ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Shevtsov Alexander
Μεταπτυχιακός Φοιτητής

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Ε. Μαρκάτος

Δευτέρα, 25/02/2019, 09:00
Αίθουσα Β108,Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

"DomQuery: Ανάλυση ευρείας κλίμακας των επεκτάσεων προγραμμάτων περιήγησης"

ΠΕΡΙΛΗΨΗ

Τα σύγχρονα προγράμματα περιήγησης αποκτούν έξτρα λειτουργικότητα, ώστε να γίνουν πιο ευέλικτα και να προσελκύσουν πολλούς χρήστες. Πράγματι, πολλοί από τους μεγάλους παίκτες όπως το Google Chrome, Mozilla Firefox και το Microsoft Edge παρέχουν πρόσθετες βελτιωμένες δυνατότητες και λύσεις απορρήτου μεσώ της μορφής των επεκτάσεων. Αυτές οι επεκτάσεις είναι διαθέσιμες στην αγορά του προ- γράμματος περιήγησης (ηλεκτρονικό κατάστημα), η οποία προσφέρει χιλιάδες επεκτάσεις στον χρήστη. Μερικές από αυτές είναι πολύ δημοφιλής με εκατομμύρια λήψεις.

Σε αυτή τη μεταπτυχιακή εργασία παρουσιάζουμε το DOMQuery, ένα σύστημα που αναλύει τις αλληλεπιδράσεις μεταξύ των επεκτάσεων του προγράμματος περιήγησης και των στοιχείων DOM των ιστοτόπων. Επιλέξαμε 705,305 διαφορετικές εκδόσεις από 307,822 επεκτάσεις και συλλέξαμε τα κορυφαία ένα εκατομμύριο ιστότοπους της Αλέξα δημιουργώντας ένα ευρετήριο όλων των στοιχείων DOM που βρίσκονται σε αυτούς τους

ιστότοπους. Το σύστημά μας, προσδιορίζει τις ιστοσελίδες και τα συγκεκριμένα στοιχεία DOM που χειρίζονται οι επεκτάσεις, προκειμένου να εντοπίσουν περιπτώσεις κακής χρήσης επέκτασης. Επιπλέον, το σύστημά μας αναλύει τα δικαιώματα επεκτάσεών και τα JavaScripts που χρησιμοποιούνται από αυτά για την ομαδοποίηση επεκτάσεων με βάση τη λειτουργικότητά τους. Χρησιμοποιώντας μια τέτοια προσέγγιση, μπορούμε να εντοπίσουμε γρήγορα επεκτάσεις που εκτελούν ύποπτη δραστηριότητα.

Η ανάλυση χιλιάδων επεκτάσεων είναι ένα δύσκολο και (εκτελεστικά) χρονοβόρο έργο και απαιτεί την αντιμετώπιση αρκετών προκλήσεων. Αντιμετωπίζουμε αυτά τα ζητήματα με τη χρήση των Kubernetes και τρέχοντας πολλαπλά Docker παράλληλα. Η ανάλυση 6.4 δισεκατομμυρίων γραμμών HTML και 85 εκατομμυρίων γραμμών κώδικα JavaScript οδήγησε στην αναγνώριση επεκτάσεων που στοχεύουν συγκεκριμένους ιστότοπους. Επιπλέον, αναλύοντας τα δικαιώματα των επεκτάσεών βρήκαμε 8,340 περιπτώσεις με λάθος χρήση των δικαιωμάτων. Στην προσπάθειά μας να ερευνήσουμε περισσότερο και να ρίξουμε φως στο φαινόμενο αυτό, θα δημοσιεύσουμε το σύνολο δεδομένων μας.

Shevtsov Alexander
M.Sc. Thesis

Computer Science Department
University of Crete
Master's Thesis Supervisor: Professor, E. Markatos

Monday 25/02/2019, 09:00
Room B108, Computer Science Dept., University of Crete

“DOMQuery: A large-scale Analysis of Browser Extensions”

ABSTRACT

Modern browsers increased and extended their functionality in order to become more flexible and attract a plethora of users. Indeed many of the big players such as Google Chrome, Mozilla Firefox and Microsoft Edge provide additional enhanced features and privacy solutions through the form of browser extensions. These extensions are available

in the browser's market (an online store), which offers hundreds of thousands of extensions to the user; some of them being very popular with millions of downloads.

In this master thesis, we present DOMQuery, a system that analyzes the interactions between browser extensions and websites' DOM elements. We selected 705,305 different versions out of 307,822 extensions and crawled the top one million Alexa websites while creating an index of all the DOM elements found in these websites. Our system identifies the webpages and the specific DOM elements that extensions manipulate in order to identify cases of extension misuse. Moreover, our system analyzes the extension's permissions and the JavaScripts used in order to cluster extensions based on their functionality. Using such an approach we can identify quickly extensions that perform a suspicious activity.

Analyzing thousands of extensions is a problematic and (execution) time-consuming task and requires tackling several challenges. We address these issues with the use of Kubernetes and running multiple Docker containers in parallel. Our analysis of 6.4 billions lines of HTML and 85 millions lines of JavaScript code resulted in identifying extensions that target specific websites. Furthermore, by analyzing the permissions from the extension's manifest, we found 8,340 extensions with wrong permission usage. To foster more research and shed more light on this phenomenon, we will publicly release our dataset.