Δίκτυα Υπολογιστών για Δεδομένα Μεγάλης Κλίμακας

Ανάλυση Δικτυακών Ροών σε Κέντρα Δεδομένων

Μάστορας Ιωάννης

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Report

Με την χρήση του επισυναπτόμενου κώδικα, κάναμε parsing του trace, με τη βοήθεια της βιβλιοθήκης dpkt. Μόλις ολοκληρώθηκε το parsing, εξήχθησαν αρκετές πληροφορίες για τις δικτυακές ροές. Αρχικά, Δημιουργήθηκε ο παρακάτω πίνακας:

Dataframe with the size and duration of each flow, with other information:

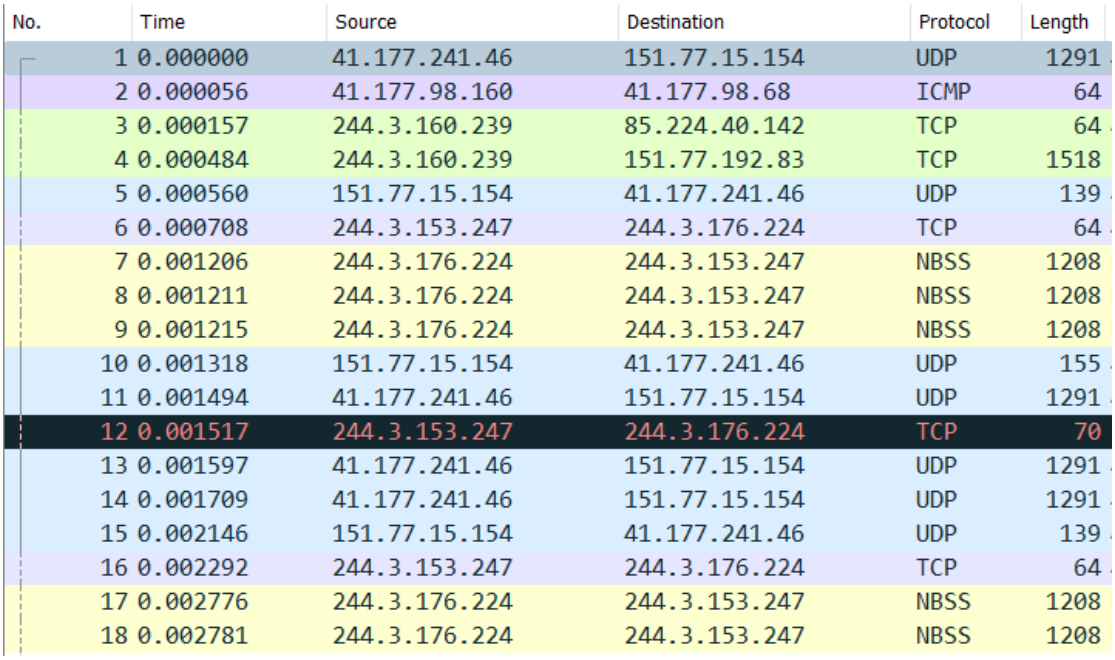
Source IP	Destination IP	Source Port	Source Destination	Protocol	
100.176.176.244	41.177.26.176	49807	80	TCP	1376
100.176.176.56	41.177.26.91	49018	80	TCP	981
105.241.75.197	244.3.31.40	3150	445	TCP	96
106.188.67.194	244.3.31.244	1734	445	TCP	96
106.204.142.139	244.3.31.41	1789	445	TCP	96
...					...
94.2.202.203	244.3.160.239	80	46009	TCP	427
94.2.202.98	244.3.160.239	80	44723	TCP	927
			45341	TCP	879
			45568	TCP	929
95.197.89.124	41.177.98.176	1292	80	TCP	808

Source IP	Destination IP	Source Port	Source Destination	Protocol	Duration
100.176.176.244	41.177.26.176	49807	80	TCP	1.831280
100.176.176.56	41.177.26.91	49018	80	TCP	1.165123
105.241.75.197	244.3.31.40	3150	445	TCP	3.099047
106.188.67.194	244.3.31.244	1734	445	TCP	3.030550
106.204.142.139	244.3.31.41	1789	445	TCP	2.934029
...					...
94.2.202.203	244.3.160.239	80	46009	TCP	0.096525
94.2.202.98	244.3.160.239	80	44723	TCP	0.099653
			45341	TCP	0.092440
			45568	TCP	0.121735
95.197.89.124	41.177.98.176	1292	80	TCP	0.571020

[50426 rows x 2 columns]

Στον πίνακα αυτό αναφέρονται για κάθε ροή του trace πληροφορίες όπως η διεύθυνση IP του αποστολέα και του παραλήπτη, το πρωτόκολλο το οποίο χρησιμοποιήθηκε για την κάθε ροή κλπ. Στο πάνω μέρος έχουμε στην τελευταία στήλη το μέγεθος της ροής σε bytes και αντίστοιχα στο κάτω μέρος έχουμε τον χρόνο που χρειάστηκε να φτάσει η πληροφορία στον παραλήπτη για την ίδια ροή. Ο πίνακας αυτός μας δίνει μια γενική εικόνα των δικτυακών ροών που έχουμε.

Αντίστοιχη εικόνα είδαμε αρχικά μέσω του λογισμικού Wireshark:



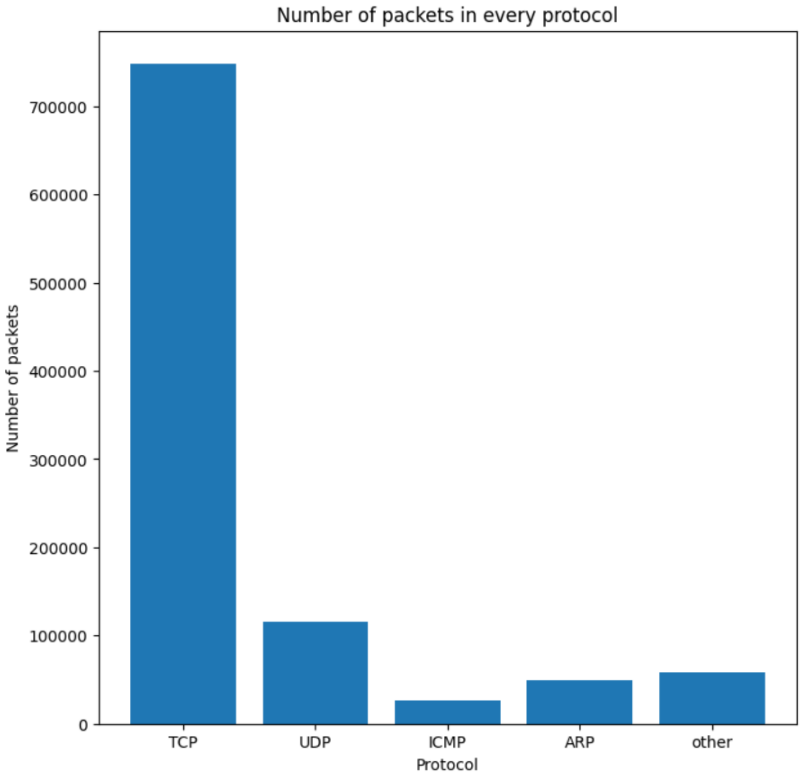
No.	Time	Source	Destination	Protocol	Length
1	0.000000	41.177.241.46	151.77.15.154	UDP	1291
2	0.000056	41.177.98.160	41.177.98.68	ICMP	64
3	0.000157	244.3.160.239	85.224.40.142	TCP	64
4	0.000484	244.3.160.239	151.77.192.83	TCP	1518
5	0.000560	151.77.15.154	41.177.241.46	UDP	139
6	0.000708	244.3.153.247	244.3.176.224	TCP	64
7	0.001206	244.3.176.224	244.3.153.247	NBSS	1208
8	0.001211	244.3.176.224	244.3.153.247	NBSS	1208
9	0.001215	244.3.176.224	244.3.153.247	NBSS	1208
10	0.001318	151.77.15.154	41.177.241.46	UDP	155
11	0.001494	41.177.241.46	151.77.15.154	UDP	1291
12	0.001517	244.3.153.247	244.3.176.224	TCP	70
13	0.001597	41.177.241.46	151.77.15.154	UDP	1291
14	0.001709	41.177.241.46	151.77.15.154	UDP	1291
15	0.002146	151.77.15.154	41.177.241.46	UDP	139
16	0.002292	244.3.153.247	244.3.176.224	TCP	64
17	0.002776	244.3.176.224	244.3.153.247	NBSS	1208
18	0.002781	244.3.176.224	244.3.153.247	NBSS	1208

Αφού αποκτήσαμε μια γενική εικόνα για το trace που έχουμε, προχωρήσαμε σε μια πιο αναλυτική περιγραφή των ροών αυτών. Αρχικά, μετρήσαμε το πλήθος των πακέτων που στάλθηκαν με την χρήση βασικών πρωτοκόλλων, όπως το TCP, το UDP, το ICMP και το ARP, ενώ μαζέψαμε συγκεντρωτικά σε μια κατηγορία όλα τα υπόλοιπα πρωτόκολλα. Στον παρακάτω πίνακα έχουμε το πλήθος και το ποσοστό του όγκου των πακέτων της κάθε κατηγορίας:

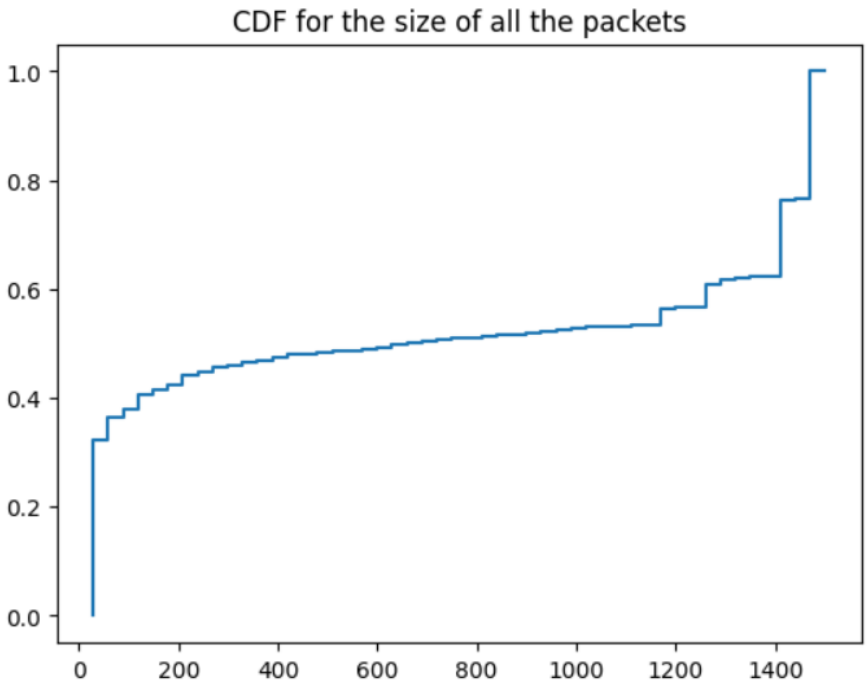
Πρωτόκολλο	**Πλήθος πακέτων**	**Ποσοστό %**
TCP	747990	74.95%
UDP	116012	11.63%
ICMP	26130	2.62%
ARP	49368	4.95%
Other	58369	5.85%
Συνολικά:	997869	100%

Είναι ξεκάθαρο ότι η συντριπτική πλειοψηφία των πακέτων ήταν TCP πρωτοκόλλου, με το UDP να ακολουθεί δεύτερο.

Αυτή η πληροφορία αποτυπώνεται και στο παρακάτω ραβδόγραμμα:

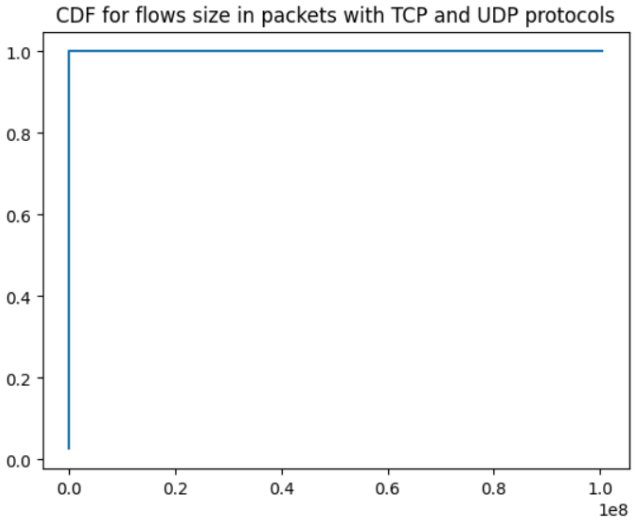


Συνεχίζουμε με κάποια διαγράμματα κατανομής. Αρχικά, αφού μελετήσαμε το πλήθος των πακέτων, ας ρίξουμε μια ματιά και στο μέγεθός τους, το οποίο από τους προηγούμενους πίνακες περιμένουμε να είναι μικρό. Έχουμε:

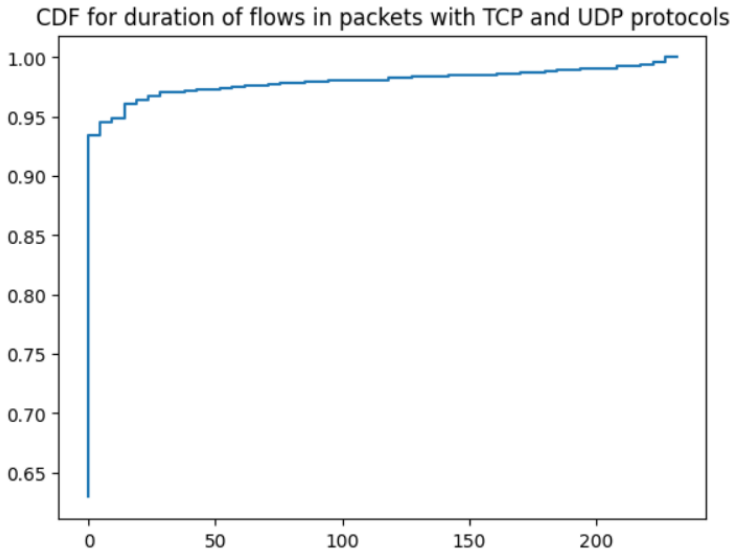


Το παραπάνω διάγραμμα μας επιβεβαιώνει, αφού βλέπουμε ότι τα περισσότερα πακέτα είναι πολύ μικρά σε μέγεθος (το 80% των πακέτων φαίνεται να είναι λιγότερο από 1400 bytes). Συνεπώς έχουμε μια δικτυακή ροή πακέτων μικρού μεγέθους.

Στη συνέχεια, εφόσον περίπου το 85% των πακέτων χρησιμοποιούσαν TCP και UDP πρωτόκολλα, θα εστιάσουμε σε αυτά. Για τα πρωτόκολλα αυτά, δημιουργήσαμε διαγράμματα κατανομής του μεγέθους της κάθε ροής που γίνεται με τα πρωτόκολλα αυτά καθώς και της χρονικής διάρκειας διέλευσής της. Έχουμε:



Το παραπάνω διάγραμμα έχει να κάνει με το μέγεθος των ροών των πρωτοκόλλων αυτών. Βλέπουμε ότι το 100% αυτών είναι πάρα πολύ μικρό. Το διάγραμμα αυτό δεν είναι ιδιαίτερα χρήσιμο, καθώς μόνο αυτήν η πληροφορία μπορεί να εξαχθεί. Συνεπώς, καταλαβαίνουμε ότι τα μεγαλύτερα σε μέγεθος πακέτα στο δίκτυο προέρχονται από άλλα πρωτόκολλα και όχι από το TCP και το UDP. Ως αναφορά τη διάρκεια κίνησης, έχουμε:



Περίπου το 95% των ροών έχουν διάρκεια γύρω στα 10 δευτερόλεπτα, γεγονός που πιθανόν να οφείλεται και στο μικρό μέγεθος τους, ενώ βλέπουμε ότι πολύ μικρό ποσοστό (κοντά στο 3%) έχουν διάρκεια πάνω από 50 δευτερόλεπτα.

Οδηγίες για την εκτέλεση του κώδικα

Για την εκτέλεση του κώδικα, αποσυμπιέστηκε η tar.gz μορφή του αρχείου που δόθηκε, για να έχουμε το trace.pcap αρχείο που θέλουμε. Επίσης, πριν εκτελεστεί ο κώδικας στο command prompt, θα πρέπει να εγκατασταθεί η βιβλιοθήκη dpkt μέσω της εντολής:

pip install dpkt

Προς αποφυγή κάποιου λάθους, θα επισυνάψω εκτός του .py αρχείου με τον κώδικα, τον ίδιο κώδικα σε μορφή .ipynb notebook, όπου θα είναι ήδη εκτελεσμένος ο κώδικας με τα αποτελέσματά του.