών του πίνακα δρομολόγησης.

1.1.2 Αυτόνομα Συστήματα

Το Διαδίκτυο είναι ένα πολύπλοκο δίκτυο που αποτελείται από συσκευές δρομολόγησης (δρομολογητές) οι οποίες είναι οργανωμένες σε διοικητικές οντότητες, τα αυτόνομα συστήματα (autonomous systems). Κάθε αυτόνομο σύστημα (AS) έχει τις δικές του πολιτικές δρομολόγησης και κριτήρια σχεδιασμού, ανάλογα με τους τεχνολογικούς, γεωγραφικούς

και οικονομικούς περιορισμούς και στοχεύει στη μεγιστοποίηση της απόδοσης από την άποψη του εύρους ζώνης και της καθυστέρησης. Τα αυτόνομα συστήματα είναι σήμερα ταυτισμένα με μεγάλους φορείς, εταιρίες παροχής υπηρεσιών Διαδικτύου (Internet Service Providers), πανεπιστήμια, σημεία ανταλλαγής Διαδικτύου (Internet Exchange Points) IXP και δικτύων διανομής περιεχομένου (Content Delivery Networks) CDN, ανάμεσα σε άλλα, και συνδέονται όλα μεταξύ τους ως αποτέλεσμα εμπορικών συμφωνιών είτε ως ομότιμοι (peers) είτε με σχέσεις παρόχου – πελάτη (customer – client). Τα αυτόνομα συστήματα μπορούν επίσης να ταξινομηθούν σε αυτόνομα συστήματα κορμού (core) και περιφερειακά. Τα αυτόνομα συστήματα κορμού είναι τυπικά οι μεγάλοι φορείς και μερικά δίκτυα διανομής περιεχομένου και είναι πυκνά συνδεδεμένα μεταξύ τους. Αντ' αυτού, οι περισσότερες εταιρίες παροχής υπηρεσιών Διαδικτύου βρίσκονται στη περιφέρεια και είναι συνδεδεμένες σε ένα ή και περισσότερα αυτόνομα συστήματα κορμού. Αυτή η διαίρεση κορμού περιφέρειας παρέχει στο Διαδίκτυο μία ενδιαφέρουσα ιεραρχία. Πολλές ιδιότητες μεγάλης κλίμακας όπως η επεκτασιμότητα και η ελαστικότητα του Διαδικτύου προκύπτουν σε μεγάλο βαθμό ως αποτέλεσμα της ιεραρχικής δομής και αυτό-οργάνωσής του [4]. Η κατανόησή τους θα μπορούσε να βοηθήσει στην ανάπτυξη καλύτερων μοντέλων του Διαδικτύου και τη βελτίωση της απόδοσής του, βελτιστοποιώντας τη δρομολόγηση ή μειώνοντας τη κυκλοφοριακή συμφόρηση για παράδειγμα.

Σε κάθε αυτόνομο σύστημα δίδεται ένας μοναδικός αριθμός 16 bit μήκους που ονομάζεται αριθμός αυτόνομου συστήματος (ASN) και αποτελεί το διακριτικό του αυτόνομου συστήματος στο Διαδίκτυο [5]. Κάθε οργανισμός διαχειρίζεται το δικό του αυτόνομο σύστημα και εφαρμόζει για τη δρομολόγηση των πακέτων στο εσωτερικό του συστήματος μια συγκεκριμένη πολιτική. Τα περισσότερο γνωστά πρωτόκολλα εσωτερικής δρομολόγησης (Interior Routing Protocols) IGP που χρησιμοποιούνται είναι το OSPF και το RIP. Πιο συγκεκριμένα το RIP αποτελεί το πιο δημοφιλές πρωτόκολλο στα μικρά ιδιωτικά δίκτυα ενώ το OSPF στα μεγάλα δίκτυα.

Κάθε αυτόνομο σύστημα που επικοινωνεί με άλλα έχει ένα ή περισσότερους παραμεθόριους δρομολογητές (border routers) που είναι υπεύθυνοι για δρομολόγηση πακέτων που απευθύνονται ή έρχονται από έξω από το αυτόνομο σύστημα αυτό[2]. Ο παραμεθόριος δρομολογητής ενημερώνει τους παραμεθόριους δρομολογητές των άλλων αυτόνομων συστημάτων σχετικά με τα δίκτυα που υπάρχουν εντός του. Ως εκ τούτου είναι εφικτή η αποστολή πακέτων δεδομένων μεταξύ γειτονικών αυτόνομων συστημάτων. Η δρομολόγηση των πακέτων ανάμεσα σε διαφορετικά αυτόνομα συστήματα ονομάζεται εξωτερική δρομολόγηση (exterior routing). Ένα πρωτόκολλο εξωτερικής δρομολόγησης (Exterior Gateway Protocol) EGP είναι υπεύθυνο για την ανταλλαγή πληροφοριών δρομολόγησης μεταξύ των αυτόνομων συστημάτων. Κάθε αυτόνομο σύστημα πρέπει να διαθέτει μια έγκυρη και ενημερωμένη λίστα με τους προορισμούς που μπορεί να φθάσει ένα πακέτο εάν διαπεράσει το αυτόνομο σύστημα και ένα αξιόπιστο πλάνο διαδρομών εντός του. Το πρωτόκολλο BGP είναι το ντε φάκτο πρότυπο πρωτόκολλο διατομεακής (interdomain) δρομολόγησης στο Διαδίκτυο σήμερα.

1.1.3 Πρωτόκολλο BGP

Το πρωτόκολλο δρομολόγησης BGP αποτελεί το κύριο πρωτόκολλο δρομολόγησης του Διαδικτύου [1]. Ως ένα πρωτόκολλο δρομολόγησης διαυτόνομου συστήματος, παρέχει σε κάθε αυτόνομο σύστημα ένα τρόπο για:

1. Τη λήψη πληροφοριών προσέγγισης υποδικτύου από γειτονικά αυτόνομα συστήματα.

2. Τη διάδοση των πληροφοριών προσέγγισης σε όλους τους δρομολογητές που είναι εσωτερικοί μέσα στο αυτόνομο σύστημα.
3. Το καθορισμό των “καλών” διαδρομών προς υποδίκτυα, με βάση τις πληροφορίες προσέγγισης και μια πολιτική αυτόνομου συστήματος.

Το σημαντικότερο είναι ότι το πρωτόκολλο BGP επιτρέπει σε κάθε υποδίκτυο να δημοσιοποιεί την ύπαρξή του στο υπόλοιπο Διαδίκτυο. Το υποδίκτυο ανακοινώνει την ύπαρξή του και το BGP σιγουρεύει ότι όλα τα αυτόνομα συστήματα μέσα στο Διαδίκτυο γνωρίζουν για το υποδίκτυο και πώς να φθάσουν εκεί. Αν δεν υπήρχε το BGP κάθε υποδίκτυο θα ήταν απομονωμένο και άγνωστο για το υπόλοιπο Διαδίκτυο. Στο πρωτόκολλο BGP ζεύγη δρομολογητών ανταλλάσσουν πληροφορίες δρομολόγησης πάνω σε ημιμόνιμες συνδέσεις TCP. Τυπικά υπάρχει μία τέτοια σύνδεση BGP TCP για κάθε ζεύξη που συνδέει απευθείας δύο δρομολογητές διαφορετικών αυτόνομων συστημάτων. Αλλά και ημιμόνιμες συνδέσεις BGP TCP μεταξύ δρομολογητών μέσα σε ένα αυτόνομο σύστημα. Για κάθε σύνδεση AS οι δύο δρομολογητές στα δύο άκρα της σύνδεσης ονομάζονται ομότιμοι BGP και η σύνδεση TCP πάνω στην οποία στέλνονται όλα τα μηνύματα BGP ονομάζεται σύνοδος BGP. Επιπλέον, μια σύνοδος (session) BGP που διατρέχει δύο αυτόνομα συστήματα ονομάζεται σύνοδος εξωτερικού BGP (eBGP) και μία σύνοδος ανάμεσα σε δύο δρομολογητές που ανήκουν στο ίδιο αυτόνομο σύστημα καλείται σύνοδος εσωτερικού BGP (iBGP) [1].

Στο BGP οι προορισμοί δεν είναι υπολογιστές αλλά είναι προθέματα τύπου CIDR, όπου το κάθε πρόθεμα παριστά ένα υποδίκτυο ή μια συλλογή υποδικτύων. Όταν ένας δρομολογητής πύλης (σε οποιοδήποτε αυτόνομο σύστημα) δέχεται προθέματα που έμαθε από eBGP, ο δρομολογητής πύλης χρησιμοποιεί τις συνόδους του iBGP ώστε να διανείμει τα προθέματα σε άλλους δρομολογητές μέσα στο αυτόνομο σύστημα. Ακόμα όταν ένας δρομολογητής μαθαίνει για ένα νέο πρόθεμα, δημιουργεί μια καταχώριση για το πρόθεμα στο πίνακα προώθησής του.

Σχετικά με την επιλογή διαδρομής BGP όπως περιγράφηκε παραπάνω το BGP χρησιμοποιεί eBGP και iBGP ώστε να διανείμει διαδρομές σε όλους τους δρομολογητές μέσα σε ένα αυτόνομο σύστημα. Από αυτή τη διανομή ένας δρομολογητής μπορεί να μάθει για περισσότερες από μία διαδρομές προς ένα πρόθεμα, ως εκ τούτου ο δρομολογητής πρέπει να επιλέξει μία από αυτές τις διαδρομές. Οι είσοδοι σε αυτή τη διεργασία επιλογής διαδρομής είναι όλες οι διαδρομές, που ο δρομολογητής έχει μάθει και έχει αποδεχθεί. Αν υπάρχουν περισσότερες της μίας διαδρομές προς το ίδιο πρόθεμα, τότε το BGP εκτελεί τους παρακάτω κανόνες επιλογής με τη σειρά, μέχρι να μείνει μια μοναδική διαδρομή:

- Στις διαδρομές εκχωρείται μια τοπική τιμή προτίμησης, ως ένα από τα ιδιοχαρακτηριστικά τους. Η τοπική προτίμηση μιας διαδρομής μπορεί να τεθεί από το δρομολογητή ή μπορεί να έχει δημοσιοποιηθεί στο δρομολογητή από ένα άλλο δρομολογητή μέσα στο ίδιο αυτόνομο σύστημα. Αυτή είναι μια απόφαση πολιτικής που εξαρτάται από το διαχειριστή του δικτύου του αυτόνομου συστήματος. Επιλέγονται οι διαδρομές με τις υψηλότερες τοπικές τιμές προτίμησης.
- Από τις διαδρομές που επελέγησαν (οι οποίες έχουν την ίδια τοπική τιμή προτίμησης) επιλέγεται η διαδρομή με το βραχύτερο μονοπάτι (path) αυτόνομου συστήματος AS-PATH. Αν αυτός ήταν ο μοναδικός κανόνας για επιλογή διαδρομής, τότε το BGP θα χρησιμοποιούσε ένα αλγόριθμο διανύσματος απόστασης για καθορισμό διαδρομής, όπου ο δείκτης μέτρησης απόστασης χρησιμοποιεί τον αριθμό αλμάτων αυτόνομου συστήματος αντί του αριθμού

αλμάτων δρομολογητών.

- Από τις διαδρομές που επελέγησαν (οι οποίες έχουν την ίδια τοπική τιμή προτίμησης και το ίδιο μήκος AS-PATH), επιλέγεται η διαδρομή με το πλησιέστερο δρομολογητή επόμενου άλματος (NEXT-HOP) Στη συγκεκριμένη περίπτωση ο όρος πλησιέστερος σημαίνει το δρομολογητή, για τον οποίο το κόστος της διαδρομής ελάχιστου κόστους, που καθορίζεται από τον αλγόριθμο ενδο-AS, είναι το μικρότερο. Η διεργασία αυτή ονομάζεται δρομολόγηση καυτής πατάτας.
- Αν έχουν επιλεγεί πάλι περισσότερες από μια διαδρομές, ο δρομολογητής χρησιμοποιεί ταυτότητες BGP για να επιλέξει τη διαδρομή.

2. ΜΑΚΡΟΣΚΟΠΙΚΗ ΤΟΠΟΛΟΓΙΑ (CAIDA)

Ο CAIDA είναι μία συλλογική επιχείρηση μεταξύ οργανώσεων στον εμπορικό τομέα, το κυβερνητικό τομέα και τον τομέα της έρευνας, βάση του οποίου αποτελεί το San Diego Supercomputer Center το οποίο είναι μία οργανωμένη ερευνητική μονάδα στο Πανεπιστήμιο της Καλιφόρνιας, Σαν Ντιέγκο, με στόχο την προώθηση μεγαλύτερης συνεργασίας στους τομείς της μηχανικής και τη διατήρηση μίας στιβαρής, επεκτάσιμης παγκόσμιου διαδικτύου υποδομής. Ειδικότερα αποστολή του CAIDA αποτελεί η διερεύνηση των πρακτικών και θεωρητικών στοιχείων του Διαδικτύου με σκοπό να βελτιώσει την ακεραιότητα στο τομέα της επιστήμης του διαδικτύου, να παρέχει μακροσκοπική γνώση σχετικά με την υποδομή του διαδικτύου, τη συμπεριφορά της, τη χρήση της και την εξέλιξή της, να προάγει ένα συνεργατικό περιβάλλον στο οποίο τα δεδομένα μπορούν να αποκτηθούν, να αναλυθούν και (κατά περίπτωση) να μοιραστούν και τέλος να ενημερώνει τις επιστημονικές, τεχνολογικές και επικοινωνιακές δημόσιες πολιτικές.

2.1 Επισκόπηση των Δραστηριοτήτων του CAIDA

Ο CAIDA το 2014 εκτόνωσε ερευνητικές δραστηριότητες προσανατολισμένες στη τοπολογία, τη δρομολόγηση, τη κίνηση, την ασφάλεια και τη σταθερότητα του διαδικτύου, στο μέλλον της αρχιτεκτονικής διαδικτύου, την οικονομία και τη πολιτική. Τις παραπάνω δραστηριότητες υποστήριξαν βασισμένες σε μετρήσεις μελέτες του διαδικτύου, τόσο στον CAIDA όσο και στον υπόλοιπο κόσμο, με έμφαση στην ακεραιότητα και τη υγεία του παγκόσμιου οικοσυστήματος του Διαδικτύου.

2.1.1 Ερευνητικά Προγράμματα Χαρτογράφησης του Διαδικτύου

Πιο συγκεκριμένα έλαβαν χώρα 2 ερευνητικά προγράμματα στην επιστημονική περιοχή της χαρτογράφησης του Διαδικτύου [6]. Το πρώτο ερευνητικό πρόγραμμα είχε σαν αντικείμενο τις χαρτογραφικές δυνατότητες για μία σημαντική κυβερνο-υποδομή. Οι χάρτες της τοπολογίας του Διαδικτύου αποτελούν ένα πολύ σημαντικό εργαλείο για το χαρακτηρισμό της σημαντικής αυτής υποδομής και τη κατανόηση των μακροσκοπικών της ιδιοτήτων, της δυναμικής της συμπεριφοράς, της επίδοσής της και της εξέλιξής της. Οι χάρτες αυτοί είναι επίσης καίριας σημασίας για τη προσομοίωση, την ανάλυση και τη ρεαλιστική μοντελοποίηση του Διαδικτύου και άλλων μεγάλης κλίμακας πολύπλοκων δικτύων. Ενώ ακόμα μπορούν να κατασκευαστούν για διαφορετικά στρώματα ή βαθμούς ανάλυσης (granularities), π.χ IP διεύθυνση, δρομολογητή, σημείων παρουσίας (points of presence) PoPs, οπτικής ίνας/χάλκινο καλώδιο, αυτόνομο σύστημα(AS), ISP/οργανισμός. Οι επιπέδου (level) σημείων παρουσίασης και επιπέδου δρομολογητή χάρτες τοπολογίας μπορούν πολύ δυναμικά να ενημερώσουν και να βαθμονομήσουν εκτιμήσεις ευπάθειας και να υπάρξει επίγνωση της κατάστασης μιας σημαντικής δικτυακής υποδομής. Οι ISP επιπέδου τοπολογίες μερικές φορές καλούνται AS επιπέδου ή διατομεακής δρομολόγησης τοπολογίες (αν και ένας ISP μπορεί να κατέχει πολλά ASes οπότε ένα επιπέδου AS γράφημα είναι μία ελαφρώς πιο λεπτομερής ανάλυση) παρέχουν γνώσεις σχετικά με οικονομικές, τεχνικές, πολιτικές και ασφάλειας ανάγκες, του σε μεγάλο βαθμό ανεξέλεγκτου οικοσυστήματος ομοτιμίας (peering). Ο CAIDA πραγματοποιεί μετρήσεις του Διαδικτύου μακροσκοπικής τοπολογίας από το 1998. Τα τρέχοντα εργαλεία που χρησιμοποιούνται παρακολουθούν την παγκόσμια IP επιπέδου συνδεσιμότητα στέλνοντας πακέτα ανίχνευσης από ένα σύνολο από πηγές παρακολούθησης σε εκατομμύρια γεωγραφικά καταναμημένους προορισμούς σε όλο

το φάσμα των IPv4 διευθύνσεων (addresses). Από το 2008 γίνεται ανίχνευση και στο φάσμα των IPv6 διευθύνσεων επίσης.

Ενώ το δεύτερο ερευνητικό πρόγραμμα είχε σαν αντικείμενο την απογραφή του Διαδικτύου. Αποτελεί πρόκληση για τη κατανόηση της εξέλιξης της Διαδικτυακής υποδομής η έλλειψη μηχανισμών συστηματικής παρακολούθησης του βαθμού στον οποίο οι διευθύνσεις IP που έχουν παραχωρηθεί χρησιμοποιούνται στη πραγματικότητα. Η χρήση IPv4 διευθύνσεων έχει παρακολουθηθεί μέσω ενεργής σάρωσης ολόκληρου του φάσματος των IPv4 διευθύνσεων. Ωστόσο, πολλά δίκτυα δεν ανταποκρίνονται σε τέτοια ερευνητικά πακέτα και ως εκ τούτου, υπάρχει ανάγκη εναλλακτικών προσεγγίσεων για την εκτίμηση της χρήσης αυτών των διευθύνσεων. Στόχος αυτού του προγράμματος ήταν να βελτιωθούν οι τρέχουσες μεθοδολογίες για την εξαγωγή συμπερασμάτων σχετικά με τη χρήση του φάσματος των IPv4 (και στο μέλλον IPv6) διευθύνσεων.

Ακόμα εκπονήθηκε ερευνητικό πρόγραμμα με θέμα «Χαρτογράφηση διασύνδεσης στο Διαδίκτυο: Συντοπισμός, Συνδεσιμότητα και Συμφόρηση» το οποίο αποτέλεσε αντικείμενο συνεργασίας του CAIDA με το τμήμα επιστήμης υπολογιστών του MIT και το εργαστήριο Τεχνητής Νοημοσύνης(MIT/CSAIL) και αποτέλεσε ένα χρηματοδοτημένο από το NSF πρόγραμμα [7]. Το πρόγραμμα αυτό επικεντρώνεται σε δύο συναφείς μετασχηματισμούς του οικοσυστήματος του Διαδικτύου:

- Στην ανάδειξη των ανταλλαγών στο Διαδίκτυο (IXes) ως σημεία αναφοράς στο πλέγμα της διασύνδεσης.
- Στο ολοένα σημαντικότερο ρόλο των παρόχων περιεχομένου (content providers) και των δικτύων διανομής περιεχομένου (CDNs) ως κύριες πηγές ροής κυκλοφορίας (traffic flowing) στο Διαδίκτυο.

Στο πρόγραμμα αυτό έγινε προσπάθεια να δημιουργηθεί ένας νέο είδος σημασιολογικά πλούσιου Διαδικτυακού χάρτη ώστε να αποσαφηνίσει το ρόλο των ανταλλαγών στο Διαδίκτυο στην ισχυρά διευκολυμένη και γεωγραφικά ποικίλη αλλά σύνθετη διατομεακή συνδεσιμότητα και να αναπτυχθούν τεχνικές για τη μέτρηση της συμφόρησης (congestion) σε διατομεακές συνδέσεις ανάμεσα σε ISPs. Τέλος έγινε διερεύνηση ζητημάτων σχετικών με τα επιστημονικά μοντέλα, τη δημόσια πολιτική, τα οικονομικά των δικτύων και την ανθεκτικότητα των υποδομών.

2.1.2 Επιστημονικά Προγράμματα για την Ασφάλεια του Διαδικτύου

Η επιστημονική περιοχή της ασφάλειας του Διαδικτύου αποτέλεσε επίσης τομέα στον οποίο εκπονήθηκαν επιστημονικά προγράμματα. Με σημείο αναφοράς τη παρακολούθηση της ασφάλειας και της σταθερότητας του Διαδικτύου πραγματοποιήθηκαν δύο επιστημονικά προγράμματα από τον CAIDA.

Το πρώτο είχε σαν αντικείμενο την ανίχνευση και ανάλυση των μεγάλης κλίμακας διακοπών λειτουργίας στη Διαδικτυακή υποδομή, για το σκοπό αυτό έγινε μια προσπάθεια να δημιουργηθεί μια τεχνολογία που να μπορεί να χρησιμεύσει ως βάση για αυτοματοποιημένα συστήματα ανίχνευσης έγκαιρης προειδοποίησης για μεγάλης κλίμακας διακοπές λειτουργίας του Διαδικτύου και να χαρακτηρίζει τέτοια συμβάντα (υποδομή και πρωτόκολλα που επηρεάστηκαν, χρονοδιάγραμμα, ASes και γεωγραφικές περιοχές που εμπλέκονται, κλπ.). Ειδικότερα έγινε προσπάθεια εύρεσης μεθόδων για την ανάλυση και την ομαδοποίηση των δεδομένων των μετρήσεων στο Διαδίκτυο από πολλαπλές διαθέσιμες πηγές που φέρνουν στο φώς διακοπές σύνδεσης στο Διαδίκτυο, συμπεριλαμβανομένων των

παγκόσμιων διακοπών σύνδεσης εξαιτίας πολιτικών ή καταστροφικών αιτιών. Τα αποτελέσματα όλης αυτής της προσπάθειας ήταν η ανάπτυξη του BGPStream ενός περιβάλλοντος λογισμικού που επεξεργάζεται μεγάλο όγκο ιστορικών και ζωντανών BGP δεδομένων μέτρησης και επικουρικά του BGPStream η ανάπτυξη επιπλέον λογισμικού ικανού να παρακολουθεί τη BGP προσβασιμότητα των χωρών και των ASes. Η συνέχιση της έρευνας της κρυφής κίνησης στο Διαδίκτυο και των όποιων ιδεών μπορούν να εξαχθούν από αυτή τη πηγή δεδομένων και τέλος η ανάπτυξη και η βελτίωση των τεχνικών απεικόνισης και των διαδραστικών παγκόσμιου ιστού διεπαφών για τη παρακολούθηση και διερεύνηση διαταραχών όπως διακοπές λειτουργίας στο Διαδίκτυο.

Το δεύτερο επιστημονικό πρόγραμμα είχε σαν αντικείμενο την εύρεση και την εφαρμογή μεθόδων για την ανίχνευση και το χαρακτηρισμό της παρεμπόδισης κίνησης στο Διαδίκτυο με βάση τη BGP κλοπή (hijacking). Οι μέθοδοι ανίχνευσης πειρατείας-διαδρομής που κατα κύριο λόγο χρησιμοποιούνται συνδυάζουν παθητικές BGP μετρήσεις και ενεργές μετρήσεις (όπως traceroutes), δεδομένου ότι ο μηχανισμός ενεργοποίησης της επίθεσης λειτουργεί στο διατομεακής δρομολόγησης επίπεδο ελέγχου, αλλά η πραγματική επίπτωση (impact) είναι επαληθεύσιμη μόνο στο επίπεδο των δεδομένων. Με γνώμονα τα παραπάνω άρχισε η ανάπτυξη και η αξιολόγηση νέων μεθόδων και η επέκταση των υποδομών μέτρησης για τη υποστήριξή τους.

2.1.3 Επιστημονικό Πρόγραμμα στο Τομέα των Οικονομικών

Τέλος επιστημονικό πρόγραμμα εκπονήθηκε και στο τομέα των οικονομικών και της πολιτικής. Στον οικονομικό τομέα υπήρξε στόχευση στη κατανόηση της δομής και των δυναμικών (dynamics) του οικοσυστήματος του Διαδικτύου από οικονομικής άποψης, συλλαμβάνοντας σημαντικές αλληλεπιδράσεις ανάμεσα σε δικτυακές οικονομικές σχέσεις, ενδοδικτυακή τοπολογία, πολιτικές δρομολόγησης και ως αποτέλεσμα διατομεακών ροών κίνησης. Από τη πλευρά της πολιτικής έγινε προσπάθεια για ανταπόκριση σε αιτήματα από κρατικούς φορείς και φορείς χάραξης πολιτικής για παρατηρήσεις και θέσεις οι οποίες εμπειρικά πληροφορούν διαμάχες στη βιομηχανία και τηλεπικοινωνιακές πολιτικές. Ενώ τέλος έγινε προσπάθεια για παροχή τεχνογνωσίας για δεοντολογικά ζητήματα που αφορούν την έρευνα της τεχνολογίας των πληροφοριών και των επικοινωνιών.

2.1.4 Έρευνες για το Μέλλον του Διαδικτύου

Ο CAIDA στη προσπάθειά του να ερευνήσει το μέλλον του Διαδικτύου επικεντρώθηκε σε δύο κυρίως επιστημονικά προγράμματα. Συμμετείχε στο ερευνητικό πρόγραμμα της μελλοντικής Διαδικτυακής αρχιτεκτονικής «Named Data Networking Next Phase (NDN-NP)».[8]. Καθώς το Διαδίκτυο έχει υπερβεί κατά πολύ τις προσδοκίες που υπήρχαν, έχει επίσης υπερβεί και τις αρχικές παραδοχές, δημιουργώντας συχνά διαμάχες που αμφισβητούν το υποκείμενο μοντέλο επικοινωνίας. Οι χρήστες και οι εφαρμογές λειτουργούν ως πρὸς το περιεχόμενο, καθιστώντας ολόένα και πιο περιοριστικό και δύσκολο να συμμορφώνονται με την απαίτηση του IP για να επικοινωνούν ανακαλύπτοντας και καθορίζοντας τη τοποθεσία. Για να περάσει το Διαδίκτυο στο μέλλον, απαιτείται μία εννοιολογικά απλή αλλά μετασχηματική αρχιτεκτονική μεταβολή, από τη σημερινή εστίαση στο πού –διευθύνσεις και εξυπηρετητές- στο τί – το περιεχόμενο που οι χρήστες και οι εφαρμογές ενδιαφέρονται. Το συγκεκριμένο πρόγραμμα στοχεύει στην ανάπτυξη μίας νέας Διαδικτυακής αρχιτεκτονικής που μπορεί να κεφαλαιοποιήσει τις δυνάμεις και να

αντιμετωπίσει τις αδυναμίες της σημερινής βασισμένης στο δρομολογητή, από σημείο σε σημείο αρχιτεκτονική επικοινωνίας του Διαδικτύου, προκειμένου να εξυπηρετήσει φυσικά τις αναδυόμενες μορφές επικοινωνίας.

Με την ονοματοδοσία των δεδομένων αντί των τοποθεσιών τους, το NDN καθιστά τα δεδομένα σε κύρια οντότητα. Το σημερινό Διαδίκτυο ασφαλίζει το χώρο μέσα στον οποίο βρίσκονται τα δεδομένα. Το NDN ασφαλίζει τα περιεχόμενα, μια σχεδιαστική επιλογή που αποσυνδέει την εμπιστοσύνη στα δεδομένα από την εμπιστοσύνη στους εξυπηρετητές, η οποία επιτρέπει πολλαπλούς ριζικά κλιμακούμενους μηχανισμούς επικοινωνίας όπως η αυτόματη προσωρινή αποθήκευση για τη βελτιστοποίηση του εύρους ζώνης. Επιπλέον το πρόγραμμα αυτό μελετά τις τεχνικές προκλήσεις που πρέπει να αντιμετωπιστούν για την επικύρωση του ως μία μελλοντική αρχιτεκτονική του Διαδικτύου:

- Επεκτασιμότητα της δρομολόγησης
- Ασφάλεια του δικτύου
- Γρήγορη προώθηση
- Προστασία του περιεχομένου και της ιδιωτικότητας
- Θεμελιώδη θεωρία της επικοινωνίας
- Μοντέλα εμπιστοσύνης

Τέλος το NDN χρησιμοποιεί από άκρη σε άκρη ανάπτυξη πεδίου δοκιμών, προσομοίωση και θεωρητική ανάλυση για την αξιολόγηση της προτεινόμενης αρχιτεκτονικής και αναπτύσσει τις προδιαγραφές και τα πρότυπα εφαρμογής των NDN πρωτοκόλλων και εφαρμογών. Η συμμετοχή του CAIDA σε αυτό το πρόγραμμα περιλάμβανε συμμετοχή στο πεδίο δοκιμών, έρευνα σε σχετικά νέα NDN-συμβατά πρωτόκολλα δρομολόγησης και συνολική διοικητική υποστήριξη.

Το δεύτερο ερευνητικό πρόγραμμα στο οποίο συμμετείχε ο CAIDA ήταν το πρόγραμμα εξερεύνησης της εξέλιξης του IPv6 στους τομείς της τοπολογίας, της απόδοσης και της κυκλοφορίας [9]. Ο κύριος στόχος του συγκεκριμένου προγράμματος είναι να βελτιωθεί η ακρίβεια, το πεδίο εφαρμογής και η χρησιμότητα της τεχνολογίας μέτρησης του IPv6:

- Σχεδιάζοντας ξεπερασμένες μετρήσεις για προσαρμοστική και ευφυής ανίχνευση
- Δημιουργώντας εργαλεία για τη μέτρηση των χαρακτηριστικών της υιοθέτησης του IPv6 στην ακμή
- Συγκρίνοντας τα χαρακτηριστικά της IPv4 και IPv6 συνδεσιμότητας

Τα σύνολα δεδομένων που προκύπτουν ενημερώνουν το έργο, με το συσχετισμό αυτών των παρατηρήσεων με άλλα τεχνικά και κοινωνικοοικονομικά στοιχεία:

1. Κατανομή διευθύνσεων
2. Διαθέσιμα γεωγραφικά και κυκλοφοριακά δεδομένα
3. Οργανωτική δομή του ISP (κυβερνητική, εμπορική, εκπαιδευτική)
4. Πολιτικοί/ρυθμιστικοί παράγοντες που επηρεάζουν την ανάπτυξη του IPv6

Τέλος, τα αυστηρά εμπειρικά δεδομένα που συλλέγονται θα συμβάλλουν στη βελτίωση της κατάστασης των ποσοτικών μοντέλων της μετάβασης στο IPv6.

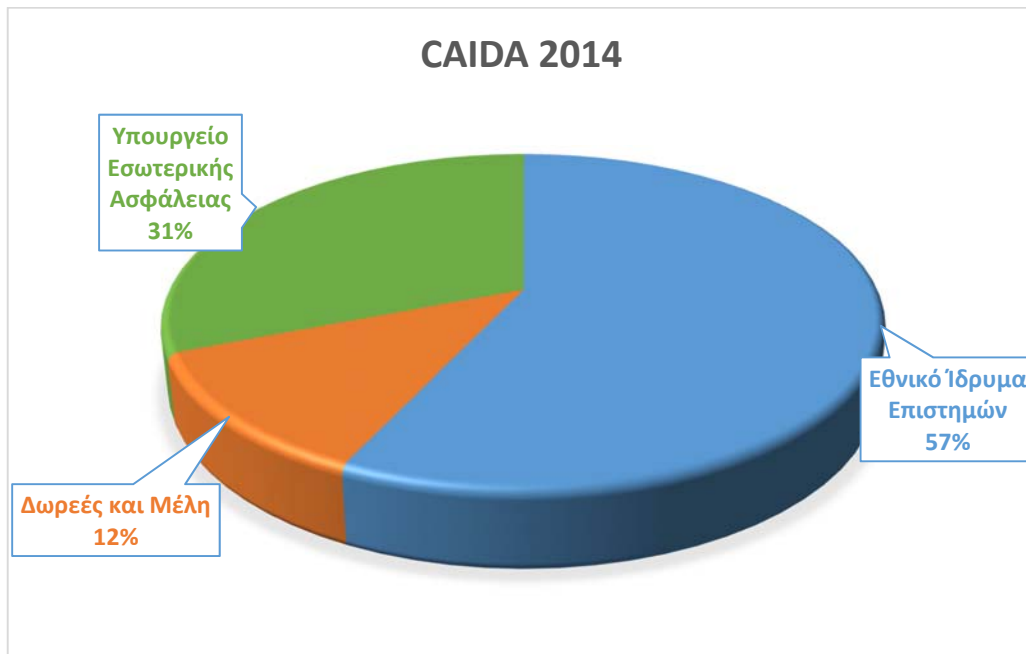
2.2 Ο CAIDA σε Αριθμούς

Οι ερευνητές του CAIDA το 2014 δημοσίευσαν 24 ομότιμες δημοσιεύσεις και 4 τεχνικές αναφορές και παρουσίασαν τα αποτελέσματα και τα πορίσματά τους στο IMC (Βανκούβερ,

Καναδάς), UbiComp (Σιάτλ, Ουάσινγκτον), HCI International (Κρήτη, Ελλάδα), στην Ευρυζωνικού Διαδικτύου Τεχνική Συμβουλευτική Ομάδα (Broadband Internet Technical Advisory Group) (BITAG) (Ντένβερ, Κολοράντο), PAM (Λος Άντζελες, Καλιφόρνια) και NANOG (Ατλάντα, Τζόρτζια) όπως επίσης και σε διάφορα άλλα σεμινάρια, προγράμματα και ερευνητικά συνέδρια. Ο CAIDA οργάνωσε και φιλοξένησε 6 συνέδρια [6]:

- AIMS 2014: Ημερίδα με θέμα τις ενεργές Διαδικτυακές μετρήσεις
- CREDS II: Κυβερνο-ασφάλεια, διάλογος για την ερευνητική δεοντολογία και ημερίδα στρατηγικής, μια μικρή ημερίδα στη τοπολογία, το BGP και το traceroute των δεδομένων
- NDNcomm 2014: Συνεδρίαση της NDN κοινότητας (φιλοξενία από κοινού με το UCLA)
- IMAPS: Βλάβες Δικτύου
- WIE 2014: 5^η Ημερίδα για τα οικονομικά του Διαδικτύου

Το 2014 η ιστοσελίδα του CAIDA www.caida.org προσέλκυσε 328.110 μοναδικούς επισκέπτες με μέσο όρο 2,38 επισκέψεις ανά επισκέπτη. Μέχρι το τέλος του Δεκεμβρίου 2014, στον CAIDA απασχολούνταν 17 υπάλληλοι, 4 επισκέπτες ερευνητές, 3 μεταδιδακτορικοί φοιτητές, 2 μεταπτυχιακοί φοιτητές και 10 προπτυχιακοί φοιτητές. Ο CAIDA έλαβε \$3.40M για την υποστήριξη των ερευνητικών του δραστηριοτήτων από τις ακόλουθες πηγές:



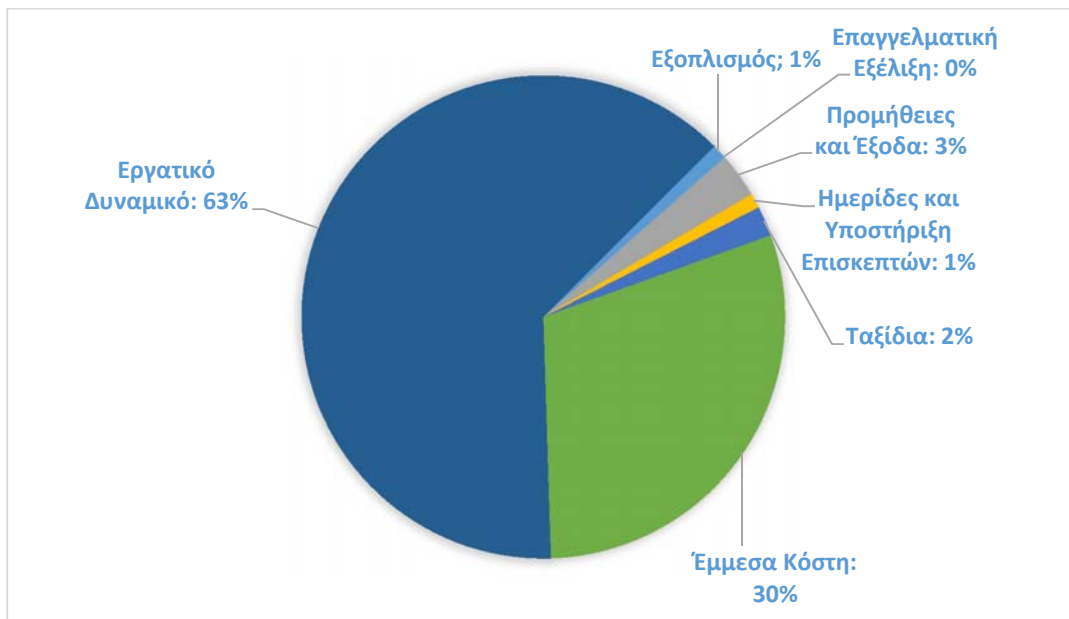
Σχήμα 1: Κατανομή ανά πηγή χρηματοδότησης

Πίνακας 1:Ακριβές ποσά χρηματοδότησης

Πηγή Χρηματοδότησης	Ποσό(\$)	Ποσοστό
Εθνικό Ίδρυμα Επιστημών	\$1,930,353	57%
Υπουργείο Εθνικής Ασφάλειας	\$1,060,698	31%
Δωρεές και Μέλη	\$407,430	12%
Σύνολο	\$3,398,481	100%

Από τα παραπάνω μπορεί να εξαχθεί με ασφάλεια το συμπέρασμα ότι κύριοι χορηγοί του CAIDA αποτελούν οι κρατικοί φορείς σε ποσοστό 88%.

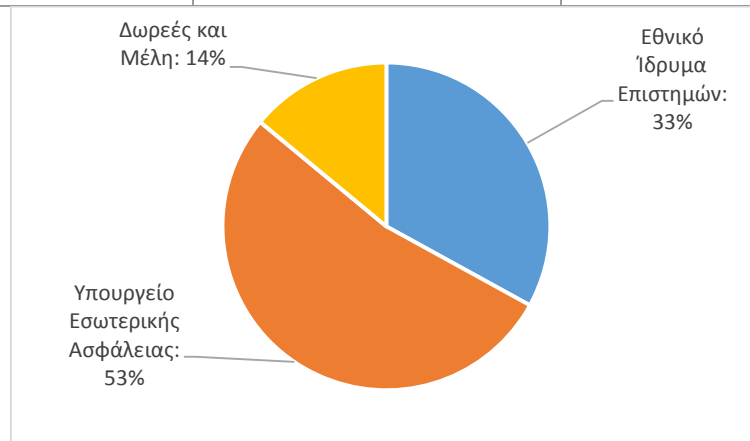
Στα επόμενα σχήματα και πίνακες παρουσιάζονται τα έξοδα του CAIDA, με βάση τη πηγή χρηματοδότησης (funding source), το τύπο των λειτουργικών εξόδων (operating expenses) και τη περιοχή του προγράμματος:



Σχήμα 2: Λειτουργικά έξοδα

Πίνακας 2: Λειτουργικά έξοδα(ποσά)

Είδος Εξόδου	Ποσό(\$)	Ποσοστό
Εργατικό Δυναμικό	\$2,008,054	63%
Έμμεσα Κόστη	\$952,240	30%
Επαγγελματική Εξέλιξη	\$11,306	0%
Προμήθειες και Έξοδα	\$86,887	3%
Ημερίδες και Υποστήριξη Επισκεπτών	\$25,682	1%
Ταξίδια	\$51,870	2%
Εξοπλισμός	\$50,074	2%
Σύνολο	\$3,186,114	100%

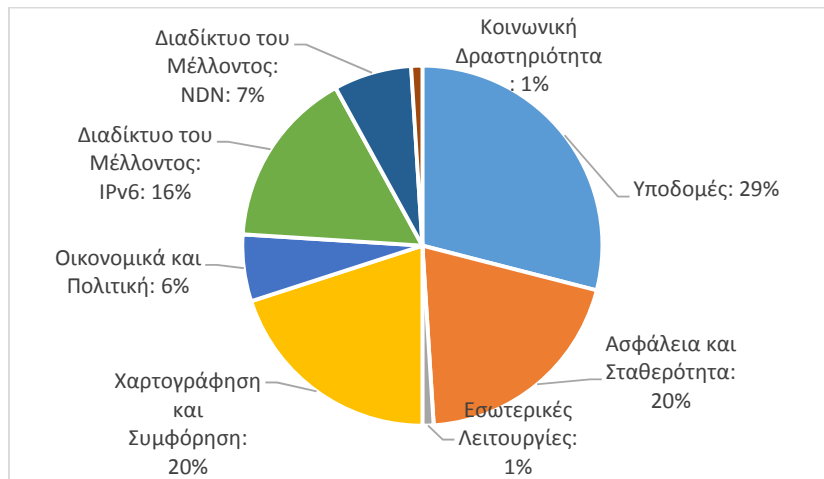


Σχήμα 3: Έξοδα ανα πηγή χρηματοδότησης

Πίνακας 3: Έξοδα ανά πηγή(ποσά)

Πηγή Χρηματοδότησης	Ποσό(\$)	Ποσοστό
Εθνικό Ίδρυμα Επιστημών	\$1,064,358	33%
Υπουργείο Εσωτερικής Ασφάλειας	\$1,672,058	52%
Δωρεές και Μέλη	\$449,698	14%
Σύνολο	\$3,186,114	100%

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida



Σχήμα 4: Έξοδα ανα περιοχή προγράμματος

Πίνακας 4: Έξοδα ανα περιοχή προγράμματος(ποσά)

Περιοχή Προγράμματος	Ποσό(\$)	Ποσοστό
Οικονομικά και Πολιτική	\$193,343	6%
Διαδίκτυο του Μέλλοντος: IPv6	\$504,456	16%
Διαδίκτυο του Μέλλοντος: NDN	\$229,724	7%
Χαρτογράφηση και Συμφόρηση	\$618,889	19%
Υποδομές	\$929,713	29%
Ασφάλεια και Σταθερότητα	\$647,606	20%
Κοινωνική Δραστηριότητα	\$38,823	1%
Εσωτερικές Λειτουργίες	\$23,560	1%
Σύνολο	\$3,186,114	100%

2.3 Δεδομένα του CAIDA

2.3.1 Στατιστικά της Συλλογής των Δεδομένων

Ο CAIDA το 2014 συνέλεξε τα παρακάτω ανεπεξέργαστα (raw) δεδομένα:

- Traceroutes διερεύνησης του φάσματος των IPv4 διευθύνσεων που συλλέχθηκαν από την υποδομή του Ark (922.3 GiB (συμπίεσμένα) / 2.7 TiB (μη συμπίεσμένα)), και τη κίνηση από την αντίστροφη αναζήτηση DNS για την ανακάλυψη IPv4 διευθύνσεων (45.0 GiB / 168.9 GiB).

- Traceroutes διερεύνησης του φάσματος των IPv6 διευθύνσεων που συλλέχθηκαν από ένα υποσύνολο των IPv6-ενεργοποιημένων Ark συστημάτων παρακολούθησης (17.4 GiB / 89.3 GiB).
- Ίχνη παθητικής κίνησης από το equinix-chicago και equinix-sanjose συστήματα παρακολούθησης συνδεδεμένα με το Tier-1 ISP δίκτυο κορμού συνδεδεμένο στις Equinix εγκαταστάσεις στο Σικάγο (Ιλινόι) και Σαν Χοσέ (Καλιφόρνια) (633.4 GiB / 1.5 TiB).
- Κρυφά ίχνη παθητικής κίνησης που συλλέχθηκαν από το UCSD τηλεσκόπιο δικτύου (84.2 TiB / 204.3 TiB).

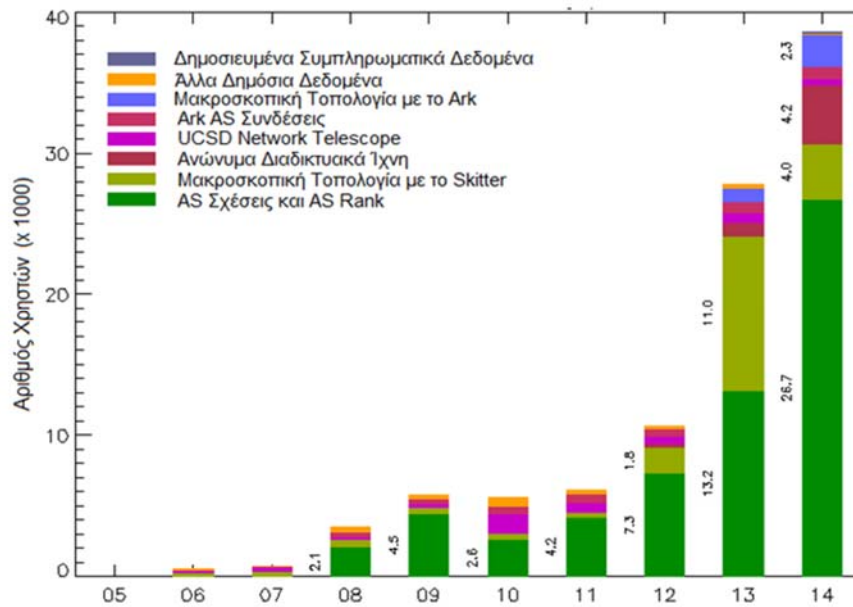
Από τα παραπάνω ανεπεξέργαστα προέκυψαν τα ακόλουθα σύνολα δεδομένων:

- IPv4 Routed /24 Topology
- IPv4 Routed /24 DNS Names (includes DNS traffic)
- IPv4 Routed /24 AS Links
- Macroscopic Internet Topology Data Kits (ITDKs): ITDK-2014-04 and ITDK-2014-12
- IPv6 Topology
- The IPv6 DNS Names
- IPv6 AS Links
- Anonymized Internet Traces 2014

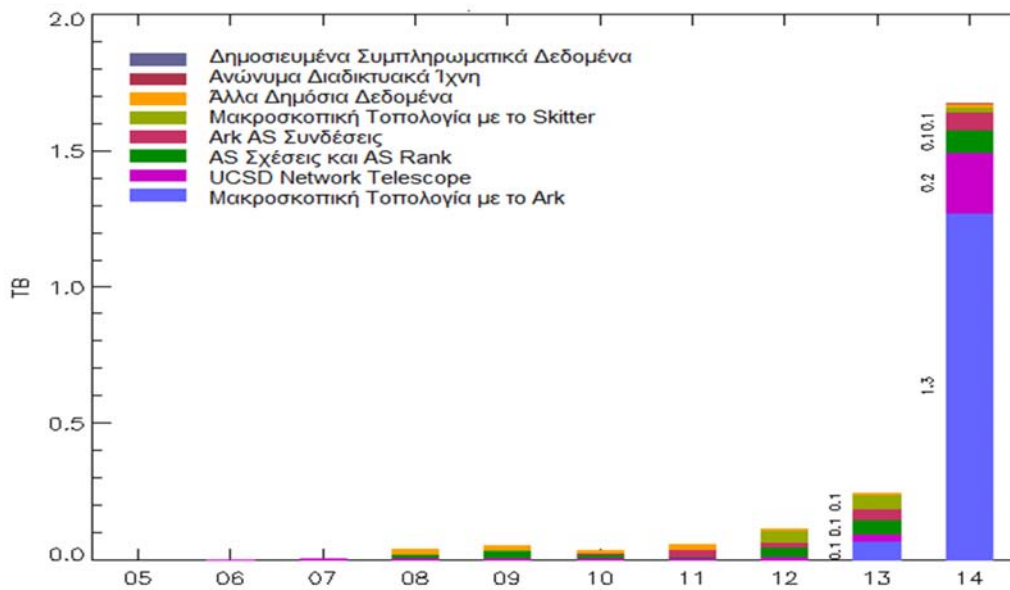
Για να αυξηθεί η δημοτικότητα και η χρήση των συνόλων των δεδομένων από την ερευνητική κοινότητα, από τον Φεβρουάριο του 2014, τα προγενέστερα των δύο ετών δεδομένα τοπολογίας του IPv4 από το Ark, όλα τα δεδομένα τοπολογίας, ονομάτων dns και συνδέσεων AS του IPv6 από το Ark και όλα τα ανεπεξέργαστα ίχνη του skitter έγιναν δημοσίως διαθέσιμα. Μόνο η πρόσβαση στα δεδομένα τοπολογίας του IPv4 από το Ark των τελευταίων 2 ετών παραμένει περιορισμένη. Η πρόσβαση στα σχεδόν πραγματικού χρόνου πακέτα δεδομένων του διαδικτυακού τηλεσκοπίου (Network Telescope) του CAIDA έγινε εφικτή για διάφορους ακαδημαϊκούς ερευνητές. Η ομάδα του CAIDA διατήρησε την AS Rank ιστοσελίδα και τα σχετικά πακέτα δεδομένων των συσχετισμών των AS που χρησιμοποιούν BGP δεδομένα και τους εμπειρικούς αλγόριθμους ανάλυσης που έχουν αναπτυχθεί από τους ερευνητές του CAIDA για να εξάγουν συμπεράσματα για τις σχέσεις μεταξύ των ASes.

2.3.2 Στατιστικά Κατανομής των Δεδομένων

- Τα Δημοσίως Διαθέσιμα πακέτα δεδομένων, η χρήση των οποίων προϋποθέτει την αποδοχή των όρων χρήσης των δεδομένων του CAIDA, αλλά κατά τα άλλα διατίθενται ελεύθερα. Ο πίνακας που ακολουθεί αναφέρει τον αριθμό των μοναδικών επισκεπτών και το συνολικό ποσό των δεδομένων που έχουν ληφθεί το 2014.

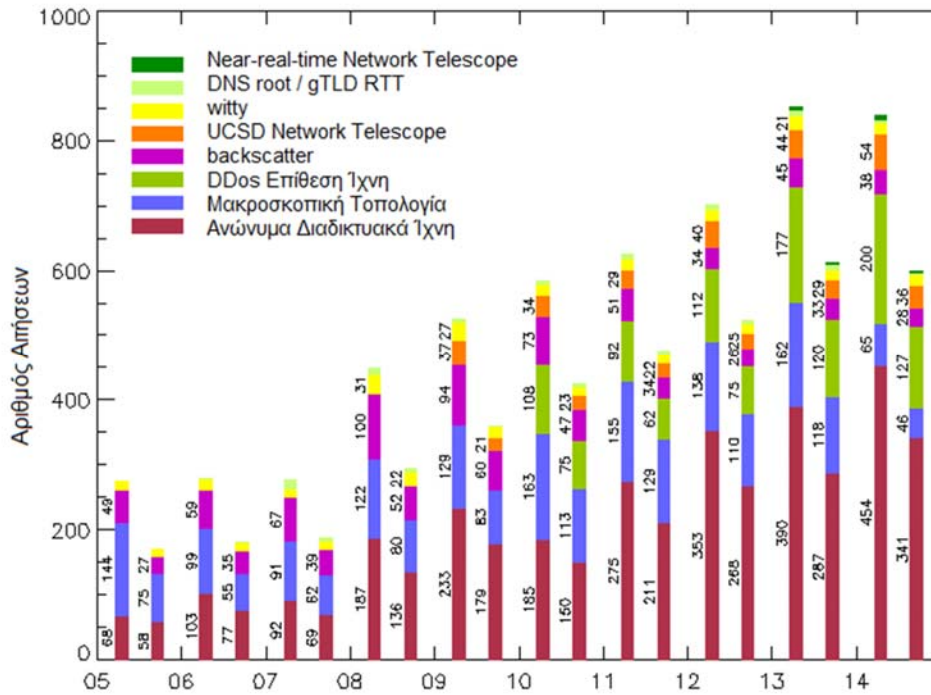


Σχήμα 5: Αριθμός των χρηστών που κατεβάζουν δημόσια δεδομένα

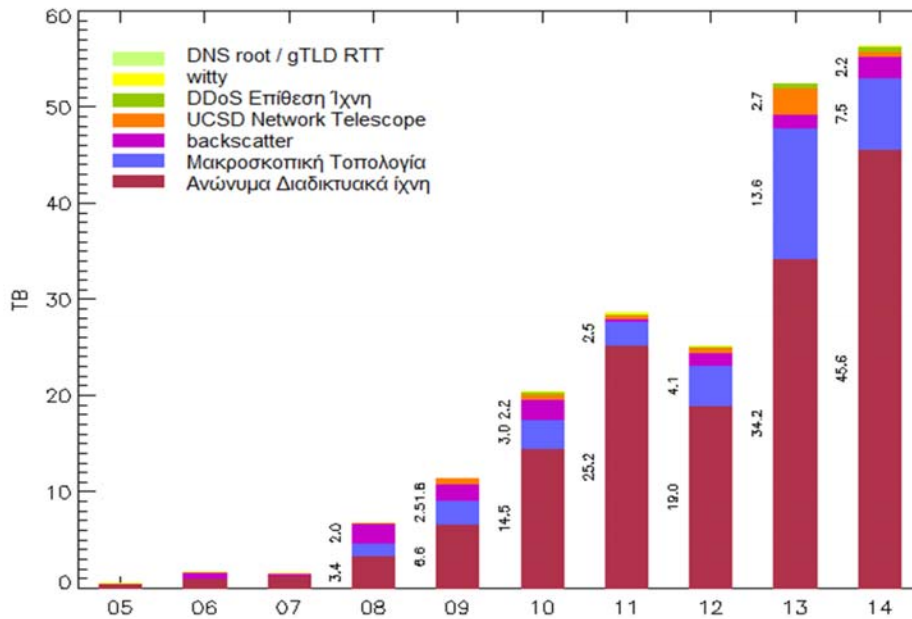


Σχήμα 6: Πλήθος των δημόσιων δεδομένων που έχουν ληφθεί

- Τα περιορισμένης πρόσβασης πακέτα δεδομένων απαιτούν οι χρήστες να είναι ακαδημαϊκοί ή κυβερνητικοί ερευνητές, ή χορηγοί του CAIDA, να κάνουν αίτηση για παροχή λογαριασμού, να δώσουν μια σύντομη περιγραφή της προβλεπόμενης χρήσης των δεδομένων και να συμφωνήσουν με τη πολιτική χρήσης.

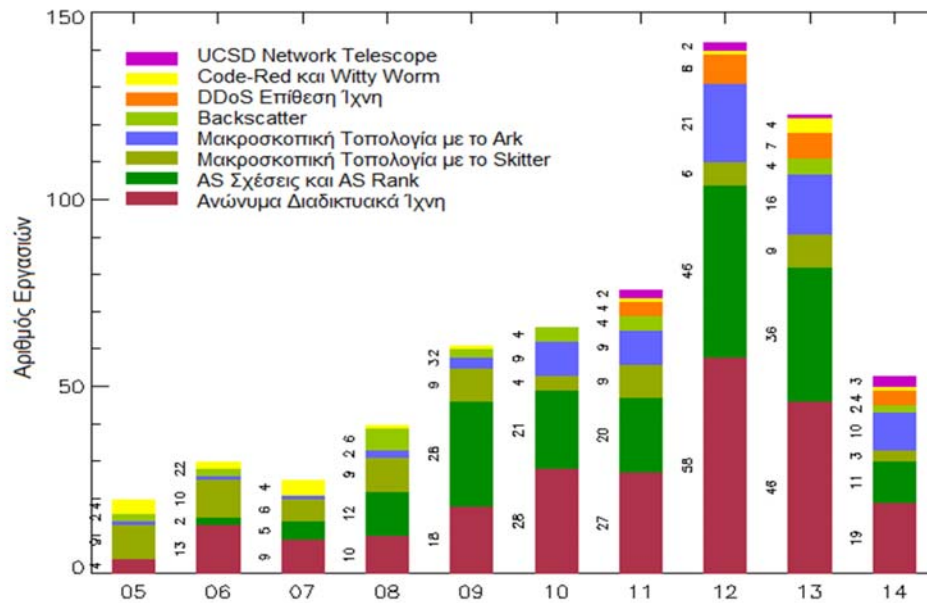


Σχήμα 7: Ληφθείσες/εγκεκριμένες αιτήσεις για περιορισμένης πρόσβασης πακέτα δεδομένων



Σχήμα 8: Πλήθος των περιορισμένης πρόσβασης δεδομένων που έχουν ληφθεί

- Δημοσιεύσεις που χρησιμοποίησαν δημόσια και/ή περιορισμένης πρόσβασης δεδομένα CAIDA (από μη-CAIDA συγγραφείς).



Σχήμα 9: Αριθμός των δημοσιευμένων εργασιών

2.4 Archipelago (Ark)

Το Archipelago (Ark) είναι μία υποδομή ενεργής μέτρησης του CAIDA που εξυπηρετεί τη δικτυακή ερευνητική κοινότητα από το 2007 διαδεχόμενη την υποδομή Skitter. Αποτελεί μία παγκόσμια κατανομημένη (distributed) πλατφόρμα μετρήσεων που έχει σαν στόχο την αποδέσμευση της ανάπτυξης εργαλείων μετρήσεων και εποπτείας του Διαδικτύου, από θέματα ασφάλειας, κατανομημένης λειτουργίας, συλλογής δεδομένων (dataset) κ.ά.

Καίρια σημεία της αρχιτεκτονικής της υποδομής Ark, είναι η αντικατάσταση του εργαλείου μετρήσεων Skitter από το Scamper και η υλοποίηση μιας κατανομημένης διαμοιραζόμενης συσχετιστικής μνήμης για το συντονισμό των διαφόρων εργαλείων μέτρησης. Το Scamper αποτελεί ένα ευέλικτο και ισχυρό εργαλείο μετρήσεων, που υποστηρίζει τα προγράμματα ping και traceroute όπως και τα πρωτόκολλα IPv4 και IPv6. Επιπλέον, δύναται να πραγματοποιήσει μετρήσεις βασισμένες στα πρωτόκολλα ICMP, TCP, UDP και περιλαμβάνει διαφοροποιημένες εκδόσεις του προγράμματος Paris traceroute, για την εύρεση IP μονοπατιών όταν προκύπτουν περιπτώσεις κατανομής δικτυακού φόρτου (load balancing).

Ο CAIDA κάνει συνεχή προσπάθεια να αναπτύξει την υποδομή του μέσω της διανομής εξοπλισμού μέτρησης (2^{ης} γενιάς Raspberry Pi) με όσο το δυνατόν μεγαλύτερη τοπολογική και γεωγραφική ποικιλία ώστε να βελτιωθεί η εικόνα μας για το παγκόσμιο Διαδίκτυο. Οι κύριοι στόχοι του CAIDA με την Ark υποδομή είναι να [10]:

- Μειώσει την προσπάθεια που απαιτείται για την εξέλιξη και την ανάπτυξη εξελιγμένων μεγάλης κλίμακας μετρήσεων.

- Παρέχει ένα βήμα προς τη κατεύθυνση μιας προσανατολισμένης προς τη κοινότητα υποδομής επιτρέποντας στους συνεργάτες να τρέχουν τις δικές τους εγκεκριμένες μετρήσεις σε μία ανθεκτική ως προς την ασφάλεια κατανομημένη πλατφόρμα.

Το Ark είναι ειδικά προσαρμοσμένο για την ενεργό μέτρηση του δικτύου. Το γεγονός αυτό επιτρέπει στο Ark να είναι πιο απλό από μερικές άλλες γενικού σκοπού (general-purpose) κατανομημένες πειραματικές πλατφόρμες και αυτό επιτρέπει στον CAIDA να επικεντρωθεί στη παροχή λειτουργιών που καλύπτουν άμεσα τις ανάγκες της δικτυακής έρευνας. Ειδικότερα, ο CAIDA παρέχει μια λειτουργία για την επικοινωνία και το συντονισμό που καθιστά ευκολότερο να γραφούν κατανομημένες μετρήσεις που πρέπει να υπολογιστούν από κοινού για την επίτευξη ενός στόχου. Ο CAIDA εργάζεται για τη παροχή ενός υψηλού επιπέδου API για να διευκολύνει τις προκλήσεις της γραφής εργαλείων μέτρησης. Στόχος του CAIDA είναι να χαμηλώσει το φράγμα φέρνοντας καινοτόμες και ενδιαφέρουσες τεχνικές μέτρησης σε λειτουργία.

2.4.1 Είδη Μετρήσεων

Οι μετρήσεις που γίνονται από τον CAIDA ομαδοποιούνται σε διαρκείς μετρήσεις, ειδικές μετρήσεις και φιλοξενούμενες (hosted) μετρήσεις.

2.4.1.1 Φιλοξενούμενες Μετρήσεις

1. The Spoofer Project: Σε συνεργασία με το NPS και το MIT, τα συστήματα παρακολούθησης του Ark συμμετέχουν στο Spoofer Project [11] βοηθώντας στη μέτρηση της ευαισθησίας (susceptibility) του Διαδικτύου σε πακέτα από πλαστογραφημένη (spoofed) IP διεύθυνση πηγής. Τα συστήματα παρακολούθησης συλλέγουν δεδομένα σχετικά με τη πλαστογραφημένη IP λαμβάνοντας τη πιθανώς πλαστογραφημένη κίνηση και προωθώντας τη στον εξυπηρετητή του Spoofer Project στο MIT για ανάλυση.
2. IPv4 και IPv6 σταθερότητα: Σε συνεργασία με ερευνητές στο Simula Research Laboratory, ο CAIDA μελετά τη σταθερότητα και τις επιδόσεις του IPv4 και IPv6: Από διπλή στοίβα Ark συστημάτων παρακολούθησης, ο CAIDA κάνει μετρήσεις (υψηλής συχνότητας rings και traceroutes πάνω στο IPv4 και IPv6) προς διπλής στοίβας (dual-stack) εξυπηρετητές από τη λίστα Alexa. Ο στόχος είναι η σύγκριση της προσβασιμότητας και των επιδόσεων (από την άποψη του RTT) στους διπλής στοίβας στόχους στο IPv4 και IPv6.
3. TCP Behavior Inference (Tbit): Αξιολογώντας την ανάπτυξη των TCP αλγορίθμων είναι ζωτικό για τη κατανόηση της ικανότητας του TCP να εκτελεί. Ο CAIDA χρησιμοποιεί την υποδομή του Ark για να συμπεράνει την ανάπτυξη των TCP αλγορίθμων και τα χαρακτηριστικά του στο σύγχρονο διαδίκτυο χρησιμοποιώντας μία προσέγγιση βασισμένη στο TCP Behavior Inference Tool (TBIT), με μία τρέχουσα εστίαση στην ανάπτυξη των αργής εκκίνησης συμπεριφορών, όπως επίσης και τη σκοπιμότητα (feasibility) των τυφλού (blind) παραθύρου TCP επιθέσεων.
4. Υγεία του Domain Name System (DNS): Δουλεύοντας με ερευνητές στα Verisign Labs, ο CAIDA χρησιμοποιεί κόμβους από το Ark έργο για να καθιερώσει τακτική διαγνωστική παρακολούθηση από ποικίλες προοπτικές, να καθιερώσει μια βασική γραμμή συμπεριφοράς απόκρισης και να ποσοτικοποιήσει τα τρέχοντα θέματα σύνδεσης όπως επίσης και εκείνα που μπορεί να προκύψουν μετά από ενδεχόμενη ανατροπή. Οι κόμβοι του Ark δεν παρέχουν μόνο μία διαφορετική διαδρομή δειγματοληψίας (sampling) για

την επίτευξη των στόχων αλλά είναι επίσης κατάλληλα για τη ταχεία και ευέλικτη ανάπτυξη, καθώς επιτρέπουν τη συνέχεια των υπάρχοντων εργαλείων, από τα επανασχεδιασμένα ή τα εξαρτόμενα από τη πλατφόρμα, τα οποία αποτελούν φραγμούς σε παρόμοιες πλατφόρμες.

Από κάθε κόμβο του Ark ο CAIDA περιοδικά εκδίδει μια σειρά από διαγνωστικά αιτήματα (diagnostic queries) για κάθε έναν από τους εξυπηρετητές στους οποίους έχει ανατεθεί ένας TLD χώρος ονομάτων (namespace). Τα αιτήματα περιλαμβάνουν δοκιμές επιπέδου μεταφοράς για να εντοπίζονται προβλήματα συνδεσιμότητας στο TCP και στο UDP, PMTU δοκιμές οριοθέτησης για την αναγνώριση εξυπηρετητών επηρεασμένων από ένα μικρότερο PMTU παρά από το ωφέλιμο φορτίο που αυτοί προσπαθούν να στείλουν, δοκιμές συνοχής της έκδοσης για τον εντοπισμό παλαιών δεδομένων από εξυπηρετητές που είναι εκτός συγχρονισμού και DNSSEC δοκιμές ορθότητας και συνεκτικότητας (consistency). Τα διαγνωστικά αιτήματα εκτελούνται τρεις φορές την ημέρα και τα αποτελέσματα συγκεντρώνονται και αποθηκεύονται για στοχευμένη (ταυτόχρονη συνοχή) και προσωρινή ανάλυση (αλλαγές συμπεριφοράς στη πάροδο του χρόνου).

2.4.1.2 Διαρκείς Μετρήσεις

Οι διαρκείς μετρήσεις έχουν σα στόχο να παρέχουν έγκαιρη, τακτική υποβολή δεδομένων εκτελώντας διάφορες κατηγορίες μετρήσεων σε συνεχή βάση καθοδηγούμενες (driven) από την αποστολή του CAIDA που περιλαμβάνει μεταξύ άλλων τη παροχή μακροσκοπικής γνώσης σχετικά με την υποδομή του Διαδικτύου, τη συμπεριφορά του, τη χρήση και την εξέλιξή του.

1. Ανακάλυψη Τοπολογίας Διαδικτύου: Χρησιμοποιώντας πολλαπλές ομάδες από γεωγραφικά κατανομημένα Ark συστήματα παρακολούθησης ο CAIDA δυναμικά και στρατηγικά διαιρεί την ερευνητική εργασία ανάμεσα στις ομάδες για να διενεργούν συντονισμένες, μεγάλης κλίμακας βασισμένες στο traceroute τοπολογικές μετρήσεις. Υποστηριζόμενο από το πρόγραμμα «Χαρτογραφικές Δυνατότητες για Κρίσιμη Κυβερνο-υποδομή», [12] ο CAIDA ενσωματώνει αυτές τις μετρήσεις με δυνατότητες ανάλυσης δεδομένων για τη παροχή διεξοδικά σχολιασμένων χαρτών Διαδικτυακής τοπολογίας που θα βελτιώσουν τις ικανότητές του να εντοπίζει, παρακολουθεί και προβάλλει μια κρίσιμη κυβερνο-υποδομή.
 - IPv4: Η παραλληλοποίηση (parallelization) του Ark επιτρέπει στον CAIDA να επιτύχει μια traceroute μέτρηση σε όλα τα δρομολογημένα /24 δίκτυα στο IPv4 χώρο διευθύνσεων σε περίπου 2-3 μέρες για μία ομάδα από 17-18 συστήματα παρακολούθησης που διερευνούν πάνω από 10 εκατομμύρια /24 (αυτό είναι, ο πλήρης δρομολογήσιμος χώρος διευθύνσεων υποδιαιρεμένος στα /24) στα 100rps. Ο CAIDA αυτή τη στιγμή έχει τρεις ομάδες ενεργές και κάθε ομάδα ερευνά αυτόνομα. Οι μετρήσεις αυτές είναι διαθέσιμες από τον CAIDA για κατέβασμα ως «The Ark IPv4 Routed /24 Topology Dataset».
 - IPv6: Για κάθε διερευνημένο μονοπάτι, ο CAIDA συλλέγει την IP διεύθυνση, το RTT, την απόκριση TTL και τις ICMP αποκρίσεις για όλα τα hops (τα τμήματα της διαδρομής μεταξύ πηγής και προορισμού), συμπεριλαμβανομένων των ενδιάμεσων hops. Κάθε Ark σύστημα παρακολούθησης διερευνά όλα τα ανακοινωμένα IPv6 προθέματα (prefixes) (/48 ή μικρότερα) ένα κάθε 48 ώρες. Ένα διερευνητικό

πέρασμα από όλα τα ανακοινωμένα προθέματα ονομάζεται κύκλος. Σε κάθε κύκλο, ένα σύστημα παρακολούθησης διερευνά μόνο ένα τυχαίο προορισμό για κάθε πρόθεμα. Διαφορετικά συστήματα παρακολούθησης διερευνούν προθέματα σε ανεξάρτητα-επιλεγμένες τυχαίες σειρές και διερευνούν ένα ανεξάρτητα-επιλεγμένο τυχαίο προορισμό για κάθε πρόθεμα. Τα προθέματα είναι τυχαία διατεταγμένα με τέτοιο τρόπο όπου ένα συγκεκριμένο σύστημα παρακολούθησης ποτέ δεν διερευνά το ίδιο πρόθεμα εντός 16 ωρών πέρα από τα όρια του κύκλου (ένα σύστημα παρακολούθησης δεν μπορεί ποτέ να ξαναερευνήσει ένα πρόθεμα μέσα στον ίδιο κύκλο, εξ ορισμού). Οι μετρήσεις αυτές είναι διαθέσιμες από τον CAIDA για κατέβασμα ως «The Ark IPv6 Topology Dataset».

Η ερευνητική ομάδα πειραματίζεται εκτελώντας traceroute μετρήσεις χρησιμοποιώντας το scamper ένα ισχυρό και ευέλικτο εργαλείο [13] που έχει σχεδιαστεί για να εξετάζει ενεργά προορισμούς στο Διαδίκτυο παράλληλα (σε καθορισμένο ρυθμό πακέτων ανά δευτερόλεπτο) έτσι ώστε μαζικά δεδομένα να μπορούν να συλλέγονται εγκαίρως. Τα αρχεία εξόδου του scamper ονομάζονται warts: ένα αρχείο warts περιέχει σημαντικά μεταδεδομένα (metadata) που περιβάλλουν κάθε μεμονωμένη μέτρηση που διεξήχθη, καθώς και σημαντικές λεπτομέρειες των απαντήσεων που ελήφθησαν. Οι μετρήσεις που διενεργήθηκαν μπορούν να κυμαίνονται από απλές σε σύνθετες. Ένα παράδειγμα μιάς απλής μέτρησης είναι όταν μια απλή μέθοδος μέτρησης (π.χ traceroute) χρησιμοποιεί μία λίστα από IP διευθύνσεις για να διενεργήσει μια μαζική μέτρηση. Μία περισσότερο σύνθετη μέτρηση μπορεί να είναι όταν το αποτέλεσμα της προηγούμενης δοκιμής επηρεάζει το τί θα συμβεί στη συνέχεια: για παράδειγμα, για κάθε hop σε ένα μονοπάτι traceroute, να συναγάγει τη διεύθυνση της εξερχόμενης διεπαφής από το προηγούμενο hop. Οι σύνθετες μετρήσεις διεξάγονται με τη σύνδεση σε μία διαδικασία του scamper που βρίσκεται σε εξέλιξη με ένα πρόγραμμα οδήγησης που περιέχει τη λογική. Το scamper υποστηρίζει τα πρωτόκολλα IPv4 και IPv6, καθώς και τα προγράμματα ping και traceroute. Επιπλέον, μπορεί να πραγματοποιήσει μετρήσεις βασισμένες στα πρωτόκολλα ICMP, TCP, UDP και περιλαμβάνει διαφοροποιημένες εκδόσεις του προγράμματος Paris traceroute, για την εύρεση IP μονοπατιών όταν ανακύπτουν περιπτώσεις ισορροπίας φόρτου.

2. Συμφόρηση: Όπως περιγράφεται στο πρόγραμμα «Χαρτογράφηση Διασύνδεσης στο Διαδίκτυο: Συντοπισμός, Συνδεσιμότητα και Συμφόρηση» καθώς το παγκόσμιο Διαδίκτυο επεκτείνεται για να ικανοποιήσει τις απαιτήσεις και τις προσδοκίες ενός συνεχώς αυξανόμενου τμήματος του παγκόσμιου πληθυσμού [7], βαθιές αλλαγές επέρχονται στη δομή της διασύνδεσης, στις δυναμικές της κυκλοφορίας και στην οικονομική και πολιτική δύναμη διαφορετικών παραγόντων του οικοσυστήματος. Αυτές οι αλλαγές δεν αφορούν μόνο τη μηχανική του δικτύου και τις λειτουργίες του, αλλά επίσης παρουσιάζουν ευρύτερες προκλήσεις για επένδυση σε νέες τεχνολογίες, για το μελλοντικό σχεδιασμό του δικτύου, για δημόσια πολιτική και επιστημονική μελέτη του ίδιου του Διαδικτύου. Και όμως και από επιστημονική και από πολιτική προοπτική, η εξέλιξη του οικοσυστήματος είναι σε μεγάλο βαθμό ανεξερεύνητο έδαφος.

Ο CAIDA θα εστιάσει τη προσοχή του σε δύο σχετικούς μετασχηματισμούς του οικοσυστήματος: την ανάδυση των IXes ως σημεία αναφοράς στο πλέγμα της διασύνδεσης και τον αυξανόμενο ρόλο των παρόχων περιεχομένου και των Δικτύων Διανομής Περιεχομένου (CDNs) ως σημαντικές πηγές της κυκλοφορίας που ρέει στο Διαδίκτυο. Με κάποιους υπολογισμούς πάνω από το μισό του όγκου της κυκλοφορίας στη Βόρεια Αμερική έρχεται αυτή τη στιγμή από μόλις δύο διανομές περιεχομένου

(Youtube και Netflix). Η μετατόπιση αυτή αποτελεί την άνοδο ενός νέου είδους ιεραρχίας στο οικοσύστημα, φέρνοντας ουσιαστικά νέους περιορισμούς (constraints) στους υπάρχοντες παίκτες που πρέπει να διαχειριστούν τη κυκλοφορία στα δίκτυά τους για να ελαχιστοποιήσουν τη συμφόρηση. Οι αποδείξεις του προβλήματος έχουν αυξηθεί δραματικά τα τελευταία 5 χρόνια, με αποτέλεσμα διαμάχες μεταξύ των εμπορικών παραγόντων, καθώς και μεταξύ του ιδιωτικού τομέα και των ρυθμιστικών φορέων, σε βάρος των χρηστών που βιώνουν εμπειρία υποβαθμισμένης ποιότητας.

Η προτεινόμενη έρευνα διαρθρώνεται σε δύο θεμελιώδεις εργασίες και μια σειρά από ερευνητικά ερωτήματα που βασίζονται σε αυτές τις εργασίες. Κατ' αρχάς, ο CAIDA θα κατασκευάσει ένα νέου τύπου σημασιολογικά πλούσιο Διαδικτυακό χάρτη, ο οποίος θα αποσαφηνίσει το ρόλο των IXes στην ισχυρά διευκολυμένη και γεωγραφικά ποικίλα αλλά σύνθετη διατομεακή συνδεσιμότητα. Ο χάρτης αυτός θα καθοδηγήσει τη δεύτερη εργασία του CAIDA: μία μελέτη για τον υπολογισμό της δυναμικής της κυκλοφοριακής συμφόρησης που προκαλείται από τις εξελισσόμενης ανταλλαγής κίνησης και διαχείρισης της κυκλοφορίας πρακτικές των CDNs και ISPs. Το CAIDA θα χρησιμοποιήσει την υποδομή μετρήσεων του όπως και τις μετρήσεις από τέσσερις βιομηχανικούς συνεργάτες: Akamai, Google, Comcast και Netflix. Τέλος ο CAIDA θα χρησιμοποιήσει τα αποτελέσματα από αυτές τις δύο εργασίες για τη διερεύνηση ερωτημάτων σχετικών με την ανθεκτικότητα των υποδομών, την επιστημονική μοντελοποίηση, τα οικονομικά του δικτύου και τη δημόσια πολιτική. Οι μέθοδοι που χρησιμοποιούνται είναι:

- Ping με χρονική ακολουθία (TSP) που περιλαμβάνει την αποστολή μίας ακολουθίας από pings κατά μήκος της διαδρομής σε ερώτημα, ψάχνοντας για μία ημερήσια (diurnal) διακύμανση στη καθυστέρηση που ενδεχομένως υποδηλώνει συμφόρηση σε ένα σύνδεσμο.
- Χαρτογράφηση των συνόρων χρησιμοποιώντας τις οπτικές γωνίες του Ark για την έρευνα και την ανάπτυξη τεχνικών μέτρησης για τη συναγωγή με ακρίβεια της παρουσίας των διατομεακών συνδέσμων για το δίκτυο που φιλοξενεί το Ark. Η προσέγγιση του CAIDA συνδυάζει την εμπειρία των μελών του στην ανακάλυψη Διαδικτυακής κλίμακας τοπολογίας και επίλυσης ψευδωνύμου, καθώς και αλγορίθμων των μελών του CAIDA για τη συναγωγή σχέσεων δρομολόγησης ανάμεσα στα δίκτυα, για να αντιστοιχίσει με ακρίβεια τους διατομεακούς δρομολογητές με τους ιδιοκτήτες τους.

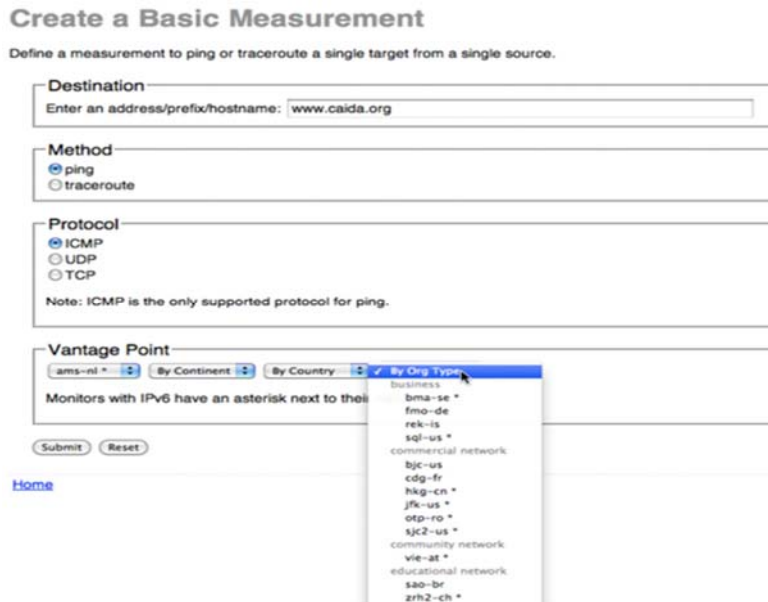
2.4.1.3 Ειδικές Μετρήσεις

Εκτός των διαρκών μετρήσεων, οι ερευνητές μπορούν να εκτελέσουν ειδικού σκοπού μετρήσεις στα συστήματα παρακολούθησης του Ark μέσω είτε μίας διεπαφής γραμμής εντολών είτε μίας διεπαφής προγράμματος περιήγησης [10].

1. Ο tod-client (τοπολογία αν και όταν ζητηθεί) δίνει στους χρήστες που εργάζονται σε ένα κέλυφος (shell) γραμμής εντολών μία διεπαφή μεταβλητή για την εκτέλεση IPv4 και IPv6 traceroute και ping μετρήσεις.
2. Vela: Διεπαφή διαδικτύου: Η υπηρεσία Vela παρέχει πρόσβαση στη τοπολογία κατά παραγγελία λειτουργικότητα του Ark μέσω ενός προγράμματος περιήγησης Διαδικτύου. Οργανισμοί έχουν πρόσβαση στην Ark πλατφόρμα μέσω της Vela διεπαφής Διαδικτύου για να τρέχουν "εφάπαξ" μετρήσεις. Η διεπαφή επιτρέπει στους χρήστες να επιλέγουν

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

ένα υποσύνολο των συστημάτων παρακολούθησης (π.χ, όλα τα συστήματα παρακολούθησης στην Ευρώπη, ή ένα Ark σύστημα παρακολούθησης από κάθε ήπειρο με συνδεσιμότητα IPv6) χρησιμοποιώντας ping ή traceroute. Στην εικόνα 1 βλέπουμε ένα υπόδειγμα δημιουργίας μίας μέτρησης στο Vela, εισάγεται η διεύθυνση του προορισμού, επιλέγεται η μέθοδος που θα χρησιμοποιηθεί (ping ή traceroute), το πρωτόκολλο και το σύστημα παρακολούθησης και εκτελείται η μέτρηση



Create a Basic Measurement

Define a measurement to ping or traceroute a single target from a single source.

Destination
Enter an address/prefix/hostname:

Method
 ping
 traceroute

Protocol
 ICMP
 UDP
 TCP
Note: ICMP is the only supported protocol for ping.

Vantage Point
ams-nl * | By Continent | By Country | By Orig Type
Monitors with IPv6 have an asterisk next to their name

[Home](#)

- business
- bma-se *
- fmo-de
- rek-is
- sql-us *
- commercial network
- bjc-us
- cdg-fr
- hkg-cn *
- jfk-us *
- otp-ro *
- sjc2-us *
- community network
- vie-at *
- educational network
- sao-br
- zrh2-ch *

Εικόνα 1: Δημιουργία μέτρησης στο Vela

Στην εικόνα 2 μπορεί κάποιος να δει ένα υπόδειγμα εκτέλεσης μέτρησης στο Vela με χρήση της μεθόδου ping στο προορισμό www.google.com από όλα τα συστήματα παρακολούθησης IPv4 με χρήση του πρωτοκόλλου ICMP.

Measurement Results

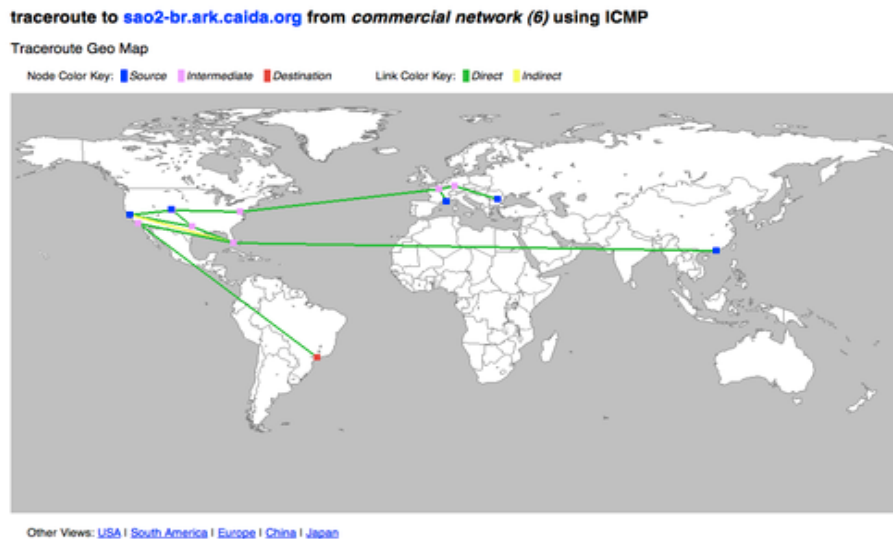
ping to www.google.com from All IPv4 (59) using ICMP

Source	Target	RTT (ms)
syd-au	74.125.237.115	1.0
sao-br	74.125.234.240	1.1
sao2-br	74.125.234.18	1.1
sjc2-us	74.125.224.82	1.2
zrh2-ch	173.194.35.18	5.2
sea-us	173.194.79.99	8.0
lej-de	173.194.69.103	9.3
bma-se	173.194.32.50	10.9
eug-us	173.194.79.104	13.3
bjc-us	74.125.227.17	15.7
cdg-fr	173.194.35.179	18.0
jfk-us	173.194.73.104	18.3
sql-us	173.194.79.103	19.6
scl-cl	173.194.42.18	20.7
nap-it	173.194.35.148	23.0
vie-at	173.194.35.148	24.0
pna-es	74.125.230.176	24.0
lax-us	173.194.79.106	26.9
rno-us	173.194.79.106	31.1
otp-ro	173.194.35.178	34.0
mtm-mx	74.125.227.147	34.6

Εικόνα 2: Αποτελέσματα μέτρησης με χρήση της μεθόδου Ping

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

Τέλος στην εικόνα 3 μπορεί κανείς να δει ένα υπόδειγμα γεωγραφικού χάρτη των traceroute μονοπατιών προς τη διεύθυνση προορισμού `sao2-br.ark.caida.org` με χρήση του πρωτοκόλλου ICMP.



Εικόνα 3: Γεωγραφικός χάρτης των traceroute μονοπατιών

Οι παρακάτω οργανισμοί έχουν εκτελέσει ειδικές μετρήσεις στο Ark:

- Department of Homeland Security (DHS S&T)
- Naval Postgraduate School (NPS)
- The Réseaux IP Européens Network Coordination Centre (RIPE NCC)
- Jacobs University, Bremen, Germany
- Eurocom, France
- Fraunhofer AISEC
- The University of Cape Town

3. ΤΟΠΟΛΟΓΙΚΗ ΕΡΕΥΝΑ

Ο CAIDA κάνει μετρήσεις, αναλύει, μοντελοποιεί και οπτικοποιεί τη Διαδικτυακή τοπολογία από το 1998. Επιδιώκει να χαρακτηρίζεται η μακροσκοπική Διαδικτυακή συνδεσιμότητα με χρήση και τοπολογικών και γεωγραφικών αναπαραστάσεων σε πολλαπλά επίπεδα λεπτομέρειας της συνάθροισης. Αυτός ο ερευνητικός στόχος είναι ιδιαίτερα απαιτητικός λόγω αντιφάσεων σε διάφορα εργαλεία μέτρησης και αντικινήτρων (counterincentives) για τους ISPs να υποστηρίξουν [14] εργαλεία που παράγουν συμπεράσματα σχετικά με τη συνδεσιμότητα. Αυτοί οι περιορισμοί καθιστούν δύσκολο και συχνά αδύνατο να επιβεβαιωθούν επιστημονικά συμπεράσματα σχετικά με την τοπολογία του Διαδικτύου σε σχέση με την πραγματικότητα. Και όμως, μια εμπειρικά θεμελιωμένη (grounded) κατανόηση του σχήματος, της δομής και της εξέλιξης του Διαδικτύου ήδη είχε σοβαρές συνέπειες για τη θεωρία και τη πρακτική της επιστήμης των δικτύων.

3.1 Επίλυση Ψευδωνύμων (Alias Resolution)

Το εργαλείο traceroute και οι παραλλαγές του χρησιμοποιούνται ευρέως από πολλές τοπολογικές ερευνητικές μελέτες για την ανακάλυψη των IP διεπαφών και πώς αυτές συνδέονται, συμπεριλαμβανομένου και των μακροσκοπικών τοπολογικών μετρήσεων του CAIDA που τρέχει το εργαλείο scamper στις υποδομές του Ark. Το traceroute λειτουργεί ανιχνεύοντας μια διεύθυνση προορισμού με μία σειρά από πακέτα με αυξανόμενες αρχικές τιμές στο πεδίο IP TTL (Time To Live). Όταν ένας IP δρομολογητής λαμβάνει ένα πακέτο που πρόκειται να διαβιβαστεί, αρχικά μειώνει το πεδίο TTL [15]. Εάν το καινούριο TTL είναι μηδενικό, ο δρομολογητής δεν προωθεί το πακέτο, αλλά αντιθέτως στέλνει ένα ICMP υπέρβασης χρόνου σφάλμα πίσω στον αποστολέα. Η διεύθυνση πηγής του ICMP μηνύματος αναγνωρίζει μία διεπαφή στο δρομολογητή που έστειλε το μήνυμα. Μια σειρά ανιχνευτών με αυξανόμενες αρχικές τιμές TTL θα αποκαλύψει κανονικά μια διεύθυνση σε κάθε κόμβο κατά μήκος της διαδρομής, με ορισμένες εξαιρέσεις. Η επανάληψη αυτής της διαδικασίας από πολλαπλές πηγές σε πολλαπλούς προορισμούς μπορεί να αποκαλύψει πολλές διευθύνσεις δρομολογητών και συνδέσμους μεταξύ τους.

Κάθε δρομολογητής εξ' ορισμού έχει τουλάχιστον δύο διεπαφές, οι δρομολογητές Διαδικτυακού κορμού συχνά έχουν δεκάδες. Επιπλέον, κάθε διεπαφή μπορεί να έχει πολλαπλές διευθύνσεις. Αλλά κάθε traceroute ανταπόκρισης αποκαλύπτει μόνο μια διεύθυνση ενός δρομολογητή. Η IP επίλυση ψευδωνύμων είναι η διαδικασία με την οποία γίνεται η αναγνώριση των διευθύνσεων IP που ανήκουν στον ίδιο δρομολογητή και αποτελεί ένα πολύ σημαντικό βήμα για τη δημιουργία τοπολογικών χαρτών του Διαδικτύου.

3.1.1 Εργαλεία Επίλυσης Ψευδωνύμων

Τα εργαλεία που χρησιμοποιούνται ως επί το πλείστον από τον CAIDA για την επίλυση ψευδωνύμων (alias resolution) είναι τα MIDAR, kapar και iffinder.

Iffinder

Το iffinder ανακαλύπτει διευθύνσεις IPv4 που ανήκουν στον ίδιο δρομολογητή χρησιμοποιώντας την τεχνική κοινής πηγής. Πιο συγκεκριμένα το iffinder στέλνει ένα ανιχνευτικό UDP πακέτο σε μια αχρησιμοποίητη θύρα σε μια διεύθυνση διεπαφής. Πολλοί δρομολογητές θα απαντήσουν σε ένα τέτοιο πακέτο με ένα σφάλμα ICMP PORT UNREACHABLE με τη διεύθυνση πηγής να ορίζεται σε εκείνη της διεπαφής, στη μετάδοση

unicast (αποστολή πλαισίων από ένα αποστολέα σε ένα και μόνο παραλήπτη δικτύου) επιστρέφει στον “ερευνητή”. Οπότε ερευνώντας μία διεπαφή και παίρνοντας αυτό το σφάλμα από μια διαφορετική διεπαφή αποτελεί μια ισχυρή ένδειξη ότι οι δύο διεπαφές ανήκουν στον ίδιο κόμβο του δικτύου [16].

Το iffinder μπορεί επίσης να ανακαλύψει νέες διεπαφές (interfaces) με διάφορους τρόπους. Όποτε το iffinder ανακαλύπτει μία νέα διεπαφή, προσθέτει τη διεπαφή στη λίστα των ερευνητικών στόχων [17].

- **IP Record Route:** Το iffinder μπορεί να χρησιμοποιήσει την επιλογή IP RECORD ROUTE στις ανιχνεύσεις του. Δεν υποστηρίζουν όλοι οι δρομολογητές αυτή την επιλογή, και είναι ικανή αυτή μόνο για την καταγραφή 9 διευθύνσεων κατά μήκος της διαδρομής, αλλά παρ’ όλα αυτά μπορεί να οδηγήσει στην ανακάλυψη πολλών νέων διεπαφών. Και δεδομένου ότι οι δρομολογητές που υποστηρίζουν το RECORD ROUTE θα καταγράψουν τη διεύθυνση της «ακραίας» (εξερχόμενης) διεπαφής στη διαδρομή του διερευνητικού πακέτου και ένα διερευνητικό PORT UNREACHABLE σε αυτή τη διεπαφή θα βρει συνήθως μια “κοντινότερη” διεπαφή του δρομολογητή, αυτό συχνά οδηγεί στην ανακάλυψη ενός ζεύγους διεπαφών που ανήκουν στον ίδιο δρομολογητή.
- **ICMP σφάλματα από ενδιαμέσους δρομολογητές:** Εάν η πηγή του σφάλματος ICMP ήταν προηγουμένως άγνωστη, το iffinder το αποθηκεύει. Επιπλέον, δεδομένου ότι τα ICMP πακέτα σφάλματος περιέχουν την IP επικεφαλίδα του αρχικού πακέτου που προκάλεσε το σφάλμα, το iffinder ανακτά τα δεδομένα RECORD ROUTE και από το πακέτο σφάλματος και από το ενσωματωμένο αρχικό πακέτο. Το αρχικό πακέτο περιέχει δεδομένα RECORD ROUTE που θα είχαν χαθεί διαφορετικά.
- **Η πύλη δίνεται σε ένα ICMP REDIRECT σφάλμα:** Τα ICMP REDIRECT σφάλματα περιέχουν μία προτιμώμενη διεύθυνση πύλης. Αν αυτή η διεύθυνση ήταν προηγουμένως άγνωστη, το iffinder την αποθηκεύει.
- **“/30 ταίρια”:** Συχνά μια σύνδεση μεταξύ δύο εσωτερικών δρομολογητών ορίζεται ως ένα /30 υποδίκτυο (subnet), οπότε οι διεπαφές με το άγνωστο άκρο της σύνδεσης έχουν τον ίδιο 30-bit αριθμό δικτύου και έναν 2-bit αριθμό εξυπηρετητή. Μία διεπαφή έχει τον αριθμό εξυπηρετητή 1, και η άλλη έχει τον αριθμό εξυπηρετητή 2. (Οι αριθμοί εξυπηρετητή 0 και 3 δεν ισχύουν). Με δεδομένη μια έγκυρη διεύθυνση εντός ενός /30 υποδικτύου, μπορεί κανείς να ορίσει “/30 ταίρι” ως την άλλη διεύθυνση μέσα στο /30 υποδίκτυο. Για κάθε γνωστή διεπαφή που είναι μια έγκυρη /30 διεύθυνση, αν το iffinder δεν έχει ήδη /30 ταίρι του στη λίστα, το iffinder θα διερευνήσει το /30 ταίρι του. Καθώς το iffinder μόνο μαντεύει την ύπαρξη της διεύθυνσης αυτής, αν η διεύθυνση δεν ανταποκρίνεται, το iffinder θα την απορρίψει (και ποτέ δεν θα εξεταστεί και πάλι, εκτός αν κάποιος άλλο μέσο αποδεικνύει ότι υπάρχει).
- **IP Traceroute:** Όταν ένας ενδιαμέσος κόμβος προωθεί ένα πακέτο που περιέχει την επιλογή IP Traceroute, θα πρέπει να στείλει ένα πακέτο ICMP TRACEROUTE που περιέχει τη δική του διεύθυνση πίσω στη πηγή. Ωστόσο, αρχικά πειράματα δείχνουν ότι η επιλογή αυτή πολύ σπάνια υποστηρίζεται και ότι ορισμένοι εξυπηρετητές θα απορρίψουν ακόμα τα πακέτα που περιέχουν αυτή την επιλογή. Επίσης, μειώνει κατά 3 τον αριθμό των διευθύνσεων που μπορούν να εγγραφούν στην επιλογή RECORD ROUTE. Χρησιμοποιώντας το επομένως θα ήταν πιθανό να γίνει το iffinder λιγότερο αποτελεσματικό, οπότε δεν χρησιμοποιείται αυτή η επιλογή στις περισσότερες διαδρομές.

Όταν χρησιμοποιείται το Record Route ή το IP Traceroute, ορισμένες ανιχνεύσεις οδηγούν σε μια ICMP_PARAMPROB απόκριση ή σε καμία απόκριση εντελώς, εάν ένας δρομολογητής είναι ελαττωματικός, ρυθμισμένος να αγνοεί το Record Route, ή οτιδήποτε άλλο. Σε τέτοιες περιπτώσεις, το iffinder θα ξανά προσπαθήσει την ανίχνευση χωρίς επιλογές IP, οδηγώντας συχνά σε μια χρήσιμη απόκριση.

Karar

Το εργαλείο karar είναι αποτέλεσμα της προσπάθειας των μελών του CAIDA για τη δημιουργία μιας ιδιαίτερα βελτιστοποιημένης εφαρμογής για ανάλυση που χρησιμοποιείται σε μεγάλης κλίμακας τοπολογίες Διαδικτύου, με βελτιστοποιημένους αλγόριθμους [18]. Το εργαλείο karar είναι εμπνευσμένο από το εργαλείο APAR. Το εργαλείο karar δεν αποθηκεύει το πλήρες σύνολο των διαδρομών στη μνήμη. Αντ' αυτού, το εργαλείο karar κάνει ένα μόνο πέρασμα πάνω από το σύνολο των ίχνων (traces) και εξάγει μόνο τις ελάχιστες πληροφορίες που χρειάζεται.

- Πρώτον, το karar βρίσκει όλα τα μοναδικά 3-κόμβων τμήματα για να χρησιμοποιηθούν για τη φάση επίλυσης ψευδωνύμων.
- Δεύτερον, εντοπίζει κοινά προθέματα μήκους 24 ή μεγαλύτερα ανάμεσα στις διευθύνσεις στο ίδιο ίχνος για να δημιουργήσει μια λίστα με υποδίκτυα που δεν μπορούν να υφίστανται σύμφωνα με τον όρο ακρίβειας υποδικτύου.
- Τέλος, αναθέτει ένα μοναδικό αριθμό ταυτότητας σε κάθε ίχνος, και αποθηκεύει μια λίστα όλων των παρατηρούμενων διευθύνσεων και ένα συμπιεσμένο χάρτη bit των αναγνωριστικών των ίχνων, στον οποίο εμφανίζεται κάθε διεύθυνση. Αυτά τα σύνολα αναγνωριστικών ίχνων περιλαμβάνουν επαρκείς πληροφορίες για τον έλεγχο της συνθήκης έλλειψης βρόχου.

Το karar βελτιώνει επίσης τον αλγόριθμο του APAR με διάφορους τρόπους:

- Πρώτον, μπορεί να φορτώσει ένα σύνολο από ψευδώνυμα που προέρχονται από άλλη πηγή, π.χ. τα αποτελέσματα μιας τεχνικής δακτυλικών αποτυπωμάτων (fingerprints) ή δημοσιευμένες τοπολογίες άμεσης παρατήρησης.
- Δεύτερον, κατά τη φάση σχηματισμού υποδικτύου, το karar προαιρετικά χρησιμοποιεί έναν αυστηρότερο έλεγχο για την ύπαρξη από σημείο σε σημείο υποδικτύου.
- Τρίτον, κατά τη διάρκεια της φάσης που συμπεραίνονται ψευδώνυμα, το karar χρησιμοποιεί αυστηρότερους ελέγχους για τη συνθήκη κοινού γείτονα.
- Τέλος, η εφαρμογή karar μπορεί να κάνει χρήση των δεδομένων TTL που προέρχονται από πολλαπλές οπτικές γωνίες, το οποίο περιορίζει τις φάσεις σχηματισμού υποδικτύου και επίλυσης ψευδωνύμου, μειώνοντας περαιτέρω το ποσοστό των ψευδώς θετικών στο καθένα.

MIDAR

Πολλοί δρομολογητές δημιουργούν IP τιμές ταυτότητας χρησιμοποιώντας έναν απλό μετρητή που είναι κοινός σε όλες τις διεπαφές του δρομολογητή. Αν κανείς στείλει διερευνητικά (probe) πακέτα σε δύο διευθύνσεις διεπαφής που ανήκουν στον ίδιο δρομολογητή, μπορεί να εντοπίσει ότι οι IP τιμές ταυτότητας στα πακέτα απόκρισης παρήχθησαν από ένα κοινόχρηστο μετρητή και έτσι ότι οι δύο διευθύνσεις είναι ψευδώνυμα. Το MIDAR χρησιμοποιεί αυτή τη βασική ιδέα για τον εντοπισμό ψευδωνύμων, αλλά

διαφέρει στον τρόπο που εντοπίζει τους κοινόχρηστους μετρητές. Το MIDAR επιτυγχάνει μεγαλύτερη ακρίβεια από ότι οι προηγούμενες τεχνικές που βασίζονται στην IP ταυτότητα και είναι το πρώτο που είναι πρακτικό για παραγωγική χρήση σε μεγάλης κλίμακας τοπολογίες Διαδικτύου [19], [20]. Το MIDAR, το iffinder και το karar συμπληρώνουν το ένα το άλλο, δίνοντας ένα πιο ακριβές και πλήρες αποτέλεσμα από κάθε ένα εργαλείο μόνο για την επίλυση ψευδωνύμων. Το MIDAR εισάγει αρκετές σημαντικές νέες συνεισφορές σε σχέση με τις προηγούμενες τεχνικές που βασίζονται στην IP ταυτότητα:

- **Τα μονότονα όρια ελέγχου (MBT) για ακριβή έλεγχο των κοινόχρηστων μετρητών:** Το MIDAR συγκρίνει χρονοσειρές των δειγμάτων IP ταυτοτήτων που συλλέγονται από πολλαπλές διεπαφές. Ο έλεγχος μονότονων ορίων του MIDAR ελέγχει αυστηρά για μια απαραίτητη προϋπόθεση: αν δύο χρονολογικές σειρές προέρχονται από ένα μόνο κοινόχρηστο μετρητή, τότε η ένωση αυτών των χρονολογικών σειρών πρέπει να είναι μονότονη. Για να αποφευχθεί η αβεβαιότητα στη συναγωγή του χρόνου που η απόκριση παράχθηκε από το δρομολογητή, το MIDAR χρησιμοποιεί το μετρήσιμο με ακρίβεια χρονικό εύρος μεταξύ της αποστολής του διερευνητικού και της λήψης της απόκρισης. Τα μονότονα όρια ελέγχου αποδίδουν σχεδόν μηδενικά ψευδώς αρνητικά αποτελέσματα για κοινόχρηστους μετρητές δεδομένου ότι ένας μετρητής πρέπει να είναι μονότονος εξ ορισμού και ένα πολύ χαμηλό ποσοστό ψευδώς θετικών για κοινόχρηστους μετρητές, δεδομένου ότι κρατά τα όρια τόσο αυστηρά όσο το δυνατόν περισσότερο.
- **Αλγόριθμος προγραμματισμού συρόμενου παραθύρου (Sliding window) για αύξηση της διερεύνησης:** Διερευνητικές διευθύνσεις παράλληλα εντός ενός χρονικού διαστήματος αρκετά μικρού για να δώσει χρήσιμες μετρήσεις γίνονται προβληματικές όταν ο αριθμός των διευθύνσεων φθάνει το ένα εκατομμύριο ή και περισσότερο. Για να αποφευχθεί αυτό, το MIDAR βασίζεται στο γεγονός ότι όταν οι διευθύνσεις μοιράζονται ένα μετρητή, οι IP ID χρονοσειρές τους πρέπει να έχουν παρόμοια ταχύτητα (ρυθμό μεταβολής ταυτότητας). Μετά τη συλλογή εκτιμήσεων ταχύτητας, το MIDAR διερευνά ένα συνεχώς μεταβαλλόμενο παράθυρο (υποσύνολο) στόχων, τέτοιο ώστε όλοι οι στόχοι με παρόμοια ταχύτητα θα διερευνούνται παράλληλα, αλλά οι στόχοι με ανόμοιες ταχύτητες μπορεί να μην υπάρχουν.
- **Τέσσερις ανεξάρτητες μέθοδοι ανίχνευσης:** Κατά το στάδιο της εκτίμησης της ταχύτητας, το MIDAR ανιχνεύει διευθύνσεις με τέσσερις διαφορετικές μεθόδους, και στη συνέχεια επιλέγει την καλύτερη μέθοδο για χρήση στα μεταγενέστερα στάδια:
 1. **TCP:** στέλνει TCP ACK στη θύρα 80 στο στόχο. Ο στόχος απαντά με TCP RST.
 2. **UDP:** στέλνει UDP σε αχρησιμοποίητη θύρα στο στόχο. Ο στόχος απαντά με ICMP Port Unreachable.
 3. **ICMP:** στέλνει ICMP Echo Request στο στόχο. Ο στόχος απαντά με ICMP Echo Reply.
 4. **Έμμεσα:** στέλνει ICMP Echo Request σε ένα εξυπηρετητή πέρα από το στόχο, με το TTL του ορισμένο να λήγει στο στόχο. Ο στόχος απαντά με ICMP Time Exceeded.

Η χρήση πολλαπλών μεθόδων επιτρέπει στο MIDAR να συλλέγει χρονοσειρές από στόχους που μπορεί να είναι αδρανείς σε μία συγκεκριμένη μέθοδο. Τα πειράματά δείχνουν ότι είναι συνηθισμένο (αν και όχι καθολικό) για δρομολογητές να χρησιμοποιούν ένα κοινόχρηστο μετρητή για τις αποκρίσεις σε όλες ή τις περισσότερες από αυτές τις μεθόδους, οπότε είναι αποδοτική η σύγκριση χρονοσειρών που συλλέγονται με διαφορετικές μεθόδους.

- **Ανίχνευση από πολλαπλά συστήματα παρακολούθησης:** Με ανίχνευση από πολλαπλές οπτικές γωνίες, το MIDAR μπορεί να επιτύχει ένα υψηλότερο ποσοστό ανίχνευσης συνολικά από ό,τι θα ήταν πρακτικά από ένα μόνο εξυπηρετητή και μπορεί ακόμα να αναθέσει στόχους σε συστήματα παρακολούθησης με έναν τρόπο που βελτιστοποιεί την απόκριση. Αυτό το χαρακτηριστικό εισάγει κάποια αμφιβολία στη μέτρηση του χρόνου δεδομένου ότι τα ρολόγια των συστημάτων παρακολούθησης δεν μπορούν ποτέ να συγχρονιστούν απόλυτα, αλλά μπορεί κανείς να το δικαιολογήσει αυτό με τη διεύρυνση της χρονικής κλίμακας στα μονότονα όρια ελέγχου, χωρίς να θυσιάζεται η αυστηρότητα του ελέγχου.
- **Επισημοποίηση πολλαπλών στάδιων ελέγχου για την εξάλειψη των ψευδώς θετικών:** Κατά τον έλεγχο N διευθύνσεων για ζεύγη που μοιράζονται ένα μετρητή, ο αριθμός των πραγματικών θετικών (ζεύγη που μοιράζονται ένα μετρητή) είναι ανάλογος του N , αλλά ο αριθμός των ψευδώς θετικών (ζεύγη διευθύνσεων που εμφανίζεται να μοιράζονται ένα μετρητή αλλά στην πραγματικότητα δεν το κάνουν) είναι ανάλογος του συνολικού αριθμού των πιθανών ζευγών, $O(N^2)$. Οπότε, όταν το N είναι μεγάλο, ακόμα και ένας έλεγχος που ενστικτωδώς δείχνει σαν ένα πολύ χαμηλό ποσοστό ψευδώς θετικών, θα δημιουργήσει ένα μεγάλο αριθμό ψευδώς θετικών σε σχέση με τον αριθμό των πραγματικών θετικών. Το MIDAR χρησιμοποιεί μια σειρά από διερευνητικά στάδια προσεκτικά σχεδιασμένα για να εξαλείψουν αποτελεσματικά τα ψευδώς θετικά και να αυξηθεί η εμπιστοσύνη στα αληθώς θετικά.

Μία πλήρης κλίμακας εκτέλεση του MIDAR αποτελείται από τέσσερα διερευνητικά στάδια:

- **Εκτίμηση:** για κάθε διεύθυνση, καθορίζεται η ταχύτητα και η καλύτερη μέθοδος διερεύνησης για χρήση σε επόμενα στάδια.
- **Ανακάλυψη:** εξετάζονται όλες οι διευθύνσεις στόχοι με πρόγραμμα συρόμενου παραθύρου για την αποτελεσματική εύρεση ζευγών διευθύνσεων που ενδεχομένως μοιράζονται ένα μετρητή.
- **Απόρριψη:** επανα-ερευνούνται τα ανακαλυπτόμενα πιθανά ζεύγη ψευδωνύμων για να αποκλειστούν περισσότερα ψευδώς θετικά.
- **Επιβεβαίωση:** διερευνάται κάθε υποψήφιο σύνολο ψευδωνύμων συνολικά για να αυξηθεί η εμπιστοσύνη στα πραγματικά θετικά και να αποκλειστούν τα ψευδώς θετικά που έχουν απομείνει. Το στάδιο επιβεβαίωσης μπορεί επίσης να χρησιμοποιηθεί αυτόνομα για τον έλεγχο των συνόλων ψευδωνύμων που ανακαλύφθηκαν από άλλη τεχνική επίλυσης ψευδωνύμων.

3.2 Χαρτογράφηση με Χρήση του plot-latlong

Το εργαλείο του CAIDA plot-latlong είναι ένα γραμμής εντολών εργαλείο, γραμμένο στη γλώσσα προγραμματισμού Perl, για τη χάραξη σημείων σε γεωγραφικούς χάρτες δοθέντος μιας λίστας από γεωγραφικού πλάτους / γεωγραφικού μήκους ζεύγη. Αυτό έχει σα στόχο καταστάσεις στις οποίες [21]:

- μια μέτρια ποσότητα ορθότητας και ακρίβειας είναι επαρκής
- ένας μεγάλος αριθμός τοποθεσιών θα πρέπει να σχεδιαστεί (δεκάδες χιλιάδες τοποθεσίες μπορούν εύκολα να αντιμετωπιστούν)
- η χάραξη πρέπει να είναι αυτοματοποιημένη (από ένα σενάριο κελυφους, για παράδειγμα)

- ένα ελαφρύ (lightweight) εργαλείο (και μικρό και με λίγες εξαρτήσεις) που απλά να σχεδιάζει σημεία είναι επαρκές
- η ευκολία της τροποποίησης είναι σημαντική (έτσι ώστε ειδικές απαιτήσεις να μπορούν να ικανοποιηθούν)

Το plot-latlong μπορεί να χειριστεί μη γραμμικές (linear) προβολές χάρτη (προς το παρόν τη Alber/Lambert προβολή) και είναι σκόπιμα μινιμαλιστικό, έτσι ώστε να μπορεί να χρησιμεύσει ως ένα δομικό στοιχείο. Οι χρήστες μπορούν να το αξιοποιήσουν με τρεις τρόπους:

1. να προσθέσουν νέους χάρτες
2. να εκτελέσουν τις εικόνες εξόδου μέσω πακέτων όπως το NetPBM για να προσθέσουν τίτλους, κ.λπ.
3. να τροποποιήσουν το πηγαίο κώδικα για να αλλάξουν το τρόπο με τον οποίο τα σημεία ζωγραφίζονται κ.λ.π.

Το plot-latlong μπορεί επίσης να χρησιμοποιηθεί για να υπολογιστεί απλά η χαρτογράφηση από το γεωγραφικό πλάτος / μήκος σε συντεταγμένες εικονοστοιχείου (για ένα δεδομένο χάρτη). Αυτές οι συντεταγμένες εικονοστοιχείου μπορούν στη συνέχεια να τροφοδοτηθούν σε άλλα προγράμματα για να ζωγραφιστούν πιο αναλυτικές εικόνες.

Για το σκοπό της εκτέλεσης του plot-latlong με είσοδο μεγάλο πλήθος γεωγραφικών συντεταγμένων έγινε λήψη των μακροσκοπικής τοπολογίας Διαδικτύου συνόλων δεδομένων του CAIDA, ITDK 2010-07 και ITDK 2013-07. Τα ITDK περιέχουν δεδομένα σχετικά με τη συνδεσιμότητα και τη δρομολόγηση που συγκεντρώθηκαν από μία μεγάλη διατομή του παγκόσμιου Διαδικτύου. Τα παραπάνω ITDKs αποτελούνται μεταξύ άλλων από δύο σχετικές τοπολογίες επιπέδου δρομολογητή και τις γεωγραφικές συντεταγμένες του κάθε δρομολογητή [22]. Οι δύο συμπεριλαμβανόμενες τοπολογίες επιπέδου δρομολογητή παράγονται από την ίδια τοπολογία IP επιπέδου, αλλά διαφέρουν ως προς την ακρίβεια και την πληρότητα της επίλυσης ψευδωνύμων που εκτελέστηκε για τη δημιουργία τους. Η πρώτη τοπολογία προέρχεται από επίλυση ψευδωνύμων με τα εργαλεία MIDAR και iffindex, τα οποία αποφέρουν τα υψηλότερης εμπιστοσύνης ψευδώνυμα, με ελάχιστα ψευδώς θετικά. Η δεύτερη τοπολογία χρησιμοποιεί επίσης το MIDAR και το iffindex αλλά περιλαμβάνει επιπλέον ψευδώνυμα που έχουν επιλυθεί με το εργαλείο karar, το οποίο αυξάνει σημαντικά την κάλυψη των ψευδωνύμων αλλά με το κόστος των λανθασμένων θετικών (που μεγαλώνει το μέγεθος των δρομολογητών και μειώνει τον αριθμό δρομολογητών).

Κάνοντας κάποιες αλλαγές στο κώδικα perl του plot-latlong καθίσταται εφικτή η απεικόνιση των γεωγραφικών συντεταγμένων των δρομολογητών που βρίσκονται στα πακέτα karar-midar-iff.nodes.geo και midar-iff.nodes.geo των ITDKs. Για κάθε πακέτο στα πλαίσια αυτής της μελέτης γίνεται απεικόνιση των δρομολογητών στο παγκόσμιο χάρτη και μέσω κάποιων τροποποιήσεων στο χάρτη της Ελλάδας. Ο χάρτης της Ελλάδας δεν περιλαμβάνεται στους χάρτες που προϋπάρχουν στον φάκελο mapimages κατά τη λήψη του plot-latlong. Για την ενσωμάτωση του χάρτη της Ελλάδας απαιτούνται τα παρακάτω βήματα:

1. Εύρεση κατάλληλου χάρτη (αρχείου .png).
2. Εισαγωγή του χάρτη στο φάκελο mapimages
3. Αναφορά του χάρτη στο αρχείο mapinfo. Όπου αναφέρεται το όνομα κλήσης του χάρτη από το εργαλείο plot-latlong και οι μέγιστες και ελάχιστες γεωγραφικές συντεταγμένες σημείων που ανήκουν σε αυτό το χάρτη.

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

```
razer9@ubuntu: ~/Desktop/plot-latlong-0.32
GNU nano 2.2.6 File: .mapinfo
MAP Greece Hellas_test.png 41.733333 19.383333 34.816667 29.641667
```

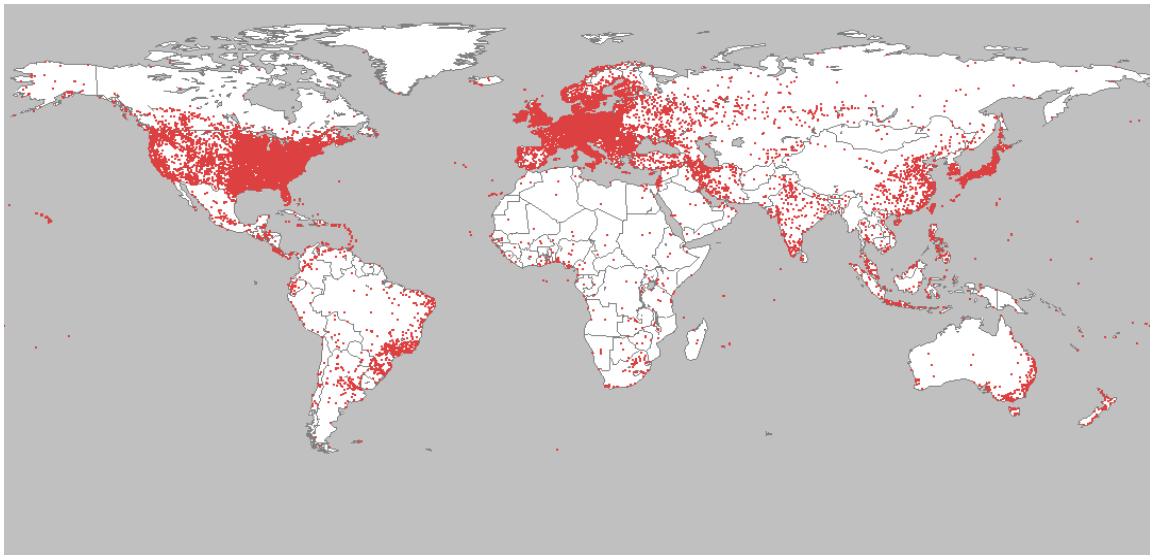
Εικόνα 4: Αναφορά του χάρτη της Ελλάδας στο αρχείο mapinfo

Με την ενσωμάτωση του χάρτη της Ελλάδας στους διαθέσιμους χάρτες είναι πλέον εφικτή η απεικόνιση των αποτελεσμάτων της εκτέλεσης του plot-latlong και στους δύο χάρτες που έχουν επιλεγεί για αυτή τη μελέτη.

Αρχικά τρέχουμε το εργαλείο plot-latlong με είσοδο το αρχείο midar-iff.nodes2010.geo και παίρνουμε τα ακόλουθα αποτελέσματα

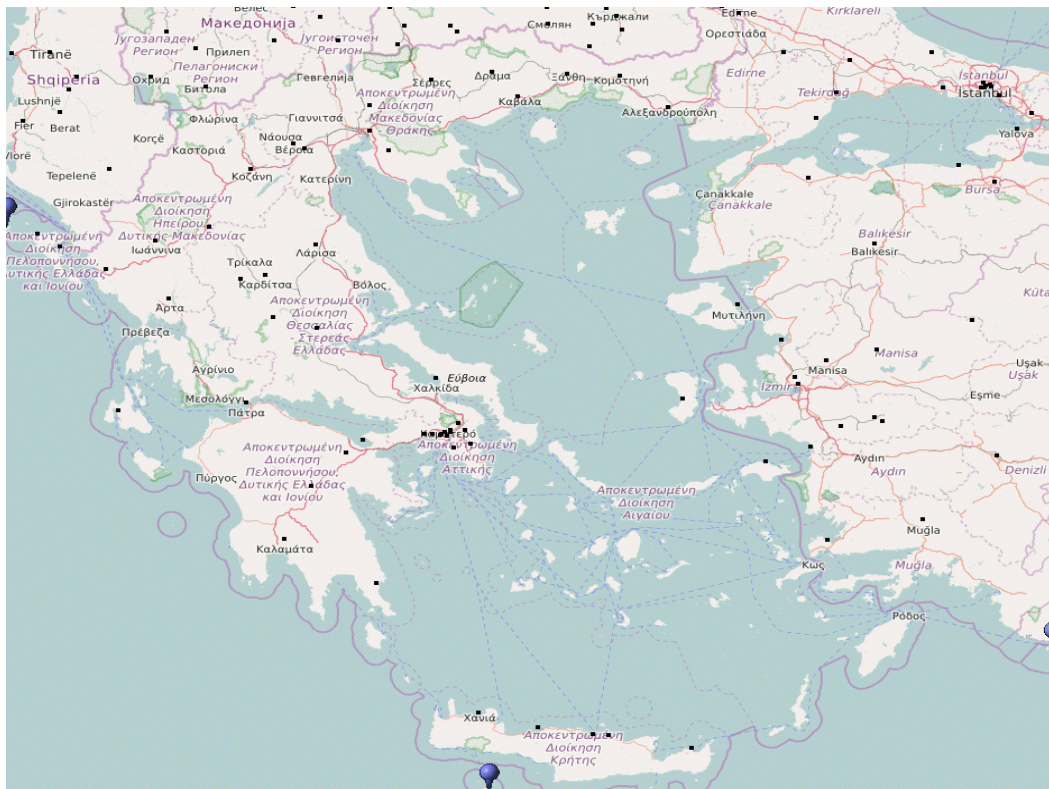
```
razer9@ubuntu: ~/Desktop/plot-latlong-0.32
razer9@ubuntu:~/Desktop/plot-latlong-0.32$ cat midar-iff.nodes2010.geo | perl plot-latlong -s 4 >Greece.png
```

Εικόνα 5: Εκτέλεση του plot-latlong με είσοδο το midar-iff.nodes2010.geo



Εικόνα 6: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το midar.iff.nodes2010.geo

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

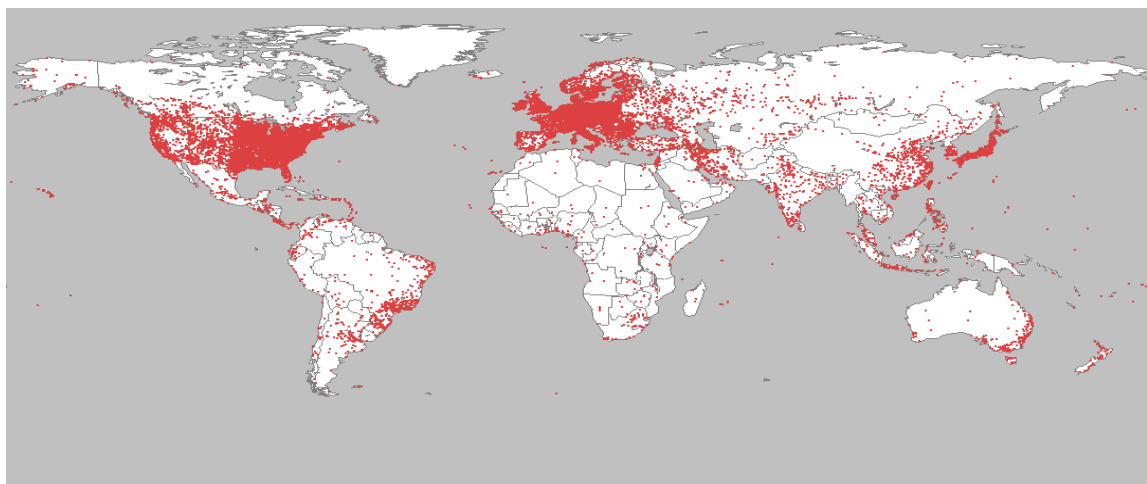


Εικόνα 7: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το midar.iff.nodes2010.geo

Στη συνέχεια τρέχουμε το εργαλείο plot-latlong με είσοδο το αρχείο kapar-midar-iff.nodes2010.geo και παίρνουμε τα ακόλουθα αποτελέσματα

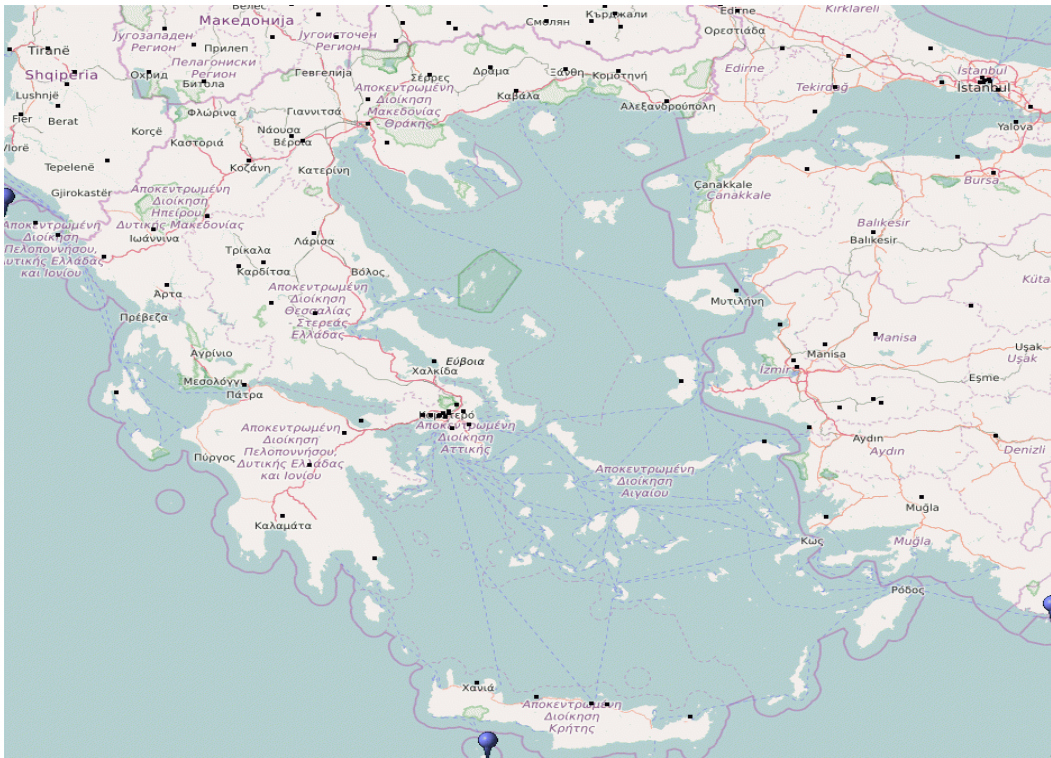
```
razer9@ubuntu: ~/Desktop/plot-latlong-0.32
razer9@ubuntu:~/Desktop/plot-latlong-0.32$ cat kapar-midar-iff.nodes2010.geo | perl plot-latlong -s 4 >Greece2.png
```

Εικόνα 8: Εκτέλεση του plot-latlong με είσοδο το kapar-midar-iff.nodes2010.geo



Εικόνα 9: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το kapar-midar.iff.nodes2010.geo

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

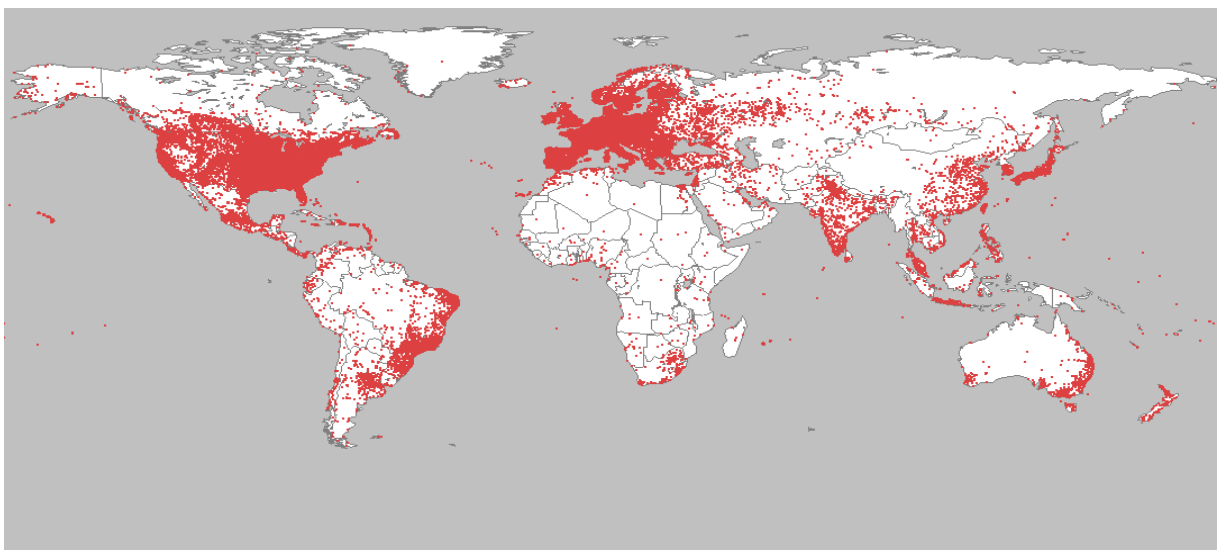


Εικόνα 10: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το `karar-midar.iff.nodes2010.geo`

Έπειτα τρέχουμε το εργαλείο `plot-latlong` με είσοδο το αρχείο `midar-iff.nodes2013.geo` και παίρνουμε τα παρακάτω αποτελέσματα

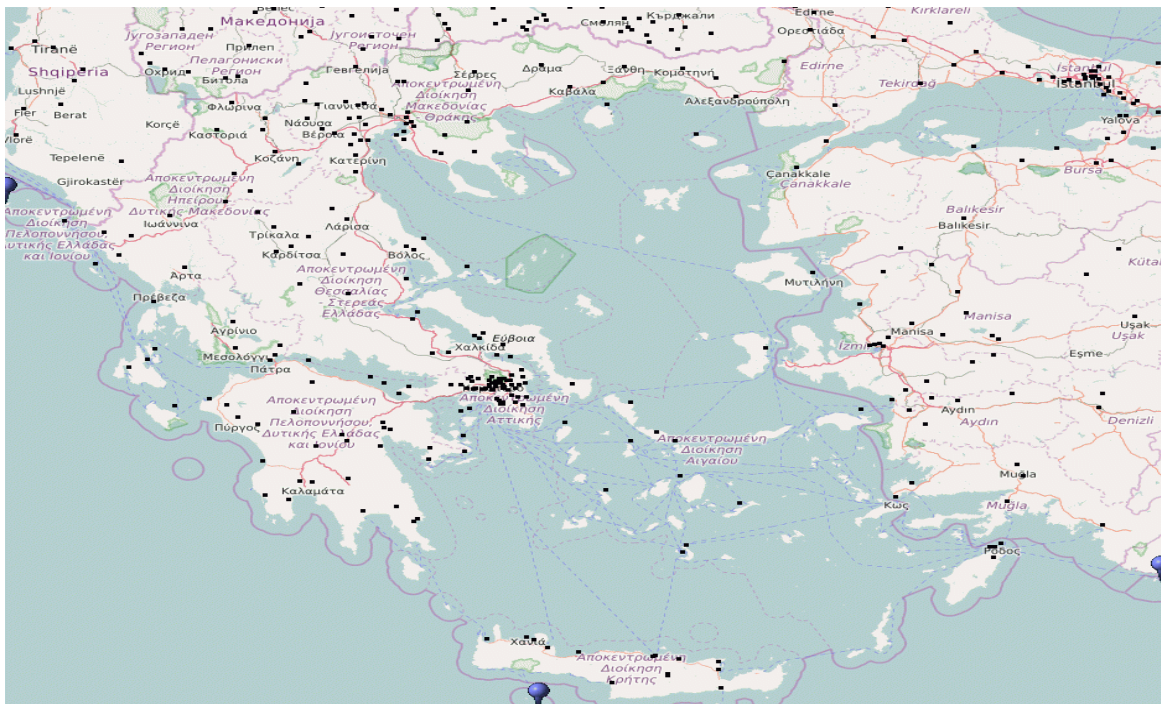
```
razer9@ubuntu: ~/Desktop/plot-latlong-0.32
razer9@ubuntu:~/Desktop/plot-latlong-0.32$ cat midar-iff.nodes2013.geo | perl plot-latlong -m World -s 2 >World3.png
```

Εικόνα 11: Εκτέλεση του `plot-latlong` με είσοδο το `midar-iff.nodes2013.geo`



Εικόνα 12: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το `midar.iff.nodes2013.geo`

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

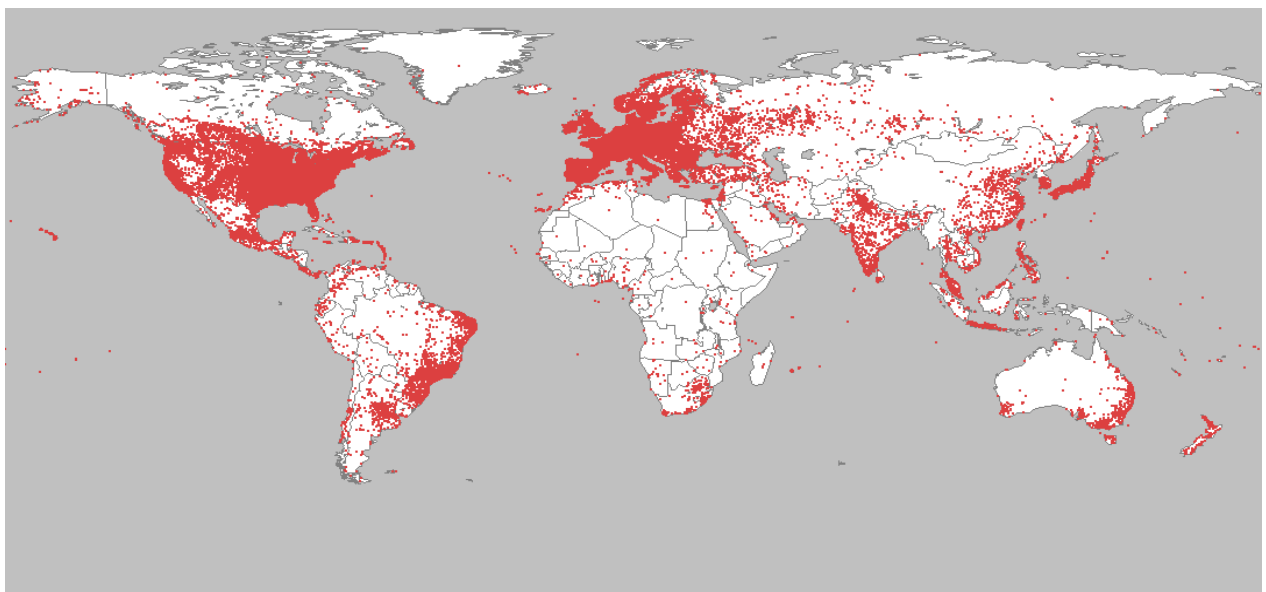


Εικόνα 13: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το midar.iff.nodes2013.geo

Τέλος τρέχουμε το εργαλείο plot-latlong με είσοδο το αρχείο kapar-midar-iff.nodes2013.geo και παίρνουμε τα κάτωθι αποτελέσματα

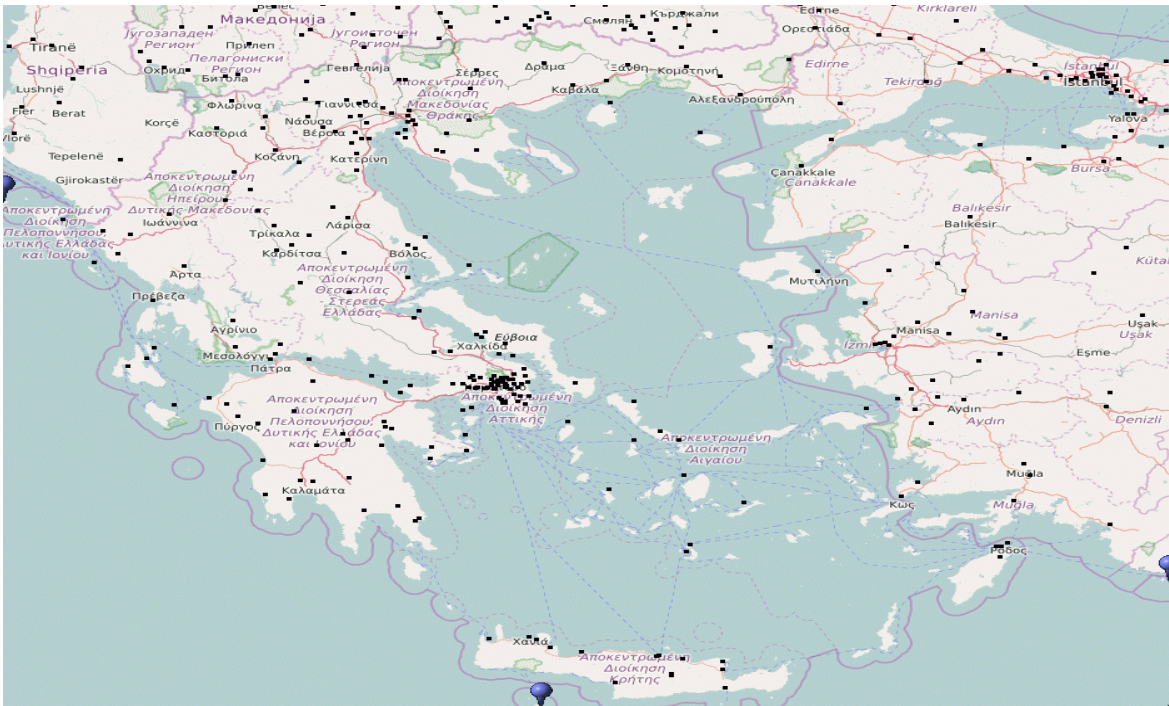
```
razer9@ubuntu: ~/Desktop/plot-latlong-0.32
razer9@ubuntu:~/Desktop/plot-latlong-0.32$ cat kapar-midar-iff.nodes2013.geo | perl plot-latlong -s 4 >Greece4.png
```

Εικόνα 14: Εκτέλεση του plot-latlong με είσοδο το kapar-midar-iff.nodes2013.geo



Εικόνα 15: Προβολή των θέσεων των δρομολογητών στο παγκόσμιο χάρτη με είσοδο το kapar-midar.iff.nodes2013.geo

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

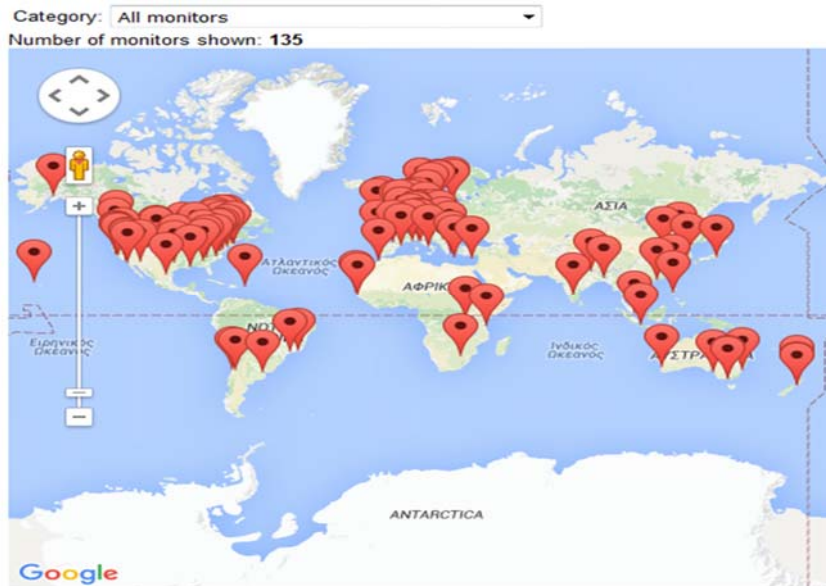


Εικόνα 16: Προβολή των θέσεων των δρομολογητών στο χάρτη της Ελλάδας με είσοδο το `karar-midar.iff.nodes2013.geo`

Όπως μπορεί κανείς εύκολα να δει ο αριθμός των προβαλλόμενων κόμβων από τα σύνολα του 2013 είναι αρκετά μεγαλύτερος από αυτό των συνόλων του 2010.

3.3 Χαρτογράφηση του Archipelago

Ο χάρτης στην εικόνα 17 απεικονίζει τη κατά προσέγγιση γεωγραφική θέση των τρέχοντων ενεργών Archipelago (Ark) συστημάτων παρακολούθησης. Το αναπτυσσόμενο μενού επιτρέπει την επιλογή διάφορων υποσυνόλων των συστημάτων παρακολούθησης [23]. Η υποδιαίρεση σύμφωνα με την ομάδα αναφέρεται στη διερευνητική μεθοδολογία της ομάδας που χρησιμοποιείται από το Ark. Τα συστήματα παρακολούθησης είναι ομαδοποιημένα σε ομάδες που δυναμικά κατανέμουν την εργασία μέτρησης ανάμεσα στα μέλη της ομάδας. Αυτή η παραλληλοποίηση επιτρέπει στον CAIDA να λαμβάνει μια traceroute μέτρηση σε όλα τα δρομολογημένα /24's σε σύντομο χρονικό διάστημα.



Εικόνα 17: Χάρτης συστημάτων παρακολούθησης του Archipelago

Ο πίνακας 5 παρέχει πρόσθετες πληροφορίες: συστήματα παρακολούθησης που διερευνούν στο IPv6 (επιπλέον του IPv4) επισημαίνονται με ένα μαύρο αστέρι (★) στη στήλη με τίτλο 'IPv6', η στήλη με τίτλο 'RAD' δείχνει εάν το σύστημα παρακολούθησης χρησιμοποιεί RADclock, που μπορεί να μετρήσει τα RTTs στο επίπεδο του μικροδευτερολέπτου, για το συγχρονισμό του χρόνου, ένα κόκκινο αστέρι (★) στη τελευταία στήλη (με τίτλο 'ORGANIZATION') υποδεικνύει ότι το υλικό του συστήματος παρακολούθησης το είχε εισφέρει ο φιλοξενούμενος οργανισμός. Η στήλη με τίτλο 'RasPi' δηλώνει ότι τα συστήματα παρακολούθησης βασίζονται σε ένα Raspberry Pi. Τα συστήματα παρακολούθησης που έχουν αναπτυχθεί σε ιδρυματικές περιοχές είναι επισημασμένα στη στήλη με τίτλο 'Res'. Η στήλη 'Status' περιέχει ένα σύνδεσμο σε ένα Xively URL με γραφήματα χρονοσειρών από αρκετές ιδιότητες του υλικού και του λογισμικού που αναφέρονται στα συστήματα παρακολούθησης.

Από τα συνολικά 135 συστήματα παρακολούθησης του Archipelago (Ark) δύο βρίσκονται στον Ελλαδικό χώρο το ath-gr που όπως μπορεί κάποιος να δει βασίζεται σε ένα Raspberry Pi, έχει αναπτυχθεί σε ιδρυματική περιοχή, έχει ενεργοποιηθεί στις 2014-04-19, βρίσκεται στη Παιανία, έχει AS αριθμό 1241 και ο οργανισμός εγκαθίδρυσής του είναι η Forthnet και το her-gr που διερευνά και στο IPv6 εκτός του IPv4, χρησιμοποιεί RADclock, ενεργοποιήθηκε στις 2008-11-04, βρίσκεται στο Ηράκλειο Κρήτης, έχει AS αριθμό 8522 και ανήκει στο Ίδρυμα Τεχνολογίας και Έρευνας (ITE).

Πίνακας 5: Συστήματα παρακολούθησης του Archipelago στην Ελλάδα

Name	IPv6	RAD	RasPi	Res	Activation Date	City	AS Number	Organization	Status
ath-gr			★	★	2014-04-19	Peania, GR	1241	Residential (Forthnet)	Monitor Data
her-gr	★	★			2008-11-04	Heraklion, Crete, GR	8522	Foundation for Research and Technology - Hellas (FORTH)	Monitor Data

3.4 AS Core: Οπτικοποιώντας την IPv4 και IPv6 Τοπολογία Διαδικτύου σε Μακροσκοπική Κλίμακα

Ο CAIDA από το 2000, έχει δημιουργήσει AS πυρήνα (core) γραφήματα με σκοπό την απεικόνιση της μεταβαλλόμενης τοπολογίας του Διαδικτύου κατά τη πάροδο του χρόνου. Για την οπτικοποίηση της μακροσκοπικής οπτικής του AS Core γίνεται χρήση του εργαλείου otter από τον CAIDA.

3.4.1 Εργαλείο Otter

Το otter είναι ένα εργαλείο του CAIDA για την οπτικοποίηση αυθαίρετων (arbitrary) δεδομένων του δικτύου που μπορούν να εκφραστούν ως ένα σύνολο κόμβων, συνδέσμων ή μονοπατιών. Τα μέλη του CAIDA ανέπτυξαν το otter για να χειριστεί τις εργασίες απεικόνισης μιας ευρείας ποικιλίας δεδομένων του Διαδικτύου με τα οποία ασχολούνται στην έρευνα τους, συμπεριλαμβανομένων συνόλων δεδομένων τοπολογίας, φόρτου εργασίας, απόδοσης και δρομολόγησης [24]. Το otter έχει χρησιμοποιηθεί για να απεικονίσει: multicast και unicast τοπολογίας βάσεις δεδομένων, πυρήνα πίνακες BGP δρομολόγησης, προσβασιμότητας και καθυστέρησης μετρήσεις, δεδομένα SNMP και δομές καταλόγου (directory) ιστοσελίδας. Η δύναμη του otter βρίσκεται στην ανεξαρτησία των δεδομένων του. Μπορεί να χειριστεί οποιοδήποτε μορφοποιημένο σύνολο δεδομένων που αποτελείται από συνδέσμους και κόμβους. Ο σχεδιασμός του otter βασίζεται σε τέσσερα αναμενόμενα χαρακτηριστικά χρήσης:

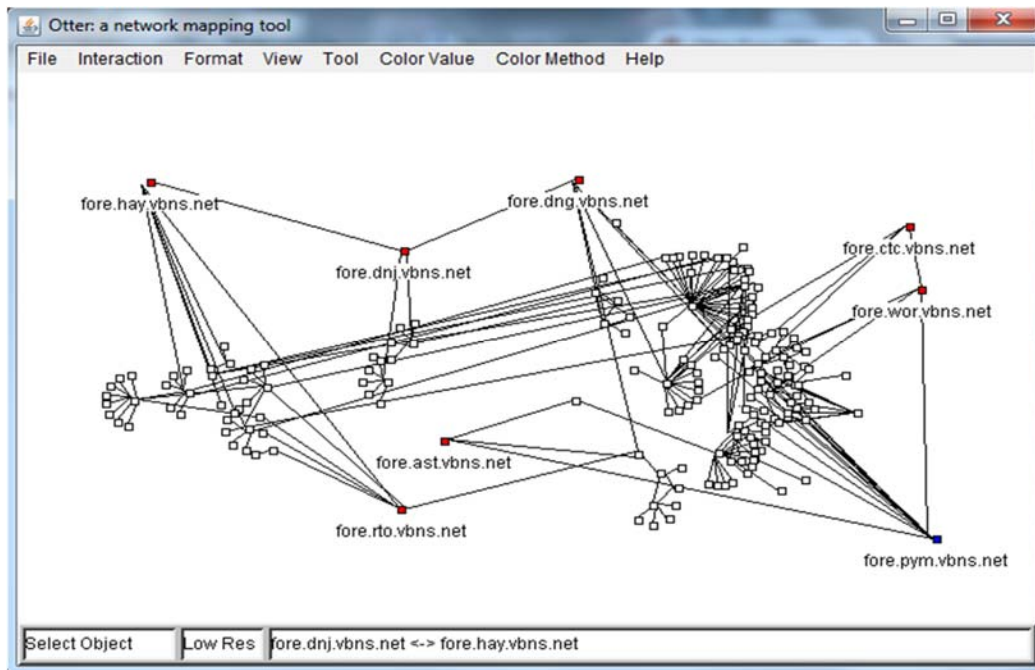
1. Οι τύποι των αντικειμένων για οπτικοποίηση είναι: κόμβοι, σύνδεσμοι, μονοπάτια
2. Τα μεγάλα σύνολα δεδομένων, τα οποία οδηγούν σε μεγάλα γραφήματα, απαιτούν αποδοτική χρήση της μνήμης
3. Οι χρήστες θέλουν και γεωγραφική και τοπολογική τοποθέτηση
4. Οι χρήστες θέλουν να είναι σε θέση να τροποποιήσουν την απεικόνιση: εστίαση, αλλαγή της διάταξης του γραφήματος

και προσαρμόζεται σύμφωνα με αυτές τις βασικές διαφορές:

1. Κάθε σύνολο δεδομένων μπορεί να έχει διαφορετικό αριθμό ιδιοτήτων ή ιδιαίτερων χαρακτηριστικών δεδομένων που ενδιαφέρουν για οπτικοποίηση
2. Οι ιδιότητες μπορεί να είναι διαφορετικού τύπου
3. Οι κόμβοι, οι σύνδεσμοι και τα μονοπάτια μπορεί να έχουν ριζικά διαφορετική σημασιολογία στα διάφορα σύνολα δεδομένων
4. Ορισμένοι αλγόριθμοι διάταξης γραφήματος μπορεί να μην έχουν νόημα για ορισμένα σύνολα δεδομένων

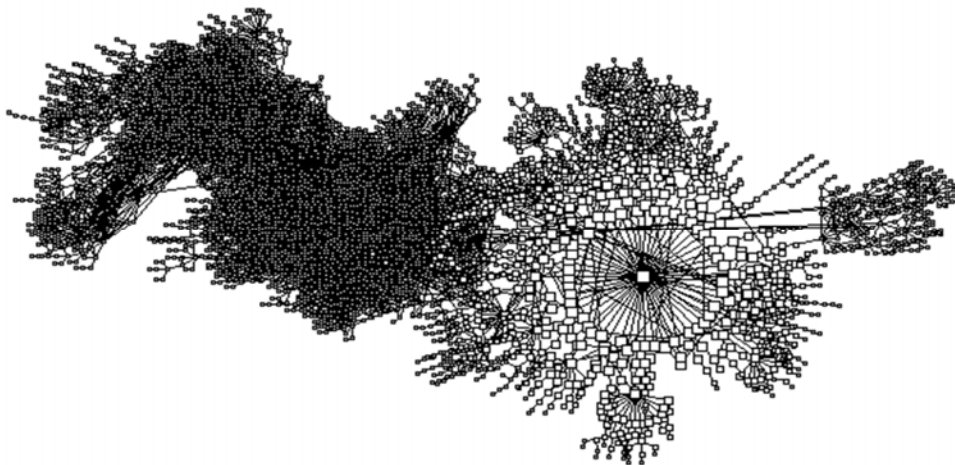
Στην εικόνα 18 γίνεται απεικόνιση του εργαλείου otter και των επιλογών που αυτό προσφέρει στη προβολή του γραφήματος που προκύπτει από τα σύνολα δεδομένων που είναι διαθέσιμα.

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida



Εικόνα 18: Κύριο παράθυρο του otter

Στην εικόνα 19 γίνεται απεικόνιση από το otter ενός μεγάλου συνόλου δεδομένων τοπολογίας, απεικονίζονται περίπου 30.000 κόμβοι



Εικόνα 19: Απεικόνιση 30.000 κόμβων από το otter

3.4.2 Πηγή Δεδομένων

Η απεικόνιση του AS Core αντιπροσωπεύει μία μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας του Διαδικτύου που έχει ληφθεί κατά τη διάρκεια ενός έτους. Το γράφημα απεικονίζει τόσο το εκτεταμένο γεωγραφικό πεδίο εφαρμογής όσο και την πλούσια διασύνδεση των κόμβων που συμμετέχουν στο παγκόσμιο σύστημα δρομολόγησης του

Διαδικτύου. Συντονισμένο από την ενεργή υποδομή μετρήσεων του CAIDA το Archipelago (Ark), ο CAIDA συλλέγει δεδομένα από συστήματα ανίχνευσης που εξετάζουν μονοπάτια προς εκατομμύρια δικτύων που καλύπτουν ένα μεγάλο μέρος των δρομολογήσιμων προθεμάτων που μπορεί κάποιος να δει στους BGP πίνακες δρομολόγησης [25]. Ο CAIDA συγκεντρώνει τα IP-επιπέδου δεδομένα για την κατασκευή IPv4 και IPv6 Διαδικτυακής συνδεσιμότητας γραφημάτων στο AS επίπεδο. Κάθε AS αντιστοιχεί περίπου σε ένα προμηθευτή υπηρεσίας Διαδικτύου (ISP). Ο CAIDA χαρτογραφεί κάθε διεύθυνση IP που παρατηρείται στο AS και είναι υπεύθυνη για τη δρομολόγηση της κυκλοφορίας σε αυτό, π.χ το προέλευσης (τέλος-του-μονοπατιού) AS για το IP πρόθεμα αντιπροσωπεύει τη καλύτερη αντιστοιχία για αυτή τη διεύθυνση στους BGP πίνακες δρομολόγησης που συλλέγονται από τις προβολές δρομολόγησης. Η θέση κάθε AS κόμβου απεικονίζεται σε πολικές (polar) συντεταγμένες (ακτίνα, γωνία) που υπολογίζονται ως (σχήμα 10):

$$\begin{aligned} \text{ακτίνα} &= 1 - \log \left(\frac{\text{βαθμός διέλευσης(AS)} + 1}{\text{μέγιστος βαθμός διέλευσης} + 1} \right) \\ \text{γωνία} &= \left(\begin{array}{l} \text{γεωγραφικό μήκος των AS's} \\ \text{BGP προθεμάτων στο Netacuity} \end{array} \right) \end{aligned}$$

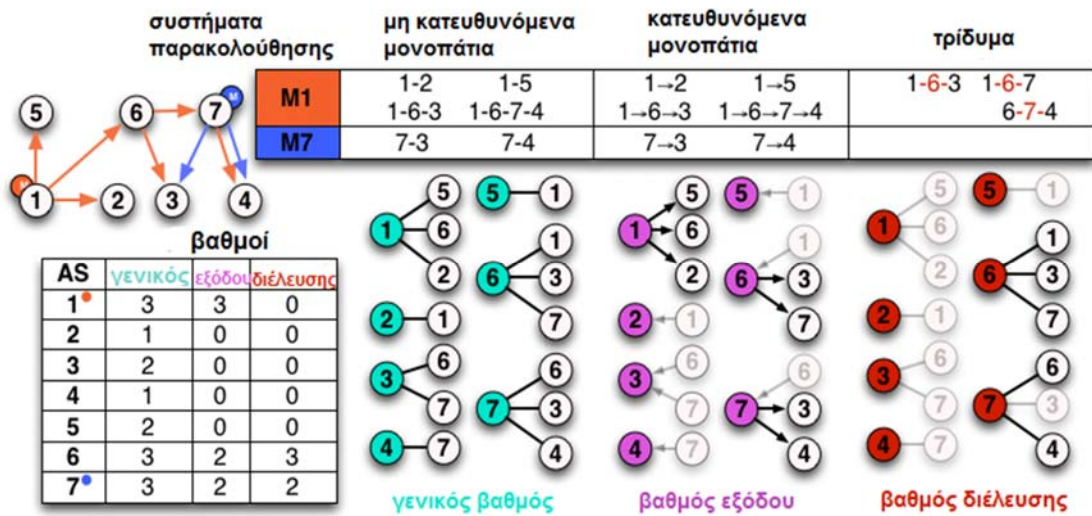
Σχήμα 10: Συντεταγμένες του AS στην AS core οπτικοποίηση

3.4.3 Μεθοδολογία Κατάταξης AS

Από το 2013 έχουν γίνει δύο σημαντικές βελτιώσεις στο τρόπο που δημιουργούνται τα γραφήματα, συμπεριλαμβανομένου του πώς κατατάσσονται τα επιμέρους ASes και τη βελτίωση της απεικόνισης του AS Core. Πρώτον, τώρα η κατάταξη των ASes βασίζεται στο βαθμό διέλευσής τους και έπειτα στο βαθμό εξόδου. Δεύτερον, τώρα ο CAIDA συμπεραίνει συνδέσμους σε ολόκληρο χώρο διευθύνσεων του Διαδικτυακού σημείου ανταλλαγής (IXP), αντί να θεωρείται το IX από μόνο του ένας κόμβος στον οποίο διάφοροι ISPs συνδέονται.

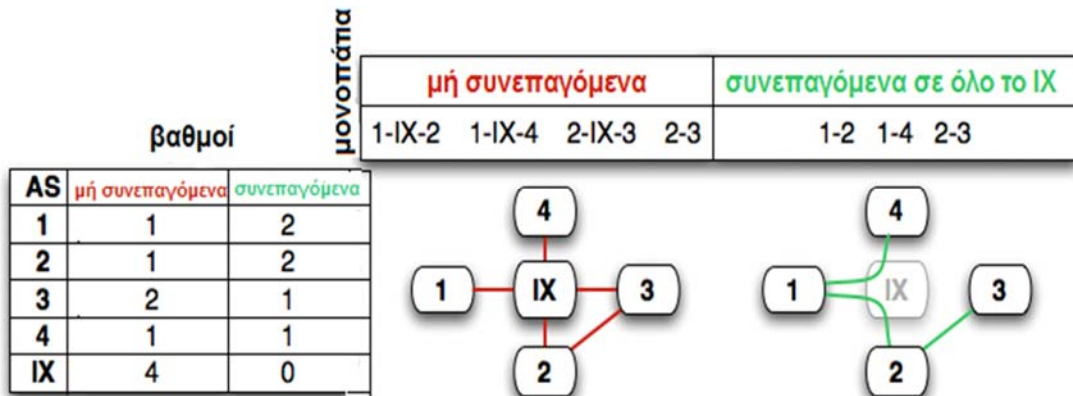
Τα παρακάτω παραδείγματα δείχνουν πώς αυτές οι αλλαγές μπορούν να επηρεάσουν τα δεδομένα που προκύπτουν, χρησιμοποιώντας δείγματα από την οπτικοποίηση του AS Core το 2013.

1. **Μεταβολή της μέτρησης της κατάταξης από το βαθμό εξόδου στο βαθμό διέλευσης (transit):** Για να τοποθετηθούν τα ASes στο γράφο, ο CAIDA αρχικά τα ταξινομεί με βάση το βαθμό εξόδου, ο οποίος είναι ο αριθμός των εξερχόμενων συνδέσμων από ένα AS όπως αυτός παρατηρήθηκε στα μονοπάτια του AS. Στο σχήμα 11, τα ASes 2, 3, 4 και 5 έχουν βαθμό εξόδου 0, δεδομένου ότι τίποτα δεν τα ακολουθεί σε οποιοδήποτε μονοπάτι. Τα ASes 6 και 7 έχουν βαθμό εξόδου 2, ενώ το AS 1 έχει βαθμό εξόδου 3 κάτι που οφείλεται μόνο στο σύστημα παρακολούθησης που φιλοξενεί. Χρησιμοποιώντας το βαθμό διέλευσης ο βαθμός του AS 1 πέφτει στο 0 και διαφοροποιείται του AS 6, το οποίο παρέχει διέλευση για 3 ASes από του AS 7 που παρέχει διέλευση για 2 μόνο ASes.



Σχήμα 11: Σύγκριση του γενικού βαθμού, του βαθμού διέλευσης και του βαθμού εξόδου των ASes

- Συμπεραίνοντας συνδέσμους σε όλο το IX χώρο διευθύνσεων:** Η δεύτερη σημαντική αλλαγή του CAIDA στη μεθοδολογία ήταν ότι συμπεράνε (infer) συνδέσμους σε όλα τα IXes όταν ο χώρος διευθύνσεων του IX εμφανίζεται στη traceroute εντολή. Τα IXes παρέχουν διάρθρωση υποδομών μέσω της οποίας άλλα ASes μοιράζονται. Αν και το IX μπορεί να έχει τα δικά του AS και χώρο διευθύνσεων, οι σύνδεσμοι που διασχίζουν το χώρο διευθύνσεων ενός IX είναι στη πραγματικότητα σύνδεσμοι μοιράσματος απευθείας μεταξύ των συνδεδεμένων ASes στο IX, και όχι στο ίδιο το IX. Παρόλα αυτά ανεπεξέργαστα traceroute δεδομένα μπορεί να περιλαμβάνουν IP διευθύνσεις που χρησιμοποιούνται από ένα IX για να διευκολυνθεί αυτή η άμεση ανταλλαγή κίνησης σε όλο το IX. Μερικές φορές ένα AS θα μπορούσε να ανακοινώσει το χώρο διευθύνσεων του IX, σε αυτή τη περίπτωση μια απλοϊκή αντιστοίχιση αυτών των ανεπεξέργαστων IP διευθύνσεων στα ASes θα εισαγάγει λανθασμένους συνδέσμους ανταλλαγής κίνησης μεταξύ του AS που ανακοίνωσε το χώρο διευθύνσεων του IX και των ASes που ανταλλάσσουν κίνηση στο IX, καθώς και να κρύψει τους πραγματικούς συνδέσμους ανταλλαγής κίνησης μεταξύ των ζευγών των ASes εκατέρωθεν του IX. Για να ανακτηθεί η πραγματική συνδεσιμότητα για την ανταλλαγή της κίνησης, ο CAIDA συμπεραίνει συνδέσμους σε όλο το χώρο διευθύνσεων IX που είναι αυτό-αναγνωριζόμενος από το IX στο PeeringDB. Αυτή η νέα μέθοδος δεν οδηγεί πάντα σε αύξηση στους βαθμούς των ASes ανακοινώνοντας το IX χώρο διευθύνσεων. Στο σχήμα 12, το AS 3 ήδη έχει άμεση σύνδεση με το AS 2 σε κάποια άλλη θέση, οπότε συμπεραίνεται ότι η ανταλλαγή κίνησης σε όλο το IX έχει σαν αποτέλεσμα την απώλεια της σύνδεσης με το AS του IX και καθόλου επιπλέον γείτονες, με αποτέλεσμα τη πτώση του βαθμού του AS 3.



Σχήμα 12: Συμπεραίνοντας AS συνδεσιμότητα σε όλο το IX χώρο διευθύνσεων

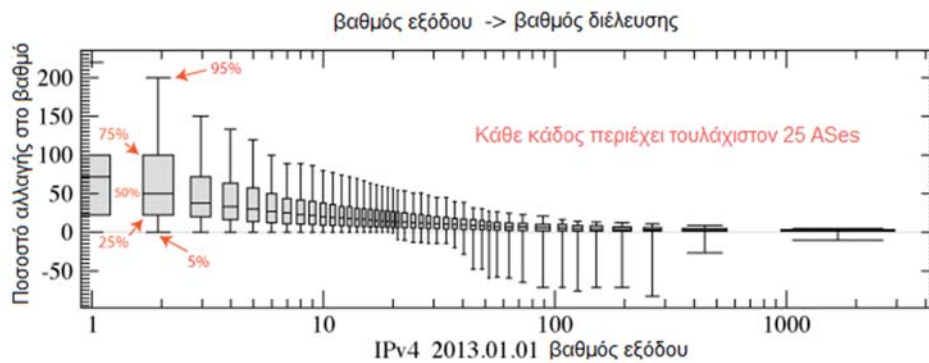
3.4.3.1 Ποσοτικοποίηση των Επιπτώσεων των Αλλαγών στους Υπολογισμούς του IPv4 AS βαθμού

Μέσω των σχημάτων (14, 15, 17, 18, 20, 21) παρουσιάζεται η επίδραση των δύο αλλαγών στη μεθοδολογία στα δεδομένα που προκύπτουν. Σε κάθε σχήμα, ο CAIDA ομαδοποιεί τα ASes σε κάδους (bins) από τουλάχιστον 25 με ένα δεδομένο εύρος βαθμού εξόδου. Κάθε κάδος αντιπροσωπεύεται από ένα διάγραμμα πλαισίου (box-and-whisker plot), το οποίο σκιαγραφεί το εύρος των αλλαγών στο βαθμό κατά τη μετακίνηση από μια παλιά μέθοδο σε μία νέα. Η κορυφή του κουτιού αντιπροσωπεύει το 3^ο τεταρτημόριο (75%), η γραμμή στο μέσο του κουτιού είναι το 2^ο τεταρτημόριο (50%) και το κάτω μέρος του κουτιού είναι το 1^ο τεταρτημόριο (25%). Οι γραμμές που εκτείνονται κάθετα από τα κουτιά αντιπροσωπεύουν το 5% και 95%. Ο CAIDA έχει υπολογίσει τα ποσοστά για τη διανομή αυτή ως εξής (σχήμα 13):

$$\text{διάγραμμα κουτιού } y = \frac{\text{βαθμός(παλιάς)} - \text{βαθμός(νέας)}}{\text{βαθμός(νέας)}}$$

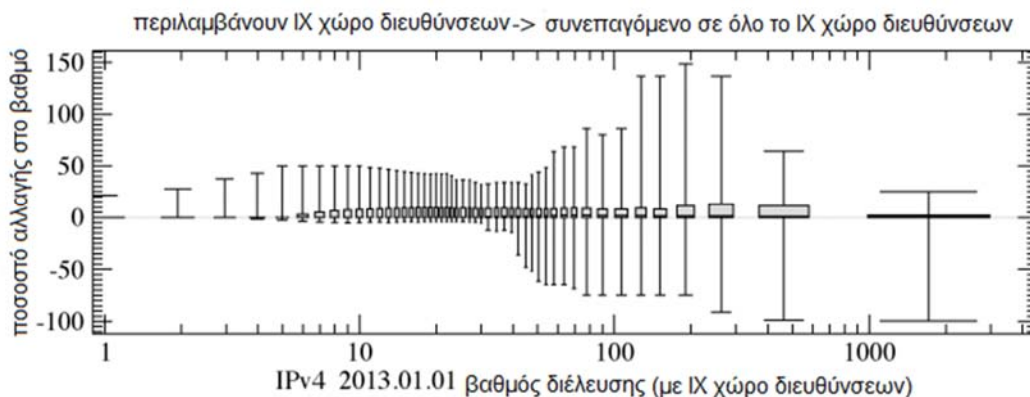
Σχήμα 13: Υπολογισμός διαγράμματος κουτιού

Το σχήμα 14 δείχνει την αλλαγή στους AS βαθμούς κατά την μετάβαση από τη χρήση του βαθμού εξόδου στο βαθμό διέλευσης για τα IPv4 ASes. Πάνω από το 75% των ASes είχαν μεγαλύτερο βαθμό διέλευσης από βαθμό εξόδου, αν και η σχετική μεταβολή μειώθηκε όσο ο βαθμός εξόδου αυξήθηκε. Για ASes με βαθμό εξόδου μεγαλύτερο από 20, ένα αυξανόμενο τμήμα (αλλά ποτέ πάνω από το 25% των ASes σε ένα κάδο) είχαν βαθμό διέλευσης μικρότερο από το βαθμό εξόδου τους.



Σχήμα 14: Αλλαγή του AS βαθμού κατά τη μετάβαση από το βαθμό εξόδου στο βαθμό διέλευσης για IPv4 ASes

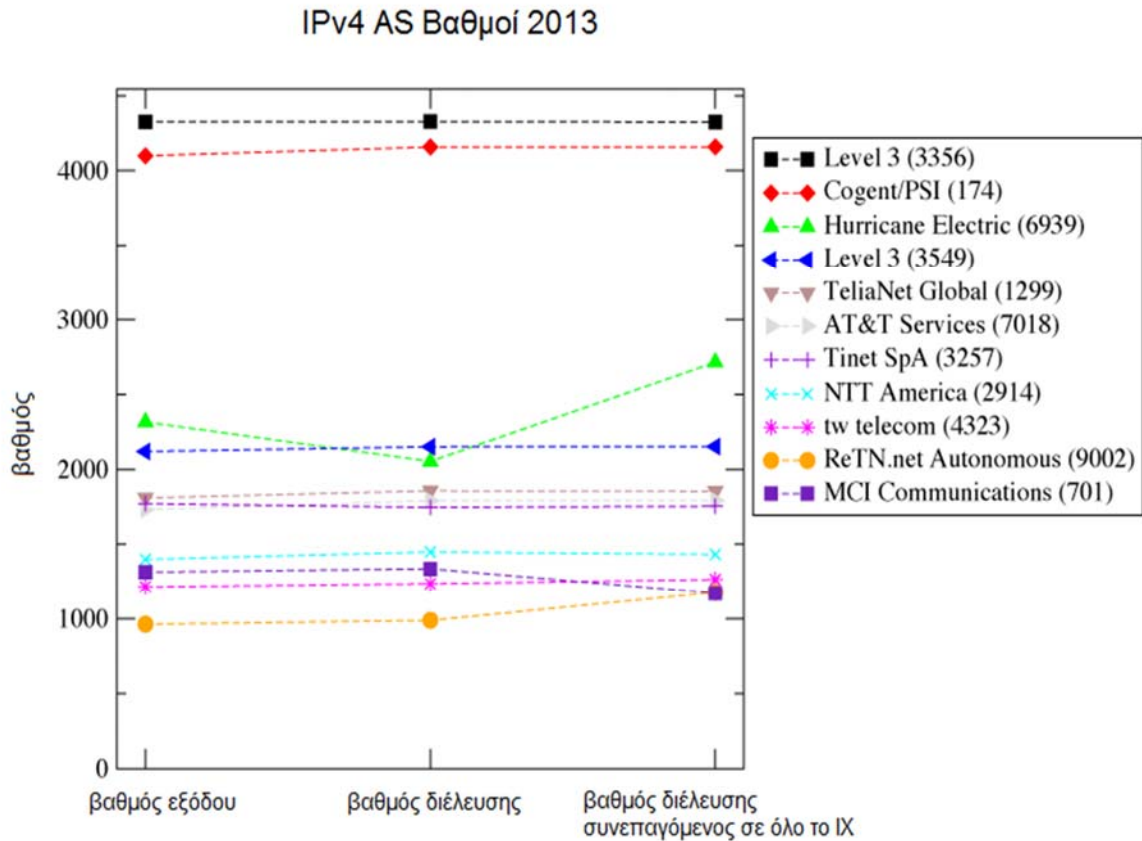
Το σχήμα 15 δείχνει τη ποσοστιαία διαφορά στο βαθμό διέλευσης κατά την αφαίρεση συνδέσμων συνδεδεμένων σε ASes, ανακοινώνοντας το IPv4 IX χώρο διευθύνσεων και προσθέτοντας AS συνδέσμους συνεπαγόμενων σε όλο το IX χώρο διευθύνσεων. Η επίδραση στους βαθμούς διέλευσης είναι λεπττή. Τα ASes πιθανώς να χάσουν συνδέσμους στο IX χώρο διευθύνσεων και ενδεχομένως να αποκτήσουν νέους συνδέσμους σε όλο το IX χώρο διευθύνσεων. Στα δεδομένα του CAIDA, γεφυρώνοντας συμπεράσματα σε όλο το IX χώρο διευθύνσεων έχει σαν αποτέλεσμα μια απώλεια 5% στους υπάρχοντες IPv4 συνδέσμους, αλλά απέδωσε επιπλέον 7,000 συνδέσμους (12% του αρχικού) για μία καθαρή αύξηση 7% στους συνδέσμους. Τα μισά από τα ASes είδαν μικρότερη από 10% μεταβολή στο βαθμό διέλευσής τους. Τα ASes με μεσαίους (ανάμεσα σε 100 και 360) βαθμούς διέλευσης χρησιμοποιώντας τη παλιά μέθοδο είδαν τη μεγαλύτερη συγκριτικά μεταβολή στο βαθμό διέλευσής τους. Ανάμεσα στα κορυφαία 50 ASes με τη παλιά μέθοδο, 4 είχαν πτώση στο βαθμό διέλευσής τους σε λιγότερο από 10% της παλιάς τιμής τους (αυτά τα 4 ήταν IXes ή ανακοινώνουν IX χώρο διευθύνσεων).



Σχήμα 15: Ποσοστιαία διαφορά βαθμού διέλευσης

Το σχήμα 16 επικεντρώνεται στα 10 κορυφαία ASes, βαθμολογημένα με βάση το βαθμό εξόδου, το βαθμό διέλευσης ή το βαθμό διέλευσης δεδομένου των συνεπαγόμενων συνδέσμων στο IPv4 γράφημα. Το Hurricane Electric (AS 6639) ήταν το μόνο AS του οποίου ο βαθμός εξόδου ήταν σημαντικά μεγαλύτερος από το βαθμό διέλευσής του, πιθανόν εξαιτίας των δύο Ark συστημάτων παρακολούθησης που φιλοξενεί. Το Hurricane Electric (AS 6639) επίσης είδε τους περισσότερους νέους συνδέσμους που συνεπάγονται σε όλο το IX χώρο διευθύνσεων, το οποίο αύξησε το βαθμό του από 2,053 σε 2,715 (32%). Το

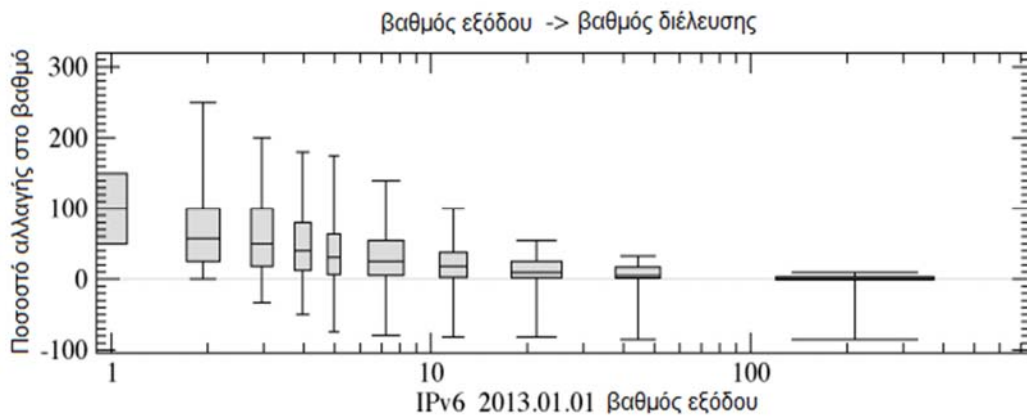
ReTN.net (AS 9002) είδε το βαθμό διέλευσης του να αυξάνεται από 988 σε 1,176 (19%). Τα περισσότερα από αυτά τα ASes είχαν βαθμούς διέλευσης στο 1% των βαθμών εξόδου τους.



Σχήμα 16: Βαθμός διέλευσης, βαθμός εξόδου και βαθμός διέλευσης συνεπαγόμενος σε όλο το IX χώρο διευθύνσεων των 10 κορυφαίων ASes

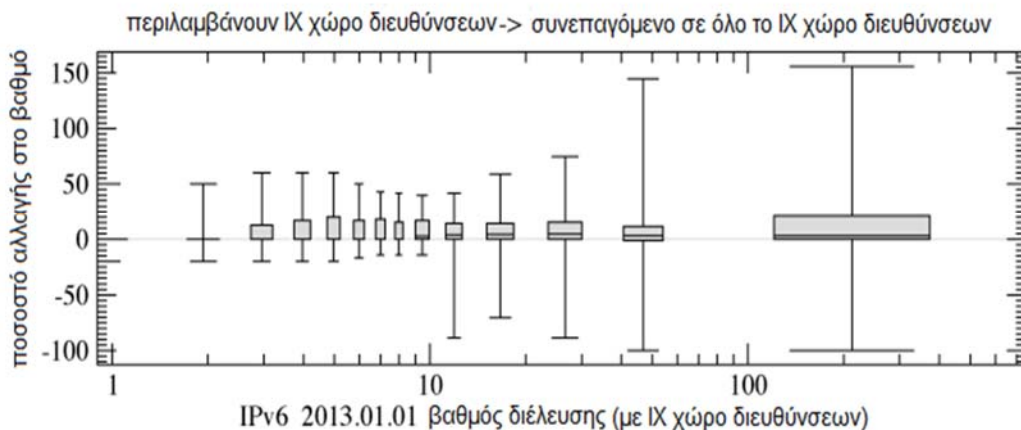
3.4.3.2 Επιδράσεις στο IPv6

Το σχήμα 17 δείχνει τις ποσοστιαίες διαφορές μεταξύ του βαθμού εξόδου και του βαθμού διέλευσης των ASes στο IPv6 γράφημα. Το IPv6 έχει μια εικόνα παρόμοια με το γράφημα IPv4: πάνω από το 75% των ASes είχαν ένα μεγαλύτερο βαθμό διέλευσης από βαθμό εξόδου, αν και η σχετική μεταβολή μειώθηκε όσο ο βαθμός εξόδου αυξήθηκε. Σε αντίθεση με το IPv4, το 5% των ASes με βαθμό εξόδου μεγαλύτερο από 5 είχαν πτώση του βαθμού διέλευσής τους περίπου στο 0, συνήθως σε περιπτώσεις όπου αυτοί φιλοξενούσαν ένα Ark σύστημα παρακολούθησης (το οποίο αύξησε τους παρατηρούμενους βαθμούς εξόδου τους αλλά δεν επηρέασε το βαθμό διέλευσής τους). Αυτά τα ASes είχαν μια παρόμοια πτώση στο IPv4 γράφημα, αλλά το μικρότερο μέγεθος του IPv6 γραφήματος αυξάνει την ορατότητά τους.



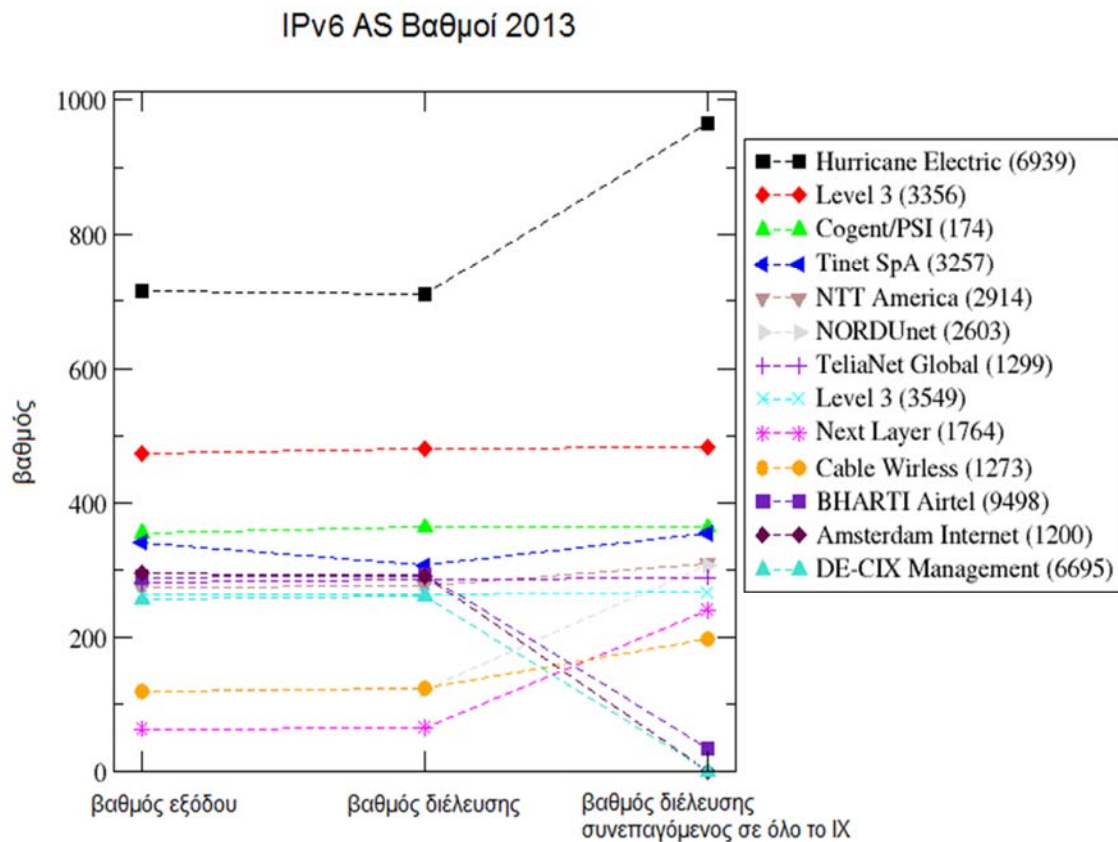
Σχήμα 17: Ποσοστιαία διαφορά μεταξύ του βαθμού εξόδου και του βαθμού διέλευσης των ASes σε IPv6

Το σχήμα 18 δείχνει ότι το 75% των ASes είχε αύξηση στο βαθμό διέλευσής τους όταν σύνδεσμοι συνήχθησαν σε όλο το IX χώρο διευθύνσεων. Για τα 50 κορυφαία ASes από τη παλιά μέθοδο, 4 από αυτά (3 IXes και ένα AS που ανακοινώνουν το IX χώρο διευθύνσεων) είδαν τους βαθμούς τους να πέφτουν περίπου στο 0. Τα μισά από τα ASes είχαν μεταβολή στους βαθμούς τους λιγότερο από 25%.



Σχήμα 18: Ποσοστιαία διαφορά βαθμού διέλευσης σε IPv6

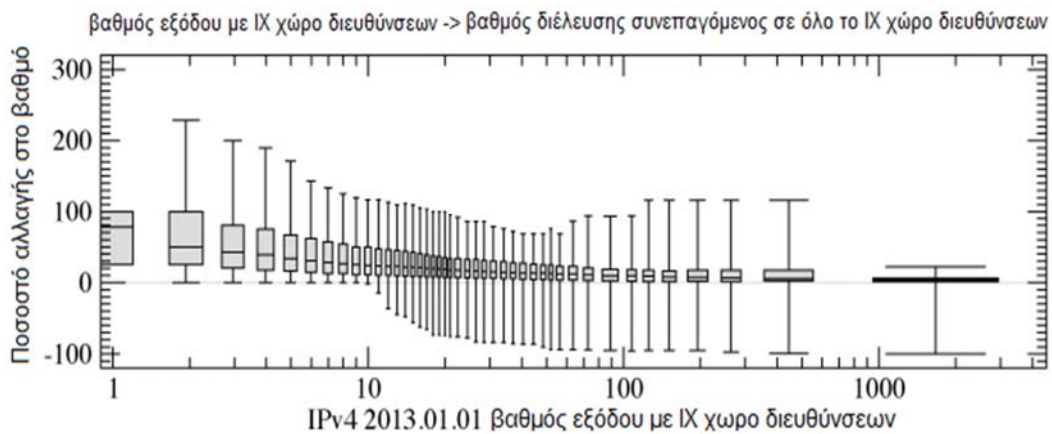
Από τα 10 κορυφαία ASes με βάση το βαθμό διέλευσης στο γράφημα (σχήμα 19), όπου ο CAIDA αντιμετωπίζει τα IXes σαν κόμβους, ο βαθμός διέλευσης του Hurricane Electric αυξήθηκε κατά 36% όταν προκύπτουν σύνδεσμοι σε όλο το IX χώρο διευθύνσεων. Όπως αναμενόταν, το Amsterdam Internet Exchange (AS 1200) και το German Internet Exchange (AS 6695) είχαν τη μεγαλύτερη πτώση στο βαθμό διέλευσης όταν ο CAIDA αφαιρέσε τους απευθείας συνδέσμους με τα IXes. Αν και το Bharti Airtel (AS 9498) δεν είναι ένα IX, είχε ανακοινώσει το χώρο διευθύνσεων ενός IX και έτσι ο βαθμός διέλευσής του μειώθηκε 88.4%. Το NORDUnet (AS 2603) και το Next Layer Telecommunication (AS 1764) εντάχθηκαν στα κορυφαία 10, με βαθμό διέλευσης αυξημένο 155.8% και 269.2% αντίστοιχα. Ο μεγαλύτερος ρόλος των IXes στη λιγότερα ώριμη IPv6 τοπολογία αντικατοπτρίζεται στην ισχυρότερη επίδραση που η αφαίρεσή τους έχει στη βασισμένη στο βαθμό AS κατάταξη.



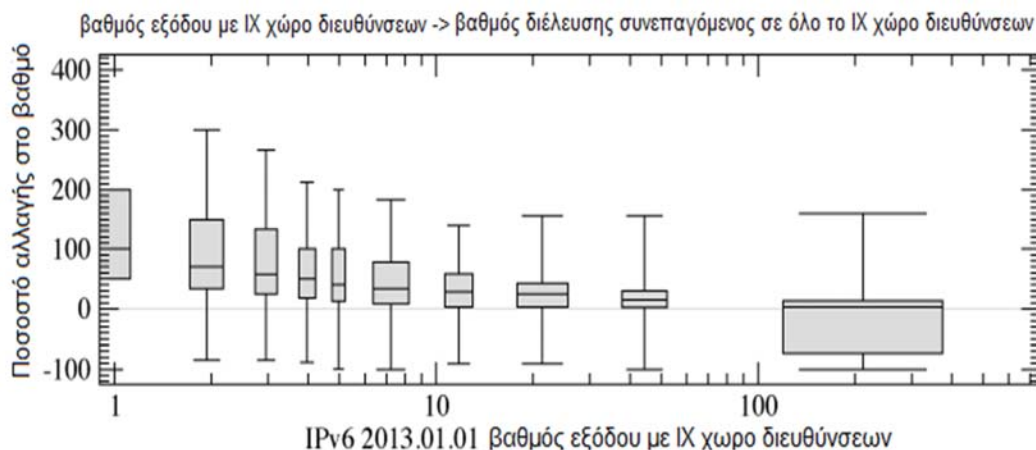
Σχήμα 19: Βαθμός διέλευσης, βαθμός εξόδου και βαθμός διέλευσης συνεπαγόμενος σε όλο το IX χώρο διευθύνσεων των 10 κορυφαίων ASes σε IPv6

3.4.3.3 Σύνοψη των Επιδράσεων των δύο Αλλαγών

Στο IPv4 και στο IPv6, τα περισσότερα συνδεδεμένα ASes γνώρισαν περιορισμένες επιδράσεις λόγω της αλλαγής από το βαθμό εξόδου στο βαθμό διέλευσης (σχήμα 20 και 21). Στους περισσότερους κάδους, πάνω από το 75% των ASes αύξησαν το βαθμό τους. Ο κορυφαίος κάδος του IPv6 (κορυφαία 25 ASes με βάση το βαθμό εξόδου) ήταν η εξαίρεση, με πάνω από το μισό μείωση στο βαθμό τους, λόγω των πολλών IX ASes που εκπροσωπούνται σε αυτό το κάδο. Στο IPv4 και IPv6, το Hurricane Electric (AS 6939) είδε το βαθμό διέλευσής του να αυξάνεται πάνω από 30% όταν σύνδεσμοι συνήχθησαν σε όλο το IX χώρο διευθύνσεων. Τα περισσότερα άλλα σε μεγάλο βαθμό συνδεδεμένα ASes στο IPv4 είδαν μικρές αλλαγές λόγω της αφαίρεσης IX κόμβου, αλλά τα κορυφαία 10 του IPv6 AS γραφήματος είδαν σημαντική αλλαγή λόγω του μεγαλύτερου ρόλου των IXes στη λιγότερο ώριμη IPv6 τοπολογία.



Σχήμα 20: Επίδραση της μετάβασης από το βαθμό εξόδου στο βαθμό διέλευσης σε IPv4



Σχήμα 21: Επίδραση της μετάβασης από το βαθμό εξόδου στο βαθμό διέλευσης σε IPv6

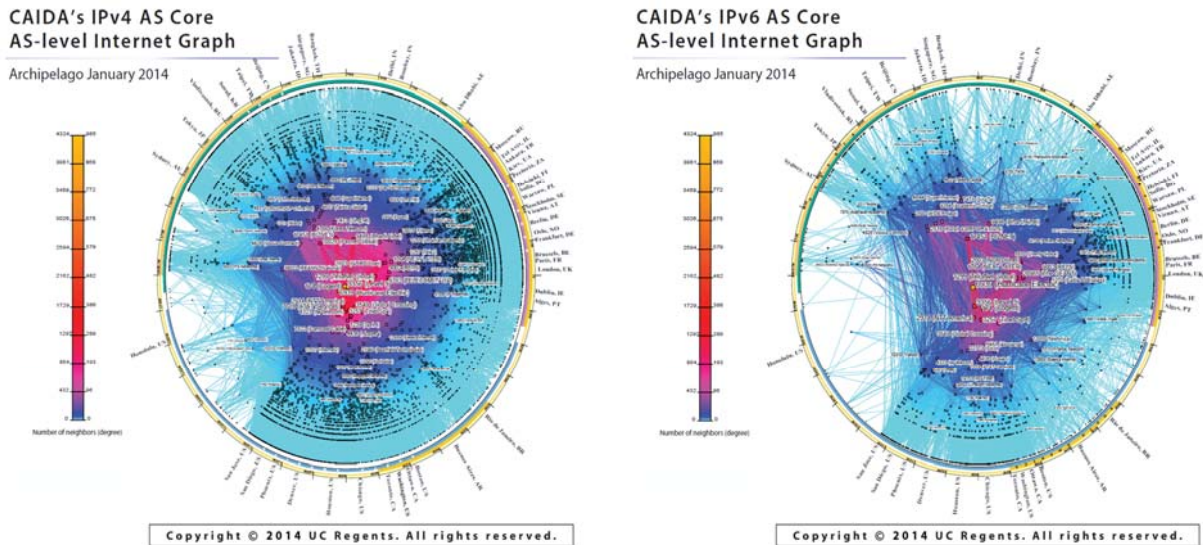
3.4.4 Μεταβολή του AS core στο Πέρασμα των Χρόνων

Η εικόνα 20 αποτελεί μια μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου που λήφθηκε το 2014. Το γράφημα απεικονίζει τόσο το εκτεταμένο γεωγραφικό πεδίο όσο και τη πλούσια διασύνδεση των κόμβων που συμμετέχουν στο παγκόσμιο σύστημα δρομολόγησης του Διαδικτύου.

Για το IPv4 χάρτη, ο CAIDA συνέλεξε δεδομένα από 74 συστήματα παρακολούθησης σε 33 χώρες σε 6 ηπείρους. Περιλαμβάνει 37,541,408 IP διευθύνσεις, 31,075,574 IP συνδέσμους, 37,197 ASES και 130,095 περιόδους ανταλλαγής κίνησης. Συντονισμένα από την ενεργή υποδομή μετρήσεων του CAIDA, το Archipelago (Ark), τα συστήματα παρακολούθησης διερεύνησαν μονοπάτια προς 263 εκατομμύρια /24 δίκτυα που καλύπτουν το 92.4% των δρομολογήσιμων προθεμάτων που υπάρχουν στους Route Views Border Gateway Protocol (BGP) πίνακες δρομολόγησης τον Ιανουάριο του 2014. Για το IPv6 χάρτη, ο CAIDA συνέλεξε δεδομένα από 33 IPv6-συνδεδεμένα Ark συστήματα παρακολούθησης που βρίσκονται σε 21 χώρες σε 6 ηπείρους. Περιλαμβάνει 51,884 IPv6 διευθύνσεις, 133,988 IPv6 συνδέσμους, 4,337 ASES και 16,856 περιόδους ανταλλαγής κίνησης. Το υποσύνολο των συστημάτων παρακολούθησης διερεύνησε μονοπάτια προς 3.3 εκατομμύρια διευθύνσεις IPv6, οι οποίες

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

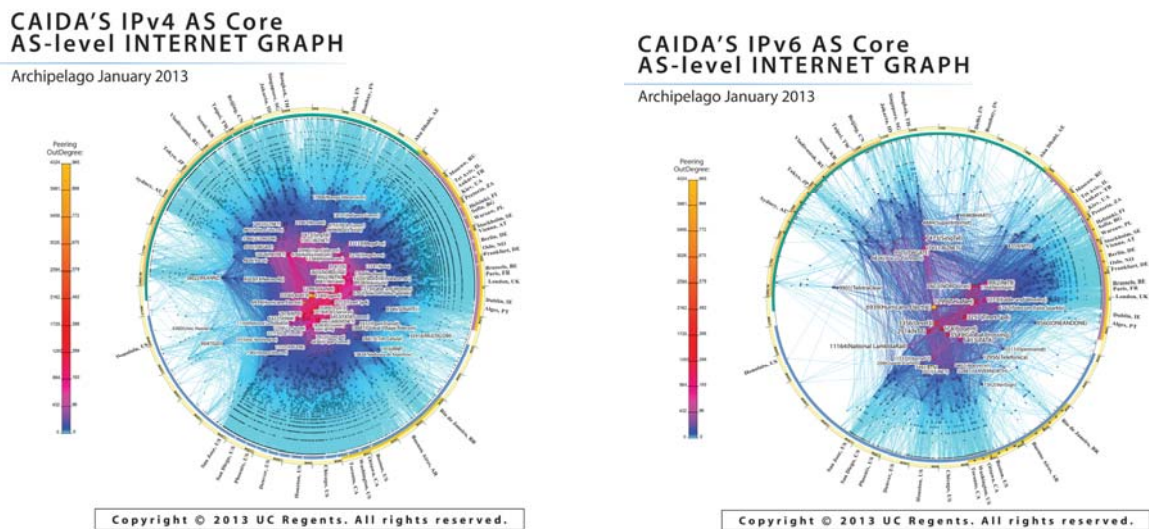
αντιπροσωπεύουν το 76,1% των συνολικά δρομολογήσιμων προθεμάτων που υπάρχουν στους Route Views Border Gateway Protocol (BGP) πίνακες δρομολόγησης τον Ιανουάριο του 2014.



Εικόνα 20: Μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου το 2014

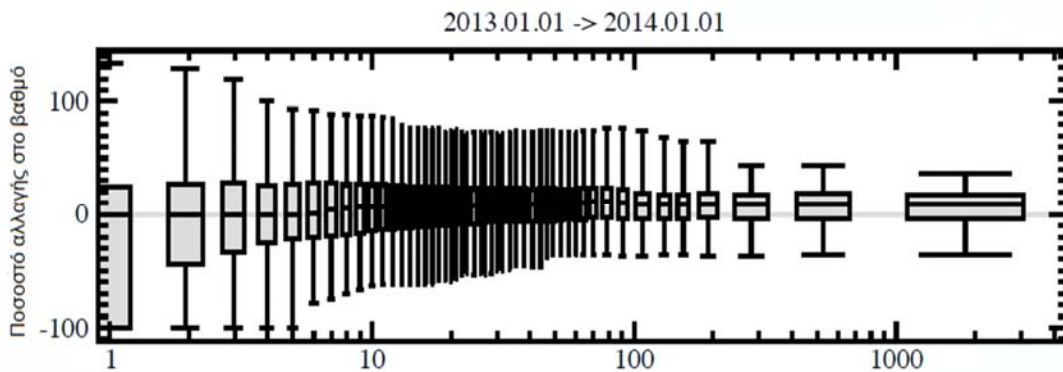
3.4.4.1 Μεταβολή σε Σχέση με το 2013

Η εικόνα 21 μας δίνει μια μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου για το 2013. Το IPv6 γράφημα όπως και τα προηγούμενα χρόνια παρουσίασε ταχύτερη αύξηση σε σχέση με το IPv4 γράφημα. Κατά τη χρονική περίοδο από Ιανουάριο 2013 σε Ιανουάριο 2014, ο αριθμός των IPv6 ASes αυξήθηκε κατά 80% και ο αριθμός των συνδέσεων μεταξύ τους αυξήθηκε κατά 63%, ενώ το IPv4 γράφημα κέρδισε 9% περισσότερα ASes και 14% περισσότερους συνδέσμους. Σε απόλυτους όρους το IPv6 γράφημα πρόσθεσε περίπου 2 χιλιάδες ASes και 5.5 χιλιάδες συνδέσμους.



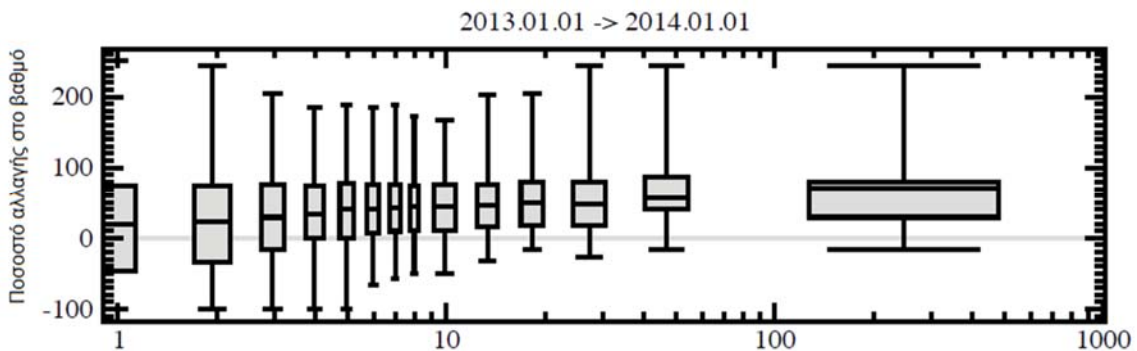
Εικόνα 21: Μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου για το 2013

Τα τρία υψηλότερου βαθμού (περισσότερο συνδεδεμένα) IPv4 και IPv6 ASes στις μετρήσεις του CAIDA παρέμειναν στις ίδιες θέσεις όπως στα δεδομένα του 2013, αλλά υπήρξαν περισσότερες μεταβολές στη κατάταξη ανάμεσα στα χαμηλότερου βαθμού IPv6 ASes απ' ό,τι ανάμεσα σε παρόμοιας κατάταξης IPv4 ASes. Η διαφορά αυτή της μεταβλητότητας συνάδει με τα νεότερα και περισσότερο δυναμικά σχήματα ανάπτυξης του IPv6. Εξετάζοντας τις δυναμικές των αλλαγών στους βαθμούς του κόμβου (αριθμός των γειτόνων) παρέχονται επιπλέον γνώσεις σχετικά με τις τάσεις εξέλιξης του Διαδικτύου. Τα διαγράμματα κουτιών στο σχήμα 22 και 23 απεικονίζουν τις κατανομές των σχετικών AS βαθμού αλλαγών που συνέβησαν ανάμεσα στο 2013 και το 2014. Κάθε κάδος σε αυτά τα γραφήματα περιλαμβάνει όλα τα ASes τα οποία είχαν τον ίδιο βαθμό στα δεδομένα του 2013 ή εκτείνονται σε αρκετές τιμές βαθμού ώστε να περιλαμβάνουν 25 ASes. Για κάθε κάδο, η μαύρη γραμμή στη μέση δείχνει το μέσο ποσοστό της μεταβολής, το κάθετο κουτί βρίσκεται ανάμεσα στη 25^η και 7^η ποσοστιαία τιμή και η κάθετη γραμμή τελειώνει στη 5^η και τη 95^η ποσοστιαία τιμή.



Σχήμα 22: Μεταβολή του βαθμού διέλευσης συνεπαγόμενου σε όλο το IX χώρο διευθύνσεων σε ένα χρόνο σε IPv4

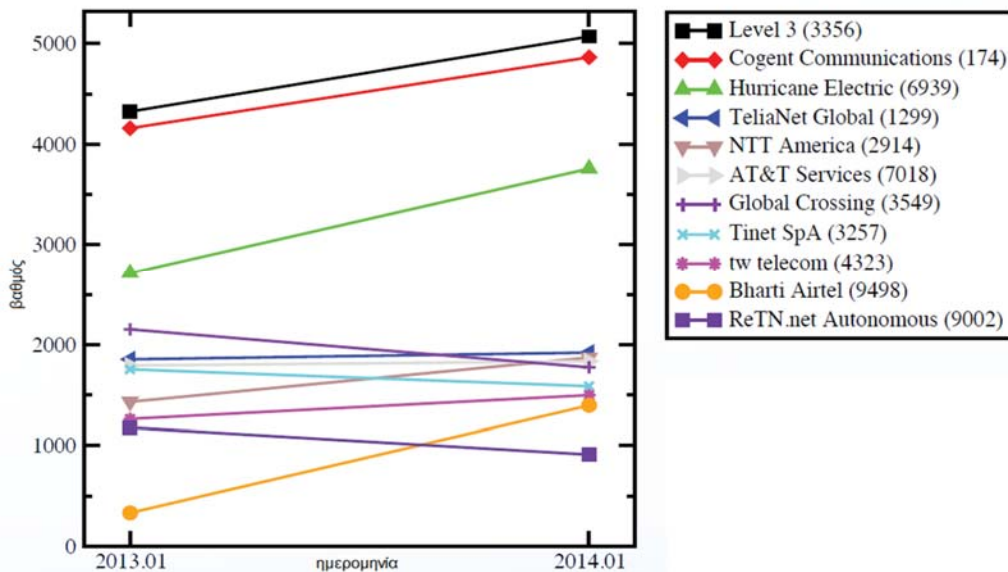
Το σχήμα 22 δείχνει ότι για τα IPv4 ASes με βαθμό μικρότερο από 6, η μέση τιμή των παρατηρούμενων αλλαγών βαθμού είναι 0, υποδεικνύοντας ότι περίπου ίσοι αριθμοί αυτών των μικρών ASes είχαν βαθμό αυξημένο ή μειωμένο. ASes με βαθμούς μεγαλύτερους από 6 είχαν τη τάση να αυξάνουν το βαθμό τους (οι διάμεσοι σε αυτούς τους κάδους είναι πάνω από το 0). Ωστόσο σε όλα τα εύρη βαθμού πάνω από το 25% των IPv4 ASes είχαν βαθμούς μειωμένους ανάμεσα στο 2013 και 2014: το κάτω άκρο σε όλα τα κουτιά στο γράφημα είναι στην αρνητική περιοχή.



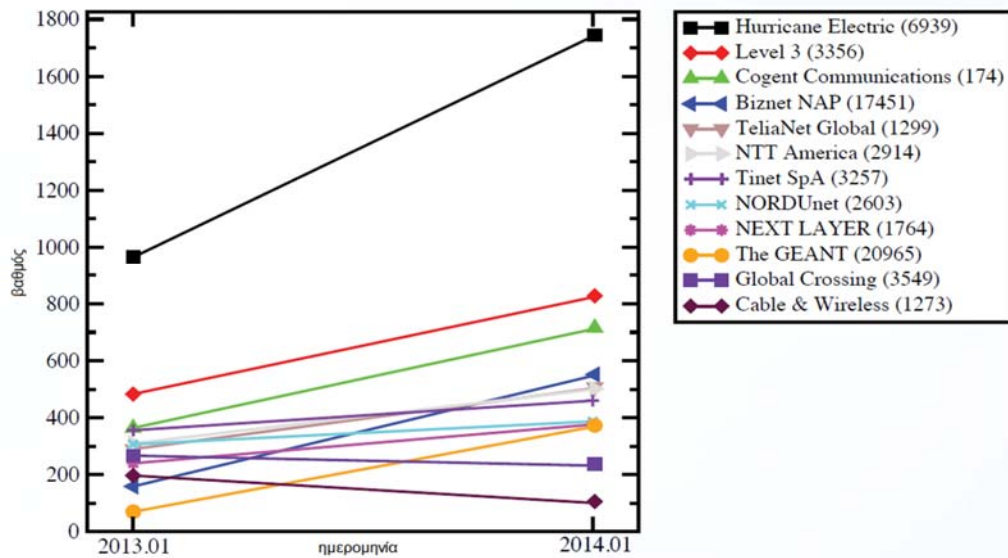
Σχήμα 23: Μεταβολή του βαθμού διέλευσης συνεπαγόμενου σε όλο το IX χώρο διευθύνσεων σε ένα χρόνο σε IPv6

Εν αντιθέσει, στο σχήμα 23 οι διάμεσες γραμμές σε όλους τους κάδους είναι πάνω από 0 υποδεικνύοντας ότι μεταξύ 2013 και 2014 περισσότερα IPv6 ASes είχαν τη τάση να αυξάνουν από το να μειώνουν τους βαθμούς τους σε όλα τα φάσματα βαθμού. Επιπλέον, σε λιγότερο από το 25% των IPv6 ASes που είχαν βαθμούς από 4 και πάνω το 2013, μειώθηκαν οι βαθμοί τους από το 2014. Τα σχήματα 22 και 23 επιβεβαιώνουν τη παρατήρηση του CAIDA ότι η IPv6 inter-AS συνδεσιμότητα μεγάλωσε σχετικά ταχύτερα από εκείνη του IPv4.

Τα σχήματα 24 και 25 δείχνουν τις αλλαγές στη συνδεσιμότητα για τα 11 IPv4 και 12 IPv6 που κατατάσσονται στα κορυφαία 10 με βάση το βαθμό διέλευσης, είτε το 2013 είτε το 2014. Στο IPv4 γράφημα, 8 ASes αύξησαν και 3 μείωσαν το βαθμό τους, η διάμεσος των παρατηρούμενων αλλαγών είναι 17% αυξημένη στο βαθμό. Η μέγιστη αύξηση 321% παρατηρήθηκε για το Bharti Airtel (AS 9498) προωθώντας αυτό το AS στο σύνολο των κορυφαίων 10. Ο βαθμός διέλευσης του RETN (AS 9002) μειώθηκε κατά 22% αφαιρώντας το έτσι από το σύνολο των κορυφαίων 10 το 2014. Στο IPv6 γράφημα, 10 από τα 12 απεικονιζόμενα IPv6 ASes αύξησαν το βαθμό τους, με μια μέση αύξηση 70%. Το Hurricane Electric (AS 6939) παρέμεινε το μεγαλύτερο IPv6 AS, αυξάνοντας το βαθμό του από 966 το 2013 σε 1740 το 2014. Το GEANT (AS 20965) και το Biznet (AS 17451) έδειξαν τη μεγαλύτερη συγκριτικά αύξηση του βαθμού στο IPv6 χώρο, αποκτώντας περίπου 425% και 244% νέους IPv6 γείτονες αντίστοιχα και έτσι εισέρχονται στο σύνολο των κορυφαίων 10. Την ίδια ώρα, το Cable & Wireless (AS 1273) και το Global Crossing (AS 3549) είδαν το IPv6 βαθμό τους να μειώνεται 49% και 13% και ως εκ τούτου να χάνουν τις θέσεις τους στα κορυφαία 10.



Σχήμα 24: Μεταβολή του βαθμού των κορυφαίων ASes σε IPv4 σε ένα χρόνο

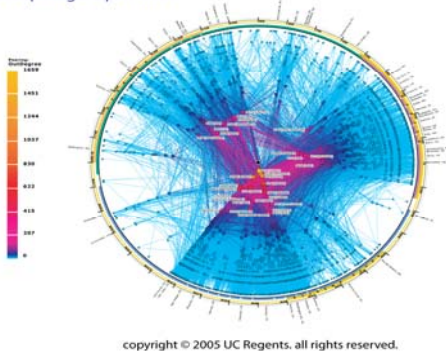


Σχήμα 25: Μεταβολή του βαθμού των κορυφαίων ASes σε IPv6 σε ένα χρόνο

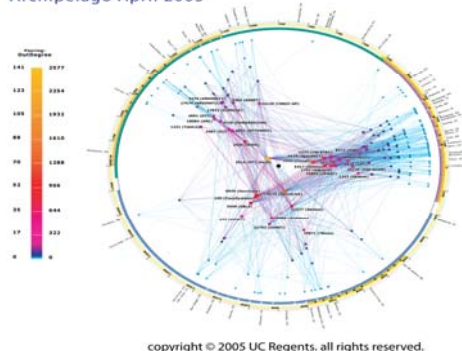
3.4.4.2 IPv4 και IPv6 AS core to 2005

Για να γίνει αντιληπτή η ραγδαία ανάπτυξη του Διαδικτύου στο πέρασμα των χρόνων είναι χρήσιμο να δοθεί μια μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου το 2005 όπου είναι η πρώτη χρονιά για την οποία έχει γίνει λήψη δεδομένων από τον CAIDA όσον αφορά το IPv6 (εικόνα 22). Ο χάρτης IPv4 περιλαμβάνει 926,201 IP διευθύνσεις, 2,000,796 IP συνδέσμους, 865,000 προορισμούς, 50% των παγκόσμιων δρομολογήσιμων προθεμάτων, 10,131 ASes (78% των ASes που υπάρχουν στα RouteViews BGP tables στις 17 Απριλίου 2005) και 29,191 περιόδους ανταλλαγής κίνησης. Αντίστοιχα ο χάρτης IPv6 που περιέχει δεδομένα που συλλέχθηκαν από 17 συστήματα παρακολούθησης σε 3 ηπείρους, περιλαμβάνει 2,913 IPv6 διευθύνσεις, 7,905 IPv6 συνδέσμους, 333 Αυτόνομα Συστήματα (ASes), και 1,304 περιόδους ανταλλαγής κίνησης. Όπως γίνεται εύκολα αντιληπτό η πυκνότητα του Διαδικτύου το 2005 όπως φαίνεται τόσο από το χάρτη IPv4 όσο και από εκείνο του IPv6 ήταν σε πολύ μεγάλο βαθμό μικρότερη από εκείνη που υφίσταται σήμερα.

CAIDA's IPv4 AS Core
AS-level INTERNET GRAPH
Archipelago April 2005



CAIDA's IPv6 AS Core
AS-level INTERNET GRAPH
Archipelago April 2005



Εικόνα 22: Μακροσκοπική εικόνα της IPv4 και IPv6 τοπολογίας Διαδικτύου το 2005

3.5 Συγκριτική Ανάλυση των Διαδικτυακών Τοπολογιών

Οι χάρτες τοπολογίας του Διαδικτύου είναι απαραίτητοι για το χαρακτηρισμό αυτής της κρίσιμης υποδομής και τη κατανόηση των ιδιοτήτων του, της δυναμικής και της εξέλιξής του. Είναι επίσης πολύ σημαντικοί για τη ρεαλιστική μοντελοποίηση, προσομοίωση και την ανάλυση της υποδομής του Διαδικτύου και άλλων πολύπλοκων δικτύων μεγάλης κλίμακας. Οι χάρτες αυτοί μπορούν να κατασκευαστούν για διαφορετικά στρώματα (ή βαθμούς ανάλυσης) π.χ οπτικές ίνες, IP διεύθυνση, δρομολογητή, PoPs, ISP/οργανισμό. Επιπέδου δρομολογητή και PoP επιπέδου χάρτες τοπολογίας μπορούν να ενημερώσουν δυναμικά και να βαθμονομήσουν τις εκτιμήσεις ευαισθησίας. Οι ISP επιπέδου τοπολογίες, μερικές φορές καλούνται AS επιπέδου ή διατομεακές τοπολογίες δρομολόγησης [26] (αν και ένας ISP μπορεί να κατέχει πολλαπλά ASes έτσι ένα γράφημα AS επιπέδου είναι μια ελαφρώς λεπτομερής ανάλυση) και παρέχουν γνώσεις σχετικά με τις τεχνικές, οικονομικές, πολιτικές και ασφάλειας ανάγκες της σε μεγάλο βαθμό ανεξέλεγκτης ανταλλαγής κίνησης οικοσυστήματος.

Κατά τη τελευταία δεκαετία, πολλές μελέτες έχουν επικεντρωθεί στη διάρθρωση των παρατηρήσιμων τοπολογιών Διαδικτύου [27] [28] συμπεριλαμβανομένου σημαντικών διαφωνιών σχετικά με τη ποιότητα των δεδομένων και των σχετικών συμπερασμάτων [29]. Επί της ουσίας διαφορετικές απόψεις σχετικά με τα αποτελέσματα του Διαδικτύου από διαφορετικές μεθόδους μέτρησης της τοπολογίας. Ως προς τα στοιχεία των μετρήσεων σε αντικείμενα και τις ιδιαιτερότητες των συλλεγόμενων δεδομένων είναι ζωτικής σημασίας για την αντικειμενική αξιολόγηση του πεδίου εφαρμογής και την εγκυρότητα των προκυπτόντων χαρτών Διαδικτύου. Το 2006 στη μελέτη των μελών του CAIDA [30], έγινε σύγκριση των AS τοπολογίας γράφων που προέκυψαν από τρεις διαφορετικές πηγές δεδομένων: traceroute (με χρήση του skitter, της προηγούμενης υποδομής του CAIDA για ενεργή μέτρηση), BGP (RouteViews) και των IRR δεδομένων. Στο πρόγραμμα αυτό τα μέλη του CAIDA επέκτειναν το πεδίο εφαρμογής της παρούσας συγκριτικής ανάλυσης για να συμπεριλάβει δύο επιπλέον τύπους γραφημάτων (IP διεπαφής και επιπέδου δρομολογητή γραφήματα) και πέντε επιπλέον πηγές δεδομένων (RIPE-RIS, Ark-IPv4-traceroute, iPlane, DIMES και IRL). Τα μέλη του CAIDA πιστεύουν ότι είναι η πιο διεξοδική μελέτη μέχρι στιγμής συγκρίνοντας και ερμηνεύοντας τα δομικά χαρακτηριστικά των τοπολογιών που έχουν συναχθεί από τις καλύτερες διαθέσιμες πηγές δεδομένων. Τέλος ελπίζουν ότι η προκύπτουσα συγκριτική ανάλυση διευκολύνει τη πιο ενημερωμένη επιλογή τοπολογικών συνόλων δεδομένων για την υποστήριξη συγκεκριμένων αναγκών έρευνας ή ανάλυσης.

3.5.1 Πηγές Δεδομένων της Μελέτης

Πίνακας 6: Σύνολα δεδομένων που χρησιμοποιούνται στη μελέτη

	διάστημα	μέρες	τύπος	επίπεδο γράφου			πλεονεκτική θέση		
				IP	Router	AS	συλλέκτες	ASes	χώρες
DIMES	2011.04.04 - 2011.04.17	14 days	traceroute	X		X	947		
iPlane	2011.04.06 - 2011.04.20	15 days	traceroute	X	X	X	251*	190	40
Ark IPv4 All Prefix /24	2011.04.01 - 2011.04.15	15 days	traceroute	X	X	X	54	54	29
RouteViews2	2011.01.16 - 2011.01.20	4 days	BGP			X	1	33	11
BGP Full	2011.01.16 - 2011.01.20	4 days	BGP			X	19	336	21
IRL	2011.04.01 - 2011.04.15	15 days	BGP			X	N/A**	N/A**	N/A**
RIPE WHOIS	2009.04.20 - 2011.04.20	2 years	IRR			X	1	20,905	183

Ο αριθμός των μοναδικών επόμενου hop IP διευθύνσεων για τους 517 PlanetLab εξυπηρετητές (πολλές περιοχές έχουν πολλαπλούς κόμβους) (*)

Η IRL τεκμηρίωση δεν διευκρινίζει πόσες πηγές ήταν στο σύνολο δεδομένων που χρησιμοποιήθηκε από τα μέλη του CAIDA (**)

3.5.1.1 Δεδομένα του traceroute

Υπόβαθρο για πολλές μελέτες Διαδικτυακής τοπολογίας είναι τα σύνολα δεδομένων που έχουν συλλεχθεί από βασισμένες στο traceroute μετρήσεις. Με την αποστολή διερευνητικών πακέτων στο προορισμό, οι traceroute μέθοδοι καλύπτουν μια ακολουθία από διεπαφές κατά μήκος της διαδρομής προς τα εμπρός από τη πηγή σε ένα συγκεκριμένο προορισμό. Η πιο διαδεδομένη τεχνική ανίχνευσης χρησιμοποιεί τα πακέτα ICMP, αν και η διερεύνηση η βασισμένη στο UDP ή στο TCP επίσης χρησιμοποιείται [31].

Σύμφωνα με το πρότυπο πρωτόκολλο IP, πριν από τη διαβίβαση ενός πακέτου κάθε δρομολογητής κατά μήκος της διαδρομής μειώνει το TTL πεδίο στην επικεφαλίδα του πακέτου. Εάν το πεδίο είναι μηδέν, ο δρομολογητής ρίχνει το πακέτο και στέλνει μία απάντηση υπέρβασης χρόνου στη IP διεύθυνση πηγής του πακέτου. Οι traceroute διερευνητικές μεθοδολογίες συναγάγουν το IP επιπέδου μονοπάτι προώθησης διαμέσου του δικτύου στέλνοντας μία σειρά από πακέτα στον ίδιο τελικό προορισμό, κάθε ένα με αύξηση των TTL τιμών και καταγράφουν τις διευθύνσεις των δρομολογητών κατά μήκος της διαδρομής επιστροφής των ICMP μηνυμάτων υπέρβασης του χρόνου. Υπάρχει κάποια μη τυποποιημένη μεταβολή στη διεπαφή που χρησιμοποιείται ως η πηγή των IP διευθύνσεων στα ICMP πακέτα απόκρισης (μερικές φορές είναι η διεύθυνση IP της εξερχόμενης διεπαφής για τη διαδρομή επιστροφής, αντί της διεπαφής της διαδρομής προώθησης) αλλά είναι πάντα μία IP διεύθυνση στο δρομολογητή όπου το TTL έληξε. Η εκτέλεση traceroute διερεύνησης από πολλαπλά σημεία προς πολλαπλούς προορισμούς αποκαλύπτει μια πληθώρα από IP διεπαφές και συνδέσμους ανάμεσά τους.

Προκειμένου να διαμορφώσουν ένα περισσότερο ρεαλιστικό χάρτη των πραγματικών φυσικών συσκευών (δρομολογητές) από αυτά τα ανεπεξέργαστα traceroute δεδομένα, τα μέλη του CAIDA πρέπει να εκτιμήσουν ποιές IP διευθύνσεις στα traceroute μονοπάτια ανήκουν στον ίδιο δρομολογητή, μία διεργασία γνωστή ως IP διεύθυνση επίλυσης ψευδωνύμου. Ένας δρομολογητής έξ ορισμού έχει τουλάχιστον δύο διεπαφές, με τους δρομολογητές του πυρήνα του Διαδικτύου να έχουν πιθανώς εκατοντάδες διεπαφές. Η διεργασία της επίλυσης ψευδωνύμων δίνει μία κατά προσέγγιση επιπέδου δρομολογητή τοπολογία. Κάποιος μπορεί επίσης να δημιουργήσει AS επιπέδου γραφήματα από προερχόμενα από το traceroute IP επιπέδου δεδομένα. Το πρώτο βήμα σε αυτή τη διεργασία είναι η χαρτογράφηση των IP διευθύνσεων στα ASes ως ακολούθως. Κάθε IP διεύθυνση ανήκει σε ένα πρόθεμα διεύθυνσης το οποίο είχε αρχικά ανακοινωθεί από ένα ανεξάρτητο φορέα δρομολόγησης στο παγκόσμιο σύστημα δρομολόγησης, που ονομάζεται αυτόνομο σύστημα (AS). Η μετατροπή του IP επιπέδου δεδομένου σε ένα AS επιπέδου γράφημα απαιτεί τη συλλογή του αρχικού AS για κάθε πρόθεμα από τα BGP δεδομένα, να σημειώνεται κάθε IP διεύθυνση με το αρχικό της AS και να χρησιμοποιηθούν αυτές οι σημειώσεις για τη συναγωγή της αντιστοιχίας των AS συνδέσμων με κάθε παρατηρούμενο από το traceroute IP σύνδεσμο. Εναλλακτικά, μπορεί κανείς να ξεκινήσει με μία επιπέδου δρομολογητή τοπολογία προερχόμενη από την επίλυση ψευδωνύμου, να σημειώσει κάθε

δρομολογητή με το AS στο οποίο ανήκει και να συναγάγει την αντιστοιχία των AS συνδέσμων με κάθε σύνδεσμο στην επιπέδου δρομολογητή τοπολογία.

Για τη μελέτη αυτή χρησιμοποιήθηκαν traceroute δεδομένα από τρεις πηγές (πίνακας 6): DIMES, iPlane, και Ark IPv4 All Prefix /24. Το DIMES είναι ένα καταμετρημένο επιστημονικό ερευνητικό πρόγραμμα που διευθύνεται από το πανεπιστήμιο του Τελ Αβίβ. Οι traceroute μετρήσεις εκτελούνται παράλληλα από εθελοντές που έχουν αναπτύξει το λογισμικό μέτρησης netDimes στους προσωπικούς τους υπολογιστές. (1065 σημεία απεικονίζονται στο πίνακα 6, αν και δεν ήταν εφικτό να βρεθεί πόσα σημεία ήταν ενεργά στο υπο-διάστημα στο οποίο έγινε η σύγκριση). Το iPlane είναι ένα ερευνητικό πρόγραμμα συλλογής τοπολογίας που αναπτύχθηκε από το πανεπιστήμιο του Ουάσιγκτον και λειτουργεί στο PlanetLab, ένα παγκόσμιο δίκτυο από ακαδημαϊκής έρευνας εξυπηρετητές (κατά το χρονικό διάστημα της μελέτης υπήρχαν 251 σημεία με 517 συστήματα παρακολούθησης, τα περισσότερα σημεία είχαν πολλαπλά συστήματα παρακολούθησης). Το iPlane κατασκευάζει ένα σχολιασμένο (annotated) χάρτη Διαδικτυακής τοπολογίας και προβλέπει την από άκρη σε άκρη απόδοση συνθέτοντας τη μετρούμενη απόδοση των τμημάτων των γνωστών Διαδικτυακών μονοπατιών, ιδιαίτερα του «πυρήνα» του δικτύου κορμού (backbone) του Διαδικτύου που περιέχει τα περισσότερα χρησιμοποιούμενα μονοπάτια. Τα Ark IPv4 All Prefix /24 είναι traceroute δεδομένα που έχουν συλλεχθεί από τη Ark υποδομή μετρήσεων του CAIDA η οποία κατά τη διάρκεια αυτής της αναφοράς αποτελείται από 54 ειδικούς υπολογιστές που λειτουργούν ως σημεία και ελέγχονται από ένα κεντρικό εξυπηρετητή στον CAIDA. Τα συστήματα παρακολούθησης του Ark προσπαθούν να διερευνούν μία μόνο τυχαία διεύθυνση σε κάθε δρομολογούμενο IPv4 /24 πρόθεμα στο Διαδίκτυο, ένας πλήρης κύκλος μέσα από το δρομολογούμενο IPv4 χώρο διευθύνσεων, λαμβάνει χώρα σε περίπου 48 ώρες.

3.5.1.2 Δεδομένα BGP για Τοπολογίες AS Επιπέδου

Τα ASes χρησιμοποιούν το Border Gateway Protocol (BGP) για να ανταλλάσσουν πληροφορίες δρομολόγησης στο Διαδίκτυο [3]. Κάθε BGP δρομολογητής διατηρεί ένα πίνακα από IP-πρόθεμα-σε-AS αντιστοιχίσεις που ορίζει τη προσβασιμότητα στα ASes περιγράφοντας μία «αλυσίδα (chain)» ή φορέα πορείας των ASes. Κάποιος μπορεί να αντλήσει ένα AS επιπέδου γράφημα του Διαδικτύου απευθείας από αυτά τα BGP δεδομένα. Δύο προγράμματα αποθήκευσης συλλέγουν και αρχειοθετούν BGP πίνακες δρομολόγησης για έρευνα: το Route Views διοικείται από το πανεπιστήμιο του Όρεγκον και η συλλογή RIS παρέχεται από το RIPE NCC. Κάθε ομότιμος εισφέρει ένα BGP πίνακα που αποθηκεύει ένα σύνολο από IP προθέματα δρομολόγησης και το υπολογισμένο καλύτερο μονοπάτι από τον εν λόγω ομότιμο σε κάθε πρόθεμα. Πρώτη πηγή των δεδομένων BGP για τη παρούσα μελέτη για τα μέλη του CAIDA είναι ο μοναδικός Route Views εξυπηρετητής με το μεγαλύτερο αριθμό ομότιμων, ο RouteViews2 (με 33 ASes ομότιμους). Η δεύτερη πηγή ο BGP Full, είναι ένας συνδυασμός από πίνακες δρομολόγησης από 5 Route Views εξυπηρετητές και 14 RIPE-NCC RIS εξυπηρετητές, δηλαδή όλοι οι διαθέσιμοι εξυπηρετητές στις 1-14 Ιανουαρίου 2011.

Η δημιουργία ενός BGP βασισμένου AS επιπέδου γράφου που χρησιμοποιεί το μέγιστο διαθέσιμο αριθμό από συλλέκτες για ένα συγκεκριμένο χρονικό διάστημα είναι η ίδια μέθοδος που χρησιμοποιούν τα μέλη του CAIDA για να παράγουν τους AS επιπέδου γράφους για

την εργασία κατάταξης των ASes [ASR]. Η τρίτη πηγή τους των BGP δεδομένων είναι η συλλογή του εργαστηρίου Διαδικτυακής Έρευνας του UCLA (IRL), η οποία περιλαμβάνει BGP δεδομένα από Route Views, PCH, traceroute.org, και το Looking Glass Wiki. Η τεκμηρίωση του IRL δεν διευκρινίζει πόσες πηγές συνέβαλαν στο σύνολο δεδομένων που χρησιμοποιήθηκε.

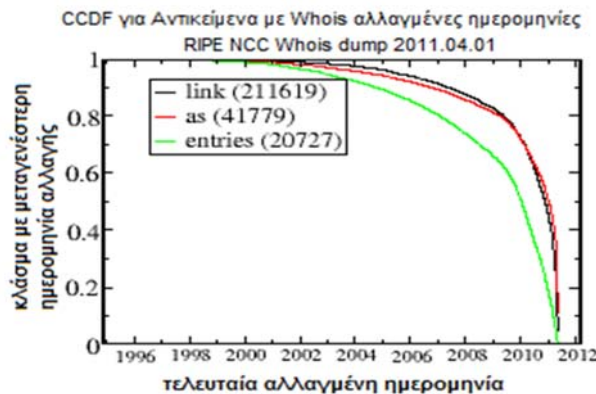
3.5.1.3 Δεδομένα IRR

Τα Περιφερειακά Διαδικτυακά Μητρώα RIR υποστηρίζουν πρόσβαση μέσω αιτημάτων στις βάσεις δεδομένων τους με διευθύνσεις του Διαδικτύου εκχωρώντας πληροφορίες μέσω του WHOIS αιτήματος και του πρωτόκολλου απόκρισης. Οι εξυπηρετητές που παρέχουν πρόσβαση μέσω αιτημάτων ονομάζονται WHOIS εξυπηρετητές. Μαζί με τις πληροφορίες ανάθεσης, τουλάχιστον μια RIR βάση δεδομένων (RIPE) αποθηκεύει εθελοντικά συμβαλλόμενες και μερικές φορές συντηρημένες πληροφορίες πολιτικών δρομολόγησης όπως το σύνολο των ανακοινώσεων που ένα AS δέχεται από τα γειτονικά του ASes. Αυτή η πληροφορία είναι χρήσιμη για τους ISPs στην ανίχνευση άκυρων (invalid) μονοπατιών (π.χ μονοπάτια τα οποία δεν ακολουθούν τις γνωστές πολιτικές των ASes στο μονοπάτι). Κάποιος μπορεί επίσης να δημιουργήσει έναν AS επιπέδου γράφο της Διαδικτυακής συνδεσιμότητας από αυτούς τους AS συνδέσμους.

Το 2004 δύο μέλη του CAIDA ανέλυσαν τη ποσότητα και τη ποιότητα των πληροφοριών που ήταν διαθέσιμα από διαφορετικές RIR βάσεις δεδομένων [32] και ανακάλυψαν ότι το RIPE NCC διατηρεί τη μεγαλύτερη βάση δεδομένων με τις πιο ακριβείς τοπολογικές πληροφορίες. Επίσης ανακάλυψαν ότι μόνο το 28% των ASes, σχεδόν στο σύνολό τους εγγεγραμμένα στο RIPE μητρώο, είχαν καταχωρημένες πολιτικές οι οποίες ήταν ταυτόχρονα εσωτερικά σύμφωνες και σύμφωνες με τους Route Views BGP πίνακες δρομολόγησης κατά τη στιγμή της ανάλυσης. Ώς εκ τούτου, τα μέλη του CAIDA επιλέγουν τη WHOIS βάση δεδομένων του RIPE NCC ως πηγή των IRR δεδομένων για ένα AS επιπέδου γράφο. Το σημαντικότερο πρόβλημα με αυτά τα δεδομένα είναι ότι οι WHOIS βάσεις δεδομένων είναι μη αυτόματα και εθελοντικά διατηρημένες, χωρίς την απαίτηση της ενημέρωσης των καταχωρημένων πληροφοριών. Έτσι πολλά αρχεία είναι πιθανόν παρωχημένα και θα πρέπει να αποφασιστεί πώς θα φιλτραριστούν παλιές και αναξιόπιστες πληροφορίες.

Τα μέλη του CAIDA έχουν λάβει τα μεταποθηκευμένα δεδομένα της WHOIS βάσης δεδομένων του RIPE NCC στις 20 Αυγούστου 2011 και χρησιμοποίησαν την ακόλουθη μέθοδο για να διατηρήσουν αρκετά καινούριες τις καταχωρήσεις. Όταν τα WHOIS αρχεία έχουν ενημερωθεί, τα αλλαγμένα πεδία τους τυπικά δείχνουν την ημερομηνία αλλαγής, παρόλο που το γεγονός αυτό δεν διευκρινίζει εάν ήταν οι πληροφορίες δρομολόγησης που ενημερώθηκαν ή κάποιο άλλο πεδίο. Παρόλα αυτά μια πρόσφατη ημερομηνία στο αλλαγμένο πεδίο τουλάχιστον σημαίνει ότι κάποιος έχει τσεκάρει την εγγραφή, αυξάνοντας τη πιθανότητα ότι οι πληροφορίες των πολιτικών δρομολόγησης εξακολουθούν να ισχύουν ή αποφάσισαν ότι η εγγραφή θέλει πολύ δουλειά για να διορθωθεί. Η πράσινη γραμμή στο σχήμα 26 δείχνει το κλάσμα των εγγραφών της WHOIS βάσης δεδομένων στο RIPE NCC που έχουν αλλαγμένα τα πεδία δεδομένων τους σε μια τιμή ίση ή μεγαλύτερη από την ημερομηνία που δίνεται στο x άξονα. Η κόκκινη γραμμή δείχνει το κλάσμα των ASes και η μαύρη γραμμή το κλάσμα των AS συνδέσμων που βρίσκονται σε αυτές τις «αλλαγμένες μετά τη συγκεκριμένη ημερομηνία» καταχωρήσεις.

Όσο παλαιότερη η ημερομηνία αλλαγής, τόσο μεγαλύτερο το κλάσμα των ASes και των AS συνδέσμων σε αυτές τις ASes εγγραφές που άλλαξαν μετά από αυτή την ημερομηνία. Το κομβικό σημείο είναι περίπου τον Ιούνιο του 2009, με μόνο το 25% των ASes και των AS συνδέσμων να έχουν αλλαγμένες ημερομηνίες τα προηγούμενα 13 χρόνια έναντι του 75% κατά τα επόμενα 2 έτη. Ός εκ τούτου, τα μέλη του CAIDA διατήρησαν όλες τις εγγραφές με τις λιγότερο των 2 χρόνων αλλαγμένες ημερομηνίες ως πηγή δεδομένων για την ανάλυσή τους, η οποία περιλαμβάνει τα στοιχεία της IRR συνδεσιμότητας για 20.905 ASes (από πάνω από 39.000 ASes στο σύνολο). Όσο τα αρχεία βάσεων δεδομένων δείχνουν μόνο συνδέσμους από κάθε AS προς τους άμεσους γείτονές του, κάθε AS ενεργεί ως ένα πλεονεκτικό σημείο παρέχοντας μια τοπική προβολή του δικτύου 1-hop μακριά (away).



Σχήμα 26: Στατιστικά στοιχεία των καταχωρήσεων στη RIPE NCC Whois βάση δεδομένων

3.5.2 Τοπολογικές Μετρήσεις

Τα μέλη του CAIDA έχουν επιλέξει τα παρακάτω τέσσερα βασικά στατιστικά χαρακτηριστικά για τη σύγκριση μεταξύ των διαθέσιμων γραφημάτων τοπολογίας του Διαδικτύου. Αυτές οι μετρήσεις είναι οι πιο δημοφιλείς στη βιβλιογραφία δικτύωσης [30] και ακόμα είναι καθορισμένο όπως φαίνεται ότι η αναπαραγωγή αυτών των μετρήσεων είναι αρκετή για να συλλάβει όλα τα απαραίτητα τοπολογικά χαρακτηριστικά των AS και δρομολογητή επιπέδου τοπολογιών.

- **Μέσος Βαθμός Κόμβου:** Οι δύο πιο βασικές ιδιότητες γραφήματος είναι ο αριθμός των κόμβων n (επίσης αναφέρεται ως μέγεθος γραφήματος) και ο αριθμός των συνδέσμων m . Η αναλογία των συνδέσμων σε κόμβους ορίζει το μέσο βαθμό κόμβου $k = 2m/n$. Ο μέσος βαθμός κόμβου είναι το πιο χοντρικό χαρακτηριστικό συνδεσιμότητας μιας δεδομένης τοπολογίας. Τα δίκτυα με υψηλότερο k είναι καλύτερα συνδεδεμένα κατά μέσο όρο και κατά συνέπεια πιθανόν να είναι πιο ισχυρά και αποτελεσματικά αλλά και δυνητικά ευάλωτα καθώς η διάχυση του κακόβουλου λογισμικού είναι επίσης πιο αποτελεσματική.
- **Βαθμός Διανομής:** Έστω $n(k)$ είναι ο αριθμός των κόμβων βαθμού k (k -βαθμού κόμβοι). Ο βαθμός διανομής κόμβου είναι η πιθανότητα ένας τυχαία επιλεγμένος κόμβος να είναι k -βαθμού: $P(k) = n(k)/n$. Ο βαθμός διανομής περιέχει περισσότερες πληροφορίες σχετικά με τη συνδεσιμότητα σε ένα συγκεκριμένο γράφημα από το μέσο βαθμό. Δοθέντος μιας ειδικής μορφής του $P(k)$ τα μέλη του CAIDA μπορούν πάντα να επαναφέρουν το μέσο βαθμό με $k = \sum_{k=1}^{k_{max}} kP(k)$, όπου

k_{max} είναι ο μέγιστος βαθμός κόμβου στο γράφημα. Στην έκθεση αυτή τα μέλη του CAIDA αναλύουν και συγκρίνουν τη συμπληρωματική αθροιστική συνάρτηση κατανομής (CCDF) του βαθμού του κόμβου, η οποία δείχνει το κλάσμα των κόμβων που έχουν βαθμό ίσο ή μεγαλύτερο από τη τιμή του ορίσματος. Ο βαθμός διανομής είναι το πιο συχνά χρησιμοποιούμενο χαρακτηριστικό τοπολογίας. Οι περισσότεροι ερευνητές δικτύου συμφωνούν ότι ο βαθμός διανομής $P(k)$ για τα AS επιπέδου γραφήματα του Διαδικτύου ακολουθεί μια συνάρτηση δύναμης $P(k) = k^{-\gamma}$ με εκθέτη γ να είναι κοντά στο 2 [30]. Καθώς ο νόμος της δύναμης είναι μια ιδιαίτερα μεταβλητή κατανομή, ο βαθμός του κόμβου είναι ένα σημαντικό χαρακτηριστικό ενός μεμονωμένου κόμβου. Τα μέλη του CAIDA ελέγχουν αν αυτή η προσέγγιση του νόμου της δύναμης ταιριάζει στα δεδομένα τους και αναφέρουν τις τιμές του εκθέτη γ . Τέλος χρησιμοποιούν δύο μεθόδους για να βρουν το κατάλληλο: ελαχίστων τετραγώνων τοποθέτηση του βαθμού CCDF και μέγιστης πιθανότητας τοποθέτηση του βαθμού αλληλουχίας.

- **Μέσος Βαθμός Γείτονα:** Έστω $a(i,k)$ είναι ο μέσος βαθμός των άμεσων γειτόνων του i -οστού κόμβου βαθμού k . Τότε ο μέσος βαθμός γείτονα βαθμού k είναι ο μέσος όρος για όλους τους κόμβους $i=1 \dots i_k$ με βαθμό k : $a_{nn}(k) = \sum_{i=1}^{i_k} a(i,k)/n(k)$. Ο μέσος βαθμός γείτονα είναι συνολική στατιστική του κοινού βαθμού διανομής. Δείχνει πόσα ASes ενός ορισμένου βαθμού κατά προτίμηση συνδέονται σε υψηλού ή χαμηλού βαθμού ASes. Σε ένα γράφημα πλήρους πλέγματος, $a_{nn}(k)$ φθάνει τη μέγιστη δυνατή τιμή του $n-1$. Ως εκ τούτου, για την ομοιόμορφη σύγκριση γραφήματος τα μέλη του CAIDA θα σχεδιάσουν κανονικοποιημένες τιμές $a_{nn}(k)/(n-1)$.
- **Ομαδοποίηση (clustering):** Έστω $m_{nn}(k)$ είναι ο μέσος αριθμός των συνδέσμων ανάμεσα στους γείτονες των k -βαθμού κόμβων. Η τοπική ομαδοποίηση είναι ο λόγος αυτού του αριθμού προς το μέγιστο δυνατό αριθμό αυτών των συνδέσμων: $C(k) = 2m_{nn}(k)/(k-1)$. Εάν δύο γείτονες ενός κόμβου συνδεθούν, τότε αυτοί οι τρεις κόμβοι μαζί σχηματίζουν ένα τρίγωνο (3-κύκλοι). Επομένως, εξ ορισμού η τοπική ομαδοποίηση είναι ο μέσος αριθμός από 3-κύκλους που περιλαμβάνει k -βαθμού κόμβους. Η μέση τοπική ομαδοποίηση είναι ο μέσος όρος των $C(k)$ επί όλων των τιμών των κόμβων βαθμού k : $C = \sum C(k)P(k)$. Η ομαδοποίηση εκφράζει τη τοπική αξιοπιστία του γραφήματος: όσο υψηλότερη η τοπική ομαδοποίηση ενός κόμβου, τόσο πιο διασυνδεδεμένοι είναι οι γείτονές της, αυξάνοντας έτσι τη ποικιλομορφία (diversity) διαδρομής τοπικά γύρω από το κόμβο.

Για όλα τα επίπεδα εστίασης (IP, δρομολογητής, AS) εξετάζεται πώς διαφορετικές πηγές δεδομένων παράγουν γραφήματα με διαφορετικά δομικά χαρακτηριστικά.

3.5.3 Κατασκευή Γραφημάτων τοπολογίας Διαδικτύου από τα Δεδομένα

3.5.3.1 IP-Επιπέδου Γραφήματα

Ένα IP διεπαφής- επιπέδου γράφημα κατασκευάζεται εξάγοντας IP συνδέσμους απευθείας από το traceroute αποτέλεσμα: δύο IP διευθύνσεις συνάγονται για να σχηματίσουν ένα σύνδεσμο εάν παρατηρήθηκαν δίπλα η μία στην άλλη σε ένα traceroute αποτέλεσμα. Το έργο DIMES δεν δημοσιεύει τα πλήρη traceroute μονοπάτια που μετρήθηκαν από τους netDimes πελάτες, αλλά εξάγουν από αυτές τις μετρήσεις ένα σύνολο τέτοιων συναγόμενων IP συνδέσμων, παράγοντας ένα IP-επιπέδου γράφημα που αναφέρεται ως DIMES IP. Εν αντιθέσει, τα iPlane και Ark IPv4 All Prefix /24 δεδομένα περιλαμβάνουν ένα πλήρες σύνολο των παρατηρούμενων IP προώθησης μονοπατιών. Προκειμένου να αποκτήσει ένα IP-

επιπέδου γράφημα από αυτά τα δεδομένα ένας ερευνητής πρέπει να αναλύσει τα ακατέργαστα μονοπάτια σε IP συνδέσμους. Παρόλο που είναι εννοιολογικά απλό να απαριθμήσεις κάθε ζεύγος από διπλανές IP διευθύνσεις σε ένα μονοπάτι που έχει συλλεχθεί, η απλότητα εξαφανίζεται μπροστά στα εκατομμύρια των πραγματικού κόσμου παρατηρήσεων. Τα ακατέργαστα μονοπάτια μπορεί να περιλαμβάνουν μη ανταποκρινόμενα hops, βρόχους, ιδιωτικές ή bogon διευθύνσεις και άλλες παρατυπίες. Διαφορετικές μέθοδοι χειρισμού αυτών των παρατυπιών θα επιφέρουν διαφορετικά αποτελέσματα στη τοπολογία που προκύπτει. Για παράδειγμα, ένα μη-αποκρινόμενο hop εμφανίζεται σε ένα traceroute μονοπάτι όταν ένας δρομολογητής προωθεί πακέτα αλλά δεν παράγει ένα μήνυμα χρονικής υπέρβασης όταν "πέφτει" ένα πακέτο. Στη περίπτωση αυτή το προκύπτον ίχνος θα έχει ένα χάσμα μεταξύ των δύο γνωστών διευθύνσεων IP στις δύο πλευρές του μη ανταποκρινόμενου δρομολογητή. Στο αποτέλεσμα του traceroute αυτά τα hops αναπαρίστανται συνήθως από έναν αστερίσκο ("*").

Τα μέλη του Caida χρησιμοποίησαν τον ακόλουθο βασικό αλγόριθμο επεξεργασίας ίχνους για να δημιουργήσουν τα Ark IPv4Pref IP και iPlane IP γραφήματα από τα Ark All Prefix /24 και iPlane σύνολα δεδομένων. Για κάθε εντοπισμένο μονοπάτι στο σύνολο δεδομένων ξεκινάνε με ένα κενό πίνακα και συμπληρώνουν τις καταχωρήσεις του πίνακα καθώς επεξεργάζονται κάθε hop ακολουθώντας τους παρακάτω κανόνες:

- Αγνοούνται αποκρίσεις από τον προορισμό.
- Αν πολλαπλές αποκρίσεις βρέθηκαν για το ίδιο hop, χρησιμοποιείται η τελευταία.
- Στη πρώτη επαναλαμβανόμενη IP διεύθυνση, συμπεραίνεται ένας βρόχος και κόβεται το μονοπάτι στο hop λίγο πριν το επαναλαμβανόμενο.
- Οι ιδιωτικές διευθύνσεις αντιμετωπίζονται ως μη ανταποκρινόμενες.
- Απορρίπτονται οι IPs χωρίς διπλανά hops.

Αφού γίνει η επεξεργασία κάθε ίχνους, δημιουργούνται IP σύνδεσμοι μεταξύ των υπόλοιπων διπλανών hops με IP διευθύνσεις. Δεν δημιουργούνται σύνδεσμοι προς ή σε μη ανταποκρινόμενα hops.

3.5.3.2 Γραφήματα Επιπέδου Δρομολογητή

Σχετικές εργασίες για τις τεχνικές επίλυσης ψευδώνυμων: Η διεργασία ταυτοποίησης των IP διευθύνσεων που ανήκουν στον ίδιο δρομολογητή ονομάζεται επίλυση ψευδώνυμων. Μια ποικιλία από τεχνικές έχουν αναπτυχθεί και εφαρμοστεί για αυτό το έργο. Το Mercator και οι παρόμοιες με το Mercator τεχνικές προσπαθούν να ταυτοποιήσουν ψευδώνυμα στέλνοντας ένα διερευνητικό πακέτο σε μία αχρησιμοποίητη θύρα σε μία διασύνδεση και να συλλέξουν τα μηνύματα σφάλματος που προκύπτουν. Η διερεύνηση μίας διασύνδεσης και η λήψη του σφάλματος από μία διαφορετική διεπαφή είναι μία ισχυρή ένδειξη ότι οι δύο διασυνδέσεις ανήκουν στον ίδιο δρομολογητή. Ωστόσο όταν εφαρμόζεται σε κλίμακας Διαδικτύου τοπολογίες αυτή η μεθοδολογία παράγει ένα υψηλό ποσοστό από ψευδώς θετικά ζεύγη ψευδώνυμων, για παράδειγμα λόγω συσκευών δικτύου στο μονοπάτι απόκρισης [33]. Άλλες τεχνικές εφαρμόζουν διαφορετικές ιδιότητες από τα υπάρχοντα Διαδικτυακά πρωτόκολλα για την μετατροπή των διεπαφών σε δρομολογητές. Συμμαχία συνάγει ότι δυο διευθύνσεις είναι ψευδώνυμα εάν διερευνητικά πακέτα που στέλνονται σε αυτές παράγουν αποκρίσεις με αυξανόμενες αλλά κατάλληλα κοντινές IP ID τιμές, καθώς το IP ID πεδίο αυξάνεται με κάθε πακέτο που στάλθηκε από το δρομολογητή. Το RadarGun ένα εργαλείο για τον εντοπισμό ψευδώνυμων επιπλέον βελτιώνει αυτή τη τεχνική αναζητώντας για

ομοιότητες στις IP ID χρονικές σειρές που συλλέγονται από πολλές διευθύνσεις. Το εργαλείο MIDAR επέκτεινε τις τεχνικές ταχύτητας IP του RadarGun εφαρμόζοντας μια εξαιρετικά ακριβή δοκιμή σύγκρισης ID βασισμένο στη μονοτονία παρά στην εγγύτητα, ενσωματώνοντας πολλαπλές διερευνητικές μεθόδους από πολλαπλά πλεονεκτικά σημεία και χρησιμοποιώντας ένα καινοτόμο ολισθαίνοντας παραθύρου διερευνητικού προγραμματισμού αλγόριθμο ο οποίος αυξάνει την επεκτασιμότητα στη Διαδικτυακή κλίμακα των εκατομμυρίων διευθύνσεων IP. Τα εργαλεία APAR και karar χρησιμοποιούν προηγμένες τεχνικές ανάλυσης γραφημάτων [34] για να συνάγουν υποδίκτυα συνδεδεμένων δρομολογητών και από αυτά ψευδώνυμα.

Τεχνικές επίλυσης ψευδωνύμων που χρησιμοποιούνται στη σύγκριση συνόλων δεδομένων: Το DIMES χρησιμοποιεί μια παρόμοια με το Mercator τεχνική για την επίλυση ψευδωνύμων. Εξαιτίας του υψηλού ποσοστού των ψευδώς θετικών αυτής της παλαιότερης μεθόδου αποφασίστηκε να μην χρησιμοποιούνται τα δεδομένα από το DIMES στις συγκρίσεις.

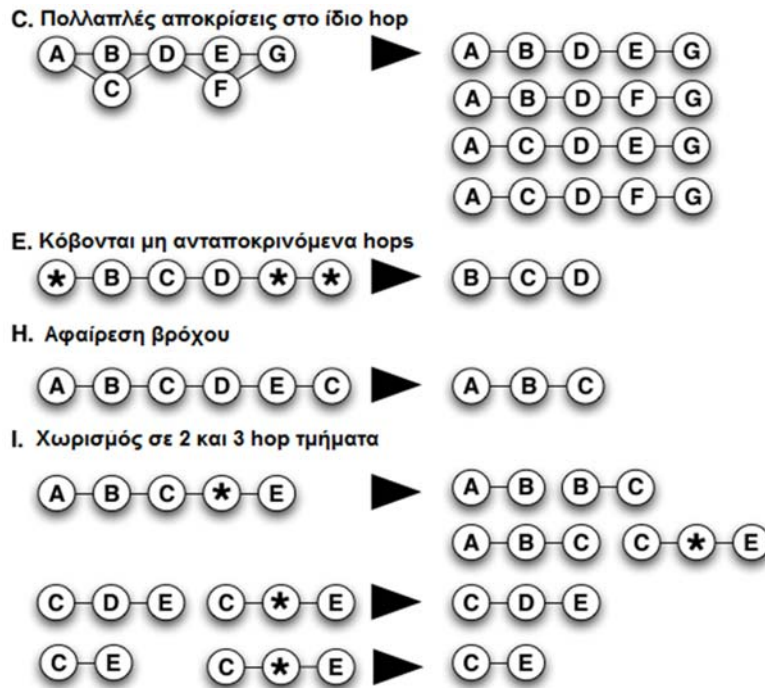
Το iPlane εφαρμόζει την ακόλουθη δύο φάσεων προσέγγιση για την επίλυση ψευδωνύμου. Η πρώτη φάση χρησιμοποιεί δύο μεθόδους για να παράγει μια λίστα από υποψήφια ζεύγη ψευδωνύμων ενώ η δεύτερη φάση δύο μεθόδους για να δοκιμάσει τη λίστα των υποψήφιων ζευγών που παράχθηκαν από τη πρώτη φάση. Η πρώτη τεχνική για τη δημιουργία μίας λίστας από υποψήφια ζεύγη ψευδωνύμων είναι η ίδια με τη παρόμοια του Mercator προσέγγιση που χρησιμοποιεί το DIMES δηλαδή στέλνοντας μια δοκιμή UDP σε κάθε διεπαφή R_j που βλέπει στο traceroute διερεύνησης. Εάν η διεύθυνση πηγής R_i της απόκρισης είναι διαφορετική από της R_j , το ζεύγος R_i, R_j προστίθεται στη λίστα των υποψήφιων ζευγών. Η δεύτερη τεχνική είναι παρόμοια με την APAR μέθοδο, εξετάζει κάθε ζεύγος από διαδοχικές διεπαφές R_k, R_l που παρατηρείται στα traceroutes, εισάγοντας τις άλλες χρήσιμες IP διευθύνσεις R_p στο ίδιο /30 υποδίκτυο όπως τις R_l (και ως εκ τούτου πιθανόν σε μία σύνδεση από σημείο σε σημείο με τις R_l), και προσθέτοντας (R_k, R_p) στη λίστα με τα υποψήφια ζεύγη. Δεδομένης της λίστας με τα υποψήφια ζεύγη και από τις δύο τεχνικές, το σύστημα στη συνέχεια δοκιμάζει κάθε υποψήφιο ζεύγος πάλι χρησιμοποιώντας δύο μεθόδους. Στη πρώτη μέθοδο ένας εξυπηρετητής που τρέχει το iPlane στέλνει ζεύγη από ICMP δοκιμές και στις δύο διευθύνσεις στο υποψήφιο ζεύγος και αποφασίζει αν οι αποκρίσεις έχουν κοντινές IP-ID τιμές και επιστρέφει το ίδιο TTL. Η δεύτερη μέθοδος έχει σαν σκοπό την αποστολή εξατομικευμένων πακέτων προς τις υποψήφιες διεπαφές με την επιλογή του να συλλέγει χρονικές σφραγίδες (timestamps). Εάν η σειρά και οι τιμές των χρονικών σφραγίδων ταιριάζουν, το iPlane θεωρεί ότι αυτές οι διεπαφές είναι στον ίδιο δρομολογητή. Τα προκύπτοντα iPlane επίλυσης ψευδωνύμων δεδομένα δείχνουν ποιές διεπαφές έχει συναχθεί ότι είναι στον ίδιο δρομολογητή αλλά σύνδεσμοι ανάμεσα στους δρομολογητές δεν περιλαμβάνονται. Για τη δημιουργία των συνδέσμων του δρομολογητή για το iPlane δρομολογητή γράφημα, η διαδικασία ξεκινά με το iPlane IP γράφημα και χρησιμοποιούνται τα ψευδώνυμα δρομολογητή του iPlane για την συγχώνευση των IP κόμβων με ψευδώνυμο και των αντίστοιχων συνδέσμων σε δρομολογητή κόμβους και συνδέσμους.

Για τη πτώση IP διευθύνσεων στα δεδομένα του Ark IPv4 All Prefix /24 σε δρομολογητές, χρησιμοποιούνται τα εργαλεία επίλυσης ψευδωνύμων του CAIDA, iffinder, karar [34] και MIDAR [33]. Οι επιπέδου δρομολογητή τοπολογίες που παράγονται από τα Ark IPv4 All Prefix /24 traceroutes χρησιμοποιώντας συνδυασμούς από τα τρία εργαλεία είναι ο πυρήνας των Internet Topology Data Kit (ITDK) συνόλων δεδομένων που δημοσιεύονται τακτικά από

τον CAIDA. Η διαδικασία κατασκευής αυτών των ITDK τοπολογιών περιλαμβάνει τα ακόλουθα βήματα. Πρώτα το *karar* σπάει τα παρατηρούμενα IP μονοπάτια σε IP συνδέσμους. Στη συνέχεια αυτοί οι συναγόμενοι IP επιπέδου σύνδεσμοι γίνονται η είσοδος για επιπλέον μετρήσεις επίλυσης ψευδώνυμων και ανάλυσης από το MIDAR και το *iffinder*. Επιπλέον το *karar* μπορεί επίσης ευρετικά να συμπεράνει το σύνολο των IP διευθύνσεων που ανήκουν στον ίδιο δρομολογητή και το σύνολο των δύο ή και περισσότερων δρομολογητών στον ίδιο "IP σύνδεσμο" (είτε ένα από σημείο σε σημείο σύνδεσμο ή LAN ή κοινόχρηστο μέσο με πολλαπλές συνημμένες IP διευθύνσεις). Ως τμήμα του ITDK υπάρχουν δύο επιπέδου δρομολογητή τοπολογίες, μια συντηρητικά συναγόμενη MIDAR-*iffinder* τοπολογία (Ark ITDK R_{mi}) και μια πιο επιθετικά συναγόμενη MIDAR-*iffinder-karar* τοπολογία (Ark ITDK R_{mik}).

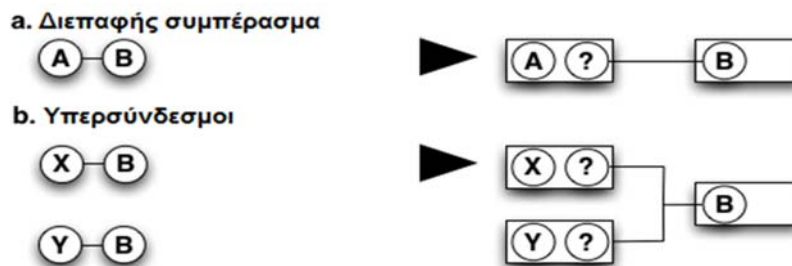
Καrar μετατροπή των IP μονοπατιών σε IP συνδέσμους: Με βάση εμπειρίες με την επεξεργασία και την διερμηνεία *traceroute* δεδομένων, έχει προκύψει η βασική προσέγγιση της εξαγωγής IP συνδέσμων από μονοπάτια που περιγράφονται στα IP γραφήματα με την ακόλουθη διαδικασία δύο φάσεων που εφαρμόζεται στο *karar* για τη κατασκευή επιπέδου δρομολογητή γραφημάτων. Η πρώτη φάση περιλαμβάνει το καθορισμό και διαχωρισμό των IP μονοπατιών σε τμήματα:

- A. Όπως στη βασική επεξεργασία IP ίχνους που περιγράφεται από πάνω, αγνοούνται οι αποκρίσεις από το προορισμό στόχο ενός ίχνους.
- B. Εάν υπάρχουν πολλαπλές αποκρίσεις για ένα μόνο δοκιμαστικό σε ένα συγκεκριμένο hop, κόβεται το ίχνος πριν από αυτό το hop.
- C. Εάν ένα hop έχει πολλαπλές προκύπτουσες αποκρίσεις από πολλαπλές δοκιμές, θεωρούνται όλοι οι συνδυασμοί ότι είναι έγκυρα μονοπάτια (σχήμα 27C). Ωστόσο εάν ο αριθμός των πιθανών συνδυαστικών μονοπατιών είναι >10 , το ίχνος απορρίπτεται.
- D. Εάν η ίδια διεύθυνση προκύπτει σε δύο γειτονικά hops, η πρώτη περίπτωση αντιμετωπίζεται σαν ένα μη ανταποκρινόμενο hop.
- E. Κόβονται τα εμπρός και πίσω μη ανταποκρινόμενα hops (σχήμα 27E).
- F. Οι ψεύτικες και ιδιωτικές διευθύνσεις αντιμετωπίζονται ως μη ανταποκρινόμενες.
- G. Εάν μόνο 2 hops απομένουν μετά από τη παραπάνω επεξεργασία, το ίχνος απορρίπτεται.
- H. Εάν εντοπιστεί ένας βρόχος, διατηρείται μόνο το τμήμα πριν το βρόχο (σχήμα 27H). Σημειώνεται ότι η βασική επεξεργασία που εφαρμόζεται για τη κατασκευή των IP επιπέδου γραφημάτων θα κράταγε το μονοπάτι A-B-C-D-E, αλλά η προσέγγιση του *karar* είναι πιο συντηρητική και αφαιρεί πλήρως τη συνδεσιμότητα που σχηματίζει ένα βρόχο.
- I. Τα μονοπάτια χωρίζονται σε αλληλεπικαλυπτόμενα 2 και 3-hop τμήματα. Εάν ένα 2-hop ή 3-hop τμήμα ταιριάζει με τις ακριανές IP διευθύνσεις ενός 3-hop τμήματος που περιέχει ένα μη ανταποκρινόμενο hop στο μέσο, απορρίπτεται το 3-hop τμήμα που περιέχει το μη ανταποκρινόμενο hop (σχήμα 27I).



Σχήμα 27: Επιλεγμένα παραδείγματα της αρχικής επεξεργασίας του karar των IP μονοπατιών σε τμήματα

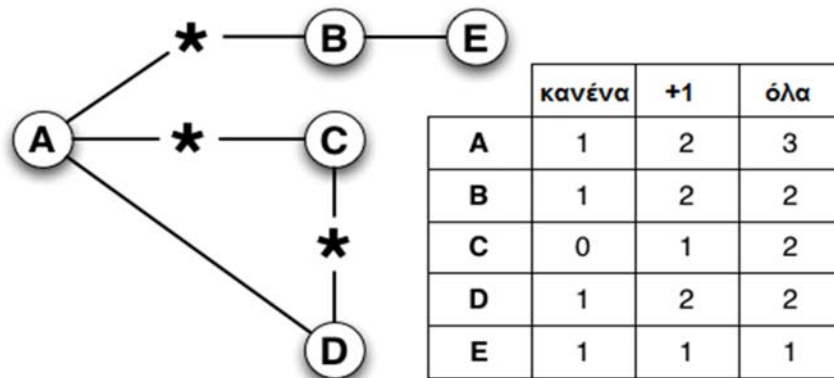
Στη δεύτερη φάση, το karar συνάγει IP συνδέσμους από τα τμήματα ως εξής. Για κάθε τμήμα μονοπατιού (A, B) αυτό αξιώνει ένα σύνδεσμο ανάμεσα στο δρομολογητή (κόμβο) R1 που περιέχει τη διεπαφή A και το δρομολογητή R2 που περιέχει τη διεπαφή B και υποθέτει ότι εάν δεν είναι ο κόμβος R2 ήδη συνδεδεμένος στο κόμβο R1, αυτός ο σύνδεσμος συνδέει τη διεπαφή B στο κόμβο R2 και μια άγνωστη διεπαφή ? που υπονοείται στο κόμβο R1 [A?] ↔ [B] (σχήμα 28a). Υπερσύνδεσμοι (hyperlinks) εμφανίζονται όταν υπάρχουν πολλαπλοί χωρίς ψευδώνυμο προκάτοχοι σε μία διεύθυνση (σχήμα 28b). Σε αντίθεση με τους παραδοσιακούς συνδέσμους, οι υπερσύνδεσμοι αντιπροσωπεύουν συνδεσιμότητα ανάμεσα σε παραπάνω από δύο κόμβους.



Σχήμα 28: Επιλεγμένα βήματα της επεξεργασίας του karar των IP τμημάτων σε συνδέσμους.

Η ενασχόληση με μη ανταποκρινόμενα hops: Εάν δεν υπάρχει μονοπάτι που θα επιλύσει μία τριπλέτα με ένα μη ανταποκρινόμενο hop στη μέση, τότε συμπεριλαμβάνεται η τριπλέτα στο τελικό γράφημα υποθέτοντας ένα προσωρινό κόμβο κράτησης ανάμεσα στους δύο γνωστούς κόμβους. Αυτό το κατασκευάσμα δεδομένων επιτρέπει να διατηρηθούν πληροφορίες σχετικά με τη συνδεσιμότητα χωρίς τη γνώση του ενδιάμεσου hop. Σημειώνεται ότι εάν ένας γνωστός κόμβος έχει περισσότερους από ένα κόμβο κράτησης θέσης ως

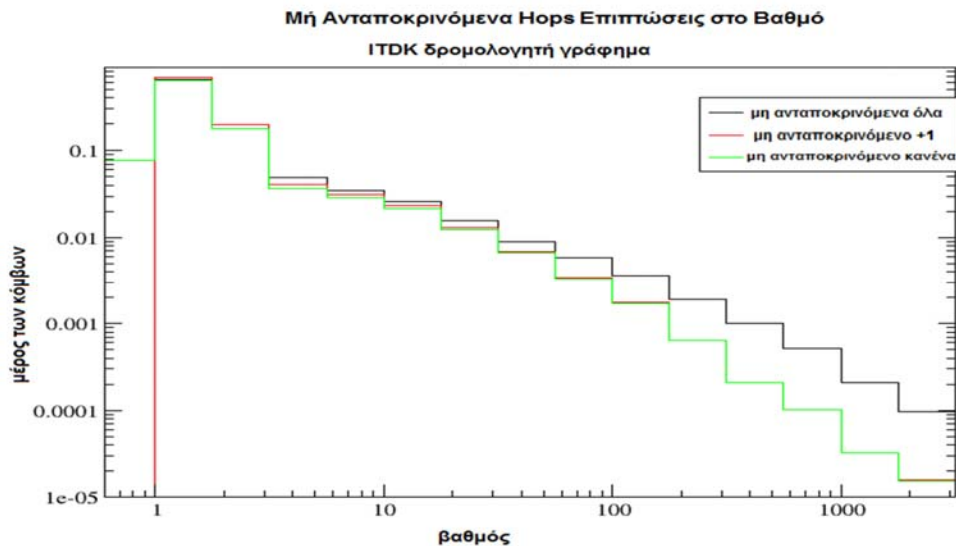
άμεσος γείτονας, τότε δεν μπορεί να διακριθεί αν αυτός είναι στη πραγματικότητα ο μοναδικός μη ανταποκρινόμενος κόμβος ή υπάρχει διαφορετικός μη ανταποκρινόμενος κόμβος για κάθε επόμενο hop που παρατηρείται στα ίχνη. 7.8% των κόμβων (συναγόμενων δρομολογητών) στο ITDK σύνολο δεδομένων έχει μόνο μη ανταποκρινόμενα hops σαν γείτονα(ες). Μερικοί από αυτούς τους συναγόμενους δρομολογητές μπορούν ενδεχομένως να πέσουν περαιτέρω σε υψηλότερου βαθμού δρομολογητές με πρόσθετα δεδομένα. Θεωρείται ότι υπάρχουν τρία σενάρια για την αντιμετώπιση μη ανταποκρινόμενων hops (σχήμα 29) που εναλλάσσει την ακρίβεια και τη πληρότητα του γραφήματος που προκύπτει. Σύμφωνα με το πρώτο και το πιο συντηρητικό σενάριο μη ανταποκρινόμενο κανένα, δημιουργείται ένα καθαρό γράφημα με την απόρριψη όλων των συνδέσμων από τους γνωστούς κόμβους στους μη ανταποκρινόμενους. Ένα τέτοιο γράφημα στερείται κάποιας πραγματικής τοπολογίας. Στο δεύτερο σενάριο μη ανταποκρινόμενο+1, υποθέτεται ότι όλη η συνδεσιμότητα που παρατηρείται μπορεί να αποδοθεί σε ένα μοναδικό άγνωστο κόμβο και έτσι να μετρηθεί μόνο η πρώτη περίπτωση ενός αντικαταστάτη γείτονα προς το βαθμό ενός συγκεκριμένου κόμβου. Στο τρίτο και λιγότερο συντηρητικό σενάριο, μη ανταποκρινόμενα όλα, υποθέτεται ότι κάθε αντικαταστάτης γείτονας είναι πραγματικά ένας διαφορετικός κόμβος και μετρούνται όλοι προς το βαθμό ενός συγκεκριμένου κόμβου. Με άλλες λέξεις το τρίτο σενάριο παραδέχεται το άνω όριο του αριθμού των πιθανών κόμβων που παριστάνεται από τα παρατηρούμενα μη ανταποκρινόμενα hops. Το σχήμα 29 απεικονίζει πώς οι τρεις διαφορετικές τεχνικές επηρεάζουν το συναγόμενο βαθμό του κόμβου A που παρατηρήθηκε στα ακόλουθα τέσσερα μονοπάτια: A*-B, A*-C, A*-D και A-E. Στη περίπτωση μη ανταποκρινόμενο κανένα όταν ρίχνονται όλα τα μη ανταποκρινόμενα hops ο A έχει ένα βαθμό ένα. Στη περίπτωση μη ανταποκρινόμενο+1 όπου μετράτε μόνο η πρώτη περίπτωση ενός μη ανταποκρινόμενου hop ο A έχει ένα βαθμό δύο. Στη περίπτωση μη ανταποκρινόμενα όλα όταν υποθέτεται ότι κάθε αντικαταστάτης αντιπροσωπεύει ένα διαφορετικό κόμβο ο βαθμός του A γίνεται τρία.



Σχήμα 29: Βαθμοί κόμβου που προκύπτουν από τις τρεις μεθόδους για τη μέτρηση των μη ανταποκρινόμενων hops

Για τη διερεύνηση της επίδρασης που αυτές οι τρεις προσεγγίσεις έχουν στο βαθμό κατανομής του γραφήματος που προκύπτει, εξετάζεται το CCDF του βαθμού κατανομής (σχήμα 30). Οι βαθμοί είναι ομαδοποιημένοι αλγοριθμικά με 4 κάδους ανά δεκαετία. Και οι τρεις μέθοδοι έχουν σαν αποτέλεσμα νόμου της δύναμης βαθμού κατανομής με τις

περισσότερες διεπαφές (ανάμεσα σε 63% και 68%) να έχουν βαθμό ένα. Το κενό ανάμεσα στο μη ανταποκρινόμενο+1 (η κόκκινη γραμμή) και στο μη ανταποκρινόμενο κανένα (η πράσινη γραμμή) συρρικνώνεται όσο ο βαθμός αυξάνεται, ενώ το κενό ανάμεσα αυτών των δύο και του μη ανταποκρινόμενα όλα (η μαύρη γραμμή) αυξάνεται σε κάθε λογαριθμικό κάδο. Στην τελευταία περίπτωση όταν μετρούνται όλοι οι αντικαταστάτες κόμβοι προς το βαθμό ενός συγκεκριμένου κόμβου, ο βαθμός κατανομής μετατοπίζεται προς τα δεξιά αφού για κάθε κόμβο ο βαθμός του μπορεί μόνο να αυξηθεί.



Σχήμα 30: Βαθμού κόμβου κατανομές που προκύπτουν από διαφορετικές μεθόδους που ασχολούνται με τα μη ανταποκρινόμενα hops

Λαμβάνοντας υπόψη τους μεγαλύτερους κόμβους στο μη ανταποκρινόμενα όλα γράφημα διαπιστώνεται ότι σχεδόν όλοι τους λαμβάνουν το μεγαλύτερο μέρος της συνδεσιμότητάς τους από μη ανταποκρινόμενα hops. Συνεπώς μετρώντας όλα τα αντικαταστάτες hops σαν διαφορετικές διεπαφές είναι απίθανο να είναι η πιο ακριβής επιλογή. Χρησιμοποιώντας τη μη ανταποκρινόμενο+1 μέθοδο φαίνεται με τη πρώτη ματιά να είναι ένας λογικός συμβιβασμός, επιτρέποντας τη σύλληψη μερικής από της ακατάληπτης συνδεσιμότητας. Ωστόσο από κατασκευής οι αντικαταστάτες κόμβοι έχουν ένα βαθμό 2 και στρεβλώνουν τις κατανομές (πρός κόμβους βαθμού 3) και τους μέσους όρους τυχόν στατιστικών μετρήσεων. Οι κόμβοι αυτοί επίσης καθιστούν δύσκολο να εξακριβωθεί η ομαδοποίηση, δεδομένου ότι η ομαδοποίησή τους είναι εντελώς άγνωστη. Για τους λόγους αυτούς επιλέχθηκε η μεθοδολογία μη ανταποκρινόμενο κανένα για τη κατασκευή γραφημάτων επιπέδου δρομολογητή από τα δεδομένα του Ark και να απορριφθεί το 7.8% των κόμβων που είχαν μόνο μη ανταποκρινόμενους γείτονες. Η συντηρητική αυτή μέθοδο έχει σαν αποτέλεσμα την απώλεια κάποιας συνδεσιμότητας.

3.5.3.3 Σύγκριση των IP και Δρομολογητή Επιπέδου Γραφημάτων Βασισμένη σε Πληροφορίες Αμεσης Παρατήρησης

Καθώς οι IP διευθύνσεις σε ένα IP επιπέδου γράφημα αντιπροσωπεύουν διεπαφές στους πραγματικούς δρομολογητές, τα IP επιπέδου γραφήματα είναι μια προσέγγιση του τι ιδανικά θα ήθελε κανείς, το οποίο είναι ένας χάρτης του πώς κάθε δρομολογητής είναι συνδεδεμένος,

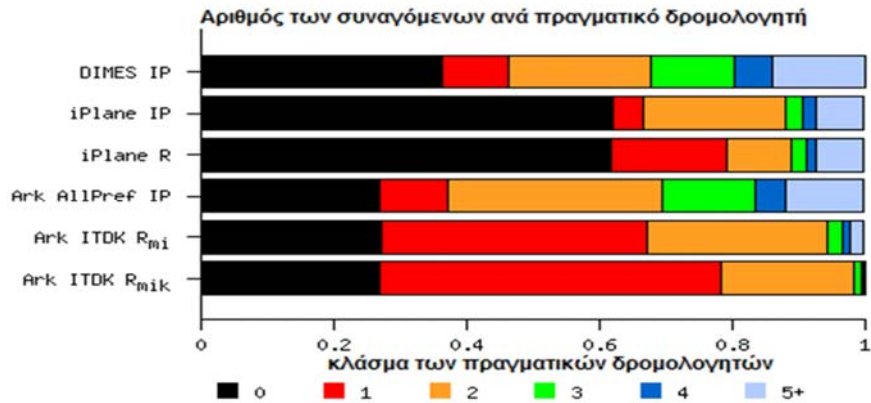
ταυτοποιώντας τις IP διευθύνσεις από όσες περισσότερες IP διεπαφές σε κάθε δρομολογητή είναι εφικτό. Με σκοπό την αξιολόγηση της ακρίβειας διάφορων συνόλων δεδομένων και των μεθόδων που χρησιμοποιούνται για τη παραγωγή αυτών των προσεγγίσεων, συγκρίθηκαν όλα τα γραφήματα επιπέδου IP και δρομολογητή που είναι διαθέσιμα για αυτή τη μελέτη σε ένα σύνολο δεδομένων με πληροφορίες άμεσης παρατήρησης δοσμένο από ένα Tier 1 ISP για το δίκτυο κορμού AS τους (2420 δρομολογητές). Ο ISP αυτός έδωσε μια ολοκληρωμένη λίστα των ονομάτων τομέα των δρομολογητών πυρήνα τους και την ευρετική που χρησιμοποιούν για να χαρτογραφούν τις διεπαφές δρομολογητή στα ονόματα τομέα. Χρησιμοποιώντας αυτό το σύνολο δεδομένων με πληροφορίες άμεσης παρατήρησης, μπορεί να αποτιμηθεί η πληρότητα των διάφορων μετρήσεων για αυτό το ISP, η ακρίβεια της επίλυσης ψευδωνύμου του δρομολογητή και πόσο καλά κάθε πειραματικό σύνολο δεδομένων συλλαμβάνει τη κατανομή βαθμού. Δυστυχώς αυτό το σύνολο δεδομένων με πληροφορίες άμεσης παρατήρησης δεν δείχνει τους πραγματικούς συνδέσμους ανάμεσα στους δρομολογητές, καθιστώντας αδύνατη την αξιολόγηση της ακρίβειας της ομαδοποίησης ή το μέσο όρο του βαθμού γεινίασης των συναγόμενων τοπολογιών. Το σχήμα 31 απεικονίζει τη κάλυψη της κάθε μεθοδολογίας δείχνοντας το κλάσμα των αληθινών δρομολογητών που:

- a. Δεν θα μπορούσε να αντιστοιχιστεί σε κάποιο δρομολογητή στη συναγόμενη τοπολογία (τα μαύρα τμήματα)
- b. Αντιστοιχίζεται σε ένα μόνο δρομολογητή (τα κόκκινα τμήματα) και επιδιώκεται η μεγιστοποίηση του
- c. Αντιστοιχίζεται σε 2 ή περισσότερους δρομολογητές (όλα τα άλλου χρώματος τμήματα), δρομολογητές που είναι υπο κατάρρευση στις συναγόμενες τοπολογίες.

Τα μαύρα τμήματα είναι τα μικρότερα για τις τοπολογίες που προέρχονται από το Ark IPv4 All Prefix /24 σύνολο δεδομένων, αδυνατώντας να συλλάβει 27% των δρομολογητών αυτού του ISP. Το Dimes σύνολο δεδομένων χάνει 37% των πραγματικών δρομολογητών για αυτό το ISP και το iPlane είναι το λιγότερο πλήρες στο 62%. Το ότι το Ark ανιχνεύει το μεγαλύτερο κλάσμα των πραγματικών δρομολογητών της τοπολογίας είναι κάπως περίεργο δεδομένου ότι το Ark έχει τα λιγότερα πλεονεκτήματα, αλλά κάθε συσκευή παρακολούθησης του Ark στέλνει ένα σημαντικά μεγαλύτερο αριθμό ανιχνευτών από τις άλλες πλατφόρμες, συλλαμβάνοντας ένα μεγαλύτερο αριθμό από IP διευθύνσεις. Αυτή η μεγαλύτερη θέα της συνολικής τοπολογίας επιτρέπει την ανίχνευση ενός μεγαλύτερου κλάσματος των άμεσα παρατηρούμενων δρομολογητών.

Το κόκκινο τμήμα από κάθε μπάρα δείχνει το κλάσμα των πραγματικών δρομολογητών που ορθά είχαν τις διεπαφές τους χαρτογραφημένες σε ένα μόνο δρομολογητή. Αυτό δεν σημαίνει βέβαια ότι το σύνολο δεδομένων περιλαμβάνει όλες τις διεπαφές ενός δεδομένου δρομολογητή, αλλά μόνο ότι όλες οι διεπαφές που πιάστηκαν χαρτογραφήθηκαν στο ίδιο δρομολογητή. Τα IP επιπέδου γραφήματα αντιμετωπίζουν κάθε παρατηρούμενη IP διεύθυνση ως ένα ξεχωριστό δρομολογητή, το οποίο σημαίνει ότι ένας πραγματικός δρομολογητής θα αντιστοιχίζεται σε τόσους δρομολογητές όσες οι IP διεπαφές που έχει. Αυτό το συμπέρασμα είναι καθαρά εσφαλμένο όπως αντικατοπτρίζεται από τα μικρά κόκκινα τμήματα στις μπάρες για όλα τα IP επιπέδου γραφήματα στο σχήμα 31: DIMES IP, iPlane IP και Ark IPv4Pref IP. Η διαδικασία επίλυσης IP ψευδωνύμων δηλαδή η συγχώνευση διευθύνσεων διεπαφών σε κοινούς δρομολογητές αυξάνει το κλάσμα των σωστών ένα προς ένα αντιστοιχίσεων. Για τα δεδομένα του iPlane το κλάσμα των πραγματικών δρομολογητών που αντιστοιχούν σε ένα μόνο συναγόμενο δρομολογητή αυξάνεται από 4.6% στα IP επιπέδου γραφήματα σε 17.4% στα επιπέδου δρομολογητή γραφήματα. Για τις επιπέδου

δρομολογητή τοπολογίες στο προερχόμενο από το Ark ITDK, αυτό το κλάσμα αυξάνεται από 10% σε 40% για την MIDAR-iffinder τοπολογία και σε 51% στη MIDAR-iffinder-kapar τοπολογία.



Σχήμα 31: Αντιστοίχιση των πραγματικών δρομολογητών σε ένα δεδομένο αριθμό συναγόμενων δρομολογητών

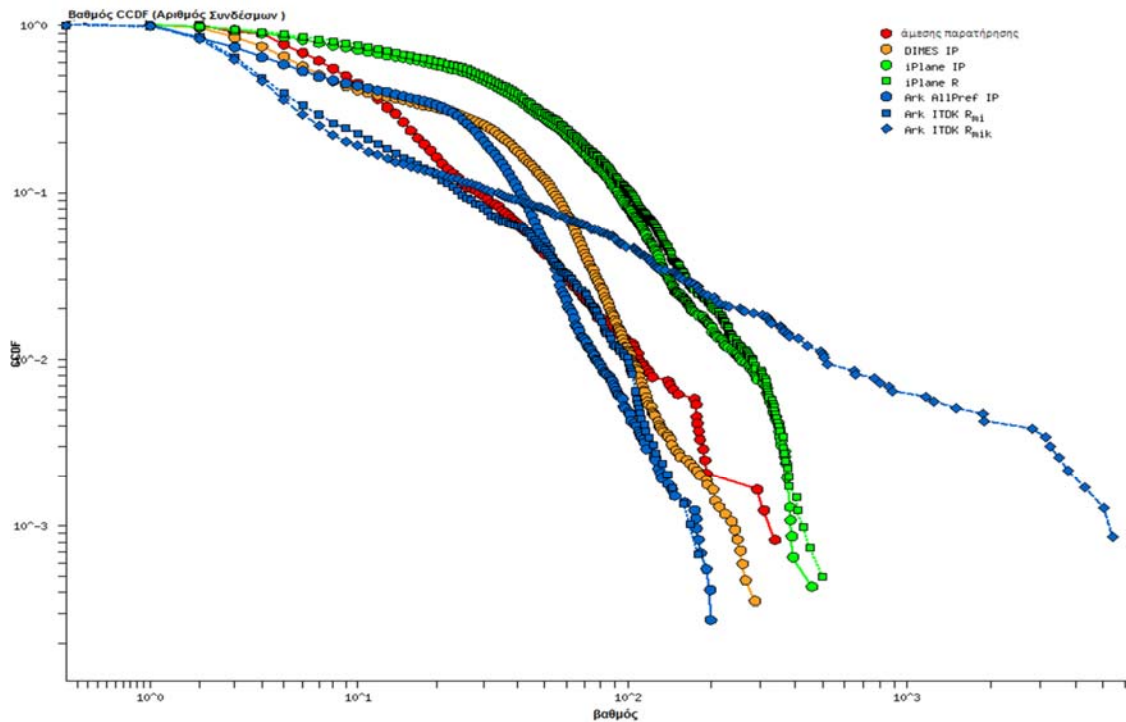
Ταυτόχρονα η επίλυση ψευδωνύμων μπορεί να υπερκαταρρεύσει δρομολογητές κατανέμοντας διεπαφές από πολλαπλούς διαφορετικούς πραγματικούς δρομολογητές στο ίδιο συναγόμενο δρομολογητή (δηλαδή ένα ψευδώς θετικό). Το σχήμα 32 απεικονίζει τη πληθώρα τέτοιων λάθος συμπερασμάτων για ένα μόνο δίκτυο κορμού ISP (με 2420 δρομολογητές). Εδώ το κόκκινο τμήμα από κάθε μπάρα δείχνει το κλάσμα των συναγόμενων δρομολογητών που ορθώς περιλαμβάνουν μόνο IP διευθύνσεις από ένα μόνο πραγματικό δρομολογητή. Καθώς τα γραφήματα IP επιπέδου πάντα ερμηνεύουν μία μόνο IP διεύθυνση ως ένα ξεχωριστό συναγόμενο δρομολογητή για αυτά τα γραφήματα τα κόκκινα κομμάτια είναι τετριμμένα 100% εξ ορισμού. Η διαδικασία επίλυσης ψευδωνύμων του iPlane δημιουργεί αναληθώς συναγόμενους δρομολογητές για 3% των πραγματικών δρομολογητών στα δεδομένα άμεσης παρατήρησης του ISP. Η επίλυση ψευδωνύμων χρησιμοποιώντας το MIDAR-iffinder οδηγεί σε ένα πολύ μικρό κλάσμα από λάθος συμπεράσματα (0.2% των πραγματικών δρομολογητών για αυτό τον ISP) ενώ η MIDAR-iffinder-kapar επεξεργασία υπερκαταρρέει 2.6% των πραγματικών δρομολογητών του ISP.



Σχήμα 32: Αντιστοίχιση των συναγόμενων δρομολογητών σε ένα δεδομένο αριθμό πραγματικών δρομολογητών

Το κλάσμα των λάθος συμπερασμάτων σε όλες τις επιπέδου δρομολογητή τοπολογίες δείχνει μικρό αλλά το σχήμα 33 δείχνει ότι αυτά μπορεί να έχουν δραματικές επιπτώσεις για τις προκύπτουσες βαθμού κόμβου κατανομές. Κανονικά μία βαθμού κόμβου κατανομή είναι ένα ιστόγραμμα του αριθμού των γειτόνων κάθε κόμβου αλλά τα δεδομένα από άμεση παρατήρηση παρέχουν μόνο τον αριθμό των ενεργών διεπαφών σε κάθε δρομολογητή, το οποίο ίσως όχι πάντα να αντιστοιχεί ένα προς ένα στον αριθμό των γειτόνων εξαιτίας των υπερσυνδέσμων. Στη περίπτωση αυτού του ISP οι περισσότερες αλλά όχι όλες οι διεπαφές είναι σε από σημείο σε σημείο συνδέσμους. Έτσι ενώ ο πραγματικός αριθμός των γειτόνων ανα κόμβο είναι άγνωστος, ο αριθμός των συνδέσμων ανα κόμβο που συνάγεται στην επιπέδου δρομολογητή τοπολογία για αυτό τον ISP είναι μια λογική προσέγγιση του αριθμού των διεπαφών, ο οποίος συγκρίνεται με τον αριθμό των διεπαφών που αναφέρονται στα δεδομένα άμεσης παρατήρησης.

Πρώτα εξάγεται το σύνολο των δρομολογητών από τη συναγόμενη τοπολογία με τουλάχιστον μία διεπαφή αντίστοιχη μίας διεπαφής στα δεδομένα άμεσης παρατήρησης και συγκρίνονται (σχήμα 33) οι CCDF του αριθμού των συνδέσμων που συνδέουν σε κάθε τέτοιο εξαγόμενο δρομολογητή έναντι του αριθμού των διεπαφών σε ένα δρομολογητή στα δεδομένα άμεσης παρατήρησης (τα κόκκινα σύμβολα) ως διακομιστές μεσολάβησης για τις CCDFs των κατανομών του βαθμού κόμβου. Και τα δύο προερχόμενα από το iPlane γραφήματα (τα πράσινα σύμβολα) υπερεκτιμούν σημαντικά τον αριθμό των δρομολογητών σε αυτό τον ISP με βαθμούς >10: 40% στο σύνολο δεδομένων άμεσης παρατήρησης έναντι 70% και 74% στις iPlane τοπολογίες. Οι DIMES IP (οι κίτρινοι κύκλοι) και οι Ark IPv4Pref IP (οι μπλέ κύκλοι) τοπολογίες αποδίδουν λογικές προσεγγίσεις του βαθμού κατανομής για το 60% των δρομολογητών άμεσης παρατήρησης που έχουν βαθμούς <10, αλλά ξεκινούν να αποκλίνουν για βαθμούς ανάμεσα σε 10 και 60, που αντιπροσωπεύουν περίπου το 37% των δρομολογητών στα δεδομένα άμεσης παρατήρησης. Το DIMES IP γράφημα είναι το κοντινότερο στις πληροφορίες άμεσης παρατήρησης στη κλίμακα των μεγάλων βαθμών (>100) αλλά αυτή η κλίμακα αντιπροσωπεύει μόνο το 1% των δρομολογητών άμεσης παρατήρησης. Ο πολύ μεγαλύτερος αριθμός των ακμών πλεονεκτικής θέσης του DIMES φυσικά θα συλλαμβάνει ένα μεγαλύτερο αριθμό των διεπαφών που εισέρχονται δρομολογητές πυρήνα από τη περιφέρεια. Και οι δύο προερχόμενες από το ITDK επιπέδου δρομολογητή τοπολογίες (οι μπλέ ρόμβοι και τετράγωνα) υποτιμούν τους βαθμούς των μικρού βαθμού κόμβων (<20), που είναι το 84% των άμεσης παρατήρησης κόμβων, αλλά η Ark ITDK Router $_{mi}$ τοπολογία που χρησιμοποιεί μόνο MIDAR-iffinder επεξεργασία (τα μπλέ τετράγωνα) ταιριάζει τέλεια στους άμεσα παρατηρούμενους στη κλίμακα των βαθμών κόμβους μεταξύ 20 και 100 ή το 15% των δρομολογητών άμεσης παρατήρησης. Σε αντίθεση η τοπολογία MIDAR-iffinder-kapar (Ark ITDK ROUTER $_{mik}$, οι μπλέ ρόμβοι) περιλαμβάνει εξωπραγματικά υπερ- υψηλού βαθμού κόμβους που εμφανίζονται όταν δύο ή περισσότεροι δρομολογητές είναι συγχωνευμένοι σε ένα μόνο υπερ-δρομολογητή: 4.6% των Ark ITDK Router $_{mik}$ δρομολογητών έχουν βαθμούς >100 έναντι μόνο 1.2% των αντιστοίχων άμεσης παρατήρησης δρομολογητών. Προσθέτοντας τα συμπεράσματα του kapar στα αποτελέσματα του MIDAR-iffinder αυξάνεται η πληρότητα της επίλυσης ψευδώνυμων (σχήμα 31) αλλά αυτή η πρόσθετη επεξεργασία επίσης υπερκαταρρέει τους δρομολογητές (σχήμα 32) στρεβλώνοντας τη κατανομή του βαθμού κόμβου προς εξωπραγματικά μεγάλους βαθμούς. Για την αποφυγή ψευδώς θετικών και σχετιζόμενων διαστρεβλωμένων (distorted) στατιστικών, χρησιμοποιείται η πιο συντηρητικά συναγόμενη Ark ITDK Router $_{mi}$ τοπολογία για το υπόλοιπο της παρούσας μελέτης.



Σχήμα 33: Το CCDF των βαθμών κόμβου για κάθε μέθοδο επεξεργασίας και πηγή δεδομένων

3.5.3.4 Γραφήματα AS Επιπέδου

Τα AS επιπέδου γραφήματα αντιπροσωπεύουν τη τοπολογία του Διαδικτύου σε επίπεδο αυτόνομων συστημάτων (ASes), τα οποία είναι κατά προσέγγιση δίκτυα κάτω από ένα ενιαίο διαχειριστικό έλεγχο. Τα ASes είναι πολύ σημαντικά καθώς αυτά είναι “η μονάδα της πολιτικής δρομολόγησης” στο σύστημα δρομολόγησης του παγκόσμιου διαδικτύου. Τα ASes ανταλλάσσουν κίνηση μεταξύ τους για να αλλάξουν τη συμφόρηση και αυτές οι σχέσεις ανταλλαγής κίνησης καθορίζουν την υψηλού επιπέδου παγκόσμια τοπολογία Διαδικτύου. Για τους σκοπούς της ανάλυσης αυτές οι σχέσεις ανταλλαγής κίνησης απεικονίζονται με ένα AS γράφημα, όπου οι κόμβοι αντιπροσωπεύουν ASes και οι σύνδεσμοι αντιπροσωπεύουν τις σχέσεις ανταλλαγής κίνησης. Η ενότητα αυτή εστιάζει στη κατασκευή των AS επιπέδου γραφημάτων από τρεις πιθανές πηγές δεδομένων: ακατέργαστα traceroute δεδομένα, BGP (Border Gateway Protocol) inter-AS πινάκων δρομολόγησης χωματερές, και WHOIS μητρώου δρομολόγησης καταχωρήσεις στη βάση δεδομένων από το RIPE (εθελοντικά συντηρημένα από ορισμένους ISPs στο IRR του RIPE).

Βασισμένα στο traceroute AS επιπέδου γραφήματα: Ένα τυπικό σημείο εκκίνησης για τη κατασκευή AS επιπέδου Διαδικτυακών τοπολογιών από δεδομένα του traceroute χρησιμοποιώντας BGP πίνακα χωματερές από Route Views ερευνητικό σχέδιο και του RIPE-NCC RIS για να αντιστοιχίσει δρομολογήσιμα προθέματα που βρέθηκαν στα ίχνη που έχουν συλλεχθεί από τα ASes στο παγκόσμιο σύστημα δρομολόγησης. Ένα μικρό ποσοστό από προθέματα IP αντιστοιχούν σε ένα σύνολο AS δηλαδή ένα σύνολο από ASes καθένα από τα οποία θα μπορούσε να ανακοινώσει το πρόθεμά. Η προέλευση αυτών των IP προθεμάτων αφήνεται ανοιχτή και απορρίπτονται αυτά τα AS σύνολα (η IETF βρίσκεται στη διαδικασία της αποδοκιμασίας των AS συνόλων). Μερικά προθέματα προέρχονται από πολλαπλά ASes, οπότε επιλέγεται το AS πιο συχνά να γίνεται αντιληπτό στους BGP πίνακες

ώς AS προέλευσης. Από 366.294 προθέματα που βρέθηκαν στους RouteViews BGP πίνακες στο πρώτο μισό του Απριλίου του 2011, 2.299 προθέματα (0.6%) προήλθαν από AS σύνολα και 18 προθέματα (0.005%) είχαν πολλαπλής προέλευσης ASes. Όταν υπάρχει αντιστοιχία ανάμεσα στον IP χώρο διευθύνσεων και στον AS χώρο, η απλούστερη μέθοδος κατασκευής ενός γραφήματος AS επιπέδου προϋποθέτει αντιστοιχία κάθε IP διεύθυνσης στα ίχνη του AS προέλευσης και συνάγονται AS συνδέσμοι αντίστοιχοι των παρατηρούμενων IP συνδέσμων. Η τεχνική αυτή χρησιμοποιείται για τη δημιουργία των iPlane AS και Ark IPv4Pref AS αρχείων AS συνδέσμων και για τη σύγκριση των AS επιπέδου τοπολογιών. Σημειώνεται ότι το DIMES παρέχει τα δικά του σύνολα από AS συνδέσμους (DIMES AS), τα οποία χρησιμοποιούνται κατευθείαν σε αντίθεση με τα δεδομένα iPlane που πρέπει πρώτα να παραχθούν.

Για τη τοπολογία Ark ITDK Router mi εξετάστηκαν δύο μέθοδοι για τη δημιουργία AS επιπέδου γραφήματα Διαδικτύου: παρατηρούμενου δρομολογητή και συναγόμενου δρομολογητή. Και στις δύο περιπτώσεις, το πρώτο βήμα είναι να εκχωρηθεί ιδιοκτησία δρομολογητή σε ASes. Για να γίνει η αντιστοιχία κάθε IP διεύθυνσης σε ένα συγκεκριμένο δρομολογητή στο AS προέλευσής του, χρησιμοποιούνται τα BGP δεδομένα. Όποιο AS έχει τις περισσότερες IP διευθύνσεις διεπαφών ανατεθειμένες σε αυτό «κατέχει» το δρομολογητή ή σε περίπτωση ισοβαθμίας ανάμεσα σε δύο ASes, ανατίθεται ο δρομολογητής στο AS με το μικρότερο βαθμό.

Η μέθοδος **AS σύνδεσμοι παρατηρούμενου δρομολογητή** ξεκινά με τις παρατηρούμενες IP διεπαφές στη διαδρομή, χρησιμοποιεί τα δεδομένα της επίλυσης ψευδωνύμων για να αντιστοιχίσει αυτές τις διεπαφές σε δρομολογητές και στη συνέχεια χρησιμοποιεί τα δεδομένα ανάθεσης δρομολογητή-AS για να αντιστοιχίσει αυτούς τους δρομολογητές σε ASes [35]. Αυτή η αντιστοιχία οδηγεί σε ένα AS μονοπάτι το οποίο στη συνέχεια χωρίζεται σε AS συνδέσμους. Το γράφημα AS που προέρχεται από αυτή τη μέθοδο καλείται Ark ITDK AS ro .

Η μέθοδος **AS σύνδεσμοι συναγόμενου δρομολογητή** ξεκινά με το ITDK γράφημα, χρησιμοποιεί τα ίδια δεδομένα εκχώρησης δρομολογητή-AS όπως παραπάνω για να αντιστοιχίσει τους δρομολογητές αυτούς σε ASes, καταλήγοντας σε ένα AS γράφημα, το οποίο στη συνέχεια χωρίζεται σε AS συνδέσμους. Η εννοιολογική διάκριση ανάμεσα στις δύο μεθόδους έγκειται στο ότι ένα AS γράφημα που κατασκευάστηκε χρησιμοποιώντας τη παρατηρούμενου δρομολογητή μέθοδο περιέχει μόνο AS συνδέσμους που αντιστοιχούν σε IP συνδέσμους που ήταν άμεσα παρατηρούμενοι μέσω συστήματος μέτρησης. Ενώ ένα γράφημα που κατασκευάστηκε από τη συναγόμενου δρομολογητή μέθοδο επίσης περιλαμβάνει συνδέσμους που δεν ήταν στη πραγματικότητα αποτελέσματα της διαδικασίας μέτρησης, αλλά μπορούν να συναχθούν από το γράφημα επιπέδου δρομολογητή. Αυτό το γράφημα ονομάζεται Ark ITDK AS ri .

Αν και αποκλείονται οι διευθύνσεις προορισμού κατά τη κατασκευή γραφημάτων επιπέδου IP και δρομολογητή (καθώς αυτά τα γραφήματα επικεντρώνονται σε δρομολογητές, όχι άκρες υποδοχής), διατηρούνται οι διευθύνσεις αυτές όταν κατασκευάζονται γραφήματα AS επιπέδου με την ακόλουθη αιτιολογία. Παρά το γεγονός ότι ο δρομολογητής ακριβώς πριν το προορισμό ίσως να διαχειρίζεται από το ίδιο AS με του προορισμού, συχνά μπορεί κάποιος να δει μόνο τη διεύθυνση παρόχου του στο traceroute αποτέλεσμα που έχει συλλεχθεί. Στη περίπτωση αυτή διατηρώντας τη διεύθυνση προορισμού παρέχεται ένας τρόπος να συλλεχθεί επιπλέον AS συνδεσιμότητα. Απορρίπτοντας τις διευθύνσεις προορισμού θα μειωθεί το μέγεθος του προκύπτοντος AS επιπέδου γραφήματος κατά 29%.

Βασισμένα στο BGP AS επιπέδου γραφήματα: Για τη δημιουργία ενός AS επιπέδου γραφήματος από τα BGP δεδομένα, γίνεται αρχή από τα AS μονοπάτια που βρέθηκαν για κάθε πρόθεμα και σπάσιμο αυτών των AS μονοπατιών σε ξεχωριστούς AS συνδέσμους. Απορρίπτονται σύνδεσμοι που περιέχουν ιδιωτικά ASes. Για τα σύνολα δεδομένων RouteViews2 και BGP Full γίνεται συλλογή ενός πίνακα δρομολόγησης πέντε συνεχόμενων ημερών και εξαγωγή AS συνδέσμων μόνο από τα συνεχή μονοπάτια κατά τη διάρκεια αυτή (μονοπάτια που μπορεί κάποιος να δει στη πλειοψηφία των πινάκων δρομολόγησης κατά τη διάρκεια του μεσοδιαστήματος (interval)). Τα σύνολα δεδομένων του IRL χρησιμοποιούν BGP δεδομένα από ενεργά Route Views, Internet2, RIPE RIS εξυπηρετητές και υπολογιστές στο Διαδίκτυο που τρέχουν μία από μία ποικιλία διαθέσιμων στο κοινό Looking Glass εφαρμογών λογισμικού, αν και η τεκμηρίωση του IRL δεν ήταν αρκετή για να εξηγήσει ακριβώς ποιιά τμήματα ποιών πηγών δεδομένων χρησιμοποιούσαν.

WHOIS AS επιπέδου γραφήματα: Για να αντληθούν AS επιπέδου γραφήματα από τα δεδομένα RIPE WHOIS IRR, χρησιμοποιούνται τα πεδία εισαγωγής και εξαγωγής που απαριθμούν τα ASes που είναι καταγεγραμμένα ως BGP γείτονες ενός συγκεκριμένου AS (αντιπροσωπεύεται από τον αριθμό αυτόνομου συστήματος του στο IRR μητρώο). Δημιουργούνται σύνδεσμοι μεταξύ του αριθμού αυτόνομου συστήματος AS και των ASes που αναφέρονται στα πεδία εισαγωγής και εξαγωγής, εξαιρουμένων των ASes που εμφανίζονται μόνο ως γείτονες αλλά δεν έχουν τις δικές τους σειρές αριθμών αυτόνομου συστήματος. Τέτοια Ases είναι εξωτερικά στη βάση δεδομένων και δεν μπορούν να εκτιμηθούν σωστά οι τοπολογικές ιδιότητές τους (π.χ ο βαθμός κόμβου). Τέλος αγνοούνται τα ιδιωτικά ASes.

3.5.4 Η Ανάλυση των Γραφημάτων Τοπολογίας του Διαδικτύου

3.5.4.1 Χαρακτηριστικά των IP και Δρομολογητή Επιπέδου Γραφημάτων

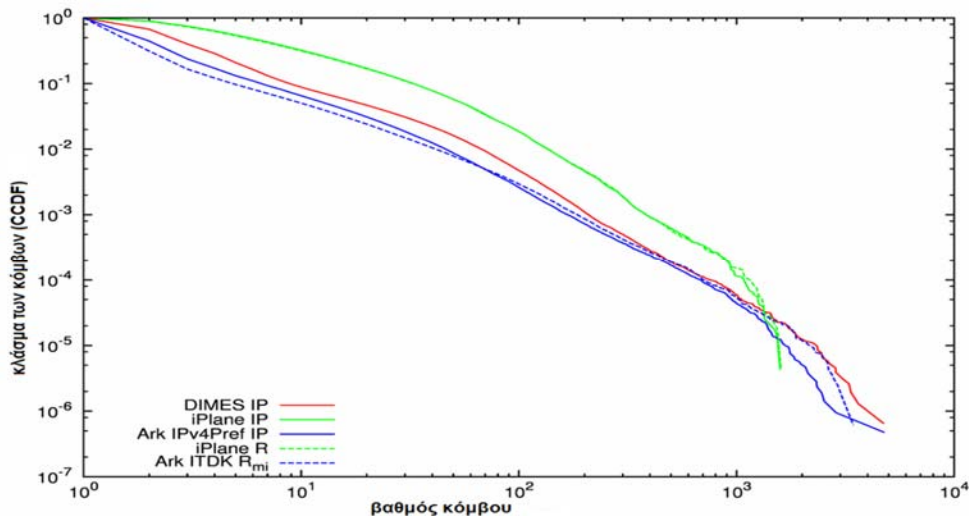
Ο πίνακας 7 συγκρίνει τα βασικά στατιστικά στοιχεία από τρία IP επιπέδου γραφήματα και δύο επιπέδου δρομολογητή γραφήματα. Ο αριθμός των συνδέσμων που παρατηρείται στα IP επιπέδου γραφήματα DIMES IP και Ark IPv4Pref IP δεδομένα είναι παρόμοιος, με μόνο 4% περισσότερους συνδέσμους στο DIMES IP, παρά το γεγονός ότι έχει 27% λιγότερους κόμβους. Το iPlane IP γράφημα έχει μόνο ένα κλάσμα (~11-15%) του αριθμού των κόμβων και 40% του αριθμού των συνδέσμων που έχουν τα άλλα δύο γραφήματα. Το μικρότερο μέγεθος του γραφήματος iPlane IP είναι σύμφωνο με την εστίασή του στη σκιαγράφηση της τοπολογίας του πυρήνα του Διαδικτύου, πράγμα το οποίο επίσης εξηγεί το μεγαλύτερο μέσο όρο του βαθμού. Το iPlane IP γράφημα έχει μικρότερο μέγιστο βαθμό κόμβου, ίσως επειδή αυτό έχει αρκετά λιγότερους κόμβους. Το iPlane IP γράφημα έχει μία τάξη μεγέθους υψηλότερου μέσου όρου τοπικής ομαδοποίησης από το Ark IPv4Pref IP γράφημα, αν και αυτή η ανομοιότητα εξαφανίζεται μετά την επίλυση ψευδωνύμων, σημείο στο οποίο το Ark IPv4Pref Router $_{mi}$ γράφημα έχει μόνο 13% υψηλότερο μέσο όρο τοπικής ομαδοποίησης από το iPlane Router γράφημα. Η διαδικασία επίλυσης ψευδωνύμων στο Ark IP επιπέδου γράφημα, Ark IPv4Pref IP, δημιουργεί ένα επιπέδου δρομολογητή γράφημα Ark ITDK Router $_{mi}$, μειώνοντας τον αριθμό των κόμβων κατά 23% και των συνδέσμων κατά 31%. Η iPlane IP επίλυση ψευδωνύμων οδήγησε στο να έχει το iPlane router γράφημα μόνο 6.7% λιγότερους κόμβους και 7.8% λιγότερους συνδέσμους. Φαίνεται ότι οι μέθοδοι επίλυσης ψευδωνύμων που χρησιμοποιούνται από το iPlane είναι λιγότερο επιθετικές και/ή αποτελεσματικές από τις MIDAR, iffinder, kapar του CAIDA.

Πίνακας 7: Βασικά στατιστικά στοιχεία των IP και δρομολογητή τοπολογίας γραφήματων

γραφήματα	αριθμός από		βαθμός		κανονικοποιη- μένος μέσος βαθμός γείτονα	μέση τοπική ομαδοποίηση
	κόμβους	ακμές	μέσος	μέγιστος		
IP επιπέδου γραφήματα						
Ark IPv4Pref IP	2,111,019	4,073,080	3.86	4,772	5.53e-05	0.012
DIMES IP	1,543,320	4,230,578	5.48	4,742	6.31e-05	0.065
iPlane IP	233,996	1,661,041	14.20	1,586	2.16e-04	0.12
επιπέδου δρομολογητή γραφήματα						
Ark ITDK Router _{mi}	1,633,126	2,729,618	3.34	3,439	8.48e-05	0.15
iPlane Router	218,399	1,531,736	14.03	1,600	2.38e-04	0.13

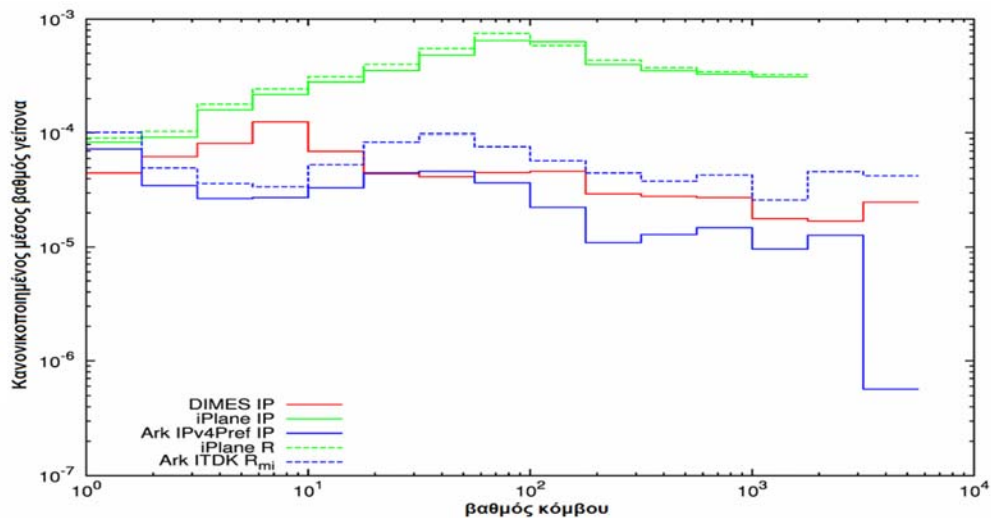
Τα γραφήματα στα σχήματα (34,35,36) είναι: ο CCDF του βαθμού κόμβου, ο μέσος βαθμός γείτονα και η τοπική ομαδοποίηση, ως συναρτήσεις του βαθμού κόμβου, για τα τρία IP επιπέδου (συνεχείς γραμμές) και τα δύο επιπέδου δρομολογητή (διακεκομμένες γραμμές) γραφήματα. Οι x-τιμές (βαθμοί) είναι ομαδοποιημένες χρησιμοποιώντας τέσσερις κάδους ανά δεκαετία. Για όλες τις τρεις στατιστικές, τα στατιστικά χαρακτηριστικά των iPlane και iPlane Router γραφημάτων (η πράσινη γραμμή) είναι παρόμοια. Σύμφωνα με τις συντηρητικές μεθόδους επίλυσης ψευδωνύμου του iPlane, λιγότερο από 6% των κόμβων στο IP επιπέδου γράφημα κατέρρευσε στους δρομολογητές.

Το σχήμα 34 αποκαλύπτει ότι η κατανομή του βαθμού κόμβου και στα δύο iPlane γραφήματα ευνοεί τους υψηλού βαθμού κόμβους: 30% των κόμβων έχει βαθμό μεγαλύτερο του 10, σε σύγκριση με το 10% ή λιγότερο και για τα δύο προερχόμενα από το Ark (οι μπλέ γραμμές) και το DIMES IP (οι κόκκινες γραμμές) γραφήματα. Έχει ήδη σημειωθεί αυτό το αποτέλεσμα στην άμεση παρατήρηση σύγκριση (σχήμα 33).



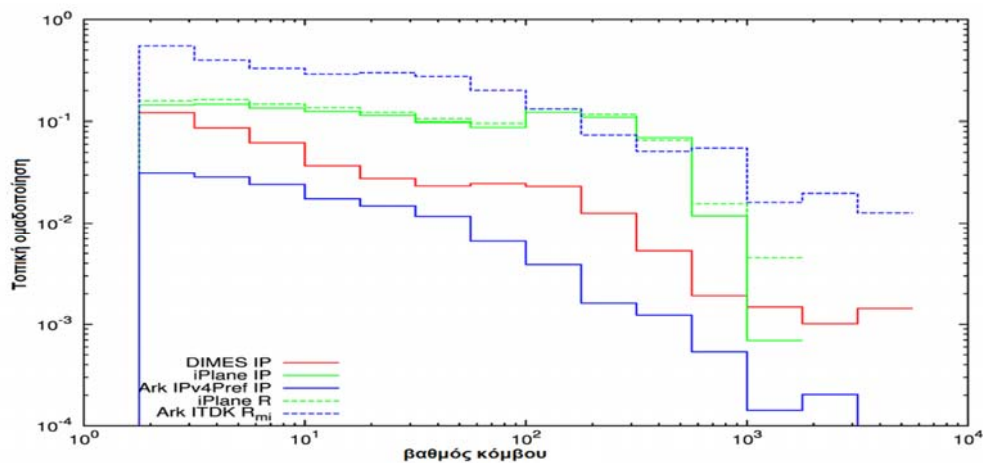
Σχήμα 34: CCDF του βαθμού κόμβου των IP και δρομολογητή επιπέδου

Οι γραφικές παραστάσεις στο σχήμα 35 ομαλοποιούν τους μέσους βαθμούς γειτόνων. Αντίθετα με τις κατανομές βαθμού, οι οποίες περιγράφουν κόμβους μεμονωμένα, ο μέσος βαθμός γείτονα αποτυπώνει πώς κόμβοι διαφορετικού βαθμού διασυνδέονται. Μπορεί κάποιος να δει δύο τύπους συμπεριφοράς. Και για τα δύο τα γραφήματα iPlane, ο μέσος βαθμός γείτονα αρχικά αυξάνεται όσο ο βαθμός κόμβου αυξάνεται, αλλά οι υψηλού βαθμού κόμβοι $k > 100$ τείνουν (tend) να συνδέονται σε μικρότερου βαθμού κόμβους και ο μέσος βαθμός γείτονα μειώνεται. Το DIMES IP γράφημα έχει παρόμοια συμπεριφορά αλλά ο μέσος βαθμός γείτονα αρχίζει να μειώνεται για $k > 10$. Σε αντίθεση, ο μέσος βαθμός γείτονα παραμένει σχεδόν σταθερός και για τα δύο προερχόμενα από το Ark γραφήματα σε όλο το φάσμα των βαθμών των κόμβων.



Σχήμα 35: Μέσος βαθμός γείτονα των IP και δρομολογητή επιπέδου

Λαμβάνοντας υπόψη τη τοπική ομαδοποίηση ως μία συνάρτηση του βαθμού του κόμβου (σχήμα 36), παρατηρείται ότι το γράφημα ITDK Router έχει γενικά τη μεγαλύτερη ομαδοποίηση, που ακολουθείται με τη σειρά του από τα δύο iPlane σύνολα δεδομένων, DIMES και Ark IPv4Pref IP. Η επίλυση ψευδώνυμων δηλαδή η συγκέντρωση IP διευθύνσεων σε ένα επίπεδο δρομολογητή γράφημα, αυξάνει την ομαδοποίηση δεδομένου ότι μειώνει τον αριθμό των κόμβων, αλλά τους καθιστά πυκνά συνδεδεμένους.



Σχήμα 36: Τοπική ομαδοποίηση των IP και δρομολογητή επιπέδου

3.5.4.2 Χαρακτηριστικά των AS Επιπέδου Γραφημάτων

Στο πίνακα 8, όλες οι άλλες πηγές δεδομένων αντί του WHOIS RIPE AS ταιριάζουν με ένα μοντέλο κατανομής AS βαθμού, ως μια συνάρτηση νόμου δύναμης με εκθέτη γ μεταξύ 2.1 και 2.2. Όσο πιο κοντά η τιμή του εκθέτη του νόμου δύναμης είναι σε 2, υπάρχουν σχετικά περισσότεροι κεντρικοί σταθμοί (υψηλού βαθμού κόμβοι) στο δίκτυο. (Σημειώνεται ότι ένα δίκτυο δεν μπορεί να έχει μία βαθμού νόμου δύναμης κατανομή με $\gamma < 2$ σε όλο το φάσμα των βαθμών λόγω διαρθρωτικών περιορισμών, ειδικότερα για τη νόμου δύναμης βαθμού κόμβου κατανομή, ο μέγιστος δυνατός βαθμός σε ένα δείγμα από N κόμβους είναι $N^{1/(\gamma-1)}$. Εάν το γ είναι < 2 , τότε ο μέγιστος βαθμός θα μπορούσε να είναι $> N$, το οποίο είναι αδύνατο. Ως εκ τούτου, μια προσπάθεια για τη δημιουργία γραφημάτων με μία νόμου δύναμης βαθμού κατανομή με εκθέτη < 2 οδηγεί σε μία εκθετική αποκοπή (cutoff) στους μεγάλους βαθμούς. Αυτές οι κατανομές βαθμού δεν έχουν ποτέ παρατηρηθεί για πραγματικά δίκτυα.) Λόγω του μεγάλου αριθμού των πηγών δεδομένων που χρησιμοποιούνται για σύγκριση AS επιπέδου γραφημάτων, πρώτα αναλύονται τα AS γραφήματα σε κάθε υπο-ομάδα: Ark, traceroute, BGP και στη συνέχεια επιλέγεται ένας μόνο αντιπρόσωπος από κάθε υπο-ομάδα για τη συνολική σύγκριση, η οποία επίσης περιλαμβάνει ένα AS γράφημα προερχόμενο από τα WHOIS δεδομένα.

Πίνακας 8: Βασικά στατιστικά των AS γραφημάτων

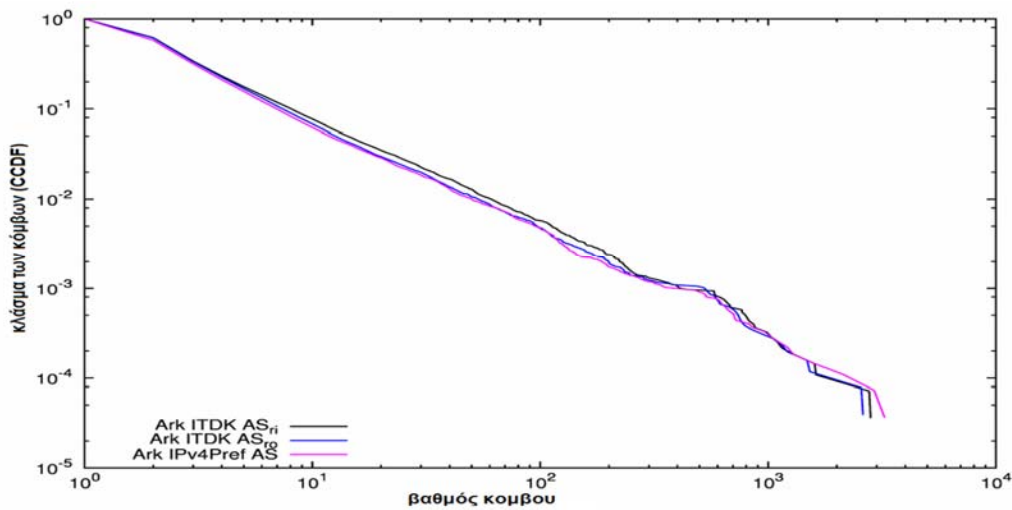
γραφήματα	τύπος δεδομένων	αριθμός από		βαθμός		κανονικοποιημένος μέσος βαθμός γείτονα	μέση τοπική ομαδοποίηση	ελαχίστων τετραγώνων προσαρμογή του CCDF	γ	μέγιστης πιθανότητας προσαρμογή του βαθμού ακολουθίας
		κόμβους	ακμές	μέσος	μέγιστος					
Ark IPv4Pref AS	traceroute	27,399	68,685	5.01	3,245	0.019	0.35			
Ark ITDK AS _{ro}	traceroute	25,578	66,401	5.19	2,607	0.016	0.33	2.19	2.18	
Ark ITDK AS _{ri}	traceroute	27,797	77,965	5.61	2,815	0.018	0.36	2.11	2.20	
DIMES AS	traceroute	25,774	78,373	6.08	4,386	0.029	0.43	2.12	2.18	
iPlane AS	traceroute	17,937	61,218	6.83	3,753	0.042	0.50	2.11	2.22	
RouteViews2 AS	BGP	37,606	80,051	4.26	3,100	0.016	0.21	2.15	2.12	
BGP full AS	BGP	36,876	103,481	5.61	2,972	0.014	0.24	2.12	1.97	
IRL AS	BGP	38,524	125,105	6.49	3,211	0.015	0.30	2.13	1.90	
WHOIS RIPE AS	WHOIS	22,898	134,448	11.74	3,727	0.027	0.37			

Διαφορές μεταξύ των βασισμένων στο Ark AS γραφημάτων: Αρχικά αναλύονται οι διαφορές ανάμεσα στα γραφήματα AS που κατασκευάζονται απευθείας από τα δεδομένα του Ark (Ark IPv4Pref AS) και από τα επιπέδου δρομολογητή γραφήματα στο ITDK χρησιμοποιώντας τις δύο διαφορετικές μεθόδους που περιγράφηκαν παραπάνω (Ark ITDK AS_{ro} και Ark ITDK AS_{ri}).

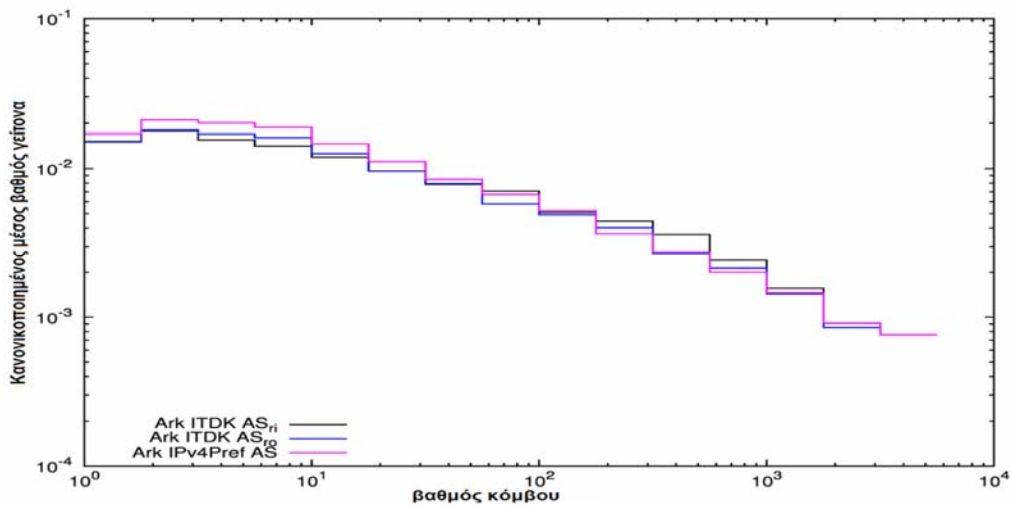
Τα σχήματα (37, 38, 39) απεικονίζουν την ομοιότητα των τριών τοπολογικών μετρήσεων για τα τρία προερχόμενα από το Ark/ITDK AS γραφήματα, αν και το Ark ITDK AS_{ri} γράφημα (οι μαύρες γραμμές) επιδεικνύει υψηλότερους βαθμούς και υψηλότερη τοπική ομαδοποίηση από τα άλλα δυο γραφήματα εξαιτίας της ενσωμάτωσης των πρόσθετων συνδέσμων που ανάγονται στη διαδικασία της αντιστοίχισης IP σε δρομολογητή και δρομολογητή σε AS.

Οι κατανομές βαθμού των Ark IPv4Pref (η μώβ γραμμή) και Ark ITDK AS_{ro} (η μπλέ γραμμή) γραφημάτων είναι αισθητά διαφορετικές για τους μεγαλύτερους κόμβους με $k > 1000$ (σχήμα 37). Το Ark ITDK AS_{ro} γράφημα μπορεί να επιλεγεί ως το αντιπροσωπευτικό του συνόλου των προερχόμενων από το Ark/ITDK AS επιπέδου γραφημάτων Διαδικτύου για σύγκριση με

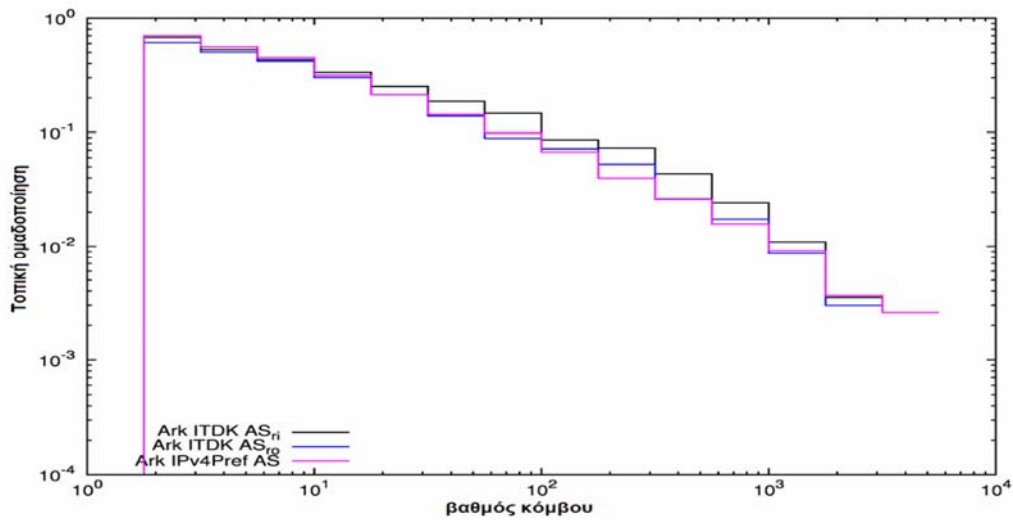
άλλα προερχόμενα από το traceroute AS επιπέδου γραφήματα. Αυτό το γράφημα είναι πιθανόν πιο ακριβές από το Ark IPv4Pref AS γράφημα επειδή το προηγούμενο προέρχεται από το επιπέδου δρομολογητή γράφημα του Διαδικτύου το οποίο είναι μια πιο ακριβής αναπαράσταση του Διαδικτύου από το IP επιπέδου γράφημα. Ανάμεσα στα δύο βασισμένα στο δρομολογητή AS επιπέδου γραφήματα, το παρατηρούμενο δρομολογητή πιο κοντά αντικατοπτρίζει τα παρατηρούμενα μονοπάτια και έτσι συλλαμβάνει κάποιους περιορισμούς της πολιτικής που δεν εκφράζονται στο συναγόμενου δρομολογητή γράφημα.



Σχήμα 37: CCDF του βαθμού κόμβου των βασισμένων στο Ark AS

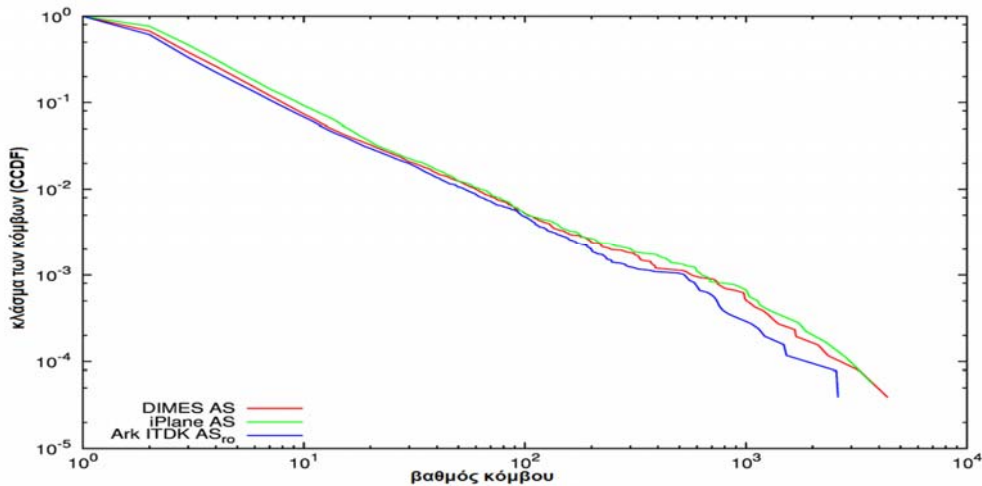


Σχήμα 38: Μέσος βαθμός γείτονα των βασισμένων στο Ark AS

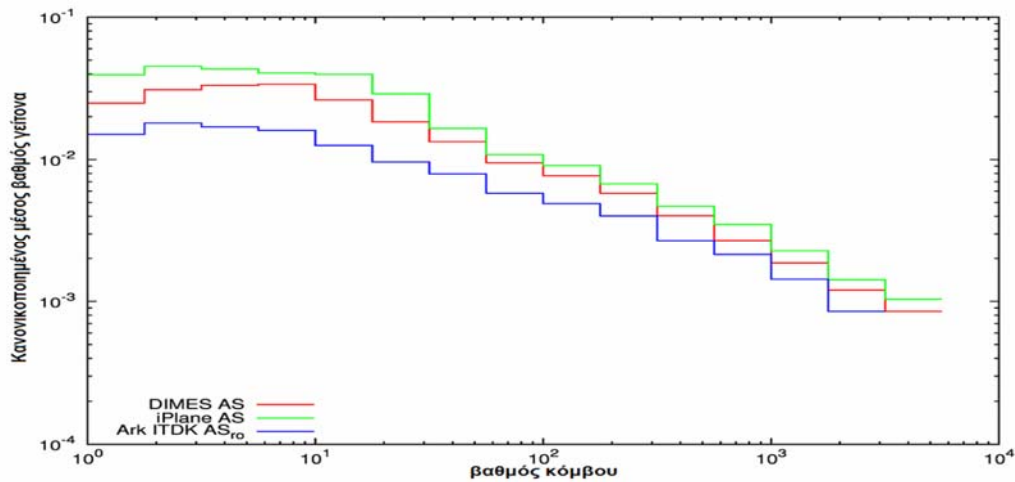


Σχήμα 39: Τοπική ομαδοποίηση των βασισμένων στο Ark AS

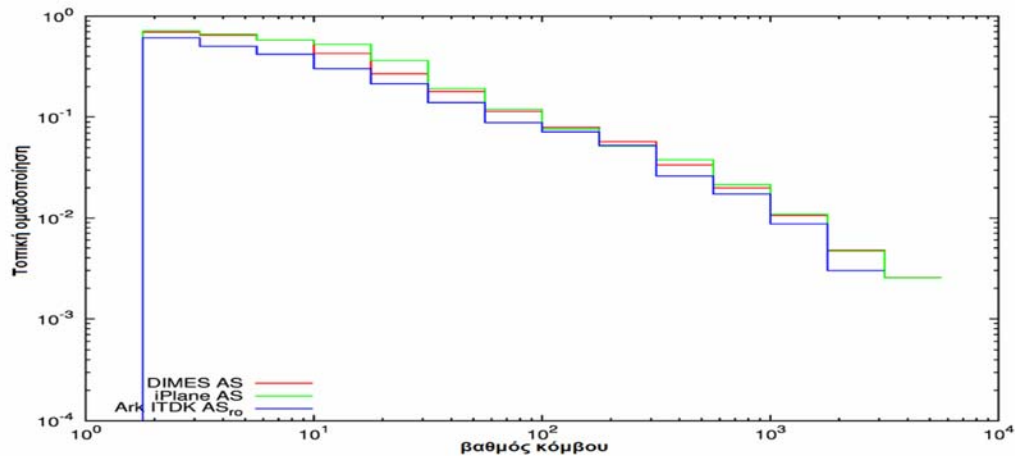
Διαφορές μεταξύ των βασισμένων στο traceroute AS γραφημάτων: Στα σχήματα (40, 41, 42) συγκρίνεται η Ark ITDK AS_{r0} AS επιπέδου τοπολογία (η μπλέ γραμμή) με τα άλλα δύο βασισμένα στο traceroute AS γραφήματα, το DIMES AS (η κόκκινη γραμμή) και το iPlane AS (η πράσινη γραμμή). Οι CCDFs του βαθμού κόμβου (σχήμα 40) και η τοπική ομαδοποίηση (σχήμα 42) είναι παρόμοιες για τα τρία γραφήματα. Για κάθε τιμή του βαθμού κόμβου, ο μέσος βαθμός γείτονα είναι ο υψηλότερος για το iPlane AS γράφημα και ο χαμηλότερος για το Ark ITDK AS_{r0} γράφημα (σχήμα 41). Υπάρχουν τρεις πιθανοί λόγοι για αυτή τη παρατηρούμενη αύξηση του μέσου βαθμού γείτονα μεταξύ των Ark ITDK AS_{r0} και iPlane AS γραφημάτων: η μεταβολή στο σύνολο των γειτόνων, η αύξηση στους βαθμούς αυτών των γειτόνων ή ένας συνδυασμός και των δύο.



Σχήμα 40: CCDF του βαθμού κόμβου των βασισμένων στο traceroute AS



Σχήμα 41: Μέσος βαθμός γείτονα των βασισμένων στο traceroute AS



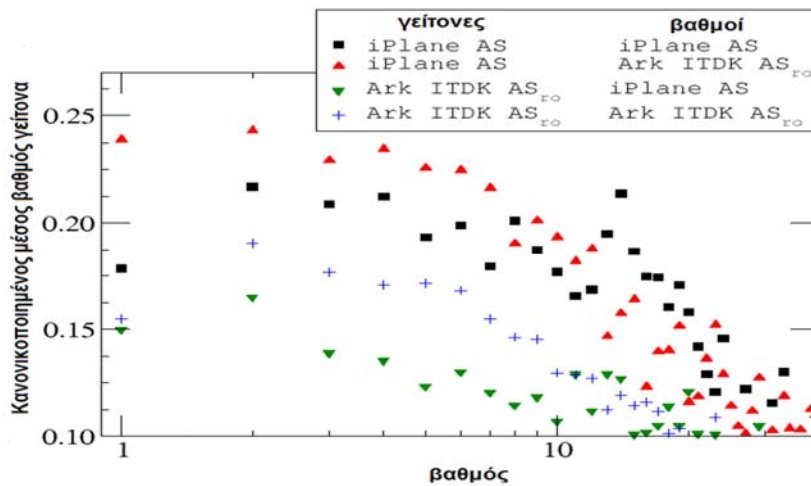
Σχήμα 42: Τοπική ομαδοποίηση των βασισμένων στο traceroute AS

Προσπαθώντας κανείς να αποσαφηνίσει (untangle) τη σχετική συμβολή αυτών των δύο παραγόντων στο μέσο βαθμό γείτονα, μπορεί να εκτελέσει την ακόλουθη προσομοίωση. Μπορεί κάποιος να θεωρήσει το Ark ITDK AS_{r0} γράφημα και το iPlane AS γράφημα και δύο προσομοιωμένα γραφήματα που λαμβάνονται από την εναλλαγή είτε του συνόλου των γειτόνων που έχει κάθε AS είτε των βαθμών αυτών των γειτόνων ανάμεσα στα δύο πραγματικά γραφήματα. Το σχήμα 43 παρουσιάζει τέσσερα σχέδια, ένα για κάθε μία από τις πιθανές παραλλαγές:

- ■ το πραγματικό iPlane AS γράφημα
- ▲ το σύνολο των γειτόνων από το iPlane AS γράφημα με τους βαθμούς για αυτούς τους γείτονες που μπορεί να δει κανείς στο Ark ITDK AS_{r0} γράφημα
- ▼ το σύνολο των γειτόνων από το Ark ITDK AS_{r0} γράφημα με τους βαθμούς για αυτούς τους γείτονες στο iPlane AS γράφημα
- + το πραγματικό Ark ITDK AS_{r0} γράφημα

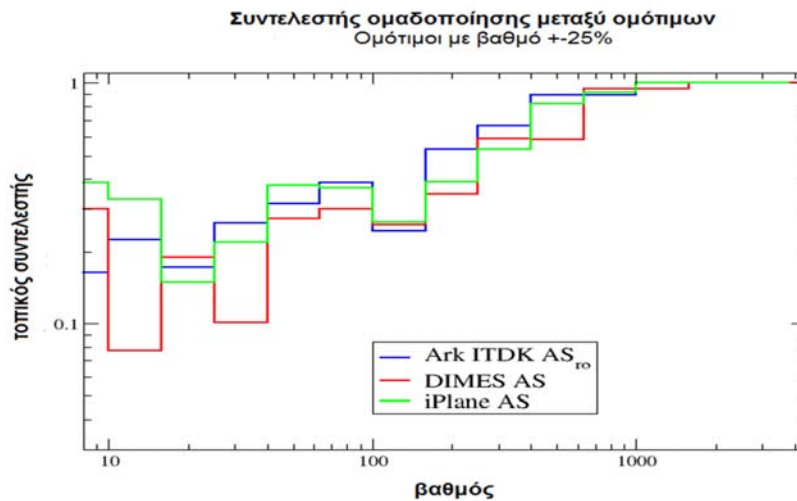
Εάν υπολογισθεί ο μέσος βαθμός γείτονα χρησιμοποιώντας το σύνολο των γειτόνων που μπορεί κάποιος να δει στο iPlane AS, αλλά κρατηθεί το σύνολο των βαθμών αυτών των

γειτόνων που μπορεί να δει κανείς στο Ark ITDK AS_{ro} μπορεί κάποιος να δει το μεγαλύτερο μέσο βαθμό γείτονα (▲) για κόμβους με βαθμούς $k < 20$. Αυτός ο συνδυασμός έχει μεγαλύτερο μέσο βαθμό από αυτό που μπορεί κάποιος να δει είτε στο iPlane AS (■) είτε στο Ark ITDK AS_{ro} (+). Εάν κρατήσει κανείς το σύνολο των γειτόνων που βρέθηκαν στο Ark ITDK AS_{ro}, αλλά χρησιμοποιήσει τους βαθμούς αυτών των γειτόνων όπως αυτοί βρέθηκαν στο iPlane AS, παρατηρεί το μικρότερο μέσο κανονικοποιημένο βαθμό γείτονα (▼) για όλα τα k . Αυτή η ακολουθία σημαίνει ότι από τις δύο πιθανές μεταβολές: μεταβολές στους AS βαθμούς ή μεταβολές σε όποια σύνολα από ASes είναι συνδεδεμένα (οι γείτονές τους), είναι η μεταβολή στις συνδέσεις προς κόμβους με υψηλότερους βαθμούς που είναι υπεύθυνη για τη διαφορά στο μέσο βαθμό γείτονα μεταξύ των δύο γραφημάτων.



Σχήμα 43: Συμβολή του βαθμού των γειτόνων στο μεγαλύτερο μέσο βαθμό γείτονα στο iPlane AS γράφημα έναντι του Ark ITDK ASro

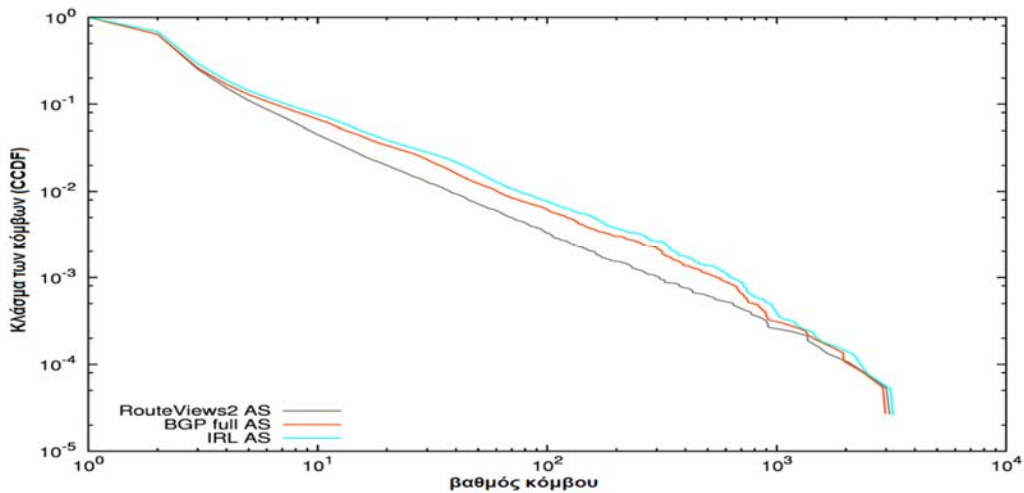
Κατά τη μελέτη των AS σχέσεων στο πραγματικό κόσμο, συχνά μπορεί κάποιος να υποθέσει ότι τα ASes που είναι σε παρόμοια επίπεδα στην AS ιεραρχία εμπλέκονται σε σχέσεις ανταλλαγής κίνησης για να μειωθούν τα κόστη διέλευσης. Η εκδήλωση αυτής της εικασίας στα AS επίπεδα γραφήματα είναι μια τάση να σχηματίζονται κλίκες ανάμεσα σε AS παρόμοιου μεγέθους. Στο σχήμα 44 εξετάζεται η συμπεριφορά της τοπικής ομαδοποίησης εάν συμπεριλαμβάνονται μόνο κόμβοι περίπου του ίδιου μεγέθους στους υπολογισμούς ομαδοποίησης, σε αυτή τη περίπτωση, συγκεκριμένα οι γείτονες που έχουν βαθμούς εντός $\pm 25\%$ ο ένας από τον άλλο. Σε αντίθεση με το σχήμα 42 όπου η τοπική ομαδοποίηση είναι μία φθίνουσα συνάρτηση του βαθμού κόμβου, η τοπική ομαδοποίηση γίνεται μία αύξουσα συνάρτηση του βαθμού κόμβου στο σχήμα 44. Αυτή η προσομοίωση επιβεβαιώνει την υπόθεση σχηματισμού κλίκας μεταξύ ASes παρόμοιου μεγέθους, που περιγράφεται παραπάνω.



Σχήμα 44: Τοπική ομαδοποίηση υπολογισμένη για υπογραφήματα σχηματισμένα από κόμβους με βαθμούς εντός 25% του ενός από του άλλου

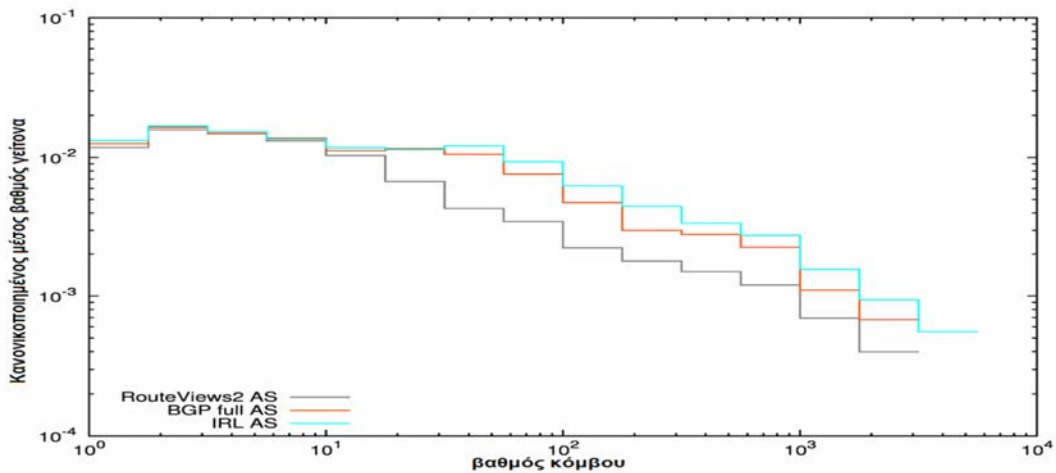
Διαφορές ανάμεσα σε βασισμένα στο BGP AS γραφήματα: Στη συνέχεια τα τρία βασισμένα στο BGP γραφήματα: το RouteViews2 που δημιουργείται από το μεγαλύτερο BGP συλλέκτη, RouteViews2 εξυπηρετητή, το BGP Full που προέρχεται από όλους τους διαθέσιμους BGP εξυπηρετητές (5 στο RouteViews2 και 14 στο RIPE NCC RIS) και το IRL δημιουργημένο από το IRL από πολλαπλές πηγές. Ο πίνακας 8 δείχνει ότι είναι οι πιο συνεισφέροντες σε ένα δεδομένο σύνολο δεδομένων, σε ακμές και με τον υψηλότερο μέσο βαθμό και ομαδοποίηση της προκύπτουσας τοπολογίας. Το αποτέλεσμα αυτό είναι ευκολονόητο: περισσότερες πλεονεκτικές θέσεις, περισσότερες ακμές μπορούν να παρατηρούν, συγκεκριμένα εφραπτόμενους συνδέσμους ανάμεσα σε χαμηλού και μέσου βαθμού κόμβους [30].

Τα σχήματα (45, 46, 47) απεικονίζουν τις τρεις τοπολογικές μετρήσεις για τα τρία βασισμένα στο BGP AS επιπέδου γραφήματα: το CCDF του βαθμού κόμβου, το μέσο βαθμό γείτονα και τη τοπική ομαδοποίηση ως συνάρτηση του βαθμού κόμβου. Η x τιμές (βαθμοί) είναι ομαδοποιημένες χρησιμοποιώντας τέσσερις κάδους ανά δεκαετία. Στο σχήμα 45 η CCDF του βαθμού κόμβου, επιβεβαιώνει ότι το IRL γράφημα (η γαλάζια γραμμή), είναι δημιουργημένο από το μεγαλύτερο αριθμό διαφορετικών συντελεστών (Route Views, RIPE-NCC RIS, PCH, traceroute.org, bgr4.net) έχει ένα ελαφρώς μεγαλύτερο ποσοστό από υψηλού βαθμού κόμβους από τα άλλα δύο γραφήματα: 0.77% των IRL AS κόμβων έχει βαθμό μεγαλύτερο από 100, σε σύγκριση με το 0.62% των δεδομένων του BGP Full AS και το 0.33% του RouteViews2 AS. Αν και αυτοί οι υψηλού βαθμού κόμβοι αποτελούν ένα μικρό μόνο κλάσμα των συνολικών γραφημάτων, αντιπροσωπεύουν τη κορυφή της ιεραρχίας δρομολόγησης του Διαδικτύου, εξυπηρετώντας μία κρίσιμη λειτουργία δρομολόγησης. Αξίζει να σημειωθεί ότι το AS 3356 (Level 3) και το AS 174 (Cogent) είναι σταθερά κατατασσόμενα πρώτο και δεύτερο και τα ASes 7018(ATT) και 3549 (Global Crossing) είναι κατατασσόμενα τρίτο και τέταρτο σε όλα τα σύνολα δεδομένων εκτός από του WHOIS. Ωστόσο τα κλάσματα των κόμβων με μία τάξη μεγέθους μεγαλύτερων βαθμών (> 1000) είναι παρόμοια και στα τρία γραφήματα: 0.04%, 0.03% και 0.02% για το IRL AS, BGP Full AS και RouteViews2 AS αντίστοιχα. Αυξάνοντας τον αριθμό και τη ποικιλία των χορηγών BGP δεδομένων φαίνεται να αποκαλύπτεται επιπλέον συνδεσιμότητα κυρίως για κόμβους με μέσους βαθμούς.



Σχήμα 45: CCDF του βαθμού κόμβου των βασισμένων στο BGP AS

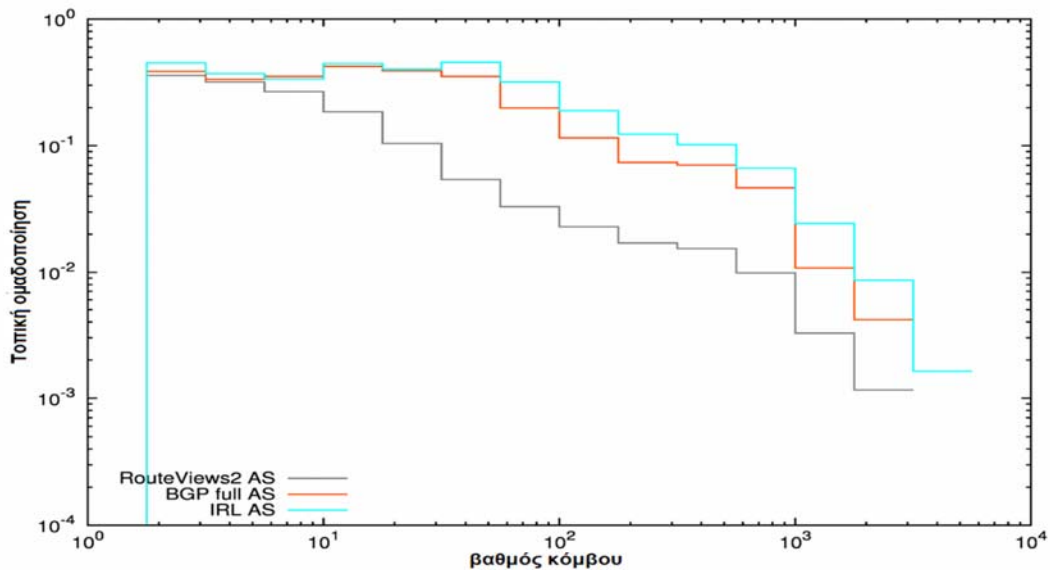
Στο σχήμα 46 απεικονίζεται ο μέσος βαθμός γείτονα έναντι του βαθμού κόμβου. Για μικρούς βαθμούς ($k < 10$ για το RouteViews2 AS γράφημα, $k < 70$ για το IRL AS και BGP Full AS γράφημα) ο μέσος βαθμός γείτονα είναι σχεδόν σταθερός και γίνεται μία φθίνουσα συνάρτηση του βαθμού κόμβου σε μεγαλύτερους βαθμούς. Στα AS επιπέδου γραφήματα είναι γνωστό ότι οι κόμβοι με χαμηλό βαθμό είναι πιο πιθανό να συνδέονται με κόμβους υψηλού βαθμού [30]. Οι επίπεδες περιοχές του BGP Full AS και του IRL AS για ASes με βαθμούς ανάμεσα σε 10 και 50 δείχνουν μία πυκνότερη συνδεσιμότητα ανάμεσα σε μέσης βαθμίδας ASes που παρατηρείται, οφείλεται σε ένα μεγαλύτερο αριθμό πλεονεκτικών θέσεων που χρησιμοποιούνται για τη συλλογή ακατέργαστων δεδομένων.



Σχήμα 46: Μέσος βαθμός γείτονα των βασισμένων στο BGP AS

Το σχήμα 47 δείχνει ότι όσο ο βαθμός κόμβου αυξάνεται, η τοπική ομαδοποίηση πέφτει πολύ πιο γρήγορα για το RouteViews2 (η μαύρη γραμμή) από ότι για τα άλλα δύο γραφήματα. Σε αντίθεση, για τις BGP Full AS και IRL AS τοπολογίες η τοπική ομαδοποίηση είναι περίπου σταθερή ή ακόμα αυξάνεται αμυδρά (slightly) για μικρού βαθμού κόμβους και αρχίζει να μειώνεται μόνο για βαθμούς πάνω από 50. Ο μεγαλύτερος αριθμός από πλεονεκτικά σημεία συλλαμβάνει περισσότερους εφαιπόμενους συνδέσμους ανάμεσα σε μικρούς κόμβους. Το BGP Full AS γράφημα που προέρχεται από ένα συνδυασμό πολλαπλών BGP πινάκων είναι σημαντικά πιο ολοκληρωμένο από το RouterViews2 AS γράφημα που προέρχεται από ένα μόνο BGP πίνακα, αλλά χρησιμοποιώντας ένα συνδυασμό των επτά διαφορετικών χορηγών

στη περίπτωση του IRL AS γραφήματος δεν προσθέτει αρκετά στη συνδεσιμότητα που έχει ήδη συλληφθεί από τους BGP πίνακες.



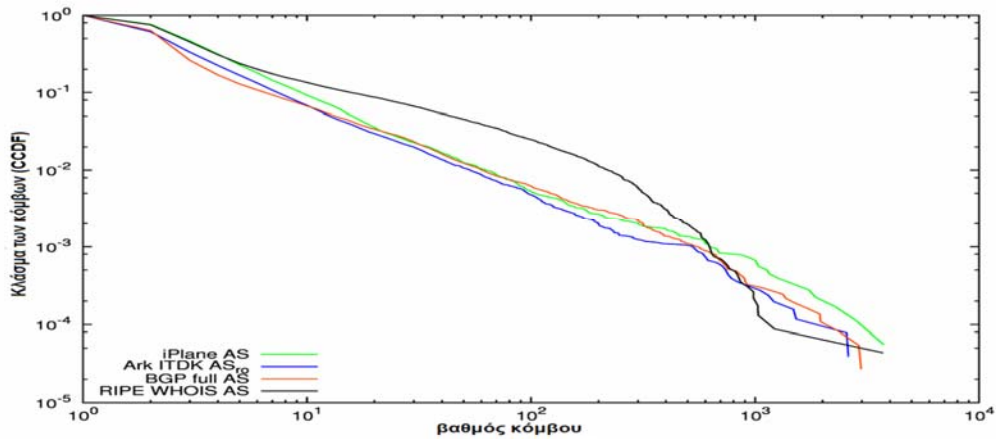
Σχήμα 47: Τοπική ομαδοποίηση των βασισμένων στο BGP AS

Όλα τα χαρακτηριστικά των BGP Full AS (η κόκκινη γραμμή) και IRL AS (η γαλάζια γραμμή) γραφημάτων που παρουσιάζονται στα σχήματα (45, 46, 47) είναι παρόμοια για όλους τους βαθμούς κόμβου, γεγονός που υποδηλώνει ότι ο συνδυασμός των BGP πινάκων που χρησιμοποιείται στο BGP Full AS σύνολο δεδομένων συλλαμβάνει ένα αντιπροσωπευτικό δείγμα της υποβόσκουσας AS τοπολογίας ακόμα και με λιγότερους συντελεστές από το IRL AS σύνολο δεδομένων. Συνεπώς επιλέγεται το BGP Full AS γράφημα ως ένα αντιπροσωπευτικό προερχόμενο από το BGP AS επιπέδου γράφημα για τη συνολική σύγκριση.

Όλα τα AS επιπέδου γραφήματα: Η τελευταία σύγκριση περιλαμβάνει ένα μόνο αντιπροσωπευτικό AS τοπολογίας γράφημα από κάθε μία από τις προηγούμενες τρεις AS επιπέδου συγκρίσεις και το RIPE WHOIS AS γράφημα το οποίο υφίσταται ως μία κατηγορία από μόνο του. Πρέπει να σημειωθεί ότι η περιοχή της RIPE NCC υπηρεσίας αποτελείται από χώρες της Ευρώπης, της Μέσης Ανατολής και της Κεντρικής Ασίας, οπότε το AS γράφημα που προέρχεται από τη WHOIS βάση δεδομένων τους αντιπροσωπεύει κυρίως την συνδεσιμότητα στην Ευρώπη. Συγκρίνοντας τις στατιστικές ιδιότητες των AS επιπέδου γραφημάτων [30] που προέρχονται από BGP πίνακες, traceroute μετρήσεις και WHOIS δεδομένα τα μέλη του Caida διερεύνησαν εάν η σημαντική διαφορά στις τοπολογικές ιδιότητες ανάμεσα στο βασισμένο στο WHOIS γράφημα και στα άλλα δύο γραφήματα μπορεί να εξηγηθεί από τις γεωγραφικές διακρίσεις στα δεδομένα. Το συμπέρασμα στο οποίο κατέληξαν, ήταν ότι η γεωγραφική διάκριση δεν μπορεί να εξηγήσει πλήρως τη διαφορά καθώς όταν λήφθηκε το υποσύνολο της τοπολογίας σύνδεσης των κόμβων και από το BGP και από το WHOIS γράφημα, τα προκύπτοντα γραφήματα διατήρησαν τις κανονικοποιημένες τοπολογικές ιδιότητες των αρχικών γραφημάτων.

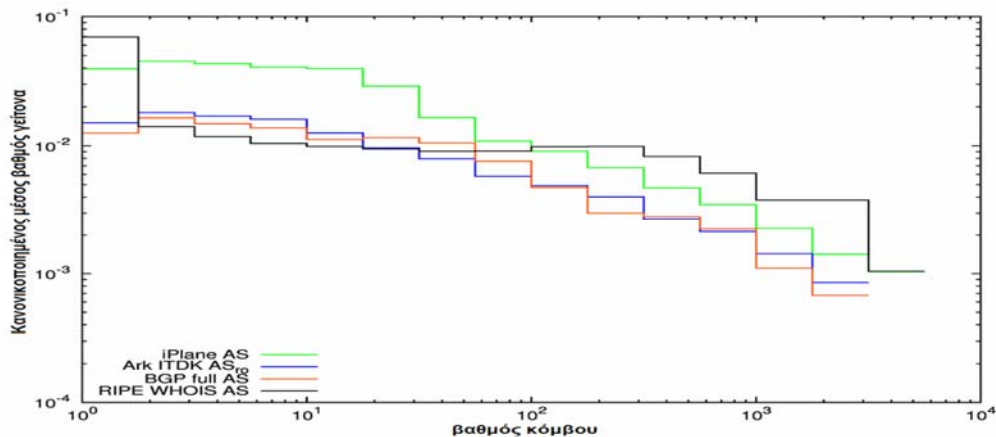
Στο σχήμα 48 η CCDF των βαθμών κόμβου, δείχνει ότι τα BGP Full AS (η κόκκινη γραμμή) και Ark ITDK AS_{ro} (η μπλέ γραμμή) γραφήματα έχουν σχετικά υψηλότερα κλάσματα ακμών ASes με βαθμούς 1 και 2 (36% και 39% έναντι 25% στο RIPE WHOIS AS γράφημα (η μαύρη γραμμή) και 23% στο iPlane AS γράφημα (η πράσινη γραμμή)). Σε σύγκριση με τα άλλα τρία γραφήματα, το RIPE WHOIS AS γράφημα έχει περισσότερους κόμβους με μέσους βαθμούς

ανάμεσα σε 5 και 500. Το γράφημα αυτό είναι πολύ διαφορετικό από τα άλλα τρία στο ότι στη κατανομή βαθμού δεν ταιριάζει η συνάρτηση του νόμου δύναμης, ενώ τα άλλα τρία μπορούν να χαρακτηριστούν ως νόμου της δύναμης με εκθέτη $\gamma = 2.1 - 2.2$. Το iPlane AS έχει το μεγαλύτερο κλάσμα από ASes με βαθμό >1000 (0.07% σε σύγκριση με το 0.03% ή λιγότερο για τα Ark ITDK AS r_o , BGP Full AS και WHOIS RIPE AS γραφήματα.



Σχήμα 48: CCDF του βαθμού κόμβου του συνόλου των AS επιπέδου

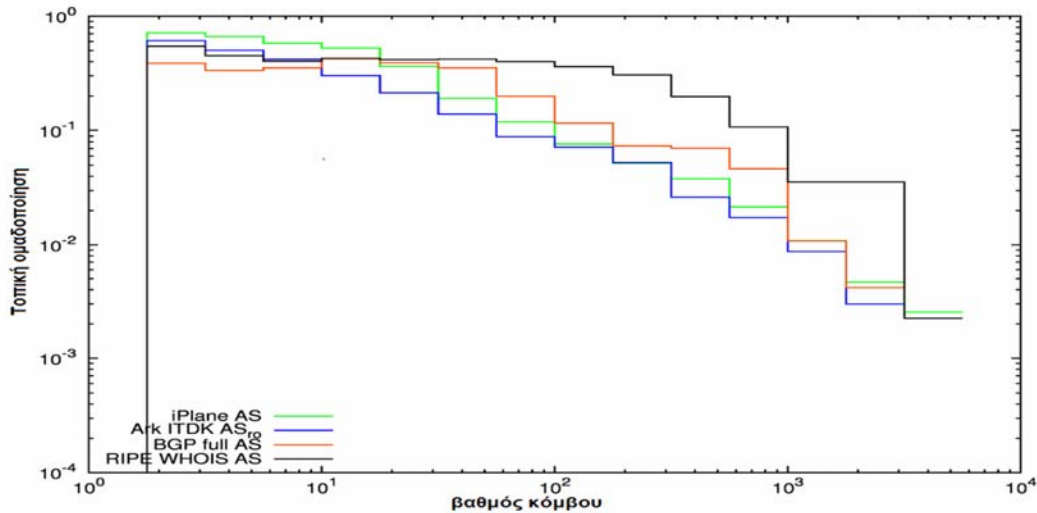
Λαμβάνοντας υπόψη το μέσο βαθμό γείτονα (σχήμα 49) παρατηρούμε ότι το RIPE WHOIS AS γράφημα (η μαύρη γραμμή) έχει το μεγαλύτερο μέσο βαθμό γείτονα για ASes με βαθμό 1. Και στα τέσσερα γραφήματα ο μέσος AS βαθμός γείτονα μειώνεται όσο ο AS βαθμός αυξάνεται (δηλαδή στα AS γραφήματα οι κόμβοι χαμηλού βαθμού είναι πιο πιθανό να συνδεθούν με κόμβους υψηλού βαθμού), αν και για το RIPE WHOIS AS γράφημα παραμένει σχεδόν σταθερός για βαθμούς μεταξύ 2 και 200 και μόνο αρχίζει να μειώνεται για μεγαλύτερους βαθμούς. Η συμπεριφορά αυτή αντικατοπτρίζει ένα σχετικά μεγάλο αριθμό από μέσου βαθμού κόμβους στο γράφημα. Ανάμεσα στα άλλα τρία γραφήματα, το iPlane AS (η πράσινη γραμμή) έχει τον υψηλότερο μέσο βαθμό γείτονα σε όλες τις κλίμακες βαθμού ενώ η ίδια κατανομή για τα BGP Full AS και Ark ITDK AS r_o γραφήματα είναι χαμηλότερα και παρόμοια μεταξύ τους.



Σχήμα 49: Μέσος βαθμός γείτονα του συνόλου των AS επιπέδου

Το RIPE WHOIS AS γράφημα (η μαύρη γραμμή) βρίσκεται επίσης εκτός από τα άλλα γραφήματα στο σχήμα 50, το οποίο αναπαριστά τη τοπική ομαδοποίηση ως συνάρτηση του βαθμού κόμβου. Για αυτό το γράφημα, η τοπική ομαδοποίηση παραμένει σχεδόν σταθερή (και ως επί το πλείστον υψηλότερη απ' ό,τι για τα άλλα τρία γραφήματα) για βαθμούς κόμβου

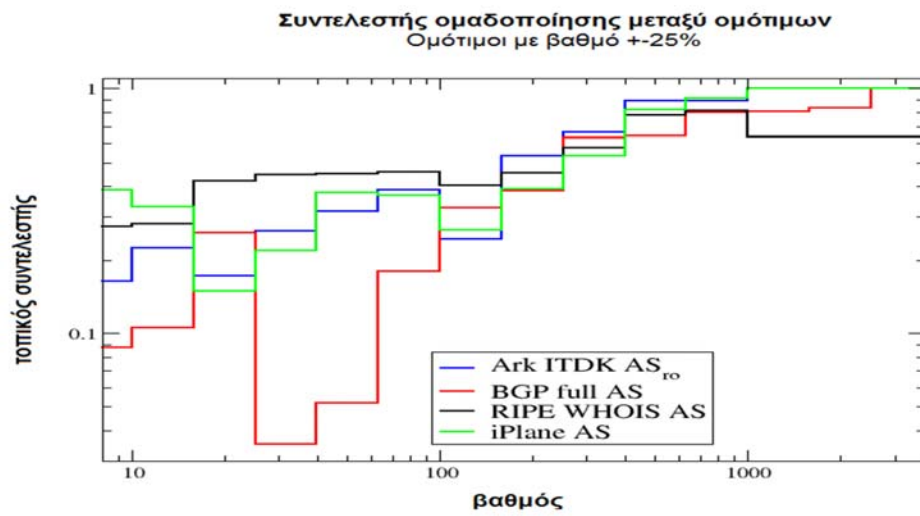
< 200. Συγκρίνοντας τις μαύρες RIPE WHOIS AS γραμμές στα σχήματα 49 και 50, μπορεί κανείς να δει ότι τα σημεία καμπής και στις δύο γραφικές παραστάσεις προκύπτουν γύρω από βαθμό κόμβου 200. Αυτή η σύμπτωση μπορεί να σημαίνει ότι όσο ο μέσος βαθμός γείτονα μειώνεται, οι γείτονες αυτοί δεν έχουν ένα αρκετά υψηλό βαθμό για να σχηματίσουν ομάδες από τη σύνδεση με άλλους (πάρα πολλούς) γείτονες ενός συγκεκριμένου υψηλού βαθμού κόμβου.



Σχήμα 50: Τοπική ομαδοποίηση του συνόλου των AS επιπέδου

Το σχήμα 51 δείχνει ξανά τη τοπική ομαδοποίηση υπολογισμένη για τα υπογραφήματα που σχηματίζονται από κόμβους με παρόμοιους βαθμούς, δηλαδή μέσα σε ένα 25% του ενός από το άλλο. Σε αυτή τη περίπτωση, η RIPE WHOIS AS ομαδοποίηση είναι παρόμοια με εκείνη των άλλων γραφημάτων. Όπως στα σχήματα (45, 46, 47), η τοπική ομαδοποίηση αυξάνεται με το βαθμό κόμβου και οι σύνδεσμοι μεταξύ των μικρότερων κόμβων ρίχνουν τη μέση ομαδοποίηση. Δεδομένου ότι το RIPE WHOIS AS γράφημα έχει ένα χαμηλότερο ποσοστό από μικρού βαθμού κόμβους (το σχήμα 48 δείχνει ένα σχετικά μεγαλύτερο αριθμό από κόμβους στη βαθμολογική κλίμακα 100 – 600), λιγότεροι από τους συνδέσμους θα συνδεθούν σε αυτούς, κάτι το οποίο καθιστά τη τοπική ομαδοποίηση υψηλότερη. Λαμβάνοντας υπόψιν τα άλλα τρία γραφήματα στο σχήμα 50, μπορεί κανείς να παρατηρήσει ότι η τοπική ομαδοποίηση στο iPlane AS γράφημα (η πράσινη γραμμή) είναι ελαφρώς υψηλότερη ή η ίδια όπως στο Ark ITDK AS $_{ro}$ γράφημα (η μπλέ γραμμή) σε κάθε τιμή βαθμού. Η τοπική ομαδοποίηση του βασισμένου στο BGP BGP Full AS γραφήματος (η κόκκινη γραμμή) είναι χαμηλότερη από εκείνη των βασισμένων στο traceroute iPlane AS και Ark ITDK AS $_{ro}$ γραφημάτων για χαμηλούς βαθμούς < 10, αλλά είναι υψηλότερη για μέσους βαθμούς $10 < k < 800$. Σύμφωνα με το πώς τα BGP δεδομένα συλλέγονται έναντι των traceroute δεδομένων, τα BGP γραφήματα ρίχνουν περισσότερο φως στα υψηλού βαθμού ASes απ' ότι στα περιφερειακά και αντιστρόφως, οι traceroute υποδομές με πλεονεκτικά σημεία διάσπαρτα στη περιφέρεια θα συλλάβουν σχετικά περισσότερους μικρού βαθμού κόμβους.

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida



Σχήμα 51: Τοπική ομαδοποίηση για υπογραφήματα από κόμβους με βαθμούς εντός 25% του ενός από του άλλου

4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην πτυχιακή αυτή εργασία έγινε προσπάθεια χαρτογράφησης με μετρήσεις βασισμένες στο CAIDA του πολύπλοκου αυτού δικτύου που ονομάζεται Διαδίκτυο. Μέσα από τη διαδικασία αυτή της χαρτογράφησης έγινε δυνατή σε μεγάλο βαθμό η αποσαφήνιση της δομής και της οργάνωσης της υποδομής του Διαδικτύου. Η χαρτογράφηση πρόσφερε μια καλή εικόνα της συνδεσιμότητας στο Διαδίκτυο και επέτρεψε τη κατανόηση των ιδιοτήτων του, της δυναμικής και της εξέλιξής του.

Το Διαδίκτυο όμως συνεχώς μεγαλώνει το πλήθος των διασυνδεδεμένων συσκευών και η πολυπλοκότητα του αυξάνεται. Ως εκ τούτου η έγκυρη και πλήρης χαρτογράφηση του Διαδικτύου δεν είναι μια εύκολη υπόθεση. Ο CAIDA με το πρόγραμμα της Μακροσκοπικής Τοπολογίας του Διαδικτύου, με το μεγάλο αριθμό εξειδικευμένων εργαλείων λογισμικού που προσφέρει και με το μεγάλο δίκτυο συσκευών και υποδομών παρακολούθησης που διαθέτει, αποτελεί τη μεγαλύτερης κλίμακας και μεγαλύτερης ακρίβειας προσπάθεια της επιστημονικής κοινότητας για τη χαρτογράφηση του Διαδικτύου. Η προσπάθεια αυτή του CAIDA είναι πράγματι σπουδαία αν αναλογιστεί κανείς το πλήθος των δεδομένων που επεξεργάζεται και προσφέρει και τις συνεχείς μελέτες που εκπονούν τα μέλη του στη προσπάθεια χαρτογράφησης του Διαδικτύου δίνοντας έμφαση στην ακεραιότητα και την υγεία του παγκόσμιου οικοσυστήματος του Διαδικτύου. Παρόλα αυτά το Διαδίκτυο παραμένει λόγω του μεγέθους του σε κάποιο βαθμό αχαρτογράφητο και οι χάρτες που παράγονται δεν χαρακτηρίζονται από πλήρη εγκυρότητα καθώς τα σύνολα δεδομένων από τα οποία προκύπτουν δεν χαρακτηρίζονται αντίστοιχα από πλήρη εγκυρότητα.

Ο CAIDA επιδιώκει να χαρακτηρίζεται η μακροσκοπική Διαδικτυακή συνδεσιμότητα με χρήση και τοπολογικών και γεωγραφικών αναπαραστάσεων σε πολλαπλά επίπεδα λεπτομέρειας της συνάθροισης. Ο ερευνητικός αυτός όμως στόχος είναι ιδιαίτερα δύσκολος λόγω αντιφάσεων σε διάφορα εργαλεία μέτρησης και αντικινήτρων για τους ISPs να υποστηρίξουν εργαλεία που παράγουν συμπεράσματα σχετικά με τη συνδεσιμότητα. Παρόλα αυτά γίνεται συνεχής προσπάθεια από τον CAIDA να βελτιωθούν τα παραγόμενα αποτελέσματα και να ανταποκρίνονται σε μεγαλύτερο βαθμό στη πραγματικότητα. Αυτό προσπαθούν να το επιτύχουν μέσα από μια συστηματική σύγκριση και μελέτη των τοπολογιών που έχουν συναχθεί από τις καλύτερες διαθέσιμες πηγές δεδομένων. Συγκρίνονται διαγράμματα τοπολογίας διαφορετικών βαθμών ανάλυσης που προκύπτουν από μεγάλο αριθμό διαφορετικών πηγών δεδομένων. Κάθε ένα από τα αποτελέσματα που προκύπτουν φωτίζει διαφορετικό κομμάτι της τοπολογίας του Διαδικτύου. Δίνοντας έτσι στο σύνολό τους μια αρκετά έγκυρη και ασφαλής εικόνα της τοπολογίας του Διαδικτύου. Εντούτοις υπάρχει ακόμα απόσταση από τη πλήρως έγκυρη και πλήρη χαρτογράφηση του Διαδικτύου.

Χρειάζεται ακόμα μεγαλύτερη συμμετοχή ιδρυμάτων του εμπορικού, του κυβερνητικού και του ερευνητικού τομέα στο πρόγραμμα του CAIDA με την εγκαθίδρυση στο δίκτυο τους συσκευών παρακολούθησης που ανήκουν στην υποδομή του CAIDA το Archipelago. Έτσι θα διευρυνθεί το οπτικό πεδίο του παγκόσμιου Διαδικτύου για την ερευνητική κοινότητα του Διαδικτύου. Κάθε επιπλέον κόμβος παρακολούθησης συμβάλλει σημαντικά στη προσπάθεια του CAIDA και αυξάνει τη πληρότητα και την ακρίβεια των δεδομένων που αντιπροσωπεύουν τη τοπολογική δομή του Διαδικτύου.

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	Ελληνικός Όρος
network layer	επίπεδο δικτύου
routing	δρομολόγηση
routing tables	πίνακες δρομολόγησης
dynamic routing protocols	πρωτόκολλα δυναμικής δρομολόγησης
Link – State	Κατάστασης Ζεύξης
Distance – Vector	Διανύσματος Απόστασης
Open Shortest Path First	ανοικτό – πρώτα η βραχύτερη διαδρομή
Routing Information Protocol	πρωτόκολλο Δρομολόγησης Πληροφοριών
Interior Gateway Routing Protocol	Εσωτερικής Πύλης πρωτόκολλο δρομολόγησης
autonomous systems	αυτόνομα συστήματα
Internet Service Providers	εταιρίες παροχής υπηρεσιών Διαδικτύου
Internet Exchange Points	σημεία ανταλλαγής Διαδικτύου
Content Delivery Networks	δικτύων διανομής περιεχομένου
peers	ομότιμοι
customer – client	παρόχου – πελάτη
core	κορμού
Interior Routing Protocols	πρωτόκολλα εσωτερικής δρομολόγησης
border router	Παραμεθόριος δρομολογητής
exterior routing	εξωτερική δρομολόγηση
Exterior Gateway Protocol	πρωτόκολλο εξωτερικής δρομολόγησης
interdomain	διατομεακής
session	σύννοδος
path	μονοπάτι
NEXT-HOP	επόμενου άλματος
granularities	Βαθμοί ανάλυσης
points of presence	σημείων παρουσίας
level	επιπέδου
peering	ομοτιμία
addresses	διευθύνσεων
content providers	παρόχων περιεχομένου
traffic flowing	ροής κυκλοφορίας
congestion	συμφόρησης
hijacking	κλοπή
impact	επίπτωση
dynamics	δυναμικών

Broadband Internet Technical Advisory Group	Ευρυζωνικού Διαδικτύου Τεχνική Συμβουλευτική Ομάδα
funding source	πηγή χρηματοδότησης
operating expenses	λειτουργικών εξόδων
raw	ανεπεξεργαστα
Network Telescope	διαδικτυακού τηλεσκοπίου
distributed	κατανεμημένη
dataset	συλλογής δεδομένων
load balancing	δικτυακού φόρτου
general-purpose	γενικού σκοπού
hosted	φιλοξενούμενες
spoofed	πλαστογραφημένη
susceptibility	ευαισθησία
dual-stack	διπλής στοίβας
feasibility	σκοπιμότητα
blind	τυφλού
sampling	δειγματοληψίας
diagnostic queries	διαγνωστικά αιτήματα
namespace	χώρος ονομάτων
consistency	συνεκτικότητας
driven	καθοδηγούμενες
parallelization	παραλληλοποίηση
prefixes	προθέματα
metadata	μεταδεδομένα
constraints	περιορισμούς
diurnal	ημερήσια
shell	κέλυφος
counterincentives	αντικινήτρων
grounded	θεμελιωμένη
Alias Resolution	Επίλυση ψευδωνύμων
interfaces	διεπαφές
subnet	υποδίκτυο
traces	ιχνών
fingerprints	Δακτυλικά αποτυπώματα
probe	διερευνητικό
Sliding window	συρόμενου παραθύρου
lightweight	ελαφρύ
linear	γραμμικές
arbitrary	αυθαίρετων
directory	καταλόγου
polar	πολικές
transit	διέλευση
infer	συμπέρανε
bins	κάδους
box-and-whisker plot	διάγραμμα πλαισίου
annotated	σχολιασμένο
backbone	δίκτυο κορμού

Χαρτογράφηση του διαδικτύου με μετρήσεις βασισμένες στο caida

chain	αλυσίδα
invalid	άκυρων
away	μακριά
clustering	ομαδοποίηση
diversity	ποικιλομορφία
timestamps	χρονικές σφραγίδες
hyperlinks	υπερσύνδεσμοι
distorted	διαστρεβλωμένων
interval	μεσοδιαστήματος
tend	τείνουν
cutoff	αποκοπή
untangle	αποσαφηνίσει
slightly	αμυδρά

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

OSI	Open Systems Interconnection
IS	Intermediate Systems
OSPF	Open Shortest Path First
RIP	Routing Information Protocol
IGRP	Interior Gateway Routing Protocol
AS	Autonomous System
IXP	Internet Exchange Point
CDN	Content Delivery Network
ASN	Autonomous System Number
IGP	Interior Routing Protocols
EGP	Exterior Routing Protocol
BGP	Border Gateway Protocol
TCP	Transmission Control Protocol
eBGP	External Border Gateway Protocol
iBGP	Internal Border Gateway Protocol
CIDR	Classless Inter Domain Routing
CAIDA	Center for Applied Internet Data Analysis
IP	Internet Protocol
PoPs	Points-of-Presence
ISP	Internet Service Provider
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
MIT	Massachusetts Institute of Technology
CSAIL	Computer Science and Artificial Intelligence Laboratory
NSF	National Science Foundation
IX	Internet exchange
NDN-NP	Named Data Networking Next Phase
NDN	Named Data Networking
IMC	Internet Measurement Conference

UbiComp	Ubiquitous Computing
HCI	Human-Computer Interaction
BITAG	Broadband Internet Technical Advisory Group
PAM	Passive and Active Measurements
NANOG	North American Network Operators' Group
AIMS	Active Internet Measurements
CREDS	Cyber-security Research Ethics Dialog & Strategy
NDNcomm	Named Data Networking Community Meeting
UCLA	University of California, Los Angeles
IMAPS	Internet Measurement And Political Science
WIE	Workshop on Internet Economics
GiB	Gibibyte
TiB	Tebibyte
DNS	Domain Name System
UCSD	University of California, San Diego
ITDK	Internet Topology Data Kit
gTLD	Generic top-level domains
RTT	Round-Trip Time
DDos	Distributed Denial of Service
ICMP	Internet Control Message Protocol
UDP	User Datagram Protocol
API	Application Programming Interface
NPS	Naval Postgraduate School
TBIT	TCP Behavior Inference Tool
TLD	top-level domain
PMTU	Path Maximum Transmission Unit
DNSSEC	Domain Name System Security Extensions
TTL	Time To Live
TSP	Time Sequence Ping
DHS	Department of Homeland Security
S&T	Science and Technology
RIPE	Réseaux IP Européens

NCC	Network Coordination Centre
MIDAR	Monotonic ID-based Alias Resolution
APAR	Analytic and Probe-based Alias Resolver
MBT	Monotonic Bounds Test
ACK	Acknowledgement
RST	Reset
RasPi	Raspberry Pi
URL	Uniform Resource Locator
ITE	Ίδρυμα Τεχνολογίας και Έρευνας
SNMP	Simple Network Management Protocol
IRR	Internet Routing Registry
RIS	Routing Information Service
IRL	Internet Research Lab
ASR	AS-ranking
PCH	Packet Clearing House
RIR	Regional Internet Registries
WHOIS	Who Is
CCDF	Complementary Cumulative Distribution Function
ID	Identifier
LAN	Local area network
IETF	Internet Engineering Task Force

ΑΝΑΦΟΡΕΣ

- [1] James F. Kurose, Keith W. Ross, Computer Networking: A Top-Down Approach (6th Edition), Pearson 2012
- [2] A. Athanassiadis (2007), *Interdomain Routing*, University of Macedonia, http://conta.uom.gr/conta/ekpaideysh/metaptyxiaka/technologies_diktywn/ergasies/2007/Interdomain%20Routing.pdf, [Προσπελάστηκε 5/7/15]
- [3] Ι. Στ. Βενιέρης, Ε. Νικολούζου, Τεχνολογίες Διαδικτύου, www.icbnet.ntua.gr/Mathimata/InternetTech/kef2.pdf, [Προσπελάστηκε 6/9/16]
- [4] M.G., Berio, S.P., Grynberg, and J.I., Alvarez-Hamelin, Router-level community structure of the Internet Autonomous Systems, EPJ Data Science, 2015
- [5] T. Sheldon, ed. *Encyclopedia of Networking and Telecommunications*, McGraw-Hill, 2001. <http://www.linktionary.com>.
- [6] CAIDA Researchers, CAIDA's Annual Report for 2014, <http://www.caida.org/home/about/annualreports/2014/> [Προσπελάστηκε 7/7/15]
- [7] CAIDA Researchers, Mapping Interconnection in the Internet: Colocation, Connectivity and Congestion, <https://www.caida.org/funding/nets-congestion/> [Προσπελάστηκε 16/7/15]
- [8] CAIDA Researchers, Named Data Networking Next Phase (NDN-NP), <https://www.caida.org/funding/ndn-ndp/>, [Προσπελάστηκε 25/7/15]
- [9] CAIDA Researchers and R. Beverly, Exploring the evolution of IPv6: topology, performance, and traffic, <https://www.caida.org/funding/nets-ipv6/>, [Προσπελάστηκε 1/8/15]
- [10] CAIDA Researchers, Archipelago (Ark) Measurement Infrastructure, <http://www.caida.org/projects/ark/>, [Προσπελάστηκε 8/8/15]
- [11] M. Luckie, K. Keys, R. Koga, Y. Hyun and K. Claffy, Spoofer, <https://www.caida.org/projects/spoofer/>, [Προσπελάστηκε 8/8/15]
- [12] CAIDA Researchers, Internet Mapping: Cartographic Capabilities for Critical Cyberinfrastructure, <http://www.caida.org/funding/c4/>, [Προσπελάστηκε 12/8/15]
- [13] CAIDA Researchers, Scamper, <http://www.caida.org/tools/measurement/scamper/>, [Προσπελάστηκε 14/8/15]
- [14] CAIDA Researchers, Topology Research, <http://www.caida.org/research/topology/>, [Προσπελάστηκε 18/8/15]
- [15] K. Keys, IP Alias Resolution Techniques, http://www.caida.org/publications/papers/2008/alias_resolution_techreport/alias_resolution_techreport_1.0.pdf, [Προσπελάστηκε 6/9/16]
- [16] R. Govindam and H. Tangmunarunkit, Heuristics for Internet Map Discovery, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.46.1484&rep=rep1&type=pdf>, [Προσπελάστηκε 23/8/15]
- [17] CAIDA Researchers, iffinder, <http://www.caida.org/tools/measurement/iffinder/>, [Προσπελάστηκε 25/8/15]
- [18] CAIDA Researchers, kapar, <http://www.caida.org/tools/measurement/kapar/>, [Προσπελάστηκε 29/8/15]
- [19] CAIDA Researchers, MIDAR, <http://www.caida.org/tools/measurement/midar/>, [Προσπελάστηκε 5/9/15]
- [20] K. Keys, Y. Hyun, M. Luckie, and K. Claffy, Internet-Scale IPv4 Alias Resolution with MIDAR, IEEE/ACM Transactions on Networking, vol. 21, no. 2, pp. 383–399, Apr 2013.
- [21] CAIDA Researchers, plot-latlong, <http://www.caida.org/tools/visualization/plot-latlong/>, [Προσπελάστηκε 20/8/16]

- [22] CAIDA Researchers, CAIDA's Macroscopic Internet Topology Data Kit, <http://data.caida.org/datasets/topology/ark/ipv4/itdk/2013-07/README>, [Προσπελάστηκε 27/8/16]
- [23] CAIDA Researchers, Archipelago Monitor Locations, <http://www.caida.org/projects/ark/locations/>, [Προσπελάστηκε 23/10/15]
- [24] Bradley Huffaker, Evi Nemeth and K Claffy, Otter: A general-purpose network visualization tool, <http://www.caida.org/tools/visualization/otter/paper/>, [Προσπελάστηκε 5/11/15]
- [25] CAIDA Researchers, IPv4 and IPv6 AS Core: Visualizing IPv4 and IPv6 Internet Topology at a Macroscopic Scale, https://www.caida.org/research/topology/as_core_network/, [Προσπελάστηκε 23/9/15]
- [26] Bradley Huffaker, Marina Fomenkov, K. Claffy, Internet Topology Data Comparison, <http://www.caida.org/publications/papers/2012/topocompare-tr/topocompare-tr.pdf>, [Προσπελάστηκε 20/6/16]
- [27] T. Bu and D. Towsley, On Distinguishing between Internet Power Law Topology Generators, IEEE INFOCOM, 2002., <http://digg.cs.tufts.edu/readings/pdf/005.pdf>, [Προσπελάστηκε 25/6/16]
- [28] S. Jaiswal, A. L. Rosenberg, and D. Towsley, Comparing the Structure of Power-Law Graphs and the Internet AS Graph, IEEE ICNP, 2004. <http://www.ieee-icnp.org/2004/papers/8-1.pdf>, [Προσπελάστηκε 26/6/16]
- [29] L. Li, D. Alderson, W. Willinger, and J. Doyle, A First-Principles Approach to Understanding the Internet Router-Level Topology, ACM SIGCOMM, 2004. <http://dl.acm.org/citation.cfm?id=1015470>, [Προσπελάστηκε 29/6/16]
- [30] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. Claffy, and A. Vahdat, The Internet AS-Level Topology: Three Data Sources and One Definitive Metric, ACM SIGCOMM Computer Communications Review (CCR), v. 36, pp. 17-26, January 2006. http://www.caida.org/publications/papers/2006/as_topology/as_topology.pdf, [Προσπελάστηκε 3/8/16]
- [31] M. Luckie, Y. Hyun, and B. Huffaker, Traceroute Probe Method and Forward IP Path Inference, in Internet Measurement Conference (IMC), Oct 2008, pp. 311—324, http://www.caida.org/publications/papers/2008/traceroute_probe_method/traceroute_probe_method.pdf, [Προσπελάστηκε 17/8/16]
- [32] G. Siganos, and M. Faloutsos, Analyzing BGP Policies: Methodology and Tool, IEEE INFOCOM, 2004. <http://www.cs.ucr.edu/~siganos/papers/Nemecis.pdf>, [Προσπελάστηκε 25/8/16]
- [33] K. Keys, Y. Hyun, M. Luckie, and K. Claffy, Internet-Scale IPv4 Alias Resolution with MIDAR: System Architecture, CAIDA Technical Report, 2011., <http://www.caida.org/publications/papers/2011/midar-tr/>, [Προσπελάστηκε 30/7/16]
- [34] K. Keys, Internet-Scale IP Alias Resolution Techniques, ACM SIGCOMM CCR, 40, pp. 50-55, 2010., http://www.caida.org/publications/papers/2010/alias_resolution/alias_resolution.pdf, [Προσπελάστηκε 12/8/16]
- [35] B. Huffaker, A. Dhamdhere, M. Fomenkov, and K. Claffy, Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers, Passive and [Active](#) Network Measurement Workshop (PAM), April 2010., http://www.caida.org/publications/papers/2010/as_assignment/as_assignment.pdf, [Προσπελάστηκε 17/8/16]