Ανίχνευση κακόβουλης κίνησης DNS over HTTPS

Το DNS αποτελεί ένα από τα σημαντικότερα πρωτόκολλα για την ομαλή λειτουργία του σύγχρονου Διαδικτύου. Ωστόσο, ο σχεδιασμός του DNS επιτρέπει στους Internet Service Providers (ISP's) να συλλέγουν δεδομένα για τις ιστοσελίδες που επισκέπτονται οι πελάτες τους και να τις χρησιμοποιούν για να εξυπηρετούν συμφέροντα, π.χ. διαφημιστικούς σκοπούς.

Για την απόκρυψη των παραπάνω πληροφοριών από τους ISP's έχουν προταθεί πολυάριθμες επεκτάσεις στο πρωτόκολλο DNS. Συγκεκριμένα, το DNS over HTTPS (DoH) αποτελεί την πιο δημοφιλή επέκταση. Βασισμένο σε μεθόδους κρυπτογραφίας, το DoH προστατεύει την ιδιωτικότητα των χρηστών, κρύβοντας από τους ISP's τα ονόματα DNS για τα οποία πραγματοποιείται αναζήτηση.

Παρά τα πλεονεκτήματα που προσφέρει το DoH, οι ιδιότητές του το έχουν καταστήσει κατάλληλο για την εξυπηρέτηση κακόβουλων σκοπών. Συγκεκριμένα, το DoH παρέχει τη δυνατότητα σε κακόβουλους χρήστες του Διαδικτύου να αποκρύψουν τις δραστηριότητές τους, όπως είναι η μεταφορά ευαίσθητων προσωπικών πληροφοριών από μολυσμένους υπολογιστές σε ένα δίκτυο, π.χ. κωδικών που έχουν υποκλαπεί.

Για τον εντοπισμό κακόβουλης κίνησης DoH έχουν προταθεί πολυάριθμες προσεγγίσεις, οι οποίες βασίζονται σε στατιστική ανάλυση ή μεθόδους μηχανικής μάθησης (machine learning), συμπεριλαμβανομένων μεθόδων βαθιάς μηχανικής μάθησης (deep learning) [1, 2]. Παράλληλα, για τη διευκόλυνση πραγματοποίησης έρευνας πάνω στο πρόβλημα, έχει δημιουργηθεί το σύνολο δεδομένων CIRA-CIC-DoHBrw-2020 [3], που χρησιμοποιείται ως κοινό σημείο για την αξιολόγηση των διαφόρων μεθόδων ανίχνευσης κακόβουλης κίνησης DoH που προτείνονται.

Στα πλαίσια της διπλωματικής θα πρέπει να πραγματοποιηθεί:

- Ανασκόπηση των μεθόδων που έχουν χρησιμοποιηθεί στη βιβλιογραφία για το διαχωρισμό καλόβουλης και κακόβουλης κίνησης DoH.
- Αξιολόγηση της καταλληλότητας του CIRA-CIC-DoHBrw-2020. Στην κατεύθυνση αυτή ενδέχεται να χρειαστεί εμπλουτισμός του συνόλου δεδομένων ή, ανάλογα με τα συμπεράσματα της αξιολόγησης, παραγωγή ενός καταλληλότερου συνόλου δεδομένων.
- Αναζήτηση μεθόδων που δεν έχουν χρησιμοποιηθεί για το χαρακτηρισμό κίνησης DoH και αξιολόγησή τους πάνω στο CIRA-CIC-DoHBrw-2020 ή στο σύνολο δεδομένων που θα παραχθεί στα πλαίσια της διπλωματικής.

[1] G. C. Amaizu, D. J. S. Agron, J. M. Lee and D. S. Kim, "Hybrid Deep Neural Network for Malicious DNS Infiltration Detection System", pp. 169-170, 2021
[2] Y. Khodjaeva and N. Zincir-Heywood, "Network flow entropy for identifying malicious behaviours in DNS tunnels", in the 16th International Conference on Availability, Reliability and Security, pp. 1-7, 2021
[3] M. T. Jafar, M. Al-Fawa'reh, Z. Al-Hrahsheh and S. T. Jafar, "Analysis and Investigation of Malicious DNS Queries Using CIRA-CIC-DoHBrw-2020 Dataset", in Manchester Journal of Artificial Intelligence and Applied Sciences, pp. 65-70, 2021