Ανάλυση Δικτυακών Ροών σε Κέντρα Δεδομένων

Στόχοι

Στόχος της εργασίας είναι η ανάλυση δικτυακής κίνησης σε κέντρα δεδομένων με βάση trace σε μορφή PCAP, η εξαγωγή χαρακτηριστικών της κίνησης με τη μορφή κατανομών, καθώς και η σύγκριση αυτών των χαρακτηριστικών της κίνησης με δικτυακή κίνηση που παρατηρείται στο Διαδίκτυο. Η εργασία είναι ατομική.

Παραδοτέα

- Κώδικας για την ανάγνωση (parsing) του trace με σύντομη περιγραφή του (documentation)
- Κείμενο με τις γραφικές παραστάσεις των κατανομών, την ανάλυση των αποτελεσμάτων και σύγκριση των χαρακτηριστικών της κίνησης στο κέντρο δεδομένων με αυτά της κίνησης στο Διαδίκτυο
- Οδηγίες για την εκτέλεση του κώδικα και για την εγκατάσταση των απαιτούμενων βιβλιοθηκών

Ημερομηνίες

Καταληκτική ημερομηνία παράδοσης: **11/4/2021**

Περιγραφή Εργασίας

Να αναλύσετε τη δικτυακή κίνηση με βάση τα καταγεγραμμένα πακέτα στο παρακάτω trace http://pages.cs.wisc.edu/~tbenson/IMC_DATA/univ1_trace.tgz
Η καταγραφή της κίνησης έχει πραγματοποιηθεί με βάση το γνωστό πρότυπο PCAP.

Ανάγνωση trace. Να γραφτεί κώδικας σε Python για την εξαγωγή χαρακτηριστικών κίνησης από το trace με τη μορφή κατανομών. Θα πρέπει να εξαχθούν τουλάχιστον τα παρακάτω χαρακτηριστικά σε επίπεδο ροής (flow):

- Μέγεθος ροής (bytes)
- Διάρκεια ροής (sec)
- Αριθμός ροών (δείτε τη σχετική μεθοδολογία που χρησιμοποιήθηκε στο [3])

Η ομαδοποίηση πακέτων σε ροές θα πρέπει να γίνει με βάση τα παρακάτω πεδία επικεφαλίδας:

- Διεύθυνση IP αποστολέα
- Διεύθυνση IP παραλήπτη
- Αριθμός θύρας αποστολέα
- Αριθμός θύρας παραλήπτη

- Πρωτόκολλο (TCP ή UDP)

Επίσης θα πρέπει να εξαχθεί και η κατανομή του μεγέθους πακέτου συνολικά από όλες τις ροές. Επιπλέον, θα πρέπει να πραγματοποιηθεί και κατηγοριοποίηση της κίνησης σε (α) TCP, (β) UDP, (γ) ICMP και (δ) ARP και να υπολογιστούν τα αντίστοιχα ποσοστά της κάθε κατηγορίας σε σχέση με το συνολικό όγκο της κίνησης.

Για την ανάγνωση του trace συστήνεται η χρήση βιβλιοθήκης για PCAP, όπως το DPKT [1] (δείτε το [2] ως ένα ενδεικτικό παράδειγμα κώδικα για το χειρισμό trace σε PCAP με το DPKT). Για λόγους εξοικείωσης με τη μορφή των καταγεγραμμένων πακέτων πριν τη συγγραφή του κώδικα, τα επιμέρους αρχεία του trace μπορούν να προβληθούν μέσω της εφαρμογής Wireshark [4].

Γραφικές παραστάσεις. Θα πρέπει να γίνει γραφική παράσταση των κατανομών (π.χ. Cumulative Distribution Function – CDF) όλων των χαρακτηριστικών της κίνησης που έχουν εξαχθεί. Αντίστοιχα παραδείγματα γραφικής παραστάσεων υπάρχουν στην εργασία [3], από όπου προέρχεται και το trace. Προσοχή στο γεγονός ότι η καταγεγραμμένη κίνηση στο trace είναι υποσύνολο της κίνησης που χρησιμοποιήθηκε ως βάση σε αυτή την εργασία, οπότε ενδέχεται να προκύψει διαφοροποίηση στα προσδιοριζόμενα χαρακτηριστικά μεταξύ του trace και της εργασίας.

Περιγραφή αποτελεσμάτων. Στο κείμενο της εργασίας, πέρα από τις γραφικές παραστάσεις των κατανομών, θα πρέπει να συμπεριληφθεί και ο σχολιασμός των αποτελεσμάτων της μελέτης και να αναδειχτούν συγκεκριμένες τάσεις (π.χ. κυρίαρχα μεγέθη ροών). Επίσης θα πρέπει να πραγματοποιηθεί σύγκριση των αποτελεσμάτων με χαρακτηριστικά της δικτυακής κίνησης, που παρατηρείται ευρύτερα στο Διαδίκτυο. Σχετικά με αυτό, θα απαιτηθεί η κατάλληλη βιβλιογραφική μελέτη για τη συγκέντρωση έγκυρων βιβλιογραφικών πηγών, που θα πρέπει να παρατεθούν στο τέλος του κειμένου της εργασίας.

Αναφορές

[1] DPKT, https://pypi.org/project/dpkt/

[2] https://dpkt.readthedocs.io/en/latest/_modules/examples/print_packets.html

[3] T. Benson, et al., Network Traffic Characteristics of Data Centers in the Wild , ACM IMC 2010

[4] Wireshark, https://www.wireshark.org/