Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής

Πρόγραμμα Μεταπτυχιακών Σπουδών

«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	**Ιχνηλάτηση στον Παγκόσμιο Ιστό: σκοπός, τρόποι λειτουργίας και αντιμετώπισης** **Web Tracking: objective, mechanisms and defenses**
Ονοματεπώνυμο Φοιτητή	**Δημήτριος Μέρμυγκας**
Πατρώνυμο	**Γεώργιος**
Αριθμός Μητρώου	**ΜΠΣΠ12045**
Επιβλέπων	**Χρήστος Δουληγέρης, Καθηγητής**

Ημερομηνία Παράδοσης **Μάιος 2016**

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)	(υπογραφή)	(υπογραφή)
Χρήστος Δουληγέρης Καθηγητής	Δημήτριος Βέργαδος Επίκουρος Καθηγητής	Παναγιώτης Κοτζανικολάου Επίκουρος Καθηγητής

Περίληψη

Όλο και περισσότερο τα τελευταία χρόνια, οι προσπάθειες των διαδικτυακών υπηρεσιών εστιάζονται στην απόκτηση όσο το δυνατόν περισσότερων πληροφοριών που αφορούν τους χρήστες τους. Οι πληροφορίες αυτές συνήθως έχουν σχέση με το περιεχόμενο των αναζητήσεων των χρηστών στο διαδίκτυο, τα είδη των ιστοσελίδων που επισκέπτονται, τους ανθρώπους με τους οποίους έρχονται σε επαφή και τα προϊόντα που αγοράζουν διαδικτυακά. Αυτή η παρακολούθηση (tracking) των χρηστών, στις περισσότερες περιπτώσεις, πραγματοποιείται από παρόχους υπηρεσιών ή περιεχομένου για καθαρά εμπορικούς σκοπούς. Στην πράξη όμως, έχει αποδειχθεί ότι η παρακολούθηση στον ιστό (Web Tracking) χρησιμοποιείται από διάφορους φορείς και για πολλούς και διάφορους λόγους. Οι λαμβανόμενες πληροφορίες δεν χρησιμοποιούνται πάντα απευθείας από τον εκάστοτε φορέα. Μια πολύ κοινή πρακτική είναι τα συλλεγόμενα δεδομένα να πωλούνται σε τρίτους, όπως ασφαλιστικές εταιρείες ή ηλεκτρονικά καταστήματα. Δεν αποκλείονται βέβαια και οι περιπτώσεις στις οποίες στα δεδομένα αυτά αποκτούν πρόσβαση κυβερνητικές υπηρεσίες ή κλέφτες ταυτότητας. Είναι σαφές επομένως ότι η ιδιωτικότητα αποτελεί την αχίλλειο πτέρνα του ιστού της σημερινής εποχής. Η παρακολούθηση των χρηστών μπορεί να πραγματοποιηθεί με πολλούς τρόπους. Για το σκοπό αυτό έχουν αναπτυχθεί κατά καιρούς διάφοροι μηχανισμοί που ομαδοποιούνται σε κατηγορίες, ανάμεσα στους οποίους αξιοσημείωτοι είναι οι μηχανισμοί παρακολούθησης που βασίζονται αποκλειστικά στην πλοήγηση, στην αποθήκευση, στην μνήμη cache (διαδικτυακή ή λειτουργική), σε αποτυπώματα (fingerprinting) κ.α. Όλοι αυτοί οι μηχανισμοί παρακολούθησης έχουν ως στόχο την αναγνώριση των χρηστών στο διαδίκτυο και τη σύνδεσή τους με τα πραγματικά τους στοιχεία, όπως ονοματεπώνυμο, διεύθυνση κατοικίας, αριθμό τηλεφώνου, διεύθυνση ηλεκτρονικού ταχυδρομείου, κ.α. Η αποκάλυψη όλων αυτών των στοιχείων μπορεί να έχει συνέπειες, κάποιες φορές δυσάρεστες, για τους χρήστες. Επομένως, η πρόκληση που τίθεται είναι το πως γίνεται να αντιμετωπιστεί η παρακολούθηση των χρηστών στο διαδίκτυο. Κατά καιρούς έχουν παρουσιαστεί διάφορες προσεγγίσεις, κάποιες από τις οποίες αποσκοπούν στην αντιμετώπιση συγκεκριμένου μηχανισμού παρακολούθησης, ενώ άλλες είναι πιο γενικής φύσεως. Σκοπός της διατριβής αυτής είναι η έρευνα και η καταγραφή των σύγχρονων μεθόδων Web Tracking, η παρουσίαση των στόχων και των τρόπων λειτουργίας τους αλλά και η ανάλυση των τρόπων αντιμετώπισής τους. Για το λόγο αυτό, η παρούσα διατριβή χωρίζεται σε τρία βασικά μέρη. Στο πρώτο μέρος γίνεται μια γενική παρουσίαση του web tracking, μια ανάλυση των σκοπών των μεθόδων παρακολούθησης μέσα από τις επιπτώσεις που έχει η υλοποίησή τους για τους χρήστες, αλλά και μία σύντομη αναφορά της επίδρασης του web tracking στην ιδιωτικότητα των χρηστών στο διαδίκτυο. Στην συνέχεια, στο δεύτερο μέρος, παρουσιάζεται μια σύνοψη όλων των γνωστών μεθόδων και μηχανισμών παρακολούθησης. Στο τρίτο μέρος, συζητούνται οι πιθανοί τρόποι αντιμετώπισής τους μέσα από τις προσεγγίσεις που έχουν παρουσιαστεί κατά καιρούς για το σκοπό αυτό, καθώς και κάποιες σκέψεις και προτάσεις σχετικά με το μέλλον του web tracking. Η εργασία ολοκληρώνεται με την παρουσίαση των συμπερασμάτων που προέκυψαν από τη μελέτη.

Abstract

In the recent years, the efforts of web services have focused on the acquisition of as more information related to their users as possible. This information is usually related to the content of users' online searches, the kind of websites they visit, the people they come in touch and the products they buy online. The tracking of the users, in most of the cases, is performed by content or service providers for purely commercial reasons. In reality though, it has been proven that web tracking is used by different carriers for various reasons. The acquired information is not always used directly by the carrier itself. A very common practice is the collected data to be sold to third parties, such as insurance companies or e-shops. There are also cases that government services or identity thieves get access to this data. It is therefore obvious that privacy constitutes the Achilles' heel of today's web. Users' web tracking can be performed in many ways. For this reason, many mechanisms, which can be grouped to categories, have been developed, among which remarkable are the ones which are based exclusively on browsing, on storage, on cache memory (web or operational), on fingerprinting and more. All these tracking mechanisms' target is the online recognition of users and the connection with their real identity attributes, such as name, home address, phone number, email address etc. The revelation of all this information may have consequences which are sometimes unpleasant for the users. Thus, the challenge ahead is how users' web tracking can be tackled. From time to time, many different approaches have been presented, some of which are focusing on a specific web tracking mechanism, whereas others are more generic. The main goal of the present thesis is the research and the register of recent web tracking methods, the presentation of their goals and working ways, as well as the analysis of ways to address them. For this reason, this thesis is divided in three basic parts. In the first part there is a generic presentation of web tracking, an analysis of the purposes of web tracking methods through the consequences that their implementation has for the users, as well as a brief mention of the impacts of web tracking on the users' online privacy. In the second part, there is a synopsis of all known web tracking methods and mechanisms. In the third part, the possible ways of defenses are discussed through the approaches that have been presented from time to time for this purpose and also some thoughts and proposals for the future of web tracking. The thesis is completed with the presenting of conclusions drawn from the study.

Περιεχόμενα

Περίληψη 3
Abstract 4
Περιεχόμενα 5
Κατάλογος Εικόνων 7
Εισαγωγή 8
1 Web Tracking 11
1.1 Θεωρητικό υπόβαθρο 11
1.2 Συνοπτική ιστορική αναφορά μεθόδων παρακολούθησης 12
2 Σκοποί & Επιπτώσεις του Web Tracking 14
2.1 Διαδικτυακό marketing και διαφήμιση 14
2.1.1 Συναφής διαφήμιση και σημασιολογική στόχευση 14
2.1.2 Στόχευση συμπεριφοράς 15
2.1.3 Χρέωση της online διαφήμισης 16
2.1.4 Διαδικτυακά συστήματα διαφημίσεων 16
2.2 Καθορισμός τιμών προϊόντων 17
2.3 Προσδιορισμός ασφαλιστικής κάλυψης 18
2.4 Web analytics 20
2.5 Δοκιμές χρηστικότητας 21
2.6 Κυβερνητική παρακολούθηση 22
2.7 Κλοπή ταυτότητας 25
3 Αναγνώριση Εντοπισμένου Χρήστη 26
3.1 Web tracking & ιδιωτικότητα 26
3.2 Το δικαίωμα της προστασίας των προσωπικών δεδομένων 27
3.3 Τρόποι αναγνώρισης εντοπισμένων χρηστών 28
3.3.1 Παρακολούθηση νόμιμων υπηρεσιών ως τρίτοι 28
3.3.2 Διαρροή πληροφοριών προς τρίτους 29
3.3.3 Πώληση πληροφοριών προς τρίτους 30
3.3.4 Απο-ανωνυμοποίηση δεδομένων 30
4 Εξέλιξη των Μηχανισμών Web Tracking 31
4.1 Client-side μηχανισμοί παρακολούθησης 31
4.2 Μηχανισμοί παρακολούθησης βάσει πλοήγησης 32

4.3 Τα cookies ως μηχανισμός παρακολούθησης 33
4.4 Web beacons 37
4.5 HTML5 Web Storage 39
4.6 Ετικέτες οντότητας 40
4.7 Λειτουργικές μνήμες cache 42
4.8 Έγχυση κεφαλίδων HTTP 45
5 Μηχανισμοί Αποτύπωσης 48
5.1 Μηχανισμοί αποτύπωσης προγράμματος περιήγησης / συσκευής 48
5.1.1 Παθητικοί μηχανισμοί 50
5.1.2 Ενεργητικοί μηχανισμοί 52
5.1.3 Η μοναδικότητα του προγράμματος περιήγησης 53
5.1.4 Μηχανισμός αποτύπωσης HTML5 Canvas 56
5.1.5 Ανθεκτικότητα των αποτυπωμάτων του προγράμματος περιήγησης 57
5.2 Μηχανισμοί αποτύπωσης λειτουργικού συστήματος 57
5.3 Μηχανισμοί αποτύπωσης θέσης 58
6 Αντιμετώπιση Μηχανισμών Web Tracking 59
6.1 Καθορισμός των κριτηρίων αξιολόγησης 60
6.2 Ιδιωτικές πρωτοβουλίες 61
6.2.1 Εργαλεία αντιμετώπισης μηχανισμών web tracking 62
6.2.2 Αυτορρύθμιση 65
6.2.3 Πρωτοβουλία Do Not Track 66
6.3 Γενική αξιολόγηση των μέτρων αντιμετώπισης των μηχανισμών web tracking 68
7 Web Tracking: Αναζητώντας Λύση 70
7.1 Τυποποίηση του μηχανισμού ρητής συγκατάθεσης του χρήστη 70
7.2 Ιδιωτικότητα από σχεδιασμό 71
8 Συμπεράσματα 73
Βιβλιογραφία 74

Κατάλογος Εικόνων

Εικόνα 1: Καθορισμός ασφαλιστικής κάλυψης βάσει δεδομένων marketing (Πηγή: Annual medical report)..... 19
Εικόνα 2: Παράδειγμα απεικόνισης πληροφοριών στο Google Analytics (Πηγή: Google Analytics)..... 21
Εικόνα 3: Αιτήματα κατά περίοδο αναφοράς και ποσοστό αιτημάτων για τα οποία δόθηκαν ορισμένα δεδομένα (Πηγή: Google)..... 23
Εικόνα 4: Η αγορά παρακολούθησης και τα θύματά της (Πηγή: BloombergBusiness)..... 24
Εικόνα 5: Αναπαράσταση λειτουργίας των cookies (Πηγή: R. Tirtea και συν.)..... 34
Εικόνα 6: Μηχανισμός προηγμένης παρακολούθησης με χρήση evercookies (Πηγή: Wall Street Journal)..... 36
Εικόνα 7: Σενάριο χρήσης web beacons ως μηχανισμό third-party web tracking (Πηγή: The Free Dictionary)..... 38
Εικόνα 8: Σενάριο λειτουργίας ετικέτας οντότητας (Πηγή: Dweeb's Lair)..... 41
Εικόνα 9: Επικοινωνία προγράμματος περιήγησης και διαδικτυακού διακομιστή μέσω HTTPS (Πηγή: INCIBE)..... 43
Εικόνα 10: Παράδειγμα καταχωρήσεων και παραμετροποίησης του HTTPS στο Chrome (Πηγή: INCIBE)..... 44
Εικόνα 11: Upstream HTTP έγχυση (Πηγή: Leo Le Taro)..... 45
Εικόνα 12: Downstream HTTP έγχυση (Πηγή: Leo Le Taro)..... 46
Εικόνα 13: Τρόπος έγχυσης κεφαλίδων HTTP της Verizon (Πηγή: Jonathan Mayer)..... 47
Εικόνα 14: Συμβολοσειρές user-agent δημοφιλών προγραμμάτων περιήγησης (Πηγή: Christof Ferreira Torres)..... 50
Εικόνα 15: Παραδείγματα σειράς αποστολής κεφαλίδων HTTP από δημοφιλή προγράμματα περιήγησης (Πηγη: Christof Ferreira Torres)..... 51
Εικόνα 16: Τυπικό παράδειγμα κεφαλίδας HTTP αιτήματος (Πηγή: Niklas Schmucker)..... 52
Εικόνα 17: Παράδειγμα ανάλυσης Panopticlick (Πηγή: Panopticlick)..... 54
Εικόνα 18: Παράδειγμα σελίδας αποτελεσμάτων της Panopticlick 2.0 ανάλυσης (Πηγή: Aaron Stuart)..... 55
Εικόνα 19: Παράδειγμα HTML5 canvas fingerprint (Πηγή: INCIBE)..... 56
Εικόνα 20: Απόκτηση ιδιωτικής διεύθυνσης IP με χρήση του WebRTC (σε Chrome, Firefox) (Πηγή: INCIBE)..... 59
Εικόνα 21: Μείωση της online παρακολούθησης από τη χρήση εργαλείων αντιμετώπισης των μηχανισμών web tracking (Πηγή: Jonathan Mayer)..... 64

Εισαγωγή

Είναι ευρέως γνωστό ότι πάροχοι υπηρεσιών, όπως το YouTube, πάροχοι περιεχομένου, όπως το Google, το Facebook ή το Amazon, αλλά και τρίτα μέρη, όπως το DoubleClick, συλλέγουν τεράστιες ποσότητες προσωπικών δεδομένων των χρηστών τους κατά την περιήγησή τους στο διαδίκτυο. Μάλιστα, η συλλογή και η ανάλυση των προσωπικών αυτών πληροφοριών αποτελούν τον επιχειρησιακό πυρήνα των περισσότερων από τις εταιρίες αυτές με σκοπό την χρήση των δεδομένων αυτών για εμπορικούς και διαφημιστικούς λόγους, όπως μπορεί να είναι η στοχευμένη διαφήμιση [1], ο καθορισμός τιμών διάφορων προϊόντων [2], η εκτίμηση της σωματικής και ψυχικής κατάστασης των χρηστών [3], κ.α. Στα προσωπικά στοιχεία των χρηστών ενδέχεται επίσης να αποκτήσουν πρόσβαση κυβερνητικοί οργανισμοί αλλά και κλέφτες ταυτότητας. Πολλά προγράμματα θυγατρικών (affiliate programs), όπως το pay per sale [4], απαιτούν την παρακολούθηση των χρηστών από την ιστοσελίδα όπου τοποθετήθηκε κάποια διαφήμιση ενός προϊόντος μέχρι την ιστοσελίδα όπου γίνεται η πραγματική αγορά του [5].

Οι προσωπικές πληροφορίες μπορεί να δοθούν οικειοθελώς από τον ίδιο το χρήστη, όταν για παράδειγμα συμπληρώνει κάποιο διαδικτυακό έντυπο, ή να συλλεχθούν έμμεσα, εν αγνοία του χρήστη, μέσω των αναζητήσεών του στις μηχανές αναζήτησης, των επισκέψεών του σε ιστοσελίδες ή ακόμα και από τη χρήση των προγραμμάτων JavaScript και Flash που είναι ενσωματωμένα στις ιστοσελίδες [6]. Μεταξύ των συλλεχθέντων δεδομένων, μπορεί να υπάρχουν πληροφορίες τεχνικής φύσης, όπως το πρόγραμμα περιήγησης που χρησιμοποιείται, το λειτουργικό σύστημα, η διεύθυνση IP ή ακόμα και λεπτομέρειες σχετικά το hardware υλικό που διαθέτει ο υπολογιστής του χρήστη, καθώς και πολύ πιο σημαντικές και ευαίσθητες πληροφορίες, όπως η γεωγραφική θέση των χρηστών, οι προτιμήσεις τους ή ακόμη και το ιστορικό των ιστοσελίδων που έχουν επισκεφθεί. Στα δεδομένα που συλλέγονται μπορεί επίσης να περιλαμβάνονται και άλλα στοιχεία για τα οποία ο χρήστης δεν έχει δώσει άδεια πρόσβασης. Κάτι τέτοιο μπορεί να συμβεί, για παράδειγμα, με τις υπηρεσίες web mail, οι οποίες επεξεργάζονται όλα τα email των χρηστών τους, ακόμη και αν αυτά προέρχονται από χρήστες που έχουν δηλώσει ότι δεν επιτρέπουν κάτι ανάλογο.

Ο εντοπισμός και η παρακολούθηση της ψηφιακής ταυτότητας ενός χρήστη είναι ιδιαίτερα σημαντικές διαδικασίες για πολλούς και διάφορους λόγους. Ακόμη πιο χρήσιμη όμως και ίσως και πιο επικερδής θα ήταν η δυνατότητα συνδυασμού της ψηφιακής με την πραγματική του ταυτότητα. Ο συνδυασμός αυτός, βέβαια, σε καμιά περίπτωση δεν μπορεί να είναι επωφελής για τους χρήστες, αλλά αντίθετα, θα μπορούσε ενδεχομένως να τους δημιουργήσει σοβαρά προβλήματα ασφάλειας και παραβίασης της ιδιωτικότητάς τους. Εάν κάποιος χρήστης επιλέξει να αποκαλύψει την πραγματική του ταυτότητα σε κάποια υπηρεσία, τότε διακινδυνεύει άμεσα να χάσει την ανωνυμία του σε άλλες οντότητες με τις οποίες γίνεται κοινή χρήση της βάσης δεδομένων των στοιχείων του [7]. Η χρήση του web tracking βέβαια παρέχει και ωφέλειες για τους χρήστες. Υπάρχουν περιπτώσεις όπου η παρακολούθηση ιστοσελίδων μπορεί να χρησιμοποιηθεί για μη επεμβατικούς λόγους, όπως στις περιπτώσεις των web analytics και των δοκιμών χρηστικότητας, τα οποία είναι σε θέση να βελτιώσουν την εμπειρία περιήγησης των χρηστών.

Κατά καιρούς έχουν αναπτυχθεί διάφοροι μηχανισμοί οι οποίοι χρησιμοποιούνται με σκοπό την παρακολούθηση των χρηστών στο διαδίκτυο και τη συλλογή όλων των παραπάνω δεδομένων [8]. Οι μηχανισμοί αυτοί ομαδοποιούνται σε κατηγορίες, ανάμεσα στις οποίες αξίζει

να αναφερθούν οι μηχανισμοί παρακολούθησης που βασίζονται αποκλειστικά στην πλοήγηση, στην αποθήκευση, στην μνήμη cache (διαδικτυακή ή λειτουργική), σε αποτυπώματα (fingerprinting), κ.α. Κάθε μία από τις κατηγορίες αυτές περιλαμβάνει πολλές και διαφορές τεχνικές.

Οι μηχανισμοί αυτοί έχουν εξελιχθεί με την πάροδο των χρόνων. Αρχικά την εμφάνισή τους έκαναν οι μηχανισμοί που βασίζονται αποκλειστικά στην πλοήγηση (session tracking mechanisms) [9]. Οι τεχνικές που χρησιμοποιούσαν οι μηχανισμοί αυτοί ήταν αρκετά απλές και δεν αποτελούσαν σημαντικές απειλές για τους χρήστες. Η επόμενη γενιά μηχανισμών παρακολούθησης περιλάμβανε μηχανισμούς που βασίζονται στην αποθήκευση (storage-based tracking mechanisms) [10]. Οι μηχανισμοί αυτοί ήταν πολύ πιο προηγμένοι από αυτούς που βασίζονται στην πλοήγηση. Καθένας τους παρουσίασε μεγαλύτερη απειλή για την ιδιωτικότητα των χρηστών. Η επόμενη ομάδα μηχανισμών παρακολούθησης, οι μηχανισμοί που βασίζονται στη μνήμη cache (cache-based tracking mechanisms), βασίστηκε επίσης στην αποθήκευση [6]. Η διαφορά της ομάδας αυτής σε σχέση με την προηγούμενη έγκειται στη χρήση της μνήμης cache (διαδικτυακής ή λειτουργικής) με σκοπό τον εντοπισμό στιγμιότυπων των περιηγητών ιστού και τον καθορισμό των ιστοσελίδων που έχουν επισκεφθεί οι χρήστες. Η τελευταία ομάδα μηχανισμών παρακολούθησης αποτελείται από τους πρόσφατα αναδυόμενους μηχανισμούς αποτυπωμάτων (fingerprinting) [11], που είναι σε θέση να εντοπίζουν τους χρήστες χωρίς να βασίζονται σε κανενός είδους μνήμης cache και ανεξάρτητα από τον ενεργοποιημένο τρόπο λειτουργίας των περιηγητών ιστού. Εκτός από τις ομαδοποιημένες αυτές κατηγορίες, κατά καιρούς έχουν εμφανιστεί και άλλοι μηχανισμοί παρακολούθησης οι οποίοι δεν κατατάσσονται σε κατηγορίες καθώς οι τεχνικές που χρησιμοποιούν ποικίλουν [6]. Οι μηχανισμοί αυτοί λειτουργούν πλήρως εν αγνοία των χρηστών και ενέχουν πολλά ηθικά ζητήματα.

Για την αντιμετώπιση των όποιων προβλημάτων δημιουργούνται από το web tracking, έχουν προταθεί διάφορες μέθοδοι και στρατηγικές οι οποίες έχουν ως στόχο την ανάπτυξη τεχνικών τέτοιων που θα εμποδίζουν τον άμεσο εντοπισμό του χρήστη και θα διαφυλάττουν στο βέλτιστο δυνατό βαθμό τα προσωπικά του δεδομένα [6]. Οι πιο δημοφιλείς από αυτές τις μεθόδους είναι το μπλοκάρισμα υπηρεσιών διαφημίσεων (Ad Block), η απόκρυψη της διεύθυνσης IP μέσω εικονικών ιδιωτικών δικτύων VPN, η χρήση διακομιστή μεσολάβησης (proxy) ή λογισμικών πελάτη TOR (The Onion Router) ή ακόμα και η χρήση λειτουργίας ιδιωτικής περιήγησης. Πολλές από αυτές τις τεχνικές χρησιμοποιούν ήδη υπάρχοντα εργαλεία, ενώ άλλες προσπαθούν να οδηγήσουν το χρήστη στην ανάληψη πρωτοβουλιών και στη λήψη άμεσων από αυτόν δράσεων.

Στόχος της συγκεκριμένης διατριβής είναι να καταστήσει γνωστή την ύπαρξη των μηχανισμών του web tracking αλλά και των συνεπειών που απορρέουν από αυτούς. Οι μηχανισμοί αυτοί έχουν διαφορετικούς στόχους, χρησιμοποιούν διαφορετικές τεχνικές και καταφέρνουν να αποκρύψουν τη λειτουργία τους με διαφορετικούς και άγνωστους για την πλειοψηφία των χρηστών τρόπους. Μάλιστα, κάποιοι από τους μηχανισμούς αυτούς δεν είναι σε καμία περίπτωση εύκολοι στο να εντοπιστούν από τους μέσους χρήστες ενώ και οι προσπάθειες αποφυγής τους είναι πολλές φορές μάταιες. Επιπλέον, πολλοί από αυτούς είναι κακόβουλοι και επικίνδυνοι καθώς έχουν στόχο την εκμαίευση ευαίσθητων προσωπικών δεδομένων και την περαιτέρω εκμετάλλευση τους. Τα τελευταία χρόνια, βέβαια, έχουν γίνει σημαντικές και αξιόλογες προσπάθειες δημιουργίας εργαλείων και υπηρεσιών με στόχο την προστασία των χρηστών.

Σκοπός της διατριβής αυτής είναι η έρευνα και η καταγραφή των σύγχρονων μεθόδων web tracking, η παρουσίαση των στόχων και των τρόπων λειτουργίας τους αλλά και η ανάλυση των τρόπων αντιμετώπισής τους. Για το λόγο αυτό, η παρούσα διατριβή χωρίζεται σε τρία βασικά μέρη. Στο πρώτο μέρος, γίνεται μια γενική παρουσίαση του web tracking (Κεφάλαιο 1), μια ανάλυση των σκοπών του Web Tracking και των επιπτώσεων που έχει η υλοποίησή του για τους χρήστες (Κεφάλαιο 2), καθώς και μία σύντομη αναφορά της επίδρασης του web tracking στην ιδιωτικότητα των χρηστών στο διαδίκτυο (Κεφάλαιο 3). Στην συνέχεια, στο δεύτερο μέρος, παρουσιάζεται η εξέλιξη των γνωστών μεθόδων και μηχανισμών παρακολούθησης (Κεφάλαιο 4) και δίνεται έμφαση τους πρόσφατα αναδυόμενους μηχανισμούς αποτυπωμάτων (fingerprinting) (Κεφάλαιο 5). Στο τρίτο μέρος, συζητούνται οι πιθανοί τρόποι αντιμετώπισής τους μέσα από τις προσεγγίσεις που έχουν παρουσιαστεί κατά καιρούς για το σκοπό αυτό (Κεφάλαιο 6) καθώς και κάποιες σκέψεις και προτάσεις σχετικά με το μέλλον του web tracking (Κεφάλαιο 7). Τέλος, η διατριβή ολοκληρώνεται με τα χρήσιμα και ουσιαστικά συμπεράσματα που προκύπτουν από την μελέτη των σύγχρονων μεθόδων web tracking.

1 Web Tracking

Παρά το γεγονός ότι δεν υπάρχει κάποιος σαφής ορισμός, το web tracking θα μπορούσε να οριστεί ως συλλογή, ανάλυση και εφαρμογή των δεδομένων που αφορούν τη δραστηριότητα του χρήστη, ο οποίος από έναν υπολογιστή ή συσκευή χρησιμοποιεί διάφορες υπηρεσίες της κοινωνίας της πληροφορίας (δηλ. το web), με σκοπό τον συνδυασμό και την ανάλυσή τους για διάφορους σκοπούς [12].

Στο κεφάλαιο αυτό γίνεται μια περιληπτική αναφορά του υπόβαθρου της έννοιας του web tracking με σκοπό την καλύτερη κατανόηση των όσων θα εξεταστούν στα πλαίσια της παρούσας διατριβής.

1.1 Θεωρητικό υπόβαθρο

Το web tracking περιλαμβάνει τη συλλογή και τη μετέπειτα διατήρηση, χρήση ή ανταλλαγή δεδομένων που αφορούν τη συμπεριφορά των χρηστών στο διαδίκτυο. Η συλλογή των πληροφοριών αυτών γίνεται μέσω της χρήσης των cookies, της JavaScript ή οποιουδήποτε είδους αποτυπώματος (fingerprint) των συσκευών που χρησιμοποιούν οι χρήστες για να επισκεφθούν ιστοσελίδες [13]. Οι τεχνολογίες που χρησιμοποιούνται για την υλοποίηση του web tracking επιτρέπουν τη συνεχή ροή σε πραγματικό χρόνο πληροφοριών που αφορούν τους χρήστες, όπως στοιχεία εγγραφής, δραστηριότητες αναζήτησης, δεδομένα συμπεριφοράς, στατιστικά επισκεψιμότητας και δεδομένα μετατροπής, στοιχεία που αντικατοπτρίζουν το πώς ένας χρήστης συμπεριφέρεται στο διαδίκτυο. Τα δεδομένα αυτά μπορεί να χρησιμοποιηθούν για να εξαχθούν συμπεράσματα σχετικά με τα ενδιαφέροντα, τα πολιτικά φρονήματα ή τη σωματική και ψυχολογική κατάσταση των χρηστών. Τα δεδομένα αυτά επίσης μπορεί να υποστούν επεξεργασία με σκοπό την αξιολόγηση, την υπόδειξη ή ακόμα και την επιρροή της κατάστασης ή της συμπεριφοράς των χρηστών. Δεδομένα σχετικά με τη συμπεριφορά των χρηστών μπορεί να οδηγήσουν επίσης και σε επιχειρησιακές αποφάσεις με βάση τα προφίλ των πελατών. Οι καταναλωτικές προθέσεις των χρηστών μπορεί να εξαχθούν από τις ψηφιακές τους ταυτότητες. Η αξία ενός πιθανού πελάτη σχετίζεται με την πιθανότητα να πειστεί για την αγορά ενός προϊόντος.

Οι τεχνολογίες επικοινωνιών που αφορούν τις φορητές συσκευές δίνουν στο web tracking ένα επιπλέον πλεονέκτημα. Μια έξυπνη φορητή συσκευή είναι απίθανο να μοιράζεται μεταξύ ατόμων και, ως εκ τούτου, η σχέση μεταξύ συσκευής και χρήστη καθίσταται ισχυρότερη από ότι, για παράδειγμα, μεταξύ χρήστη και υπολογιστή desktop. Οι φορητές συσκευές περιέχουν μοναδικά αναγνωριστικά, όπως συγκεκριμένο advertising identifier, μοναδικό αναγνωριστικό συσκευής (UDID), διεύθυνση ελέγχου πρόσβασης μέσου (MAC address), διεύθυνση ελέγχου πρόσβασης μέσου Bluetooth (Bluetooth MAC address), διεύθυνση ελέγχου πρόσβασης μέσου επικοινωνίας κοντινού πεδίου (MAC NFC address), αναγνωριστικό κινητής τηλεφωνίας IMSI (International Mobile Subscriber Identity) και αναγνωριστικό IMEI (International Mobile Equipment Identifier) [12]. Όλα αυτά τα αναγνωριστικά των φορητών συσκευών δεν μπορεί να αλλάξουν από τους απλούς χρήστες. Εκτός από μοναδικά αναγνωριστικά, οι έξυπνες φορητές συσκευές ενδέχεται να περιέχουν ένα πλούσιο σύνολο δεδομένων, όπως όνομα χρήστη, κωδικό πρόσβασης, ηλικία, φύλο και βιβλίο διευθύνσεων. Οι έξυπνες φορητές συσκευές μπορεί να παρουσιάζουν ακριβή δεδομένα που αφορούν τη συμπεριφορά ενός χρήστη. Επίσης στις έξυπνες φορητές συσκευές είναι εύκολα προσβάσιμα ακριβή στοιχεία γεωγραφικού εντοπισμού (geolocation) που χρησιμοποιούν τα προγράμματα περιήγησης.

Το web tracking έχει αναπτυχθεί με διάφορους τρόπους. Ψηφιακά ίχνη δεδομένων μπορεί να προκύψουν από ακούσια ή μη επιθυμητή δημοσιοποίηση πληροφοριών, γεγονός που μπορεί να οδηγήσει σε αποκάλυψη προσωπικών δεδομένων. Υπάρχουν πολλοί τρόποι δημιουργίας ενός ψηφιακού ίχνους δεδομένων. Για παράδειγμα, μια εμπορική καμπάνια ψηφιακών διαφημίσεων θα μπορούσε να εκχωρήσει ένα μοναδικό αναγνωριστικό σε κάποιον χρήστη, πρόγραμμα περιήγησης ή συσκευή. Ένας άλλος τρόπος είναι η προσωποποίηση πληροφοριών παραπομπής με την προσθήκη πληροφοριών τμηματοποίησης του κοινού (audience segmentation) κατά την πλοήγηση στον Ιστό, έτσι ώστε άλλες ιστοσελίδες που συμμετέχουν στην εμπορική καμπάνια να είναι σε θέση με τη σειρά τους να παρακολουθούν το χρήστη, το πρόγραμμα περιήγησης ή τη συσκευή. Ένα τρίτο παράδειγμα είναι ο συνδυασμός των μοναδικών αναγνωριστικών με τα δεδομένα που έχουν συλλεχθεί κατά το παρελθόν από επισκέψεις σε ένα συγκεκριμένο ιστότοπο. Ένα τέταρτο παράδειγμα είναι η πραγματοποίηση του web tracking για μια εμπορική καμπάνια μέσω συνδυασμού νέων δεδομένων παρακολούθησης (ενός χρήστη, προγράμματος περιήγησης ή συσκευής δεδομένων) με τα δεδομένα που συλλέχθηκαν κατά το παρελθόν από ένα συγκεκριμένο ιστότοπο ή με δεδομένα που συγκεντρώθηκαν από τρίτους. Ένα τελευταίο παράδειγμα περιλαμβάνει τη χρήση cookies (Cookie Matching Service) που συνδέουν ψηφιακά ίχνη του ίδιου χρήστη, προγράμματος περιήγησης ή συσκευής με τη χρήση διαφόρων ιστότοπων [15].

Το web tracking αποτελείται από διάφορα αυτοματοποιημένα στάδια, ξεκινώντας με τη συλλογή δεδομένων από τον Παγκόσμιο Ιστό, τη διατήρηση τους και τη χρήση τους. Ο συνδυασμός, ο συσχετισμός και η αποσυγκειμενοποίηση (decontextualization) των δεδομένων αυτών, μπορεί να δημιουργήσει πολύ λεπτομερή προγνωστικά προφίλ της συμπεριφοράς των χρηστών [16]. Τα δεδομένα αποθηκεύονται σε βάσεις δεδομένων γραφήματος από διάφορες υπηρεσίες στο διαδίκτυο. Η δομή του γραφήματος επιτρέπει την εμφάνιση των μοντέλων συμπεριφοράς που διαφορετικά θα παρέμεναν απαρατήρητα. Τα δεδομένα που προκύπτουν από το web tracking είτε χρησιμοποιούνται αυτόνομα είτε συνδυάζονται με άλλα δεδομένα που συλλέγονται από διάφορες πηγές. Με τον τρόπο αυτό δημιουργούνται πρότυπα σχετικά με τη συμπεριφορά των χρηστών. Ενώ τα μοναδικά αναγνωριστικά που συνδέονται άμεσα ή έμμεσα με έναν χρήστη ή υπολογιστή παρέχουν λίγες πληροφορίες για τον χρήστη, μια συλλογή από μοναδικά αναγνωριστικά αποκαλύπτει μια διάχυτη άποψη των συνηθειών κάποιου και της συμπεριφοράς περιήγησής του στο διαδίκτυο. Με τον τρόπο αυτό, η συλλογή των μοναδικών αναγνωριστικών μπορεί να χρησιμοποιηθεί για τη δημιουργία ψηφιακών ταυτοτήτων.

1.2 Συνοπτική ιστορική αναφορά μεθόδων παρακολούθησης

Ανατρέχοντας στην ιστορική πορεία του web tracking, ένα ορόσημο που προκύπτει είναι η ανάπτυξη της τεχνολογίας των cookies, σχεδόν 20 χρόνια πριν [17]. Τα HTTP cookies εισήχθησαν το 1994, με αρχικό σκοπό την επίλυση του «μικρού» προβλήματος της αξιόπιστης εφαρμογής του εικονικού καλαθιού αγορών. Λόγω του γεγονότος ότι συνήθως το πρωτόκολλο υπερκειμενικής μεταφοράς (Hypertext Transfer Protocol - HTTP) δεν «θυμάται» τίποτα σχετικά με προηγούμενες συναλλαγές, οι πράκτορες των χρηστών (user agents), μέχρι τότε, δεν ήταν σε θέση να διατηρήσουν ιστορικό συναλλαγών (state information). Η διατήρηση ενός ιστορικού συναλλαγών ήταν απαραίτητη για το εικονικό καλάθι αγορών, καθώς μόνο έτσι υπήρχε τρόπος αποθήκευσης των επιλεγμένων στοιχείων κατά τη διάρκεια των αγορών. Εκείνη την εποχή, τα cookies ενεργοποιούνταν από προεπιλογή στις ρυθμίσεις του προγράμματος περιήγησης και ο χρήστης δεν λάμβανε καμιά ειδοποίηση σχετικά με τη χρήση τους [18]. Η μη μεταφορά των

cookies στο συνηθισμένο χρήστη, αποτέλεσε θέμα διαφάνειας και επομένως θέμα προστασίας της ιδιωτικότητας. Για την αντιμετώπιση του κινδύνου παραβίασης της ιδιωτικότητας και της ασφάλειας από πιθανή διαρροή πληροφοριών σε άλλους δικτυακούς τόπους μέσω των cookies, υλοποιήθηκε η πολιτική της ιδίας προέλευσης (Same Origin Policy – SOP) [19]. Με βάση την πολιτική αυτή, τα cookies μπορούσαν να διαβαστούν μόνο από το ίδιο domain που τα δημιουργούσε.

Το 1998, η Διεθνής Ομάδα για την Προστασία των Προσωπικών Δεδομένων στις Τηλεπικοινωνίες (International Working Group on Data Protection in Telecommunications – IWGDPT) έθεσε επί τάπητος διάφορα ζητήματα προστασίας της ιδιωτικότητας που συνδέονταν με τη συστηματική συλλογή ή χρήση των προσωπικών δεδομένων στον Παγκόσμιο Ιστό [20]. Στο έγγραφο εργασίας της, παρουσίασε το P3P (Platform for Privacy Preferences Project), ένα πρωτόκολλο που αναπτύχθηκε από το W3C με σκοπό την παρεμπόδιση cookies τρίτων, εκτός της περίπτωσης στην οποία η ιστοσελίδα που επισκέφθηκε ο χρήστης προσέφερε μια αποδεκτή από αυτόν πολιτική P3P [21]. Ωστόσο, μόνο ένας μεγάλος κατασκευαστής προγράμματος περιήγησης εφάρμοσε το πρότυπο, με αποτέλεσμα, το P3P να μην έχει υιοθετηθεί ευρέως στο διαδίκτυο.

Τα cookies τρίτων έχουν γίνει η ψυχή της βιομηχανίας της ψηφιακής διαφήμισης. Το 2008, στελέχη μάρκετινγκ εταιρειών που χρησιμοποιούν web tracking συζήτησαν για το μέλλον των analytics και της στατιστικής των ιστότοπων. Το μέλλον, πέντε χρόνια μετά, ήταν η υλοποίηση του οράματος που αφορούσε την ενσωμάτωση των παραδοσιακών στατιστικών επισκεψιμότητας των ιστότοπων (First & Third Party Analytics) στην ανάλυση δεδομένων από άλλες υπηρεσίες στο διαδίκτυο, όπως βίντεο, widgets, κοινωνική δικτύωση, παιχνίδια και μηχανές αναζήτησης (Web Analytics) [22].

Η εξέλιξη των μεθόδων παρακολούθησης θα περιγραφεί αναλυτικότερα στο δεύτερο μέρος της παρούσας διατριβής.

2 Σκοποί & Επιπτώσεις του Web Tracking

Η παρακολούθηση είναι μία διαδικασία που συνήθως χρησιμοποιείται από φορείς οι οποίοι στοχεύουν στην επίτευξη ενός ευρέως φάσματος στόχων που έχουν θέσει από την υλοποίησή της. Όσον αφορά τον χρήστη, οι στόχοι αυτοί έχουν, τις περισσότερες φορές, επιπτώσεις στην ιδιωτικότητά του. Υπάρχουν όμως και περιπτώσεις όπου η παρακολούθηση ιστοσελίδων μπορεί να χρησιμοποιηθεί για μη επεμβατικούς σκοπούς, όπως τα Web Analytics και οι δοκιμές χρηστικότητας, τα οποία έχουν δημιουργηθεί με σκοπό τη βελτίωση της εμπειρίας του χρήστη στην περιήγηση.

Οι λαμβανόμενες πληροφορίες δεν χρησιμοποιούνται πάντα απευθείας από τον ίδιο τον φορέα παρακολούθησης. Μια αρκετά κοινή πρακτική είναι η πώληση των δεδομένων αυτών σε τρίτους (π.χ. ασφαλιστικές εταιρείες ή ηλεκτρονικά καταστήματα). Στα δεδομένα αυτά μπορεί να αποκτήσουν επίσης πρόσβαση κυβερνητικές υπηρεσίες αλλά και κλέφτες ταυτότητας.

Στο κεφάλαιο αυτό αναλύεται ο σκοπός χρησιμοποίησης του Web Tracking καθώς και οι πιθανές επιπτώσεις του στους χρήστες. Η ανάλυση αυτή θα γίνει μέσα από τις περιπτώσεις των πιο συχνών εφαρμογών του, όπως το διαδικτυακό marketing και η διαφήμιση, ο καθορισμός τιμών προϊόντων, ο προσδιορισμός ασφαλιστικής κάλυψης, τα Web Analytics, οι δοκιμές χρηστικότητας, η κυβερνητική παρακολούθηση και η κλοπή ταυτότητας.

2.1 Διαδικτυακό marketing και διαφήμιση

Η βασική φιλοσοφία πάνω στην οποία αναπτύχθηκε αρχικά το web tracking ήταν η διευκόλυνση του τομέα του marketing και η αύξηση των κερδών από τις πωλήσεις μέσω της διαδικτυακής διαφήμισης. Με την πάροδο του χρόνου και την χρήση πιο προηγμένων τεχνικών επικεντρώθηκε σε αντικείμενα όπως η παρακολούθηση της συμπεριφοράς των χρηστών-καταναλωτών, η τμηματοποίηση του κοινού και η στόχευση. Σήμερα το web tracking χρησιμοποιείται από την πλειονότητα των διαδικτυακών τόπων, αποκλειστικά για τους σκοπούς αυτούς [5].

2.1.1 Συναφής διαφήμιση και σημασιολογική στόχευση

Η συναφής διαφήμιση ή διαφήμιση περιεχομένου (contextual advertising) αναφέρεται στην παρουσίαση διαφημίσεων που σχετίζονται με το περιεχόμενο μιας ιστοσελίδας [5]. Συνήθως, λέξεις-κλειδιά που αντιπροσωπεύουν το συνολικό θέμα της σελίδας εξάγονται από το πρωτεύον κείμενο. Στη συνέχεια, επιλέγονται διαφημίσεις σύμφωνα με αυτές τις λέξεις-κλειδιά και με τον τρόπο αυτό έχουν άμεση συσχέτιση με το περιεχόμενο. Η συνολική διαδικασία επιτυγχάνεται μέσω αλγορίθμων οι οποίοι συνδυάζουν το περιεχόμενο της ιστοσελίδας με μια σχετική διαφήμιση [23]. Για παράδειγμα, σε ένα χρήστη που επισκέπτεται ιστοσελίδες για αυτοκίνητα, θα του παρουσιάζονται διαφημίσεις σχετικές με αυτά. Η σχετικότητα αυτή της διαφήμισης αυξάνει την πιθανότητα της διαφημιστικής αλίευσης των πιθανολογούμενων αγοραστικών ενδιαφερόντων του χρήστη.

Η σημασιολογική στόχευση (semantic targeting) είναι μια τεχνική που επιτρέπει την παροχή στοχευμένων διαφημίσεων σε ιστοσελίδες και χρησιμοποιείται από δημιουργούς online περιεχομένου και από διαφημιστές για να αυξήσει την αποτελεσματικότητα της διαφημιστικής τους καμπάνιας. Χρησιμοποιείται στην επιλογή και την ταξινόμηση ενός στοχευμένου κοινού με βάση την σημασία και τις σημασιολογικές σχέσεις μέσα στο περιεχόμενο που δημιουργείται

από τους χρήστες (πχ. αναζητήσεις) ή από τους δημιουργούς των ιστοσελίδων [24]. Η δυνατότητα αυτή επιτρέπει την παροχή των διαφημίσεων από αυτοματοποιημένα συστήματα που βασίζονται στο περιεχόμενο που προβάλλεται στον χρήστη. Η σημασιολογική στόχευση τελειοποιεί την έννοια της συναφούς διαφήμισης, προσπαθώντας να εντοπίσει τη σημασιολογία του εμφανιζόμενου περιεχομένου, συμπεριλαμβανομένης της αποσαφήνισης των λέξεων με πολλαπλές πιθανές σημασίες. Με τον τρόπο αυτό, μειώνεται ιδανικά ο αριθμός των άστοχων διαφημίσεων [5].

Μια συνηθισμένη πρακτική υλοποίησης του web tracking για λόγους συναφούς διαφήμισης είναι μέσω επιλεγμένων λέξεων που προέρχονται από το ηλεκτρονικό ταχυδρομείο των χρηστών. Παράδειγμα τέτοιου τρόπου υλοποίησης αποτελεί το Gmail. Το Gmail χρησιμοποιεί λέξεις, μέσα από τα εξερχόμενα και εισερχόμενα μηνύματα ηλεκτρονικού ταχυδρομείου, τις οποίες χρησιμοποιεί για την εμφάνιση στοχευμένων διαφημίσεων. Τα εισερχόμενα μηνύματα σαρώνονται από την υπηρεσία παρά το γεγονός ότι ο αποστολέας δεν έχει δώσει ρητή άδεια για κάτι τέτοιο. Με τον τρόπο αυτό, το Gmail σαρώνει το περιεχόμενο του συνόλου των εισερχομένων μηνυμάτων ενός χρήστη με αποκλειστικό σκοπό τον προσδιορισμό των ενδιαφερόντων του χρήστη και τη δημιουργία αποκλειστικά στοχευμένων διαφημίσεων βάσει αυτών [25]. Η διαδικασία περιγράφεται ξεκάθαρα στους Όρους Παροχής Υπηρεσιών της Google: *«Τα αυτοματοποιημένα συστήματά μας αναλύουν το περιεχόμενό σας (συμπεριλαμβανομένων των μηνυμάτων ηλεκτρονικού ταχυδρομείου) για να σας προσφέρουν εξατομικευμένες λειτουργίες προϊόντων, όπως προσαρμοσμένα αποτελέσματα αναζήτησης, προσαρμοσμένες διαφημίσεις και εντοπισμό ανεπιθύμητων μηνυμάτων και κακόβουλου λογισμικού. Αυτή η ανάλυση πραγματοποιείται κατά την αποστολή, τη λήψη και την αποθήκευση του περιεχομένου.»* [26].

2.1.2 Στόχευση συμπεριφοράς

Η στόχευση συμπεριφοράς (behavioural targeting) ή διαφήμιση συμπεριφοράς (behavioural advertising), όπως αλλιώς είναι γνωστή, είναι μια μορφή στοχευμένης διαφήμισης που προσπαθεί να μαντέψει το κατάλληλο περιεχόμενο διαφήμισης με βάση τα προφίλ χρηστών που έχουν συλλεχθεί. Τα προφίλ αυτά περιέχουν πληροφορίες όπως το φύλο, την ηλικιακή ομάδα, την τοποθεσία, το εκτιμώμενο εισόδημα και τα ενδιαφέροντα [5].

Η συλλογή αυτών των πληροφοριών επιτρέπει στους διαφημιστές να χρησιμοποιούν τον προϋπολογισμό της διαφημιστικής τους καμπάνιας πιο αποτελεσματικά, καθώς απευθύνονται μόνο στο κοινό που είναι περισσότερο πιθανό να μετατραπεί σε πελάτες. Μελέτες έχουν αποδείξει ότι η στόχευση συμπεριφοράς αυξάνει σημαντικά την αποτελεσματικότητα της online διαφήμισης, καθιστώντας πιο πιθανή την αγορά ενός διαφημιζόμενου προϊόντος [28].

Αυτά τα προφίλ έχουν δημιουργηθεί κυρίως μέσω των αναζητήσεων και του ιστορικού περιήγησης των χρηστών [27]. Περισσότερα στοιχεία αντλούνται από τα προσωπικά δημόσια προφίλ των χρηστών σε κοινωνικά δίκτυα, όπως το Facebook, με αποτέλεσμα τα δημιουργημένα προφίλ να είναι λεπτομερέστατα. Τα περισσότερα κοινωνικά δίκτυα απλοποιούν αυτή τη διαδικασία για τους διαφημιστές, μερικά από αυτά ακόμα και σκόπιμα, αφού περιλαμβάνουν, για παράδειγμα, κώδικα παρακολούθησης που είναι σε θέση να δίνει σε τρίτους πληροφορίες σχετικές με το ποιοι λογαριασμοί χρηστών σε ένα κοινωνικό δίκτυο αντιστοιχούν σε κάποιο από τα υπάρχοντα προφίλ [29].

2.1.3 Χρέωση της online διαφήμισης

Διαφημιζόμενοι και διαφημιστές χρησιμοποιούν διάφορα μοντέλα υπολογισμού τιμολόγησης της online διαφήμισης. Τα πιο συχνά χρησιμοποιούμενα μοντέλα είναι τα εξής [30]:

- Πληρωμή ανά εμφάνιση (Pay per impression): Στο μοντέλο αυτό, η χρέωση των διαφημιζόμενων γίνεται ανάλογα με το πόσο συχνά εμφανίζονται οι διαφημίσεις τους στους επισκέπτες των ιστοσελίδων.
- Πληρωμή ανά κλικ (Pay per click): Στο μοντέλο αυτό, ένα από τα δημοφιλέστερα αυτή τη στιγμή, η χρέωση των διαφημιζόμενων γίνεται μόνο όταν οι χρήστες κάνουν κλικ στις διαφημίσεις που εμφανίζονται στις σελίδες που επισκέπτονται.
- Πληρωμή ανά πώληση (Pay per sale): Στο μοντέλο αυτό, οι διαφημιστές παίρνουν ένα ποσοστό επί των πωλήσεων των προϊόντων όταν αυτές γίνονται μέσω των διαφημίσεων που υπάρχουν σε ιστοσελίδες.

Για την χρέωση της online διαφήμισης, η παρακολούθηση των χρηστών είναι απαραίτητη για να είναι δυνατή η χρήση των προγραμμάτων θυγατρικών (affiliate programs) [31]. Η παρακολούθηση μέσω προγραμμάτων θυγατρικών χρησιμοποιείται από τις επιχειρήσεις που είναι ιδιοκτήτριες δικτύων θυγατρικών εταιρειών, με σκοπό τη διαχείριση τους και την απόκτηση εικόνας σχετικά με την απόδοση της κάθε θυγατρικής καθώς και του δικτύου στο σύνολό του. Η παρακολούθηση αυτή επιτρέπει στον ιδιοκτήτη του δικτύου, καθώς και σε κάθε θυγατρική εταιρεία την απόκτηση γνώσεων σχετικά με το ποια τμήματα του δικτύου έχουν τη μεγαλύτερη επισκεψιμότητα (που μετράται με αριθμό κλικ) και με το αν μια εκστρατεία marketing έχει επιτυχία ή όχι. Τα περισσότερα από τα προγράμματα θυγατρικών στηρίζονται σε διάφορες μετρικές βάσει των οποίων απεικονίζουν την παρακολούθηση που υλοποιούν μέσω γραφικών παραστάσεων. Δημοφιλή δίκτυα θυγατρικών είναι τα Amazon και eBay, τα οποία χρησιμοποιούν προγράμματα θυγατρικών όπως το pay per sale, όπου οι διαφημιστές παίρνουν μερίδιο επί των πωλήσεων όταν η αγορά ενός προϊόντος γίνεται μέσω της διαφήμισής του [5], [32]. Η παρακολούθηση στην περίπτωση αυτή αφορά την παρακολούθηση των χρηστών από την ιστοσελίδα όπου τοποθετήθηκε κάποια διαφήμιση ενός προϊόντος μέχρι την ιστοσελίδα όπου γίνεται η πραγματική αγορά του.

Εάν δεν ληφθούν προληπτικά μέτρα, τα μοντέλα αυτά μπορεί εύκολα να υπονομευθούν. Για παράδειγμα, σε ένα μοντέλο pay per click, μπορεί να γίνουν επανειλημμένα κλικ σε διαφημίσεις ενός ανταγωνιστή από το ίδιο άτομο, μια πρακτική που ονομάζεται click fraud, δηλαδή απάτη στον αριθμό των κλικ που σημειώνονται για κάθε διαφήμιση [33]. Η υπονόμευση αυτή καθιστά επιτακτική την ανάγκη χρήσης μεθόδων ταυτοποίησης του χρήστη, με σκοπό τον εντοπισμό και την αποτροπή τέτοιων ενεργειών.

2.1.4 Διαδικτυακά συστήματα διαφημίσεων

Τα διαδικτυακά συστήματα διαφημίσεων προσφέρουν υπηρεσίες που επιτρέπουν στους δημιουργούς και τους υπεύθυνους των ιστοσελίδων, να πωλούν διαφημιστικό χώρο σε διαφημιζόμενους. Οι βασικές λειτουργίες και απαιτήσεις τέτοιων συστημάτων είναι οι εξής [5]:

- Ομαδοποίηση περιεχομένου διαφήμισης μέσω πληρωμής των διαφημιστών και εξεύρεση εκδοτών πρόθυμων να προβάλλουν αυτό το περιεχόμενο.
- Παροχή υποδομής διακομιστών (server infrastructure) όπου φιλοξενούνται οι διαφημίσεις.

- Διάθεση λύσεων λογισμικού που να επιτρέπουν την εύκολη ενσωμάτωση διαφημιστικού χώρου σε υφιστάμενες ιστοσελίδες.
- Διάθεση λύσεων εμφάνισης των διαφημίσεων ανάλογα με τις προτιμήσεις των διαφημιζόμενων. Για παράδειγμα, μπορεί να μην είναι επιθυμητή η εμφάνιση του λογοτύπου μιας εταιρείας μαζί με ορισμένους τύπους περιεχομένου. Κάτι τέτοιο απαιτεί τη χρήση αλγόριθμων συναφούς διαφήμισης.
- Παρακολούθηση των χρηστών του διαδικτύου και δημιουργία προφίλ ανά χρήστη με σκοπό τη στόχευση συμπεριφοράς.
- Συλλογή στατιστικών στοιχείων και δημιουργία αναφορών με σκοπό την πληροφόρηση των διαφημιζόμενων σχετικά με το είδος των διαφημίσεων που παρουσιάζουν απήχηση στο κοινό. Η δυνατότητα αυτή των διαδικτυακών συστημάτων διαφημίσεων βοηθά στη βελτιστοποίηση των διαφημιστικών εκστρατειών.

Στην πράξη, οι ιστοσελίδες περιέχουν ένα μικρό κομμάτι κώδικα JavaScript το οποίο στη συνέχεια φέρνει και εμφανίζει περιεχόμενο από ένα διακομιστή διαφημίσεων τρίτου μέρους (third party ad server). Οι διαφημιζόμενοι, από την άλλη πλευρά, συνήθως δημιουργούν και παρακολουθούν τις εκστρατείες marketing μέσω διαδικτυακών διεπαφών (web interface).

Παραδείγματα δημοφιλών διαδικτυακών συστημάτων διαφήμισης είναι το DoubleClick (που ανήκει στην Google), το ValueClick και το AdBrite [34].

2.2 Καθορισμός τιμών προϊόντων

Το web tracking μπορεί επίσης να χρησιμοποιηθεί για την τροποποίηση της διαφημιζόμενης τιμής των προϊόντων σύμφωνα με την εκτιμώμενη οικονομική κατάσταση των εν δυνάμει πελατών. Στο [35] αποδεικνύεται ότι η εμφανιζόμενη τιμή διαφέρει με βάση τη γεωγραφική θέση του χρήστη που επισκέπτεται μια ιστοσελίδα έως και 166%, ενώ με βάση την οικονομική του κατάσταση έως και 4 φορές υψηλότερες τιμές. Επίσης αποδείχθηκε ότι η τιμή σε έναν ιστότοπο που διαφημίζει κάποιο προϊόν άλλης ιστοσελίδας η διαφορά μπορεί να φτάσει έως και 50%. Η μελέτη αυτή βασίστηκε στη συλλογή δεδομένων μέσα από 200 διαφορετικές ιστοσελίδες online προμηθευτών και για διάστημα 20 ημερών. Οι συγγραφείς κατέληξαν στο συμπέρασμα ότι το λειτουργικό σύστημα ή το πρόγραμμα περιήγησης που χρησιμοποιείται από τον χρήστη δεν παίζει κανένα ρόλο στον καθορισμό των τιμών.

Μέσω του web tracking εκτός από τον καθορισμό των τιμών των προϊόντων μπορεί να καθοριστούν και τα επιτόκια των πιστωτικών καρτών. Στο [36] αναφέρεται ότι τα επιτόκια των πιστωτικών καρτών που προσφέρονται από διάφορες τράπεζες ποικίλλουν ανάλογα με τα στοιχεία του χρήστη που αναζητά προσφορές πιστωτικών καρτών στο διαδίκτυο. Για παράδειγμα, παρατηρήθηκε ότι το επιτόκιο της κάρτας Chase Sapphire διαφημιζόταν με δύο τιμές, 13,24% και 12,24%. Σε άλλο άρθρο αναφέρεται ότι η Capital One Financial Corporation χρησιμοποιεί web tracking για να αποφασίσει ποιες πιστωτικές κάρτες θα πρέπει να εμφανίζει στους χρήστες που επισκέπτονται στην ιστοσελίδα της για πρώτη φορά. Οι υπολογισμοί αυτοί γίνονται από την εταιρεία [x+1] Inc., η οποία χρησιμοποιεί online τεχνικές παρακολούθησης με σκοπό την απόκτηση και συγκέντρωση πληροφοριών σχετικά με τους χρήστες του διαδικτύου [37]. Οι τεχνικές αυτές αφορούν την πρόσβαση και ανάλυση χιλιάδων δεδομένων για ένα και μόνο χρήστη. Τα πιο πολύτιμα δεδομένα για τους υπολογισμούς αυτούς αφορούν τον

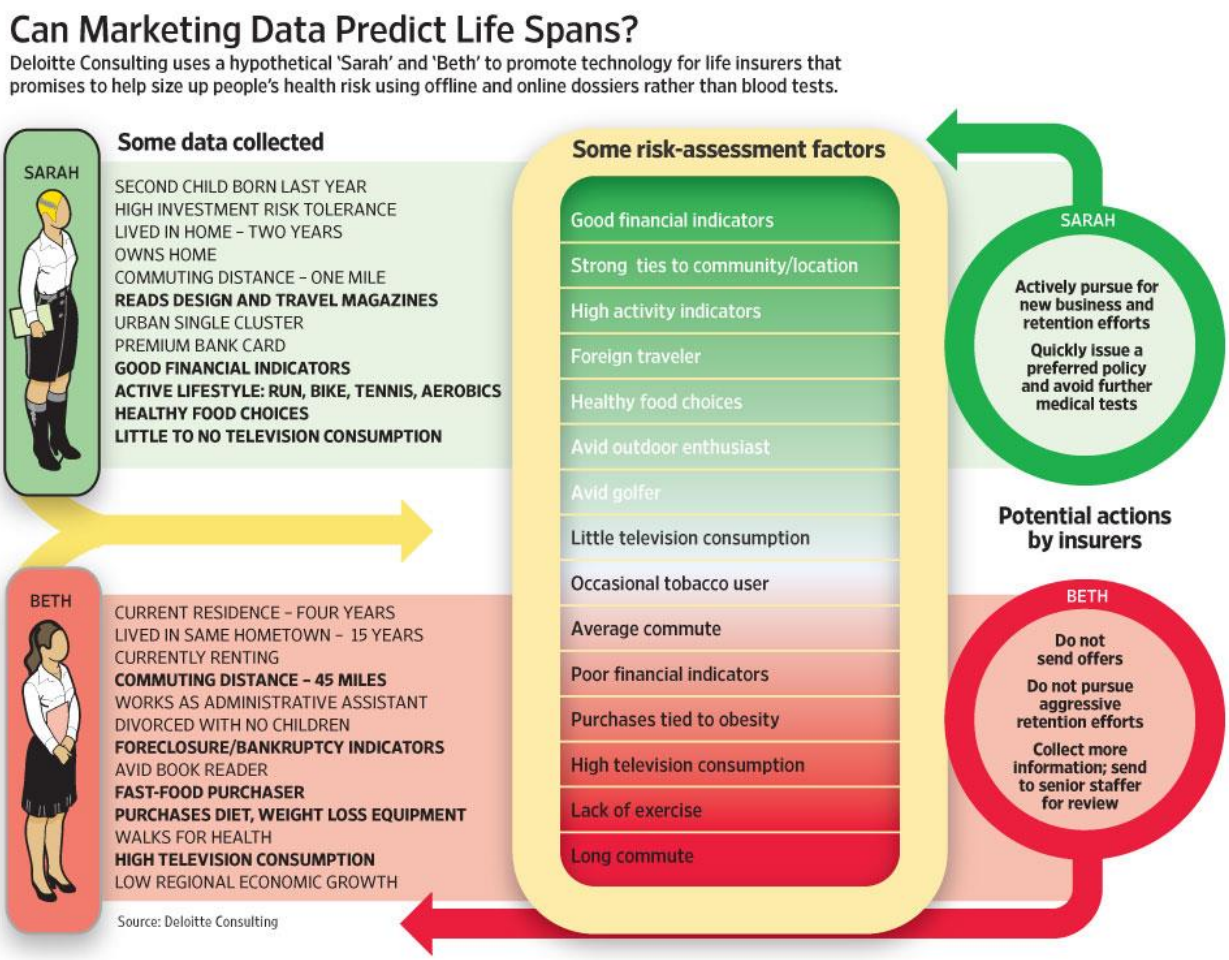
ταχυδρομικό κώδικα και την ημερομηνία γέννησης των χρηστών, καθώς οι άνθρωποι που ζουν στην ίδια περιοχή τείνουν να έχουν παρόμοια εισοδήματα και βιοτικό επίπεδο.

Ο καθορισμός τιμών μέσω web tracking εμφανίζεται και σε άλλες περιπτώσεις. Το 2010 για παράδειγμα, αποκαλύφθηκε ότι η Capital One Financial Corporation διαφοροποιεί τα επιτόκια για τα δάνεια αυτοκινήτων με βάση το πρόγραμμα περιήγησης που χρησιμοποιείται από τον υποψήφιο πελάτη (3,5% για τον Firefox, 2,7% για το Safari, 2,3% για το Chrome και 3,1% για το Opera) [38]. Επίσης, το 2009 οι τιμές ενοικίασης ενός αυτοκινήτου μέσω της ιστοσελίδας Hotwire ήταν διαφορετικές αν η κράτηση γινόταν από υπολογιστή του χώρου εργασίας (χαμηλότερη τιμή) ή του τόπου κατοικίας ενός χρήστη (υψηλότερη τιμή) [39]. Η ώρα κράτησης ήταν και στις δύο περιπτώσεις η ίδια, γεγονός που αποδεικνύει ότι η διάκριση στις τιμές έγινε με βάση την ταυτότητα του χρήστη σε σχέση με το εκάστοτε χρησιμοποιούμενο πρόγραμμα περιήγησης και όχι με βάση την ώρα που γινόταν η κράτηση.

Τέλος, το web tracking φαίνεται να επηρεάζει ακόμα και τιμές δωματίων στα ξενοδοχεία. Η εταιρεία Orbitz Worldwide Inc. ταξινομεί με διαφορετικό τρόπο τις διαφημίσεις ξενοδοχείων, ανάλογα με τον τύπο του υπολογιστή που χρησιμοποιείται από τον πελάτη [40]. Η εταιρεία διαπίστωσε ότι οι χρήστες υπολογιστών Mac έχουν την τάση να ξοδεύουν περίπου 30% περισσότερο για κρατήσεις ξενοδοχείων από ότι οι χρήστες PC. Χρησιμοποιώντας αυτό το στοιχείο, τα πιο ακριβά ξενοδοχεία διαφημίζονται σε χρήστες υπολογιστών Mac, ενώ τα φθηνότερα στους χρήστες PC. Επιπλέον, τα αποτελέσματα αναζήτησης κατατάσσονται σύμφωνα με την εκτιμώμενη υψηλότερη τιμή που ο χρήστης της κάθε κατηγορίας σκοπεύει να πληρώσει για το δωμάτιο.

2.3 Προσδιορισμός ασφαλιστικής κάλυψης

Οι online δραστηριότητες των χρηστών αποκαλύπτουν πολλά στοιχεία για τον τρόπο ζωής, τα ενδιαφέροντα, τις συνήθειες και τα χόμπι τους. Ασφαλιστικές εταιρείες μπορούν να χρησιμοποιήσουν τις εν λόγω πληροφορίες για την αξιολόγηση των κινδύνων που ίσως διατρέχει η ζωή των χρηστών και να καθορίσουν επομένως το είδος της ασφαλιστικής κάλυψης που θα προσφέρουν (Εικόνα 1).



Εικόνα 1: Καθορισμός ασφαλιστικής κάλυψης βάσει δεδομένων marketing (Πηγή: Annual medical report)[1]

Μεγάλες ασφαλιστικές ανέπτυξαν λογισμικό, το οποίο είναι σε θέση, μέσα από διαφορετικά δεδομένα marketing, να αξιολογήσει τα ποσοστά εμφάνισης σε σύντομο χρονικό διάστημα μιας πολύ βαριάς αρρώστιας ή πρόκλησης ατυχήματος σε ένα χρήστη [41]. Τα δεδομένα αυτά προέρχονται από εγγυήσεις προϊόντων, έρευνες καταναλωτών, συνδρομές σε περιοδικά, και δαπάνες πιστωτικών καρτών. Με βάση τον υποτιθέμενο τρόπο ζωής των χρηστών, μερικές ασφαλιστικές εταιρείες τροποποιούν τη συχνότητα των ιατρικών εξετάσεων των πελατών τους.

Άλλος τρόπος απόκτησης δεδομένων από τις ασφαλιστικές εταιρείες είναι η αγορά τους από εταιρείες συλλογής δεδομένων. Η Acxiom Corporation, μία από τις μεγαλύτερες εταιρείες συλλογής δεδομένων, είναι γνωστή για την αγορά δεδομένων από online εκδότες σχετικά με τα είδη των online άρθρων που διαβάζονται από τους συνδρομητές τους [3]. Τα δεδομένα αυτά συσχετίζονται με τις διαθέσιμες για το κοινό πληροφορίες από κοινωνικά δίκτυα και άλλα online προφίλ των χρηστών. Η Acxiom φαίνεται να έχει στην κατοχή της πληροφορίες που αφορούν πάνω από 100 εκατομμύρια νοικοκυριά στις Η.Π.Α.

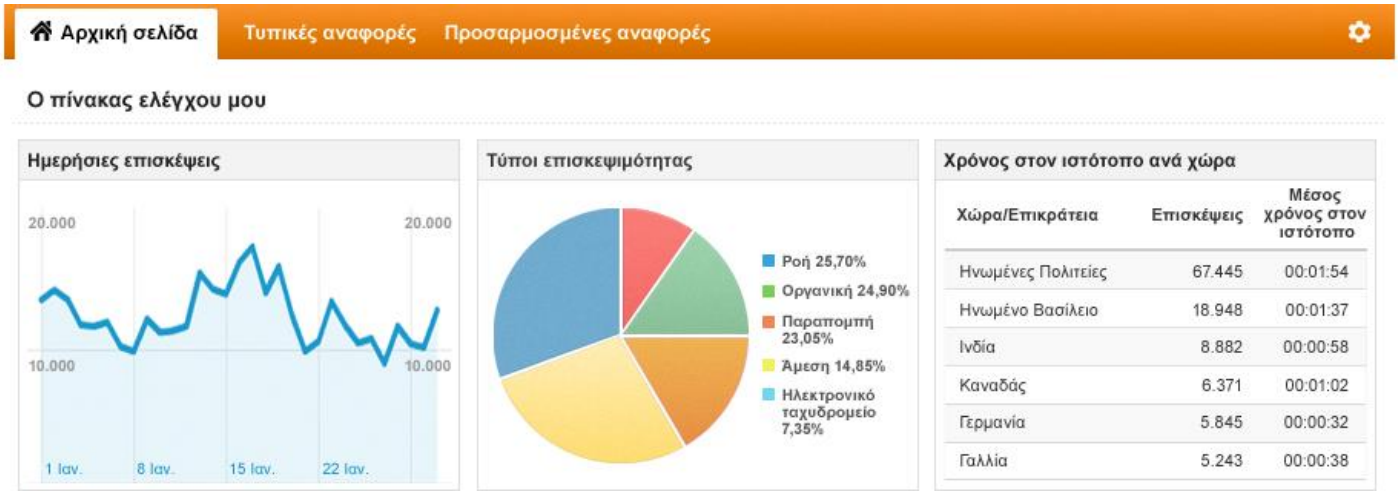
[1] Annual medical report: https://www.annualmedicalreport.com/insurance-companies-mining-online-data-to-predict-medical-health-lifespan-video/

2.4 Web analytics

Ο τομέας των web analytics ασχολείται με τη μέτρηση και την ερμηνεία της χρήσης των δεδομένων μιας ιστοσελίδας [5]. Οι πληροφορίες που δυνητικά ενδιαφέρουν τους διαχειριστές των ιστοσελίδων είναι οι εξής:

- Ο αριθμός των επισκεπτών με την πάροδο του χρόνου, ο οποίος μπορεί περαιτέρω να διαιρεθεί στον αριθμό των χρηστών που επισκέπτονται τη σελίδα τακτικά και στο αριθμό των νέων επισκεπτών. Η πληροφορία αυτή περιλαμβάνει επίσης τον χρόνο παραμονής των επισκεπτών στην ιστοσελίδα και το ποιες σελίδες έχουν περισσότερους επισκέπτες. Αυτά τα δεδομένα προέρχονται από την ανάλυση των clickstreams (ιστορικά περιήγησης που συλλέγονται από εταιρείες), η οποία αφορά τη συλλογή και αξιολόγηση όλων των καταγεγραμμένων δραστηριοτήτων ενός χρήστη σε μια ιστοσελίδα [42].
- Ο τρόπος ενημέρωσης των επισκεπτών για μια ιστοσελίδα και ο τρόπος επίσκεψης τους σε αυτή. Συνήθως οι τρόποι αυτοί είναι τρεις: η άμεση επίσκεψη (ο χρήστης πληκτρολογεί τη διεύθυνση της ιστοσελίδας στη γραμμή διευθύνσεων του προγράμματος πλοήγησης), η επίσκεψη μέσω αναφοράς από άλλες ιστοσελίδες (μέσω http referer) και η αναζήτηση της ιστοσελίδας μέσω κάποιας μηχανής αναζήτησης (στην περίπτωση αυτή εξάγονται και σχετικές αναζητήσεις λέξεις-κλειδιών).
- Η αποτελεσματικότητα των εκστρατειών marketing που μετράται με το ποσοστό της κίνησης δεδομένων του διαδικτύου των αντίστοιχων διαφημίσεων που οδήγησαν στην επίσκεψη μιας σελίδας.
- Η γεωγραφική θέση των επισκεπτών, η οποία συνήθως εξάγεται από τις IP διευθύνσεις τους. Για το λόγο αυτό υπάρχουν διάφορες βάσεις δεδομένων, τόσο εμπορικές όσο και δωρεάν (open source), που αντιστοιχούν εύρος IP διευθύνσεων σε γεωγραφική περιοχή.
- Αναγνώριση εταιρειών, η οποία μπορεί, για παράδειγμα, να δώσει πληροφορίες για το ποιοι ανταγωνιστές έχουν επισκεφθεί την ιστοσελίδα μιας εταιρείας. Τα δεδομένα αυτά βασίζονται σε πληροφορίες IP διεύθυνσης επίσης.
- Τεχνικές λεπτομέρειες, όπως το λειτουργικό σύστημα, η ανάλυση οθόνης και η έκδοση του προγράμματος περιήγησης των επισκεπτών.

Τα λογισμικά web analytics μπορεί να είναι αυτό-φιλοξενούμενα (self-hosted), αν και συνήθως χρησιμοποιούνται υπηρεσίες τρίτων, όπως το Google Analytics [43]. Τα δεδομένα που συλλέγονται παρουσιάζονται οπτικά. Στην εικόνα 2 παρουσιάζεται ένα παράδειγμα απεικόνισης πληροφοριών στο Google Analytics. Στη συγκεκριμένη περίπτωση εμφανίζονται οι ημερήσιες επισκέψεις, οι τύποι επισκεψιμότητας και ο χρόνος παραμονής στον ιστότοπο ανά χώρα.



Εικόνα 2: Παράδειγμα απεικόνισης πληροφοριών στο Google Analytics (Πηγή: Google Analytics)[2]

Οι υπηρεσίες αυτές συνήθως απαιτούν την ύπαρξη κάποιου snippet κώδικα JavaScript στις ιστοσελίδες, το οποίο στη συνέχεια λαμβάνει περισσότερους κώδικες παρακολούθησης από έναν διακομιστή τρίτου μέρους. Οποτεδήποτε ένας χρήστης πραγματοποιεί μια συγκεκριμένη ενέργεια, όπως η μετάβαση σε μια άλλη σελίδα, ο κώδικας παρακολούθησης ενημερώνει το διακομιστή ανάλυσης σχετικά.

Ένα άλλο, διαφορετικό μοντέλο των web analytics, δεν βασίζεται σε κώδικα που βρίσκεται στην πλευρά του πελάτη (client-side model), αλλά εξάγει πληροφορίες απευθείας από τα αρχεία καταγραφής του web server (server-side model). Ένα πλεονέκτημα αυτής της προσέγγισης είναι ότι λειτουργεί ακόμα και στην περίπτωση που η Javascript στην πλευρά του πελάτη έχει απενεργοποιηθεί. Ωστόσο, στο μοντέλο client-side συλλέγονται περισσότερες πληροφορίες στον υπολογιστή του χρήστη από αυτές που αποστέλλονται από το πρόγραμμα περιήγησης από προεπιλογή.

2.5 Δοκιμές χρηστικότητας

Το web tracking μπορεί επίσης να χρησιμοποιηθεί και για την αξιολόγηση της χρηστικότητας των διαδικτυακών εφαρμογών, με σκοπό την ενίσχυση των σχεδιαστικών και αναπτυξιακών τους διαδικασιών. Με τη χρήση του JavaScript στην πλευρά του διακομιστή (server-side model), είναι δυνατή η συλλογή όλων των ενεργειών ενός χρήστη κατά τη διάρκεια χρήσης του πληκτρολογίου και του ποντικιού [44]. Με τον τρόπο αυτό, η αλληλεπίδραση με ιστοσελίδες μπορεί να αναλυθεί με κάθε λεπτομέρεια. Για παράδειγμα, τα ακόλουθα σενάρια μπορεί να καταστούν εφικτά:

- Η εγγραφή σε πραγματικό και αναπαραγωγή σε μεταγενέστερο χρόνο όλων των κινήσεων του κέρσορα με σκοπό την ανίχνευση τυχόν προβλημάτων που μπορεί να συναντούν οι χρήστες κατά τον εντοπισμό ή τη χρήση ορισμένων λειτουργιών κάποιας ιστοσελίδας. Τα δεδομένα κίνησης μπορεί να δώσουν πληροφορίες ακόμα και υψηλότερου επιπέδου σχετικά με την αλληλεπίδραση με ορισμένα στοιχεία της σελίδας, όπως τα κουμπιά και οι γραμμές κύλισης της ιστοσελίδας.

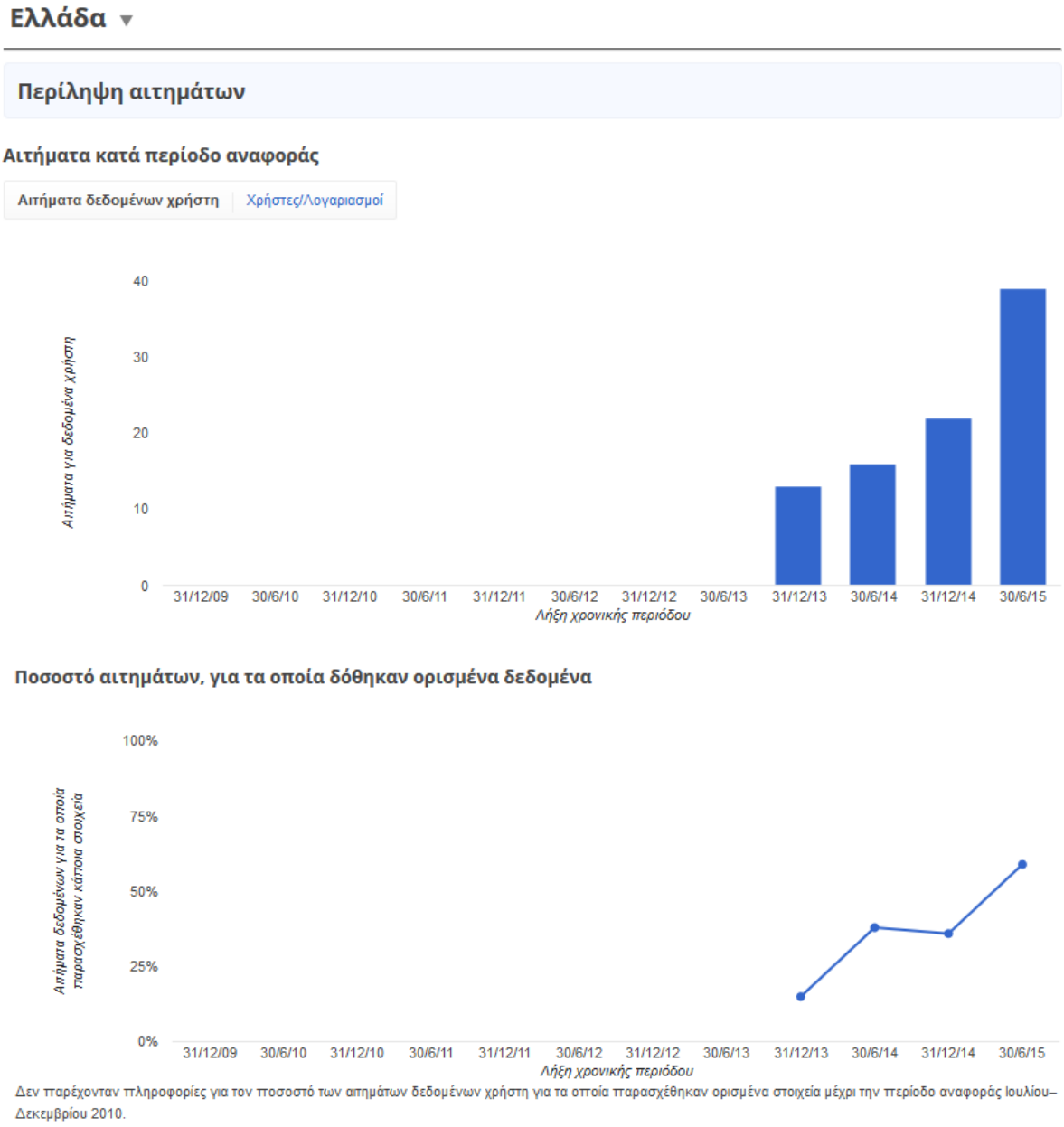
[2] Google Analytics: http://www.google.com/intl/el/analytics/

- Κατά τη συμπλήρωση ερωτηματολογίων από χρήστες, η σειρά με την οποία γίνεται η διαδικασία και ο χρόνος που χρειάζεται για κάθε βήμα μπορεί να χρησιμοποιηθούν με σκοπό την άντληση πληροφοριών για το ποιες απαντήσεις προκαλούν προβλήματα και το πώς θα μπορούσαν να αναδιαρθρωθούν τα πεδία της φόρμας.
- Τα δεδομένα που αντλούνται μέσω της αλληλεπίδρασης αυτής με τις ιστοσελίδες θα μπορούσαν ακόμη να χρησιμοποιηθούν και για την ταξινόμηση των δεξιοτήτων του εκάστοτε επισκέπτη με σκοπό την ανάλογη προσαρμογή της σελίδας. Με τον τρόπο αυτό, για παράδειγμα, έμπειροι χρήστες θα μπορούσαν να έχουν πρόσβαση σε πιο αναβαθμισμένη έκδοση μιας ιστοσελίδας ενώ οι άπειροι χρήστες θα μπορούσαν να λαμβάνουν πρόσθετες υποδείξεις για τη χρήση της.

Η χρήση του JavaScript στην πλευρά του διακομιστή, παρουσιάζει το πλεονέκτημα ότι δίνει λύσεις παρακολούθησης που δεν απαιτούν ειδικό λογισμικό ή hardware υλικό. Ένα επιπλέον πλεονέκτημα είναι η δυνατότητα παρακολούθησης των χρηστών μέσα στον ίδιο τον υπολογιστή που χρησιμοποιούν, γεγονός που επίσης μειώνει το κόστος των δοκιμών [44]. Η χρήση συσκευών οφθαλμικής ιχνηλάτησης (eye-tracking devices) μπορεί να δώσει ακόμα περισσότερες δυνατότητες παρακολούθησης με σκοπό την ανάλυση των αντικειμένων μιας ιστοσελίδας που τραβούν περισσότερο την προσοχή των χρηστών [45].

2.6 Κυβερνητική παρακολούθηση

Τα δεδομένα που λαμβάνονται από την παρακολούθηση των χρηστών αποτελούν πολύτιμη πηγή πληροφοριών για κυβερνητικές υπηρεσίες και αρχές επιβολής του νόμου. Μεταξύ Ιανουαρίου και Ιουνίου του 2015, η ελληνική κυβέρνηση κατέθεσε 39 αιτήματα στη Google για παροχή πληροφοριών, όπως το ιστορικό αναζητήσεων, που αφορούσαν 59 άτομα και η απάντηση της Google ήταν θετική στο 59% των περιπτώσεων αυτών (Εικ. 3) [46]. Η Google επίσης σαρώνει ενεργά τις εικόνες που περνούν μέσα από τους λογαριασμούς του Gmail για να εξακριβώσει αν ταιριάζουν με περιπτώσεις παιδικής πορνογραφίας [47].



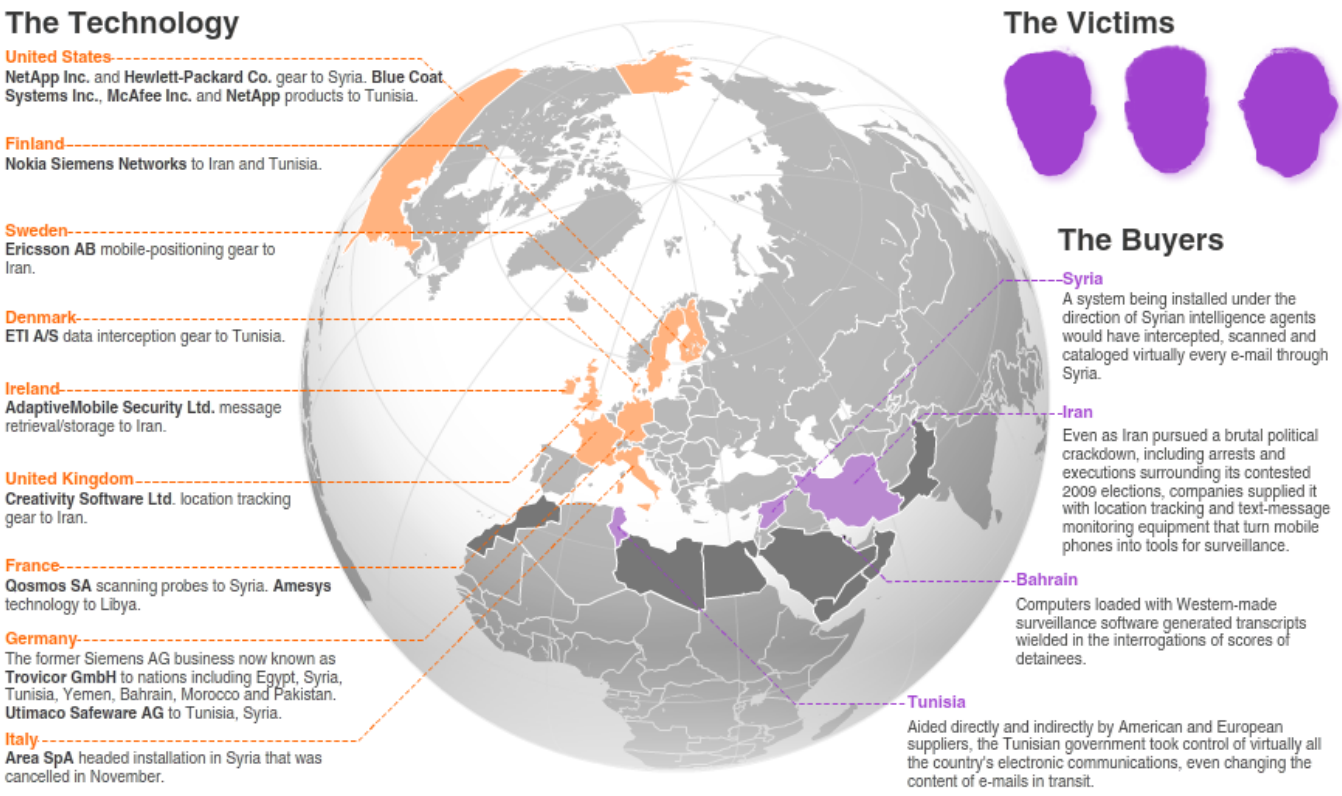
Εικόνα 3: Αιτήματα κατά περίοδο αναφοράς και ποσοστό αιτημάτων για τα οποία δόθηκαν ορισμένα δεδομένα (Πηγή: Google)[3]

Τα τελευταία χρόνια, ένας συνδυασμός τεχνικών έρευνας, διαρροών και άρσεων απορρήτου αποκάλυψε πρωτοφανή στοιχεία σχετικά με την παρακολούθηση του διαδικτύου από τις κυβερνήσεις [48]. Μερικά κράτη, όπως το Ιράν και το Μπαχρέιν (Εικ. 4) [49], εφαρμόζουν σχεδόν πλήρη παρακολούθηση διαδικτύου. Άλλα κράτη, όπως οι Ηνωμένες Πολιτείες και η Μ. Βρετανία, έχουν τη δυνατότητα για κάτι τέτοιο από τεχνικής άποψης, αλλά υπόκεινται σε νομικά πλαίσια.

Από έγγραφα που διέρρευσαν αποδεικνύεται ότι η Αμερικανική Υπηρεσία Πληροφοριών (NSA) χρησιμοποιεί cookies τρίτων για παρακολούθηση του διαδικτύου. Οι τρόποι με τους οποίους χρησιμοποιούνται αυτά τα cookies είναι τουλάχιστον τρεις. Κατ' αρχάς, η υπηρεσία

[3] Google: https://www.google.com/transparencyreport/userdatarequests/GR/

διεξήγαγε έρευνα αναγνωρίζοντας παθητικά χρήστες του δικτύου Tor συσχετίζοντάς cookies με μη Tor συνόδους. Συγκεκριμένα, η NSA επιχείρησε να συνδέσει ένα cookie διαφήμισης τρίτων της Google (DoubleClick cookie) μεταξύ Tor και μη Tor συνόδων [50]. Δεύτερον, η υπηρεσία διαθέτει ένα ενεργό, man-in-the-middle σύστημα, που ονομάζεται « QUANTUMCOOKIE», το οποίο προκαλεί την αποκάλυψη των cookies. Η συγκεκριμένη τεχνολογία εφαρμόζεται στην αναγνώριση χρηστών του δικτύου Tor και στην αντιμετώπιση κακόβολου λογισμικού [51]. Τρίτον, η υπηρεσία χρησιμοποιεί παθητικά λαμβανόμενα cookies ως μέσο παρακολούθησης, εκμεταλλευόμενη την περίπτωση man-in-the-middle. Σε μια τουλάχιστον περίπτωση, σύμφωνα με την εφημερίδα Washington Times, η NSA εκμεταλλεύτηκε τις πληροφορίες που περιέχουν τα cookies της Google για τον εντοπισμό και την παρακολούθηση υπόπτων [51]. Εκτός από αυτές τις εφαρμογές, χρησιμοποιούνται εργαλεία HTTP ανάλυσης, όπως το «XKEYSCORE», που ενσωματώνουν τα δεδομένα των cookies. Με τον τρόπο αυτό, ένας αναλυτής θα μπορούσε εύκολα να επωφεληθεί από τα cookies τρίτων υποβάλλοντας ερωτήματα στα δεδομένα που υποκλέπτονται [48].



Εικόνα 4: Η αγορά παρακολούθησης και τα θύματά της (Πηγή: BloombergBusiness)[4]

Από έγγραφα που διέρρευσαν αποκαλύφθηκαν επίσης δύο προγράμματα επιτήρησης και στόχευσης των χρηστών από την Βρετανική Κρατική Υπηρεσία Επικοινωνιών (GCHQ), μέσω της παρακολούθησης δεδομένων τρίτου μέρους τόσο μέσα από προγράμματα περιήγησης στο web όσο και από mobile εφαρμογές. Το πρώτο πρόγραμμα, το «MUTANT BROTH», είναι ένα αποθετήριο των cookies παρακολούθησης που συνδέονται με επιπλέον μεταδεδομένα, όπως διευθύνσεις IP και User-Agent strings. Αυτό το αποθετήριο φέρεται να έχει χρησιμοποιηθεί σε

[4] Bloomberg Business: http://www.bloomberg.com/data-visualization/wired-for-repression/

περιπτώσεις στοχευμένου κακόβουλου λογισμικού [52]. Το άλλο πρόγραμμα, το «BADASS», παρέχει ένα παρόμοιο αποθετήριο και μια διεπαφή αναζήτησης για επερωτήσεις σε πληροφορίες που έχουν διαρρεύσει από mobile εφαρμογές. Το σύστημα συλλέγει και εξάγει αναγνωριστικά που έχουν διαρρεύσει, λεπτομέρειες για συσκευές και λειτουργικά συστήματα καθώς και επιπλέον πληροφορίες που διαβιβάζονται μέσα σε απλό κείμενο [53].

2.7 Κλοπή ταυτότητας

Μια ακόμα επίπτωση της χρήσης του web tracking είναι η χρησιμοποίηση των συλλεχθέντων πληροφοριών για την κλοπή της ταυτότητας των χρηστών. Ως κλοπή ταυτότητας ορίζεται η μη εξουσιοδοτημένη χρήση των προσωπικών στοιχείων κάποιου με σκοπό την επίτευξη παράνομων δραστηριοτήτων. Μελέτες του Carnegie Mellon University απέδειξαν ότι ο συνδυασμός των δεδομένων που αποκαλύπτουν οι χρήστες του διαδικτύου σε διάφορες ιστοσελίδες είναι σε πολλές περιπτώσεις αρκετός για να βρεθούν άλλα στοιχεία ζωτικής σημασίας, όπως, για παράδειγμα, ο αριθμός κοινωνικής ασφάλισής τους. Η ανεύρεση αυτών των στοιχείων, ανοίγει τον δρόμο για την κλοπή της ταυτότητας κάποιου, καθώς, ο αριθμός κοινωνικής ασφάλισης χρησιμοποιείται ευρέως για τον έλεγχο ταυτότητας σε ευαίσθητες ιστοσελίδες (π.χ. τραπεζικές ή δανειακές υπηρεσίες) [54].

Το 2012, έρευνα της Javelin Strategy & Research σχετικά με την εξαπάτηση ταυτότητας στο διαδίκτυο αποκάλυψε ότι οι χρήστες ιστοσελίδων όπως LinkedIn, Google+ και Facebook είναι πολύ πιο πιθανό να πέσουν θύματα απάτης σε σύγκριση με χρήστες του διαδικτύου που δεν χρησιμοποιούν τις παραπάνω υπηρεσίες. Ο λόγος θα μπορούσε να είναι η αποκάλυψη στοιχείων, όπως η ημερομηνία γέννησης και το σχολείο στο οποίο φοίτησαν (σε ποσοστό πάνω από 60%) ή ο αριθμός τηλεφώνου τους (σε ποσοστό κοντά στο 20%) [55]. Επίσης, οι χρήστες φορητών συσκευών που κάνουν check-in μέσω του ενσωματωμένου GPS αποδείχθηκε ότι είναι πάνω από το διπλάσιο πιθανό να πέσουν θύματα απάτης ή κλοπής ταυτότητας σε σύγκριση με χρήστες που χρησιμοποιούν άλλες mobile εφαρμογές [56].

3 Αναγνώριση Εντοπισμένου Χρήστη

Η ανίχνευση και παρακολούθηση της ψηφιακής ταυτότητας των χρηστών μπορεί να αποδειχθούν πολύτιμες για πολλούς λόγους. Ακόμη πιο επικερδής θα μπορούσε να αποδειχθεί η δυνατότητα σύνδεσης της ψηφιακής με την πραγματική ταυτότητά τους (ονοματεπώνυμο, αριθμός κοινωνικής ασφάλισης, κλπ.). Έχει αποδειχθεί ότι το 87% του πληθυσμού των ΗΠΑ μπορεί να αναγνωριστεί επαρκώς με βάση μόνο 3 στοιχεία: την ημερομηνία γέννησης, το φύλο και τον ταχυδρομικό κώδικα [57]. Η δυνατότητα αυτή μπορεί να αποτελέσει σοβαρή απειλή για την ασφάλεια των χρηστών. Εάν ένας χρήστης αποφασίσει να αποκαλύψει την πραγματική του ταυτότητα σε μία υπηρεσία (π.χ., κατά τη διάρκεια της εγγραφής του σε ένα λογαριασμό ηλεκτρονικού ταχυδρομείου), τότε κινδυνεύει να αναγνωριστεί και από όλες τις άλλες οντότητες με κοινή βάση δεδομένων των χρηστών, στην οποία επίσημα περιέχονται ανώνυμες πληροφορίες, όπως η ημερομηνία γέννησης και το φύλο.

Στο κεφάλαιο αυτό γίνεται μια αναφορά στους συνηθέστερους τρόπους αναγνώρισης των χρηστών με βάση τις online δραστηριότητές τους καθώς και το τι αντίκτυπο έχει αυτή η αναγνώριση στην παραβίαση της ιδιωτικότητάς τους.

3.1 Web tracking & ιδιωτικότητα

Οι τεχνικές δυνατότητες παρακολούθησης των δραστηριοτήτων των χρηστών σε ιστοσελίδες έχουν πολλαπλασιαστεί κατά την τελευταία δεκαετία και η αναδυόμενη «Κοινωνία της Πληροφορίας» σε αυτό το διάστημα έχει βιώσει αρκετές αλλαγές [12]. Το web tracking αναπτύχθηκε κυρίως όταν πάροχοι online υπηρεσιών άρχισαν να παρακολουθούν τους χρήστες τους για να διαπιστώσουν αν ένας συγκεκριμένος χρήστης είχε ξαναχρησιμοποιήσει τη συγκεκριμένη υπηρεσία και ποιες ήταν οι δραστηριότητές του. Από τότε η εξέλιξή του ήταν ραγδαία αφού πλέον η μεγαλύτερη πλειοψηφία των εμπορικών φορέων μπορεί να έχει μια ολοκληρωμένη άποψη των δραστηριοτήτων των χρηστών. Με τον τρόπο αυτό, κάθε φορέας φαίνεται να είναι σε θέση να παρακολουθεί κάθε πτυχή της συμπεριφοράς ενός αναγνωρίσιμου χρήστη μέσω των ιστοσελίδων που επισκέπτεται. Αυτό που επιτυγχάνεται τελικά είναι μια πλήρης καταγραφή του ιστορικού της χρήσης του διαδικτύου από τους χρήστες, εμπλουτισμένο με στοιχεία του προφίλ της πραγματικής τους ζωής, όπως χρηματοπιστωτικές πληροφορίες καθώς και πληροφορίες σχετικά με την ψυχαγωγία, την υγεία, τις πολιτικές και θρησκευτικές πεποιθήσεις ή πληροφορίες της θέσης στην οποία βρίσκεται [12].

Αυτή η εξέλιξη του web tracking χαιρετήθηκε και ενισχύθηκε από τους εμπορικούς φορείς αλλά και από άλλα ενδιαφερόμενα μέρη της ευρύτερης επιχειρηματικής κοινότητας, καθώς και από ορισμένους φορείς χάραξης πολιτικής σε εθνικό και περιφερειακό επίπεδο. Όσον αφορά τον ίδιο τον χρήστη όμως, εγκυμονεί πρωτοφανείς κινδύνους για την προστασία της ιδιωτικότητας του στην κοινωνία της πληροφορίας. Οι πιθανές επιπτώσεις μιας τέτοιας εξέλιξης είναι προφανείς και δεν πρέπει να υποτιμούνται σε σχέση με το δυναμικό της βαρύτητας τους, αφού μπορεί να ακυρώσουν μερικές από τις βασικές αρχές της προστασίας της ιδιωτικότητας του ατόμου, όπως είναι η διαφάνεια και ο έλεγχος.

Οι υποστηρικτές αυτής της εξέλιξης του web tracking, από την άλλη πλευρά, ισχυρίζονται ότι είτε οι κίνδυνοι αυτοί δεν υπάρχουν καθόλου είτε ότι έχουν προσπαθήσει να τους αντιμετωπίσουν και να τους αμβλύνουν, τουλάχιστον εν μέρει. Επιπλέον, ισχυρίζονται ότι τα δεδομένα που συλλέγονται μέσω του web tracking δεν αποτελούν προσωπικές πληροφορίες των χρηστών. Ένας ισχυρισμός που συχνά προβάλλεται είναι ότι μεγάλο μέρος των δεδομένων

που χρησιμοποιείται έχει ήδη ανωνυμοποιηθεί, γεγονός που σημαίνει ότι δεν τίθεται κανένα θέμα όσον αφορά την ασφάλεια της ιδιωτικότητας των χρηστών. Επίσης, υπάρχει ο ισχυρισμός ότι οποιαδήποτε στοιχεία συμπεριφοράς συνδέονται αποκλειστικά και μόνο με το μέσο περιήγησης στο διαδίκτυο και σε καμία περίπτωση τα στοιχεία αυτά δεν ανάγονται στους χρήστες.

Ωστόσο, αυτοί οι ισχυρισμοί δεν έχουν καμία απολύτως επιστημονική απόδειξη, ενώ παράλληλα δεν θα πρέπει να αγνοηθεί το γεγονός ότι συσκευές, όπως τα έξυπνα τηλέφωνα, γίνονται ολοένα και περισσότερο προσωπικές και επιτρέπουν τον εύκολο συσχετισμό με οποιοδήποτε μεμονωμένο χρήστη. Εξάλλου, έχει αποδειχθεί ότι πολλά φαινομενικά ανώνυμα δεδομένα, όπως οι πληροφορίες εντοπισμού των κινητών τηλεφώνων, μπορεί να αναχθούν στον εντοπισμό οποιουδήποτε συγκεκριμένου χρήστη, εάν η βάση δεδομένων και το χρονικό πλαίσιο είναι αρκετά ευρεία [58]. Η πιο πρόσφατη ακαδημαϊκή έρευνα απέδειξε ότι είναι αδύνατη η διατήρηση της ανωνυμίας των δεδομένων αν η χρονική περίοδος που απεικονίζει την online συμπεριφορά των χρηστών είναι αρκετά μεγάλη, δηλαδή είναι σχεδόν αδύνατη η εγγύηση ότι τα «ανώνυμα» δεδομένα δεν γίνεται να αναχθούν σε ένα συγκεκριμένο άτομο με την πάροδο του χρόνου. Δεδομένου του γεγονότος αυτού, είναι αδύνατη η παραδοχή ότι η συνολική συμπεριφορά των χρηστών στο διαδίκτυο δεν μπορεί να επηρεάσει την προστασία της ιδιωτικότητας τους [12].

Επιπρόσθετα, η γνώση που προκύπτει από την καθημερινή πρακτική ακυρώνει τους ισχυρισμούς των υποστηρικτών της εξέλιξης του web tracking. Μπορεί ναι μεν οι διαφημίσεις, σε τεχνικό επίπεδο, να απευθύνονται στις συσκευές σύνδεσης στο διαδίκτυο, η αγορά όμως των προϊόντων γίνεται από τους χρήστες. Έτσι, ο ισχυρισμός ότι η επεξεργασία των δεδομένων συμπεριφοράς marketing απευθύνεται εξ αρχής «μόνο» στις συσκευές, μπορεί να θεωρηθεί ως μια προσπάθεια αποκήρυξης του αληθινού προβλήματος, αφού σε τελική ανάλυση είναι το άτομο και όχι η συσκευή η μόνη περίπτωση «επιτυχίας» όλων των διεργασιών εντοπισμού για τους υποστηρικτές του. Η «επιτυχία» αυτή έγκειται στο γεγονός της αγοράς των διαφημιζόμενων προϊόντων.

3.2 Το δικαίωμα της προστασίας των προσωπικών δεδομένων

Μια βασική αρχή στο ευρύ φάσμα των διεθνών νομοθετικών πλαισίων, είναι το δικαίωμα της προστασίας των προσωπικών δεδομένων των χρηστών του διαδικτύου. Βασικά στοιχεία είναι η διαφάνεια, ο έλεγχος και ο σεβασμός για το περιεχόμενο [12]. Το γεγονός και μόνο ότι οι χρήστες αγνοούν ότι παρακολουθούνται, αποτελεί από μόνο του κίνδυνο της προστασίας της ιδιωτικότητάς τους. Το web tracking ως διαδικασία χρησιμοποιεί μια σειρά τεχνικών μέσων που περιορίζουν τη δυνατότητα ενημέρωσης των χρηστών. Για παράδειγμα, pixels, όπως τα web beacons, και μίνι ιστοσελίδες, όπως τα iFrames, αγνοούνται από τον μέσο χρήστη και η ύπαρξή τους σε μια ιστοσελίδα δημιουργεί αυτόματα αιτήματα HTTP, όπως η πρόσβαση σε cookies που περιέχουν μοναδικά αναγνωριστικά.

Πολλές από τις τεχνολογίες web tracking έχουν αναπτυχθεί και εφαρμοστεί για επιχειρηματικούς σκοπούς, χωρίς την σχετική παροχή πληροφοριών στους χρήστες των οποίων τα δεδομένα συλλέγονται χωρίς να έχουν κάποια εναλλακτική. Οι όποιες προσπάθειες, που θα μπορούσαν να εκληφθούν ως έκφραση αντίρρησης στην παρακολούθηση, μέσω τεχνικών μηχανισμών έναντι ορισμένων μηχανισμών παρακολούθησης έχουν παρακαμφθεί ενεργά, όπως, για παράδειγμα, η εκ νέου δημιουργία διαγραφόμενων cookies, η παθητική λήψη

αποτυπωμάτων και η παράκαμψη των ρυθμίσεων του εκάστοτε χρησιμοποιούμενου προγράμματος περιήγησης. Ο εντοπισμός αυτής της παράκαμψης και η δημόσια κατακραυγή της οδήγησε τα ενδιαφερόμενα μέρη να αποδεχθούν την υποχρέωσή τους στο να σέβονται την ελεύθερη βούληση των χρηστών [12]. Όμως, η μετέπειτα προσθήκη μηχανισμών αυτοεξαίρεσης (opt-out schemes) οδήγησαν στην περιορισμένη παρεμβατική ικανότητα των χρηστών, γεγονός που έχει προκαλέσει μεγάλη ζημιά στην εμπιστοσύνη των χρηστών όσον αφορά την αξιοπιστία και την εντιμότητα όλων των παρόχων υπηρεσιών διαδικτύου και έχει υπονομεύσει την υγιή ανάπτυξη καινοτόμων διαδικτυακών υπηρεσιών.

Σύμφωνα με τις διατάξεις της οδηγίας 95/46/ΕΚ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 24ης Οκτωβρίου 1995, για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων αυτών, το web tracking ουσιαστικά αποτελεί επεξεργασία προσωπικών δεδομένων λόγω του γεγονότος ότι επιτρέπει την εξατομίκευση και την ταυτοποίηση των χρηστών ή/και τη λήψη αυτοματοποιημένων αποφάσεων γι' αυτούς [59]. Ένα παράδειγμα τέτοιας πρακτικής αποτελούν οι αυτόματοι μηχανισμοί αποφάσεων με αλγορίθμους σε πλατφόρμες προσφορών σε πραγματικό χρόνο με σκοπό τη στοχευμένη διαφήμιση.

3.3 Τρόποι αναγνώρισης εντοπισμένων χρηστών

Με βάση τα όσα αναφέρθηκαν στις προηγούμενες ενότητες γίνεται φανερό ότι τα ιστορικά περιήγησης που συλλέγονται από εταιρείες (clickstreams) σε καμιά περίπτωση δεν θεωρούνται ανώνυμα, αλλά μάλλον ψευδώνυμα [60]. Ο τελευταίος όρος, όχι μόνο μπορεί να θεωρηθεί ως καταλληλότερος τεχνικά αλλά αντανακλά καλύτερα το γεγονός πως μετά τη συλλογή των δεδομένων η εταιρεία παρακολούθησης μπορεί να προσπαθήσει να συσχετίσει την πραγματική ταυτότητα των χρηστών με το ψευδώνυμο (username ως μοναδικό αναγνωριστικό) με το οποίο έχουν συνδυαστεί τα δεδομένα αυτά. Έτσι, η αναγνώριση ενός χρήστη επηρεάζει όχι μόνο μελλοντικές παρακολουθήσεις αλλά επιτρέπει και την αναδρομική συσχέτιση με τα δεδομένα που έχουν ήδη συλλεχθεί. Η ταυτοποίηση των στοιχείων αρκεί να γίνει μία και μόνο φορά για τον κάθε χρήστη.

Οι τρόποι με τους οποίους μπορεί να γίνει αυτή η ταυτοποίηση των στοιχείων είναι πλέον πολλοί. Οι πιο συνηθισμένοι είναι οι εξής [60]:

- Παρακολούθηση νόμιμων υπηρεσιών ως τρίτοι
- Διαρροή πληροφοριών προς τρίτους
- Πώληση πληροφοριών προς τρίτους
- Απο-ανωνυμοποίηση δεδομένων

3.3.1 Παρακολούθηση νόμιμων υπηρεσιών ως τρίτοι

Δεν είναι σπάνιο το φαινόμενο, ιστοσελίδες κοινωνικής δικτύωσης ή μεγάλων εταιρειών με ευρεία απήχηση να λειτουργούν ως τρίτοι και να παρακολουθούν τους ίδιους τους χρήστες τους. Κάτι τέτοιο ισχύει για το Facebook ή τη Google, όπου η χρήση ενσωματωμένων κουμπιών («Like» και «+1», αντίστοιχα) σε άλλες ιστοσελίδες χρησιμοποιείται με σκοπό την παρακολούθηση των δικών τους χρηστών σε δικτυακούς τόπους τρίτων. Για τη λειτουργία αυτών των κουμπιών οι χρήστες θα πρέπει να συνδεθούν ή να είναι συνδεδεμένοι στην αντίστοιχη ιστοσελίδα, η παρακολούθησή τους όμως συνεχίζεται και μετά την αποσύνδεση. Η

Google αναφέρει ότι διατηρεί ορισμένες πληροφορίες σχετικά με την επίσκεψή του χρήστη, συνήθως για διάστημα δύο εβδομάδων, για τη συντήρηση και τη διόρθωση σφαλμάτων του συστήματός της και πως αυτές οι πληροφορίες δεν οργανώνονται ανά μεμονωμένα προφίλ, ονόματα χρήστη ή διευθύνσεις URL [61]. Το Facebook αναφέρει ότι ανάμεσα στα δεδομένα που λαμβάνει είναι το αναγνωριστικό του χρήστη, ο ιστότοπος που επισκέπτεται, η ημερομηνία και η ώρα, καθώς και άλλες πληροφορίες που σχετίζονται με το πρόγραμμα περιήγησης [62].

Η παρακολούθηση των χρηστών αφορά τις δραστηριότητες τους στο διαδίκτυο και έχει ως σκοπό τη διαφήμιση μέσω στόχευση συμπεριφοράς σε συνδυασμό με προσωπικές πληροφορίες αναγνώρισης και μπορεί να θεωρηθεί ως παραβίαση της ασφάλειας των προγραμμάτων περιήγησης και της ιδιωτικότητας των χρηστών. Για το λόγο αυτό πολλές αμερικανικές οργανώσεις, τον Ιούνιο του 2010, έστειλαν ανοιχτή επιστολή στο Facebook ώστε να μην διατηρεί δεδομένα για συγκεκριμένους επισκέπτες σε δικτυακούς τόπους τρίτων που ενσωματώνουν κουμπιά «κοινωνικών plugins» ή «Like», παρά μόνο μετά από απόφαση του ίδιου του επισκέπτη της ιστοσελίδας [63]. Τον Νοέμβριο του 2015, η κυβέρνηση του Βελγίου ζήτησε από το Facebook να σταματήσει τον εντοπισμό ανθρώπων που δεν είχαν συνδεθεί στο Facebook. Η ένσταση αφορούσε τη συλλογή προσωπικών πληροφοριών από όλους όσους επισκέπτονται ιστοσελίδες που περιέχουν το κουμπί «Like» του Facebook, ανεξάρτητα από το αν είχαν λογαριασμούς σε αυτό[64].

3.3.2 Διαρροή πληροφοριών προς τρίτους

Προσωπικές πληροφορίες ενδέχεται να διαρρεύσουν προς τρίτους με πολλούς τρόπους [60]. Σχεδόν τα 3/4 των ιστοσελίδων παρουσιάζουν διαρροή ευαίσθητων πληροφοριών ή ακόμα και αναγνωριστικά στοιχεία των χρηστών. Ανάμεσα στους τρόπους αυτούς, ο πιο εύκολος είναι μέσω του πεδίου referer της κεφαλίδας του HTTP [65]. Το referer είναι ένα πεδίο στην κεφαλίδα των αιτήσεων HTTP που κάνει το πρόγραμμα περιήγησης του χρήστη. Στο πεδίο αυτό αναγράφεται η διεύθυνση URL μιας σελίδας στην οποία υπήρχε σύνδεσμος για την σελίδα στην οποία γίνεται η αίτηση. Ελέγχοντας το referer η σελίδα προορισμός μπορεί να γνωρίζει την σελίδα αφετηρίας του χρήστη. Με άλλα λόγια, όταν ο χρήστης βρίσκεται στη σελίδα Α και κάνει κλικ σε ενεργό σύνδεσμο προς την σελίδα Β, τότε το πρόγραμμα περιήγησης καθώς ανοίγει την σελίδα Β αναγράφει στο referer ότι προήλθε από την σελίδα Α. Έτσι ο διαχειριστής μιας σελίδας, λόγω του referer, μπορεί να γνωρίζει από ποιες σελίδες προέρχονται οι επισκέπτες (μηχανή αναζήτησης, social media / blogpost κλπ.). Κάτι τέτοιο μπορεί να είναι ανεπιθύμητο για έναν χρήστη καθώς με τον τρόπο αυτό μπορεί να τεθούν σε κίνδυνο προσωπικά δεδομένα, όπως ονοματεπώνυμο, μηνύματα ηλεκτρονικού ταχυδρομείου, μοναδικά αναγνωριστικά και δεδομένα θέσης, τα οποία μπορεί να διαρρεύσουν χωρίς λόγο σε τρίτους, ως αποτέλεσμα λαθών ή πρακτικών κακού προγραμματισμού σε ιστοσελίδες και mobile εφαρμογές.

Εκτός από το πεδίο referer της κεφαλίδας του HTTP, άλλοι τρόποι μέσω των οποίων μπορεί να υπάρξει διαρροή δεδομένων προς τρίτους είναι [60]:

- Αιτήσεις URL που μπορεί να περιέχουν στοιχεία όπως φύλο, ταχυδρομικός κώδικας ή ενδιαφέροντα.
- Αναγνωριστικά σε cookies κοινής χρήσης που προέρχονται από διακομιστές «κρυφών τρίτων μερών».
- Όνομα Χρήστη ή πραγματικό όνομα στον τίτλο της σελίδας.

3.3.3 Πώληση πληροφοριών προς τρίτους

Είναι συχνό το φαινόμενο, να επισκέπτονται χρήστες μια ιστοσελίδα για πρώτη φορά και στην οθόνη του υπολογιστή τους να εμφανίζεται μια pop-up σελίδα με τον τίτλο «Έρευνα: "Κερδίστε ένα iPad!"». Το επιχειρηματικό μοντέλο για πολλές από αυτές τις σελίδες είναι η συλλογή και η πώληση των προσωπικών στοιχείων των χρηστών που συμμετέχουν σε τέτοιου είδους «έρευνες». Ολοένα και περισσότερο, αυτές οι σελίδες αναπτύσσουν δεσμούς με εταιρείες παρακολούθησης. Άλλοι τύποι εταιρειών, όπως οι λίστες άμεσου μάρκετινγκ (direct marketing lists) και οι εταιρείες ανταλλαγής/συγκέντρωσης δεδομένων των καταναλωτών (consumer data exchanges/aggregators) παίζουν ακριβώς τον ίδιο ρόλο με τις εταιρείες έρευνας [60]. Με την αποκάλυψη της ταυτότητας σε μια σελίδα έρευνας, η ταυτότητα αυτή μπορεί με δύο τρόπους να συσχετιστεί με το ιστορικό περιήγησης των χρηστών.

Πρώτον, η ίδια η σελίδα έρευνας μπορεί να λειτουργεί ως τρίτος σε άλλους δικτυακούς τόπους. Με τον τρόπο αυτό, η επίσκεψη και η εγγραφή στην ιστοσελίδα έρευνας σημαίνει αυτόματη σύνδεση των παρεχόμενων πληροφοριών εγγραφής με την clickstream των ήδη συλλεχθέντων δεδομένων του χρήστη και επομένως την μετέπειτα παροχή της ταυτότητάς του στις σελίδες όπου η σελίδα έρευνας λειτουργεί ως τρίτος. Εναλλακτικά, η ταυτότητα των χρηστών περνάει, μέσω των μεθόδων που αναφέρθηκαν στην προηγούμενη υποενότητα, σε trackers που είναι ενσωματωμένοι στη σελίδα έρευνας, γεγονός που επιτρέπει την σύνδεση των πληροφοριών αυτών με cookies αναγνώρισης τα οποία με τη σειρά τους τις συνδέουν με το ιστορικό περιήγησης. Με άλλα λόγια, με την εγγραφή σε μια σελίδα έρευνας, ο tracker αποκτά το ιστορικό περιήγησης, η σελίδα έρευνας αποκτά την ταυτότητα και ο συνδυασμός των δύο μπορεί να γίνει μέσω της κεφαλίδας referer ή άλλων τρόπων διαρροής πληροφοριών.

3.3.4 Απο-ανωνυμοποίηση δεδομένων

Οι προαναφερθέντες τρόποι αναγνώρισης των εντοπισμένων χρηστών αφορούν την άμεση ή έμμεση αλληλεπίδρασή τους με τρίτους. Τα ίδια τα δεδομένα όμως υπάρχει η δυνατότητα να απο-ανωνυμοποιηθούν μέσω εξωτερικής πληροφόρησης και πιο συγκεκριμένα μέσω πτυχών του ιστορικού περιήγησης των χρηστών που κατά καιρούς αποκαλύπτονται δημοσίως. Επομένως, μια εταιρεία παρακολούθησης μπορεί πολύ εύκολα να χρησιμοποιήσει τις βάσεις δεδομένων των clickstreams και να ανακαλύψει την πραγματική ταυτότητα των χρηστών. Η φιλοσοφία του τρόπου αυτού αναγνώρισης των εντοπισμένων χρηστών είναι πολύ απλή [60].

Κάθε χρήστης στη διάρκεια της μέρας κάνει κάποιες διαδικτυακές δραστηριότητες, όπως να σχολιάσει ένα άρθρο, να ανεβάσει μια ανάρτηση στο Facebook, να απαντήσει σε κάποια ανακοίνωση στο Twitter ή να πάρει μέρος σε μια συζήτηση σε κάποιο blog. Με τις δραστηριότητες αυτές έχει αυτομάτως δημιουργήσει ένα δημόσιο αρχείο που περιέχει τις διευθύνσεις URL των σελίδων που έχει επισκεφθεί. Το αρχείο αυτό είναι μοναδικό καθώς είναι απίθανο κάποιος άλλος χρήστης να επισκέφθηκε τις ίδιες ιστοσελίδες την ίδια ακριβώς στιγμή της ημέρας. Αυτό σημαίνει ότι ένας αλγόριθμος που σαρώνει τις βάσεις δεδομένων των ανώνυμων clickstreams μπορεί εύκολα να ταυτίσει την clickstream του εκάστοτε χρήστη με την πραγματική του ταυτότητα.

4 Εξέλιξη των Μηχανισμών Web Tracking

Το γενικό συμπέρασμα που προκύπτει μετά τα τρία πρώτα κεφάλαια της παρούσας διατριβής είναι ότι το web tracking έχει εξελιχθεί πλέον σε διαδικασία που αφορά την παρακολούθηση των χρηστών του διαδικτύου και την συλλογή δεδομένων που τους αφορούν με απώτερο σκοπό τη δημιουργία του διαδικτυακού τους προφίλ. Οι διαδικασίες παρακολούθησης και συλλογής δεδομένων πραγματοποιούνται μέσω διάφορων μηχανισμών web tracking. Η ανάπτυξη της ψηφιακής βιομηχανίας του μάρκετινγκ τα τελευταία χρόνια έχει οδηγήσει στην εξέλιξη πιο προηγμένων και δυσκολότερα εντοπίσιμων μηχανισμών παρακολούθησης.

Η παρακολούθηση των χρηστών μπορεί να πραγματοποιείται σε ένα δικτυακό τόπο, στον οποίο υπάρχει και ο αντίστοιχος μηχανισμός, ή σε διάφορες ιστοσελίδες, όπου οι μηχανισμοί παρακολούθησης συλλέγουν δεδομένα μέσω των δραστηριοτήτων περιήγησης των χρηστών σε διαφορετικές ιστοσελίδες. Η δεύτερη περίπτωση είναι η πιο χρήσιμη στη διαμόρφωση των προφίλ των χρηστών, καθώς παρέχει τη δυνατότητα συλλογής μιας ευρείας γκάμας πληροφοριών. Η περίπτωση αυτή επίσης αφορά ένα είδος παρακολούθησης που τις περισσότερες φορές είναι δύσκολο να εντοπιστεί.

Σκοπός του παρόντος κεφαλαίου είναι η παρουσίαση της εξέλιξης των συνηθέστερων μηχανισμών web tracking που έχουν αναπτυχθεί την τελευταία 20ετία και χρησιμοποιούνται για να «ακολουθούν» τους χρήστες του διαδικτύου και να συλλέγουν δεδομένα μέσω της διαδικτυακής συμπεριφορά τους. Οι μηχανισμοί αυτοί θα μπορούσαν να χωριστούν γενικά σε δύο μεγάλες κατηγορίες: τους μηχανισμούς παρακολούθησης και αναγνώρισης των χρηστών που λειτουργούν στη μεριά του πελάτη (client-side mechanisms) και τους μηχανισμούς αποτύπωσης (fingerprinting). Οι μηχανισμοί αποτύπωσης θα εξεταστούν στο επόμενο κεφάλαιο.

4.1 Client-side μηχανισμοί παρακολούθησης

Σε αυτή την κατηγορία των μηχανισμών παρακολούθησης τα στοιχεία παρακολούθησης (δεδομένα, αρχεία, πρόσθετα (add-ons), κλπ.) αποθηκεύονται τοπικά από τα προγράμματα περιήγησης σε διαφορετικούς χώρους του υπολογιστή-πελάτη. Τα στοιχεία αυτά διαβιβάζονται στους διαδικτυακούς διακομιστές και χρησιμοποιούνται για την αναγνώριση του χρήστη και την πραγματοποίηση ενεργειών σύμφωνα με το προφίλ του. Η αφαίρεση των αποθηκευμένων αυτών στοιχείων δεν αποτελεί πάντα έναν αυτόματο ή προγραμματισμένο μηχανισμό, γεγονός που καθιστά τη διαδικασία κατάργησής τους δύσκολη και παράλληλα ευνοεί τη μονιμότητά τους.

Όπως αναφέρθηκε και στην εισαγωγή της παρούσας διατριβής οι μηχανισμοί αυτοί έχουν εξελιχθεί με την πάροδο των χρόνων. Το αποτέλεσμα αυτής της εξέλιξης είναι η δημιουργία τριών διαφορετικών κατηγοριών client-side μηχανισμών παρακολούθησης:

- Μηχανισμοί βάσει πλοήγησης (session tracking mechanisms)
- Μηχανισμοί βάσει αποθήκευσης (storage-based tracking mechanisms)
- Μηχανισμοί βάσει προσωρινής αποθήκευσης (cache-based tracking mechanisms)

Βασική διαφορά των μηχανισμών αυτών, εκτός βέβαια από το χώρο αποθήκευσης των δεδομένων παρακολούθησης, αποτελεί και η μονιμότητα (persistence) που παρουσιάζουν τα στοιχεία παρακολούθησης. Μια γενική εικόνα της διαφοράς αυτής παρουσιάζεται στον πίνακα

1, στον οποίο εκτός από τη μονιμότητα των στοιχείων εμφανίζεται και ο τρόπος τερματισμού τους (elimination).

Πίνακας 1: Μονιμότητα και τερματισμός στοιχείων παρακολούθησης client-side μηχανισμών

Μηχανισμός	Μονιμότητα	Τερματισμός
Πλοήγησης (Session)	Ενεργή κατά την περιήγηση	Κλείσιμο περιήγησης
Αποθήκευσης (Storage)	Απροσδιόριστη	Τερματισμός των περιεχομένων της μνήμης cache και του προγράμματος περιήγησης
Προσωρινής Αποθήκευσης (Cache)	Εξαρτάται από χρονικές παραμέτρους	Τερματισμός της μνήμης cache

Όπως θα γίνει κατανοητό από την εξέταση των διαφόρων μηχανισμών παρακολούθησης οι τρόποι τερματισμού των δεδομένων στα προγράμματα πλοήγησης δεν είναι τόσο αποτελεσματικοί όσο θα περίμενε κανείς.

4.2 Μηχανισμοί παρακολούθησης βάσει πλοήγησης

Το W3C Document Object Model (DOM) είναι μια διεπαφή ανάμεσα σε πλατφόρμες με σκοπό την πρόσβαση και την αλληλεπίδραση με το περιεχόμενο, την δομή και το στυλ των Web Documents (δηλαδή HTML, XHTML και XML). Είναι ανεξάρτητο από τις γλώσσες προγραμματισμού και οργανώνει όλα τα αντικείμενα σε μια δομή δέντρου. Κάθε αντικείμενο (πχ. ένα παράθυρο) έχει έναν αριθμό ιδιοτήτων (πχ. όνομα) [6].

Οι μηχανισμοί παρακολούθησης που βασίζονται αποκλειστικά στην πλοήγηση (session tracking mechanisms) χρησιμοποιούν αναγνωριστικά τα οποία αποθηκεύονται προσωρινά όσο ο χρήστης χρησιμοποιεί το πρόγραμμα περιήγησης και παραμένουν για όσο χρόνο διαρκεί η συνεδρία. Τα αναγνωριστικά αυτά είναι συνήθως στοιχεία που περιέχονται στις ιστοσελίδες ως κρυμμένα πεδία στις ιδιότητες του DOM της σελίδας ή σε φόρμες διαδικτυακής ταυτοποίησης που απλά ταυτοποιούν το χρήστη κατά τη διάρκεια μιας ενεργής συνεδρίας και δεν αποτελούν σημαντικές απειλές παραβίασης της ιδιωτικότητάς του [6]. Σε αντίθεση με τα cookies και άλλους μηχανισμούς παρακολούθησης, τα αναγνωριστικά αυτά δεν αποθηκεύονται και εξαφανίζονται όταν η συνεδρία ή η σελίδα που επισκέπτεται ο χρήστης τερματίζεται. Οι μηχανισμοί παρακολούθησης αυτοί θεωρούνται πια παρωχημένοι και γενικά δεν χρησιμοποιούνται πολύ, ιδιαίτερα όταν υπάρχει η δυνατότητα χρήσης cookies ή άλλων μηχανισμών αποθήκευσης που παρουσιάζουν μεγαλύτερη μονιμότητα.

Εξαίρεση ίσως αποτελεί η ιδιότητα window.name του DOM. Όλα τα σύγχρονα προγράμματα περιήγησης μπορούν να αποθηκεύσουν μια αρκετά μεγάλη ποσότητα δεδομένων (2-32 MB) μέσω της JavaScript χρησιμοποιώντας την ιδιότητα window.name του DOM. Τα δεδομένα αυτά χρησιμοποιούνται για παρακολούθηση από την ίδια την ιστοσελίδα αλλά και από τρίτα μέρη, καθώς η ιδιότητα window.name είναι ανθεκτική σε επαναφορτώσεις σελίδων και προσβάσιμη και από άλλα domains. Η τεχνική αυτή μπορεί να συνδυαστεί με αντικείμενα

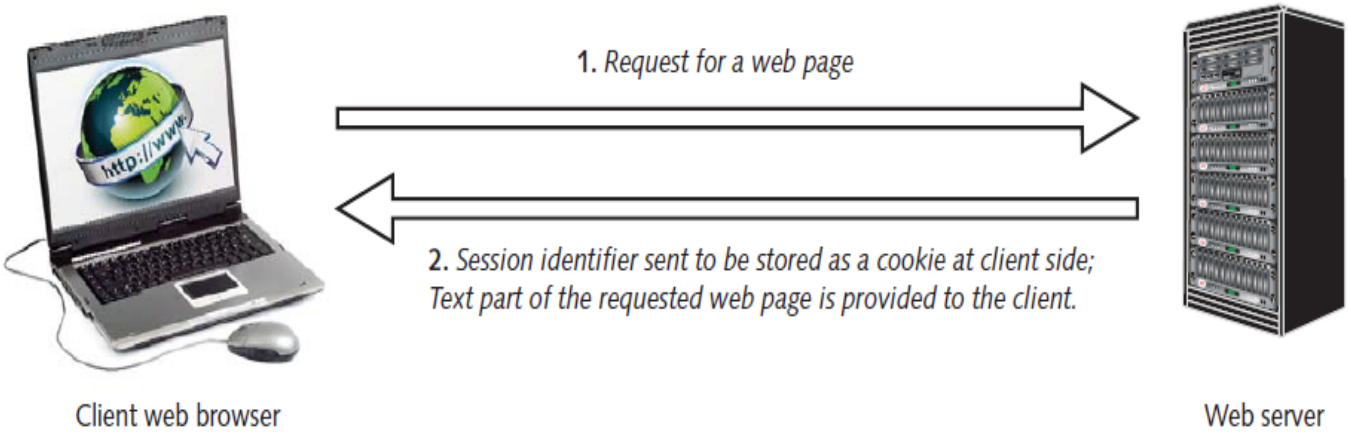
JSON/Javascript, όπως το JSON stringifier, για να αποθηκεύσει ολόκληρα σύνολα μεταβλητών συνεδρίας στον υπολογιστή-πελάτη [66].

Μειονέκτημα αυτού του μηχανισμού παρακολούθησης είναι ότι κάθε ξεχωριστό παράθυρο ή καρτέλα που ανοίγει, έχει αρχικά ένα άδειο window.name. Επιπλέον, η ιδιότητα window.name μπορεί να χρησιμοποιηθεί για την παρακολούθηση των χρηστών όταν επισκέπτονται διάφορες ιστοσελίδες, όπως αναφέρθηκε, καθιστώντας το μηχανισμό αυτό απειλή της ιδιωτικότητας στο διαδίκτυο. Παρ' όλα αυτά, ο συγκεκριμένος μηχανισμός παρακολούθησης είναι πιο ασφαλής από ότι τα cookies λόγω της μη συμμετοχής του διακομιστή, γεγονός που τον καθιστά λιγότερο ευάλωτο σε επιθέσεις cookie sniffing. Ωστόσο, εάν δεν ληφθούν ειδικά μέτρα για την προστασία των δεδομένων, είναι ευάλωτος σε άλλες επιθέσεις, επειδή τα δεδομένα είναι διαθέσιμα σε ιστοσελίδες που ανοίγουν στο ίδιο παράθυρο ή καρτέλα [67].

4.3 Τα cookies ως μηχανισμός παρακολούθησης

Τα cookies είναι αρχεία κειμένου (έως 4 KB) που αποθηκεύονται τοπικά στον υπολογιστή του χρήστη, στον web browser ή σε υποφακέλους προγραμμάτων [68]. Ένα HTTP cookie είναι ένα μικρό κομμάτι δεδομένων που στέλνεται από μια ιστοσελίδα και αποθηκεύεται στον web browser του χρήστη κατά την διάρκεια της περιήγησης. Αυτά τα μοναδικά αρχεία κειμένου δημιουργούνται όταν το πρόγραμμα περιήγησης του χρήστη φορτώνει μια ιστοσελίδα για πρώτη φορά [67].

Στην εικόνα 5 παρουσιάζεται ο τρόπος λειτουργίας των cookies. Όταν το πρόγραμμα περιήγησης επικοινωνεί με το διακομιστή και ζητά την εμφάνιση μιας συγκεκριμένης ιστοσελίδας, αυτός αναζητεί το μοναδικό κωδικό αναγνώρισης που αποθηκεύεται στο αρχείο του εκάστοτε cookie. Ο κωδικός αναγνώρισης είναι ένας μοναδικός κωδικός που αποδίδεται σε συγκεκριμένο cookie το οποίο αποθηκεύεται στη συσκευή του χρήστη ώστε να καταστεί διακριτό από άλλα cookies. Σε περίπτωση που ο χρήστης έχει ήδη επισκεφθεί την ιστοσελίδα, οι πληροφορίες που έχουν αποθηκευτεί στο cookie κοινοποιούνται στη σελίδα αυτή, η οποία με τον τρόπο αυτό διατηρεί ιστορικό επισκέψεων. Λόγω της ύπαρξης του κωδικού αναγνώρισης, τα cookies θεωρούνται μηχανισμοί εντοπισμού, ικανοί να αποκαλύπτουν συγκεκριμένα χαρακτηριστικά του χρήστη [6].



Εικόνα 5: Αναπαράσταση λειτουργίας των cookies (Πηγή: R. Tirtea και συν.)[5]

Οι πληροφορίες που συνήθως περιέχουν τα cookies είναι οι εξής [69]:

- Όνομα (Name): Το όνομα του cookie
- Τιμή (Value): Συνήθως ένας τυχαία δημιουργούμενος αριθμός
- Ημερομηνία λήξης (Expiration Date): Αποτελεί την ημερομηνία λήξης του cookie. Ημερομηνία λήξης έχουν μόνο τα μόνιμα cookies (persistent cookie) τα οποία διαγράφονται χειροκίνητα από τον χρήστη ή από τον περιηγητή με βάση την ημερομηνία αυτή. Αντίθετα τα cookies συνεδρίας (session cookies) διαγράφονται μετά το πέρας του χρόνου επίσκεψης του χρήστη στη σελίδα
- Domain Name: Αφορά το όνομα του domain που δημιούργησε το cookie και λαμβάνει τις πληροφορίες από αυτό
- Path: Χρησιμοποιείται για να καθορίσει πότε ένα cookie αποστέλλεται πίσω στο διακομιστή

Τα cookies μπορεί να εξυπηρετούν πολλούς σκοπούς. Ο βασικός σκοπός τους είναι η διευκόλυνση της λειτουργικότητας μιας ιστοσελίδας. Ο δεύτερος σκοπός των cookies είναι η προσαρμοσμένη απεικόνιση της ιστοσελίδας στον εκάστοτε χρήστη. Για παράδειγμα το Yahoo μπορεί να παρουσιάσει υπηρεσίες όπως το Ημερολόγιο (Calendar), η Πρόγνωση Καιρού (Weather Forecast) κλπ., που βασίζονται στα cookies. Οι σκοποί αυτοί καθιστούν τα cookies χρήσιμα τόσο για τους χρήστες του διαδικτύου όσο και για τις οντότητες που τα χειρίζονται.

Τα cookies χρησιμοποιούνται επίσης και στην περίπτωση των web analytics. Στην περίπτωση αυτή αποτελούν εργαλεία στατιστικής για να εκτιμήσουν τον αριθμό των επισκεπτών μιας σελίδας, να ανιχνεύσουν τις δημοφιλέστερες λέξεις-κλειδιά, κλπ. [70]. Τα cookies μπορεί να δημιουργηθούν από τον πάροχο της ιστοσελίδας (first party analytics) ή άλλη οντότητα με την οποία ο χρήστης δεν έχει καμία αλληλεπίδραση (third party analytics). Ενώ τα first party analytics θεωρείται ότι περιορίζονται σε καθαρά ανώνυμους στατιστικούς σκοπούς που χρησιμοποιούνται από τον πάροχο της ιστοσελίδας που δημιούργησε το cookie, κάτι τέτοιο δεν ισχύει στην περίπτωση των analytics τρίτων [71]. Τα cookies των analytics τρίτων αποτελούν χρήσιμο εργαλείο διαδικτυακών συστημάτων διαφήμισης για τη συλλογή δεδομένων

[5] https://www.enisa.europa.eu/activities/identity-and-trust/library/pp/cookies

που αφορούν τους χρήστες-πελάτες, όπως για παράδειγμα ποιος πελάτης έκανε κλικ σε ποιες διαφημίσεις, με σκοπό την αποτελεσματικότερη προσαρμογή των διαφημίσεών τους [72].

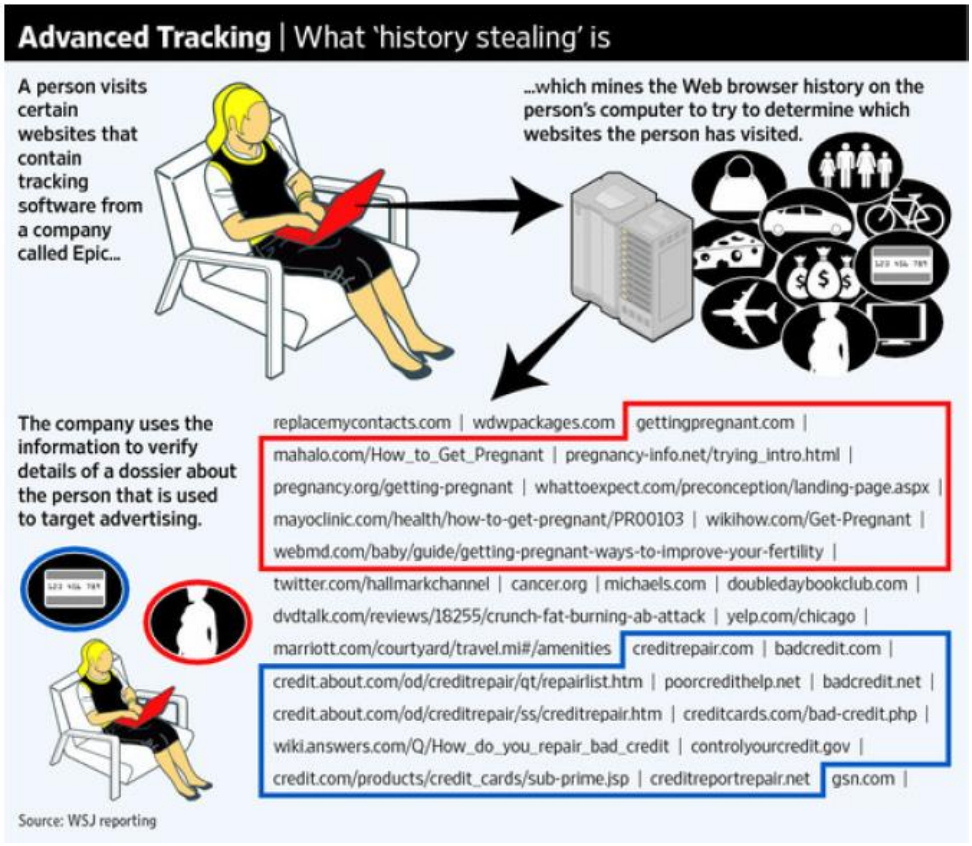
Όσον αφορά τη δημιουργία προφίλ των χρηστών, τα first party cookies παρακολούθησης επιτρέπουν στους παρόχους των διαδικτυακών συστημάτων διαφήμισης να αναγνωρίζουν ένα χρήστη που έχει ήδη επισκεφθεί μια ιστοσελίδα και να δημιουργούν προφίλ με βάση την επαναλαμβανόμενη επισκεψιμότητα και τα δεδομένα που συλλέγονται κατά τη διάρκεια αυτών των επισκέψεων. Στην περίπτωση των cookies τρίτων, τα δεδομένα που συλλέγονται δεν αφορούν αποκλειστικά τη σελίδα που επισκέφθηκε ο χρήστης αλλά τη συνολική του συμπεριφορά σε διαφορετικές ιστοσελίδες.

Εκτός από τα παραδοσιακά HTTP cookies, για σκοπούς παρακολούθησης χρησιμοποιούνται επίσης και τα flash cookies [72]. Τα flash cookies, γνωστά και ως Local Shared Objects (LSO), χρησιμοποιούνται και διαχειρίζονται από το Adobe Flash, ένα δημοφιλές plug-in πρόγραμμα περιήγησης που χρησιμοποιείται για κινούμενο, διαδραστικό, διαδικτυακό περιεχόμενο (animated interactive web content). Τα flash cookies έχουν την δυνατότητα να αποθηκεύσουν μέχρι και 100 KB δεδομένων, να παρέχουν τόσο προσωπικές όσο και τεχνικές πληροφορίες και δεν έχουν ημερομηνία λήξης. Αυτό που κάνει τα LSO χρήσιμο εργαλείο για την παρακολούθηση είναι ότι επιτρέπουν την ανταλλαγή δεδομένων μεταξύ όλων των περιηγητών του ίδιου υπολογιστή που χρησιμοποιούν το plug-in. Επιπλέον, το γεγονός ότι τα cookies αυτά αποθηκεύονται έξω από το πρόγραμμα περιήγησης καθιστά πιο δύσκολο για το χρήστη να τα ελέγχει ή να τα διαγράψει [5]. Τα flash cookies χρησιμοποιούνται ακόμη για την αναδημιουργία άλλων cookies, τα οποία ο χρήστης έχει ήδη διαγράψει [73]. Ερευνητές έχουν κατά καιρούς διατυπώσει ότι πολλές ιστοσελίδες χρησιμοποιούν flash cookies με κωδικούς, που έχουν ως αποτέλεσμα την αποδέσμευση των cookies του Internet Explorer που δεσμεύονται από προεπιλογή [74].

Τα HTTP και flash cookies αποτελούν παραδείγματα μηχανισμών παρακολούθησης που βασίζονται στην αποθήκευση (storage-based tracking mechanisms), καθώς όπως γίνεται κατανοητό από όσα έχουν περιγραφεί στην ενότητα αυτή, η λειτουργία τους εξαρτάται από τη ρητή αποθήκευση δεδομένων στους υπολογιστές των χρηστών. Οι μέθοδοι παρακολούθησης που βασίζονται στην αποθήκευση θεωρούνται ίσως οι πιο διαδεδομένοι και πιο συχνά χρησιμοποιούμενοι [6].

Σύμφωνα με την έρευνα Web Privacy Census του 2015, οι χρήστες που επισκέπτονται τις 100 πιο δημοφιλείς ιστοσελίδες συλλέγουν πάνω από 6000 HTTP cookies κατά την επίσκεψή τους στις σελίδες αυτές, εκ των οποίων το 5% είναι flash cookies και το 83% cookies τρίτων [75].

Μια τελευταία κατηγορία cookies αποτελούν τα evercookies (zombie cookies ή supercookies όπως αλλιώς είναι γνωστά), τα οποία θεωρούνται ένα ακόμα πιο μόνιμο είδος cookies [5]. Μεγάλο τους πλεονέκτημα αποτελεί η δυνατότητα χρήσης πολλών χώρων αποθήκευσης (Internet Explorer userData storage, HTML5 Session Storage, HTML5 Local Storage, HTML5 Global Storage, κλπ.) με σκοπό την αναδόμησή τους μετά τη διαγραφή ή ακόμα και την αναπαραγωγή τους σε άλλα προγράμματα περιήγησης του ίδιου υπολογιστή [6]. Μπορούν να προσδιορίσουν ένα χρήστη, ακόμη και όταν όλα τα άλλα είδη cookies (συμπεριλαμβανομένων των flash cookies) έχουν διαγραφεί. Στην εικόνα 6 δίνεται μια γραφική απεικόνιση παραδείγματος προηγμένης παρακολούθησης με χρήση των evercookies.



Εικόνα 6: Μηχανισμός προηγμένης παρακολούθησης με χρήση evercookies (Πηγή: Wall Street Journal)[6]

Στον πίνακα 2 δίνεται μια σύνοψη των διαφόρων ειδών cookies που χρησιμοποιούνται ως μηχανισμοί παρακολούθησης, των τεχνολογιών που χρησιμοποιούν καθώς και τον στόχο χρήσης τους.

Πίνακας 2: Σύνοψη ειδών cookies ως μηχανισμοί web tracking

Μηχανισμός Παρακολούθησης	Τεχνολογία	Στόχος
HTTP cookies	Κεφαλίδες HTTP, JavaScript	Αναγνωριστικό προγράμματος περιήγησης
Flash cookies	Flash	Αναγνωριστικό λειτουργικού συστήματος
Evercookies (supercookies)	Συνεδρία διαδικτυακού διακομιστή, κεφαλίδες HTTP, HTML5, JavaScript, Flash, Silverlight, Java	Αναγνωριστικά προγράμματος περιήγησης & λειτουργικού συστήματος

[6] http://www.wsj.com/articles/SB10001424053111903480904576508382675931492

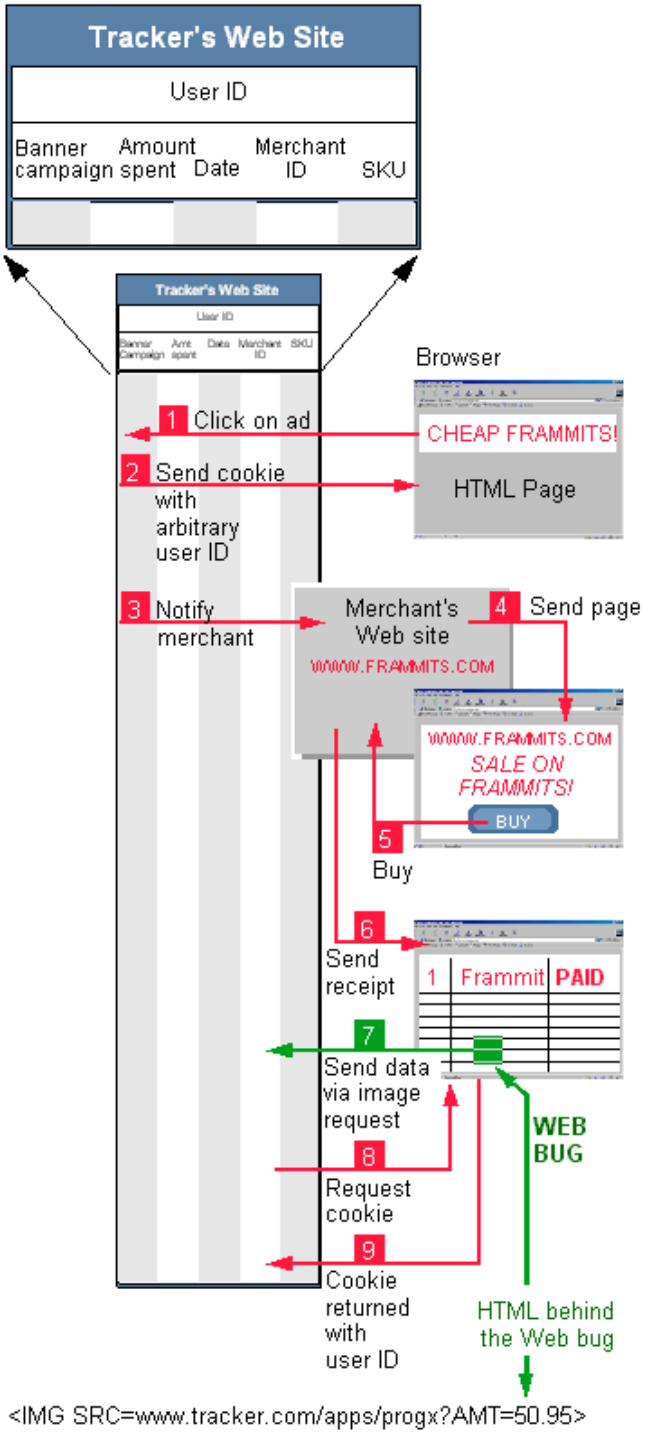
4.4 Web beacons

Τα web beacons (γνωστά και ως web bugs, διαφανή αρχεία GIF, pixels ή action tags), είναι αντικείμενα ενσωματωμένα σε ιστοσελίδες ή μηνύματα ηλεκτρονικού ταχυδρομείου [76]. Τις περισσότερες φορές τα web beacons είναι γραφικές εικόνες 1x1 και περιλαμβάνονται σε κομμάτι κώδικα HTML που είναι συνήθως αόρατος στο χρήστη, λόγω του μεγέθους του. Χρησιμοποιούνται για την παρακολούθηση της συσκευής την οποία χρησιμοποιεί ο χρήστης για να προσπελάσει μια ιστοσελίδα ή να ανοίξει το ηλεκτρονικό ταχυδρομείο του καθώς και τον χρόνο αυτής της προσπέλασης ή ανάγνωσης. Τα web beacons φορτώνονται την ώρα της προσπέλασης ή της ανάγνωσης αυτής. Για κάθε εικόνα αποστέλλεται και ένα αίτημα, ενώ η ιστοσελίδα που παραλαμβάνει το αίτημα είναι σε θέση να γνωρίζει την προέλευσή του.

Υπάρχουν πλήθος τρόπων με τους οποίους τα web beacons χρησιμοποιούνται ως μηχανισμοί παρακολούθησης τον ιστό. Στην εικόνα 7 παρουσιάζεται ένα παράδειγμα τέτοιας χρήσης. Στο παράδειγμα αυτό παρουσιάζεται μια ιστοσελίδα παρακολούθησης τρίτου μέρους για τον καθορισμό της απήχησης που είχε ένα banner διαφημιστικής καμπάνιας όσον αφορά τις πωλήσεις των διαφημιζόμενων προϊόντων. Σε σενάρια όπως αυτό, οι χρήστες παραμένουν ανώνυμοι, ακόμη και αν αποκαλύπτονται οι αγοραστικές τους συνήθειες.

Το 2010, οι J. Angwin και συν. δημοσίευσαν στην Wall Street Journal ότι εξέτασαν τις 50 πιο δημοφιλείς ιστοσελίδες της χρονιάς εκείνης και ανακάλυψαν 3.180 «αρχεία παρακολούθησης» (HTTP cookies, flash cookies και web beacons) [37].

Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο, τα web beacons χρησιμοποιούνται εκτενώς από τις ιστοσελίδες κοινωνικής δικτύωσης δεδομένου του ενδιαφέροντός τους να γνωρίζουν σε ποιες ιστοσελίδες οι χρήστες τους δημιουργούν προφίλ με σκοπό την παροχή εξατομικευμένων διαφημίσεων. Καλό παράδειγμα τέτοιου είδους web beacons αποτελούν τα κουμπιά «Like» του Facebook, «tweet» του Twitter και «+1» της Google. Σχεδόν κάθε ιστοσελίδα σήμερα παρέχει αυτά τα κουμπιά κοινωνικής δικτύωσης. Οι δυνατότητες αυτών των κουμπιών είναι μεγάλες καθώς ακόμα και στην περίπτωση που δεν πατηθούν, οι σελίδες κοινωνικής δικτύωσης είναι σε θέση να γνωρίζουν τις σελίδες που επισκέπτονται οι χρήστες τους. Όταν το πρόγραμμα περιήγησης ενός χρήστη επισκέπτεται μια ιστοσελίδα, στέλνει αίτημα τέτοιας εικόνας στην εκάστοτε σελίδα κοινωνικής δικτύωσης ενώ παράλληλα στέλνει και ένα cookie αναγνώρισης.



Εικόνα 7: Σενάριο χρήσης web beacons ως μηχανισμό third-party web tracking (Πηγή: The Free Dictionary)[7]

[7] http://encyclopedia2.thefreedictionary.com/Web+beacons

4.5 HTML5 Web Storage

Το HTML5 web storage αποτελεί έναν μηχανισμό web tracking που βασίζεται στην αποθήκευση (storage-based tracking mechanism). Το web storage API χρησιμοποιείται για την αποθήκευση δεδομένων στον web browser. Η μέθοδος αυτή υποστηρίζει μόνιμη αποθήκευση δεδομένων, όμοια με αυτή των cookies, αλλά με σημαντικά βελτιωμένη χωρητικότητα [77] και χωρίς την απαίτηση αποθήκευσης πληροφορίας στην κεφαλίδα των HTTP αιτημάτων [78]. Υπάρχουν δύο κύριοι τύποι web storage: η τοπική αποθήκευση (local storage) και η αποθήκευση συνεδρίας (session storage), που συμπεριφέρονται παρόμοια με τα μόνιμα cookies και τα cookies συνεδρίας, αντίστοιχα.

Η HTML5 είναι μια γλώσσα σήμανσης (markup language), η πέμπτη έκδοση του προτύπου W3C, που επιτρέπει την εύκολη ενσωμάτωση εικόνων και βίντεο σε ιστοσελίδες [70]. Η HTML5 web storage δίνει τη δυνατότητα αποθήκευσης και ανάκτησης μεγάλης ποσότητας δεδομένων (5MB) σε τοπικό επίπεδο, στα πλαίσια ενός διακομιστή, χωρίς την ανάγκη χρήσης κάποιου cookie [79]. Οι αρχικές εκδόσεις της HTML5 παρείχαν μηχανισμούς παγκόσμιας αποθήκευσης (global storage) και δυνατότητες αποθήκευσης δεδομένων στις ιστοσελίδες, ένα σχέδιο όμως που δεν εφαρμόστηκε από κανένα πρόγραμμα περιήγησης λόγω παραβίασης της πολιτικής SOP (same-origin policy) [80].

Σε αντίθεση με τα cookies, στα οποία πρόσβαση μπορεί να έχει τόσο η πλευρά του διακομιστή (server side) όσο και αυτή του πελάτη (client side), η HTML5 web storage εμπίπτει αποκλειστικά στην επίβλεψη του client-side scripting. Η τεχνολογία client-side scripting γενικά αναφέρεται στην κλάση των διαδικτυακών προγραμμάτων που εκτελούνται στην πλευρά του πελάτη από το διακομιστή και όχι στην πλευρά του διακομιστή [81]. Τα client-side scripts, που συχνά ενσωματώνονται σε έγγραφα HTML, επιτρέπουν την τελειοποίηση της επιθυμητής αλληλεπίδρασης πελάτη – διακομιστή, με αποτέλεσμα τη δυνατότητα αυτόματης μετάδοσης στο διακομιστή της διαδικτυακής αποθήκευσης δεδομένων σε κάθε HTTP αίτημα και την απευθείας εγγραφή των δεδομένων από τον web server στο χώρο διαδικτυακής αποθήκευσης.

Η τοπική αποθήκευση HTML5 (HTML5 local storage), η οποία υπακούει στην ίδια πολιτική, παρέχει μια ακόμη δυνατότητα παρακολούθησης των χρηστών. Η αποθήκευση των ζευγών κλειδιού-τιμής (key-value pairs) δεν απαιτεί κανένα plug-in. Τα ζεύγη αυτά αποθηκεύονται μόνιμα (δεν υπάρχει καμία ημερομηνία λήξης) και παραμένουν στο χώρο αποθήκευσης μέχρι να αφαιρεθούν από την ιστοσελίδα ή από το χρήστη. Το γεγονός ότι τα δεδομένα αυτά μπορεί να φτάνουν τα 5 MB δίνει ένα σημαντικό πλεονέκτημα στον μηχανισμό αυτό σε σχέση με τα HTTP και τα flash cookies. Το περιεχόμενο της τοπικής αποθήκευσης μπορεί να μοιραστεί μεταξύ των διαφόρων παραθύρων του προγράμματος περιήγησης [80] και αδειάζει αυτόματα όταν διαγράφονται τα cookies.

Σύμφωνα με την έρευνα Web Privacy Census του 2015, το 76% των δημοφιλέστερων ιστοσελίδων χρησιμοποιεί εκτός από flash cookies και μηχανισμό τοπικής αποθήκευσης HTML5 για την αποθήκευση 877 ζευγών κλειδιού-τιμής [75]. Σε αρκετές περιπτώσεις, οι τιμές που αποθηκεύτηκαν σε μορφή τοπικής αποθήκευσης HTML5 ταίριαζαν με τα HTTP cookies των αντίστοιχων ιστοσελίδων αλλά και ιστοσελίδων τρίτων.

Η αποθήκευση συνεδρίας HTML5 (HTML5 session storage) έχει παρόμοια λειτουργία με την τοπική αποθήκευση HTML5 διατηρώντας την πολιτική ιδίας προέλευσης (SOP) και αποθηκεύοντας αντικείμενα μεγάλου μεγέθους έως 5 MB. Η διαφορά τους έγκειται στο γεγονός

ότι στην αποθήκευση συνεδρίας τα αντικείμενα παραμένουν διαθέσιμα μόνο για το τρέχον παράθυρο του προγράμματος περιήγησης και διαγράφονται όταν το παράθυρο κλείσει [80].

Στον πίνακα 3 παρουσιάζονται οι κύριες διαφορές των βασικών χαρακτηριστικών του μηχανισμού HTML5 web storage με τα cookies.

Πίνακας 3: Βασικά χαρακτηριστικά HTTP Cookies, Flash Cookies και HTML5 Storage

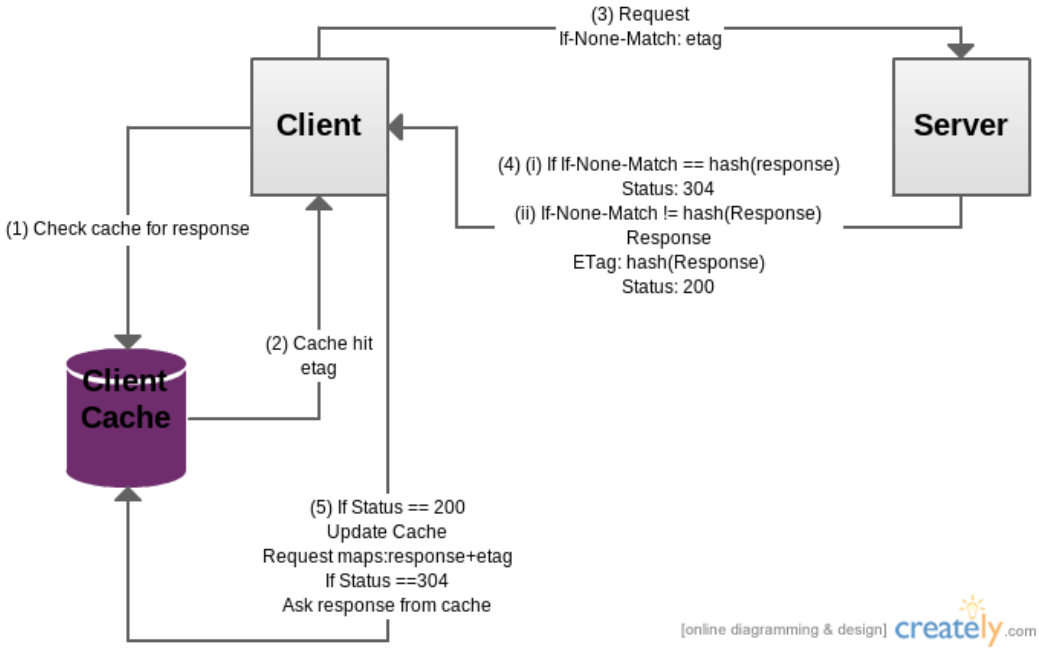
	HTTP Cookies	Flash cookies	HTML5 storage
Χωρητικότητα	4KB	100KB (από προεπιλογή)	5MB (από προεπιλογή)
Λήξη	Συνεδρία (από προεπιλογή)	Μόνιμα (από προεπιλογή)	Μόνιμα (από προεπιλογή)
Χώρος αποθήκευσης	Αρχείο SQL (Firefox)	Εκτός προγράμματος περιήγησης	Αρχείο SQL (Firefox)
Πρόσβαση	Μόνο από το πρόγραμμα περιήγησης	Από διάφορα προγράμματα περιήγησης στον ίδιο υπολογιστή	Μόνο από το πρόγραμμα περιήγησης

Άλλες εφαρμογές API που χρησιμοποιούνται ως μηχανισμοί web tracking μέσω αποθήκευσης αρχείων και διαχείρισης των βάσεων δεδομένων στον υπολογιστή-πελάτη είναι οι File και IndexedDB, αντίστοιχα [6]. Ο τερματισμός της λειτουργίας των μηχανισμών αυτών, στις περισσότερες περιπτώσεις, απαιτεί χειροκίνητη παρέμβαση του χρήστη.

4.6 Ετικέτες οντότητας

Οι ετικέτες οντότητας (Entity Tags ή Etags) είναι μια τυποποιημένη μέθοδος ελέγχου της προσωρινής αποθήκευσης περιεχομένου (cache) που παρέχει το HTTP. Προορίζονται να χρησιμοποιηθούν ως υπογραφές που επικυρώνουν το αν η μνήμη cache του πελάτη (client) ενημερώνεται, αλλά και για την αποθήκευση και ανάκτηση αυθαίρετων (arbitrary) πληροφοριών με σκοπό την παρακολούθηση. Για το λόγο αυτό αποτελούν παράδειγμα μηχανισμών παρακολούθησης που βασίζονται στη μνήμη cache (cache-based tracking mechanisms).

Σε μια κανονική HTTP ροή δεδομένων που περιλαμβάνει το περιεχόμενο προσωρινής αποθήκευσης, σε περίπτωση που ο πελάτης στείλει αίτημα για πόρους που έχουν ρυθμιστεί ώστε να χρησιμοποιούν τα περιεχόμενα της μνήμης cache, ο διαδικτυακός διακομιστής θέτει μια ETag στην κεφαλίδα της HTTP απόκρισης, την τιμή της οποίας ο πελάτης αποθηκεύει μαζί με το επιστρεφόμενο περιεχόμενο στη δική του μνήμη cache. Την επόμενη φορά που ο πελάτης θα στείλει αίτημα στην ίδια πηγή, η τιμή του ETag θα είναι «If-None-Match». Αν η τρέχουσα ETag ταιριάζει με αυτή που αποστέλλεται από τον πελάτη, επιστρέφεται ο κωδικός κατάστασης 304 (Μη Τροποποιημένο – Not Modified), διαφορετικά το περιεχόμενο αποστέλλεται με νέα ETag (Εικ. 8) [82].



Εικόνα 8: Σενάριο λειτουργίας ετικέτας οντότητας (Πηγή: Dweeb's Lair)[8]

Σύμφωνα με το πρωτόκολλο HTTP, τα πεδία που περιέχει μια κεφαλίδα HTTP είναι τα εξής [82]: Last-Modified, ETag, Cache-Control και Expires. Το πεδίο Last-Modified εκτός από μια έγκυρη ημερομηνία (όπως ορίζει το πρωτόκολλο HTTP), δέχεται και οποιαδήποτε τυχαία συμβολοσειρά (string) [83]. Σε αντίθεση με το πεδίο Last-Modified, το πραγματικό περιεχόμενο και η μορφή μιας ETag είναι εντελώς αυθαίρετα. Το πεδίο ETag μπορεί να φτάσει τα 81864 bit και επομένως μπορεί να αποδειχθεί πολύ χρήσιμο για την αποθήκευση των αναγνωριστικών των χρηστών αλλά και για την αναγέννηση των cookies, στοιχεία τα οποία εκμεταλλεύονται οι εταιρείες παρακολούθησης.

Ο μηχανισμός παρακολούθησης με χρήση της ETag χρησιμοποιείται από διάφορες μεγάλες ιστοσελίδες, ακόμη και αν οι χρήστες έχουν διαγράψει τα cookies. Η πρώτη χρήση των ETag ως μηχανισμός παρακολούθησης χρηστών διαπιστώθηκε το 2011, όταν η σελίδα Hulu παρακολουθούσε τους χρήστες της μέσω αποθήκευσης πληροφοριών αναγνώρισης σε ETag και flash LSO [84]. Η αποθήκευση αυτή γινόταν με χρήση της εμπορικής υπηρεσίας KISSmetrics, η οποία βοηθούσε στην αναδημιουργία των HTTP και HTML5 cookies. Την ίδια υπηρεσία χρησιμοποιούσαν για παρόμοιους σκοπούς και άλλες λιγότερο γνωστές ιστοσελίδες. Η αποκάλυψη αυτού του μηχανισμού παρακολούθησης καταδεικνύει την ευκολία απόκρυψης και ανάκτησης πληροφοριών παρακολούθησης από τους περισσότερους χρήστες εκτός και αν αυτοί έχουν την ικανότητα να εξετάζουν σκόπιμα κάθε μέσο αποθήκευσης.

Αν και αυτός ο μηχανισμός θα πρέπει να λειτουργεί με κάθε HTTP συμβατό πρόγραμμα περιήγησης, υπάρχουν διαφορές στον τρόπο με τον οποίο οι διακομιστές επαναφορτώνουν στην πραγματικότητα το περιεχόμενο, κάτι που τελικά επηρεάζει το αν τα αιτήματα για πρόσβαση στο προσωρινά αποθηκευμένο περιεχόμενο αποστέλλονται στο διακομιστή. Κάθε πρόγραμμα περιήγησης εφαρμόζει τη διαδικασία προσωρινής αποθήκευσης διαφορετικά και ο τρόπος με τον οποίο ο κάθε χρήστης φορτώνει μια σελίδα έχει συχνά συνέπειες για το αν η μνήμη cache

[8] http://dweebslair.blogspot.gr/2014_03_01_archive.html

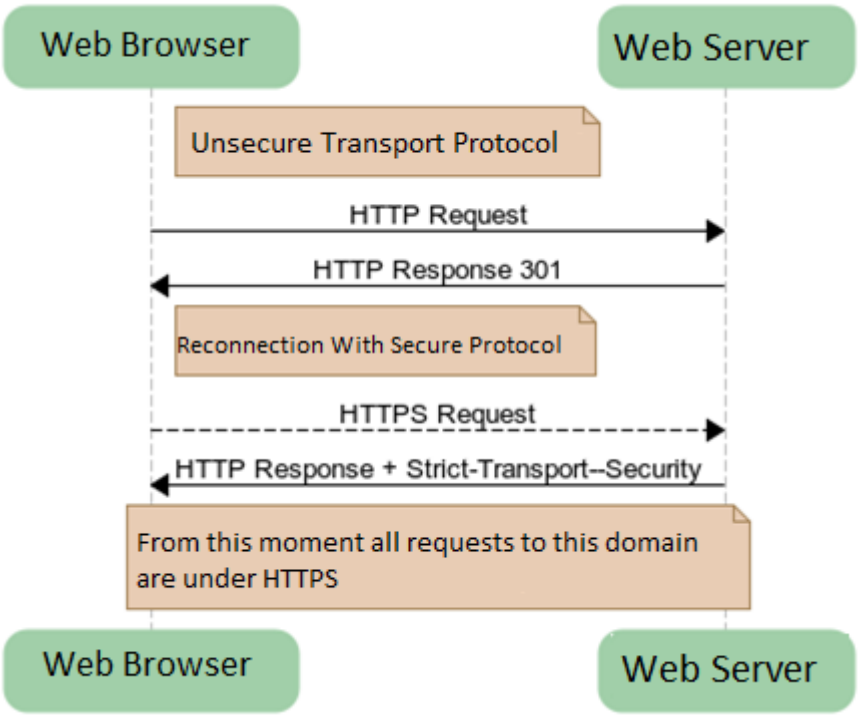
επικυρώνεται εκ νέου. Για παράδειγμα, εάν ένας χρήστης πατήσει απλώς το 'ENTER' στη γραμμή διευθύνσεων URL για να φορτώσει μια ιστοσελίδα χρησιμοποιώντας τον Chrome, το συγκεκριμένο πρόγραμμα περιήγησης θα φορτώσει στατιστικά πόρων από τη μνήμη cache χωρίς να τα επικυρώσει. Εάν ο χρήστης κάνει κλικ στο κουμπί Ανανέωση ή φορτώσει μια σελίδα χρησιμοποιώντας κάποιο hotkey, τότε ο Chrome θα επικυρώσει τη μνήμη cache. Από αυτό το παράδειγμα γίνεται κατανοητό ότι ακόμα και όταν το πρόγραμμα περιήγησης στείλει ένα HTTP αίτημα για δεδομένους πόρους, οι ETag εξακολουθούν να μην είναι απολύτως αξιόπιστες ακόμη και σε ένα πλήρως συμβατό πρόγραμμα περιήγησης. Το πρότυπο HTTP αναφέρει σχετικά ότι οι ETag μπορεί να χρησιμοποιηθούν για τη σύγκριση πόρων, αλλά οι πελάτες δεν είναι απαραίτητο να τις προσθέτουν ως πεδίο της κεφαλίδας HTTP [82].

Όταν η τιμή της ETag αποστέλλεται στο διακομιστή παρακολούθησης, ο τρόπος επεξεργασίας της είναι παρόμοιος ενός HTTP cookie. Σε αντίθεση με τα cookies, όμως, οι ETag είναι πιο δύσκολο να απομιμηθούν στους HTTP πελάτες και έτσι μπορεί να θεωρηθούν πιο αξιόπιστες. Η τιμή τους βέβαια μπορεί να αφαιρεθεί εύκολα με το σβήσιμο της μνήμης cache του προγράμματος περιήγησης, ενώ οι χρήστες θα μπορούσαν να εμποδίσουν τα ETag παρακολούθησης πλήρως, απενεργοποιώντας την προσωρινή αποθήκευση ή χρησιμοποιώντας έναν πελάτη που δεν υποστηρίζει τη διαδικασία αυτή. Λόγω της ευκολίας απομάκρυνσής τους επομένως, οι ETag παρουσιάζουν χαμηλότερη δυνατότητα μονιμότητας σε σχέση με άλλες τεχνικές και δεν γίνεται να αποτελέσουν μόνη μέθοδος παρακολούθησης με βάση τη μνήμη αποθήκευσης. Το «πρόβλημα» αυτό των ETag έχει λυθεί από τις εταιρείες παρακολούθησης μέσω της χρήσης και του πεδίου Last-Modified ως μηχανισμό παρακολούθησης εκμεταλλευόμενοι το γεγονός ότι το πεδίο αυτό δέχεται εκτός από έγκυρες ημερομηνίες και τυχαίες συμβολοσειρές, καθιστώντας το ως ένα μηχανισμό παρακολούθησης ισχυρότερο σε σύγκριση με το ETag, αφού μπορεί να λειτουργεί ακόμα και όταν μηχανισμοί όπως τα cookies ή το ETag έχουν εμποδιστεί από τους χρήστες [83].

4.7 Λειτουργικές μνήμες cache

Οι λειτουργικές μνήμες cache (operational cache) είναι συστατικά που χρησιμοποιούνται για την αποθήκευση πληροφοριών που σχετίζονται με λειτουργίες των περιηγητών. Τέτοιου είδους πληροφορίες μπορεί να περιλαμβάνουν μόνιμες ανακατευθύνσεις, πιστοποιήσεις ταυτότητας ή μια λίστα των domain που θα πρέπει να χρησιμοποιούνται από κοινού με το πρωτόκολλο αυστηρής ασφάλειας μεταφοράς HTTP (HTTP Strict Transport Security – HSTS) [6].

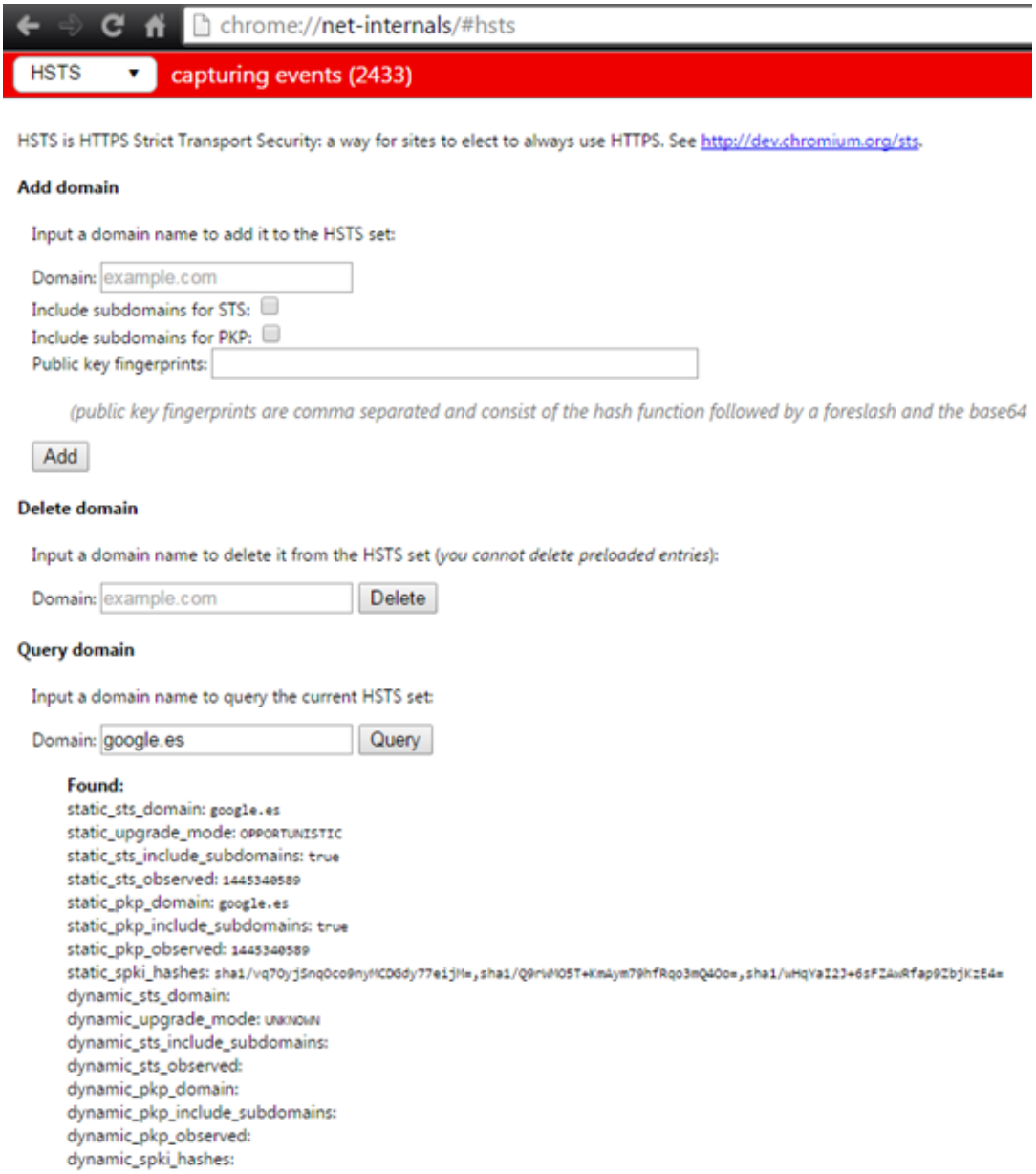
Το HSTS είναι ένας μηχανισμός ασφάλειας, στόχος του οποίου είναι να εξασφαλίζει ότι η αλληλεπίδραση μεταξύ διαδικτυακών διακομιστών και προγραμμάτων πλοήγησης πραγματοποιείται μόνο με ασφαλείς συνδέσεις HTTPS και ποτέ μέσω του πρωτοκόλλου HTTP που δεν παρέχει καμία ασφάλεια [85]. Για το σκοπό αυτό, το πρόγραμμα περιήγησης δημιουργεί μια βάση δεδομένων στην οποία αποθηκεύει μια λίστα των αρχικά καταχωρημένων ιστοσελίδων και στη συνέχεια, προσθέτει σταδιακά νέες ιστοσελίδες τις οποίες ο χρήστης επισκέπτεται μόνο μέσω αιτημάτων HTTPS. Τα αιτήματα HTTPS δημιουργούνται όταν στα αρχικά αιτήματα HTTP των προγραμμάτων περιήγησης επιστρέφει μια HTTP απόκριση μόνιμης αναδρομολόγησης 301, υποδεικνύοντας ότι ο συγκεκριμένος πόρος βρίσκεται μόνιμα διαθέσιμος σε άλλη URL, η οποία χρησιμοποιεί ασφαλή σύνδεση HTTPS (Εικ. 9).



Εικόνα 9: Επικοινωνία προγράμματος περιήγησης και διαδικτυακού διακομιστή μέσω HTTPS (Πηγή: INCIBE)[9]

Οι κεφαλίδες των αιτημάτων HTTPS περιέχουν το πεδίο ασφάλειας Strict Transport Security με την παράμετρο max-age, η οποία καθορίζει (σε δευτερόλεπτα) το χρονικό διάστημα για το οποίο ο χρήστης θα μπορεί να επισκέπτεται την εκάστοτε σελίδα μόνο με σύνδεση HTTPS. Με τον τρόπο αυτό, το πρόγραμμα περιήγησης δεν μπορεί να πραγματοποιήσει άλλη σύνδεση εκτός από HTTPS, για το χρονικό διάστημα που ορίζει η max-age, το οποίο μπορεί να φτάσει και τον ένα χρόνο (Strict-Transport-Security: max-age=31536000). Κάθε φορά που αποθηκεύεται μια νέα καταχώριση, παραμένει στη βάση δεδομένων μέχρι τη λήξη της και οποιαδήποτε εξάλειψη των cookies ή τερματισμός της μνήμης cache ή των προσωρινών αρχείων δεν αλλάζει το χρόνο αυτό. Τερματισμός του χρόνου αυτού μπορεί να πραγματοποιηθεί μόνο μέσω των προηγμένων επιλογών του προγράμματος περιήγησης με μη τετριμμένο τρόπο. Στην εικόνα 10 παρουσιάζεται παράδειγμα καταχωρήσεων και παραμετροποίησης του HTTPS στον Chrome.

[9] https://www.incibe.es/blogs/post/Security/SecurityBlog/Article_and_comments/web_tracking_en



Εικόνα 10: Παράδειγμα καταχωρήσεων και παραμετροποίησης του HTTPS στο Chrome (Πηγή: INCIBE)[10]

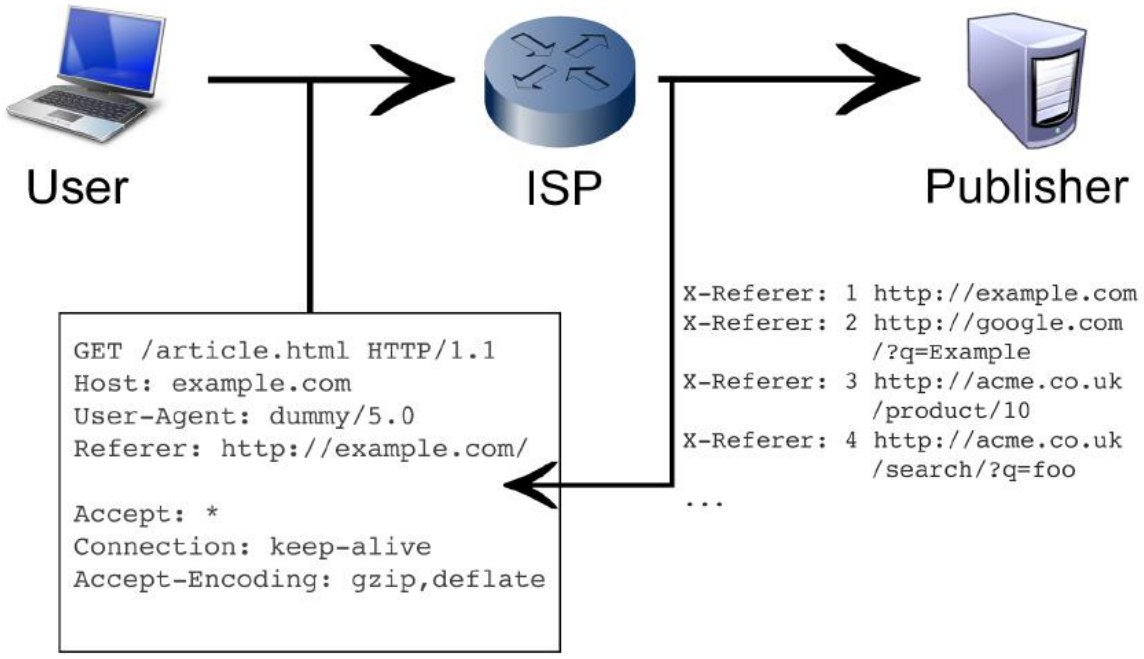
Με τη χρήση αυτού του μηχανισμού, είναι δυνατή η δημιουργία καταχωρήσεων HSTS στο πρόγραμμα περιήγησης του χρήστη και, επομένως, η δημιουργία ενός συνόλου αναγνωριστικών, το οποίο έχει ονομαστεί HSTS supercookies [86]. Επομένως, το HSTS, ανεξάρτητα από την αρχική αποστολή του, έχει γίνει ένα εργαλείο που μπορεί να χρησιμοποιηθεί για τη λήψη πληροφοριών (ακόμα και του ιστορικού περιήγησης), γεγονός που θέτει σε κίνδυνο την προστασία της ιδιωτικότητας του χρήστη [87].

[10] https://www.incibe.es/blogs/post/Security/SecurityBlog/Article_and_comments/web_tracking_en

4.8 Έγχυση κεφαλίδων HTTP

Η έγχυση κεφαλίδων HTTP (HTTP header injection) είναι μια γενική κατηγορία ευπάθειας της ασφάλειας των διαδικτυακών εφαρμογών και πραγματοποιείται μέσω της δυναμικής δημιουργίας κεφαλίδων HTTP με βάση τα στοιχεία του χρήστη [88]. Η έγχυση κεφαλίδων στις HTTP αποκρίσεις μπορεί να επιτρέψει τη διάσπαση της ίδιας της απόκρισης, την καθήλωση της συνεδρίας μέσω της επικεφαλίδας Set-Cookie, το Cross-site scripting (XSS) καθώς και κακόβουλες επιθέσεις ανακατεύθυνσης μέσω της κεφαλίδας θέσης.

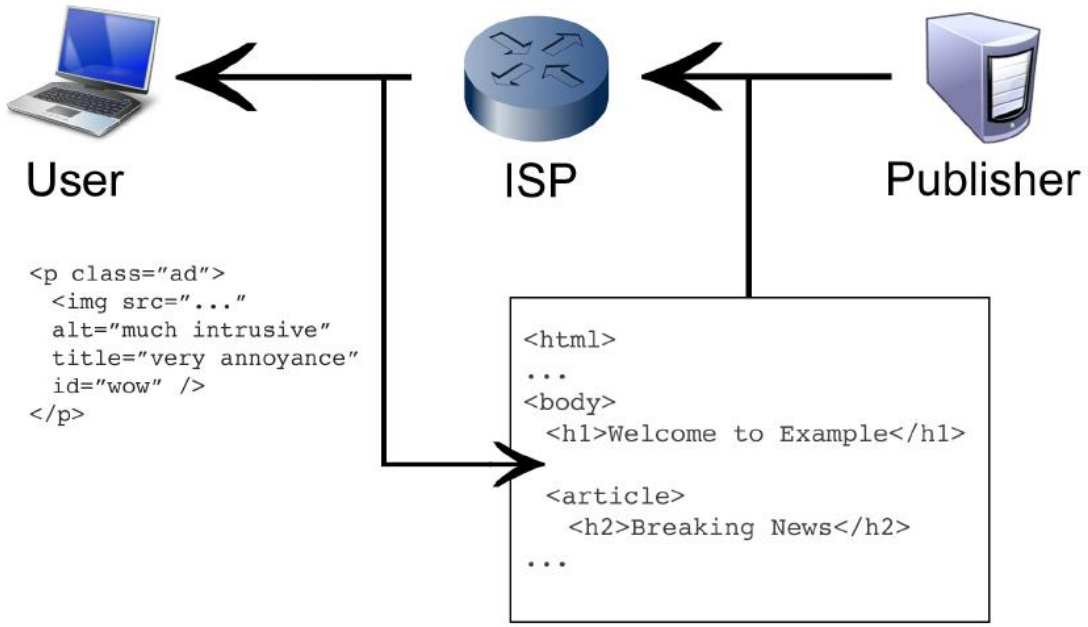
Η έγχυση κεφαλίδων HTTP μπορεί να διακριθεί σε δύο περιπτώσεις: την upstream και την downstream έγχυση. Βασική αρχή της upstream HTTP έγχυσης αποτελεί η παρέμβαση μιας υπηρεσίας διαδικτυακής παροχής (ISP) ή κάποιου ενδιάμεσου δικτύου σε ένα αίτημα HTTP με σκοπό την προσθήκη στοιχείων ταυτοποίησης των χρηστών πριν από τη διαβίβαση του αιτήματος στον προορισμό του. Τα στοιχεία ταυτοποίησης συνήθως περιλαμβάνουν μια σειρά από αναφορές (X-Referer) που περιέχουν το ιστορικό περιήγησης των χρηστών (Εικ. 11).



Εικόνα 11: Upstream HTTP έγχυση (Πηγή: Leo Le Taro)[11]

Η downstream HTTP έγχυση βασίζεται στην ίδια αρχή με την περίπτωση της upstream HTTP έγχυσης, με τη διαφορά ότι λαμβάνει χώρα κατά την αντίθετη κατεύθυνση. Μη κρυπτογραφημένες αποκρίσεις HTTP, όπως σελίδες HTML, εικόνες και JavaScript μπορεί να παραποιηθούν με σκοπό την προσθήκη ή αφαίρεση συγκεκριμένων στοιχείων (Εικ. 12).

[11] https://hal.inria.fr/hal-01167493/document



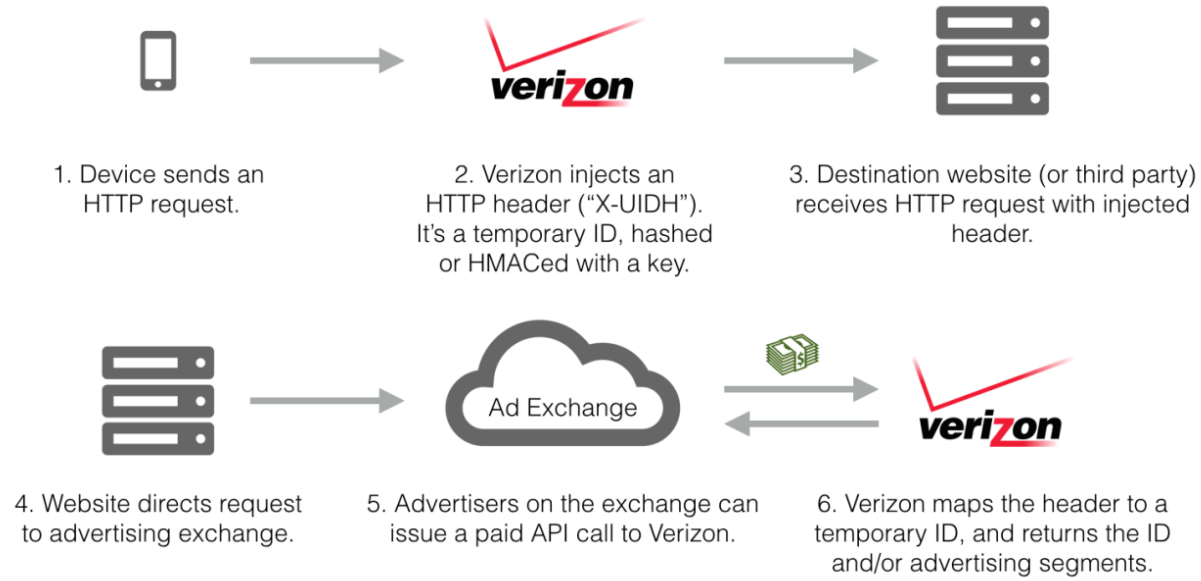
Εικόνα 12: Downstream HTTP έγχυση (Πηγή: Leo Le Taro)[12]

Αν και η διαδικασία της έγχυσης κεφαλίδων HTTP, όσον αφορά το web tracking, θεωρητικά χρησιμοποιείται για διαφημιστικούς λόγους, μια υπηρεσία διαδικτυακής παροχής (ISP) είναι δυνατό να εγχύσει, αντί για διαφημιστικό υλικό, αυθαίρετο κώδικα JavaScript που θα εκτελεστεί στο περιεχόμενο του προγράμματος περιήγησης του χρήστη [88]. Από τη στιγμή που το πρόγραμμα περιήγησης του χρήστη θεωρεί ότι το script προέρχεται από την πραγματική, «αληθινή» ιστοσελίδα, ο κώδικας αυτός μπορεί να χρησιμοποιηθεί για παρακολούθηση των αλλαγών του DOM χωρίς την ανάγκη εκτέλεσης μηχανών JavaScript στην υποδομή του ISP.

Η παρακολούθηση χρηστών μέσω του μηχανισμού της έγχυσης κεφαλίδων HTTP βγήκε για πρώτη φορά στην επιφάνεια το 2014, όταν δημοσιεύτηκε ότι η Verizon, ένας αμερικάνικος φορέας τηλεπικοινωνιών, προσθέτει κεφαλίδες που περιέχουν δείκτες της κυκλοφορίας HTTP των πελατών της, με στόχο τη δημιουργία αναγνωριστικών για καθένα από αυτούς (Εικ. 13) [89]. Ο μηχανισμός αυτός ονομάστηκε «permacookies».

Επιπλέον, και πιο πρόσφατα, η accessnow.org δημοσίευσε τη μελέτη «The Rise of Mobile Tracking Headers: How Telcos Around the World Are Threatening Your Privacy», η οποία αναλύει τον ίδιο μηχανισμό αναγνώρισης σε φορητές συσκευές που χρησιμοποιείται από τηλεπικοινωνιακούς φορείς [90].

[12] https://hal.inria.fr/hal-01167493/document



Εικόνα 13: Τρόπος έγχυσης κεφαλίδων HTTP της Verizon (Πηγή: Jonathan Mayer)[13]

[13] http://webpolicy.org/2014/10/24/how-verizons-advertising-header-works/

5 Μηχανισμοί Αποτύπωσης

Οι μηχανισμοί αποτύπωσης είναι μια κατηγορία μεθόδων που χρησιμοποιούν ένα ευρύ φάσμα τεχνολογιών με σκοπό την παρακολούθηση των χρηστών στον ιστό μέσω του καθορισμού ειδικών χαρακτηριστικών και μοναδικών αναγνωριστικών του περιβάλλοντος του υπολογιστή-πελάτη. Τα χαρακτηριστικά και αναγνωριστικά αυτά, όπως είναι το μοναδικό αναγνωριστικό μιας συσκευής, ενός λειτουργικού συστήματος και της έκδοσης του προγράμματος περιήγησης, είναι γνωστά και ως αποτυπώματα (fingerprints) [6]. Κάποια από τα αναγνωριστικά αυτά, που συνήθως αποτελούνται από μία ή περισσότερες τιμές, λαμβάνονται παθητικά κατά τη διάρκεια της περιήγησης του χρήστη σε μια ιστοσελίδα. Άλλα ζητούνται με ενεργό τρόπο από το διακομιστή και δεν γίνεται να ληφθούν διαφορετικά. Ανεξαρτήτως τρόπου λήψης, τα αναγνωριστικά αυτά παρέχουν πολύτιμες πληροφορίες σχετικά με τον πελάτη χωρίς την απαίτηση ύπαρξης αποθηκευμένων δεδομένων σε αυτόν.

Μέσω του αποτυπώματος οι χρήστες παρακολουθούνται σε πολλές διαφορετικές ιστοσελίδες που ανήκουν σε διαφορετικές οντότητες, γεγονός που δεν είναι δυνατό να γίνει άμεσα με τη χρήση των cookies. Στους μηχανισμούς αποτύπωσης δεν δημιουργούνται cookies ενώ ο χρήστης δεν χρειάζεται να συνδεθεί σε κάποια ιστοσελίδα για να ενεργοποιηθούν. Με τον τρόπο αυτό η παρακολούθηση αποτελεί μια διαφανή διαδικασία για το χρήστη και λειτουργεί ανεξάρτητα από το αν το πρόγραμμα περιήγησης αποδέχεται τα cookies ή όχι. Ως εκ τούτου, ένας μέσος χρήστης δεν έχει κανένα τρόπο για να αντιληφθεί την παρακολούθηση, την οποία βέβαια δεν μπορεί και να αποτρέψει. Φυσικά υπάρχουν κάποιοι τρόποι αντιμετώπισης όπως η απενεργοποίηση της υποστήριξης του JavaScript, της Java και του Flash, τρόποι που ωστόσο δεν εμποδίζουν την λειτουργία των μηχανισμών παθητικής αποτύπωσης.

Σε γενικές γραμμές οι μηχανισμοί αποτύπωσης χωρίζονται στις εξής κατηγορίες:

- Μηχανισμοί αποτύπωσης προγράμματος περιήγησης / συσκευής (Browser / device fingerprinting)
- Μηχανισμοί αποτύπωσης λειτουργικού συστήματος (OS fingerprinting)
- Μηχανισμοί αποτύπωσης θέσης (Location fingerprinting)

Σκοπός του κεφαλαίου αυτού είναι η παρουσίαση των μηχανισμών αποτύπωσης ως μέθοδοι παρακολούθησης των χρηστών στον ιστό.

5.1 Μηχανισμοί αποτύπωσης προγράμματος περιήγησης / συσκευής

Οι μηχανισμοί αποτύπωσης του προγράμματος περιήγησης (browser fingerprinting) αποτελούν διαδικασίες συλλογής ιδιοτήτων ή χαρακτηριστικών του προγράμματος περιήγησης του πελάτη για διάφορους λόγους, κυρίως, για την ταυτοποίηση του χρήστη. Παραδείγματα αυτών των ιδιοτήτων ή των χαρακτηριστικών αποτελούν οι διαστάσεις της οθόνης, η γραμματοσειρά του συστήματος, τα plug-in που υποστηρίζει το πρόγραμμα περιήγησης, η ζώνη ώρας, η έκδοση του προγράμματος περιήγησης, κ.λπ. Ο συνδυασμός αυτών των ιδιοτήτων ονομάζεται αποτύπωση του προγράμματος περιήγησης (browser fingerprint) και μπορεί να χρησιμεύσει ως μοναδικό αναγνωριστικό τόσο του προγράμματος περιήγησης όσο και της συσκευής στην οποία λειτουργεί. Ως εκ τούτου, μπορεί να χρησιμοποιηθεί και ως μέθοδος παρακολούθησης για την ταυτοποίηση του χρήστη που τα χρησιμοποιεί.

Τα αποτυπώματα γίνεται επίσης να συνδυαστούν με τις διευθύνσεις IP ως αναγεννητές των cookies. Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, ένας σημαντικός αριθμός δικτυακών τόπων κάνουν χρήση των evercookies (supercookies) ως τρόπο αναγέννησης των HTTP cookies τα οποία ο χρήστης έχει προηγουμένως διαγράψει [91]. Μια ακόμα χρήση των αποτυπωμάτων είναι ο συνδυασμός τους με τη διεύθυνση IP (κατά την απουσία των cookies), ως μέσο διάκρισης των χρηστών που χρησιμοποιούν μια ενιαία διεύθυνση IP, ακόμη και αν οι χρήστες αυτοί έχουν απενεργοποιήσει όλα τα cookies Ειδικότερα, ένα αποτύπωμα που δεν μεταφέρει περισσότερα από 15-20 bit παρεχόμενης πληροφορίας μπορεί σε όλες σχεδόν τις περιπτώσεις, με δεδομένη, βέβαια, τη διεύθυνση IP, να είναι αρκετό για να προσδιορίσει μοναδικά ένα συγκεκριμένο πρόγραμμα περιήγησης [92].

Οι περισσότεροι μηχανισμοί αποτύπωσης του προγράμματος περιήγησης υποθέτουν ότι μετά την ταύτισή του μέσω αυτών των μηχανισμών, το πρόγραμμα περιήγησης παραμένει σταθερό (δηλαδή, τα χαρακτηριστικά του δεν αλλάζουν με την πάροδο του χρόνου). Στην πραγματικότητα, ιδιότητες με τιμές που παρουσιάζουν μεγαλύτερη σταθερότητα ή αλλάζουν σταδιακά με την πάροδο του χρόνου (π.χ., η έκδοση του λειτουργικού συστήματος) χρησιμοποιούνται για την δημιουργία των μηχανισμών αποτύπωσης. Επιπλέον, ιδιότητες με περισσότερες μεταβλητές τιμές (π.χ. οι γραμματοσειρές του συστήματος) μπορεί να δημιουργήσουν πιο μοναδικές αποτυπώσεις σε σχέση με ιδιότητες με λιγότερες μεταβλητές τιμές (π.χ. το όνομα του λειτουργικού συστήματος). Το γεγονός αυτό αποδεικνύει την ποικιλομορφία των τρόπων προσδιορισμού του προγράμματος περιήγησης με τη χρήση της μεθόδου αποτύπωσης.

Μέσω του προγράμματος περιήγησης, μια ιστοσελίδα είναι σε θέση να αντλήσει πληροφορίες για τα περισσότερα από τα χαρακτηριστικά που χρησιμοποιούνται στους μηχανισμούς αποτύπωσης προγράμματος περιήγησης αλλά και για το βασικό μηχανισμό επικοινωνίας της με το πρόγραμμα περιήγησης καθώς και για το ίδιο το λειτουργικό σύστημα που εκτελείται στον υπολογιστή-πελάτη. Η άντληση αυτών των στοιχείων είναι απαραίτητη για την ορθή λειτουργία των ίδιων των ιστοσελίδων ή των διαδικτυακών εφαρμογών API [11]. Για παράδειγμα, μια διαδικτυακή εφαρμογή μπορεί να χρειαστεί να προσαρμόσει την εμφάνισή της με βάση τις διαστάσεις της οθόνης του χρήστη ή μπορεί να χρειαστεί να ζητήσει την παρουσία ενός συγκεκριμένου plug-in του προγράμματος περιήγησης (π.χ. Adobe Flash Player) για να μπορεί να φορτώνει επιπλέον πόρους (π.χ. flash ταινίες και παιχνίδια). Για το λόγο αυτό κάποια προγράμματα περιήγησης έχουν ενσωματωμένα αντικείμενα JavaScript, όπως το navigator ή το screen, που δίνουν τη δυνατότητα σε διαδικτυακές εφαρμογές API να αντλούν χαρακτηριστικά των προγραμμάτων περιήγησης, όπως το όνομα, την έκδοση, την πλατφόρμα και την ανάλυση της οθόνης που εμφανίζει την εφαρμογή API. Επιπλέον, plug-in του προγράμματος περιήγησης, τα οποία είναι συστατικά στοιχεία λογισμικού που προσθέτουν χαρακτηριστικά στο πρόγραμμα περιήγησης όπως η υποστήριξη ενός συγκεκριμένου τύπου αρχείων (π.χ., αρχεία με επέκταση .swf), περιέχουν εφαρμογές API για την πρόσβαση σε βασικές ιδιότητες του συστήματος, όπως το όνομα του λειτουργικού συστήματος, η ανάλυση της οθόνης, κ.λπ. Παρόλο που αυτές οι εφαρμογές API είναι χρήσιμες για την ανάπτυξη προσαρμοσμένων διαδικτυακών εφαρμογών, την ίδια στιγμή, παρέχουν πολύτιμες πληροφορίες και δεδομένα σε φορείς παροχής μηχανισμών αποτύπωσης, επιτρέποντάς τους να δημιουργήσουν μοναδικά αναγνωριστικά των προγραμμάτων περιήγησης.

Σε γενικές γραμμές, οι μηχανισμοί αποτύπωσης προγράμματος περιήγησης / συσκευής χωρίζονται σε παθητικούς (passive browser fingerprinting) και ενεργητικούς (active browser

fingerprinting). Σαφώς και τίποτα δεν απαγορεύει το συνδυασμό των δύο κατηγοριών, γεγονός που θα μπορούσε να οδηγήσει στη δημιουργία ακόμα πιο αποτελεσματικών μηχανισμών αποτύπωσης.

5.1.1 Παθητικοί μηχανισμοί

Οι παθητικοί μηχανισμοί βασίζονται στα χαρακτηριστικά που βρίσκονται στα περιεχόμενα των διαδικτυακών αιτημάτων και που παρατηρούνται χωρίς τη χρήση οποιουδήποτε εκτελέσιμου κώδικα στην πλευρά του πελάτη. Η φιλοσοφία αυτή εμπεριέχει επομένως τη χρήση cookies, κεφαλίδων HTTP αιτημάτων, διευθύνσεων IP και άλλων στοιχείων μέσω των οποίων αντλούνται οι απαραίτητες διαδικτυακές πληροφορίες.

Για παράδειγμα, η συμβολοσειρά user-agent (User-agent string), είναι ένα πεδίο της κεφαλίδας των HTTP αιτημάτων που συνήθως περιέχει αναγνωριστικά του προγράμματος περιήγησης, της μηχανής απόδοσης (rendering engine), της έκδοσης του προγράμματος περιήγησης και του λειτουργικού συστήματος [93]. Σε ορισμένες περιπτώσεις, ο συνδυασμός του user-agent string με τη διεύθυνση IP προσδιορίζει μονοσήμαντα ένα συγκεκριμένο πρόγραμμα περιήγησης του χρήστη. Αυτό οφείλεται στο γεγονός ότι το user-agent string αποκαλύπτει πολλές μοναδικές πληροφορίες, όπως για παράδειγμα, τους αριθμούς έκδοσης του προγράμματος περιήγησης και του λειτουργικού συστήματος. Στην εικόνα 14 παρουσιάζονται τα user-agent string δημοφιλών προγραμμάτων περιήγησης. Μια ματιά στον πίνακα της εικόνας αποδεικνύει ότι τα προγράμματα περιήγησης που βασίζονται στη μηχανή απόδοσης WebKit της Apple, δηλαδή τα Safari, Opera και Chrome, επηρεάζονται σημαντικά από τη συγκεκριμένη μηχανή απόδοσης, καθώς αποκαλύπτουν τις δευτερεύουσες εκδόσεις του λειτουργικού συστήματος και της μηχανής απόδοσης, γεγονός το οποίο τα καθιστά πιο ευάλωτα στους μηχανισμούς αποτύπωσης.

Browser	User-Agent
Mozilla Firefox	Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9; rv:29.0) Gecko/20100101 Firefox/29.0
Apple Safari	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.75.14 (KHTML, like Gecko) Version/7.0.3 Safari/537.75.14
Opera	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/34.0.1847.132 Safari/537.36 OPR/21.0.1432.57
Google Chrome	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/34.0.1847.131 Safari/537.36
Microsoft Internet Explorer	Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko

Εικόνα 14: Συμβολοσειρές user-agent δημοφιλών προγραμμάτων περιήγησης (Πηγή: Christof Ferreira Torres)[14]

Η σειρά με την οποία τα προγράμματα περιήγησης στέλνουν τις κεφαλίδες HTTP και η διακύμανση αποδοχής των κεφαλίδων αυτών από τους διάφορους τύπους περιεχομένου, αποτελεί ένα ακόμη παράδειγμα παθητικού μηχανισμού αποτύπωσης [94]. Στην εικόνα 15 παρουσιάζεται ένα παράδειγμα της σειράς με την οποία αποστέλλουν διάφορα δημοφιλή προγράμματα περιήγησης τις κεφαλίδες των HTTP αιτημάτων σε ιστοσελίδες, ενώ στην εικόνα 16 παρουσιάζεται ένα τυπικό παράδειγμα τέτοιας κεφαλίδας HTTP αιτήματος [95].

[14] http://www.open.ou.nl/hjo/supervision/c.torres14-bsc-thesis.pdf

Mozilla Firefox	Apple Safari, Opera, Google Chrome	Microsoft Internet Explorer
Host	Host	Accept
User-Agent	Accept	Accept-Language
Accept	User-Agent	User-Agent
Accept-Language	Accept-Language	Accept-Encoding
Accept-Encoding	Accept-Encoding	Host

Εικόνα 15: Παραδείγματα σειράς αποστολής κεφαλίδων HTTP από δημοφιλή προγράμματα περιήγησης (Πηγή: Christof Ferreira Torres)[15]

Η σειρά αυτή εξαρτάται από τη μηχανή απόδοσης που χρησιμοποιεί το εκάστοτε πρόγραμμα περιήγησης. Ο Mozilla Firefox χρησιμοποιεί την μηχανή απόδοσης Gecko και ως εκ τούτου το αίτημα GET που αποστέλλει ακολουθείται από την εξής σειρά κεφαλίδων: του τερματικού (host), του user-agent και τέλος της αποδοχής (accept, accept-language και accept-encoding) [96]. Τα προγράμματα περιήγησης Safari, Opera και Google Chrome χρησιμοποιούν τη μηχανή αποδοχής WebKit της Apple και η σειρά των κεφαλίδων μετά το αίτημα GET έχει ως εξής: host, accept, user-agent, accept-encoding, accept-language [97]. Τέλος, ο Microsoft Internet Explorer χρησιμοποιεί τη δική του μηχανή απόδοσης που ονομάζεται Trident, η οποία στέλνει τις κεφαλίδες HTTP με έναν εντελώς διαφορετικό τρόπο από τα προηγούμενα προγράμματα περιήγησης. Μετά το αίτημα GET η σειρά αποστολής των κεφαλίδων έχει ως εξής: accept, accept-language, user-agent, accept-encoding, host [98]. Δυστυχώς, δεν υπάρχει κάποιο δεδομένο πρότυπο με το οποίο να καθορίζεται ο τρόπος αποστολής των κεφαλίδων HTTP από τις μηχανές απόδοσης. Το γεγονός αυτό αποτελεί ζήτημα για την προστασία της ιδιωτικότητας των χρηστών, δεδομένου ότι η σειρά αποστολής των κεφαλίδων αποκαλύπτει επιπλέον διακριτές πληροφορίες ενώ παράλληλα μια ιστοσελίδα μπορεί με τη σειρά αποστολής των κεφαλίδων HTTP να ανακαλύψει στοιχεία για το πρόγραμμα περιήγησης του χρήστη, ακόμη και εάν αυτός είναι σε θέση να έχει πλαστογραφήσει (spoofing) το πεδίο του user-agent [93].

[15] http://www.open.ou.nl/hjo/supervision/c.torres14-bsc-thesis.pdf

Request parameter	Value
Requested URI	/headers
Request Method	GET
Remote IP Address	85.177.91.142
Remote IP Port	34561
Protocol version	HTTP/1.1
HTTP Header*	Value
Accept	text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Charset	ISO-8859-1,utf-8;q=0.7,*;q=0.7
Accept-Encoding	gzip, deflate
Accept-Language	en-us,en;q=0.5
Connection	keep-alive
Dnt	1
Host	www.xhaus.com
Keep-Alive	115
User-Agent	Mozilla/5.0 (X11; Linux x86_64; rv:2.0.1) Gecko/20100101 Firefox/4.0.1

Εικόνα 16: Τυπικό παράδειγμα κεφαλίδας HTTP αιτήματος (Πηγή: Niklas Schmucker)[16]

5.1.2 Ενεργητικοί μηχανισμοί

Οι ενεργητικοί μηχανισμοί αποτύπωσης είναι τεχνικές, όπου μια ιστοσελίδα εκτελεί έναν κώδικα JavaScript ή άλλο κώδικα στον τοπικό υπολογιστή-πελάτη με σκοπό την παρατήρηση και άντληση δεδομένων που αφορούν πρόσθετα χαρακτηριστικά του προγράμματος περιήγησης ή του τερματικού. Οι μηχανισμοί αυτής της κατηγορίας περιλαμβάνουν πρόσβαση στο μέγεθος της οθόνης, καταμέτρηση (enumeration) των γραμματοσειρών (fonts) ή των plug-in, αξιολόγηση των χαρακτηριστικών απόδοσης ή απόδοση γραφικών μοτίβων. Παραδείγματα τέτοιων ενεργητικών μηχανισμών είναι τα εξής [91], [99]:

- Plug-in API εφαρμογές, όπως το Silverlight API της Microsoft, το Flash API της Adobe Systems ή η Java API της Oracle χρησιμοποιούνται με σκοπό τη συλλογή αναγνωριστικών πληροφοριών της συσκευής (π.χ. ταχύτητα της CPU, κατασκευαστής της CPU, εγκατεστημένα προγράμματα, κλπ.).
- Η μέτρηση της απόκλισης που παρουσιάζουν τα ρολόγια του υπολογιστή και του διακομιστή (clock skew measurement) μπορεί να γίνει μέσω αλγορίθμων, όπως του αλγόριθμου της εταιρείας 41st Parameter, η οποία ισχυρίζεται ότι βάση του αλγορίθμου της αποτελεί μια παράμετρος χρονικής απόκλισης, γεγονός που επιτρέπει στον αλγόριθμο να ελέγχει περισσότερες από 100 παραμέτρους [100].
- Αποτύπωση της στοίβας (stack) του TCP, δηλαδή η ανίχνευση των παρατυπιών στην στοίβα του TCP/IP, γίνεται με σκοπό την ανάκτηση της πραγματικής διεύθυνσης IP των πελατών ακόμα και στην περίπτωση που αυτοί συνδέονται σε ιστοσελίδες μέσω διακομιστών μεσολάβησης (proxy servers).

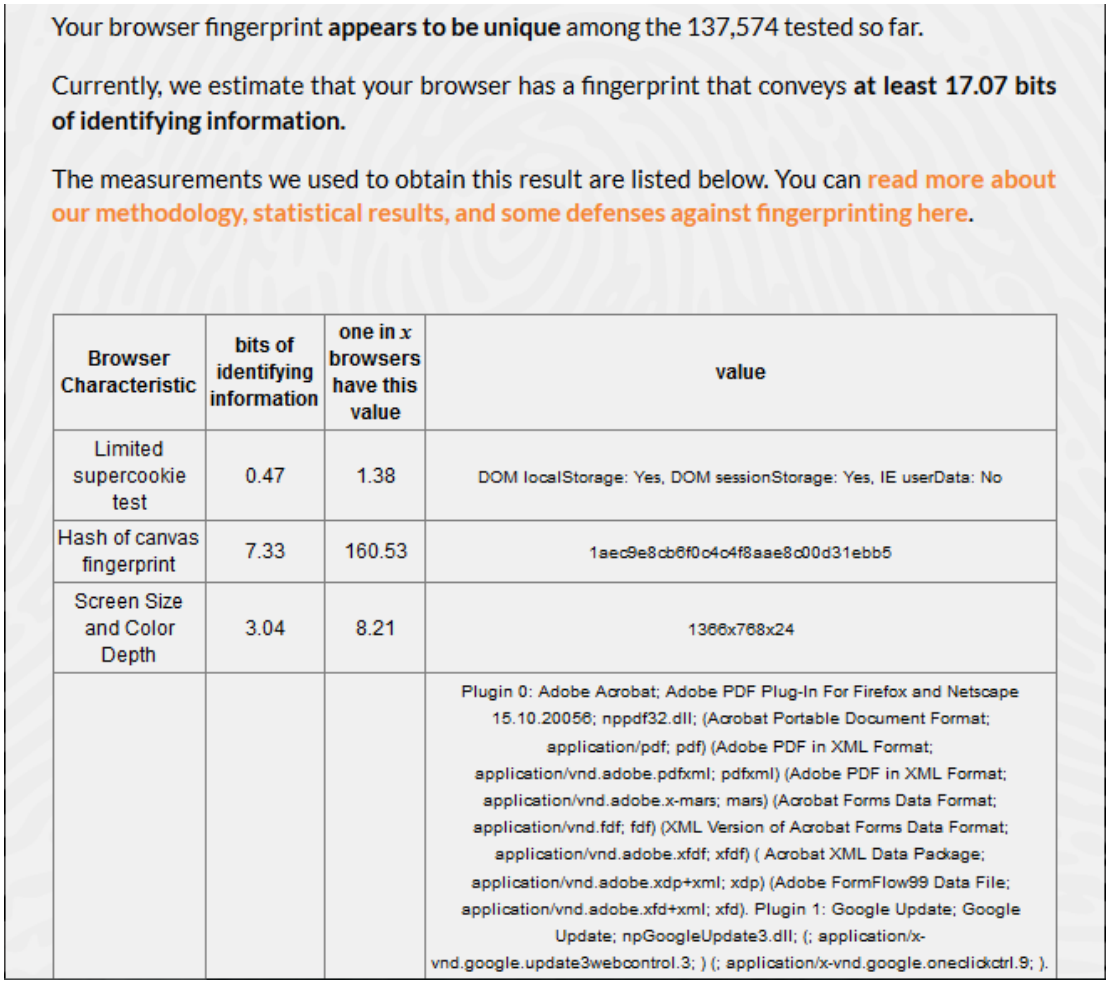
[16] http://www.snet.tu-berlin.de/fileadmin/fg220/courses/SS11/snet-project/web-tracking_schmuecker.pdf

- Η ανίχνευση του ιστορικού της αλληλουχίας φύλλων στυλ (CSS history detection hack) μπορεί να χρησιμοποιηθεί για τον εντοπισμό των ιστοσελίδων που έχει επισκεφθεί ο χρήστης.
- Η ανίχνευση των εγκατεστημένων γραμματοσειρών στο σύστημα μπορεί να πραγματοποιηθεί με τη χρήση του Flash, της Java ή οποιουδήποτε άλλου plug-in, αλλά και μέσω χρήσης JavaScript και CSS [101].
- Ένα ευρύ φάσμα ελέγχων συμπεριφοράς του JavaScript μπορεί να χρησιμοποιηθεί για τη μέτρηση της υπογραφής της απόδοσης της μηχανής JavaScript του προγράμματος περιήγησης, επιτρέποντας την ανίχνευση της έκδοσης του προγράμματος περιήγησης, του λειτουργικού συστήματος αλλά και της αρχιτεκτονικής δομής του τερματικού [102], [103].
- Ο μηχανισμός αποτύπωσης HTML5 Canvas μπορεί να παράγει διαφορετικά εικονοστοιχεία (pixel) σε διαφορετικά προγράμματα περιήγησης, ανάλογα με το σύστημα στο οποίο εκτελείται [104].

Θα πρέπει να σημειωθεί ότι οι ενεργητικοί μηχανισμοί αποτύπωσης που αναφέρθηκαν παραπάνω είναι μόνο η κορυφή του παγόβουνου και ότι υπάρχουν πολλοί περισσότεροι που βοηθούν στη συλλογή πληροφοριών που περιλαμβάνουν τα αποτυπώματα.

5.1.3 Η μοναδικότητα του προγράμματος περιήγησης

Το Σεπτέμβρη του 2010, η πρωτοβουλία Panopticlick του οργανισμού Electronic Frontier Foundation (EFF) εφάρμοσε μια τεχνική ταυτοποίησης ενός διαδικτυακού πελάτη κατά τη στιγμή της πρόσβασης σε μια ιστοσελίδα [91]. Χρησιμοποιώντας έναν αλγόριθμο μέσω αιτημάτων HTTP και AJAX συνέλεξε πληροφορίες σχετικά με τα εγκατεστημένα plug-in, την ανάλυση της οθόνης, τις πηγές, τη ζώνη ώρας, τα cookies και τα αντικείμενα flash και καθόρισε ένα αποτύπωμα που επιτρέπει σε ένα διαδικτυακό πελάτη να είναι διακριτός από τα εκατομμύρια των άλλων. Αποδείχθηκε ότι το ποσοστό της αξιόπιστης ταυτοποίησης ενός προγράμματος περιήγησης που έχει ενεργοποιημένα την Java και τον Flash μπορεί να φτάσει σχεδόν το 95%. Αυτό το υψηλό ποσοστό αποδεικνύει την ακρίβεια με την οποία μπορεί να κατασκευαστεί ένα αποτύπωμα, προσδιορίζοντας την ταυτότητα ενός προγράμματος περιήγησης ή μιας συγκεκριμένης συσκευής. Στην εικόνα 17 παρουσιάζεται ένα παράδειγμα απεικόνισης της ανάλυσης Panopticlick ενός διαδικτυακού πελάτη.



Your browser fingerprint **appears to be unique** among the 137,574 tested so far.

Currently, we estimate that your browser has a fingerprint that conveys **at least 17.07 bits of identifying information.**

The measurements we used to obtain this result are listed below. You can read more about our methodology, statistical results, and some defenses against fingerprinting here.

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
Limited supercookie test	0.47	1.38	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No
Hash of canvas fingerprint	7.33	160.53	1aec9e8cb6f0c4c4f8aae8c00d31ebb5
Screen Size and Color Depth	3.04	8.21	1366x768x24
			Plugin 0: Adobe Acrobat; Adobe PDF Plug-In For Firefox and Netscape 15.10.20056; nppdf32.dll; (Acrobat Portable Document Format; application/pdf; pdf) (Adobe PDF in XML Format; application/vnd.adobe.pdfxml; pdfxml) (Adobe PDF in XML Format; application/vnd.adobe.x-mars; mars) (Acrobat Forms Data Format; application/vnd.fdf; fdf) (XML Version of Acrobat Forms Data Format; application/vnd.adobe.xfdf; xfdf) (Acrobat XML Data Package; application/vnd.adobe.xdp+xml; xdp) (Adobe FormFlow99 Data File; application/vnd.adobe.xfd+xml; xfd). Plugin 1: Google Update; Google Update; npGoogleUpdate3.dll; (; application/x-vnd.google.update3webcontrol.3;) (; application/x-vnd.google.oneclickctrl.9;).

Εικόνα 17: Παράδειγμα ανάλυσης Panopticlick (Πηγή: Panopticlick)[17]

Από το Δεκέμβρη του 2015, ο οργανισμός EFF κυκλοφόρησε την νέα έκδοση, το Panopticlick 2.0 [105]. Η νέα έκδοση του Panopticlick προσθέτει επιπλέον δυνατότητες ανάλυσης στο online εργαλείο που βελτιώνουν την αξία του. Οι επιπρόσθετες αυτές δυνατότητες είναι οι ακόλουθες:

- Έλεγχος του μηχανισμού αποτύπωσης HTML5 Canvas
- Έλεγχος προστασίας από παρακολούθηση μέσω διαφημίσεων ή αόρατων web beacons
- Έλεγχος συμβατότητας με την πολιτική «Do Not Track»

Στις περισσότερες περιπτώσεις οι έλεγχοι λειτουργούν ικανοποιητικά, αλλά αποτυγχάνουν εάν το λογισμικό ασφαλείας του υπολογιστή ή τα add-ons του προγράμματος περιήγησης δεν επιτρέπουν την λειτουργία συγκεκριμένων τεχνολογιών.

Η νέα σελίδα αποτελεσμάτων εμφανίζει τις ακόλουθες πληροφορίες (Εικ. 18):

- Εάν το πρόγραμμα περιήγησης αποκλείει την παρακολούθηση διαφημίσεων
- Εάν το πρόγραμμα περιήγησης αποκλείει αόρατους trackers

[17] https://panopticlick.eff.org/

- Αν το πρόγραμμα περιήγησης επιτρέπει την πρόσβαση σε τρίτους που συμμορφώνονται με την πολιτική «Do Not Track» [106]
- Εάν το πρόγραμμα περιήγησης προστατεύει τη δημιουργία αποτυπωμάτων

Εικόνα 18: Παράδειγμα σελίδας αποτελεσμάτων της Panopticlick 2.0 ανάλυσης (Πηγή: Aaron Stuart)[18]

Επίσης το Panopticlick 2.0 επιτρέπει την λεπτομερή εμφάνιση των αποτελεσμάτων καθενός από τους ελέγχους που πραγματοποιούνται από την υπηρεσία. Κάποιοι από τους ελέγχους αυτούς είναι οι εξής:

- Έλεγχος για supercookies και cookies
- Έλεγχος του μηχανισμού αποτύπωσης HTML5 Canvas
- Έλεγχος μεγέθους οθόνης και βάθους χρωμάτων
- Έλεγχος των plug-in του προγράμματος περιήγησης
- Έλεγχος ζώνης ώρας
- Έλεγχος ενεργοποιημένης κεφαλίδας «Do Not Track»
- Έλεγχος κεφαλίδων HTTP Accept
- Έλεγχος για WebGL αποτύπωση
- Έλεγχος γλώσσας και γραμματοσειρών συστήματος
- Έλεγχος πλατφόρμας
- Έλεγχος user-agent

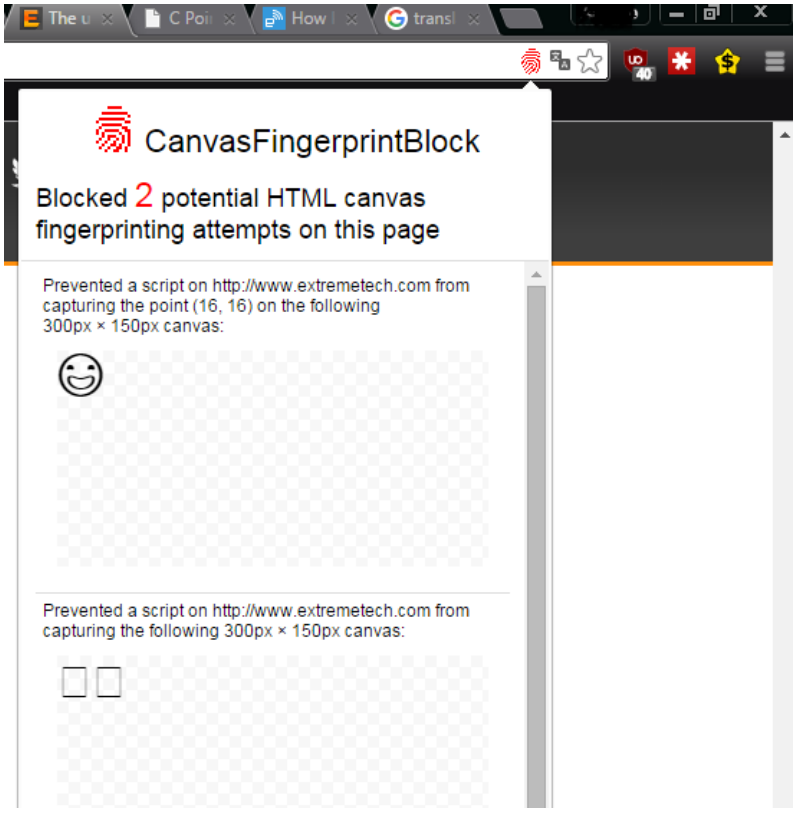
[18] https://websetnet.com/eff-launches-panopticlick-2-tracking-fingerprinting-tests

Σε περίπτωση που το JavaScript του υπολογιστή είναι απενεργοποιημένο, οι μοναδικοί έλεγχοι που λειτουργούν είναι αυτοί της ενεργοποιημένης κεφαλίδας «Do Not Track», των κεφαλίδων HTTP Accept, των user agents και του αν είναι ενεργοποιημένα τα cookies. Όλοι οι άλλοι έλεγχοι αποτυγχάνουν χωρίς το JavaScript. Η προσθήκη νέων ελέγχων δίνει μεγαλύτερες δυνατότητες στην πρωτοβουλία Panopticlick.

Τέλος, το Panopticlick 2.0 μπορεί να προτείνει εργαλεία, ανάλογα με τα αποτελέσματα της ανάλυσης. Σύμφωνα με τον οργανισμό EFF, τέτοια εργαλεία είναι τα Privacy Badger, Adblock ή Disconnect, ανάλογα με την πλατφόρμα και τα αποτελέσματα των ελέγχων [107].

5.1.4 Μηχανισμός αποτύπωσης HTML5 Canvas

Το αποτύπωμα canvas είναι μία από τις πιο πρόσφατες μεθόδους που χρησιμοποιούνται για τον εντοπισμό συσκευών με βάση τα χαρακτηριστικά του hardware υλικού τους. Η τεχνολογία WebGL χρησιμοποιείται για τη δημιουργία γραφημάτων (συνήθως 3D) και δημιουργεί μια εικόνα στο στοιχείο canvas της HTML του πελάτη [104]. Η εικόνα αυτή εξαρτάται άμεσα από το hardware υλικό και παρουσιάζει αρκετά υψηλό βαθμό εντροπίας, ικανό να δημιουργήσει αποτύπωμα από τον υπολογιστή του χρήστη. Με την ανάλυση των χαρακτηριστικών των εικονοστοιχείων που συνθέτουν την εικόνα που παράγεται στο διαδικτυακό πελάτη, είναι δυνατόν να ληφθεί ένα αποτύπωμα αναγνώρισης με υψηλό βαθμό ακρίβειας.



Εικόνα 19: Παράδειγμα HTML5 canvas fingerprint (Πηγή: INCIBE)[19]

[19] https://www.incibe.es/blogs/post/Security/SecurityBlog/Article_and_comments/web_tracking_en

5.1.5 Ανθεκτικότητα των αποτυπωμάτων του προγράμματος περιήγησης

Το αποτύπωμα ενός προγράμματος περιήγησης είναι δυνατόν να αλλάξει για πολλούς λόγους. Κάποιοι από τους λόγους αυτούς θα μπορούσε να είναι οι ενημερώσεις του προγράμματος περιήγησης, η εγκατάσταση ή η αναβάθμιση των plug-in, η απενεργοποίηση των cookies ή του JavaScript, η εγκατάσταση μιας νέας γραμματοσειράς ή η εγκατάσταση μιας εξωτερικής εφαρμογής που περιλαμβάνει νέες γραμματοσειρές, η αγορά μιας νέας οθόνης ή, στην περίπτωση ενός φορητού υπολογιστή, η σύνδεση με μια εξωτερική οθόνη η οποία στη συνέχεια αλλάζει την ανάλυση της οθόνης. Επί της ουσίας όμως, η ανθεκτικότητα των αποτυπωμάτων του προγράμματος περιήγησης δεν πρόκειται να αλλάξει ή ακόμα χειρότερα μπορεί να αυξηθεί λόγω του γεγονότος ότι υπάρχουν κάποιες τιμές που παραμένουν σταθερές ή σπάνια αλλάζουν μετά την εγκατάσταση ενός συστήματος.

Όλες οι αναβαθμίσεις του προγράμματος πλοήγησης, των plug-in και των γραμματοσειρών είναι δυνατόν να ανιχνευθούν και να συνδυαστούν με προηγούμενα αποτυπώματα μέσω ενός μόνο μέρους των διευθύνσεων IP ή μέσω προηγούμενων αποτυπωμάτων που περιέχονται σε μια βάση δεδομένων. Οι K. Boda και συν. απέδειξαν την αποτελεσματικότητα μιας τέτοιας ανίχνευσης χρησιμοποιώντας μια βάση δεδομένων που περιείχε παλαιότερα αποτυπώματα προγραμμάτων πλοήγησης βάσει ενός συγκεκριμένου συνόλου γραμματοσειρών, την ζώνη της ώρας, και την ανάλυση της οθόνης [108]. Οι συγγραφείς απέδειξαν ότι τα στοιχεία αυτά αρκούν για την ανίχνευση παλαιότερων αποτυπωμάτων ανεξαρτήτως προγράμματος πλοήγησης και ότι τα user-agent string μπορεί να είναι πιο αποτελεσματικά αλλά ταυτόχρονα και εύθραυστα αναγνωριστικά.

Ο Peter Eckersley εφάρμοσε έναν αλγόριθμο ευρετικής εκτίμησης με σκοπό την ανίχνευση του κατά πόσον ένα δεδομένο αποτύπωμα θα μπορούσε να είναι μια εξελιγμένη έκδοση ενός παλαιότερου αποτυπώματος [91]. Ο αλγόριθμος έκανε σωστή πρόβλεψη για το 65% των περιπτώσεων, λανθασμένη πρόβλεψη για το 0,56% των περιπτώσεων και καμία πρόβλεψη για το 35% των περιπτώσεων. Το ποσοστό των ψευδώς θετικών αποτελεσμάτων έφτασε μόλις το 0,86%.

Από τα παραπάνω προκύπτει το συμπέρασμα ότι παρά το γεγονός ότι τα αποτυπώματα έχουν αποδειχθεί ότι δεν είναι ιδιαίτερα σταθερά, τα προγράμματα περιήγησης αποκαλύπτουν τόσο πολλές πληροφορίες σχετικά με την έκδοση αλλά και τις ρυθμίσεις ενός συστήματος που τελικά αποτελούν ανεξάντλητες πηγές ορθής λειτουργίας των μηχανισμών αποτύπωσης προγραμμάτων πλοήγησης και συσκευών.

5.2 Μηχανισμοί αποτύπωσης λειτουργικού συστήματος

Αν και το λειτουργικό σύστημα που χρησιμοποιεί ένας διαδικτυακός πελάτης δεν είναι γενικά μοναδικό, εξακολουθεί να είναι χρήσιμο για τους σκοπούς του web tracking. Όπως αναφέρθηκε σε προηγούμενη ενότητα η αποτύπωση της στοίβας του TCP (TCP/IP stack) γίνεται με σκοπό την ανάκτηση της πραγματικής διεύθυνσης IP των πελατών. Η αποτύπωση αυτή περιλαμβάνει τη συγκέντρωση πληροφοριών σχετικά με τις ρυθμιστικές ιδιότητες του επιπέδου TCP, παράμετροι που όταν συνδυαστούν μπορεί να χρησιμεύσουν και για την αποτύπωση του λειτουργικού συστήματος του υπολογιστή-πελάτη [109].

Κάποιες από τις παραμέτρους του πρωτοκόλλου TCP παραμένουν αναλλοίωτες μέχρι την υλοποίηση κάποιας δικτυακής σύνδεσης. Διαφορετικά λειτουργικά συστήματα και διαφορετικές εκδόσεις του ίδιου λειτουργικού συστήματος, θέτουν διαφορετικές

προεπιλεγμένες τιμές στις παραμέτρους αυτές. Η συλλογή και η ανάλυση των τιμών αυτών μπορεί να δώσει πολύτιμες πληροφορίες σχετικά με τη διαφοροποίηση των διαφόρων λειτουργικών συστημάτων αλλά και των υλοποιήσεων του πρωτοκόλλου TCP/IP. Τα πεδία του TCP/IP που χρησιμεύουν σε αυτή την περίπτωση είναι τα ακόλουθα:

- Αρχικό μέγεθος του πακέτου (16 bits)
- Αρχικό TTL (8 bits)
- Μέγεθος παραθύρου (16 bits)
- Μέγιστο μέγεθος segment (16 bits)
- Τιμή κλιμάκωσης παραθύρου (8 bits)
- "don't fragment" flag (1 bit)
- "SackOK" flag (1 bit)
- "NOP" flag (1 bit)

Ο συνδυασμός των παραμέτρων αυτών μπορεί να σχηματίσει ένα αποτύπωμα των 67bit που αφορά ένα συγκεκριμένο υπολογιστή-πελάτη. Από τα πεδία αυτά, ο έλεγχος του αρχικού TTL και του μεγέθους του παραθύρου συχνά αρκεί, προκειμένου να προσδιοριστεί επιτυχώς ένα λειτουργικό σύστημα [110].

Όπως συμβαίνει και στην περίπτωση της αποτύπωσης του προγράμματος πλοήγησης, το λειτουργικό σύστημα μιας συσκευής μπορεί επίσης να αποτυπωθεί με χρήση JavaScript και Flash [6]. Η Javascript διευκολύνει επίσης την αναγνώριση της γλώσσας του συστήματος, τη γλώσσα που χρησιμοποιεί ο χρήστης, την τοπική ώρα αλλά και την τοπική ημερομηνία και ώρα με απόκλιση έως 1 χιλιοστού του δευτερολέπτου. Η χρήση της Flash διευκολύνει επίσης την ανίχνευση των ηχητικών δυνατοτήτων του συστήματος, τη δυνατότητα πρόσβασης του χρήστη σε κάμερα ή μικρόφωνο, την υποστήριξη εκτύπωσης και τη δυνατότητα πρόσβασης ανάγνωσης του σκληρού δίσκου.

5.3 Μηχανισμοί αποτύπωσης θέσης

Η διεύθυνση IP της συσκευής ή του δικτύου στο οποίο βρίσκεται αποτελεί ένα από τα πιο κοινά χρησιμοποιούμενα δεδομένα στην προσπάθεια εντοπισμού των χρηστών του διαδικτύου. Οι μηχανισμοί αποτύπωσης θέσης που χρησιμοποιούνται σε αυτή την περίπτωση είναι οι εξής [111]:

- Η ανάλυση της κυκλοφορίας δεδομένων μεταξύ ιστοσελίδας και συσκευής
- Οι κεφαλίδες HTTP
- Η χρήση των Java, Flash, JavaScript και HTML5

Εικόνα 20: Απόκτηση ιδιωτικής διεύθυνσης IP με χρήση του WebRTC (σε Chrome, Firefox) (Πηγή: INCIBE)[20]

Πληροφορίες σχετικές με τις διευθύνσεις IP των χρηστών μπορεί ακόμα να προέρχονται από τους διακομιστές μεσολάβησης (proxy servers) ή από ιδιωτικά δίκτυα των πελατών με χρήση API Javascript εφαρμογών, που παρέχονται, για παράδειγμα, από το WebRTC (Web Real-Time Communication), ένα API definition που δημιουργήθηκε από το W3C για να υποστηρίξει εφαρμογές φωνής, βιντεοκλήσεις και P2P διαμοιρασμό αρχείων μεταξύ προγραμμάτων περιήγησης, χωρίς plug-in (Εικ. 19) [112].

Ο γεωγραφικός εντοπισμός (geolocation) είναι ένα ακόμα από τα δεδομένα που εξάγονται και χρησιμοποιούνται για τη δημιουργία προφίλ και την ταυτοποίηση των χρηστών [111]. Για το σκοπό αυτό, οι μηχανισμοί αποτύπωσης θέσης χρησιμοποιούν δημόσιες βάσεις δεδομένων ή εφαρμογές API HTML5 που αφορούν τον γεωγραφικό εντοπισμό. Παράδειγμα δημόσιας βάσης δεδομένων γεωγραφικού εντοπισμού αποτελεί η βάση δεδομένων GeoLite [113]. Οι εφαρμογές API που αφορούν τον γεωγραφικό εντοπισμό δίνουν τη δυνατότητα στους χρήστες να παρέχουν την γεωγραφική τους θέση σε διαδικτυακές εφαρμογές αν το επιθυμούν [114]. Ωστόσο, ο γεωγραφικός εντοπισμός είναι από τα δεδομένα που δεν είναι επαρκώς ακριβή και επηρεάζεται από περιστάσεις όπως η χρήση δικτύου VPN ή Tor, τα οποία παραποιούν την πραγματική θέση του χρήστη.

6 Αντιμετώπιση Μηχανισμών Web Tracking

Τα προβλήματα που προκύπτουν από το web tracking, ως μέρος της διαδικασίας δημιουργίας προφίλ της συμπεριφοράς των χρηστών στις διαδικτυακές τους περιηγήσεις, είναι πολλά και επηρεάζουν την προστασία των θεμελιωδών δικαιωμάτων των χρηστών, όπως συζητήθηκε στο πρώτο μέρος της παρούσας διατριβής. Ο συνδυασμός των σκοπών της λειτουργίας των μηχανισμών παρακολούθησης, όπως περιεγράφηκαν στο δεύτερο μέρος της διατριβής, καθώς και των νομικών απαιτήσεων για την προστασία του δικαιώματος της ιδιωτικότητας και του δικαιώματος προστασίας των δεδομένων σε σχέση με την παρακολούθηση και τη δημιουργία προφίλ της online συμπεριφοράς των χρηστών, καθιστά επιτακτική την ανάγκη για λήψη μέτρων, τα οποία μπορεί να επιτρέψουν στους χρήστες τη δυνατότητα επιλογής όσον αφορά την online παρακολούθησή τους.

Σε αυτό το κεφάλαιο, θα εξεταστούν τα μέτρα αντιμετώπισης των μηχανισμών web tracking και το πώς αυτά παρέχουν στο χρήστη προστασία αλλά και τη δυνατότητα επιλογής ώστε να αποφεύγει την παρακολούθηση. Η δυνατότητα επιλογής απαιτεί τη συνειδητοποίηση της

[20] https://www.incibe.es/blogs/post/Security/SecurityBlog/Article_and_comments/web_tracking_en

δραστηριότητας της παρακολούθησης εκ μέρους των χρηστών, δηλαδή την ύπαρξη επαρκούς πληροφόρησης και την ικανότητα των χρηστών να αντιλαμβάνονται τις συνέπειες που έχει η αποκάλυψη πληροφοριών που αφορούν την προσωπική τους ζωή καθώς και το δικαίωμα τους να εκφράσουν ελεύθερα την επιθυμία της μη παρακολούθησής τους. Ο όρος «μέτρα αντιμετώπισης» χρησιμοποιείται με την ευρεία έννοια και δεν περιορίζεται μόνο στα μέτρα αντιμετώπισης των μηχανισμών web tracking, όπως υπονοείται. Για τους σκοπούς της παρούσας εργασίας, ο όρος αυτός περιλαμβάνει τις τεχνολογικές δυνατότητες, τις εμπορικές και κοινωνικές πρωτοβουλίες καθώς και τις νομικές διατάξεις που αποσκοπούν στη διευκόλυνση της επιλογής του χρήστη για ενεργή και αποτελεσματική αναφορά του στον φορέα παρακολούθησης σχετικά με το αν επιθυμεί να παρακολουθείται και αν συμφωνεί με την δημιουργία προφίλ της online συμπεριφοράς του.

6.1 Καθορισμός των κριτηρίων αξιολόγησης

Για την αξιολόγηση των μέτρων αντιμετώπισης των μηχανισμών web tracking απαιτείται ο καθορισμός κάποιων κριτηρίων. Με βάση τα όσα έχουν αναφερθεί στα προηγούμενα κεφάλαια, τα κριτήρια που επιλέχθηκαν είναι τα εξής: α) η κοινοποίηση, β) οι πληροφορίες, γ) το επίπεδο προστασίας και δ) η εφαρμοσιμότητα. Η αξιολόγηση των μέτρων αντιμετώπισης των μηχανισμών web tracking μέσω των κριτηρίων αυτών χωρίζεται σε δύο στάδια. Τα κριτήρια της κοινοποίησης και των πληροφοριών εφαρμόζονται πριν από οποιαδήποτε δραστηριότητα παρακολούθησης με το σκεπτικό ότι ο χρήστης δεν είναι εξοικειωμένος με την ύπαρξη των μηχανισμών παρακολούθησης. Τα άλλα δύο κριτήρια εφαρμόζονται σε ένα δεύτερο στάδιο, αφού ο χρήστης έχει επιλέξει να μην παρακολουθείται. Η υλοποίηση του δεύτερου σταδίου προϋποθέτει ότι τα δύο κριτήρια του πρώτου σταδίου πληρούνται, δηλαδή ο χρήστης έχει επιλέξει να μην παρακολουθείται με δεδομένο ότι έχει πληροφορηθεί για την ύπαρξη των μηχανισμών παρακολούθησης. Η περίπτωση κατά την οποία ο χρήστης επιλέγει να παρακολουθείται δεν έχει κανένα ενδιαφέρον για το δεύτερο στάδιο, αφού, εφόσον ο χρήστης ενημερώθηκε και επέλεξε θετικά υπέρ της παρακολούθησης, τότε το πρώτο στάδιο λειτούργησε ικανοποιητικά, παρέχοντας στο χρήστη το δικαίωμα να ασκήσει την επιλογή του. Επομένως, ο χρήστης αυτός δεν χρειάζεται τα διάφορα μέτρα αντιμετώπισης των μηχανισμών web tracking, πράγμα που σημαίνει ότι το κριτήριο της εφαρμοσιμότητας δεν υπάρχει λόγος να εξετασθεί. Ο κίνδυνος υπάρχει για τον χρήστη που επέλεξε να μην παρακολουθείται και τα δικαιώματα του μπορεί να παραβιάζονται από οντότητες που δεν σέβονται την επιθυμία του αυτή. Η έκταση της προστασίας και της αποτελεσματικότητας του κάθε μέτρου αντιμετώπισης των μηχανισμών web tracking εξαρτάται από το πώς ανταποκρίνεται σε κάθε κριτήριο. Με βάση το σκεπτικό αυτό, αποτελεσματικά θεωρούνται τα μέτρα που ανταποκρίνονται σε περισσότερα από ένα κριτήρια.

Το κριτήριο της κοινοποίησης είναι ζωτικής σημασίας και αποτελεί το πρώτο βήμα προστασίας των θεμελιωδών δικαιωμάτων του χρήστη από την άποψη ότι ο χρήστης πρέπει να γνωρίζει ότι υπάρχει μια πιθανή απειλή, πριν λάβει οποιαδήποτε μέτρα για την προστασία των δικαιωμάτων του. Με βάση το κριτήριο αυτό, η αξιολόγηση των μέτρων αντιμετώπισης των μηχανισμών web tracking αφορά το αν και κατά πόσο είναι σε θέση να αντιμετωπίσουν την πτυχή του προβλήματος της κοινοποίησης στον χρήστη για την παρακολούθηση της online δραστηριότητας και συμπεριφοράς του. Ο τρόπος της κοινοποίησης στον χρήστη ότι παρακολουθείται θα πρέπει να είναι σαφής και εύκολα ορατός.

Ο χρήστης θα πρέπει επίσης να έχει στη διάθεσή του πληροφορίες, κυρίως για τους σκοπούς της παρακολούθησης, για τον τύπο των πληροφοριών και των δεδομένων που συλλέγονται, για τον φορέα παρακολούθησης ή τον υπεύθυνο επεξεργασίας των δεδομένων, για τα δικαιώματα πρόσβασης, διόρθωσης, κλπ. όπως αυτά ορίζονται στην οδηγία για την προστασία των δεδομένων (Data Protection Directive 95/46/EC) [115], για το δικαίωμά του να κάνει άρση της οποιασδήποτε συναίνεσης, για τους μηχανισμούς παρακολούθησης, κλπ. Όλες αυτές οι πληροφορίες θα τον οδηγήσουν στη σωστή λήψη απόφασης σε σχέση με το θέμα της παρακολούθησης. Το κριτήριο για την αποτελεσματική διαχείριση αυτής της πτυχής του ζητήματος είναι η παροχή των παραπάνω πληροφοριών στο χρήστη, ώστε να διευκολύνεται η συνειδητή επιλογή.

Το κριτήριο της προστασίας αφορά τους χρήστες που έχουν επιλέξει να μην παρακολουθούνται. Το δυνητικό επίπεδο προστασίας που προσφέρουν τα μέτρα αντιμετώπισης των μηχανισμών web tracking μπορεί να εξεταστεί ως προς τα αποτελέσματα:

- έναντι όλων ή ορισμένων φορέων παρακολούθησης,
- έναντι του συνόλου ή ορισμένων μηχανισμών και
- με βάση την καθολική ή γεωγραφικά περιορισμένη προστασία που παρέχουν.

Τέλος, το κριτήριο της εφαρμοσιμότητας αφορά το κατά πόσο το εκάστοτε μέτρο αντιμετώπισης των μηχανισμών web tracking μπορεί να επιβάλει τη συμμόρφωση με το σεβασμό της επιθυμίας του χρήστη.

Στο σημείο αυτό θα πρέπει να σημειωθεί ότι σίγουρα υπάρχουν και άλλα κριτήρια αξιολόγησης των μέτρων αντιμετώπισης των μηχανισμών web tracking. Ωστόσο, στα πλαίσια της διατριβής αυτής, τα τέσσερα αυτά κριτήρια επιλέχθηκαν, κυρίως επειδή ανταποκρίνονται τόσο στις κύριες απαιτήσεις της νομοθεσίας της ΕΕ, όπως η συγκατάθεση (E-Privacy Directive 2009/136/EC) [116] και η ποιότητα των δεδομένων και των δικαιωμάτων των δεδομένων (Data Protection Directive 95/46/EC) όσο και στην πραγματική πρακτική της παρακολούθησης που αφορά την απόκρυψη λειτουργίας καθώς και τον πλουραλισμό των μηχανισμών web tracking. Επιπλέον, το ζήτημα της επιλογής του τρόπου παρακολούθησης, δηλαδή αν η παρακολούθηση του χρήστη γίνεται μετά από δική του συναίνεση ή από προεπιλογή κάποιου από τα εμπλεκόμενα μέρη, δεν έχει οριστεί ως κριτήριο, καθώς η επιλογή του ίδιου του χρήστη όσον αφορά το θέμα της παρακολούθησης θεωρείται ως ο πιο λογικός και δίκαιος τρόπος να εκφράσει τη βούλησή του. Προεπιλεγμένες ρυθμίσεις ενός προγράμματος περιήγησης ή μιας ιστοσελίδας που είτε επιτρέπουν είτε απορρίπτουν την παρακολούθηση δεν εγγυώνται τη συνειδητή επιλογή του χρήστη πάνω στο θέμα αυτό. Η προεπιλεγμένη ρύθμιση θα πρέπει να είναι ουδέτερη, προκειμένου να εκφράζεται η πραγματική προτίμηση του χρήστη ή εάν αυτό δεν είναι δυνατό, τότε μια μέση λύση θα ήταν προτιμητέα: για παράδειγμα, η μη παρακολούθηση από προεπιλογή, αλλά η υπενθύμιση στο χρήστη ότι θα πρέπει να ρυθμίσει τις προτιμήσεις του.

6.2 Ιδιωτικές πρωτοβουλίες

Με τον όρο ιδιωτικές πρωτοβουλίες εννοούνται όλοι οι τρόποι λήψης μέτρων από τους χρήστες με σκοπό την αντιμετώπιση των μηχανισμών web tracking. Με βάση τη βιβλιογραφία και τις τεχνολογικές εξελίξεις πάνω στο θέμα του web tracking, τα μέτρα που μπορούν να λάβουν οι

ίδιοι οι χρήστες προκειμένου να αντιμετωπίσουν τους μηχανισμούς της παρακολούθησης στον ιστό χωρίζονται στις εξής κατηγορίες [117]:

- Εργαλεία αντιμετώπισης μηχανισμών web tracking
- Αυτορρύθμιση (Self- regulation)
- Πρωτοβουλία Do Not Track (Do Not Track Initiative)

6.2.1 Εργαλεία αντιμετώπισης μηχανισμών web tracking

Η εξέλιξη των μηχανισμών web tracking έχει οδηγήσει και στην ανάπτυξη διάφορων εργαλείων αντιμετώπισής τους. Τα εργαλεία αυτά στοχεύουν στην αποτροπή λειτουργίας των τεχνολογιών που υποστηρίζουν τους μηχανισμούς παρακολούθησης και περιλαμβάνονται στις λίστες που δημιουργούνται για την παρακολούθηση της online συμπεριφοράς των χρηστών. Κατά καιρούς έχει δημιουργηθεί μια πληθώρα τέτοιων εργαλείων τα οποία θα μπορούσαν να ομαδοποιηθούν στις εξής κατηγορίες:

- Λίστες προστασίας από παρακολούθηση (Tracking protection lists)
- Λειτουργία ιδιωτικής περιήγησης (Private browsing feature)

Λίστες προστασίας από παρακολούθηση

Οι λίστες προστασίας από παρακολούθηση (tracking protection lists) είναι εργαλεία που εμποδίζουν την παρακολούθηση, όπως η επέκταση προγραμμάτων πλοήγησης Ghostery [118], και περιέχουν λίστες με τεχνολογίες που υποστηρίζουν τους μηχανισμούς web tracking. Τα εργαλεία αυτά, προσφέρουν στο χρήστη έναν ορισμένο βαθμό ελέγχου, υπό την έννοια ότι ο χρήστης μπορεί να επιλέξει αν θέλει να παρακολουθείται από ορισμένες εταιρείες και ορισμένους τύπους τεχνολογιών (διαφήμισης, analytics, κλπ.). Επιπλέον, τα εργαλεία αυτά εμποδίζουν τους μηχανισμούς παρακολούθησης, εφαρμόζοντας με τον τρόπο αυτό την προτίμηση του χρήστη να μην παρακολουθείται (κριτήριο εφαρμοσιμότητας). Ωστόσο, είναι σημαντικό, ο χρήστης να γνωρίζει πρώτα ότι παρακολουθείται για να κατεβάσει και να εγκαταστήσει τα εργαλεία αυτά (κριτήριο κοινοποίησης). Ο χειρισμός και η χρήση τέτοιων εργαλείων προϋποθέτει μια μέση ικανότητα του χρήστη. Για παράδειγμα, τα εργαλεία πρέπει τακτικά να ενημερώνονται, κάτι που απαιτεί μια πρόσθετη προσπάθεια από το χρήστη. Η ανάγκη για ενημερώσεις θέτει επίσης το ερώτημα σχετικά με το τι μπορεί να συμβεί αν ο χρήστης δεν ενημερώσει τα λογισμικά αυτά. Το πιο λογικό είναι η έκθεση του χρήστη σε δραστηριότητες παρακολούθησης, καθώς είναι σίγουρη η ανάπτυξη νέων μηχανισμών παρακολούθησης, οι τεχνολογίες των οποίων δεν θα περιλαμβάνονται στις λίστες των μη ενημερωμένων εργαλείων. Επίσης, οι φορείς που χειρίζονται τις λίστες των εργαλείων αυτών θα πρέπει να τις ενημερώνουν συνεχώς, για να περιλαμβάνουν κάθε νέα τεχνολογία παρακολούθησης. Ένα άλλο ζήτημα είναι ότι τα εργαλεία αυτά εμποδίζουν μεν τη λειτουργία των μηχανισμών web tracking, όμως δεν προσφέρουν περαιτέρω πληροφορίες σχετικά με τους σκοπούς της παρακολούθησης, τα δεδομένα ή τις πληροφορίες που συλλέγονται από τις οντότητες παρακολούθησης (κριτήριο πληροφορίας). Οι λίστες προστασίας από παρακολούθηση φαίνεται να δίνουν μια «λύση» χωρίς όμως προηγουμένως να έχουν ασχοληθεί με τα βασικά ζητήματα που οδήγησαν στη δημιουργία του προβλήματος.

Στις λίστες προστασίας από παρακολούθηση προκύπτουν συχνά ζητήματα που αφορούν τη διαφάνεια της διαδικασίας χαρακτηρισμού μιας τεχνολογίας ως «tracker» και ενσωμάτωσής της σε μία από τις κατηγορίες της λίστας. Επιπλέον, ο χρήστης σπάνια γνωρίζει ποιος φορέας

βρίσκεται πίσω από καθένα από αυτά τα εργαλεία. Τα στοιχεία αυτά οδηγούν στο συμπέρασμα ότι τα κριτήρια κοινοποίησης και πληροφόρησης αποτελούν τα αδύνατα σημεία των εργαλείων αποκλεισμού των μηχανισμών web tracking [118].

Κατά καιρούς, βέβαια, έχουν εμφανιστεί εργαλεία που προσπαθούν να αντιμετωπίσουν ορισμένα από τα προαναφερθέντα ζητήματα χωρίς ιδιαίτερη επιτυχία. Παράδειγμα αποτελεί το project Cookie Clearinghouse, μια online ιδιωτική πρωτοβουλία του Centre for Internet & Society της Νομικής Σχολής του Στάνφορντ σε συνεργασία με το Mozilla [119]. Σε αντίθεση με άλλα εργαλεία αποκλεισμού των μηχανισμών web tracking, το Cookie Clearinghouse διέθετε μια «ανοικτή» διαδικασία σύνθεσης των κατηγοριών της λίστας. Επίσης, δινόταν η δυνατότητα στους χρήστες να αναφέρουν οποιαδήποτε περίπτωση είναι διαφορετική από τη συνηθισμένη, ενώ παράλληλα οι φορείς που αναφέρονταν ως trackers θα είχαν τη δυνατότητα να παρουσιάσουν στοιχεία που να αποδεικνύουν το αντίθετο. Το συγκεκριμένο project διέκοψε την εξέλιξη του [120], [121].

Λειτουργία ιδιωτικής περιήγησης

Η λειτουργία ιδιωτικής περιήγησης (Private browsing feature) είναι μια ρύθμιση του προγράμματος περιήγησης, η οποία αποσκοπεί στη διευκόλυνση των χρηστών στην απόκρυψη της ταυτότητάς τους από τις ιστοσελίδες που επισκέπτονται και δεν επιτρέπει στις ιστοσελίδες να «αφήσουν ίχνη» (π.χ. cookies) στα τερματικά των χρηστών. Τα πέντε πιο δημοφιλή προγράμματα περιήγησης (Mozilla Firefox, Internet Explorer, Google Chrome, Opera και Safari) διαθέτουν τέτοια λειτουργία περιήγησης. Οι Xianyi Gao και συν. εξέτασαν τις ανακολουθίες μεταξύ των στόχων και της υλοποίησης των λειτουργιών ιδιωτικής περιήγησης [122]. Στη συγκεκριμένη μελέτη αποδείχθηκε ότι κανένα από τα πέντε προγράμματα περιήγησης δεν μπόρεσε να εγγυηθεί ότι οι ιστοσελίδες δεν ήταν σε θέση να καθορίσουν αν το πρόγραμμα περιήγησης βρισκόταν σε κατάσταση ιδιωτικής λειτουργίας. Αυτό σημαίνει ότι η ιστοσελίδα μπορούσε να λάβει τουλάχιστον κάποιες πληροφορίες που αφορούσαν το χρήστη, γεγονός που σημαίνει ότι το κριτήριο του επιπέδου προστασίας δεν ικανοποιείται.

Σε πιο πρόσφατη μελέτη, οι Rodrigo de S. Ruiz και συν. κατέληξαν στο ίδιο συμπέρασμα [123]. Οι συγγραφείς, μετά από δοκιμές που διενέργησαν, διαπίστωσαν ότι όλα τα προγράμματα περιήγησης που έλεγξαν παρουσιάζουν ελαττώματα στη λειτουργία ιδιωτικής περιήγησής τους. Τα ελαττώματα αυτά αφορούν την παραγωγή δεδομένων που παραμένουν διαθέσιμα στο σύστημα και επιτρέπουν όχι μόνο τον εντοπισμό των ιστοσελίδων που έχουν επισκεφθεί οι χρήστες, αλλά σε ορισμένες περιπτώσεις ακόμα και την αναδημιουργία ολόκληρου του ιστορικού περιήγησής τους. Έτσι, ένα άλλο ζήτημα που προκύπτει σχετικά με τη χρήση της λειτουργίας ιδιωτικής περιήγησης είναι ότι το απόρρητο των χρηστών, όπως αυτό διαφημίζεται, ουσιαστικά δεν παρέχεται.

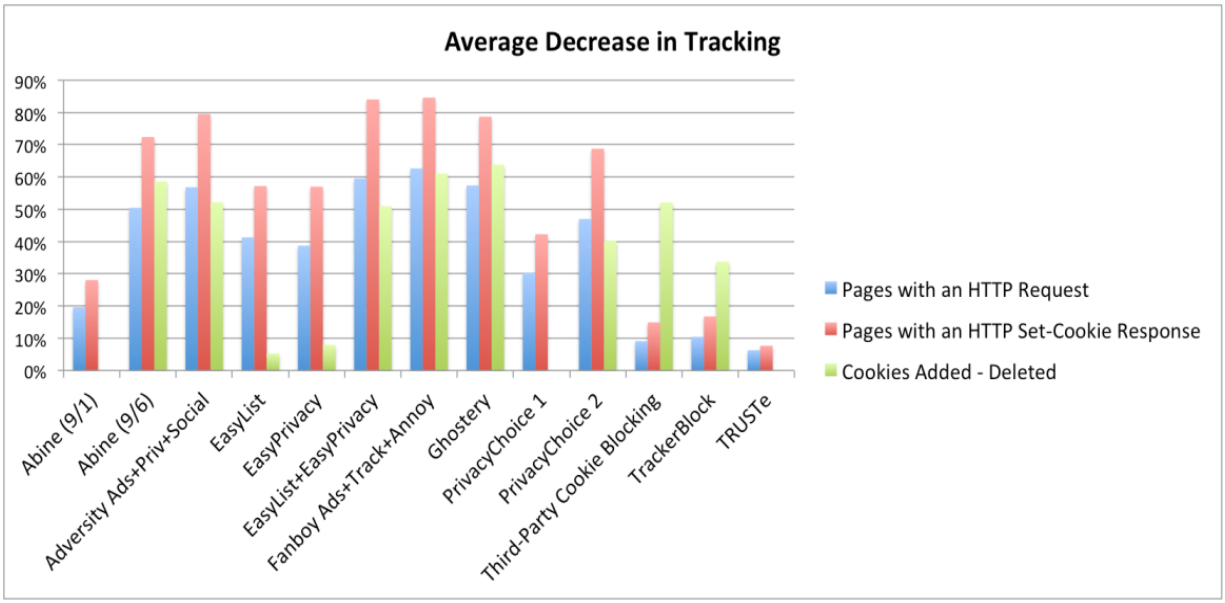
Ο Mozilla Firefox περιλαμβάνει ένα είδος αποποίησης ευθυνών, για την αποφυγή οποιασδήποτε παρανόησης σχετικά με την ουσιαστικά αποτελεσματική δράση της λειτουργίας ιδιωτικής περιήγησης [124]: *«Προειδοποίηση: Η λειτουργία Ιδιωτικής Περιήγησης δε σας κάνει ανώνυμο στο διαδίκτυο. Ο πάροχός σας, ο εργοδότης σας, ή καθαυτές οι ιστοσελίδες που επισκέπτεσθε μπορούν να ανιχνεύσουν σε ποιες σελίδες έχετε περιηγηθεί. Επίσης, η λειτουργία Ιδιωτικής Περιήγησης δε σας προστατεύει από keyloggers ή spyware που μπορεί να εγκατασταθούν στον υπολογιστή σας.»*

Επιπλέον, η λειτουργία ιδιωτικής περιήγησης προϋποθέτει ότι ο χρήστης έχει ήδη επίγνωση της δραστηριότητας των μηχανισμών web tracking, καθώς δεν ειδοποιεί το χρήστη ότι υπάρχει

ο κίνδυνος παρακολούθησης της online δραστηριότητάς του. Επίσης, απαιτεί ενεργό χειρισμό από μεριάς χρήστη, ο οποίος θα πρέπει να επιλέξει την εν λόγω λειτουργία [126]. Η συμμόρφωση με την επιθυμία των χρηστών να μην παρακολουθούνται δεν μπορεί να εφαρμοστεί (κριτήριο εφαρμοσιμότητας). Οι φορείς παρακολούθησης μπορούν να χρησιμοποιήσουν τις τεχνικές ατέλειες στο σχεδιασμό του συγκεκριμένου μέτρου αντιμετώπισης των μηχανισμών web tracking και να συλλέξουν πληροφορίες σχετικά με τις προτιμήσεις ιδιωτικότητας των χρηστών χωρίς ιδιαίτερη δυσκολία. Στην πραγματικότητα, η χρήση της λειτουργίας ιδιωτικής πλοήγησης οδηγεί σε μια ψευδαίσθηση της ιδιωτικότητας, παρόμοια με την ολοκληρωτική ή ελλιπή κοινοποίηση του προβλήματος, καθώς και στις δύο περιπτώσεις, ο χρήστης περιηγείται στο διαδίκτυο, χωρίς να λάβει κανένα μέτρο προστασίας των δικαιωμάτων του.

Αξιολόγηση των εργαλείων αντιμετώπισης μηχανισμών web tracking

Ο J. Mayer διεξήγαγε μια εμπειρική έρευνα για την αποτελεσματικότητα ορισμένων εργαλείων ως μέτρα αντιμετώπισης των μηχανισμών web tracking και παρατήρησε μία μείωση στην δραστηριότητα παρακολούθησης λόγω της χρήσης των εργαλείων αυτών (Εικ. 21) [126]. Το συμπέρασμα αυτό είναι σίγουρα ένα θετικό αποτέλεσμα για τους χρήστες, καθώς αποδεικνύεται ότι τα εργαλεία αυτά προσφέρουν μια κάποια λύση κατά της ανεπιθύμητης παρακολούθησης της online συμπεριφοράς τους.



Εικόνα 21: Μείωση της online παρακολούθησης από τη χρήση εργαλείων αντιμετώπισης των μηχανισμών web tracking (Πηγή: Jonathan Mayer)[21]

Ωστόσο, τα εργαλεία αντιμετώπισης των μηχανισμών web tracking από μόνα τους δεν μπορούν να διαχειριστούν όλα τα ζητήματα σε σχέση με την online παρακολούθηση. Όπως συζητήθηκε σε προηγούμενες ενότητες, τα εργαλεία αυτά ασχολούνται μόνο με έναν αριθμό φορέων παρακολούθησης, δεν προσφέρουν πλήρη προστασία στους χρήστες που δεν επιθυμούν να παρακολουθούνται, δεν παρέχουν πάντοτε λεπτομέρειες σχετικά με τη λειτουργία τους και είναι προϊόντα ιδιωτικών εταιρειών. Επιπλέον, ένα άλλο σημαντικό ζήτημα είναι η παρακολούθηση από τρίτους. Οι B. Krishnamurthy και C. Wills [29] επισήμαναν ότι ο

[21] http://cyberlaw.stanford.edu/blog/2011/09/tracking-trackers-self-help-tools

εντοπισμός και ο αποκλεισμός του περιεχομένου τρίτων είναι δύσκολος λόγω του γεγονότος ότι τα τρίτα μέρη χρησιμοποιούν τεχνολογίες όπως Javascript για να διαβάζουν και να χρησιμοποιούν τα cookies των ίδιων των δικτυακών τόπων ή χρησιμοποιούν άλλους τρόπους για να εμφανίζονται ως first party οντότητες. Τα ζητήματα αυτά καθιστούν τα εργαλεία αντιμετώπισης των μηχανισμών web tracking ανεπαρκή στο να χειριστούν από μόνα τους το θέμα της έλλειψης επιλογής και ελέγχου από τους χρήστες όσον αφορά την online παρακολούθησή τους.

6.2.2 Αυτορρύθμιση

Η αυτορρύθμιση (self-regulation) είναι υποκινούμενη, καταρχάς, από την απαίτηση συμμόρφωσης με την επικείμενη νομοθεσία και κατά δεύτερο λόγο από το ενδιαφέρον των οντοτήτων αυτορρύθμισης να βελτιώσουν την εικόνα τους στους καταναλωτές. Επιπλέον, υπάρχουν περιπτώσεις κατά τις οποίες όποτε στο καταναλωτικό κοινό υπάρχει αυξανόμενη ανησυχία για ένα συγκεκριμένο θέμα, οι εταιρείες είναι πρόθυμες να αποδείξουν ότι έχουν επίγνωση του θέματος και λαμβάνουν εύλογα μέτρα χειρισμού του. Μια τέτοια περίπτωση αποτελεί το web tracking και η δημιουργία προφίλ των χρηστών με βάση την online συμπεριφορά τους. Η αυξανόμενη ανησυχία των χρηστών για την παραβίαση της ιδιωτικότητάς τους και για την προστασία των προσωπικών τους δεδομένων, έχει αποτελέσει κίνητρο για μεγάλες εταιρείες, όπως το Pinterest και το Twitter, να συμμετάσχουν από μόνες τους στην αντιμετώπιση του ζητήματος της προστασίας της ιδιωτικότητας των χρηστών είτε με την έκδοση κανόνων και πολιτικών ή ανακοινώνοντας δημόσια τη δέσμευσή τους να συμμορφώνονται με τους κανόνες άλλων πρωτοβουλιών, όπως η πρωτοβουλία Do Not Track, που θα συζητηθεί αργότερα σε αυτό το κεφάλαιο [127].

Το 2011, η Ευρωπαϊκή Ένωση Προτύπων Διαφήμισης (European Advertising Standards Alliance - EASA) και η ο Οργανισμός Διαδραστικής Επικοινωνίας (Interactive Advertising Bureau - IAB) εξέδωσαν μια σύσταση βέλτιστων πρακτικών για την επιγραμμική συμπεριφορική διαφήμιση (Best Practice Recommendation on Online Behavioral Advertising - EASA/IAB Code) [128]. Σκοπός του κώδικα EASA/IAB είναι η προσφορά υποστήριξης και συμβουλών σχετικά με την πρακτική διαφημιστική αυτορρύθμιση. Εκτός από συστάσεις, ο κώδικας EASA/IAB εισήγαγε μια ιστοσελίδα και ένα μηχανισμό, ο οποίος χαρακτηρίστηκε ως *«μηχανισμός υποστήριξης των χρηστών του διαδικτύου για την άσκηση της επιλογής τους σχετικά με τη συλλογή και χρήση δεδομένων για σκοπούς συμπεριφορικής διαφήμισης στο διαδίκτυο από ένα ή περισσότερα τρίτα μέρη ή συνδέσεις με ένα μηχανισμό που επιτρέπει την επιλογή του χρήστη πάνω στη συμπεριφορική διαφήμιση στο διαδίκτυο»*. Η επιλογή του χρήστη σχετικά με τη συλλογή προσωπικών δεδομένων, εφαρμόζεται μέσω της εγκατάστασης του λεγόμενου opt-out cookie [125], το οποίο στέλνει την προτίμηση της μη συλλογής προσωπικών δεδομένων και της μη λήψης στοχευμένων διαφημίσεων με βάση τα προσωπικά δεδομένα, στις διαφημιστικές εταιρείες.

Ωστόσο, η Γνώμη 16/2011 (Opinion 16/2011), σχετικά με τη σύσταση βέλτιστων πρακτικών EASA/IAB για την επιγραμμική συμπεριφορική διαφήμιση του άρθρου 29 για την προστασία των δεδομένων (Article 29 Working Party) βρήκε ότι ο κώδικας δεν συμμορφώνεται με την τροποποιημένη οδηγία «e-Privacy» [129]. Ένας από τους λόγους που αναφέρονται είναι ότι το opt-out cookie, μπορεί εύκολα να ανασχεδιαστεί και ως εκ τούτου να παρακάμψει την προτίμηση του χρήστη. Η Γνώμη καταλήγει στο συμπέρασμα ότι ο κώδικας μπορεί να

δημιουργήσει τη ψευδαίσθηση στους χρήστες ότι δεν παρακολουθούνται ενώ βρίσκονται στο διαδίκτυο και στη βιομηχανία διαφήμισης ότι συμμορφώνεται με το νόμο.

Από τα παραπάνω προκύπτει το συμπέρασμα ότι η αυτορρύθμιση είναι πολύτιμη, καθώς το μέρος (first ή third party), που δημιουργεί ζήτημα με πιθανές αρνητικές συνέπειες, όταν αυτορυθμίζεται, τότε αναγνωρίζει την ύπαρξη αυτού του ζητήματος και την ανάγκη για καθορισμό κανόνων που να το ρυθμίζουν και να εξαλείφουν ή τουλάχιστον να μειώνουν τις συνέπειες αυτές. Στην περίπτωση του web tracking ως διαδικασία δημιουργίας προφίλ της online συμπεριφοράς των χρηστών οι φορείς, που προωθούν και επωφελούνται από την επιγραμμική συμπεριφορική διαφήμιση, αναγνωρίζουν μέσω της αυτορρύθμισης την ύπαρξη των αρνητικών πτυχών της δραστηριότητάς τους αυτής, δηλαδή την ανεπιθύμητη παρακολούθηση, την παραβίαση της προστασίας των προσωπικών δεδομένων, κ.λπ. Επιπλέον, με τους κώδικες δεοντολογίας και τις οδηγίες βέλτιστων πρακτικών, αναγνωρίζουν την ανάγκη ρύθμισης των αρνητικών αυτών πτυχών, γεγονός που σίγουρα αποτελεί σημαντικό βήμα προς μια κάποια λύση του προβλήματος. Η πρόθεση της αυτορρύθμισης μπορεί να χρησιμοποιηθεί ως μοχλός πίεσης όλων των ενδιαφερόμενων μερών σε διάλογο για μια αποτελεσματική λύση στο πρόβλημα, όπως θα συζητηθεί αργότερα στο παρόν κεφάλαιο.

Ένα άλλο σημαντικό πλεονέκτημα της αυτορρύθμισης είναι ότι δεν περιορίζεται από γεωγραφικά όρια, πράγμα το οποίο συμβαίνει με τους νόμους (π.χ. διαφορετικές δικαιοδοσίες ανά κράτος) [130]. Βέβαια, δεν θα πρέπει να διαφεύγει το γεγονός ότι η αυτορρύθμιση δεν αποτελεί από μόνη της λύση στο πρόβλημα της αθέμιτης παρακολούθησης των χρηστών μέσω των μηχανισμών web tracking. Συχνά, η αυτορρύθμιση δεν περιέχει μηχανισμούς ελέγχου της συμμόρφωσης ή παρουσιάζει ελλείψεις κατά το σχεδιασμό της. Ως εκ τούτου, η υλοποίησή της θα μπορούσε να προσφέρει απλά μια ψευδαίσθηση ιδιωτικότητας, στην οποία ο χρήστης πιστεύει λανθασμένα ότι δεν παρακολουθείται. Η κατάσταση αυτή ισοδυναμεί με άγνοια και έλλειψη ενημέρωσης των χρηστών σχετικά με τη διαδικασία της παρακολούθησης και συνεπάγεται ένα εξίσου υψηλό κίνδυνο για την προστασία της ιδιωτικότητας και των προσωπικών δεδομένων των χρηστών. Αυτό που είναι επίσης αμφισβητήσιμο, είναι η παροχή κοινοποίησης και σχετικών πληροφοριών (αντίστοιχα κριτήρια) καθώς και η ποιότητα των πληροφοριών αυτών.

6.2.3 Πρωτοβουλία Do Not Track

Η πρωτοβουλία «Μη Παρακολούθησης» (Do Not Track - DNT) ξεκίνησε στις ΗΠΑ. Η απουσία νομοθεσίας για την προστασία των προσωπικών δεδομένων οδήγησε στην ανάπτυξη αυτής της πολιτικής με σκοπό το σεβασμό της ιδιωτικότητας των χρηστών από τους φορείς παρακολούθησης στο διαδίκτυο και τη παροχή δυνατότητας στους χρήστες να αποφασίζουν για το αν επιθυμούν να παρακολουθούνται. Η πρωτοβουλία DNT θα μπορούσε να περιλαμβάνεται στις δύο προαναφερθείσες κατηγορίες μέτρων αντιμετώπισης των μηχανισμών web tracking, όμως λόγω της σημασίας της πρωτοβουλίας και της δυνατότητάς της να προσφέρει ουσιαστική λύση στο πρόβλημα της ανεπιθύμητης παρακολούθησης στο διαδίκτυο, εξετάζεται ως ξεχωριστή κατηγορία.

Ο μηχανισμός DNT είναι μια ρύθμιση του προγράμματος περιήγησης που προσφέρει στους χρήστες του την επιλογή να επιτρέπουν (ή όχι) την παρακολούθηση των online δραστηριοτήτων τους. Η κύρια ιδέα είναι ότι το πρόγραμμα περιήγησης μπορεί να επιτρέπει προστασία της ιδιωτικότητας, προσφέροντας έτσι τη δυνατότητα στο χρήστη να επιλέξει προστασία από κάθε

μηχανισμό web tracking, οντότητα παρακολούθησης και όλα αυτά χωρίς εδαφικούς περιορισμούς. Ο μηχανισμός στέλνει μια κεφαλίδα στην οντότητα παρακολούθησης ανάλογα με την επιθυμία παρακολούθησης του χρήστη. Η συγκεκριμένη κεφαλίδα DNT μπορεί να περιέχει τρεις τιμές [132]:

- «1», δηλώνει την επιθυμία του χρήστη να μην παρακολουθείται,
- «2», σημαίνει ότι ο χρήστης επιθυμεί να παρακολουθείται και
- «null» σημαίνει ότι ο χρήστης δεν έχει ορίσει την προτίμησή του

Η πρωτοβουλία DNT είναι στενά συνδεδεμένη με την προσέγγιση της προστασίας της ιδιωτικότητας από τον σχεδιασμό (privacy by design), σύμφωνα με την οποία, η προστασία της ιδιωτικότητας θα πρέπει να είναι ενσωματωμένη στις προδιαγραφές σχεδιασμού των τεχνολογιών των πληροφοριών, των επιχειρηματικών πρακτικών και των δικτυακών υποδομών [133]. Η προσέγγιση αυτή απαιτεί τη μετάβαση από την κατάσταση αντίδρασης στην κατάσταση πρόληψης. Τον Ιανουάριο του 2016, το Ευρωπαϊκό Κοινοβούλιο οριστικοποίησε τον κοινοτικό κανονισμό για την προστασία των δεδομένων (General Data Protection Regulation) που περιλαμβάνει την προσέγγιση privacy by design ως μια από τις βασικές αρχές της προστασίας των δεδομένων [134].

Με την εφαρμογή της πρωτοβουλίας DNT δημιουργήθηκε ένα ζήτημα που αφορά το ποια θα πρέπει να είναι η προεπιλεγμένη ρύθμισή του. Το ζήτημα ξεκίνησε όταν η Microsoft έθεσε την τιμή «1» (δεν επιτρέπω την παρακολούθηση) ως προεπιλεγμένη ρύθμιση απορρήτου του Internet Explorer 10 [135]. Οι υποστηρικτές της προσέγγισης απορρήτου από προεπιλογή (privacy by default), υποστηρίζουν ότι η προτίμηση Do Not Track (τιμή «1»), θα πρέπει να αποτελεί προεπιλεγμένη ρύθμιση σε ένα πρόγραμμα περιήγησης, προκειμένου να προστατευθούν οι χρήστες που δεν έχουν επίγνωση της διαδικασίας του web tracking. Από την άλλη πλευρά, κατά της ιδιωτικότητας από προεπιλογή είναι τα δίκτυα διαφήμισης και οι διαφημιστές, οι οποίοι βλέπουν τη ρύθμιση αυτή ως μια σημαντική απώλεια κέρδους. Οι χρήστες, συνήθως, τείνουν να είναι απρόθυμοι στην αλλαγή των προεπιλεγμένων ρυθμίσεων του οποιουδήποτε εξοπλισμού τους είτε για λόγους ασφαλείας είτε επειδή απλά δεν ξέρουν πώς να το κάνουν [136]. Ως αποτέλεσμα, τα δίκτυα διαφήμισης δεν θα έχουν την ευκαιρία να προσεγγίσουν υποψήφιους καταναλωτές, οι οποίοι θα μπορούσαν να ενδιαφέρονται για προσαρμοσμένη διαφήμιση. Παρά το γεγονός ότι και οι δύο πλευρές έχουν λογικά επιχειρήματα, φαίνεται ότι το απόρρητο από προεπιλογή κερδίζει έδαφος, τουλάχιστον στην Ευρωπαϊκή Ένωση, αφού περιλαμβάνεται στον κανονισμό για την προστασία των δεδομένων ως βασική αρχή επεξεργασίας δεδομένων [134].

Η συμμόρφωση των φορέων παρακολούθησης με την DNT προτίμηση των χρηστών δεν δείχνει να εφαρμόζεται (κριτήριο εφαρμοσιμότητας). Η συμμόρφωση αυτή εξαρτάται από τις ίδιες τις εταιρείες και η μη εφαρμοσιμότητά της δείχνει να οφείλεται σε μεγάλο βαθμό και από την ποικιλία υλοποίησης της πρωτοβουλίας DNT στα διάφορα προγράμματα περιήγησης. Το πλέον χαρακτηριστικό παράδειγμα αποτελεί ο προαναφερθείς Microsoft Internet Explorer 10 και η αντίδραση των εταιρειών στις προεπιλεγμένες ρυθμίσεις απορρήτου του. Χαρακτηριστικότατη αντίδραση ήταν ο σχεδιασμός ενός patch για τον διαδικτυακό διακομιστή Apache, τον πιο κοινά χρησιμοποιούμενο διακομιστή σε ιστοσελίδες, το οποίο απενεργοποιεί την προεπιλεγμένη ρύθμιση του DNT αν το πρόγραμμα πλοήγησης του χρήστη είναι ο Microsoft Internet Explorer 10 [137]. Η κατάσταση αυτή προέτρεψε τους υποστηρικτές του

DNT να κινήσουν διαδικασίες τεχνικής προτυποποίησης της προσέγγισης, κάτι που θα εξασφάλιζε δύο βασικούς παράγοντες:

- τη δημιουργία ενός κοινού χαρακτηριστικού του μηχανισμού DNT από όλα τα προγράμματα περιήγησης, και
- τη συμμόρφωση της διαφημιστικής βιομηχανίας και διάφορων άλλων ενδιαφερόμενων μερών με το πρότυπο DNT, δεδομένου ότι θα συμμετέχουν στην διαδικασία προτυποποίησης της προσέγγισης

Οι διαδικασίες έχουν ξεκινήσει από το Σεπτέμβρη του 2011 και παρά την ευρεία συμμετοχή και την θετική ένδειξη για ένα κοινώς αποδεκτό αποτέλεσμα, η εξέλιξή τους κινείται τόσο αργά ώστε η οριστικοποίηση ενός συμφωνημένου προτύπου τίθεται υπό αμφισβήτηση [138]. Ο λόγος είναι η διαφωνία των συμμετεχόντων σε κρίσιμα θέματα, τα οποία θα καθορίσουν αν το πρότυπο θα διευκολύνει την εύρεση μιας λύσης για το υπό συζήτηση θέμα ή θα είναι τελικά μια ατελής προσπάθεια [139].

Ανεξάρτητα από την εξέλιξη του DNT ως προτύπου, η ιδέα ενός τυποποιημένου μηχανισμού ρύθμισης των προτιμήσεων των χρηστών έχει πολλές θετικές πτυχές. Η διαδικασία αυτή περιλαμβάνει κυρίως τη συμμετοχή όλων των φορέων που επωφελούνται από το web tracking, ενώ η δέσμευσή τους να σεβαστούν την προτίμηση του χρήστη λειτουργεί ως ένα είδος μηχανισμού «αυτό-εφαρμοσιμότητας». Ένα τέτοιο πρότυπο υπερνικά τους περιορισμούς των λιστών προστασίας από την παρακολούθηση, με βάση το κριτήριο επίπεδου προστασίας, καθώς δεν απαιτεί από το χρήστη να ενημερώνει τον μηχανισμό ή δεν εξαρτάται από την τακτική ενημέρωση των λιστών από το φορέα διαχείρισής τους. Ο χρήστης απλά θέτει την επιθυμία του κατά των μηχανισμών web tracking. Επιπλέον, ένα πρότυπο μπορεί δυνητικά να είναι εφαρμόσιμο σε παγκόσμιο επίπεδο.

6.3 Γενική αξιολόγηση των μέτρων αντιμετώπισης των μηχανισμών web tracking

Στο κεφάλαιο αυτό, παρουσιάστηκαν κάποια από τα κυριότερα μέτρα αντιμετώπισης των μηχανισμών web tracking. Η επιμέρους αξιολόγησή τους έγινε με βάση τέσσερα κριτήρια:

- την κοινοποίηση στον χρήστη και την ευαισθητοποίηση του,
- την παροχή πληροφοριών
- το επίπεδο της προστασίας του χρήστη, ο οποίος δεν επιθυμεί να παρακολουθείται και
- την εφαρμοσιμότητα αυτής της προτίμησης του χρήστη.

Με βάση τα όσα συζητήθηκαν στις προηγούμενες ενότητες, μία από τις πιο προβληματικές περιοχές αποδείχθηκε ότι είναι η κοινοποίηση στον χρήστη. Η ευαισθητοποίησή του σχετικά με τη δραστηριότητα των μηχανισμών παρακολούθησης είναι το πρώτο βήμα για την επιλογή του. Το κριτήριο της κοινοποίησης αποτελεί βασική προϋπόθεση της Οδηγίας Προστασίας Προσωπικών Δεδομένων (e-Privacy Directive) και δεν θεωρείται μηχανισμός εξαίρεσης (opt out). Τα pop-up παράθυρα, για παράδειγμα, που εμφανίζονται όταν οι χρήστες επισκέπτονται μια ιστοσελίδα είναι ένας άμεσος και φιλικός προς το χρήστη τρόπος με τον οποίο ειδοποιείται ότι παρακολουθείται. Αντίθετα τα εργαλεία και η αυτορρύθμιση ως μέτρα αντιμετώπισης των μηχανισμών web tracking δεν κοινοποιούν στους χρήστες τη διαδικασία της παρακολούθησης. Ειδικά τα εργαλεία εμπλοκής των μηχανισμών web tracking αλλά και η κεφαλίδα DNT είναι

μέτρα αντιμετώπισης που προϋποθέτουν την κοινοποίηση στον χρήστη και την ευαισθητοποίηση του.

Όσον αφορά το κριτήριο της παροχής πληροφοριών οι οποίες θα πρέπει να είναι επαρκείς για να επιτρέπουν μια συνειδητή επιλογή του χρήστη, η κατάσταση φαίνεται να είναι καλύτερη. Ο χρήστης, ο οποίος έχει ήδη επίγνωση της δραστηριότητας της παρακολούθησης, μπορεί να βρει κάποιες πληροφορίες σχετικά με την ταυτότητα του tracker, τον τύπο του μηχανισμού παρακολούθησης που χρησιμοποιεί ή ακόμα και τον σκοπό της παρακολούθησης αν αναζητήσει τα στοιχεία αυτά σε κάποιο εργαλείο αποκλεισμού ή σε μια λίστα προστασίας από την παρακολούθηση.

Ωστόσο, πηγή των πληροφοριών αυτών αποτελούν είτε οι πάροχοι των ιστοσελίδων (οι οποίοι αποτελούν μέρος της διαφημιστικής βιομηχανίας του διαδικτύου) είτε άλλα μέρη, όπως οι οντότητες ανάπτυξης των εργαλείων αποκλεισμού. Το γεγονός αυτό δημιουργεί δύο κινδύνους:

- οι πληροφορίες αυτές μπορεί να είναι ανακριβείς ή ψευδείς
- οι πληροφορίες είναι ανεπαρκείς.

Ως εκ τούτου, εάν ένας χρήστης επιλέξει να παρακολουθείται και να λαμβάνει στοχευμένες διαφημίσεις βασιζόμενος στις πληροφορίες που παρέχονται από αυτά τα μέτρα αντιμετώπισης των μηχανισμών web tracking, τότε η επιλογή του μπορεί να μην θεωρηθεί έγκυρη.

Η εξέταση των πρωτοβουλιών που δεν επικεντρώνονται στον αποκλεισμό και την εξαίρεση συγκεκριμένων φορέων και μηχανισμών παρακολούθησης αλλά που εκφράζουν με γενικό τρόπο την προτίμηση των χρηστών όσον αφορά την παρακολούθηση της online συμπεριφοράς τους, προσφέρουν ένα υψηλότερο επίπεδο προστασίας. Κάτι τέτοιο δεν πρέπει να αποτελεί έκπληξη, λαμβάνοντας υπόψη το γεγονός ότι στον ανταγωνισμό μεταξύ των προγραμματιστών που εργάζονται για τους φορείς παρακολούθησης και αυτούς που δημιουργούν εργαλεία αποκλεισμού των φορέων αυτών, οι πρώτοι, συνήθως, βρίσκονται ένα βήμα μπροστά.

Όσον αφορά το κριτήριο της εφαρμοσιμότητας, τα περισσότερα από τα ισχύοντα μέτρα αντιμετώπισης των μηχανισμών web tracking δεν δείχνουν να επιτυγχάνουν. Η λειτουργία τους δείχνει να βασίζεται στον εθελοντικό σεβασμό των φορέων παρακολούθησης στην προτίμηση του χρήστη. Ο μηχανισμός DNT δεν έδειξε να μπορεί να εφαρμοστεί. Απλά ενημερώνει το αντίστοιχο πεδίο της κεφαλίδας των αιτημάτων HTTP σχετικά με το τι θέλει ο χρήστης. Αντίθετα τα εργαλεία αποτροπής των μηχανισμών web tracking δείχνουν να είμαι άμεσα εφαρμόσιμα. Εμποδίζουν, για παράδειγμα, συγκεκριμένα cookies παρακολούθησης, χωρίς άλλες πρόσθετες διαδικασίες, αιτήματα ή δικαιώματα.

Σε γενικές γραμμές, τα μέτρα αντιμετώπισης των μηχανισμών web tracking δεν ανταπεξέρχονται με επιτυχία στο πρόβλημα της ελεύθερης επιλογής του χρήστη που αφορά την παρακολούθησή του μέσω της online συμπεριφοράς του. Καθένα από τα μέτρα αυτά παρουσιάζει δυνατά και αδύνατα σημεία. Για τα αδύνατα σημεία ορισμένων μέτρων υπάρχει προοπτική βελτίωσης (π.χ. η δυνατότητα εφαρμοσιμότητας). Επιπλέον, μερικά από τα ισχυρά σημεία που παρουσιάζουν, κάποια από τα μέτρα αυτά θα μπορούσαν να συνδυαστούν για να επιτευχθεί ένα αποτέλεσμα πιο κοντά στο επιθυμητό.

7 Web Tracking: Αναζητώντας Λύση

Σύμφωνα με τα όσα συζητήθηκαν στα προηγούμενα κεφάλαια, εφικτή λύση στο ζήτημα του web tracking θα μπορούσε να υπάρξει μέσω της προσέγγισης της επιλογής του ίδιου του χρήστη του διαδικτύου σχετικά με την παρακολούθησή της online συμπεριφοράς του. Για να υλοποιηθεί μια τέτοια προσέγγιση θα πρέπει να πληρούνται τα κριτήρια που αναφέρθηκαν στο προηγούμενο κεφάλαιο: κοινοποίηση των δραστηριοτήτων παρακολούθησης, επαρκείς πληροφορίες που να επιτρέπουν τη συνειδητή επιλογή, προστασία ενάντια σε όλες τις οντότητες και τους μηχανισμούς παρακολούθησης καθώς και δυνατότητα εφαρμοσιμότητας.

Μια βραχυπρόθεσμη λύση στο ζήτημα της παρακολούθησης των χρηστών του διαδικτύου μέσω μηχανισμών web tracking θα μπορούσε να αποτελέσει η βελτίωση του τρόπου αποκόμισης της συγκατάθεσης του χρήστη, όπως αυτός παρουσιάζεται στην παράγραφο 3 του Άρθρου 5 της Οδηγίας του e-Privacy, μέσω της προτυποποίησής της [140]. Μια μακροπρόθεσμη λύση θα μπορούσε να είναι η εφαρμογή της προσέγγισης της ιδιωτικότητας από σχεδιασμό (privacy by design) μέσω ενσωμάτωσης των εγγυήσεων προστασίας της ιδιωτικότητας των χρηστών στα αρχικά στάδια σχεδιασμού ενός προϊόντος ή μιας υπηρεσίας.

Σκοπός του τελευταίου αυτού κεφαλαίου είναι η ανάπτυξη κάποιων ιδεών σχετικά με το τι θα μπορούσε να γίνει ώστε να προστατευθεί η ιδιωτικότητα των χρηστών από τους μηχανισμούς web tracking. Ουσιαστικά θα γίνει μια παρουσίαση των δύο προαναφερθέντων λύσεων: της βραχυπρόθεσμης λύσης μέσω προτυποποίησης των όσων αναφέρει η παράγραφος 3 του Άρθρου 5 της Οδηγίας του e-Privacy και της μακροπρόθεσμης λύσης μέσω εφαρμογής της προσέγγισης της ιδιωτικότητας από σχεδιασμό.

7.1 Τυποποίηση του μηχανισμού ρητής συγκατάθεσης του χρήστη

Η Οδηγία 2009/136/ΕΕ περί δικαιωμάτων των πολιτών (2009/136/EC Citizens Rights Directive) άλλαξε το ήδη υπάρχον καθεστώς της εξαίρεσης (opt-out) εισάγοντας το καθεστώς της ρητής εντολής συγκατάθεσης (opt-in) [141]. Ο μηχανισμός opt-in έχει τη δυνατότητα να εκπληρώσει τα κριτήρια της κοινοποίησης και πληροφόρησης, καθώς οι οντότητες παρακολούθησης θα πρέπει να ενημερώνουν τους χρήστες για τη χρήση μηχανισμών web tracking και να συμμορφώνονται με τις απαιτήσεις για τη ρητή συγκατάθεσή τους παρέχοντάς τους όλες τις απαραίτητες πληροφορίες που θα διευκολύνουν την τεκμηριωμένη επιλογή τους. Ωστόσο, η εφαρμογή του μηχανισμού opt-in είναι προβληματική επειδή πρώτον, οι πάροχοι ιστοσελίδων ερμηνεύουν με διαφορετικό τρόπο τις απαιτήσεις της Οδηγίας «e-Privacy» και, δεύτερον, το επίπεδο εναρμόνισης της νομοθεσίας μεταξύ των κρατών-μελών είναι χαμηλό.

Επομένως μια λύση στο ζήτημα της εφαρμογής του μηχανισμού opt-in θα ήταν η προτυποποίησή του, η οποία θα αποτελούσε και βραχυπρόθεσμη λύση του ζητήματος της παρακολούθησης της online συμπεριφοράς των χρηστών. Με μια τέτοια τυποποίηση ο χρήστης δεν θα αναγκαζόταν, για παράδειγμα, να εκτίθεται σε διαφορετικά είδη παραθύρων pop-up αλλά θα μπορούσε να δέχεται διαφημίσεις με βάση προκαθορισμένες σε τεχνικό πρότυπο μεθόδους, συμβατές με τις απαιτήσεις της Οδηγίας «e-Privacy». Ένα ευρωπαϊκό τεχνικό πρότυπο δημιουργούμενο από τις αρμόδιες Ευρωπαϊκές Επιτροπές Τυποποίησης (CEN), θα μπορούσε να συμβάλει στην εναρμόνιση της εφαρμογής του μηχανισμού opt-in από τους παρόχους των ιστοσελίδων [142]. Το πρότυπο αυτό θα πρέπει να περιγράφει τους τρόπους με τους οποίους η συγκατάθεση του χρήστη για online παρακολούθηση θα πρέπει να ζητείται με ειδικές τεχνικές απαιτήσεις. Θα πρέπει επίσης να περιγράφει τον τύπο των πληροφοριών που θα

πρέπει να παρέχονται στο χρήστη (φορέας παρακολούθησης, μηχανισμός και χρησιμοποιούμενη τεχνολογία, συγκεκριμένος σκοπός χρήσης των συλλεγόμενων μέσω παρακολούθησης πληροφοριών ή στοιχείων, κλπ.). Επίσης θα πρέπει να ορίζεται και κάποιος συγκεκριμένος τρόπος αποδοχής της online παρακολούθησης από τους χρήστες, ο οποίος θα πρέπει να είναι φιλικός προς αυτούς, χωρίς να εμποδίζεται η περιήγηση στις ιστοσελίδες. Σκοπός δεν είναι να γίνει η περιήγηση δυσάρεστη, αλλά να διευκολύνεται η επιλογή για παρακολούθηση, αν αυτό είναι επιθυμητό από τον χρήστη.

Η Ευρωπαϊκή Ένωση (ΕΕ) θα μπορούσε να στείλει ένα αίτημα προτυποποίησης (mandated standard) της όλης διαδικασίας προς την Ευρωπαϊκή Επιτροπή Τυποποίησης (CEN), σύμφωνα με τη διαδικασία που προβλέπεται από το άρθρο. 10 του κανονισμού 1025/2012 (Regulation 1025/2012/EU), γεγονός που θα εξασφάλιζε την παρακολούθηση των δραστηριοτήτων τυποποίησης από την ΕΕ [143]. Τέτοια αιτήματα υποστηρίζονται από την ευρωπαϊκή νομοθεσία, που στην περίπτωση του web tracking είναι οι Οδηγίες «e-Privacy»και προστασίας των προσωπικών δεδομένων. Την ίδια στιγμή, οι φορείς που θα κληθούν να συμμορφωθούν με το πρότυπο, δεν αποκλείονται από τη διαδικασία προτυποποίησης καθώς αυτοί είναι που συμμετέχουν στις Τεχνικές Επιτροπές, όπως ακριβώς συμβαίνει άλλωστε και με κάθε διαδικασία προτυποποίησης κάποιου Ευρωπαϊκού Προτύπου. Μια τέτοια κίνηση της ΕΕ, δηλαδή η αποστολή αιτήματος προτυποποίησης που να αφορά το μηχανισμό ρητής συγκατάθεσης του χρήση, θα μπορούσε να εξασφαλίσει τα εξής:

- ότι η ΕΕ καθορίζει τους στόχους και τηρεί τη διαδικασία
- οι φορείς παρακολούθησης θα συμμετέχουν στη διαδικασία προτυποποίησης ενός προτύπου που θα χρησιμοποιήσουν

Ένα επιπλέον πλεονέκτημα μιας τέτοιας λύσης είναι ότι οι φορείς παρακολούθησης είναι ελεύθεροι να χρησιμοποιήσουν οποιαδήποτε άλλη μέθοδο που να συνάδει με το καθεστώς opt-in, κάτι που δεν απαγορεύεται στα Ευρωπαϊκά Πρότυπα, γεγονός που θα αποτελεί απόδειξη συμμόρφωσης με τις απαιτήσεις του καθεστώτος της ρητής εντολής συγκατάθεσης.

Αυτός που θα ωφεληθεί περισσότερο από την αποστολή ενός αιτήματος τυποποίησης θα είναι ο χρήστης του διαδικτύου, καθώς ένας εναρμονισμένος μηχανισμός opt-in θα καθιστούσε ευκολότερη την έκφραση της επιλογής του ώστε να δώσει (ή όχι), τη συναίνεσή του. Κατά την επίσκεψή τους στις διάφορες ιστοσελίδες, οι χρήστες θα έχουν να κάνουν με μια συγκεκριμένη μέθοδο (π.χ. pop-up window), αντί να προσπαθούν να κατανοήσουν τους διαφορετικούς μηχανισμούς opt-in σε κάθε ιστοσελίδα και να βρουν τις σχετικές πληροφορίες για το πώς θα δεχτούν ή θα απορρίψουν τη διαδικασία παρακολούθησής τους.

Σε γενικές γραμμές, ένα πρότυπο σαν αυτό θα ήταν επωφελές για κάθε ενδιαφερόμενο μέρος.

7.2 Ιδιωτικότητα από σχεδιασμό

Θα ήταν σημαντικό οι προγραμματιστές να λαμβάνουν υπόψη την ιδιωτικότητα των χρηστών από τα πρώτα στάδια σχεδιασμού ενός προϊόντος ή μιας υπηρεσίας. Η διαδικασία «δράση-αντίδραση», δηλαδή η εγκατάσταση κάποιου εργαλείου εμπλοκής των μηχανισμών web tracking αφού ο χρήστης έχει ήδη αποτελέσει αντικείμενο παρακολούθησης της online συμπεριφοράς του, αποτελεί μερική μόνο λύση του προβλήματος. Με τον τρόπο αυτό μένει

ανοιχτό το ενδεχόμενο, ότι ο χρήστης παρακολουθείται για μεγάλο χρονικό διάστημα, χωρίς να το γνωρίζει ή και να το επιθυμεί.

Η πρωτοβουλία DNT, όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, ήδη υποστηρίζει την ιδέα της προστασίας της ιδιωτικότητας από τον σχεδιασμό, με την προϋπόθεση ότι οι μηχανές αναζήτησης έχουν ενσωματωμένο τον μηχανισμό αυτόν. Ένα άλλο μέτρο που συμμερίζεται την ιδέα της προστασίας της ιδιωτικότητας από το σχεδιασμό είναι οι Τεχνολογίες Ενίσχυσης της Ιδιωτικότητας (Privacy Enhancing Technologies - PET) [144], όπως τα λογισμικά κρυπτογράφησης, τα anonymizer και οι επεκτάσεις των προγραμμάτων περιήγησης [145].

Η Ann Cavoukian έχει υποστηρίξει ότι η προστασία της ιδιωτικότητας από σχεδιασμό, προχωρεί πέρα από τις τεχνολογίες PET: *«τα οποία είναι απαραίτητα μεν εργαλεία ενδυνάμωσης των χρηστών ενάντια στους μηχανισμούς παρακολούθησης, αποτελούν όμως μόνο ένα κομμάτι ενός ευρύτερου πλαισίου το οποίο θα πρέπει να λαμβάνεται υπόψη όταν γίνεται συζήτηση για το πώς η τεχνολογία μπορεί να χρησιμοποιηθεί στην υπηρεσία της προστασίας της ιδιωτικότητας των χρηστών»* [146]. Η πρώην επίτροπος Πληροφορίας και Προστασίας Προσωπικών Δεδομένων του Καναδά, μία από τους πρώτους υποστηρικτές της προστασίας της ιδιωτικότητας από τον σχεδιασμό, έχει θέσει επτά θεμελιώδεις αρχές, που διέπουν την πρωτοβουλία αυτή, μεταξύ των οποίων είναι ο προληπτικός και όχι ο αντιδραστικός τρόπος δράσης, η προστασία της ιδιωτικότητας ως προεπιλεγμένη ρύθμιση, η ενσωμάτωση της προστασίας της ιδιωτικότητας στο σχεδιασμό και η υιοθέτηση της ιδιωτικότητας από το σχεδιασμό με επίκεντρο τον χρήστη [147]. Η Πρόταση για νέο Κανονισμό του Ευρωπαϊκού Κοινοβουλίου σε θέματα προστασίας των ατόμων για την επεξεργασία προσωπικών δεδομένων και την ελεύθερη διακίνηση τους (General Data Protection Regulation) έχει ήδη υιοθετήσει την ιδέα της προστασίας της ιδιωτικότητας από τον σχεδιασμό και την προστασία της ιδιωτικότητας από προεπιλογή, γεγονός που δείχνει μια τάση ενίσχυσης της προστασίας των δεδομένων στην ΕΕ [148].

8 Συμπεράσματα

Σκοπός της διατριβής αυτής ήταν η έρευνα και η καταγραφή των σύγχρονων μεθόδων web tracking, η παρουσίαση των στόχων και των τρόπων λειτουργίας τους αλλά και η ανάλυση των τρόπων αντιμετώπισής τους.

Οι μέθοδοι web tracking που παρουσιάστηκαν έχουν διαφορετικά πεδία εφαρμογής και σκοπούς. Αποδείχθηκε ότι με την πάροδο του χρόνου οι μέθοδοι αυτές εξελίχθηκαν από φιλικές προς την ιδιωτικότητα του χρήστη, όπως τα HTTP cookies, τα οποία αφαιρούνται εύκολα από τους χρήστες στην περίπτωση που δεν είναι επιθυμητά, σε πραγματικές απειλές της ιδιωτικότητας και της προστασίας των προσωπικών τους δεδομένων στο διαδίκτυο. Μάλιστα, κάποιες από τις πιο πρόσφατες τεχνικές παρακολούθησης, όπως το fingerprinting, είναι αδύνατο να εμποδιστούν από τους μέσους χρήστες και αποδεικνύονται αρκετά επιβλαβείς παραβιάζοντας τα δικαιώματά τους.

Ωστόσο, κατά τα τελευταία χρόνια, έχει παρατηρηθεί μια ανάλογη αύξηση ανάπτυξης εργαλείων, υπηρεσιών και πρωτοβουλιών που έχουν ως στόχο την αντιμετώπιση των μηχανισμών παρακολούθησης των χρηστών στο διαδίκτυο. Στα πλαίσια της παρούσας διατριβής έγινε μια περιγραφή των προσεγγίσεων αντιμετώπισης των μεθόδων web tracking και μια αξιολόγησή τους με βάση τέσσερα κριτήρια που τέθηκαν. Αποδείχθηκε ότι οι σύγχρονες προσπάθειες αντιμετώπισης των μηχανισμών web tracking είναι ανεπαρκείς για την πλήρη προστασία του χρήστη από την ανεπιθύμητη παρακολούθηση της διαδικτυακής συμπεριφοράς του.

Παρόλα αυτά, φαίνεται ότι η προστασία της ιδιωτικότητας και η επιγραμμική συμπεριφορική παρακολούθηση είναι δύο έννοιες που γίνεται να συνυπάρξουν υπό προϋποθέσεις. Η συνύπαρξη αυτή είναι εφικτή όταν η προστασία της ιδιωτικότητας και των προσωπικών δεδομένων των χρηστών αποτελέσει εγγενές στοιχείο του σχεδιασμού των προϊόντων, των υπηρεσιών και των χαρακτηριστικών του διαδικτύου, στοιχεία τα οποία μπορεί να διευκολύνουν τη συνειδητή επιλογή παρακολούθησης της διαδικτυακής συμπεριφοράς.

Βιβλιογραφία

[1] Arslan Aziz, Rahul Telang "What is a Cookie Worth?", Carnegie Mellon University, Mar. 2015, http://www.law.northwestern.edu/research-faculty/searlecenter/events/internet/documents/Telang_What%20is%20Cookie%20Worth_III.pdf

[2] Thorin Klosowski "How Web Sites Vary Prices Based on Your Information (and What You Can Do About It)", lifehacker.com, July 2013, http://lifehacker.com/5973689/how-web-sites-vary-prices-based-on-your-information-and-what-you-can-do-about-it

[3] Leslie Scism, Mark Maremont "Insurers Test Data Profiles to Identify Risky Clients", The Wall Street Journal - Europe Edition, Nov. 2010, http://www.wsj.com/news/articles/SB10001424052748704648604575620750998072986

[4] Deborah Preston "Affiliate Networks and Cost Per Sale (CPS) Explained", Linkedin.com, June 2015, https://www.linkedin.com/pulse/affiliate-networks-cost-per-sale-cps-explained-deborah-preston

[5] Niklas Schmücker "Web Tracking", Berlin University of Technology, SNET2 Seminar Paper - Summer Term 2011, http://www.snet.tu-berlin.de/fileadmin/fg220/courses/SS11/snet-project/web-tracking_schmuecker.pdf

[6] Tomasz Bujlow, Valentín Carela-Español, JosepSolé-Pareta, Pere Barlet-Ros "Web Tracking: Mechanisms, Implications, and Defenses", Broadband Communications Research Group, Department of Computer Architecture, Universitat Politècnica de Catalunya, Barcelona, July 2015, http://arxiv.org/pdf/1507.07872.pdf

[7] Arvind Narayanan "There is no such thing as anonymous online tracking", Stanford Center for Internet and Society, July 2011, http://cyberlaw.stanford.edu/blog/2011/07/there-no-such-thing-anonymous-online-tracking

[8] Károly Boda, Ádám Máté Földes, Gábor György Gulyás, Sándor Imre "Tracking and Fingerprinting in E-Business: New Storageless Technologies and Countermeasures", Research and Development in E-Business through Service-Oriented Solutions, Chapter 7, pp: 134-166, IGI Global, 2013, ISBN: 978-1-466-64182-2

[9] Avinash Chugh, Jon Mountjoy "Session Tracking", WebLogic: The Definitive Guide, Chapter 2, pp: 45-49, O'Reilly Media, Inc., 2004, ISBN: 978-0-596-55225-1

[10] Sonal Mittal "User privacy and the evolution of third-party tracking mechanisms on the world wide web", Thesis, Harvard University, Law School, May 2010, https://stacks.stanford.edu/file/druid:hw648fn9717/SonalMittal_Thesis.pdf

[11] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, Claudia Diaz "The Web never forgets: Persistent tracking mechanisms in the wild", Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp: 674-689, Nov. 2014, https://securehomes.esat.kuleuven.be/~gacar/persistent/the_web_never_forgets.pdf

[12] International Working Group on Data Protection in Telecommunications "Web Tracking and Privacy: Respect for context, transparency and control remains essential", Working paper, Apr. 2013, http://www.datenschutz-berlin.de/attachments/949/675.46.13.pdf

[13] Gabi Nakibly, GiladShelef, ShiranYudilevich "Hardware Fingerprinting Using HTML5", Computer Science Department, Technion – Israel Institute of Technology, Mar. 2015, http://arxiv.org/pdf/1503.01408.pdf

[14] Allaboutcookies.org "Mobile Technology Tracking Methods other than cookies", http://www.allaboutcookies.org/mobile/mobile-tracking.html

[15] Google Developers "Cookie Matching", Real-Time Bidding Protocol, https://developers.google.com/ad-exchange/rtb/cookie-guide#background

[16] Mireille Hildebrandt, Antoinette Rouvroy "Law, Human Agency and Autonomic Computing: The Philosophy of Law Meets the Philosophy of Technology", Routledge, 2011, ISBN: 978-0-415-59323-6

[17] John Schwartz "Tracks in Cyberspace: Giving the Web a Memory Cost Its Users Privacy", NYTimes.com, Sep. 2001, http://www.nytimes.com/2001/09/04/technology/04COOK.html

[18] RFC 2109 "HTTP State Management Mechanism", Feb. 1997, https://tools.ietf.org/html/rfc2109

[19] W3C "Same Origin Policy", https://www.w3.org/Security/wiki/Same_Origin_Policy

[20] International Working Group on Data Protection in Telecommunications "Essentials for privacy-enhancing technologies (e.g. P3P) on the WorldWideWeb", Common Position, Apr. 2013, www.datenschutz-berlin.de/attachments/178/priv_en.pdf

[21] W3C "Platform for Privacy Preferences (P3P) Project", https://www.w3.org/P3P/

[22] Marshall Sponder "Measurement 3.0 (in the next 5 years) OMMA Global – day 2", Webmetricsguru INC., Sep. 2008, http://www.webmetricsguru.com/archives/2008/09/measurement-30-on-the-next-5-years-omma-global-day-2/#sthash.0bIPrJyZ.dpuf

[23] M. Malheiros, C. Jennett, S. Patel, S. Brostoff, M.A. Sasse "Too Close for Comfort: A Study of the Effectiveness and Acceptability of Rich-Media Personalized Advertising", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp: 579-588, May 2012, http://discovery.ucl.ac.uk/1349989/1/Malheiros_et_al._-_2012_-_Too_Close_for_Comfort_A_Study_of_the_Effectivenes-preprint.pdf

[24] B. Zamanzadeh, N. Ashish, C. Ramakrishnan, J. Zimmerman "Semantic Advertising", Computer Science: Artificial Intelligence, Cornell University Library, Sep. 2013, http://arxiv.org/ftp/arxiv/papers/1309/1309.5018.pdf

[25] The Guardian "Gmail does scan all emails, new Google terms clarify", Apr. 2014, https://www.theguardian.com/technology/2014/apr/15/gmail-scans-all-emails-new-google-terms-clarify

[26] Google "Όροι Παροχής Υπηρεσιών της Google", Google Απόρρητο και όροι, Απρ. 2014, http://www.google.gr/policies/terms/regional.html

[27] GresiSanje, IsilSenol "The Importance of Online Behavioral Advertising for Online Retailers", International Journal of Business and Social Science, Vol. 3, No. 18, pp: 114-121, Sep. 2012, http://ijbssnet.com/journals/Vol_3_No_18_Special_Issue_September_2012/13.pdf

[28] A. Goldfarb, C. Tucker "Online advertising, behavioral targeting, and privacy", Communications of the ACM, Vol. 54, No. 5, pp: 25–27, May 2011, http://portal.acm.org/citation.cfm?id=1941498

[29] B. Krishnamurthy, C.E. Wills "On the leakage of personally identifiable information via online social networks", Proceedings of the 2nd ACM workshop on Online social networks, pp: 7-12, Aug. 2009, http://conferences.sigcomm.org/sigcomm/2009/workshops/wosn/papers/p7.pdf

[30] Wikipedia "Compensation methods", https://en.wikipedia.org/wiki/Compensation_methods

[31] Techopedia "Affiliate Tracking", https://www.techopedia.com/definition/23106/affiliate-tracking

[32] Smart Insights "Selling through Amazon and eBay in 2014", Dec. 2013, http://www.smartinsights.com/marketplace-analysis/partner-analysis/selling-amazon-ebay-2014/

[33] Wikipedia "Click fraud", https://en.wikipedia.org/wiki/Click_fraud

[34] Wikipedia "List of advertising networks", https://secure.wikimedia.org/wikipedia/en/wiki/List_of_advertising_networks

[35] J. Mikians, L. Gyarmati, V. Erramilli, N. Laoutaris "Detecting price and search discrimination on the internet", 11th ACM Workshop on Hot Topics in Networks (Hotnets'12), New York, Seattle, Washington, USA, pp: 79-84, Oct. 2012, http://conferences.sigcomm.org/hotnets/2012/papers/hotnets12-final94.pdf

[36] Tony Mecia "Credit card issuers watch online how you shop, customize offers", CreditCard.com, 2011, http://www.creditcards.com/credit-card-news/credit-card-issuers-watching-online-shopping-offers-1273.php

[37] Emily Steel, Julia Angwin "On the Web's Cutting Edge, Anonymity in Name Only", The Wall Street Journal, Aug. 2010, http://www.wsj.com/articles/SB10001424052748703294904575385532109190198

[38] Phil Villarreal "Capital One Made Me Different Loan Offers Depending On Which Browser I Used", Consumerist, Nov. 2010, https://consumerist.com/2010/11/01/capital-one-made-me-different-loan-offers-depending-on-which-browser-i-used/

[39] Christopher Elliott "Are online travel agencies quoting higher rates because of your Web cookies?", Feb. 2009, http://elliott.org/blog/are-online-travel-agencies-quoting-higher-rates-because-of-your-web-cookies/

[40] Dana Mattioli "On Orbitz, Mac Users Steered to Pricier Hotels", The Wall Street Journal, Aug. 2012, http://www.wsj.com/news/articles/SB10001424052702304458604577488822667325882

[41] The Economist ""Insurance data: Very personal finance", June 2012, http://www.economist.com/node/21556263

[42] Wikipedia "Clickstream", https://en.wikipedia.org/wiki/Clickstream

[43] Google "Google Analytics", http://www.google.com/intl/el/analytics/

[44] Luis A. Leiva "Diverse Contributions to Implicit Human-Computer Interaction", Phd Thesis on Philosophy in Computer Science, Universitat Politècnica de Valencia, Departament de Sistemes Informatics i Computacio, Nov. 2012, http://personales.upv.es/luileito/phd/thesis.pdf

[45] Wikipedia "Eye tracking", https://en.wikipedia.org/wiki/Eye_tracking

[46] Google "Αιτήματα αποκάλυψης στοιχείων χρηστών: Ελλάδα", Google Ασφάλεια και απόρρητο, Ιαν. 2016, https://www.google.com/transparencyreport/userdatarequests/GR/

[47] Rich McCormick "Google scans everyone's email for child porn", The Verge, Aug. 2014, http://www.theverge.com/2014/8/5/5970141/how-google-scans-your-gmail-for-child-porn

[48] S. Englehardt, D. Reisman, C. Eubank, P. Zimmerman, J. Mayer, A. Narayanan, E.W. Felten, "Cookies That Give You Away: The Surveillance Implications of Web Tracking", Proceedings of the 24th International Conference on World Wide Web, pp: 289-299, May 2015, http://senglehardt.com/papers/www15_cookie_surveil.pdf

[49] B. Elgin, V. Silver "The Surveillance Market and Its Victims", Bloomberg Business, Dec. 2011, http://www.bloomberg.com/data-visualization/wired-for-repression/

[50] The Guardian "Tor Stinks", Oct. 2013, http://www.theguardian.com/world/interactive/2013/oct/04/tor-stinks-nsa-presentation-document

[51] A. Soltani, A. Peterson, B. Gellman "NSA uses Google cookies to pinpoint targets for hacking", The Washington Post, Dec. 2013, http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/nsa-uses-google-cookies-to-pinpoint-targets-for-hacking

[52] R. Gallagher "Operation Socialist: The Inside Story of How British Spies Hacked Belgium's Largest Telco", The Intercept, Dec. 2014, https://firstlook.org/theintercept/2014/12/13/belgacom-hack-gchq-inside-story/

[53] M. Lee "Secret "BADASS" Intelligence Program Spied on Smartphones", The Intercept, Jan. 2015, https://firstlook.org/theintercept/2015/01/26/secret-badass-spy-program/

[54] A. Acquisti, R. Gross, F. Stutzman "Face Recognition and Privacy in the Age of Augmented Reality", Journal of Privacy and Condentiality, Vol. 6, No. 2, pp: 1- 20, 2014, http://repository.cmu.edu/cgi/viewcontent.cgi?article=1122&context=jpc

[55] Business Wire "Identity Fraud Rose 13 Percent in 2011 According to New Javelin Strategy & Research Report", Feb. 2012, http://www.businesswire.com/news/home/20120222005485/en/Identity-Fraud-Rose-13-Percent-2011-Javelin

[56] Ashlee Kieler "Identity Theft: Could It Be Your GPS' Fault? Probably Not, But Maybe", Consumerist, Jan. 2014, https://consumerist.com/2014/01/09/identity-theft-could-it-be-your-gps-fault-probably-not-but-maybe/

[57] Latanya Sweeney "Computational Disclosure Control: A Primer on Data Privacy Protection", Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Massachusetts, USA, 2001, http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/sweeney-thesis-draft.pdf

[58] Paul Ohm "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization", University of Colorado Law School, Vol. 57, pp: 1701-1777, 2010, http://paulohm.com/classes/techpriv13/reading/wednesday/OhmBrokenPromisesofPrivacy.pdf

[59] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281, 23/11/1995 P. 0031 – 0050, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[60] Arvind Narayanan “There is no such thing as anonymous online tracking”, The Center for Internet and Society (CIS), Stanford Law School, July 2011, http://cyberlaw.stanford.edu/blog/2011/07/there-no-such-thing-anonymous-online-tracking

[61] Google “Τρόποι διασφάλισης του απορρήτου από τη χρήση του κουμπιού +1”, Google+ Help, https://support.google.com/plus/answer/1319578?hl=el

[62] Facebook “Τι δεδομένα λαμβάνει το Facebook όταν επισκέπτομαι ιστότοπους που διαθέτουν το κουμπί "Like";”, Κέντρο Βοήθειας, https://www.facebook.com/help/186325668085084

[63] Ian Paul “Advocacy Groups Ask Facebook for More Privacy Changes”, PC World, http://www.pcworld.com/article/199099/facebook_privacy_fixes.html

[64] Stephanie Bodoni, John Martens “Belgium Tells Facebook to Stop Storing Personal Data from Non-Users”, Bloomberg Business, Nov. 2015, http://www.bloomberg.com/news/articles/2015-11-09/facebook-told-to-stop-storing-personal-data-from-belgian-surfers

[65] Simon Rice “Does your website have a leak?”, Information Commissioner’s Office Blog, Sep. 2015, https://iconewsblog.wordpress.com/2015/09/16/does-your-website-have-a-leak/

[66] Thomas Frank “JSON stringify revisited”, thomasfrank.se, July 2006, http://www.thomasfrank.se/json_stringify_revisited.html

[67] Wikipedia “HTTP cookie”, https://en.wikipedia.org/wiki/HTTP_cookie

[68] allaboutcookies.org “All About Cookies”, http://www.allaboutcookies.org/cookies/

[69] Lori MacVittie “Cookies, Sessions, and Persistence”, White Paper, F5 Networks, Inc., July 2008, https://f5.com/resources/white-papers/cookies-sessions-and-persistence

[70] Lorrie Faith Cranor, Manya Sleeper, Blasé Ur “Tracking and Surveillance”, Privacy and Information Technology, Chapter 5, IAPP Privacy and Information Technology text book, 2013, http://www.blaseur.com/papers/trackingsurveillance-draft.pdf

[71] Article 29 Data Protection Working Party “Opinion 4/2012 on Cookie Consent Exemption”, WP 194, June 2012, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp194_en.pdf

[72] A.M. McDonald “Footprints Near the Surf: Individual Privacy Decisions in Online Contexts”, Dissertation, Carnegie Mellon University, Research Showcase Carnegie Mellon University, Dec. 2010, http://repository.cmu.edu/cgi/viewcontent.cgi?article=1008&context=dissertations

[73] Jeff Chester “Cookie Wars: How New Data Profiling and Targeting Techniques Threaten Citizens and Consumers in the “Big Data” Era”, Chapter 4, European Data Protection: In Good Health?, pp: 53-77, Feb. 2012, Springer Netherlands, ISBN: 978-9-400-72902-5

[74] Pedro Giovanni Leon, Lorrie Faith Cranor, Aleecia M. McDonald, Robert McGuire “Token Attempt: The Misrepresentation of Website Privacy Policies through the Misuse of P3P Compact Policy Tokens”, CyLab, Carnegie Mellon University, Pittsburgh, USA, Sep. 2010, https://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab10014.pdf

[75] Ibrahim Altaweel, Nathaniel Good, Chris Jay Hoofnagle “Web Privacy Census”, Technology Science, Dec. 2015, http://techscience.org/a/2015121502/

[76] Wikipedia “Web beacon”, https://en.wikipedia.org/wiki/Web_beacon

[77] Shwetank Dixit “Web Storage: Easier, More Powerful Client-Side Data Storage”, Opera Software ASA, Mar. 2013, https://dev.opera.com/articles/web-storage/

[78] AndyHume.net “localStorage is not cookies”, Mar. 2011, http://blog.andyhume.net/localstorage-is-not-cookies/

[79] R. Tirtea, C. Castelluccia, D. Ikonomou “Bittersweet cookies. Some security and privacy considerations”, Technical report, ENISA European Network and Information Security Agency, Feb. 2011, https://www.enisa.europa.eu/activities/identity-and-trust/library/pp/cookies

[80] P. Verleg, M. van Eekelen, H. Vranken “Cache cookies: searching for hidden browser storage”, Bachelor thesis, Computer Science, Radboud University, June 2014, http://www.cs.ru.nl/bachelorscripties/2014/Patrick_Verleg___3049701___Cache_Cookies_searching_for_hidden_browser_storage.pdf

[81] W3C “Client-side Scripting and HTML”, WD-script-970314, W3C Working Draft, Mar. 1997, https://www.w3.org/TR/WD-script-970314

[82] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee “Hypertext Transfer Protocol -- HTTP/1.1”, June 1999, http://www.ietf.org/rfc/rfc2616.txt

[83] Nik Cubrilovic “Persistent and Unblockable Cookies Using HTTP Headers”, Aug. 2011, https://www.nikcub.com/posts/persistant-and-unblockable-cookies-using-http-headers/

[84] M. Ayenson, D.J. Wambach, A. Soltani, N. Good, C.J. Hoofnagle “Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning”, SSRN Electronic Journal, July 2011, https://fpf.org/wp-content/uploads/2011/07/Flash%20Cookies%20and%20Privacy%20II%20Now%20with%20HTML5%20and%20ETag%20Respawning.pdf

[85] C. Jackson, A. Barth, J. Hodges “HTTP Strict Transport Security (HSTS)”, RFC 6797, Nov. 2012, https://tools.ietf.org/html/rfc6797

[86] Mark Stockley “Anatomy of a browser dilemma – how HSTS ‘supercookies’ make you choose between privacy or security”, Naked Security, Feb. 2015, https://nakedsecurity.sophos.com/2015/02/02/anatomy-of-a-browser-dilemma-how-hsts-supercookies-make-you-choose-between-privacy-or-security/

[87] Lavasoft “Weird New Tricks Allow User Tracking in Firefox and Chrome”, http://www.lavasoft.com/mylavasoft/company/blog/weird-new-tricks-allow-user-tracking-in-firefox-and-chrome

[88] Leo Le Taro “New Methods for Targeted Advertising and User Tracking on the Internet”, Cryptography and Security, June 2015, https://hal.inria.fr/hal-01167493/document

[89] Jonathan Mayer “How Verizon’s Advertising Header Works”, Web Policy, Oct. 2014, http://webpolicy.org/2014/10/24/how-verizons-advertising-header-works/

[90] Nader Ammari, Gustaf Björksten, Peter Micek, Deji Olukotun “The Rise of Mobile Tracking Headers: How Telcos Around the World Are Threatening Your Privacy”, accessnow, Aug. 2015, https://www.accessnow.org/cms/assets/uploads/archive/AIBT-Report.pdf

[91] Peter Eckersley “How Unique Is Your Web Browser?”, Chapter 1, Privacy Enhancing Technologies, Vol. 6205 of the series Lecture Notes in Computer Science, Springer, pp: 1-18, 2010, https://panopticlick.eff.org/static/browser-uniqueness.pdf

[92] Ting-Fang Yen, Yinglian Xie, Fang Yu, Roger Peng Yu, Martin Abadi "Host Fingerprinting and Tracking on the Web: Privacy and Security Implications", Proceedings of the 19th Annual Network & Distributed System Security Symposium, Feb. 2012, http://research.microsoft.com/pubs/156901/ndss2012.pdf

[93] Wikipedia "User agent", https://en.wikipedia.org/wiki/User_agent

[94] Kris Jordan "Unacceptable Browser HTTP Accept Headers (Yes, You Safari and Internet Explorer)", New Media Campains, July 2009, http://www.newmediacampaigns.com/blog/browser-rest-http-accept-headers

[95] Christof Ferreira Torres "Fingerprint Privacy: A fresh perspective on web privacy", Thesis, Bachelor in Computer Science, Faculty of Science, Technology and Communication, University of Luxembourg, 2014, http://www.open.ou.nl/hjo/supervision/c.torres14-bsc-thesis.pdf

[96] Mozilla Developer Network "HTTP access control (CORS)", Feb. 2016, https://developer.mozilla.org/en-US/docs/Web/HTTP/Access_control_CORS

[97] Wade Alcorn, Christian Frichot, Michele Orru "The Browser Hacker's Handbook", John Wiley & Sons, Feb. 2014, ISBN: 978-1-118-91435-9

[98] Microsoft "Using Windows Internet Explorer Developer Tools Network Capture", https://msdn.microsoft.com/en-us/library/gg130952%28v=vs.85%29.aspx

[99] Mozilla Wiki "Fingerprinting", https://wiki.mozilla.org/Fingerprinting

[100] Experian Information Solutions, Inc "41st Parameter: Preventing fraud and enabling sales", http://www.experian.com/decision-analytics/41st-parameter.html

[101] Lalit Patel "JavaScript/CSS Font Detector", Mar. 2007, http://www.lalit.org/lab/javascript-css-font-detect/

[102] Keaton Mowery, Dillon Bogenreif, Scott Yilek, Hovav Shacham "Fingerprinting Information in JavaScript Implementations", Proceedings of W2SP 2011, IEEE Computer Society, May 2011, http://cseweb.ucsd.edu/~hovav/dist/jspriv.pdf

[103] Martin Mulazzani, Philipp Reschl, Markus Huber, Manuel Leithner, Sebastian Schrittwieser, Edgar Weippl "Fast and Reliable Browser Identication with JavaScript Engine Fingerprinting", Web 2.0 Security&Privacy 2013, San Francisco; May 2013, http://www.w2spconf.com/2013/papers/s2p1.pdf

[104] Keaton Mowery, Hovav Shacham "Pixel Perfect: Fingerprinting Canvas in HTML5", Proceedings of W2SP 2012, IEEE Computer Society, May 2012, https://cseweb.ucsd.edu/~hovav/dist/canvas.pdf

[105] Aaron Stuart "EFF launches Panopticlick 2 with new tracking and fingerprinting tests", WebSetNet, Dec. 2015, https://websetnet.com/eff-launches-panopticlick-2-tracking-fingerprinting-tests/

[106] W3C "Tracking Compliance and Scope", W3C Last Call Working Draft, July 2015, https://www.w3.org/TR/tracking-compliance/

[107] Bill Budington "Panopticlick 2.0 Launches, Featuring New Tracker Protection and Fingerprinting Tests", EFF, Dec. 2015, https://www.eff.org/deeplinks/2015/12/panopticlick-20-launches-featuring-new-tracker-protection-and-fingerprinting-tests

[108] Károly Boda, Ádám Máté Földes, Gábor György Gulyás, Sándor Imre "User Tracking on the Web via Cross-Browser Fingerprinting", Chapter 4, Information Security Technology for Applications, Vol. 7161 of the series Lecture Notes in Computer

Science, pp 31-46, Oct. 2011, https://pet-portal.eu/files/articles/2011/fingerprinting/cross-browser_fingerprinting.pdf

[109] Wikipedia "TCP/IP stack fingerprinting", https://en.wikipedia.org/wiki/TCP/IP_stack_fingerprinting

[110] NETRESEC "Passive OS Fingerprinting", Nov. 2011, http://www.netresec.com/?page=Blog&month=2011-11&post=Passive-OS-Fingerprinting

[111] Antonio Lopez "Web tracking of the Internet users", INCIBE, Dec. 2015, https://www.incibe.es/blogs/post/Security/SecurityBlog/Article_and_comments/web_tracking_en

[112] Brendan Coles "Network Discovery", Feb. 2016, GitHub Inc., https://github.com/beefproject/beef/wiki/Network-Discovery

[113] MaxMind, Inc. "GeoLite Legacy Downloadable Databases", http://dev.maxmind.com/geoip/legacy/geolite/

[114] Mozilla Developer Network "Using geolocation", https://developer.mozilla.org/en-US/docs/Web/API/Geolocation/Using_geolocation

[115] Wikipedia "Data Protection Directive", https://en.wikipedia.org/wiki/Data_Protection_Directive

[116] Wikipedia "Directive on Privacy and Electronic Communications", https://en.wikipedia.org/wiki/Directive_on_Privacy_and_Electronic_Communications

[117] Robert L. Mitchell "Ad tracking: Is anything being done?", Computerworld, Apr. 2014, http://www.computerworld.com.au/article/541921/ad_tracking_anything_being_done_/

[118] Lonneke van der Velden "The Third Party Diary: Tracking the trackers on Dutch governmental websites", Necsus, June 2015, http://www.necsus-ejms.org/third-party-diary-tracking-trackers-dutch-governmental-websites-2/

[119] Kate Kaye "Mozilla and Stanford Pitch New Cookie Blocking Approach", AdvertisingAge, June 2013, http://adage.com/article/dataworks/mozilla-stanford-pitch-cookie-blocking-approach/242553/

[120] Cookie Clearinghouse "How The Cookie Clearinghouse Works", June 2013, http://cch.law.stanford.edu/our-projects/

[121] WebTechReview "Cookie Clearinghouse pushes 'Do Not Track' addition step", Jan. 2016, http://webtechreview.com/cookie-clearinghouse-pushes-do-not-track-addition-step/

[122] Xianyi Gao, Yulong Yang, Huiqing Fu, Janne Lindqvist, Yang Wang "Private Browsing: An Inquiry on Usability and Privacy Protection", Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES '14), Noe. 2014, pp: 97-106, http://www.winlab.rutgers.edu/~janne/WPES14-privatebrowsing.pdf

[123] Rodrigo de S. Ruiz, Fernando Pompeo Amatte, Kil Jin Brandini Park D. Sc., Rogério Winter "Overconfidence: Personal Behaviors Regarding Privacy that Allows the Leakage of Information in Private Browsing Mode", International

Journal of Cyber-Security and Digital Forensics (IJCSDF), Vol. 4, Issue 3, pp: 404-416, The Society of Digital Information and Wireless Communications, 2015, http://sdiwc.us/digitlib/journal_paper.php?paper=00001410.pdf

[124] Mozilla Support "Firefox: Ιδιωτική περιήγηση", https://support.mozilla.org/el/kb/%CE%99%CE%B4%CE%B9%CF%89%CF%84%CE%B9%CE%BA%CE%AE%20%CF%80%CE%B5%CF%81%CE%B9%CE%AE%CE%B3%CE%B7%CF%83%CE%B7

[125] What is an Opt-Out cookie? http://www.allaboutcookies.org/manage-cookies/opt-out-cookies.html

[126] Jonathan Mayer "Tracking the Trackers: Self-Help Tools", Stanford Law School, The Center for Internet and Society, Sep. 2011, http://cyberlaw.stanford.edu/blog/2011/09/tracking-trackers-self-help-tools

[127] Rainey Reitman "Pinterest Commits to Respecting Do Not Track", EFF, July 2013, https://www.eff.org/deeplinks/2013/07/pinterest-commits-respecting-do-not-track

[128] European Advertising Standards Alliance "Comprehensive standards for Online Behavioural Advertising", Apr. 2011, http://www.easa-alliance.org/page.aspx/386

[129] ΟΜΑΔΑ ΕΡΓΑΣΙΑΣ ΤΟΥ ΑΡΘΡΟΥ 29 ΓΙΑ ΤΗΝ ΠΡΟΣΤΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ "Γνώμη 16/2011 σχετικά με τη σύσταση βέλτιστων πρακτικών EASA/IAB για την επιγραμμική συμπεριφορική διαφήμιση", Dec. 2011, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2011/wp188_el.pdf

[130] Giuditta De Prato, Daniel Nepelski "International patenting strategies in ICT", European IPR Helpdesk, 2013, https://www.fer.unizg.hr/_download/repository/JRC79479.pdf

[131] Wikipedia "Do Not Track legislation", https://en.wikipedia.org/wiki/Do_Not_Track_legislation

[132] How-To Geek "Why Enabling "Do Not Track" Doesn't Stop You From Being Tracked", http://www.howtogeek.com/126705/why-enabling-do-not-track-doesnt-stop-you-from-being-tracked/

[133] Ann Cavoukian "Privacy by Design in Law, Policy and Practice, A White Paper for Regulators, Decision-makers and Policy-makers", Information and Privacy Commissioner, Ontario, Canada, Aug. 2011, https://www.ipc.on.ca/images/resources/pbd-law-policy.pdf

[134] Sabba Mahmood, Leonie Power "Getting to know the General Data Protection Regulation, Part 6 – Designing for compliance", Field Fisher Waterhouse, Jan. 2016, http://privacylawblog.fieldfisher.com/2016/getting-to-know-the-general-data-protection-regulation-part-6-designing-for-compliance

[135] Chloe Albanesius "Internet Explorer 10 Released for Windows 7", PC Magazine Online, Nov. 2013, http://www.pcmag.com/article2/0,2817,2412077,00.asp

[136] Jared Spool "Do users change their settings?", User Interface Engineering, Sep. 2011, https://www.uie.com/brainsparks/2011/09/14/do-users-change-their-settings/

[137] Stephen Shankland "Apache Web software overrides IE10 do-not-track setting", CNET, Sep. 2012, http://www.cnet.com/news/apache-web-software-overrides-ie10-do-not-track-setting/

[138] Joel Hruska "Microsoft won't enable 'Do Not Track' in next-gen Spartan browser", ExtremeTech, Apr. 2015, http://www.extremetech.com/computing/202733-the-death-of-do-not-track-microsoft-wont-enable-feature-by-default-in-next-gen-spartan-browser

[139] Dawn Chmielewski "How 'Do Not Track' Ended Up Going Nowhere", Vox Media, Inc., Jan. 2016, http://recode.net/2016/01/04/how-do-not-track-ended-up-going-nowhere/

[140] European Commision "Cookies", Information Providers Guide: The EU Internet Handbook, http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm

[141] Cameron McKenna "Cookie consent: opt-in or opt-out?", Lexology, Feb. 2011, http://www.lexology.com/library/detail.aspx?g=835a5a2d-2e35-4951-a464-f0e774dbc263

[142] Wikipedia "European Committee for Standardization", https://en.wikipedia.org/wiki/European_Committee_for_Standardization

[143] CEN "Standardization requests (Mandates)", Guidance Documents, http://boss.cen.eu/reference%20material/Guidancedoc/Pages/Mandates.aspx

[144] Wikipedia "Privacy-enhancing technologies", https://en.wikipedia.org/wiki/Privacy-enhancing_technologies

[145] Center for Democracy & Technology "The Role of Privacy by Design in Protecting Consumer Privacy", Jan. 2010, https://cdt.org/insight/the-role-of-privacy-by-design-in-protecting-consumer-privacy-1/

[146] Henry Kenyon "Privacy by Design: Protect User Data From 'Get-Go'", Information Week, May 2015, http://www.informationweek.com/government/cybersecurity/privacy-by-design-protect-user-data-from-get-go/d/d-id/1318437

[147] Ann Cavoukian "Privacy by Design: The 7 Foundational Principles", Information and Privacy Commissioner of Ontario, Canada, Aug. 2009, https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

[148] DLA Piper "The EU General Data Protection Regulation: are you ready?", Publications, Dec. 2015, https://www.dlapiper.com/en/europe/insights/publications/2015/12/ipt-news-q4-2015/the-eu-general-data-protection/