

ΠΡΟΣ

- 1) Όλα τα μέλη ΔΕΠ του Τμήματος Επιστήμης Υπολογιστών
- 2) Τους εκπροσώπους των Μεταπτυχιακών φοιτητών του Τμήματος Επιστήμης Υπολογιστών
- 3) Την Επταμελή Εξεταστική Επιτροπή
- 4) Όλα τα μέλη της Πανεπιστημιακής Κοινότητας

Πρόσκληση σε Δημόσια Παρουσίαση της Διδακτορικής Διατριβής του

κ. Παπαδόπουλου Παναγιώτη

Doctoral Dissertation Defense

Mr. Panagiotis Papadopoulos

Την Παρασκευή, 21 Σεπτεμβρίου 2018 και ώρα 17:00 στην αίθουσα Τηλεδιάσκεψης Κ206 του Τμήματος Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης στο Ηράκλειο, θα γίνει η δημόσια παρουσίαση και υποστήριξη της Διδακτορικής Διατριβής του υποψηφίου διδάκτορα του Τμήματος Επιστήμης Υπολογιστών κ. Παπαδόπουλο Παναγιώτη με θέμα:

“Ανάλυση των επιπτώσεων της ψηφιακής διαφήμισης στην ιδιωτικότητα των χρηστών”

“Analyzing the Impact of Digital Advertising on User Privacy”

ΠΕΡΙΛΗΨΗ

Η ψηφιακή διαφήμιση είναι μια επιχείρηση πολλών δισεκατομμυρίων δολαρίων που έχει την δύναμη να τροφοδοτεί ολόκληρο το δωρεάν διαδίκτυο. Τα τελευταία χρόνια, προχωρά προοδευτικά προς ένα προγραμματιστικό μοντέλο στο οποίο οι διαφημίσεις ταιριάζουν με τα πραγματικά ενδιαφέροντα των ατόμων τα οποία συλλέγονται καθώς αυτοί περιηγούνται στον διαδίκτυο. Ο διαφημιστής πληρώνει ένα χρηματικό κόστος για να αγοράσει διαφημιστικό χώρο στο ψηφιακό μέσο ενός

εκδότη (π.χ. ιστότοπο) παρέχοντας έτσι την ψηφιακή του διαφήμιση μέσα στο περιεχόμενό του όπου καταλήγει στην οθόνη του επισκέπτη.

Σε αντίθεση με τις παραδοσιακές διαφημίσεις σε μέσα όπως οι εφημερίδες, η τηλεόραση ή το ραδιόφωνο, στον ψηφιακό κόσμο, οι τελικοί χρήστες πληρώνουν επίσης ένα κόστος για να λάβουν μια διαφήμιση. Ενώ το κόστος από την πλευρά του διαφημιζόμενου είναι σαφώς νομισματικό, στον τελικό χρήστη, περιλαμβάνει τόσο ποσοτικά, άμεσα προσδιορίσιμα κόστη (όπως HTTP requests και μεταφερόμενα bytes), όσο και ποιοτικά όπως η απώλεια ιδιωτικότητας μέσα στο οικοσύστημα των ψηφιακών διαφημίσεων. Πράγματι, καθώς οι διαφημίσεις γίνονται όλο και πιο εξατομικευμένες ώστε να ταιριάζουν με τα ενδιαφέροντα των χρηστών και να γίνονται όσο το δυνατόν πιο αποτελεσματικές, χρειάζονται περισσότερες προσωπικές πληροφορίες για τους επισκέπτες. Με τον τρόπο αυτό, οι εταιρείες παρακολούθησης αναπτύσσουν εξελιγμένους μηχανισμούς παρακολούθησης χρηστών που ανακτούν οποιαδήποτε πληροφορία που μπορεί να αποκαλύψει τα συμφέροντα και τις προτιμήσεις του χρήστη.

Αυτές οι πληροφορίες μπορεί να περιλαμβάνουν τρέχουσες και ιστορικές γεωγραφικές θέσεις, εγκατεστημένες εφαρμογές, ιστορικό περιήγησης, κ.ο.κ. Όλες αυτές οι πληροφορίες χρησιμοποιούνται για την δημιουργία πλούσιων προφίλ χρηστών που μπορούν να μοιραστούν ή να πουληθούν σε οποιονδήποτε ενδιαφερόμενο (π.χ. διαφημιστές, μεσίτες δεδομένων, πλατφόρμες διαχείρισης δεδομένων κ.λ.π.) πέρα από τον έλεγχο των ίδιων των χρηστών.

Για την διεξαγωγή τέτοιων αγορών δεδομένων και πριν από την πραγματοποίηση συγχώνευσης οποιασδήποτε βάσης δεδομένων, διαφορετικές οντότητες εκτελούν συγχρονισμούς των διαφορετικών αναγνωριστικών χρήστη που έχουν ορίσει για τους ίδιους χρήστες. Με αυτόν τον τρόπο μειώνουν τον αριθμό των διαφορετικών αναγνωριστικών με τα οποία γνωρίζουν έναν χρήστη, αυξάνοντας έτσι την ικανότητα τους να ταυτοποιήσουν ξανά τους χρήστες όταν αυτοί διαγράφουν όλα τα δεδομένα του browser τους (π.χ. τα cookies) ή ακόμα και όταν περιηγούνται στο δίκτυο μέσω VPN για να διατηρήσουν την ανωνυμία τους.

Εκτός από την συνεχή ανάπτυξη της ψηφιακής διαφήμισης και την επίδραση της στην καθημερινότητα μας, ελάχιστα γνωρίζουμε για την ροή πληροφοριών μεταξύ των συμμετεχουσών εταιρειών και τις διασυνδέσεις μεταξύ τους. Με αυτή την διπλωματική εργασία επιδιώκουμε να ενισχύσουμε την διαφάνεια σε αυτό το μεγάλο οικοσύστημα και να διερευνήσουμε την αμφίδρομη επίδραση μεταξύ της ιδιωτικότητας των χρηστών και της ψηφιακής διαφήμισης. Συγκεκριμένα, διερευνάμε τον αντίκτυπο της εξατομικευμένης διαφήμισης στην ιδιωτικότητα και την ανωνυμία των χρηστών, δεδομένων των περίπλοκων εφαρμογών παρακολούθησης χρηστών και τεχνικών συλλογής προσωπικών δεδομένων. Μετράμε πειραματικά τις διαρροές πληροφοριών χρηστών που εμφανίζονται κατά την περιήγηση ιστοτόπων και χρήση εφαρμογών σε κινητά. Με βάση την γνώση που αποκτήθηκε από τα πειράματα, σχεδιάζουμε αντίμετρα για να μετριάσουμε την απώλεια της ιδιωτικής ζωής.

Επιπρόσθετα, μελετάμε πως αυτά τα δεδομένα που συλλέγονται από χρήστες επηρεάζουν την δυναμική τιμολόγησης των προγραμματιστικών δημοπρασιών διαφήμισης και πόσα πληρώνουν οι διαφημιστές για να προσεγγίσουν έναν χρήστη. Στη συνέχεια, συγκρίνουμε το κόστος που επιβάλλεται από την ψηφιακή διαφήμιση τόσο στους χρήστες όσο και στους διαφημιζόμενους για την ίδια διαφήμιση. Αυτά τα κόστη μπορούν είτε να προσδιοριστούν άμεσα (π.χ. θερμοκρασία συστήματος, bytes, ενέργεια) είτε να είναι ποιοτικά, όπως η ιδιωτικότητα. Τέλος, σε μια προσπάθεια να διερευνηθούν εναλλακτικές λύσεις για την διατήρηση των διαδικτυακών εσόδων των παραγωγών περιεχομένου, οι οποίες δεν απαιτούν χρήση προσωπικών δεδομένων, διεξάγουμε μια λεπτομερή ανάλυση της αναδυόμενης τεχνολογίας του web cryptomining αναφορικά με την κερδοφορία της για τους παραγωγούς περιεχομένου και του κόστους της στον χρήστη.

Επιβλέπων: Καθηγητής, Ευάγγελος Μαρκάτος

ABSTRACT

Digital advertising is a multi-billion dollar business that has the power to fuel the entire free Internet. The recent years, it progressively moves towards a programmatic model in which ads are matched to actual interests of individuals collected as they browse the web. The advertiser pays a monetary cost to buy ad-space in a publisher's medium (e.g., website) thus delivering their digital advertisement along with the publisher's interesting content in the visitor's display.

Unlike traditional advertisements in mediums such as newspapers, TV or radio, in the digital world, the end-users are also paying a cost for the advertisement delivery. Whilst the cost on the advertiser's side is clearly monetary, on the end-user, it includes both quantifiable costs, such as network requests and transferred bytes, and qualitative costs such as privacy loss to the ad ecosystem. Indeed, as advertisements become more and more personalized to match the users interests and become as effective as possible, more personal information about the visiting users is needed. Motivated by that, tracking companies deploy sophisticated user-tracking mechanisms retrieving any piece of information can reveal the user's interests and preferences.

Such information may include current and historical geolocations, installed apps, browsing histories, and so forth. All this information is used to form rich user profiles and large audience segments that can be shared with or sold to anyone interested (e.g., advertisers, data brokers, data management platforms, etc.) beyond the control of the users. To conduct such data markets and before performing any background user database merges, different entities perform synchronisations of the different userIDs they have set for the same users. This way they reduce the number of the different "aliases" with which they know a user, increasing this way their capability of re-identifying users when they erase their browser state (i.e., cookies) or even when they browse through VPN to preserve their privacy.

Besides the continuous growth of digital advertising and its impact on our everyday lives, little we know about the flow of information within the participating companies and the interconnections between them. Motivated by that, in this dissertation, we aim to enhance the transparency in this large ecosystem and investigate the bidirectional effect between user privacy and programmatic ad-buying. In particular, we explore the impact of personalized advertising on the users privacy and anonymity given the elaborate deployed user tracking and personal data collecting techniques. We experimentally measure the user information leaks appeared while using websites and mobile apps. Based on the insight gained from these experiments, we design countermeasures to mitigate the privacy loss.

Towards the opposite direction, we study how these collected user data affect the pricing dynamics of programmatic ad-auctions and how much advertisers pay to reach a user. Then, we compare the costs imposed by digital advertising to both users and advertiser for the very same delivered ad traffic. These costs include network overhead, temperature, energy consumption, loss of privacy. Finally, in an attempt to investigate privacy-preserving alternatives for web monetization that can be completely detached from any personal data requirement, we perform a detailed analysis of the profitability and the user-side overheads of the emerging technology of web mining.

Supervisor: Professor, Evangelos Markatos