

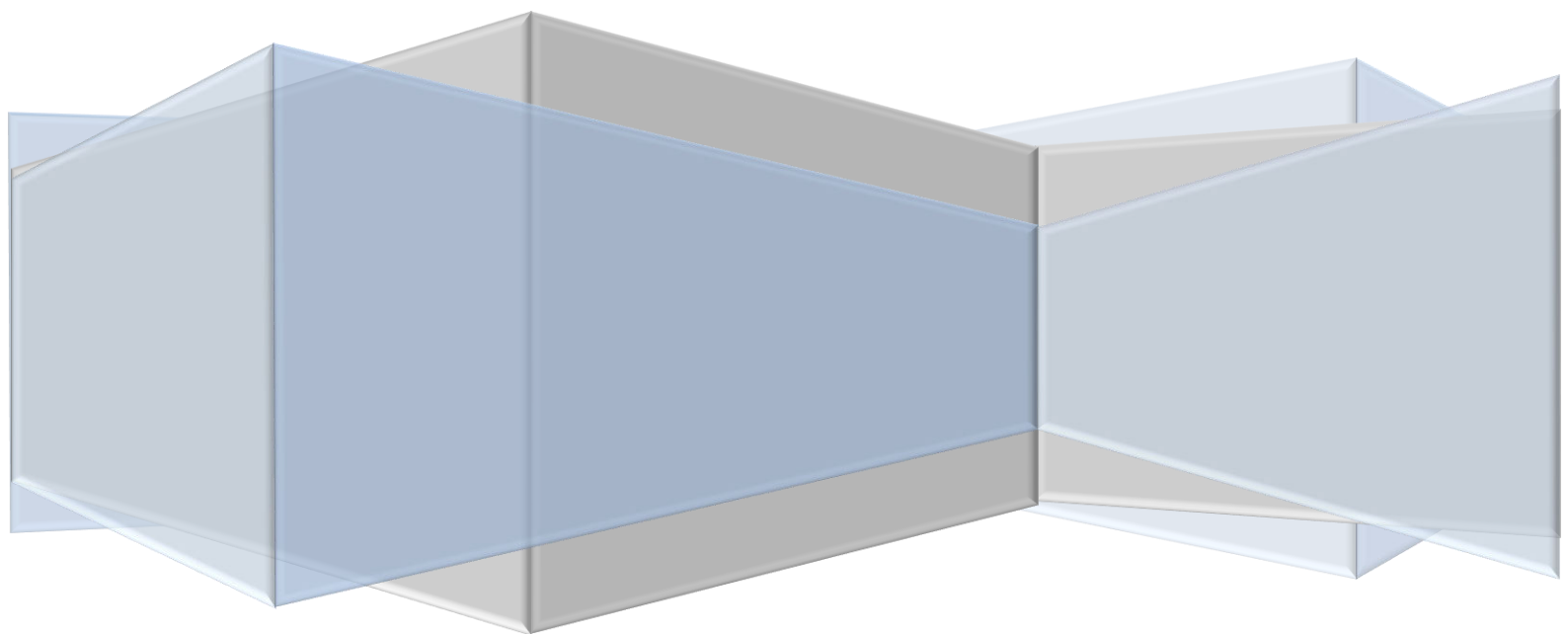
ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΤΑΣΟΥΛΑΣ ΘΕΟΦΑΝΗΣ

**Συστήματα παρακολούθησης και συλλογής
πληροφοριών χρηστών διαδικτύου**

Πτυχιακή Εργασία

Επιβλέπουσα καθηγήτρια: Τσώχου Αγγελική



Μάιος 2019

Περιεχόμενα

Πρόλογος.....	3
Κεφάλαιο 1 – Θεμελιώδεις έννοιες	
1.1) Συστήματα καταγραφής και παρακολούθησης δεδομένων.....	5
1.2) Τεχνολογικό υπόβαθρο.....	6
1.3) Νομικές πτυχές.....	7
1.4) Συστήματα ιχνηλάτησης σε εταιρείες.....	9
Κεφάλαιο 2 - Σύστημα καταγραφής και παρακολούθησης δεδομένων χρηστών διαδικτύου	
2.1) Cookies.....	11
2.2) HTML και HTML5.....	14
2.3) Web Beacons.....	16
2.4) Λογισμικό κατασκοπείας - Spyware.....	18
2.5) Παρακολούθηση ηλεκτρονικής αλληλογραφίας - Email Tracking.....	20
2.6) Υπηρεσίες εντοπισμού θέσης LBS.....	22
2.7) Εξακρίβωση ταυτότητας.....	25
2.8) Παρακολούθηση χρωματισμένων δεδομένων - Taint tracking	27
2.9) Ιχνηλάτηση των ιχνηλατών - Web Privacy Measurement.....	30
2.10) Άλλες μέθοδοι ιχνηλάτησης	32
Κεφάλαιο 3 - Συστήματα αντι-ιχνηλάτησης (anti-tracking)	
3.1) Μέθοδοι αποτροπής παρακολούθησης από τους περιηγητές ιστού.....	36
3.2) Μεσολαβητές επικοινωνίας.....	38
3.3) Αποκλεισμός κώδικα Script.....	40
3.4) Αποτροπή fingerprints.....	41
3.5) Άλλα συστήματα αποτροπής ιχνηλάτησης.....	41
Κεφάλαιο 4 - Σύστημα ταξινόμησης (classification scheme).....	44
Κεφάλαιο 5 - Ενημερότητα χρηστών του διαδικτύου σχετικά με τις τεχνολογίες παρακολούθησης και συλλογής πληροφοριών χρηστών διαδικτύου	
5.1) Σχεδίαση και ανάπτυξη ερωτηματολογίου.....	49
5.2) Εξαγωγή αποτελεσμάτων.....	57
Κεφάλαιο 6 - Συμπεράσματα και μελλοντική έρευνα.....	64
Βιβλιογραφικές και διαδικτυακές ερευνητικές πηγές.....	65

Πρόλογος

Στις μέρες μας οι τεχνολογίες παρακολούθησης και αυτόματης καταγραφής δεδομένων και συμπεριφορών των χρηστών διαδικτύου ανθίζουν σε μεγάλο βαθμό. Πολλές από αυτές είναι γνωστές και διαδεδομένες, όπως τα λογισμικά καταγραφής πλοήγησης (cookies) σε έναν φυλλομετρητή. Εντούτοις, με την άνοδο των τεχνολογικών γνώσεων (π.χ. γνώσεις προγραμματισμού) αναπτύσσονται διαρκώς προηγμένα συστήματα, όπως για παράδειγμα οι τεχνολογίες καταγραφής χρήσης ηλεκτρονικού ταχυδρομείου (email trackers), που χρησιμοποιούνται προκειμένου να συλλέγονται αυτόματα δεδομένα των παραληπτών μηνυμάτων ηλεκτρονικού ταχυδρομείου (λ.χ. ώρα ανάγνωσης μηνύματος, τοποθεσία ανάγνωσης μηνύματος, κτλ.). Αυτά συνήθως αναπτύσσονται για εταιρείες, ώστε να δημιουργηθούν νέες στρατηγικές διαφημίσεων και προώθησης των προϊόντων και υπηρεσιών τους. Ταυτοχρόνως υφίσταται νομικό πλαίσιο στην προστασία προσωπικών δεδομένων και της ιδιωτικότητας, καθιστώντας την ένα σοβαρό ζήτημα. Για τις εταιρείες είναι αναγκαίο να λαμβάνουν κατάλληλα μέτρα προστασίας και να συμμορφώνονται με τα νομικά αυτά πλαίσια, προκειμένου να μην ζημιωθούν οικονομικά. Από τη σκοπιά των πλοηγών του διαδικτύου η γνώση των προαναφερθέντων τεχνολογιών είναι επίσης πολύ σημαντική, εφόσον η προστασία της επεξεργασίας των προσωπικών τους δεδομένων καθίσταται ως θεμελιώδες δικαίωμά τους σύμφωνα με το νομικό πλαίσιο του κανονισμού (ΕΕ) 2016/679.

Στην παρούσα πτυχιακή εργασία θα σημειωθούν οι βασικές έννοιες στον παρόντα κλάδο, όπως το τεχνολογικό υπόβαθρο του διαδικτύου και οι νομικές και οι εταιρικές πτυχές των συστημάτων. Μετέπειτα θα μελετηθούν οι σύγχρονες τεχνολογίες καταγραφής δεδομένων και παρακολούθησης χρηστών διαδικτύου. Από τις πιο διαδεδομένες (π.χ. cookies) μέχρι και τις πιο ειδικές. Στη συνέχεια θα γίνει ανάλυση των τεχνολογιών που αποτρέπουν την ιχνηλάτηση των χρηστών του διαδικτύου, ώστε να γνωρίζει ο αναγνώστης και τους κατάλληλους τρόπους προφύλαξης από τις τεχνολογίες αυτές. Ακόμη, θα αναπτυχθεί σύστημα ταξινόμησης (classification scheme) των τεχνολογιών, με κατηγορίες τεχνολογιών και κριτήρια κατάταξης στις κατηγορίες αυτές. Τέλος, θα πραγματοποιηθεί εμπειρική έρευνα σχετικά με την υπάρχουσα διάδοση αυτών των τεχνολογιών και την ενημερότητα των χρηστών Διαδικτύου.

Η παρούσα πτυχιακή εργασία στοχεύει να γνωστοποιήσει τις σύγχρονες τεχνολογίες αυτόματης καταγραφής δεδομένων και παρακολούθησης χρηστών διαδικτύου. Γνωρίζοντάς τις, οι οργανισμοί και οι εταιρείες μπορούν να βελτιστοποιήσουν τις μεθόδους προώθησης και τις στρατηγικές διαφημίσεων και προώθησης των προϊόντων και υπηρεσιών τους. Ακόμα, οι χρήστες του διαδικτύου θα έχουν επίγνωση των όσων συμβαίνουν στο διαδίκτυο και θα μπορούν να προφυλάσσονται από μη επιθυμητές συλλογές και επεξεργασίες των προσωπικών τους δεδομένων. Μάλιστα θα γνωστοποιηθούν και αντίμετρα για αυτά τα συστήματα. Επίσης, θα είναι γνωστές οι διάφορες μορφές και λειτουργίες των τεχνολογιών αυτών. Μάλιστα συναρτήσει του ποιες έχουν υλοποιηθεί, ποιες δεν έχουν υλοποιηθεί και ποιες χρήζουν περαιτέρω βελτίωσης, δύναται να δοθούν νέες ιδέες για τη δημιουργία νέων τεχνολογιών. Τέλος, σημειώνεται ότι η μελέτη αυτών των συστημάτων θα γίνει από τη σκοπιά του τεχνολογικού υπόβαθρου και

της επιστήμης της Πληροφορικής. Δε θα ερευνηθούν από άποψη συμφερόντων των οργανισμών και εταιρειών, από νομική ή κάποια άλλη οπτική γωνία.

Κεφάλαιο 1 - Θεμελιώδεις έννοιες

Είναι γεγονός στη σημερινή εποχή η ραγδαία εξέλιξη της τεχνολογίας και του διαδικτύου, όπου επιφέρουν μια νέα τάξη πραγμάτων στους χρήστες ηλεκτρονικών υπηρεσιών. Στο σύγχρονο κόσμο, όλοι οι χρήστες του διαδικτύου καταλήγουν στο να αποθηκεύουν πληροφορίες online, ακόμη και αν δεν το γνωρίζουν. Αυτό έχει ως αποτέλεσμα να διαμοιράζονται καθημερινά δεδομένα δισεκατομμυρίων χρηστών. Σχεδόν κάθε διαδικτυακός τόπος που δέχεται επίσκεψη, αποθηκεύει πληροφορίες σχετικά με τον επισκέπτη, είτε από τον επιτραπέζιο υπολογιστή, τον φορητό υπολογιστή, το κινητό τηλέφωνο ανεξαρτήτου τοποθεσίας (σπίτι, internet cafe, δασική περιοχή). Κάθε διαδικτυακή εταιρεία, με την οποία συναναστρέφεται ένα άτομο, είτε για μια διαδικτυακή αγορά, είτε για τη συναναστροφή με ένα άτομο σε ένα διαδικτυακό κοινωνικό δίκτυο (online social network), είτε απλά για την εύρεση μιας πληροφορίας σε μια μηχανή αναζήτησης, αποθηκεύει πληροφορίες για τους πελάτες της σε έναν υπολογιστή στο δίκτυό τους και συχνά σε μια βάση δεδομένων τους. Με αυτόν τον τρόπο επιτυγχάνονται οικονομικές ωφέλειες με ανάπτυξη στρατηγικών μάρκετινγκ και περισσότερες πωλήσεις, βελτίωση των διαδικτυακών και τεχνολογικών υπηρεσιών μέσω της εύρεσης αναγκών των χρηστών και μείωση της εγκληματικότητας δια της συνεχούς επίβλεψης και επιτήρησης των ενεργειών στο internet και άλλες ωφέλειες. [1]

1.1) Συστήματα καταγραφής και παρακολούθησης δεδομένων

Τα συστήματα παρακολούθησης χρηστών διαδικτύου και συλλογής πληροφοριών ή πιο απλά τα συστήματα ιχνηλάτησης (tracking systems) αποτελούν ένα σύνολο από μεθόδους και τεχνικές πληροφορικής. Με αυτά επιτυγχάνεται η παρατήρηση ενεργειών από πρόσωπα ή από προγράμματα σε ένα ή περισσότερα δίκτυα, παρέχοντας μια σειρά από δεδομένα για περαιτέρω επεξεργασία. Αυτά τα συστήματα πολλές φορές είναι προγράμματα λογισμικού, δηλαδή ένα σύνολο από εντολές προς το μηχάνημα υπολογιστή, και στηρίζονται στην επικοινωνία μεταξύ των υπολογιστών. Τα βασικά θεμέλιά τους είναι η αίτηση κάποιων πληροφοριών μέσω ενός πρωτοκόλλου επικοινωνίας και η αποθήκευσή τους σε κάποια αποθηκευτική δομή (π.χ. αρχείο, βάση δεδομένων, κλπ.).

Η σημασία τους και ο ρόλος τους στο διαδίκτυο έχει ήδη γίνει εμφανής και μεγιστοποιείται με την πάροδο του χρόνου. Και αυτό, διότι με τη συλλογή πληθώρας πληροφοριών για τους χρήστες του διαδικτύου, οι οργανισμοί και οι εταιρείες μπορούν να βελτιστοποιήσουν τις στρατηγικές διαφημίσεων και τις μεθόδους προώθησης των προϊόντων και των υπηρεσιών τους. Τούτο επιτυγχάνεται με τη ανάλυση των συλλεγμένων δεδομένων και την εξαγωγή συμπερασμάτων, όπως τις καταναλωτικές προτιμήσεις των πελατών. Γενικώς, επιτυγχάνουν καλύτερη οικονομική κατάσταση, που είναι και ένας από τους πιο βασικούς τους στόχους, αν όχι ο βασικότερός τους. Επίσης, καταπολεμάτε το ηλεκτρονικό έγκλημα, ελέγχοντας τις τυχόν ανεπιθύμητες διαδικτυακές δραστηριότητες. Εφόσον καταγράφονται πληροφορίες των ατόμων που εισέρχονται σε έναν ιστότοπο, μπορούν να εξακριβωθούν, γεγονός που επιβάλλει συμμόρφωση με την πολιτική της

ιστοσελίδας και του νόμου. Εξίσου σημαντική αιτία για τη σημασία των τεχνολογιών αυτών είναι και το δικαίωμα της ιδιωτικότητας. Οι χρήστες του διαδικτύου χρειάζονται επίγνωση των όσων συμβαίνουν στο διαδίκτυο, ώστε να μπορούν να προφυλάσσονται από μη επιθυμητές συλλογές και επεξεργασίες των προσωπικών τους δεδομένων. Η γνώση αυτών συνεισφέρει στον ερευνητικό και επιστημονικό τομέα, δίνοντας νέες ιδέες για τη δημιουργία νέων τεχνολογιών, σύμφωνα με τις ήδη υπάρχουσες τεχνολογίες και μη, αλλά και σύμφωνα με αυτές, που χρήζουν περαιτέρω βελτίωσης. [1]

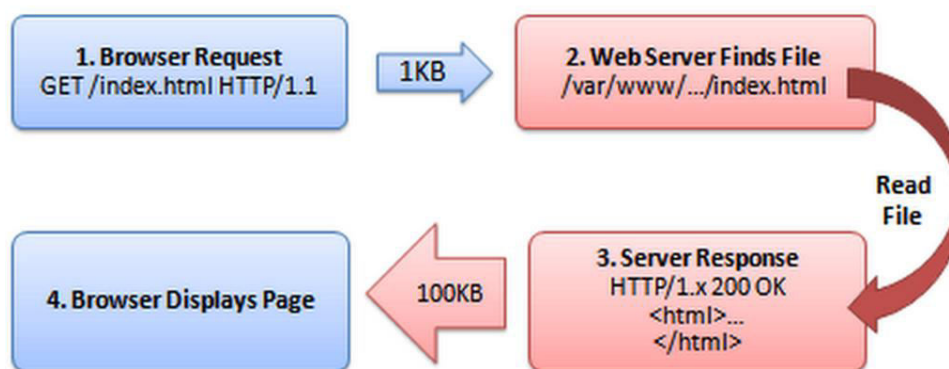
1.2) Τεχνολογικό υπόβαθρο

Για την καλύτερη κατανόηση του 2ου κεφαλαίου θα χρειαστεί να εξηγηθούν κάποιες λειτουργίες του διαδικτύου. Το διαδίκτυο με απλά λόγια είναι το παγκόσμιο σύστημα διασυνδεδεμένων δικτύων υπολογιστών. Ένα δίκτυο είναι και πάλι ένα σύνολο διασυνδεδεμένων υπολογιστών. Άρα το ένα δίκτυο είναι ένα σύνολο ενωμένων υπολογιστών και το διαδίκτυο ένα σύνολο ενωμένων δικτύων. Οι υπολογιστές μπορεί να έχουν πολλές μορφές, όπως ο προσωπικός υπολογιστής (desktop) που ανήκει σε ένα χρήστη, συμπεριλαμβανομένων των έξυπνων κινητών τηλεφώνων (smartphones), ταμπλετών (tablets), φορητών υπολογιστών (laptop). Ο προσωπικός υπολογιστής ονομάζεται πελάτης (client), διότι συνήθως αιτείται υλικό (π.χ. ιστοτόπους) από το διαδίκτυο. Οι υπολογιστές που συνήθως αποστέλλουν αυτό το υλικό ονομάζονται εξυπηρετητές (servers), συμπεριλαμβανομένων των διακομιστών, των υπολογιστικών νεφών (cloud) και εικονικών μηχανών vm (Virtual Machine). Οι υπολογιστές desktop συνήθως έχουν ένα φιλικό προς το χρήστη περιβάλλον με λειτουργικό σύστημα και επιτραπέζιες εφαρμογές για να διευκολύνει εργασίες με προσανατολισμό στην επιφάνεια εργασίας. Αντίθετα, ένας διακομιστής διαχειρίζεται όλους τους πόρους του δικτύου και συχνά δεν εκτελεί άλλα καθήκοντα εκτός από αυτό. Επειδή ένας διακομιστής έχει σχεδιαστεί για να διαχειρίζεται, να αποθηκεύει, να αποστέλλει και να επεξεργάζεται δεδομένα 24 ώρες την ημέρα, πρέπει να είναι πιο αξιόπιστος από έναν υπολογιστή γραφείου και να προσφέρει μια ποικιλία λειτουργιών και υλικού που δεν χρησιμοποιούνται συνήθως στον μέσο επιτραπέζιο υπολογιστή. Φυσικά, για να μπορεί ο διακομιστής να συλλέγει και να επεξεργάζεται δεδομένα, κάποιο άτομο τον διαχειρίζεται με κώδικα προγραμματισμού. Έστω λοιπόν μια επιχείρηση που πουλάει παπούτσια και έχει αγοράσει ένα διακομιστή. Σε αυτόν δίνει κάποια αρχεία υπολογιστή, συμπεριλαμβανομένων των φωτογραφιών από τα παπούτσια προς πώληση, κειμένου που περιγράφει τα χαρακτηριστικά τους, προγραμμάτων και τεχνολογιών καταγραφής δεδομένων, ώστε να γνωρίζει ο ιστότοπος ποιος τον επισκέφθηκε. Ως εκ τούτου, έχει ανοιχτό κάθε μέρα το διακομιστή και όποιος πελάτης ζητήσει να δει την ιστοσελίδα αυτή, ο διακομιστής είναι υπεύθυνος για να του την παραδώσει. Και όποιος πελάτης επισκεφθεί τον ιστότοπο αυτόν, ο διακομιστής είναι υπεύθυνος να συλλέξει τις πληροφορίες του πελάτη και να τις αποθηκεύσει. Εν ολίγοις, κατά αυτό τον τρόπο λειτουργεί το διαδίκτυο. [2]

Ο ιστός (web) είναι μια μπερδεμένη μάζα διασυνδεδεμένων σελίδων με περιεχόμενο, που μπορούν να προβληθούν από ένα φυλλομετρητή ιστού ή αλλιώς ένα πρόγραμμα περιήγησης διαδικτύου (web browser). Μερικά τέτοια προγράμματα είναι τα

Google chrome, Mozilla firefox και Microsoft edge. Όταν ο φυλλομετρητής του χρήστη κάνει αίτηση για τη φόρτωση μιας ιστοσελίδας, λαμβάνει ένα αρχείο HTML (Hypertext Markup Language) από τον διακομιστή που συσχετίζεται με έναν συγκεκριμένο τομέα (όπως "<http://example.com>" όπου "example.com" είναι ο τομέας). Εν ολίγοις η HTML είναι μια γλώσσα απεικόνισης του τρόπου με τον οποίο θα εμφανιστεί ένα στοιχείο (element) στο πρόγραμμα περιήγησης, περιφράσσοντάς το με αναγνωρίσιμες λέξεις που ονομάζονται ετικέτες (tags) (π.χ. <tag> element </tag>). Τα στοιχεία είναι μια απεριόριστη σειρά επιπρόσθετων περιεχομένων (εικόνες, βίντεο, κώδικας Javascript, κώδικας HTML (iFrame)), που εγκαθιδρύουν την ομορφιά της HTML και του ίδιου του ιστού. Όλα αυτά τα περιεχόμενα του αρχείου HTML, χρειάζεται να φορτωθούν και να ληφθούν από το πρόγραμμα περιήγησης του χρήστη, κατά την κλήση του ιστοτόπου. Προκειμένου να γίνει η μεταφορά των αρχείων από το διακομιστή, που φιλοξενεί το ζητηθέν περιεχόμενο, προς το φυλλομετρητή, χρησιμοποιείται το πρωτόκολλο μεταφοράς υπερκειμένου HTTP (ή η παραλλαγή του HTTPS, που προσθέτει και ασφάλεια). Στη συναλλαγή αυτή συμμετέχουν μόνο ο χρήστης ως πελάτης (client) και ο διακομιστής ως εξυπηρετητής (server) ανεξαρτήτου μορφής (λ.χ. server, cloud, vm, κλπ.). Παράλληλα, οφείλει ο φυλλομετρητής να στείλει κάποιες πληροφορίες στον ιστότοπο, όπως τα cookies που έχει αποθηκευμένα ο χρήστης ή τη διεύθυνση IP για να γνωρίζει το πού θα στείλει τα περιεχόμενα της ζητηθείσας σελίδας. Η αποστολή αυτών των πληροφοριών, καθώς και η αποθήκευσή τους, αποτελούν τη μέθοδο καταγραφής δεδομένων με HTTP, η οποία είναι καθίσταται βασικό θεμέλιο της ιχνηλάτησης. [2, 3]

HTTP Request and Response



Εικόνα 1: Απλή αναπαράσταση του τρόπου λειτουργίας του HTTP πρωτοκόλλου.

1.3) Νομικές πτυχές

Από τη νομική οπτική γωνία το θέμα της παρακολούθησης και καταγραφής δεδομένων, αν και δεν αφορά την παρούσα πτυχιακή εργασία, είναι κρίσιμο και ευαίσθητο. Πάνω από 80 χώρες και ανεξάρτητα εδάφη, συμπεριλαμβανομένων σχεδόν όλων των χωρών της Ευρώπης και πολλών πολιτειών της Λατινικής Αμερικής και της Καραϊβικής, της Ασίας και της Αφρικής, έχουν πλέον υιοθετήσει εκτεταμένους νόμους για την προστασία των προσωπικών δεδομένων. Η πιο σημαντική και ευρέως γνωστή νομοθεσία είναι ο

ΓΚΠΔ (Γενικός Κανονισμός Προστασίας Προσωπικών Δεδομένων) ή αλλιώς ο κανονισμός (ΕΕ) 2016/679. Ο ΓΚΠΔ εφαρμόστηκε το Μάιο του 2018 από το Ευρωπαϊκό Κοινοβούλιο και αντικατέστησε την Οδηγία 95/46/ΕΚ, που είχε εφαρμοστεί το 1997, για την προστασία των προσωπικών δεδομένων βάσει της Ευρωπαϊκής Σύμβασης των Ανθρώπινων Δικαιωμάτων. Έτσι, παρόλο που σε πολλούς φαίνεται καινούργιο το ζήτημα αυτό, δεν είναι. Η Ευρωπαϊκή διάσταση στην προστασία δεδομένων προσωπικού χαρακτήρα έχει ως απώτερο σκοπό τη βελτιστοποιημένη προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων των Ευρωπαίων πολιτών σε όλα τα κράτη μέλη της Ευρωπαϊκής Ένωσης. Θέτει σημαντικές υποχρεώσεις σε φορείς και βασικά δικαιώματα στους χρήστες του διαδικτύου. Λόγου χάρη, ο νόμος επιβάλλει η επεξεργασία των προσωπικών δεδομένων να γίνεται με τρόπο ασφαλή, προστατεύοντας τα προσωπικά δεδομένα από μη εξουσιοδοτημένες προσβάσεις, τροποποιήσεις ή διαγραφές και να περιορίζεται μόνο στους σκοπούς που εξυπηρετούν το φορέα. Ακόμη, δίνει το δικαίωμα και της αποφυγής της ιχνηλάτησης, όπως για παράδειγμα με τη χρήση ψευδωνυμοποίησης στα δεδομένα προσωπικού χαρακτήρα που το αφορούν. Έτσι, μπορούν τα ίδια τα υποκείμενα των δεδομένων να μειώσουν σε κάποιο βαθμό μερικούς κινδύνους, καθιστώντας τα υπεύθυνα μέλη της ασφάλειας των δεδομένων τους. Αυτή αποτελεί μια βασική αφορμή για την ανάλυση τεχνολογιών που αποτρέπουν την παρακολούθηση από ιστοτόπους. [4]

Η ιδιωτικότητα των ατόμων τίθεται ως ένα βασικό θεμέλιο, το οποίο χρήζει προστασίας. Η Ιδιωτικότητα (privacy είναι η αξίωση των ατόμων να καθορίζουν το χρόνο, τον τρόπο και την έκταση αναφορικά με τη συλλογή και επεξεργασία των προσωπικών τους δεδομένων. Τα δεδομένα προσωπικού χαρακτήρα θεωρούνται οποιεσδήποτε πληροφορίες, όπου άμεσα ή έμμεσα αναφέρονται σε φυσικό πρόσωπο και μπορούν να το εξακριβώσουν εξολοκλήρου ή έστω και κάποια συγκεκριμένα στοιχεία που χαρακτηρίζουν την υπόστασή του από φυσική, βιολογική, ψυχολογική, οικονομική, πολιτιστική ή κοινωνική άποψη. Τέτοια δεδομένα μπορεί να είναι αποθηκευμένα σε ένα διαδικτυακό κοινωνικό δίκτυο ή να μπορούν να εντοπισθούν σύμφωνα με το ιστορικό περιήγησης του χρήστη. Δηλαδή, εάν ένας χρήστης του διαδικτύου έκανε πολλές και συχνές αναζητήσεις στο διαδίκτυο σχετικά με οικονομικές ενισχύσεις που προσφέρονται σε πολύτεκνους, ένας ιχνηλάτης μπορεί να υποθέσει ότι αυτό το άτομο ανήκει σε πολύτεκνη οικογένεια. Παρόλο που αυτή η πληροφορία δεν είναι 100% δεδομένη, υπάρχει μεγάλη επιτυχίας κατανόησης των συμπεριφορών, χαρακτηριστικών και προτιμήσεων των χρηστών του διαδικτύου. Επιπλέον, ο νόμος ορίζει ότι αν τα προσωπικά δεδομένα σχετίζονται με φυλετική ή εθνική προέλευση, πολιτικά φρονήματα, θρησκευτικές ή φιλοσοφικές πεποιθήσεις, συμμετοχή σε συνδικαλιστικές οργανώσεις, υγεία, κοινωνική πρόνοια, ερωτική ζωή, ποινικές διώξεις ή καταδίκες, καθώς και συμμετοχή σε συναφείς με τα ανωτέρω ενώσεις προσώπων, τότε χαρακτηρίζονται ως ευαίσθητα προσωπικά δεδομένα. Και αυτό, διότι η γνώση μιας τέτοιας πληροφορίας συνήθως φέρει μεγαλύτερη ενόχληση. Λόγου χάρη, ένα άτομο θα ενοχλούνταν περισσότερο, εάν ένας τρίτος γνώριζε την ερωτική ζωή ή το ιστορικό ιατρικό απόρρητό του, παρά εάν γνώριζε απλά το όνομά του ή την πόλη που κατοικεί. Κάποιοι δίκαιοι θα μπορούσε να πει ότι η διάκριση απλών και ευαίσθητων προσωπικών δεδομένων είναι υποκειμενικό ζήτημα, καθώς η ψυχρούσταση των ανθρώπων ποικίλλει και διαφέρει. Μολαταύτα, η διάκρισή τους είναι ζήτημα αντικειμενικό, το οποίο ορίζεται από τα νομικά πλαίσια που εφαρμόζονται ανά χώρα και ανεξάρτητο έδαφος. [4]

Η επεξεργασία προσωπικών δεδομένων είναι κάθε είδους εργασία συλλογής, ανάκτησης, αναζήτησης, καταχώρησης, οργάνωσης, αποθήκευσης, προσαρμογής, τροποποίησης, διαγραφής, καταστροφής, διαβίβασης ή δημοσίευσης αυτών, είτε εκτελούνται με αυτοματοποιημένες ή όχι διαδικασίες, είτε με τη συναίνεση ή όχι του υποκειμένου των δεδομένων. Η ιχνηλάτηση απαιτεί αρκετή επεξεργασία δεδομένων, καθώς συχνά δεν αρκείται στην απλή συλλογή και αποθήκευση δεδομένων. Αλλά, επιδιώκει να μεταβιβάσει, τροποποιήσει, αναλύσει και γενικότερα να αξιοποιήσει τα δεδομένα αυτά, με όποιον τρόπο μπορεί να προβεί σε περαιτέρω βελτίωση. Ένα συχνό παράδειγμα είναι η στατιστική ανάλυση για την εκτίμηση των μελλοντικών δραστηριοτήτων, τόσο από τη μεριά του πελάτη, όσο και από τη μεριά του εξυπηρετητή. Για αυτό και κάθε είδους ιχνηλάτηση αλλά και κάθε είδους επεξεργασία προσωπικών δεδομένων είναι αναγκαίο να προστατεύεται με κατάλληλα μέτρα ασφάλειας. Και αυτό συμβαίνει, προκειμένου να αποφευχθούν ανεπιθύμητες διαρροές προσωπικών δεδομένων και παραβιάσεις της ιδιωτικότητας των ατόμων των δεδομένων. [4]

1.4) Συστήματα ιχνηλάτησης σε εταιρείες

Η αδιάκοπη εξάρτηση των υπηρεσιών, που παρέχονται από ιδιωτικούς ή κρατικούς φορείς, με τις Τεχνολογίες Πληροφορίας και Επικοινωνίας μεγεθύνει σημαντικά τις ευπάθειες των πληροφοριακών συστημάτων, το πλήθος και το εύρος των απειλών και των επιθέσεων. Επίσης, αυξάνει και τη σοβαρότητα των συνεπειών για τον πάροχο, αλλά και τον τελικό χρήστη της εκάστοτε υπηρεσίας σε περίπτωση κάποιου περιστατικού ασφάλειας. Οι σημερινές επιθέσεις ενδέχεται να έχουν ποινικό, οικονομικό ή ακόμα και τρομοκρατικό κίνητρο, θέτοντας σε κίνδυνο στρατηγικά συμφέροντα κρατών και οικονομικά συμφέροντα εταιρειών. Αυξανόμενος είναι και ο ρυθμός των χρηστών διαδικτύου που εκφράζουν ανησυχία αναφορικά με το πλήθος των προσωπικών δεδομένων τους, που βρίσκονται διαθέσιμα στο Διαδίκτυο. Το 2012 περίπου το 92% του Αμερικανικού πληθυσμού δήλωσε ότι ανησυχεί για θέματα ιδιωτικότητας έστω και μερικώς (TRUSTe 2014 US Consumer Confidence Privacy Report). Οι διαρροές κρίσιμων πληροφοριών, η τροποποίηση ευαίσθητων δεδομένων και η μη διαθεσιμότητα βασικών λειτουργιών δημιουργούν στη σύγχρονη κοινωνία φόβο, αποσταθεροποίηση και άλλα προβλήματα. Συμπερασματικά, οι επιχειρήσεις και οι οργανισμοί οφείλουν να επανεξετάζουν τον τρόπο επεξεργασίας αυτών, καθώς και να λαμβάνουν προληπτικά μέτρα για τη διασφάλιση της προστασίας τους. Δεδομένου, μάλιστα ότι οι επιθέσεις εναντίον των τεχνολογιών πληροφορικής και επικοινωνίας εξελίσσονται συνεχώς και η ανίχνευσή τους γίνεται όλο και πιο δύσκολη, η εξασφάλιση ενός επαρκούς επιπέδου ασφάλειας και προστασίας της ιδιωτικότητας των χρηστών κρίνεται αναγκαία. Βασικό κίνητρο για την υλοποίηση μέτρων ασφάλειας αποτελούν τα νομικά πλαίσια στο χώρο της ιδιωτικότητας και της προστασίας των προσωπικών δεδομένων. Άλλωστε, η προστασία των προσωπικών δεδομένων αποτελεί ένα επίκαιρο θέμα παγκοσμίως. Όλα αυτά δημιουργούν ένα περίπλοκο ζήτημα, κυρίως επειδή η μετατροπή των νομικών απαιτήσεων των κανονισμών σε τεχνικές απαιτήσεις λογισμικού και υλικού για τα πληροφοριακά συστήματα και τις υποκείμενες επιχειρησιακές διαδικασίες είναι πολύ δύσκολη. [1]

Τα συστήματα ιχνηλάτησης ή τα συστήματα αντί-ιχνηλάτησης (anti-tracking), δηλαδή τα συστήματα που έρχονται σε εναντίωση με τις ενέργειες των συστημάτων ιχνηλάτησης, ενδέχεται να αποτελέσουν μέτρο ασφάλειας στους οργανισμούς. Αρχικά, επειδή καταγράφονται και αποθηκεύονται τα δεδομένα του κάθε ατόμου ή/και των ενεργειών αυτού, τα άτομα του διαδικτύου, που επιθυμούν να δραστηριοποιηθούν παράνομα, φοβούνται. Έπειτα, μέθοδοι στις οποίες στηρίζονται είτε τα συστήματα ιχνηλάτησης, είτε τα αντι-ιχνηλάτησης είναι κοινές με άλλα συστήματα ασφάλειας ή μπορούν να γίνουν κοινές. Για παράδειγμα, το taint tracking (ιχνηλάτηση χρωματισμένων δεδομένων) στηρίζεται στη διαδικασία ελέγχου πληροφορίας. Αυτή η διαδικασία είχε χρησιμοποιηθεί πρώτα στον τομέα της ασφάλειας υπολογιστών και μετά στην ιχνηλάτηση. Άλλο παράδειγμα, είναι η αποτροπή διαδικτυακού κώδικα script, που εκτός από την απαγόρευση χρήσης κώδικα ενός ιστοτόπου για να ιχνηλατήσει, απαγορεύει και τη χρήση κώδικα που είναι επιβλαβές (π.χ. εγκατάσταση ιών). Μολαταύτα, κάποιος θα μπορούσε να επισημάνει ότι οι μέθοδοι ιχνηλάτησης μπορούν να συνδεθούν και με τα επιβλαβή λογισμικά. Αυτό είναι αληθές και συνεπώς η γνώση του τρόπου λειτουργίας τους είναι αξιοσημείωτη και κάθε τμήμα Ηλεκτρονικών Υπολογιστών σε κάθε φορέας οφείλει να γνωρίζει περί των συστημάτων ιχνηλάτησης. Για αυτό το λόγο, αλλά και για όλα τα προαναφερθέντα, η παρούσα εργασία επιδιώκει την εξακρίβωση του τεχνολογικού υπόβαθρου των συστημάτων αυτών, αναλύοντας το πώς λειτουργούν. [\[1\]](#)

Κεφάλαιο 2 - Συστήματα καταγραφής δεδομένων **και παρακολούθησης χρηστών διαδικτύου**

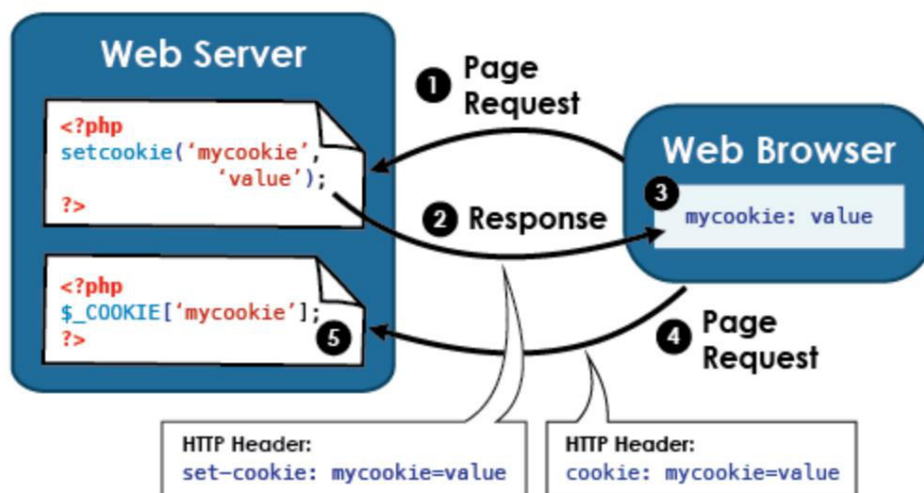
Τα συστήματα καταγραφής δεδομένων και παρακολούθησης χρηστών, όπως προαναφέρθηκε, αποτελούν ένα σύνολο από προγράμματα λογισμικού, πρωτόκολλα επικοινωνίας υπολογιστών, καθώς και άλλες τεχνικές Πληροφορικής, ώστε να συλλέγουν και να αποθηκεύουν δεδομένα των χρηστών του διαδικτύου. Κύριο ενδιαφέρον παρουσιάζει η αρχιτεκτονική τους, δηλαδή ο τρόπος λειτουργίας αυτών, και κάποιες παραλλαγές αυτών. Ο αριθμός των παραλλαγών αυτών είναι άπειρος, αλλά η επισήμανση και κατανόηση μερικών από αυτών, βοηθάει στην περαιτέρω αντίληψη του τρόπου λειτουργίας των συστημάτων αυτών. Επίσης, ενδέχεται να δώσει ιδέες για τη δημιουργία νέων παραλλαγών ή τη βελτίωση των ήδη υπαρχόντων.

2.1) Cookies

Τα HTTP cookies χρησιμοποιούνται ευρέως από τους κατόχους ιστοσελίδων και τα τρίτα μέρη αυτών για την παρακολούθηση των χρηστών και των συμπεριφορών τους. Ένα cookie δεν είναι πρόγραμμα, αλλά είναι μια συμβολοσειρά κειμένου, που αποτελείται από ζευγάρια κλειδιού-τιμής (σαν ζευγάρια ερώτησης-απάντησης), που περιγράφουν στοιχεία και δραστηριότητες του επισκέπτη (λ.χ. Κλειδί: Ώρα Επίσκεψης - Τιμή: 17:45, Κλειδί: Αριθμός click σε Διαφημίσεις - Τιμή: 3). Για παράδειγμα, μπορεί να αποθηκεύει τον κωδικό πρόσβασης ενός χρήστη (Κλειδί: Κωδικός Πρόσβασης - Τιμή: Μαρία1234), ώστε να μη χρειαστεί να πληκτρολογηθεί ξανά στην επόμενη επίσκεψη σε αυτόν τον ιστότοπο. Κατά συνέπεια, ο ιστότοπος αντλεί πληροφορίες για τους επισκέπτες του, οι οποίες αποτελούν σημαντική πηγή εσόδων. Μάλιστα, άλλες τρίτες υπηρεσίες επενδύουν οικονομικώς προκειμένου να μοιραστούν αυτή την πληροφορία. Έτσι, τα cookies πρώτου μέρους τοποθετούνται (στη συσκευή του επισκέπτη) και αναπτύσσονται από τον ιστότοπο που επισκέφθηκε ο χρήστης, ενώ τα cookies τρίτων μερών από τρίτες υπηρεσίες, συνήθως από εταιρείες μεσετιείας δεδομένων, διαφημιστές στο διαδίκτυο και εφαρμογές παρακολούθησης (π.χ. Analytics). Αυτό έχει ως αποτέλεσμα μια μεγάλη συλλογή προσωπικών δεδομένων, που σχετίζονται με τις προτιμήσεις και συμπεριφορές των χρηστών του διαδικτύου, αλλά και μια μεγάλη αλληλοσυσχέτιση αυτών. Αξίζει να σημειωθεί ότι τα cookies και η τοπική αποθήκευση HTML5 εξυπηρετούν διαφορετικούς σκοπούς. Τα cookies είναι κυρίως για την ανάγνωση από την πλευρά του διακομιστή, ενώ η τοπική αποθήκευση μπορεί να διαβαστεί μόνο από την πλευρά του πελάτη. Χωρίς τη μεταφορά πληροφοριών μέσω cookies, ο ιστός αποτελεί μια απλή μάζα ασύνδετων πληροφοριών, που οι συναλλαγές του δεν συσχετίζονται. [\[1, 5, 6, 7, 8, 9, 10\]](#)

Η μέθοδος τοποθέτησης των cookies σε μια συσκευή (υπολογιστής, smartphone, κλπ.) είναι απλή και χρησιμοποιεί συνήθως (αν όχι πάντα) το πρωτόκολλο μεταφοράς υπερκειμένου HTTP ή HTTPS. Όταν πληκτρολογηθεί μια διεύθυνση URL μιας τοποθεσίας Web (έστω <https://www.amazon.com>) σε ένα πρόγραμμα περιήγησης, αυτό στέλνει αίτημα στο διακομιστή ιστού (της Amazon) για προβολή της ζητηθείσας σελίδας. Έπειτα, από το

πρόγραμμα περιήγησης θα ζητηθεί ένα ή παραπάνω αρχεία cookies που έχει ορίσει ο ιστότοπος (η Amazon), τα οποία είναι αποθηκευμένα στο σκληρό δίσκο της συσκευής του αιτούντα. Στην περίπτωση που αυτά εντοπιστούν, το πρόγραμμα περιήγησης θα στείλει όλα τα ζεύγη κλειδιού-τιμής από τα αρχεία cookies ή τα ίδια τα αρχεία στον εξυπηρετητή (του Amazon) μαζί με την αιτηθείσα διεύθυνση URL. Ειδάλλως, εάν δεν εντοπίσει κανένα αρχείο cookie, δε θα στείλει τίποτα πίσω. Στη συνέχεια ο διακομιστής προβάλλει την αιτηθείσα σελίδα. Και παράλληλα, εάν ο διακομιστής Web (της Amazon) λάβει τα δεδομένα των cookies, τότε μπορεί να τα χρησιμοποιήσει για να αναγνωρίζει και να καταγράφει τις προτιμήσεις και δραστηριότητες των χρηστών. Τούτο γίνεται συχνά, προκειμένου να επιτευχθούν διάφορα εμπορικά και χρηματοοικονομικά συμφέροντα (λ.χ. η καλύτερη αγοραστική εμπειρία και η πιο εύστοχη αποστολή διαφημίσεων). Διαφορετικά, στην περίπτωση που δε λάβει δεδομένα από cookies, ο ιστότοπος θεωρεί ότι ο χρήστης αυτός δεν τον έχει επισκεφθεί ξανά προηγουμένως. Άρα, ο διακομιστής δημιουργεί ένα νέο αναγνωριστικό στη βάση δεδομένων του για τον επισκέπτη και μετά στέλνει ζεύγη κλειδιού-τιμής στη συσκευή. Η συσκευή αποθηκεύει τα cookies στον σκληρό δίσκο. Μετέπειτα, ο διακομιστής Web μπορεί να αλλάξει τα ζεύγη κλειδιού-τιμής ή να προσθέσει καινούργια σε κάθε νέα επίσκεψη. Μια εναλλακτική περίπτωση, είναι η κλήση των cookies από ένα API, το οποίο ενεργοποιείται μέσω script του ιστοτόπου. Αφού κληθούν, στη συνέχεια αποστέλλονται μέσω οποιουδήποτε καναλιού, όπως μέρος μιας αίτησης HTTP. Αυτή η μέθοδος χρησιμοποιείται σε ιστοτόπους, ώστε να συμμορφώνονται με τα νομικά πλαίσια προστασίας προσωπικών δεδομένων. Πιο συγκεκριμένα, κατά την επίσκεψη στέλνουν ένα ή πολύ λίγα cookies και έπειτα δίνουν στον επισκέπτη τη δυνατότητα, να επιλέξει (σε ένα βαθμό) ποια cookies να χρησιμοποιηθούν από τον ιστότοπο. Μετέπειτα καλούνται τα επιλεγμένα cookies μέσω script από το API. [5, 6, 7, 8, 9, 10]



Εικόνα 2: Αλληλεπίδραση προγράμματος περιηγητή με εξυπηρετητή αιτηθείσας ιστοσελίδας.

Ενώ τα cookies αποτελούν το πρωταρχικό μέσο web εφαρμογών για τον έλεγχο των αιτημάτων HTTP και τη διατήρηση των πελατών τους, η χρήση τους έχει σημαντικές επιπτώσεις στο ιδιωτικό απόρρητο των χρηστών και χρήζει ασφάλειας. Πράγματι, ένας από

τους πρωταρχικούς στόχους των εταιρειών μεσιτείας δεδομένων και των διαφημιζόμενων στο διαδίκτυο είναι να συλλέξουν όσο το δυνατόν περισσότερες πληροφορίες σχετικά με τους χρήστες προς το στόχο της παράδοσης στοχοθετημένων διαφημίσεων. Αυτό έχει οδηγήσει στην ανάπτυξη διαφόρων εργαλείων που διευκολύνουν τη διαχείριση, την αφαίρεση, αλλά και την παρεμπόδιση εγκατάστασης των cookies, καθώς και τους νόμους που διέπουν την ιδιωτικότητα του χρήστη. Τούτο όμως ενδέχεται να μην αρκεί, αλλά απαιτείται ένα ασφαλές πρωτόκολλο για τη μετάδοση πληροφορίας των cookies, με απώτερο σκοπό να μη διαρρεύσει σε μη εξουσιοδοτημένες οντότητες. Υπάρχουν πολλοί μέθοδοι υλοποίησης ασφάλειας στα cookies, όπως η χρήση κρυπτογράφησης. Για παράδειγμα, έστω ότι ο όρος (δεδομένα) k υποδηλώνει την κρυπτογράφιση των δεδομένων, χρησιμοποιώντας το κλειδί διακομιστή k και το κλειδί του διακομιστή sk . Αφού το cookie συλλέξει τα δεδομένα του, τα κρυπτογραφεί με το κλειδί k και παρέχει, ως τιμή στο πεδίο της προδιαγραφής cookie HTTP, τις εξής πληροφορίες: όνομα χρήστη, χρόνος λήξης του cookie, (δεδομένα) k και HMAC (όνομα χρήστη | χρόνος λήξης του cookie | δεδομένα | k , sk). Τα δύο πεδία, όνομα χρήστη και χρόνος λήξης του cookie, δεν κρυπτογραφούνται, ώστε να μπορεί ο διακομιστής να υπολογίσει την τιμή του $k = \text{HMAC}(\text{όνομα χρήστη} | \text{χρόνος λήξης του cookie} | sk)$. Επίσης, ας σημειωθεί ότι το HMAC (Hash Message authentication code) αποτελεί μια συνάρτηση αυθεντικοποίησης μηνύματος Hash, δηλαδή ένα συγκεκριμένο τύπο κώδικα επαλήθευσης μηνυμάτων (MAC), που περιλαμβάνει κρυπτογραφική λειτουργία κατακερματισμού και κρυπτογραφικό κλειδί (στην προκειμένη περίπτωση του διακομιστή). Το παραπάνω πρωτόκολλο είναι απλό στην υλοποίηση και πολύ αποδοτικό στην ασφάλεια, καθώς παρέχει και τις βασικές υπηρεσίες ασφάλειας CIA (εμπιστευτικότητα, ακεραιότητα, αυθεντικότητα), αλλά και την anti-replay (υπηρεσία αποφυγής εισβολής ή πραγματοποίησης αλλαγών στα πακέτα που μετακινούνται από μια πηγή σε ένα προορισμό). Η υπηρεσία anti-replay, αν και όχι τόσο γνωστή, θεωρείται πολύ σημαντική για την υλοποίηση ασφάλειας στα cookies. [10]

Ένα είδος διαφορετικό από τα HTTP cookies είναι τα flash cookies, που σχεδιάστηκαν για να βελτιώσουν την εμπειρία περιήγησης ενός χρήστη. Τα flash cookies προορίζονται να χρησιμοποιηθούν για βασικές λειτουργίες του web, κυρίως στηριζόμενες στο adobe flash player, ένα λογισμικό που προσφέρει πλούσια περιεχόμενα διαδικτύου (π.χ. multimedia). Λόγου χάρη, μπορούν να αποθηκεύσουν τις προτιμήσεις ιδιωτικότητας ενός χρήστη ή να αποθηκεύσουν το σημείο, όπου ο χρήστης έπαψε να παίζει ένα flash βιντεοπαιχνίδι. Βασική διάκριση των flash cookies με τα HTTP cookies είναι η ισχυρότερη αλληλοσύνδεση των flash cookies. Ότι, δηλαδή, μπορούν να συνδέσουν τις δραστηριότητες ενός χρήστη σε μια τοποθεσία Web σε αυτό που ο χρήστης κάνει σε άλλη τοποθεσία Web. Ανεξάρτητα από το πόσα προγράμματα περιήγησης ή διαφορετικές εκδόσεις του προγράμματος περιήγησης έχουν εγκατασταθεί στον υπολογιστή του χρήστη, η χρήση των flash cookies θα επιτρέψει σε όλα τα προγράμματα περιήγησης να μοιραστούν ένα cookie. Για αυτό και τα flash cookies έχουν μεγαλύτερη χωρητικότητα (περίπου 20-25 φορές). Ωστόσο, η άρνηση χρήσης των flash cookies δύναται να επηρεάσει την εμπειρία περιήγησης ενός χρήστη. Φερ' ειπείν, ενδέχεται να οδηγήσει σε αδυναμία προβολής βίντεο ή άλλου περιεχομένου που χρησιμοποιεί το flash (λ.χ. βιντεοπαιχνίδια και κινούμενα σχέδια). Μάλιστα είναι αξιοσημείωτο ότι, ένα γιγάντιο μέρος του ιστού χρησιμοποιεί flash cookies. Εάν μάλιστα οι χρήστες επιχειρήσουν να διαγράψουν τα flash cookies, ενδέχεται να μη

μπορέσουν να προσδιορίσουν τη θέση τους. Ακόμη και τα εργαλεία περιήγησης συχνά δεν ελέγχουν τα flash cookies. Βέβαια, παρέχεται ένα διαδικτυακό εργαλείο στον ιστότοπο της Adobe που μπορεί να χρησιμοποιηθεί για τη διαγραφή των υφιστάμενων ή μελλοντικών Flash cookies, αλλά θεωρείται από πολλούς χρήστες δύσκολο. Παρόμοια είδη αποθηκευτικού χώρου βάσει μιας εφαρμογής, πλατφόρμας ή λογισμικού αποτελούν όλα τα local shared objects, όπως τα Google Gears, τα Java JNLP και το Silverlight Isolated Storage. [8, 11]

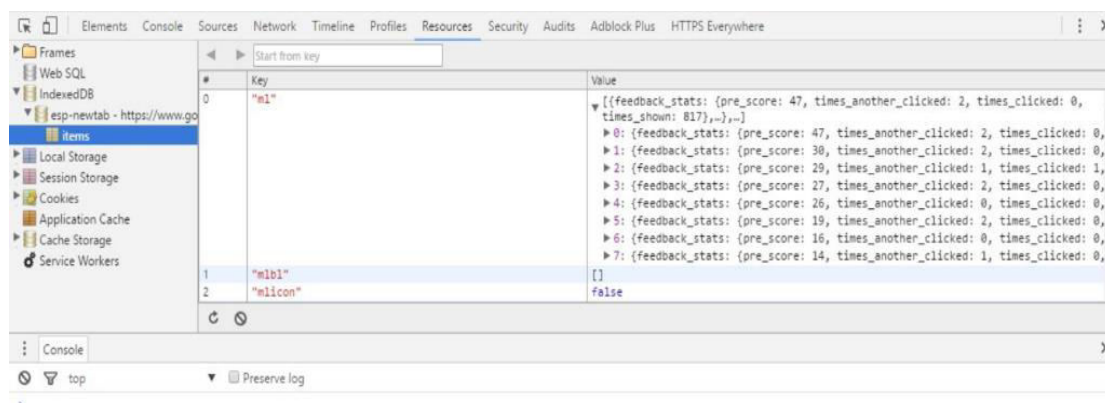
Ενώ τα HTTP cookie εξακολουθούν να αποτελούν τη συνηθέστερη μέθοδο παρακολούθησης σε απευθείας σύνδεση από τρίτους, έχουν καταδειχθεί, βελτιωθεί και αναπτυχθεί μια σειρά από πιο ενοχλητικές παραλλαγές τους. Μια από αυτές είναι τα persistent cookies ή zombie cookies ή evercookies, εκμεταλλευόμενοι διαφορετικούς μηχανισμούς αποθήκευσης του προγράμματος περιήγησης ή άλλων προγραμμάτων αποκατασταθούν τα cookies που έχουν αφαιρεθεί. Αυτά είναι απλά HTTP cookies, αλλά δημιουργούνται ξανά μετά τη διαγραφή τους από αντίγραφα ασφαλείας, που είναι αποθηκευμένα είτε στη μνήμη localStorage της HTML5, είτε στον υπολογιστή του επισκέπτη. Βέβαια, εάν έχουν αποθηκευτεί στο σκληρό δίσκο του επισκέπτη, συνήθως δε θα βρίσκονται στον αποθηκευτικό χώρο του προγράμματος περιήγησης, ώστε να είναι δύσκολο να ανιχνευθούν. Μια μέθοδος επίτευξης αυτού του γεγονότος είναι, κατά την επίσκεψη ενός χρήστη σε μια ιστοσελίδα να κατεβεί κρυφά κάποιο script ή εκτελέσιμο πρόγραμμα, το οποίο θα ελέγχει ανά χρονικά διαστήματα τα persistent cookies. Εάν κατά τον έλεγχο διαπιστωθεί ότι δεν έχουν διαγραφεί, τότε αποθηκεύει τα περιεχόμενά τους ως πρότυπο αναδημιουργίας. Παρόλα αυτά, εάν διαπιστωθεί ότι έχουν διαγραφεί, τότε δημιουργεί ένα νέο αρχείο βάσει του τελευταίου αποθηκευμένου προτύπου. Αυτό το νέο αρχείο είναι το ίδιο το persistent cookie, που υπήρχε και πριν. Άλλη παρόμοια μέθοδος αναδημιουργίας στηρίζεται στη χρήση του HTML5 localStorage API, ώστε σε κάθε επίσκεψη του χρήστη να επαναφέρονται τα παλιά cookies. Τέλος, τα flash cookies ή τα ETags (μέρος του HTTP πρωτοκόλλου) μπορούν να αποθηκεύουν τις προηγούμενες καταστάσεις των διαγραμμένων cookies, συμβάλλοντας στην επαναφορά τους. [11, 12]

2.2) HTML και HTML5

Όπως προαναφέρθηκε στην ενότητα 1.2 η HTML είναι μια γλώσσα σήμανσης υπερκειμένου. Το υπερκείμενο είναι ένα μέσο πληροφοριών, που περιλαμβάνει στην ουσία απλό κείμενο, αλλά μπορεί με τη βοήθεια ετικετών να το αναγνωρίζει ως πολυμέσα (ήχο, βίντεο, εικόνες, κλπ.), υπερσυνδέσεις, κώδικα και άλλες δομές. Η HTML είναι μια βασική και συχνή αλληλεπίδραση μεταξύ του πελάτη και του εξυπηρετητή δια μέσου του HTTP πρωτοκόλλου. Βέβαια, η HTML από μόνη της δεν ιχνηλατεί (ακόμα και με φόρμες αυθεντικοποίησης). Μόνο αξιοποιώντας τη χρήση κώδικα και τη χρήση διακομιστών, μπορεί να συλλέγει και να καταγράφει δεδομένα σε κάποια αποθηκευτική δομή, όπως είναι τα αρχεία και οι βάσεις δεδομένων. Συνοπτικά, το αρχείο είναι μια δομή που περιέχει πληροφορίες και η βάση δεδομένων είναι μια συλλογή από συστηματικά μορφοποιημένα σχετιζόμενα δεδομένα, στα οποία είναι δυνατή η ανάκτηση δεδομένων μέσω αναζήτησης κατ' απαίτηση. Οπότε, αρκεί να διαβασθούν οι κεφαλίδες του HTTP πρωτοκόλλου και να

αποθηκευτούν σε μια αποθηκευτική δομή. Ακόμα και χωρίς άμεσο χειρισμό του προγράμματος περιήγησης και μόνο με την ανάλυση κεφαλίδων HTTP σε σύνολο δεδομένων πολλαπλών σημείων, δύναται να επιτευχθεί ποσοστό αναγνώρισης χρηστών 80%, το οποίο είναι παρόμοιο με εκείνο που επιτυγχάνεται με cookies. [2, 3, 5]

Η παρακολούθηση ιστού στηρίζεται στην ικανότητα του ιστοτόπου να αποθηκεύει πληροφορίες σχετικά με τον επισκέπτη. Το HTML5 (αναβαθμισμένη έκδοση του HTML) εισήγαγε νέους μηχανισμούς αποθήκευσης από πλευράς πελάτη (client) για προγράμματα περιήγησης. Συγκεκριμένα παρέχεται η τοπική αποθήκευση (local storage), μια μόνιμη περιοχή αποθήκευσης δεδομένων, που οι ιστότοποι μπορούν να έχουν πρόσβαση με κλήσεις API. Κατ' αυτό τον τρόπο ο ιστότοπος χρησιμοποιεί μερικές γραμμές κώδικα σε Javascript, με απώτερο σκοπό να καλέσει κάποιο API, το οποίο θα αντλήσει δεδομένα από τον επισκέπτη (π.χ. με την εντολή GET) και να τα αποθηκεύσει στο local storage. Ακόμα, μπορεί να χρησιμοποιηθεί μια βάση δεδομένων IndexedDB για την αποθήκευση των αντλούμενων δεδομένων σε μορφότυπο συμβολοσειράς JSON. Αυτή είναι πολύ πιο αποδοτική για την αποθήκευση μεγαλύτερων ποσοτήτων δεδομένων, καθώς ανά κλειδί (key) μπορεί να αποθηκεύει περισσότερες από μία τιμές (value). Επιπλέον, καθιστά δυνατή τη ρύθμιση ευρετηρίων σε συγκεκριμένα στοιχεία της βάσης δεδομένων για την πιο γρήγορη εύρεσή τους. Το σύστημα αυτό δεν συμπεριφέρεται σαν μια παραδοσιακή σχεσιακή βάση δεδομένων, αλλά μοιάζει περισσότερο με τις τεχνικές NoSQL. Όπως και το local storage, έτσι και αυτός ο μηχανισμός λειτουργεί με API, που ονομάζεται IndexedDB API. Οπότε, ο browser μπορεί να επικοινωνεί με την κλήση του IndexedDB API, χρησιμοποιώντας Javascript. [3, 11, 13]



Εικόνα 3: Προβολή των πόρων της IndexedDB από τον ιστότοπο του google.nl, χρησιμοποιώντας τον περιηγητή Google Chrome.

Οι τεχνικές αποθήκευσης και ζήτησης των πληροφοριών με την HTML5 δεν είναι ιδιαίτερα δύσκολες στην υλοποίηση, διότι το πιο τεχνικό μέρος το αναλαμβάνει το API (application programming interface). Το API είναι ένα σύνολο ορισμών υπορουτίνας, πρωτόκολλα επικοινωνίας και εργαλεία για την οικοδόμηση λογισμικού, αποτελώντας μια σειρά σαφώς καθορισμένων μεθόδων επικοινωνίας μεταξύ των διαφόρων συνιστωσών. Συνοπτικά το API είναι μια μέθοδος επικοινωνίας μεταξύ συστημάτων. Δεν υπάρχει

απαραίτητα συγκεκριμένη ή περιορισμένη μέθοδος για την υλοποίησή του. Φερ' ειπείν μπορεί να αποτελείται μόνο από κώδικα (π.χ. SQL για αποθήκευση, HTML για παρουσίαση, PHP για εκτέλεση). Στην HTML5 μπορούν να χρησιμοποιηθούν πολλά API, κυρίως υλοποιημένα σε Javascript. Μερικά API γραμμένα σε Javascript για HTML5 είναι τα:

α) Web storage, που χρησιμοποιείται για μόνιμη αποθήκευση δεδομένων των δεδομένων ζευγών-κλειδιών σε πελάτες Web

β) Server-Sent Events, που χρησιμοποιείται για άνοιγμα μιας σύνδεσης HTTP για τη λήψη ειδοποιήσεων από ένα διακομιστή

γ) XMLHttpRequest2, που χρησιμοποιείται για την παροχή λειτουργικότητας για τη μεταφορά δεδομένων μεταξύ ενός διακομιστή και ενός πελάτη

δ) Geolocation, για την παροχή εφαρμογών ιστού με πρόσβαση σε πληροφορίες γεωγραφικής θέσης

ε) Canvas 2D Context, για χρήση Canva στοιχείων, τα οποία μπορούν μάλιστα να δώσουν αναγνωριστικά δακτυλικά αποτυπώματα (fingerprints) σε επισκέπτες

στ) Web Messaging, για επικοινωνίες μεταξύ των πλαισίων περιήγησης σε έγγραφα HTML

Η συνδυαστική χρήση αυτών μπορεί να επιφέρει μεγάλη συλλογή, καταγραφή, αποθήκευση και περαιτέρω επεξεργασία δεδομένων από χρήστες του διαδικτύου. Λόγου χάρη, μπορεί να κληθεί το Server-Sent Events API για να λαμβάνει τις πληροφορίες που μεταδίδονται στο HTTP πρωτόκολλο, έπειτα το Web storage API για την αποθήκευσή τους και τέλος το Web Messaging API για τη μεταβίβασή τους σε άλλους ιστοτόπους. Εντούτοις, παρόλες τις δυνατότητες της παρακολούθησης και καταγραφής δεδομένων χρηστών μέσω HTML ή HTML5, καθώς και την ευκολία υλοποίησης, δεν εφαρμόζονται τόσο συχνά όσο άλλες μέθοδοι ιχνηλάτησης. [5, 13]

2.3) Web Beacons

Ένα Web beacon, όπου εναλλακτικώς ονομάζεται Web bug, tracking bug, pixel tag και clear gif, είναι μια τεχνική προβολής μιας εικόνας πολύ μικρού μεγέθους (1x1 pixel συνήθως) μορφής .gif σε ένα αρχείο HTML ή σε κώδικα προγραμματισμού Web (π.χ. Javascript). Αυτή έχει ως σκοπό να παρακολουθεί τη συμπεριφορά των χρηστών. Προκειμένου αυτή η εικόνα να εμφανιστεί, το πρόγραμμα περιήγησης του χρήστη χρειάζεται να κάνει μια αίτηση στο διακομιστή της εικόνας. Έτσι, ο διακομιστής λαμβάνοντας την αίτηση αυτή, γνωρίζει ότι το αρχείο HTML έχει προβληθεί και από ποιο χρήστη. Ενδεχομένως να είναι οποιοδήποτε στοιχείο HTML που προορίζεται να περάσει απαρατήρητο από τους χρήστες και να υπάρχει για σκοπούς επιτήρησης. Ο χρήστης δε μπορεί να δει το Web beacon λόγω του μεγέθους του (1x1 pixel). Ωστόσο, μπορεί να ελέγξει την πηγή έκδοσης του αρχείου HTML και να διαπιστώσει μια ετικέτα (συνήθως) που

φορτώνει από ένα διαφορετικό web server από το υπόλοιπο μέρος της ιστοσελίδας. [[1](#), [6](#), [8](#), [14](#)]

Το web beacon ταυτοποιεί τη διεύθυνση IP του χρήστη και καταγράφει τα δεδομένα κίνησής του, κατά την πλοήγησή του στο διαδίκτυο. Έχει τη δυνατότητα να προσδιορίσει και να καταμετρήσει τους επισκέπτες, τους όρους αναζήτησης του χρήστη, τα κλικ σε διαφημίσεις ή άλλες παραπομπές και συνδέσμους μέσα στο HTML αρχείο. Ακόμα, μπορεί να παρακολουθήσει έναν χρήστη σε πολλές διαφορετικές τοποθεσίες Web, εξατομικεύοντας έτσι την εμπειρία τού χρήστη τού Διαδικτύου. Επίσης, δύναται να επιτρέψει σε τρίτο μέρος να παρέχει δεδομένα για τη δημιουργία αρχείων καταγραφής διακομιστή (server log files) σε μια τοποθεσία Web, που η ίδια δε μπορεί να κάνει αυτή τη λειτουργία. Όλα αυτά είναι εφικτά, επειδή μπορεί να χρησιμοποιεί cookies. Πράγματι, όταν ζητάει το browser να εμφανιστεί η εικόνα του web beacon, ο διακομιστής του web beacon μπορεί να στείλει cookies στο browser. Με αυτό τον τρόπο έχει τη δυνατότητα μέσω cookies να παραδώσει ή να ανακτήσει σε ένα τρίτο μέρος τις πληροφορίες που συλλέγονται. Μια άλλη περίπτωση είναι η ενσωμάτωση του web bug σε κάποιο ηλεκτρονικό μήνυμα (emails) σε μορφή HTML. Κατά συνέπεια, ο αποστολέας θα γνωρίζει, εάν το μήνυμά του διαβάστηκε και πότε. Επιπρόσθετα, δύναται να χρησιμοποιηθεί σε ένα δίκτυο για να αποκαλυφθεί πόσες φορές διαβάστηκε ή διακινήθηκε ένα μήνυμα αλλά και από ποιους, αποκαλύπτοντας μέχρι και τη διεύθυνση IP του παραλήπτη. Με την πάροδο του χρόνου, το web beacon κατασκευάζει ένα σημαντικό προφίλ, πολλές φορές χωρίς τη γνώση των χρηστών, χρησιμοποιώντας τις προαναφερθείσες μεθόδους. Ορισμένες εταιρείες παρακολούθησης αγοράζουν αυτά τα προφίλ και χρησιμοποιούν αναλυτικές μεθόδους, για να αναπτύξουν συμπληρωματικά στατιστικά προφίλ. Μια άλλη αξιοσημείωτη περίπτωση είναι η δημιουργία ενός web beacon με τη βοήθεια κώδικα PHP, ASP.net και άλλων γλωσσών προγραμματισμού Web σε εξυπηρετητές. Με αυτόν τον τρόπο δύναται ένα web beacon να δημιουργηθεί από μια γεννήτρια εντοπισμού διεύθυνσης (URL Tracking Generator), ενώ παράλληλα η αίτηση για προβολή της εικόνας συνεπάγεται και την αίτηση για εκτέλεση του κώδικα. Πράγματι αυτό είναι εφικτό μπορεί να επιφέρει πολλές από τις παραπάνω λειτουργίες με τη μέθοδο πλέον αλγορίθμων. Για τον εντοπισμό διευθύνσεων IP και άλλων προσωπικά αναγνωρίσιμων στοιχείων (PII) ενδέχεται να χρησιμοποιηθεί μια υπηρεσία γεωγραφικού εντοπισμού, όπως η Maxmind. [[1](#), [6](#), [8](#), [14](#)]

```

$from      = "hello@tayyiboladoja.com";
$to        = $email;
$subject   = "Tracking Test" ;

$body      = "<p>Hello Tayyib<br> this is a test email.</p> <img src='
http://tayyiboladoja.com/email-tracker/record.php?log=true&user_id=12&name=Tayyib Oladoja
&subject=This is a test&email=hello@tayyiboladoja.com&date=2016-05-10 02:29:37pm'
width='1' height='1' border='0' />|";

$textVersion = "Text Version";

$mail->Host      = $host;
$mail->Username  = $username;
$mail->Password  = $password;
$mail->From      = $from;
$mail->FromName  = "Tayyib Oladoja";
$mail->AddAddress($to);
$mail->Subject   = $subject;

```

Εικόνα 4: Μέρος κώδικα PHP που αποστέλλει ένα email. Στο email φορτώνεται ένα Web beacon με τη χρήση της ετικέτας και ως παράμετρό της το σύνδεσμο του Web beacon, που εμπεριέχει και εκτελέσιμο PHP. Έτσι, εκτελείται ταυτόχρονα και κώδικας PHP του συνδέσμου, που λαμβάνει πληροφορίες του παραλήπτη. Ο πλήρης κώδικας βρίσκεται σε αυτόν το σύνδεσμο: <http://tayyiboladoja.com/blog/how-to-track-emails.shtml>

2.4) Λογισμικό κατασκοπείας - Spyware

Ο όρος spyware αναφέρεται σε ένα λογισμικό κατασκοπείας, το οποίο εγκαθίσταται στον υπολογιστή ενός χρήστη, χωρίς τη συγκατάθεση ή τη γνώση του. Τούτο επιτυγχάνεται με τη λήψη και εγκατάσταση λογισμικών (συνήθως δωρεάν), που η προέλευσή τους είναι από τρίτους και αγνώστους κατασκευαστές. Το spyware εκτελείται στο παρασκήνιο εν αγνοία του χρήστη, καθιστώντας το δύσκολο στην ανίχνευσή του. Συγκεντρώνει στοιχεία του χρήστη, όπως ιστοσελίδες που επισκέπτεται, κωδικούς πρόσβασης, ακόμη και αριθμούς πιστωτικών καρτών, και τα μεταδίδει σε τρίτους. Υπάρχουν πολλά διαφορετικά είδη Spyware, που εκτός από τις ενέργειες κατασκοπείας, μπορούν να επιφέρουν παραβιάσεις ασφάλειας, αλλάζοντας ρυθμίσεις και εκτελώντας διάφορες ανεπιθύμητες, κακόβουλες και ενοχλητικές δραστηριότητες. [1, 15]

Τα κατασκοπευτικά προγράμματα έχουν πολλές επιλογές σε έναν υπολογιστή με λογισμικό Windows, όταν πρόκειται να αναζητήσουν “σημεία” του υπολογιστή, που συλλέγουν προσωπικές πληροφορίες για τους χρήστες, και να αποκτήσουν μη εξουσιοδοτημένη πρόσβαση σε αυτούς. Για παράδειγμα, οι πάροχοι υπηρεσιών επιπέδου (Layered Service Providers), οι οποίοι βρίσκονται ανάμεσα στο API δικτύου (στο επίπεδο εφαρμογών) και στον πυρήνα (kernel), μπορούν να φιλτράρουν την κυκλοφορία δικτύου ή/και να συλλέγουν πληροφορίες για τους χρήστες. Ένα άλλο παράδειγμα είναι οι διεργασίες υποβάθρου, που εκτελούνται αυτόματα κατά την εκκίνηση για την παρακολούθηση των ενεργειών των χρηστών. Επίσης, επειδή οι ευαίσθητες πληροφορίες είναι συχνά προσβάσιμες μέσω διαδικτυακών διεπαφών, οι επεκτάσεις (extensions) και

εφαρμογές του προγράμματος περιήγησης είναι ένας άλλος δημοφιλής μηχανισμός συλλογής ευαίσθητων δεδομένων και παρακολούθησης των ενεργειών των χρηστών. Αυτές είναι οι κύριες πηγές, που χρησιμοποιούνται στην πλειοψηφία των προγραμμάτων spyware ως μηχανισμός πρόσβασης στις πληροφορίες σχετικά με τις συνήθειες περιήγησης ενός χρήστη ή για τον έλεγχο της συμπεριφοράς του προγράμματος περιήγησης. [16]

Ο πιο άμεσος μηχανισμός παρακολούθησης της συμπεριφοράς των χρηστών είναι η εγγραφή σε συμβάντα του προγράμματος περιήγησης, χρησιμοποιώντας τη διεπαφή IConnectionPoint του προγράμματος περιήγησης. Όλες σχεδόν οι ενέργειες του προγράμματος περιήγησης ή αλλαγές κατάστασης καταγράφονται ως συμβάντα από ένα component (συστατικό/μέρος) του προγράμματος. Για παράδειγμα, τα συμβάντα δημιουργούνται, όταν ανοίγει μια νέα διεύθυνση URL, όταν δεν είναι δυνατή η εύρεση μιας σελίδας που ζητήθηκε ή όταν έχει ολοκληρωθεί η λήψη ενός πόρου. Σε απόκριση σε ένα συμβάν, μια επέκταση του προγράμματος περιήγησης μπορεί να ζητήσει περισσότερες πληροφορίες, χρησιμοποιώντας τις διεπαφές COM (Component Object Model) που προσφέρονται από το πρόγραμμα περιήγησης. Κατ' αυτόν τον τρόπο ένα spyware μπορεί να αλληλεπιδράσει με το πρόγραμμα περιήγησης ιστού, επικαλούμενο τις λειτουργίες του ως απόκριση σε συμβάντα. Αφού, λοιπόν, το spyware έχει εξαγάγει τις επιθυμητές πληροφορίες, η επόμενη εργασία του είναι να μεταδώσει τα δεδομένα σε ένα τρίτο μέρος. Για το σκοπό αυτό, οι πληροφορίες πρέπει είτε να μεταφερθούν απευθείας στο δίκτυο (ανοίγοντας μια σύνδεση με τον τρίτο στις γραμμές εντολών), είτε να μεταβιβαστούν σε μια συνεργαζόμενη διαδικασία που εκτελείται στον ίδιο κεντρικό υπολογιστή, είτε να αποθηκεύονται τοπικά (για παράδειγμα, σε ένα αρχείο στο δίσκο ή στο μητρώο των Windows). [16]

Μια ακόμα ειδική περίπτωση είναι ότι ένα spyware θα μπορούσε να επιχειρήσει να εγκαταστήσει και ρυθμίσει μια επιπρόσθετη διαδρομή, μέσω της οποίας τα γεγονότα μπορούν να διαρρεύσουν. Αυτό επιτυγχάνεται ξεκινώντας ένα άλλο νήμα ή καταχωρίζοντας ένα χρονόμετρο με μια κατάλληλη λειτουργία επανάκλησης. Σε απάντηση σε ένα συμβάν προγράμματος περιήγησης, οι συλλεγμένες πληροφορίες χρήστη δε θα διαρρεύσουν αμέσως, αλλά πρώτα θα αποθηκευτούν σε μια δομή δεδομένων προσβάσιμη σε παγκόσμιο επίπεδο. Αργότερα, το δεύτερο νήμα ή η λειτουργία επανάκλησης του χρονοδιακόπτη θα μπορούσε να εξαλείψει αυτές τις πληροφορίες, χωρίς να εντοπιστεί. Εκτός εάν, κάθε φορά που ένα component δημιουργεί ένα νήμα ή ένα χρονόμετρο, όλες οι κλήσεις API των Windows που καλεί αυτό το component καταγράφονται (όχι μόνο εκείνες που επικαλούνται ως ανταπόκριση στα συμβάντα). Εν πάση περιπτώσει, το spyware χρειάζεται να αλληλεπιδράσει με το λειτουργικό σύστημα. Και αυτό, προκειμένου με τη χρήση μιας υπηρεσίας του λειτουργικού συστήματος να μπορέσει να παραδώσει τα υποκλεμμένα δεδομένα στον τρίτο αποδέκτη. [16]

Τα Spyware ποικίλλουν και χρησιμοποιούν διάφορες μεθόδους ή λογισμικά κατασκοπίας, ή ακόμα και επίθεσης προς κάποιο χρήστη (π.χ. Malware). Μια σημαντική μέθοδος καταγραφής είναι η keystroke logger ή αλλιώς keylogging ή keyboard capturing. Αυτή η μέθοδος καταγράφει τις πληροφορίες που εισέρχονται από μια συσκευή εισόδου. Φερ' ειπείν, μπορεί να καταγράψει τα κουμπιά που πάτησε ένας χρήστης στο πληκτρολόγιο του. Συνεπώς, όταν κάποιος πληκτρολογεί τον κωδικό πρόσβασης σε μια διαδικτυακή ή

τοπική πλατφόρμα, ο κωδικός μπορεί να καταγραφεί σε ένα αρχείο, το οποίο έπειτα μπορεί να αποσταλεί σε κάποιον τρίτο με τις μεθόδους που προαναφέρθηκαν. Μια μέθοδος `keystroke logger` υλοποιείται, έχοντας πρόσβαση σε επίπεδο πυρήνα και ελέγχοντας τις πληροφορίες (σε δυαδική μορφή) που στέλνονται από τη συσκευή εισόδου (π.χ. πληκτρολόγιο), χρησιμοποιώντας API του λειτουργικού συστήματος (π.χ. με την εντολή `GetAsyncKeyState()` στο windows API), με κώδικα Javascript (π.χ. το συμβάν `onKeyUp()`) και με άλλους τρόπους. Άλλο, αλλά μάλλον πιο ακίνδυνο, είδος Spyware είναι το Adware. Οι εφαρμογές που χαρακτηρίζονται ως Adware παρακολουθούν τη δραστηριότητα περιήγησης στο Web του χρήστη (με τις μεθόδους που προαναφέρθηκαν), ώστε να στέλνουν στοχευόμενες διαφημίσεις στον υπό παρακολούθηση χρήστη. Μάλιστα, πολλές εφαρμογές Adware απομακρυσμένης παρακολούθησης χαρακτηρίζονται από τους δημιουργούς τους ως νόμιμες. Εντούτοις, υπάρχουν και μη νόμιμα Adware, που εκτελούν μη επιθυμητές ενέργειες, όπως η αλλαγή των προεπιλεγμένων ρυθμίσεων σε ένα πρόγραμμα περιήγησης. [15, 17]

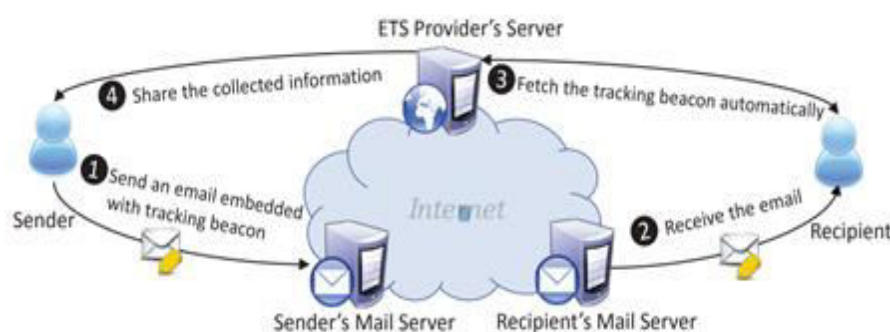
2.5) Παρακολούθηση ηλεκτρονικής αλληλογραφίας - Email Tracking

Ο όρος `email tracking` αναφέρεται σε διάφορες μεθόδους και τεχνικές, που επιτρέπουν την παρακολούθηση και καταγραφή δεδομένων μέσα από τα ηλεκτρονικά μηνύματα. Το ηλεκτρονικό ταχυδρομείο ξεκίνησε ως μη αλληλεπιδραστικό πρωτόκολλο για την αποστολή απλών μηνυμάτων κειμένου. Ωστόσο, οι σύγχρονες υπηρεσίες ηλεκτρονικού ταχυδρομείου υποστηρίζουν μεγάλο μέρος της λειτουργικότητας του ιστού και η παρακολούθηση ιστού από τρίτο μέρος επεκτάθηκε επίσης στην ηλεκτρονική αλληλογραφία. Ο φαινομενικός σκοπός της παρακολούθησης ηλεκτρονικού ταχυδρομείου είναι για τους αποστολείς, να γνωρίζουν ποια μηνύματα έχουν διαβαστεί από τους παραλήπτες. Μολαταύτα, διαπιστώνεται ότι η παρακολούθηση ηλεκτρονικού ταχυδρομείου, είναι πολύ πιο εξελιγμένη, καθώς συνδέεται με τα cookies των χρηστών και συνεπώς με τις δραστηριότητές τους στον ιστό. Κατά συνέπεια, η παρακολούθηση ηλεκτρονικού ταχυδρομείου επιτρέπει στον αποστολέα να συλλέγει λεπτομερή δεδομένα σχετικά με τη συμπεριφορά, την ώρα ανάγνωσης του μηνύματος, τη συσκευή και το λειτουργικό σύστημα που χρησιμοποιεί ο παραλήπτης, καθώς και την IP διεύθυνσή του. Μάλιστα, ενδέχεται ένα μεγάλο δίκτυο τρίτων μερών να λαμβάνει επίσης αυτή την πληροφορία. Επιπλέον, όταν οι χρήστες κάνουν κλικ σε συνδέσμους των emails ή ακόμα και στα ίδια τα emails για ανάγνωσή τους, ανεξάρτητα από την υπηρεσία ηλεκτρονικού ταχυδρομείου, αποκαλύπτεται η διεύθυνση ηλεκτρονικού ταχυδρομείου τους (π.χ. Nikos@email.com) στους ιχνηλάτες. Τοιουτοτρόπως, είναι συχνά δυνατή η σύνδεση μιας διεύθυνσης ηλεκτρονικού ταχυδρομείου με προσωπικά ηλεκτρονικά προφίλ, όπως σε ιστοτόπους κοινωνικών μέσων. Σημειώνεται ότι, παρόλη την εξάπλωση και εφαρμογή του `email tracking` στα ηλεκτρονικά μηνύματα οι έρευνες σε αυτόν το πλαίσιο είναι λίγες (μέχρι και το 2018). [18, 19]

Η παρακολούθηση μέσω ηλεκτρονικού ταχυδρομείου είναι εφικτή, επειδή οι σύγχρονες υπηρεσίες ηλεκτρονικού ταχυδρομείου επιτρέπουν τη δημιουργία email σε μορφή HTML. Τοιουτοτρόπως επιτρέπεται η προσθήκη, όχι μόνο απλού κειμένου, αλλά και

ενσωματωμένων εικόνων Web beacons (που μπορούν να αναφέρουν ένα αντικείμενο παρακολούθησης που είναι αποθηκευμένο σε ένα διακομιστή του αποστολέα ή τον παροχέα εντοπισμού του, με τη μορφή ενός URL (Uniform Resource Locator)), φύλλων στυλ (stylesheets), υπερσυνδέσμων και άλλων στοιχείων σε ένα email. Επομένως, όταν εμφανίζονται web beacons ή όταν γίνονται κλικ σε υπερκείμενα ή εικόνες με συνδέσμους, το πρόγραμμα περιήγησής του θα πρέπει να στείλει ένα αίτημα HTTP GET, για να μεταβεί στην ιστοσελίδα ή να παραλάβει την εικόνα από το διακομιστή, όπου φιλοξενείται. Γιατί, προκειμένου να παρουσιαστεί η εικόνα ή η ιστοσελίδα, ο διακομιστής της οφείλει να ξέρει το πού θα την παρουσιάσει. Άρα ο φυλλομετρητής θα στείλει και θα δεχθεί cookies τού χρήστη ή κάποιες άλλες κατ' ελάχιστον πληροφορίες (π.χ. διεύθυνση IP) μαζί με το αίτημά του. Αφού παραληφθούν οι πληροφορίες αυτές από το πρωτόκολλο επικοινωνίας τού διακομιστή, ο διακομιστής μπορεί να στείλει πίσω στον αποστολέα τού μηνύματος ή ακόμα και σε άλλα τρίτα μέρη τις πληροφορίες που έλαβε. Και αυτό γίνεται χρησιμοποιώντας μεθόδους ανακατεύθυνσης (redirection) και επικεφαλίδες παραπομπής (href). Είναι συναρπαστικό το γεγονός ότι όλα αυτά υλοποιούνται μόνο με HTML. Και μάλιστα δε χρησιμοποιείται JavaScript, καθώς απαλείφεται, αν όχι πάντοτε, σχεδόν πάντοτε από τις υπηρεσίες ηλεκτρονικού ταχυδρομείου και από τους φυλλομετρητές στα emails για λόγους ασφαλείας. [14, 18, 19, 20]

Εκτός από την HTML, είναι εφικτή και η χρήση της γλώσσας PHP σε ένα email, επιφέροντας πολλά παρόμοια σημαντικά αποτελέσματα στον αποστολέα με πιο άμεσο τρόπο. Ειδικότερα, μπορεί να λαμβάνει πληροφορίες σχετικά με τη συσκευή πρόσβασης του παραλήπτη και τη συμπεριφορά ανάγνωσης ηλεκτρονικού ταχυδρομείου με απλές εντολές, όπως η `$_SERVER['REMOTE_ADDR']` που επιτρέπει τη λήψη διεύθυνσης IP. Παράλληλα, μπορεί αυτές τις πληροφορίες να τις καταγράφει σε αρχεία log files. Με τις αυτοματοποιημένες αυτές διαδικασίες επιτρέπεται και ένα είδος cross-tracking (διασταυρωμένη ιχνηλάτηση). Εάν για παράδειγμα το email ανοίξει σε διαφορετικές συσκευές, κάθε μεμονωμένη πρόσβαση καταγράφεται με τις αντίστοιχες πληροφορίες χρήστη-παράγοντα, η οποία επιτρέπει την παρακολούθηση μεταξύ συσκευών. Εναλλακτική χρήση της PHP (ή και παρόμοιας γλώσσας) θα χρειαστεί ένα Web beacon, το οποίο κατά την προβολή του εκτελεί και ένα σύνολο εντολών PHP. Και αυτό, διότι ο σύνδεσμος που αναπαράγει το Web beacon είναι υλοποιημένος σε PHP. Οι εντολές αυτές επιτρέπουν τη λήψη και αποθήκευση προσωπικών δεδομένων από το διακομιστή που στέλνει το email. Έτσι, δε χρειάζεται να λάβει πληροφορίες από άλλο διακομιστή, αλλά ο ίδιος ο αποστολέας λαμβάνει τις επιθυμητές του πληροφορίες, χωρίς μεσολάβηση άλλου. Ένα τέτοιο παράδειγμα κώδικα σε γλώσσα PHP προαναφέρθηκε και στην εικόνα 4. [19, 21]



Εικόνα 5: Πώς γίνεται η παρακολούθηση ηλεκτρονικής αλληλογραφίας.

Πολλές εταιρείες προσφέρουν τέτοιες υπηρεσίες σε αποστολές email, επιτρέποντάς τους περαιτέρω αναλύσεις και συνεπώς βελτιώσεις των υπηρεσιών τους. Μια από αυτές είναι και η G-Lock Analytics με πιο φιλικό περιβάλλον αλληλεπίδρασης με το χρήστη. Οι χρήστες της υπηρεσίας G-Lock Analytics μπορούν να καταλάβουν με σαφήνεια τί συμβαίνει με τα μηνύματα ηλεκτρονικού ταχυδρομείου μετά την αποστολή τους. Η υπηρεσία προσφέρει πλήρη και σαφέστατη καθοδήγηση, επιτρέποντάς τους να παρακολουθήσουν τα κλικ μέσω του δικού τους τομέα, να μετρήσουν το ενδιαφέρον των συνδρομητών για το περιεχόμενό τους και την προσφορά τους με την καταγραφή επαναλαμβανόμενων προβολών και κλικ. Ακόμα, προσφέρει γεωγραφικό εντοπισμό μοναδικών μηνυμάτων που ανοίχτηκαν, αναφορές σε πραγματικό χρόνο σχετικά με την απόδοση της καμπάνιας ηλεκτρονικού ταχυδρομείου, εξαγωγή αναφορών παρακολούθησης ηλεκτρονικού ταχυδρομείου σε ένα αρχείο. Η πλατφόρμα αυτή είναι πολύ απλή στη χρήση και αποτελεί ακόμα μια απόδειξη της μεγάλης σημασίας, που έχει η τεχνολογία email tracking. [21]

2.6) Υπηρεσίες εντοπισμού θέσης LBS

Πρόσφατες τεχνολογικές εξελίξεις στην ασύρματη παρακολούθηση τοποθεσίας, όπως αυτές που βρίσκονται στα κινητά τηλέφωνα ή οι ετικέτες ταυτοποίησης μέσω ραδιοσυχνοτήτων (RFID), παρουσιάζουν πρωτοφανείς ευκαιρίες βελτίωσης των υπηρεσιών εντοπισμού θέσης LBS (Local Based Services). Οι LBS είναι υπηρεσίες που παρέχονται στον συνδρομητή με βάση την τρέχουσα γεωγραφική του θέση. Δυο βασικές εξυπηρετήσεις μιας τέτοιας τεχνολογίας είναι η εύρεση της τωρινής τοποθεσίας και ο εύκολος προσανατολισμός της επόμενης. Για τον εντοπισμό μιας τοποθεσίας στον πραγματικό κόσμο (π.χ. Ελευθερίου Βενιζέλου 18, Περιστέρι, Αθήνα, Ελλάδα), οι LBS χρησιμοποιούν συνήθως API (διεπαφή προγραμματισμού εφαρμογών) και χάρτες. Υπενθυμίζεται ότι, το API είναι ένα σύνολο ορισμών υπορουτίνας, πρωτόκολλα επικοινωνίας και εργαλεία για την οικοδόμηση λογισμικού, αποτελώντας μια σειρά σαφώς καθορισμένων μεθόδων επικοινωνίας μεταξύ των διαφόρων συνιστωσών. Εν ολίγοις το API είναι μια μέθοδος επικοινωνίας μεταξύ συστημάτων. Οι χάρτες είναι μια οπτικοποίηση γεωγραφικών δεδομένων (λ.χ. συντεταγμένες), που υλοποιείται και χρησιμοποιείται με γλώσσες προγραμματισμού markup, script και άλλες τεχνικές ανάπτυξης ιστού. Επομένως, οι χάρτες δεν είναι εικόνες, αλλά οπτικοποιημένα δεδομένα. Ο συνδυασμός των χαρτών με API προσφέρει μια ανάλυση κυκλοφορίας σε πραγματικό χρόνο (realtime traffic analysis), καθώς το API επικοινωνεί πολλαπλώς και επανειλημμένως με άλλα συστήματα, προσφέροντάς τους πληθώρα γεωγραφικών πληροφοριών. Κατά αυτόν τον τρόπο συνήθως λειτουργούν οι LBS υπηρεσίες. [22, 23]

Οι LBS μπορούν να ταξινομηθούν σε τρεις τύπους: σποραδικά ερωτήματα (sporadic queries), επίγνωση θέσης (position awareness) και εντοπισμός τοποθεσίας (location tracking). Τα σποραδικά ερωτήματα αφορούν τις υπηρεσίες, στις οποίες ένα άτομο κάνει

ερωτήματα (queries), μεταφέροντας την τρέχουσα μόνο θέση του σε εξωτερική υπηρεσία. Η υπηρεσία αυτή δίνει απαντήσεις βάσει των πληροφοριών θέσης που δέχθηκε (π.χ. πού βρίσκεται το πλησιέστερο ξενοδοχείο από αυτήν την τοποθεσία). Η επίγνωση θέσης αναφέρεται σε συσκευές που παρακολουθούν τη θέση ενός ατόμου (π.χ. τα συστήματα πλοήγησης στο αυτοκίνητο (GPS)), αλλά χρησιμοποιούν μόνο αυτές τις πληροφορίες εσωτερικά. Οι υπηρεσίες εντοπισμού τοποθεσίας λαμβάνουν συχνές ενημερώσεις της θέσης ενός ατόμου, οι οποίες επιδιώκουν τη βελτίωση των μεταφορών μέσω της χρήσης των πληροφοριών διαδρομής που συλλέγονται. Για παράδειγμα μπορούν να χρησιμοποιηθούν αυτές οι ενημερώσεις, για να εκτιμηθεί η κυκλοφορία ενός αυτοκινητόδρομου. Συνεπώς μπορούν να βελτιωθούν οι οδηγίες που δίνονται από εφαρμογές προς τους οδηγούς οχημάτων ή να αποφευχθούν οι κυκλοφοριακές συμφορές. Ας σημειωθεί ότι αυτή δεν είναι η μοναδική κατηγοριοποίηση. Θα ήταν εξίσου λογικό κάποιος να συγχωνεύσει τους τύπους επίγνωσης θέσης και εντοπισμού τοποθεσίας ή κάποιος άλλος να αναλύσει τις LBS υπηρεσίες σε παραπάνω ή λιγότερες κατηγορίες. [22, 23]

Οι LBS συνήθως περιλαμβάνουν τουλάχιστον τρία μέρη: τον αιτούντα χρήστη, την τεχνολογία παροχής εντοπισμού θέσης και τουλάχιστον έναν φορέα παροχής υπηρεσιών. Ο χρήστης είναι το άτομο που χρησιμοποιεί μια τεχνολογία παροχής εντοπισμού θέσης, όπως μια εφαρμογή στο έξυπνο κινητό τηλέφωνό του. Με αυτήν ο χρήστης μπορεί να κάνει αιτήματα (π.χ. πώς να πάει στο κέντρο της πόλης από την τρέχουσα διεύθυνσή του), τα οποία αποστέλλονται από την τεχνολογία παροχής εντοπισμού θέσης και λαμβάνονται από το φορέα παροχής υπηρεσιών. Ο οργανισμός, που παρέχει την τεχνολογία παροχής εντοπισμού θέσης, χρησιμοποιεί έναν μεσίτη/μεσολαβητή τοποθεσίας για την παράδοση της θέσης του χρήστη στους φορείς παροχής υπηρεσιών. Για παράδειγμα, οι πάροχοι κινητής τηλεφωνίας, που αναπτύσσουν τις τεχνολογίες εντοπισμού θέσης για τα κινητά τους τηλέφωνα, θα μπορούσαν να εφαρμόσουν έναν τέτοιο μεσίτη σε ένα διακομιστή στο δίκτυό τους. Με αυτόν τον τρόπο, παραδίδουν πληροφορίες θέσης κινητού τηλεφώνου προς εξωτερικούς παρόχους υπηρεσιών, οι οποίοι ενδέχεται να προσφέρουν υπηρεσίες εύρεσης ξενοδοχείων ή εφαρμογές πλοήγησης. Τα συστήματα και δίκτυα παράδοσης της πληροφορίας (π.χ. Cellular Network) στηρίζονται σε τεχνικές τηλεπικοινωνίας, οι οποίες δεν σχετίζονται με την παρούσα εργασία και τον κλάδο της. [22]

Τα προβλήματα προστασίας της ιδιωτικής ζωής ενδέχεται να παρεμποδίσουν αρκετούς χρήστες από την αποδοχή χρήσης των LBS. Παρόλο που ορισμένοι χρήστες θα μπορούσαν συνειδητά να επιτρέπουν τη συνεχή ή μερική παρακολούθηση τοποθεσίας, λίγοι θα ήταν άνετοι με τη γνωστοποίηση της θέσης τους σε όλες τις περιπτώσεις (π.χ. όταν πηγαίνουν στον ιατρό). Για αυτούς τους λόγους, διευρύνονται αλγόριθμοι ελέγχου αποκάλυψης τοποθεσίας, που ωστόσο αποκρύπτουν τις θέσεις των χρηστών στις ευαίσθητες περιοχές (που οι ίδιοι οι χρήστες έχουν ορίσει) και παρακρατούν τις πληροφορίες διαδρομής που υποδεικνύουν ποιες περιοχές έχουν επισκεφθεί. Αυτοί προσφέρουν στους χρήστες μεθόδους για τον ορισμό των ευαίσθητων περιοχών, των κατηγοριών παρόχων υπηρεσιών ή των ωρών της ημέρας που δεν επιθυμούν να γνωστοποιείται η θέση τους. Συνεπώς, οι χρήστες μπορούν να προσαρμόσουν μια πολιτική ιδιωτικότητας, προσαρμοσμένη ανάλογα με τις ανάγκες τους. Για παράδειγμα μια πολιτική, που εφαρμόζεται στον μεσολαβητή τοποθεσίας, χρησιμοποιώντας τους κατάλληλους

χάρτες, θα μπορούσε να επιτρέψει την παρακολούθηση τοποθεσίας σε δημόσιους δρόμους, αλλά να αποτρέψει την παρακολούθηση σε κτίρια ή σε ιδιωτική ιδιοκτησία. Άλλες πολιτικές θα μπορούσαν να κάνουν περαιτέρω διάκριση μεταξύ τύπων κτιρίων (όπως οικιστικών ή εμπορικών). Εντούτοις, η ισχυρή προστασία της ιδιωτικής ζωής δεν πρέπει να μειώνει άσκοπα τη διαθεσιμότητα πληροφοριών θέσης σε μη ευαίσθητες περιοχές, για τις οποίες οι χρήστες δεν ενοχλούνται για τη γνωστοποίηση της επίσκεψής τους. [22]

Χρησιμοποιείται, λοιπόν, μεσολαβητής τοποθεσίας που περιέχει δύο βασικά στοιχεία λογισμικού: α) έναν διαχειριστή ειδοποιήσεων που παρακολουθεί τις αιτήσεις θέσης από εξωτερικούς παρόχους υπηρεσιών και β) έναν διαχειριστή απορρήτου που αποθηκεύει τις προτιμήσεις απορρήτου των χρηστών και εκτελεί τους αλγόριθμους ελέγχου αποκάλυψης για να προσδιορίσει, εάν οι τοποθεσίες μπορούν να αποκαλυφθούν στον αιτούντα πάροχο υπηρεσιών. Πιο συγκεκριμένα, τα στοιχεία αλληλοεπιδρούν ως εξής. Η συσκευή του χρήστη ενημερώνει περιοδικά την τοποθεσία του στο μεσίτη τοποθεσίας. Για κάθε ενημέρωση τοποθεσίας από χρήστες, ένας διαχειριστής ειδοποίησης προσδιορίζει τους παρόχους υπηρεσιών που επιθυμούν να λάβουν αυτές τις ενημερώσεις. Πριν από την προώθηση μιας ενημέρωσης τοποθεσίας σε έναν παροχέα υπηρεσιών, ο διαχειριστής ειδοποιήσεων ζητά άδεια από τον διαχειριστή απορρήτου, ο οποίος ελέγχει την πολιτική απορρήτου του χρήστη για οποιονδήποτε περιορισμό περιγράφεται σε αυτήν. Αυτοί οι περιορισμοί αναπαρίστανται σε έναν χάρτη ευαισθησίας θέσης, ο οποίος έχει ταξινομήσει τις τοποθεσίες είτε ως μη ευαίσθητες είτε ως ευαίσθητες, σύμφωνα με τις προτιμήσεις του χρήστη (π.χ. όρισε κάθε ιατρείο ως ευαίσθητη περιοχή). Εάν αυτοί οι περιορισμοί επιτρέπουν την απελευθέρωση δεδομένων, τότε ο διαχειριστής απορρήτου διεξάγει την ανάλυση ευαισθησίας (χρησιμοποιώντας τους αλγόριθμους) για την επίτευξη τελικής απόφασης. Αυτή η διαδικασία επαναλαμβάνεται για κάθε πάροχο υπηρεσιών. Μάλιστα, προκειμένου να αυξηθεί η ασφάλεια ιδιωτικότητας, οι αλγόριθμοι μειώνουν τις πληροφορίες θέσης όχι μόνο στις ευαίσθητες περιοχές που επισκέπτεται ένα άτομο, αλλά και γύρω από αυτές. Παρόλα αυτά, σε πυκνές περιοχές, όπως οι πόλεις, ο χρήστης περιβάλλεται συχνά από τέτοιες ευαίσθητες περιοχές. Επομένως, είναι δύσκολο να προσδιοριστεί, εάν ένας χρήστης πρόκειται να εισέλθει στην περιοχή ή απλά περνάει δίπλα από αυτήν. [22]

Συγκρίνονται τρεις αλγόριθμοι, για να χαρακτηριστεί το μέγεθος αυτού του προβλήματος και να αναλυθούν οι πιθανές λύσεις σε καθέναν από αυτούς. Το πείραμα με αυτούς τους αλγόριθμους προϋποθέτει έναν χάρτη ευαισθησίας, που ταξινομεί τις τοποθεσίες είτε ως μη ευαίσθητες είτε ως ευαίσθητες.

Αλγόριθμος Base

Ο αλγόριθμος βάσης απελευθερώνει μόνο ενημερώσεις τοποθεσίας σε περιοχές που έχουν ταξινομηθεί ως μη ευαίσθητες στον χάρτη ευαισθησίας. Ο αλγόριθμος ελέγχει ανεξάρτητα τον χάρτη ευαισθησίας για κάθε ενημέρωση τοποθεσίας που αποστέλλει ο πελάτης στον μεσίτη.

Αλγόριθμος Bounded-Rate

Εκτός από τους ελέγχους απορρήτου του αλγόριθμου βάσης, ο αλγόριθμος οριακού ρυθμού διασφαλίζει επιπλέον ότι οι ενημερώσεις δεν αποστέλλονται με συχνότητα υψηλότερη από ένα προκαθορισμένο όριο. Έτσι, για μικρότερα όρια συχνότητας, αυτός ο αλγόριθμος μειώνει την ποσότητα των πληροφοριών που απελευθερώνονται σε ευαίσθητες περιοχές, ώστε να δυσχεράνει τον πάροχο υπηρεσιών να συμπεράνει τις επισκέψεις του χρήστη σε περιοχές, που είναι κοντά στις ευαίσθητες.

Αλγόριθμος k-Area

Για τον αλγόριθμο k περιοχής χρειάζεται να χωριστεί ένας χάρτης ευαισθησίας σε ζώνες που περιλαμβάνουν τουλάχιστον k ξεχωριστές ευαίσθητες περιοχές (ο k είναι ένας ακέραιος αριθμός). Για κάθε ζώνη επισκέψεων, ο αλγόριθμος καταστέλλει τις ενημερώσεις τοποθεσίας, μόνο όταν ένα άτομο εισέρχεται σε μια ευαίσθητη περιοχή. Επιπλέον, οι ενημερώσεις τοποθεσίας απελευθερώνονται, μόνο όταν δεν αποκαλύπτουν ποια από τις τουλάχιστον k ευαίσθητες περιοχές επισκέφθηκε ο χρήστης. [22]

Για να ελεγχθεί η προστασία της ιδιωτικότητας ανά τοποθεσία, όταν εισέρχονται άτομα σε ευαίσθητες περιοχές, και ο επηρεασμός της ακρίβειας της θέσης σε δημόσιους χώρους, χρησιμοποιήθηκαν σε έρευνα μοντέλα κίνησης από το City Simulator. Αυτό είναι ένα εργαλείο που εμπεριέχει μοντέλα προσομοίωσης από τμήματα πόλεων. Ο αλγόριθμος Base και ο Bounded-Rate με όριο έως και ένα λεπτό αποδείχθηκαν ότι δεν προσφέρουν σχεδόν καμία προστασία. Οι αλγόριθμοι k-περιοχής εμπόδισαν λιγότερο από το 15% των ενημερώσεων τοποθεσίας σε δημόσιους χώρους, ενώ ο αλγόριθμος οριακού ρυθμού παρακράτησε μεταξύ 50% και 75%. Εν κατακλείδι, τα αποτελέσματα του πειράματος δείχνουν ότι ο αλγόριθμος k-Area είναι ο πιο αποτελεσματικός, όσον αφορά την προστασία των ευαίσθητων τοποθεσιών, χωρίς να μειώνεται πάρα πολύ η ακρίβεια της θέσης στις μη ευαίσθητες περιοχές. Ο αλγόριθμος οριακού ρυθμού μειώνει τη συνολική ακρίβεια, επειδή καταστέλλει περισσότερες ενημερώσεις τοποθεσίας σε ευαίσθητες περιοχές. Η αντιστάθμιση, ωστόσο, είναι ότι ο αλγόριθμος k περιοχής καθυστερεί τη μετάδοση ενημέρωσης θέσης σε ευαίσθητες περιοχές, έως ότου ο χρήστης εγκαταλείψει τη ζώνη. [22]

2.7) Εξακρίβωση ταυτότητας

Η εξακρίβωση ταυτότητας (identification) είναι μια μη συγκεκριμένη, αλλά αφηρημένη τεχνική. Χρησιμοποιείται από ιχνηλάτες (trackers) για να διακρίνουν χρήστες του διαδικτύου και μετέπειτα να συλλέγουν πιο αποτελεσματικά τα δεδομένα τους. Μπορεί, επίσης να χρησιμοποιηθεί και για έλεγχο ταυτότητας πολλαπλών παραγόντων, για ενίσχυση της ασφαλείας, για παροχή εξατομικευμένων και ειδικότερων υπηρεσιών ανά χρήστη του διαδικτύου. Γενικώς, η διάκριση των διάφορων χρηστών του διαδικτύου και η δυνατότητα επικοινωνίας μεταξύ διαφορετικών ιστοτόπων επιτυγχάνεται, αναθέτοντας ένα αναγνωριστικό σε κάθε χρήστη του διαδικτύου. Κατά συνέπεια, οι ιστότοποι έχουν ένα σημείο αναφοράς, προκειμένου να διακρίνουν τους επισκέπτες τους, και μπορούν να αλληλομοιράζονται καλύτερα τις πληροφορίες, που ο καθένας τους συλλέγει. Το αναγνωριστικό είναι μια απλή συμβολοσειρά και παράγεται από μεθόδους, όπως είναι τα ETags και οι τεχνικές fingerprinting. Βέβαια, είναι σημαντικό να δημιουργείται με

τυχαιότητα, προκειμένου να μπορούν να μειωθεί η πιθανότητα εύρεσης δυο ή περισσότερων ίδιων αναγνωριστικών. Η τυχαιότητα μπορεί να οφείλεται σε πολλούς παράγοντες, όπως από την προσθήκη μιας χρονικής σήμανσης (timestamp). Το αναγνωριστικό είναι μια ακόμα πληροφορία ενός ατόμου μαζί με τις υπόλοιπες πληροφορίες που αντλούνται από τον ιστότοπο (π.χ. διεύθυνση IP, αριθμός κλικ σε διαφημίσεις, κλπ.). [5, 11]

Το ETag αποτελεί μέρος του πρωτοκόλλου HTTP και είναι ένα αδιαφανές αναγνωριστικό. Αυτό το αναγνωριστικό εκχωρείται μέσω του HTTP από έναν διακομιστή ιστού σε έναν πόρο, που βρίσκεται σε μια διεύθυνση URL. Εάν η αναπαράσταση των πόρων στη συγκεκριμένη διεύθυνση URL αλλάξει, τότε εκχωρείται ένα νέο και διαφορετικό ETag. Με αυτό τον τρόπο τα ETags μπορούν να συγκριθούν γρήγορα, για να προσδιοριστεί, αν οι δύο ή περισσότερες αναπαραστάσεις ενός πόρου είναι οι ίδιες. Οι τιμές ETag μπορούν να χρησιμοποιηθούν σε συστήματα παρακολούθησης ιστοσελίδων (π.χ. cookies), ώστε να ελέγχεται για παράδειγμα η συχνότητα επισκεψιμότητας σε ένα συγκεκριμένο πόρο. Μάλιστα, επειδή τα ETags αποθηκεύονται από το πρόγραμμα περιήγησης και επιστρέφονται με τα επόμενα αιτήματα για τον ίδιο πόρο, ένας διακομιστής παρακολούθησης μπορεί να επαναλαμβάνει το κάθε ETag που λαμβάνει από ένα συγκεκριμένο πρόγραμμα περιήγησης. Έτσι, θα διασφαλίσει ότι το ETag (ή καλύτερα η τιμή αυτού) θα παραμείνει για πάντα σε αυτόν τον υπολογιστή, όπως γίνεται και με την ανάκτηση των cookies. [24]

Στην HTML5 υπάρχει ένα στοιχείο (element) που ονομάζεται canvas (καμβάς) και επιτρέπει τη δυναμική υλοποίηση δισδιάστατων (ή ακόμα και τρισδιάστατων) σχημάτων, κινουμένων σχεδίων (animations) και εικόνων bitmap με τη χρήση Javascript. Πρόκειται για ένα μοντέλο χαμηλού επιπέδου, το οποίο χρησιμοποιεί ένα API, που ονομάζεται canvas API, για την κλήση συναρτήσεων, που σχεδιάζουν και κάνουν render. Η χρήση του στοιχείου canvas και του canvas API των σύγχρονων φυλλομετρητών μπορεί να υλοποιήσει το canvas fingerprinting (δακτυλικό αποτύπωμα καμβά), που είναι μια τεχνική διάκρισης χρηστών του διαδικτύου. Με αυτήν, ένας ιστότοπος μπορεί να εκμεταλλευτεί τις λεπτές διαφορές στην υλοποίηση των δημιουργημάτων (κειμένων, σχημάτων, WebGL σκηνών, κλπ.) του καμβά. Συνεπώς, εξάγει ένα fingerprint (δακτυλικό αποτύπωμα) του επισκέπτη, το οποίο μπορεί εύκολα να ληφθεί σε κλάσματα του δευτερολέπτου χωρίς την επίγνωσή του. Για παράδειγμα, μια συγκεκριμένη γραμματοσειρά μπορεί να αποτυπωθεί με διάφορους τρόπους ανά υπολογιστή, ανάλογα με το λειτουργικό σύστημα, τη βιβλιοθήκη γραμματοσειρών, την κάρτα γραφικών, το πρόγραμμα οδήγησης γραφικών και το πρόγραμμα περιήγησης του υπολογιστή. Αυτό μπορεί να οφείλεται στις διαφορές μεθόδων rasterization (μέθοδος μετασχηματισμού σε pixels) του συστήματος του χρήστη, όπως η εξομάλυνση της αλλοιώσεως (anti-aliasing) ή η εξομάλυνση υπο-εικονοστοιχείων (sub-pixel smoothing). Ακόμη, μπορεί να οφείλεται στις διαφορές των γραμματοσειρών, στη φυσική απεικόνιση και σε άλλους παρόμοιους παράγοντες, που επηρεάζουν τον τρόπο απεικόνισης. [11, 12]

Όταν ένας χρήστης επισκέπτεται μια σελίδα, το fingerprinting script πρώτα απεικονίζει κείμενο με τη γραμματοσειρά και το μέγεθος της επιλογής του χρήστη και προσθέτει χρώματα φόντου. Στη συνέχεια, καλεί τη μέθοδο ToDataURL του canvas API για

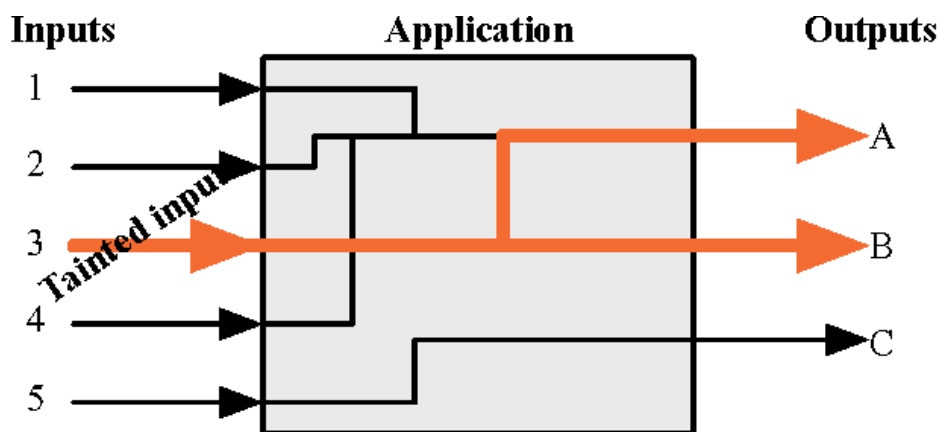
να πάρει τα δεδομένα εικονοστοιχείων καμβά σε μορφή dataURL, η οποία είναι βασικά μια κωδικοποίηση βασισμένη στο Base64 των δυαδικών δεδομένων εικονοστοιχείων. Μετέπειτα, το fingerprinting script λαμβάνει τη hash τιμή από τα δεδομένα των εικονοστοιχείων που κωδικοποιούνται με κείμενο (text encoded pixel data), η οποία αποτελεί το fingerprint αποτύπωμα. Αυτή μπορεί επίσης να συνδυαστεί με άλλες ιδιότητες του προγράμματος περιήγησης υψηλής εντροπίας, όπως η λίστα των plugin, η λίστα γραμματοσειρών, το ποσοστό μεγέθυνσης (zoom) που χρησιμοποιείται ή το user agent string. Τοιουτοτρόπως, μειώνεται η πιθανότητα να βρεθούν ίδια fingerprints σε διαφορετικούς χρήστες του διαδικτύου. Συνήθως χρησιμοποιείται Javascript για την υλοποίησή του, όπως για παράδειγμα η open source βιβλιοθήκη fingerprintJS. Ένα παράδειγμα αποτελεί ο κώδικας στον παρακάτω σύνδεσμο: <https://codepen.io/run-time/pen/XJNXWV> [11, 12, 25]

Μερικές αξιοσημείωτες μέθοδοι εξακρίβωσης ταυτότητας είναι τα Audio fingerprinting, Battery Based fingerprinting, Font fingerprinting και Panoptick fingerprinting. Το Audio fingerprinting είναι μια τεχνική που λειτουργεί με το Audio Context API. Χρησιμοποιώντας αυτό το μηχανισμό, εισάγεται ήχος από κάποια συσκευή εισόδου του υπολογιστή του χρήστη (π.χ. μικρόφωνο). Έπειτα, ο ήχος επεξεργάζεται και εξάγονται διάφορα χαρακτηριστικά του, όπως το τονικό ύψος, η ένταση του ήχου και άλλα. Βάσει των εξαχθέντων χαρακτηριστικών, παράγονται τα αποτυπώματα ενός χρήστη. Παρόμοια τεχνική είναι και η Battery Based fingerprinting, αφού η δημιουργία των αποτυπωμάτων σύμφωνα με την τρέχουσα κατάσταση μπαταρίας, το χρόνο φόρτισης ή αποφόρτισης. Χρησιμοποιεί λειτουργίες της HTML5, όπως το Battery Status API, αλλά συχνά αποτελεί αδύναμο είδος αποτυπώματος. Το Font fingerprinting είναι η εκτύπωση αναγνωριστικών ανάλογα με τα χαρακτηριστικά της γραμματοσειράς στο πρόγραμμα περιήγησης. Όπως προαναφέρθηκε, αυτή εξαρτάται από πολλές παραμέτρους, όπως την κάρτα γραφικών, το λειτουργικό σύστημα ή τους οδηγούς προγραμμάτων (drivers). Το βέβαιο είναι πως τα αποτυπώματα αυτά είναι δύσκολο να αντιμετωπιστούν. Τελευταίο, το Panoptick fingerprinting είναι μια πολύ συνδυαστική τεχνική αποτυπώματος. Προσδιορίζει τους χρήστες βάσει της ανάλυσης οθόνης, του βάθους χρώματος, της ζώνης ώρας, της έκδοσης των επεκτάσεων, της λίστας γραμματοσειρών που είναι διαθέσιμη στον υπολογιστή. Επίσης, με βάση τις πληροφορίες σχετικά με το αν είναι ενεργοποιημένη ή όχι η Javascript, καθώς και πληροφορίες σχετικά με το αν ορισμένες αιτήσεις αποθήκευσης γίνονται αποδεκτές από το πρόγραμμα περιήγησης. Ένας από τους περιορισμούς της τεχνικής Panoptick fingerprinting είναι ότι δεν είναι σε θέση να προσδιορίσει ταυτόσημα διαμορφωμένες μηχανές (identically configured machines). [26]

2.8) Παρακολούθηση χρωματισμένων δεδομένων - Taint tracking

Η παρακολούθηση των χρωματισμένων δεδομένων (taint tracking) ή αλλιώς η παρακολούθηση της ροής δεδομένων (information flow tracking) αποτελεί μια μέθοδο συνεχούς ελέγχου των πληροφοριών στο σύστημα από άκρο σε άκρο. Ο στόχος της είναι να εξασφαλίσει την προστασία της εμπιστευτικότητας των πληροφοριών που μεταφέρονται στα πληροφοριακά συστήματα, γεγονός που πολλές φορές δεν επιτυγχάνεται με

συμβατικούς μηχανισμούς ασφαλείας, όπως ο έλεγχος πρόσβασης και η κρυπτογράφηση. Με αυτόν τον τρόπο επιβεβαιώνεται ότι τα μυστικά δεδομένα εισόδου δεν μπορεί ένας εισβολέας (μη εξουσιοδοτημένο άτομο ή κακόβουλο πρόγραμμα) να τα προβλέψει και συμπεράνει μέσω των παρατηρήσεών του από την έξοδο του συστήματος. Η χρήση αυτής της μεθόδου στα δυο παρακάτω συστήματα αποτελεί ωφέλιμη λύση αυτού του ζητήματος. [27, 28]



Εικόνα 6: Μεταφορά εισερχόμενων και χρωματισμένων δεδομένων σε εφαρμογή και έξοδος των αλληλοεπιδρώντων δεδομένων.

Η παρακολούθηση της ροής δεδομένων με ευαισθησία στην ιδιωτικότητα παρέχει πολύτιμες πληροφορίες για τους χρήστες των smartphone και γενικότερα του διαδικτύου, καθώς και για τις εταιρείες παροχής υπηρεσιών ασφαλείας που επιδιώκουν να εντοπίσουν κακές εφαρμογές. Πολλές από αυτές τις εφαρμογές συνδυάζουν δεδομένα από απομακρυσμένες υπηρεσίες νέφους (cloud) με πληροφορίες από τοπικούς αισθητήρες, όπως δέκτη GPS, κάμερα, μικρόφωνο και επιταχυνσιόμετρο. Οι αιτήσεις έχουν συχνά νόμιμους λόγους για την πρόσβαση σε αυτά τα δεδομένα που είναι ευαίσθητα στην προστασία της ιδιωτικής ζωής, οι οποίοι αναγράφονται στην πολιτική ιδιωτικότητας και στους όρους χρήσης της εφαρμογής. Μολαταύτα δεν υπάρχει έγκυρη απόδειξη για την τήρηση των λόγων αυτών. Για παράδειγμα, αν επιτρέπει ένας χρήστης μια αίτηση για πρόσβαση στις πληροφορίες τοποθεσίας του, δεν έχει κανέναν τρόπο να γνωρίζει εάν η εφαρμογή θα στείλει την τοποθεσία του σε μια υπηρεσία βάσει τοποθεσίας, στους διαφημιζόμενους, στην εφαρμογή προγραμματιστή ή σε οποιαδήποτε άλλη οντότητα. [27, 28]

Σε μελέτη παρακολούθησης της συμπεριφοράς 30 δημοφιλών εφαρμογών του λειτουργικού συστήματος Android σε VM (εικονική μηχανή) το έτος 2010, βρέθηκαν 20 εφαρμογές που έκαναν δυνητικά κατάχρηση των προσωπικών πληροφοριών των χρηστών (λ.χ. αποστολή προσωπικών δεδομένων σε διακομιστές για απομακρυσμένη διαφήμιση) χάρις το TaintDroid. Το TaintDroid είναι ένα αποτελεσματικό, δυναμικό σύστημα παρακολούθησης και ανάλυσης συστημάτων, ικανό να παρακολουθεί ταυτόχρονα πολλαπλές πηγές ευαίσθητων ή και απλών προσωπικών δεδομένων. Επιτρέπει την ανάλυση σε πραγματικό χρόνο, χρησιμοποιώντας μόλις το 32% των επιδόσεων της

επεξεργαστικής ισχύος (CPU) και επιβάλλει αμελητέα επιβάρυνση σε διαλογικές εφαρμογές τρίτων. Το TaintDroid λειτουργεί με τη υπόθεση, ότι οι εφαρμογές τρίτων που έχουν ληφθεί δεν είναι αξιόπιστες, και παρακολουθεί σε πραγματικό χρόνο πώς αυτές οι εφαρμογές έχουν πρόσβαση και χειραγωγούν τα προσωπικά δεδομένα των χρηστών. Πιο συγκεκριμένα, χρησιμοποιώντας ένα διεργασμένο VM, όταν μια εφαρμογή (π.χ. facebook, calculator, web browser) ζητάει να εκτελέσει μια ή περισσότερες διαδικασίες στην IPC (Inter-process communication - επικοινωνία μεταξύ διαδικασιών), προστίθεται στη διαδικασία της εφαρμογής μια ετικέτα (tag) χρώματος σύμφωνα με συγκεκριμένους κανόνες, καθώς και στα πακέτα από πληροφορίες που στέλνουν. Έτσι, αυτόματα επισημαίνει τα δεδομένα από ευαίσθητες πηγές απορρήτου και μεταβατικά προσθέτει ετικέτες σε αυτές, καθώς τα δεδομένα μεταδίδονται μέσω μιας εφαρμογής. Όταν τα επισημασμένα δεδομένα μεταδίδονται μέσω του δικτύου ή εγκαταλείπουν με άλλο τρόπο το σύστημα, το TaintDroid καταγράφει τις ετικέτες των δεδομένων, την εφαρμογή που είναι υπεύθυνη για τη μετάδοση των δεδομένων και τον προορισμό των δεδομένων. Αυτή η ανατροφοδότηση σε πραγματικό χρόνο δίνει στους χρήστες και τις υπηρεσίες ασφαλείας μεγαλύτερη εικόνα για το τι κάνουν οι κινητές εφαρμογές. [27]

Ένα παρόμοιο σύστημα είναι και το TaintBochs, όπου παρακολουθεί τα ευαίσθητα δεδομένα με το να τα χρωματίζει σε επίπεδο υλικού. Μέσα σε ένα προσομοιωμένο περιβάλλον κάθε byte της μνήμης του συστήματος, της κατάστασης της συσκευής και της σχετικής κατάστασης του επεξεργαστή επισημαίνεται με μια χρωματισμένη κατάσταση. Τα δεδομένα είναι χρωματισμένα εάν θεωρούνται ευαίσθητα και κάθε χρωματισμός σημειώνεται σε αρχεία καταγραφής. Το TaintBochs αναπαράγει σημαίες χρώσης, όποτε οι αντίστοιχες τιμές τους στο υλικό συμμετέχουν σε μια λειτουργία. Με αυτόν τον τρόπο, τα χρωματισμένα δεδομένα παρακολουθούνται σε όλο το σύστημα καθώς ρέουν μέσα από τα προγράμματα οδήγησης συσκευών πυρήνα, τα γραφικά GUI σε επίπεδο χρήστη, τα προσωρινά buffers. Το TaintBochs ανέλυσε πολλές μεγάλες εφαρμογές πραγματικού κόσμου. Μεταξύ αυτών ήταν τα Mozilla, Apache και Perl, τα οποία χρησιμοποιούνται καθημερινά για τη διεκπεραίωση εκατομμυρίων κωδικών πρόσβασης σε οποιαδήποτε πλατφόρμα ή εφαρμογή. Η έρευνα αποκάλυψε ότι αυτές οι εφαρμογές και τα εξαρτήματα, στα οποία βασίζονται, δεν έπαιρναν πρακτικά μέτρα για να περιορίσουν τη διάρκεια ζωής των ευαίσθητων δεδομένων που χειρίζονται, αφήνοντας τους κωδικούς πρόσβασης και άλλα ευαίσθητα δεδομένα διάσπαρτα σε όλη τη μνήμη του χρήστη και του πυρήνα. Για να παρακολουθηθεί η θέση των ευαίσθητων δεδομένων με το TaintBochs, προστέθηκε μια άλλη μνήμη, σύνολα μητρώων και άλλα υλικά (hardware) που αποτέλεσαν τη μνήμη σκιάς. Η μνήμη σκιάς παρακολουθεί την κατάσταση χρώματος του κάθε byte στο σύστημα. Κάθε λειτουργία που εκτελείται κατά την κατάσταση μηχανής από τον επεξεργαστή ή τις συσκευές προκαλεί την εκτέλεση παράλληλης λειτουργίας στη μνήμη σκιάς, (π.χ. η αντιγραφή μιας λέξης από τον καταχωρητή A στη θέση B αναγκάζει την κατάσταση στον καταχωρητή σκιάς A να αντιγραφεί στη θέση σκιάς B). Επομένως, για να προσδιοριστεί αν ένα byte είναι μολυσμένο, αρκεί να ελεγχθεί στην αντίστοιχη θέση στη μνήμη σκιάς. Αν οποιοδήποτε bit σε ένα byte είναι μολυσμένο, ολόκληρο το byte θεωρείται μολυσμένο. Η μέθοδος αυτή είναι μια συντηρητική προσέγγιση, δηλαδή δεν χάνονται ποτέ τα ίχνη των ευαίσθητων δεδομένων, αν και κάποια δεδομένα μπορεί να χρωματιστούν άσκοπα. [28]

Όπως και σε παρόμοια συστήματα παρακολούθησης ροής πληροφοριών, ένας βασικός περιορισμός του TaintDroid και του TaintBochs είναι ότι μπορούν να αποτύχουν μέσω διαρροών από σιωπηρές ροές. Η χρήση σιωπηρών ροών για την αποφυγή της ανίχνευσης των χρωστικών είναι από μόνη της ένας δείκτης κακόβουλης πρόθεσης και μπορεί εύκολα να ανιχνευθεί μέσω άλλων τεχνικών όπως η αυτοματοποιημένη ανάλυση στατικού κώδικα. Ωστόσο, αποτελούν μια αποδοτική μέθοδο για ανίχνευση της ροής των δεδομένων, καθώς και για εμπιστευτικότητα αυτών. [27, 28]

2.9) Ιχνηλάτηση των ιχνηλατών - Web Privacy Measurement

Πολλοί ερευνητές επιδιώκουν να εντοπίσουν τις υπηρεσίες ιχνηλάτησης ή αλλιώς να μετρήσουν την ιδιωτικότητα ιστού, χρησιμοποιώντας διάφορα διαδικτυακά εργαλεία. Αυτά τα εργαλεία ονομάζονται Web Privacy Measurements (μετρήσεις απορρήτου ιστού). Μερικά τέτοια ανοιχτού κώδικα λογισμικά, που χρησιμοποιούνται για τέτοιες συνθετικές ανιχνεύσεις είναι τα OpenWPM, FPDetective, Chameleon Crawler, FourthParty, WebXray, TrackingObserver, Common Crawl4. Βέβαια, για πιο οξυδερκή μελέτη μπορούν να χρησιμοποιηθούν σύνολα δεδομένων από εφαρμογές προστασίας της ιδιωτικότητας (π.χ. Ghostery), που συλλέγονται από πραγματικούς χρήστες. Αφού οι ερευνητές διαλέξουν το κατάλληλο εργαλείο, στη συνέχεια επιλέγουν ένα σύνολο ιστοτόπων. Συνηθίζεται να επιλέγονται μερικές εκατοντάδες από τους δημοφιλέστερους ιστοτόπους στο www.alexacom/topsites. Μετέπειτα, το Web Privacy Measurement ελέγχει ποιοι είναι οι κύριοι ιχνηλάτες στο επιλεγμένο σύνολο ιστοτόπων. Βέβαια, κάποια εξεζητημένα Web Privacy Measurements μπορούν να αποκαλύπτουν και πώς τα προσωπικά δεδομένα χρησιμοποιούνται. Μπορούν ακόμα να αλληλεπιδρούν με τις διεπαφές COM από το πρόγραμμα περιήγησης ή με το IConnectionPoint για επιπρόσθετη συλλογή δεδομένων και κατ' επέκταση επιπλέον συμπεράσματα περί των ιχνηλατών. Δεν υπάρχει συγκεκριμένη μέθοδος για τη λειτουργία ή την αρχιτεκτονική τους, αλλά σημειώνεται παρακάτω η λειτουργία και η αρχιτεκτονική για ένα από αυτά τα εργαλεία, το OpenWPM. [29, 30]

Το Open-WPM είναι ένα εργαλείο Web Privacy Measurement κατασκευασμένο με βιβλιοθήκες της Python. Διαχωρίζεται σε τρεις βασικές υποδομές. Η μεγαλύτερη ανάλυση γίνεται στην τρίτη υποδομή, καθώς το τεχνικό κομμάτι των άλλων δυο μεθόδων δεν σχετίζεται με τεχνολογίες καταγραφής.

Διαχειριστές προγράμματος περιήγησης (browser managers)

Οι διαχειριστές των προγραμμάτων περιήγησης λειτουργούν ως επίπεδο αφάιρεσης (abstraction layer) για την αυτοματοποίηση μεμονωμένων περιστατικών προγράμματος περιήγησης. Ο διαχειριστής προγράμματος είναι υπεύθυνος για τη μετατροπή εντολών πλατφόρμας υψηλού επιπέδου (π.χ. επισκέψεις σε έναν ιστότοπο) σε συγκεκριμένες υπορουτίνες. Αυτό γίνεται σύμφωνα με μια καθορισμένη διαμόρφωση των προτιμήσεων χρήστη (π.χ. ενεργοποίηση του Do Not Track).

Διαχειριστής εργασιών (task manager)

Ο διαχειριστής εργασιών επιτρέπει στον ερευνητή να διανέμει εντολές στους διαχειριστές του προγράμματος περιήγησης, να τους παρακολουθεί και να επανακινεί όσους έχουν αποτύχει να ολοκληρώσουν μια εντολή μέσα σε ένα χρονικό όριο. Αυτό επιτυγχάνεται με τη δημιουργία ενός νήματος εκτέλεσης εντολών ανά πρόγραμμα περιήγησης, κάθε φορά που ο ερευνητής εκτελεί μια εντολή σε αυτό το συγκεκριμένο πρόγραμμα περιήγησης. Αυτό το νήμα έχει συγκεκριμένη διάρκεια ζωής και έχει ρυθμιστεί να παρακολουθεί την κατάσταση της αντίστοιχης διαδικασίας διαχειριστή του προγράμματος περιήγησης.

Συσσωρευτής δεδομένων (data aggregator),

Ο συσσωρευτής δεδομένων λειτουργεί ως επίπεδο αφαίρεσης για τον εξοπλισμό του προγράμματος περιήγησης. Με αυτόν τον τρόπο δημιουργήθηκε και μια γλώσσα υψηλού επιπέδου για την έκδοση εντολών ανίχνευσης, οι οποίες μπορούν να επεκταθούν, καθώς οι ερευνητές προσθέτουν προσαρμοσμένες εντολές. Ο συσσωρευτής δεδομένων λαμβάνει δεδομένα, όπως τα λαμβάνει ένα πρόγραμμα web crawler. Στην πιο απλή εκδοχή, το web crawler, αφού κάνει αίτηση για λήψη ενός συνδέσμου (π.χ. με την εντολή `requests.get(url)` με τη βιβλιοθήκη `requests` της Python), αντιγράφει όλο το HTML κείμενο μιας ιστοσελίδας σε μια πρόχειρη μεταβλητή-πίνακα. Έπειτα ψάχνει σε αυτήν συγκεκριμένες ετικέτες (λ.χ. για το τρίτο μέρος ιχνηλάτησης `<meta name="referrer">`) και όταν τις βρει, αποθηκεύει τα περιεχόμενά τους σε άλλες μεταβλητές ή αρχεία. Αυτή η διαδικασία μπορεί να γίνει για πολλές ιστοσελίδες ενός ιστοτόπου ή και για πολλούς ιστοτόπους. Ο συσσωρευτής δεδομένων καταγράφει όλες τις αντλούμενες πληροφορίες σε μια κεντρική βάση δεδομένων SQLite, η οποία επιτρέπει στους ερευνητές να μοιράζονται δεδομένα, πιο εύκολα από ό,τι με τις βάσεις δεδομένων διακομιστή-πελάτη. Η πρόσβαση στο συσσωρευτή δεδομένων γίνεται μέσω μιας απομονωμένης διασύνδεσης υποδοχής (socket interface), η οποία μπορεί εύκολα να συνδεθεί με οποιοδήποτε αριθμό διαχειριστών του προγράμματος περιήγησης ή από διαδικασίες οργάνων. [30]

Εκτός από τη χρήση εργαλείων, υπάρχουν και αλγόριθμοι μη επιβλεπόμενης μηχανικής μάθησης για την ανίχνευση ιχνηλατών. Ένας από αυτούς, λαμβάνει υπόψη μια συλλογή αρχείων καταγραφής (log files) HS. Αυτά εμπεριέχουν τις συναλλαγές HTTP που παράγονται από ένα γνωστό σύνολο χρηστών και έναν στοχευμένο τομέα ιστοτόπου W. Ο αλγόριθμος εξάγει από τα HS όλα τα ζεύγη κλειδιών-τιμής HTTP που περιέχονται σε κάθε αίτημα HTTP. Αυτά τα αιτήματα κατευθύνονται από το σύνολο W ή αναφέρονται σε αυτό. Φερ' ειπείν, μπορούν να έχουν αποθηκευμένο το W, είτε μέσα στο πεδίο Host (άρα το αίτημα είναι πρώτου μέρους), είτε στο πεδίο Referer (τρίτου μέρους αίτημα). Στη συνέχεια, για κάθε κλειδί ο αλγόριθμος διερευνά τις μοναδικά όμοιες τιμές μεταξύ των αναγνωριστικών των χρηστών, που δημιουργούνται από τα αιτήματα (π.χ. του προφίλ του προγράμματος περιήγησης), και των τιμών που περιέχονται στα κλειδιά. Διαισθητικά, ο αλγόριθμος ψάχνει για οποιοδήποτε κλειδί, του οποίου οι τιμές είναι μοναδικά συνδεδεμένες σε συγκεκριμένους χρήστες. Δηλαδή, ψάχνει για κλειδιά με τιμές που είναι διαφορετικές για κάθε διαφορετικό χρήστη, αλλά ίδιες για τον ίδιο χρήστη. Επιπλέον, ο αλγόριθμος εξάγει το σύνολο των ζευγών υπηρεσιών-κλειδιών, τα οποία έχουν τιμή

μεγαλύτερη ή ίση με τη μεταβλητή minUsers. Η minUsers μεταβλητή ορίζει τον ελάχιστο αριθμό μοναδικών ζευγών χρηστών που ο αλγόριθμος πρέπει να τηρεί, προκειμένου να επισημάνει ένα κλειδί ως αναγνωριστικό χρήστη. Με αυτή τη μέθοδο, ελέγχεται κατά πόσο ποικίλλει ο αριθμός των επιστρεφόμενων κλειδιών, που ταξινομεί ο αλγόριθμος, με την αύξηση και με τη μείωση των τιμών minUsers. [29]

2.10) Άλλες μέθοδοι ιχνηλάτησης

Για να μεταφέρει το πρόγραμμα περιήγησης σε έναν ιχνηλάτη πληροφορίες, υπάρχουν πολλοί τρόποι. Ένας σιωπηρός τρόπος είναι μέσω της κεφαλίδας HTTP και ένας ρητός τρόπος είναι μέσω Javascript, μεταδίδοντας απευθείας τα αποτελέσματα μιας document.referrer API κλήσης. Μέσω της κεφαλίδας HTTP τα τρίτα μέρη (δηλαδή οι ενσωματωμένοι ιχνηλάτες σε μια ιστοσελίδα και όχι οι ιχνηλάτες της ίδιας της ιστοσελίδας) μπορούν να αποκτήσουν τα ιστορικά περιήγησης των χρηστών, διευθύνσεις ηλεκτρονικού ταχυδρομείου και άλλες πληροφορίες. Αρκεί να χρησιμοποιηθούν cookies, LBOs ή άλλες τεχνολογίες παρακολούθησης, που τους επιτρέπουν να αναγνωρίζουν μοναδικά τους χρήστες και την κεφαλίδα HTTP referer, που αναφέρει στο τρίτο μέρος ποιο είναι το site (πρώτο μέρος) το οποίο επισκέπτεται ο χρήστης. [5, 11]

Αναγνωριστικά συνεδρίας σε κρυφά πεδία

Πριν τα cookies γίνουν διαθέσιμα το 1994, ο μόνος τρόπος για την παρακολούθηση ενός χρήστη (εκτός από τη χρήση της διεύθυνσης IP) ήταν μέσω αναγνωριστικών συνεδρίας αποθηκευμένα σε κρυφά πεδία. Πιο συγκεκριμένα, δύναται να μεταβιβαστεί ένα αναγνωριστικό για κάποιο χρήστη σε μια άλλη ιστοσελίδα μέσω της διεύθυνσης URL (με τη μέθοδο GET) ή μέσω ενός κρυφού πεδίου μιας φόρμας ιστού (με τη μέθοδο POST). Οι τιμές που υποβάλλονται από μια φόρμα ιστού, χρησιμοποιώντας τη μέθοδο POST, αντίθετα με τις τιμές που επισυνάπτονται στη διεύθυνση URL, χρησιμοποιώντας τη μέθοδο GET, ούτε επιμένουν στο δίσκο ούτε διατηρούνται στο ιστορικό του προγράμματος περιήγησης. Παρόλο που η τεχνική αυτή εξακολουθεί να λειτουργεί, η χρησιμότητά της περιορίζεται σε μία μόνο περιήγηση. Και αυτό, επειδή οι πιο συχνά χρησιμοποιούμενοι μηχανισμοί ιχνηλάτησης εξαρτώνται από τη ρητή αποθήκευση δεδομένων στους υπολογιστές των χρηστών. [11]

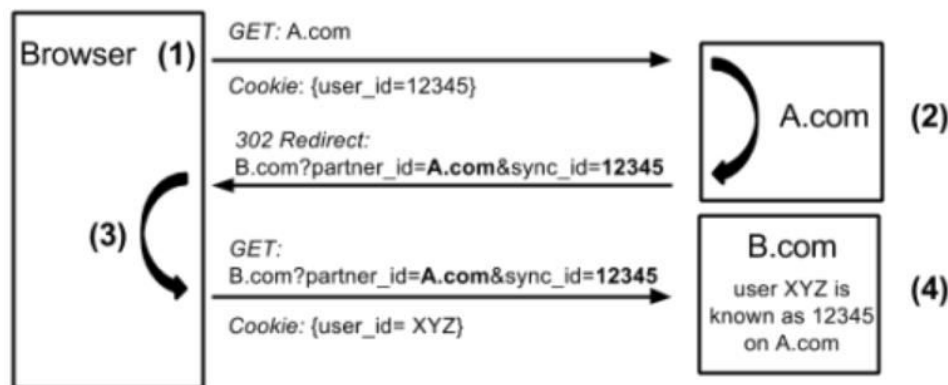
Αυθεντικοποίηση φόρμας ιστού (web-form authentication)

Μια άλλη δυνατότητα εντοπισμού ενός χρήστη είναι η αυθεντικοποίηση φόρμας ιστού (web-form authentication). Αφού ο χρήστης ζητήσει ή του ζητηθεί να εγγραφεί στον ιστότοπο (π.χ. facebook ή gmail), στη συνέχεια οι πόροι που παρέχονται από τον ιστότοπο είναι διαθέσιμοι μόνο για τον χρήστη που πραγματοποιεί είσοδο. Αυτό καθιστά την ταυτότητα χρήστη πολύ εύκολη και ακριβή. Με αυτή τη μέθοδο μπορούν να καταγραφούν όλα όσα κάνει ένας χρήστης, που έχει συνδεθεί, στον ιστότοπο. Αυτή η μέθοδος είναι

ανεξάρτητη από το πρόγραμμα περιήγησης ιστού, το λειτουργικό σύστημα ή τον υπολογιστή που χρησιμοποιείται, καθώς και από τον τόπο, όπου ο χρήστης είναι συνδεδεμένος στο Internet. Υπάρχουν, μολαταύτα, δύο σημαντικές προϋποθέσεις που συνδέονται με αυτή τη μέθοδο. Αρχικά, ο χρήστης πρέπει να ελέγχει την ταυτότητά του κάθε φορά που χρησιμοποιεί τον ιστότοπο με όνομα χρήστη (username) και κωδικό πρόσβασης (password), εκτός και αν χρησιμοποιεί cookies για την αποθήκευσή τους. Δεύτερον, ο έλεγχος ταυτότητας είναι έγκυρος μόνο μέσα από μια χρονική περίοδο (session) μετά την είσοδο του χρήστη. Αυτό γίνεται, ώστε ο εξυπηρετητής να μη σπαταλάει πόρους μνήμης και επεξεργαστικής ισχύος, περιμένοντας ένα χρήστη να συνδεθεί. [11]

Συγχρονισμός cookies (Cookies synchronization)

Μια άλλη σημαντική μέθοδος είναι ο συγχρονισμός cookies (Cookies synchronization ή cookies syncing). Με αυτήν την πρακτική οι τομείς (domain) ενός ιστοτόπου μεταδίδουν μεταξύ τους τα αναγνωριστικά ή άλλα στοιχεία από cookies του χρήστη. Για παράδειγμα, ο τομέας A θα μπορούσε να μεταβιβάσει ένα αναγνωριστικό στον τομέα B, κάνοντας ένα αίτημα σε μια διεύθυνση URL που φιλοξενείται από τον τομέα B. Το αίτημα αυτό εμπεριέχει ως παράμετρο συμβολοσειράς (String Parameter) το αναγνωριστικό του χρήστη. Κατ' αυτό τον τρόπο επιτυγχάνεται το cross-tracking, δηλαδή η διασταύρωση των πληροφοριών χρήστη μεταξύ ιστοτόπων. Ο συγχρονισμός των cookies αποτελεί ένα αποτελεσματικό μέσο για την κοινή χρήση των τιμών των cookies και ευνοεί το σχηματισμό προφίλ και την παρακολούθηση συμπεριφοράς ενός χρήστη. [9, 11, 12]



Εικόνα 7: Παράδειγμα δύο τομέων που συγχρονίζουν τα cookies τους.

Τοπικά κοινόχρηστα αντικείμενα (Local Shared Objects - LSO)

Τα τοπικά κοινόχρηστα αντικείμενα (Local Shared Objects - LSO), όπως τα Flash cookies, τα Java JNLP, τα Google Gears και το Silverlight Isolated Storage, χρησιμοποιούνται σε συνεργασία με κάποιο λογισμικό (π.χ. Adobe Flash, Java) για την αποθήκευση δεδομένων στους υπολογιστές των χρηστών. Σε σύγκριση με τους άλλους αποθηκευτικούς χώρους ιχνηλάτησης (π.χ. HTTP cookies) μπορούν να αποθηκεύουν περισσότερα δεδομένα και είναι πιο δύσκολο για τον χρήστη να τα διαγράψει. Επίσης, είναι προσβάσιμα από όλα

τα προγράμματα περιήγησης που είναι εγκατεστημένα στο σύστημα, καθώς όλα τα plug-ins του συνεργαζόμενου λογισμικού (π.χ. Adobe Flash, Java) στο πρόγραμμα περιήγησης μοιράζονται τον ίδιο κατάλογο αποθήκευσης. Αυτό αποτελεί παρόμοια ή και δυνατότερη μέθοδο παρακολούθησης από το συγχρονισμό cookies, αφού επιτρέπει την ιχνηλάτηση χρηστών σε διαφορετικά προγράμματα περιήγησης. Η καταγραφή των δραστηριοτήτων του χρήστη παράγεται μέσω της άντλησης δεδομένων από απλά cookies. Η άντληση αυτή γίνεται με API του συνεργαζόμενου λογισμικού. Με άλλα λόγια, το API αντλεί πληροφορίες από τα απλά HTTP Cookies (τη στιγμή που δημιουργούνται) και με τις αποθηκεύει στα LSO. Έπειτα, αυτές οι πληροφορίες αλληλομοιράζονται μεταξύ των LSO, του λογισμικού και του τομέα (domain) της ίδιας παραγωγού εταιρίας. Αυτό τα καθιστά και ως απομακρυσμένα κοινόχρηστα αντικείμενα (Remote Shared Objects - RSO). Τα συνεργαζόμενα RSO αντικείμενα από τον ίδιο τομέα μπορούν να έχουν πρόσβαση στα περιεχόμενα των τοπικά επιτηρούμενων αντικειμένων, που είναι αποθηκευμένα σε έναν κατάλογο δίσκων (οι flash αποθηκεύονται ως .sor αρχεία). Τέλος, το αντικείμενο LocalConnection, όταν υποστηρίζεται από το συνεργαζόμενο λογισμικό, μπορεί να χρησιμοποιηθεί για την επικοινωνία μεταξύ διαφορετικών αρχείων (π.χ. SWF), που εκτελούνται στον ίδιο υπολογιστή ταυτόχρονα. Ως εκ τούτου, μπορεί να χρησιμοποιηθεί για την επικοινωνία και την ανταλλαγή τιμών μεταξύ των LSO, είτε εκτελούνται στα κανονικά είτε και στα ιδιωτικά παράθυρα περιήγησης. [5, 11]

Υπηρεσίες ανάλυσης ως τρίτο μέρος

Για ιστοτόπους που επιθυμούν να αναλύσουν την επισκεψιμότητά τους, έχουν δημιουργηθεί υπηρεσίες ανάλυσης τρίτου μέρους, όπως το Google Analytics (GA), που συλλέγουν δεδομένα από τον ιστότοπο και πραγματοποιούν ανάλυση σε αυτά. Εκτός αυτού, ελέγχεται ο πηγαίος κώδικας του ιστοτόπου για τη χρήση τεχνολογιών καταγραφής και παρακολούθησης χρηστών. Έπειτα, τοποθετείται Script στον ιστότοπο (συνήθως ως iframe), ώστε να χρησιμοποιούνται επιπρόσθετες τέτοιες τεχνολογίες (π.χ. cookies) από το μηχανισμό ανάλυσης. Έτσι, θα εκτελείται κώδικας από το διακομιστή της υπηρεσίας στην ιστοσελίδα που επισκέπτεται ο χρήστης του διαδικτύου. Πλέον, η υπηρεσία ανάλυσης μπορεί να ορίζει cookies στο πρόγραμμα περιήγησης του χρήστη, που θα περιέχουν ένα μοναδικό αναγνωριστικό. Ακόμα, με χρήση συναρτήσεων που ελέγχουν την κατεύθυνση/πορεία και τα κλικ του κέρσορα/ποντικού μπορεί να αποθηκεύονται συγκεκριμένες πληροφορίες σε κάποια αποθηκευτική δομή. Για παράδειγμα, "όταν πατήσει κλικ ο χρήστης σε αυτό το πλαίσιο του HTML αρχείου, αύξησε τον αριθμό της μεταβλητής Αριθμός_Κλικ κατά ένα, ώστε να γνωστοποιηθεί το πόσες φορές πάτησε κλικ σε αυτό το πλαίσιο". Συνεπώς, η υπηρεσία ανάλυσης θα ιχνηλατεί παράλληλα και περαιτέρω τους μελλοντικούς επισκέπτες ως τρίτος. Στη συνέχεια, θα μεταφέρει τις πληροφορίες, που συλλέχθηκαν, στο διακομιστή της (π.χ. στο google-analytics.com), κάνοντας ρητά αιτήματα που περιλαμβάνουν προσαρμοσμένες παραμέτρους στη διεύθυνση URL. Σε αυτές μπορεί να περιέχονται οι ενσωματωμένοι ιστότοποι, το αναγνωριστικό χρήστη (από το cookie) και πληροφορίες συστήματος (λειτουργικό σύστημα, πρόγραμμα περιήγησης, γεωγραφικές πληροφορίες κ.λπ.). Έπειτα, θα αναλύσει στατιστικώς (ή και με άλλες μεθόδους) τα συλλεχθέντα δεδομένα και θα εξάγει χρήσιμες πληροφορίες προς τον ιστότοπο (λ.χ. όταν μια διαφήμιση είναι στο πάνω μέρος αριστερά, έχει 50% περισσότερη πιθανότητα να

προβληθεί). Τα αναγνωριστικά των cookies, που ορίζονται από το μηχανισμό ανάλυσης σε διαφορετικούς ιστοτόπους, είναι αναγκαίο να είναι ίδια μεταξύ τους ανά χρήστη. Ειδικά ο χρήστης θα συσχετιστεί με ένα διαφορετικό ψευδώνυμο σε δύο ή περισσότερους ιστοτόπους, περιορίζοντας την ικανότητα να δημιουργείται ένα προφίλ περιήγησης μεταξύ ιστοτόπων για αυτόν τον χρήστη. Το Google Analytics, παρόλο που θεωρείται ένας από τους πιο σημαντικούς ιχνηλάτες, δεν έχει τη δυνατότητα δημιουργίας προφίλ περιήγησης μεταξύ ιστοτόπων χρησιμοποιώντας τα μοναδικά αναγνωριστικά στα cookies του. Τέλος, σημειώνεται ότι, επειδή το script τρέχει στο περιβάλλον της ιστοσελίδας (με τη χρήση ετικέτας <script src=...>), το προκύπτον cookie είναι ιδιοκτησία του ιστοτόπου και δεν ανήκει στον ιχνηλάτη. [5]

Κεφάλαιο 3 - Συστήματα αντι-ιχνηλάτησης (anti-tracking)

Σύμφωνα με το προηγούμενο κεφάλαιο, η ηλεκτρονική παρακολούθηση αποτελεί μια σοβαρή πρόκληση για την προστασία της ιδιωτικής ζωής, η οποία έχει γίνει αντικείμενο σημαντικής προσοχής τόσο στον ακαδημαϊκό χώρο όσο και στον βιομηχανικό κλάδο. Υπάρχουν, ωστόσο, εργαλεία ενίσχυσης της ιδιωτικότητας (PETs) καθώς και άλλοι μηχανισμοί, που αποτρέπουν την παρακολούθηση χρηστών και αποτελούν τα συστήματα αντι-παρακολούθησης (anti-tracking systems). Πολλές φορές, η ανίχνευση των συστημάτων παρακολούθησης και καταγραφής είναι ο πυρήνας του συστήματος αντι-παρακολούθησης. Με άλλα λόγια, τα συστήματα αντι-παρακολούθησης χρησιμοποιούν τις μεθόδους των συστημάτων παρακολούθησης. Εντούτοις, επειδή επενδύονται οικονομικώς περισσότερο τα συστήματα ιχνηλάτησης και εξαιτίας των πολλών παραλλαγών, συνδυασμών και διαφορετικών μεθοδολογιών τους, αυτά τα συστήματα καθίστανται ισχυρότερα από ότι τα αντι-παρακολούθησης. Μολαταύτα, οι χρήστες του διαδικτύου χρησιμοποιούν συστήματα αντι-παρακολούθησης και δείχνουν ικανοποιημένοι.

3.1) Μέθοδοι αποτροπής παρακολούθησης από τους περιηγητές ιστού

Οι χρήστες του διαδικτύου πολλές φορές ανησυχούν για την προστασία της ιδιωτικότητάς τους. Τα προγράμματα περιήγησης ιστού, για να γίνουν πιο αρεστά και προτιμητέα, προσφέρουν ρυθμίσεις αντι-παρακολούθησης, ασφαρίζοντας την ιδιωτικότητα των χρηστών τους. Μια συνηθισμένη ρύθμιση είναι η προσωρινή αποθήκευση πληροφοριών σε cookies, ιστορικό περιήγησης, κρυφή μνήμη ή σε οποιοδήποτε άλλο αρχείο ή σε άλλη αποθηκευτική δομή. Μετά τη χρονική περίοδο (session) περιήγησης του χρήστη, όλα τα προαναφερθέντα θα διαγράφονται και θα καθαρίζονται αυτόματα. Ως εκ τούτου, κάθε φορά που ο χρήστης θα ανοίγει το browser και θα επισκέπτεται έναν ιστότοπο, ο ιστότοπος θα τον φιλοξενεί ως νέο επισκέπτη. Φυσικά αυτό γίνεται και χειροκίνητα, δηλαδή ο χρήστης μπορεί ανά πάσα στιγμή να καθαρίσει τα αποθηκευτικά αρχεία και χώρους του browser. Ακόμη προσφέρεται η δυνατότητα για μπλοκάρισμα των πρώτων ή/και τρίτων cookies. Αυτό γίνεται είτε με αιτήματα Do Not Track, είτε με χρήση πολιτικής P3P, είτε με DNS blocking. Υπάρχουν και πολλές άλλες ρυθμίσεις που προσφέρονται από τους browsers. Και μάλιστα μπορούν να οριστούν διαφορετικές προτιμήσεις ανά ζώνη δικτύου, η οποία ορίζεται (αυτόματα από το πρόγραμμα) ή έχει οριστεί (χειροκίνητα από το χρήστη) ως επικίνδυνη, περιορισμένη, ασφαλής ή άλλη κατηγορία. [26, 31]

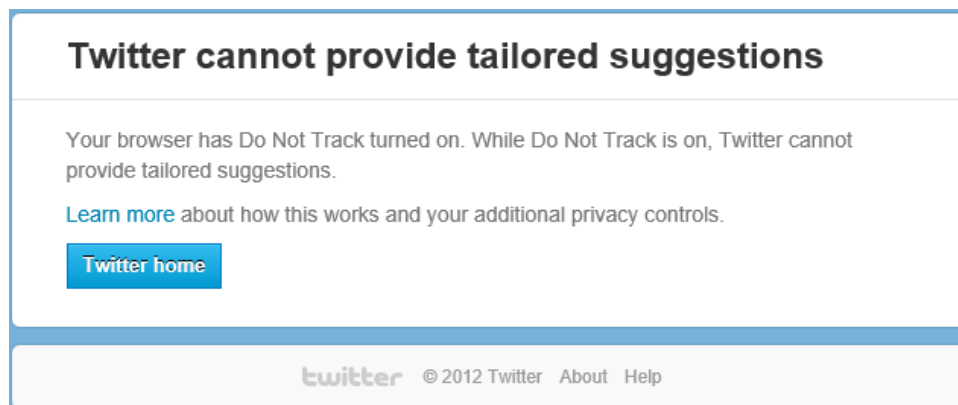
Άλλη μέθοδος αποτροπής ιχνηλάτησης είναι το Do Not Track. Το Do Not Track είναι ένα αίτημα, το οποίο προστίθεται στην κεφαλίδα του HTTP πρωτοκόλλου, από το πρόγραμμα περιήγησης προς τον ιστότοπο προς επίσκεψη. Το πεδίο κεφαλίδας Do Not Track μπορεί να είναι 0, 1 ή null. Αυτές οι τιμές σημαίνουν ότι ο χρήστης συναινεί στην παρακολούθηση, δεν θέλει να παρακολουθείται ή δεν έχει προτίμηση αντίστοιχα. Όταν το πεδίο Do Not Track έχει τιμή 1, θα πρέπει να περιορίζονται οι ιστοσελίδες από την εγκατάσταση cookies ή από άλλους τρόπους παρακολούθησης. Παρόλα αυτά, θα χρειαστεί

και ο ανάλογος κώδικας ή API από τη μεριά του διακομιστή, προκειμένου να αποτραπεί ή να περιοριστεί η ιχνηλάτηση. Το πρόγραμμα περιήγησης δεν ελέγχει, εάν το αίτημα υπολογίζεται από το διακομιστή ή όχι από κάποιο κώδικα ή API. Αυτή είναι η γενική περιγραφή του Do Not Track, καθώς αυτό μπορεί να σχεδιαστεί και να υλοποιηθεί με διάφορους τρόπους. [31, 32]

Η πολιτική P3P (Privacy Preferences Project) δεν είναι ακριβώς εργαλείο anti-tracking. Στην πραγματικότητα είναι ένα εργαλείο, με το οποίο ο χρήστης μπορεί να δηλώνει τα δεδομένα που επιτρέπει να συλλέγονται από ιστοτόπους. Άρα κατ' επέκταση μπορεί να αποτρέψει την εγκατάσταση κάποιων trackers. Αποτελείται από ένα αρχείο XML που ανακτάται και χρησιμοποιείται από τους περιηγητές ιστού και τους διακομιστές ιστοτόπων. Η P3P πολιτική καθορίζει ποιες πληροφορίες θα αποθηκεύονται στα μηχανήματα των χρηστών και για πόσο καιρό, καθώς επίσης και πώς επιτρέπεται να αντλούνται από ένα διακομιστή. Οι χρήστες μπορούν επίσης να ορίσουν μια πολιτική για τον εαυτό τους στο πρόγραμμα περιήγησης τους, γνωστοποιώντας τα δεδομένα που οι χρήστες είναι διαθέσιμοι να παραχωρήσουν σε κάποιο ιστότοπο. Ομοίως και ο ιστότοπος ορίζει τα δεδομένα που επιθυμεί να συλλέξει. Κατ' αυτό τον τρόπο, το πρόγραμμα περιήγησης αιτείται μιας ιστοσελίδας και το αρχείο P3P πολιτικής της από τον διακομιστή. Στη συνέχεια, ο διακομιστής στέλνει σίγουρα το αρχείο P3P πολιτικής της ιστοσελίδας και προαιρετικά την ιστοσελίδα (ή ένα μέρος της) στο πρόγραμμα περιήγησης. Ο browser συγκρίνει την πολιτική προστασίας της ιδιωτικότητας της ιστοσελίδας με τις επιλογές του χρήστη. Εάν ο διακομιστής θέλει να αποθηκεύσει περισσότερες πληροφορίες από ό,τι επιθυμεί και έχει ορίσει ο χρήστης, ο διακομιστής δεν θα μπορέσει να λάβει τις επιθυμητές του πληροφορίες. Έπειτα θα ενημερωθεί καταλλήλως, ότι δε γίνεται να λάβει τις επιθυμητές του πληροφορίες εξαιτίας της πολιτικής P3P του χρήστη. Μετέπειτα, είναι εφικτό, ο διακομιστής του ιστοτόπου να αρνηθεί την αποστολή του αιτούμενου περιεχομένου ή ενός μέρους αυτού. Αυτό μπορεί να γίνει με χρήση κώδικα, ο οποίος ελέγχει τη ροή των αιτημάτων εκ μέρους του διακομιστή και του πελάτη. Εάν συμβεί αυτό, τότε στο χρήστη δε θα προβληθεί η ιστοσελίδα, που ζήτησε να επισκεφθεί, ή κάποιο μέρος αυτής. Ακόμη, ο διακομιστής μπορεί να στείλει κάποιο ενημερωτικό μήνυμα ή ένα άλλο HTML αρχείο στον επισκέπτη του, λέγοντάς του ότι αρνείται να προβάλει την αιτηθείσα ιστοσελίδα, επειδή ο χρήστης αρνείται να παραχωρήσει κάποιες πληροφορίες. Έτσι, ο χρήστης θα χρειαστεί είτε να παραχωρήσει τις πληροφορίες που ζητάει ο διακομιστής, είτε να μην επισκεφθεί τον ιστότοπο αυτόν. Είναι άξιο να σημειωθεί, επίσης, ότι κάποιες υπηρεσίες δε μπορούν να λειτουργήσουν χωρίς την αποθήκευση αυτών των επιπρόσθετων πληροφοριών που ζητούνται από το διακομιστή. Για παράδειγμα, μπορεί ένα μέσο κοινωνικής δικτύωσης να προβάλει δημοσιεύσεις τρίτων ατόμων στο χρήστη, βάσει του ιστορικού αναζήτησής του. Άρα, χωρίς να γνωρίζει το μέσο κοινωνικής δικτύωσης ότι ο χρήστης επισκέφθηκε προηγουμένως αθλητικές ιστοσελίδες, δε γίνεται να του προτείνει δημοσιεύσεις ποδοσφαιριστών ή άλλων αθλητών. [31, 32]

Η DNS blocking είναι μια στρατηγική που χρησιμοποιεί μαύρες λίστες, στις οποίες εμπεριέχονται οι διευθύνσεις των ανεπιθύμητων προς ιχνηλάτηση ιστοτόπων. Αυτό γίνεται, προκειμένου να αποκλειστεί η πρόσβαση σε αυτούς τους ιστοτόπους. Στο πλαίσιο του αποκλεισμού ιχνηλατών (συμπεριλαμβανομένων των διαφημίσεων) οι μαύρες λίστες της DNS διανέμονται συνήθως με τη μορφή ενός αρχείου hosts. Το φιλτράρισμα DNS μπορεί να

χρησιμοποιηθεί για να μπλοκάρει ολόκληρους τομείς ή υπο-τομείς (domains or sub domains), αλλά όχι μεμονωμένους URI. Δηλαδή, δεν μπορεί κανείς να μπλοκάρει το newyorktimes.com/tracker.js ενώ διατηρεί την πρόσβαση στο newyorktimes.com. Ο τρόπος λειτουργίας του είναι σχετικά απλός. Όταν ένας πρώτος ή τρίτος ιστότοπος αιτείται πληροφορίες, αυτός ανιχνεύεται από το HTTP πρωτόκολλο και ελέγχεται, εάν ανήκει στη μαύρη λίστα. Εάν ανήκει, διακόπτεται η επικοινωνία μαζί του και δε λαμβάνει απάντηση στα αιτήματά του. Ειδάλλως συνεχίζεται η επικοινωνία κανονικά. Αυτή η μέθοδος συχνά αποκλείει και άνοιγμα νέων παραθύρων, που εκτός από ιχνηλάτηση μπορεί να έχουν και στόχο να βλάψουν τον υπολογιστή του επισκέπτη. [31, 33]



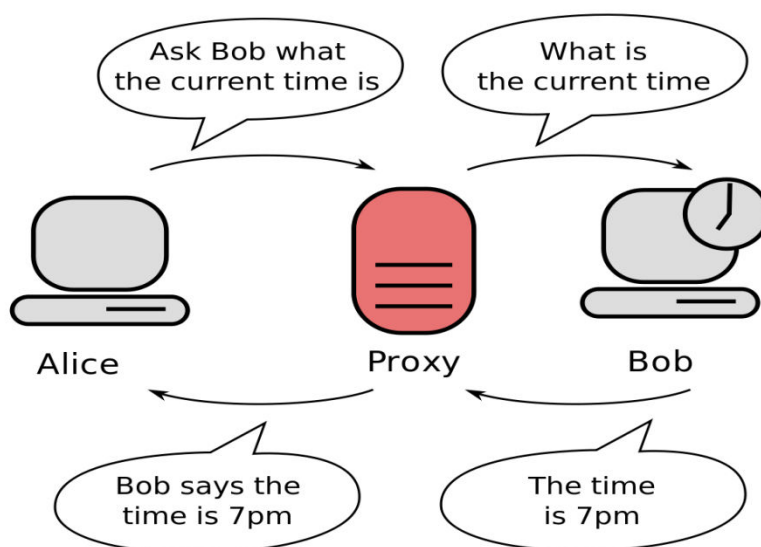
Εικόνα 8: Το 2012 το Twitter δεν πρόσφερε τη δυνατότητα προσαρμοσμένων προτάσεων, σε όσους χρήστες είχαν ενεργοποιημένο το “Do Not Track” αίτημα από το πρόγραμμα περιήγησής τους.

3.2) Μεσολαβητές επικοινωνίας

Ιδιαίτερη άνθιση κατέχουν τα εικονικά ιδιωτικά δίκτυα VPN (Virtual Private Network), που χρησιμοποιούνται για ιδιωτική πρόσβαση στο διαδίκτυο. Βασικά, ένα VPN είναι ένα ιδιωτικό δίκτυο που χρησιμοποιεί ένα δημόσιο δίκτυο (συνήθως στο Διαδίκτυο), για να συνδέσει απομακρυσμένους χρήστες ή ιστοτόπους με άλλους χρήστες ή ιστοτόπους. Αντί να χρησιμοποιεί μια αποκλειστική σύνδεση πραγματικού κόσμου, ένα VPN χρησιμοποιεί εικονικές συνδέσεις. Αυτές δρομολογούνται μέσω του Διαδικτύου από το ιδιωτικό δίκτυο της εταιρείας του στον απομακρυσμένο ιστότοπο ή χρήστη του διαδικτύου. Το VPN μεσολαβεί μεταξύ ενός πελάτη και του εξυπηρετητή, ώστε να μην υπάρχει άμεση σύνδεση μεταξύ τους, αλλά έμμεση. Αυτό γίνεται, διότι το VPN αιτείται και λαμβάνει από τον εξυπηρετητή εκ μέρους του πελάτη. Τοιουτοτρόπως, ο εξυπηρετητής δε γνωρίζει πραγματικά ποιος τον επισκέφθηκε, γιατί δεν υπάρχουν τα ίχνη εξακρίβωσής του (π.χ. διεύθυνση IP του πελάτη), παρά μόνο τα στοιχεία του VPN (π.χ. διεύθυνση IP του VPN). Τα στοιχεία του VPN (ή ίσως και το ίδιο το VPN) δημιουργούνται με τυχαιότητα ή σύμφωνα με την επιλογή του πελάτη. Λόγου χάρη, δύναται ο πελάτης να ορίσει διεύθυνση IP τέτοια, ώστε να φαίνεται ότι η προέλευσή του είναι από τη Γερμανία, ενώ κανονικά δεν είναι. Συνεπώς, εκτός από το ότι διατηρείται η ανωνυμία του επισκέπτη και κατ’ επέκταση

ασφαλιζεται η ιδιωτικότητά του, δυσχεραίνεται και η ιχνηλάτησή του, καθώς αυτή είναι παραποιημένη. Το ίδιο, όμως, δεν ισχύει με το VPN, διότι αυτό γνωρίζει τα στοιχεία του επισκέπτη, για λόγους ασφάλειας (π.χ. αντιμετώπιση τυχόν συμβάντων ηλεκτρονικού εγκλήματος). Τα προσωπικά στοιχεία ενός χρήστη αποθηκεύονται σε χώρο που ορίζει το VPN (π.χ. βάση δεδομένων), ανακτώνται από εξουσιοδοτημένους εξυπηρετητές (servers) και ασφαλιζονται με κατάλληλα μέτρα προστασίας, προκειμένου να προωθούν με ασφάλεια τα αιτήματα του χρήστη στους εξυπηρετητές των ιστοσελίδων που αιτείται να επισκεφθεί. Έτσι, θα παρουσιάζουν το ζητούμενο περιεχόμενο, διαφυλάσσοντας σε κάθε περίπτωση την ανωνυμία του χρήστη, τόσο από τον ιστότοπο όσο και από τρίτες μη εξουσιοδοτημένες προσβάσεις (τρίτες οντότητες). [34, 35]

Proxy ονομάζεται ένας διακομιστής, που ενεργεί ως ενδιάμεσος για τα αιτήματα των πελατών που αναζητούν πόρους από άλλους διακομιστές. Ένας πελάτης συνδέεται με τον Proxy, ζητώντας κάποια υπηρεσία, όπως ένα αρχείο, μια σύνδεση, μια ιστοσελίδα ή άλλο πόρο που διατίθεται από διαφορετικό διακομιστή. Ο Proxy προωθεί και τροποποιεί το αίτημα. Ως εκ τούτου, εκ μέρους του πελάτη αποστέλλει και λαμβάνει τις πληροφορίες του αιτήματός του και των απαντήσεων σε αυτό. Ο πελάτης δεν ιχνηλατείται, καθώς ο Proxy αντικαθιστά τα περιεχόμενα των κεφαλίδων HTTP με ψευδή ή δικά του, αποκρύπτοντας τα περιεχόμενα κεφαλίδων HTTP πελάτη. Ένας δημοφιλής Proxy είναι ο Privoxy. Αυτός έχει τη δυνατότητα, χρησιμοποιώντας DNS ή άλλο είδος φιλτραρίσματος URI, να εντοπίσει τρίτους ιχνηλάτες και να απορρίψει τα αιτήματά τους για άντληση πληροφοριών εκ μέρους του Proxy. Διότι, όσο περισσότερο επιτρέπεται η ιχνηλάτηση ενός Proxy, τόσο περισσότερο αυξάνεται η πιθανότητα να ανιχνευθεί ο πελάτης βάσει της συμπεριφοράς του. Ωστόσο, η πιθανότητα αυτή είναι πάντοτε πολύ μικρή. Όλα αυτά είναι δυνατά και χωρίς εξάρτηση από το πρόγραμμα περιήγησης. Σε σύγκριση με τα VPN, οι Proxies τροποποιούν μόνο την κυκλοφορία (traffic) των δεδομένων κατά την περιήγηση στο διαδίκτυο. Αντιθέτως, τα VPN μπορούν να επηρεάζουν όλες τις κυκλοφορίες (traffic) του δικτύου (π.χ. από online gaming). Κατ' επέκταση τα VPN είναι πιο αποτελεσματικά στην αποτροπή των αποτυπωμάτων (fingerprints), αφού ο σχηματισμός τους μπορεί να γίνεται βάσει πληροφοριών από πολλά διαφορετικά μέρη του υπολογιστή. Τέλος, ο Proxy ενδεχομένως να υστερεί σε θέματα ασφάλειας. [26, 33]



Εικόνα 9: Η επικοινωνία μεταξύ δύο υπολογιστών (γκρι χρώμα) συνδέεται μέσω ενός τρίτου υπολογιστή (κόκκινο χρώμα) που λειτουργεί ως Proxy. Ο Bob ξέρει ότι μιλάει με έναν Proxy, αλλά δε ξέρει σε ποιον θα μεταβιβαστούν τελικά οι πληροφορίες.

3.3) Αποκλεισμός κώδικα script

Απαγορεύοντας τη χρήση Javascript ή οποιουδήποτε άλλου διαδικτυακού script, καταργείται σε μεγάλο βαθμό η λειτουργικότητα ενός ιστοτόπου. Μια λειτουργία που αποτρέπεται είναι και η δυνατότητα ιχνηλάτησης των επισκεπτών του. Λόγου χάρη, οι μέθοδοι καταγραφής δεδομένων που στηρίζονται σε API απαιτούν την εκτέλεση κώδικα στο μηχανήμα του πελάτη. Προκειμένου να καταργηθεί ο κώδικας, αρκεί να εντοπισθούν οι ετικέτες `<src>` στο HTML αρχείο, το οποίο λαμβάνεται από το πρόγραμμα περιήγησης. Στη συνέχεια διαγράφονται οι ετικέτες αυτές μαζί με τα περιεχόμενά τους και φορτώνεται όλο το υπόλοιπο HTML αρχείο. Όπως προαναφέρθηκε, με αυτό τον τρόπο χάνεται μια μεγάλη λειτουργικότητα του ιστοτόπου. Υπάρχουν πολλές εφαρμογές και επεκτάσεις που υλοποιούν την παραπάνω διαδικασία, αλλά με πιο περιορισμένο τρόπο. Ελέγχουν περαιτέρω τα περιεχόμενα των ετικετών `<src>`, επιδιώκοντας να εντοπίσουν τί είδος κώδικας πρόκειται να εκτελεστεί. Αυτό είναι εφικτό λόγω της προέλευσης URL του κώδικα. Διότι, ο κώδικας είναι αποθηκευμένος σε κάποιον εξυπηρετητή. Εάν είναι γνωστό, ποιοι εξυπηρετητές εμπεριέχουν κώδικα που ιχνηλατεί με συγκεκριμένο ή όχι τρόπο, αυτοί μπορούν να αποθηκευτούν σε μια λίστα. Αυτή η λίστα θα ανακτάται σε κάθε έλεγχο και θα συγκρίνονται τα περιεχόμενά της με τα περιεχόμενα της κάθε ετικέτας `<src>`. Εναλλακτικώς, μπορεί να ελέγχεται η ύπαρξη μιας συγκεκριμένης γραμματοσειράς σε μια ετικέτα `<src>`. Κατ' αυτό τον τρόπο, εάν σε μια ετικέτα ανιχνευθεί ο όρος "doubleclick", μπορεί να θεωρηθεί ότι χρησιμοποιείται η υπηρεσία ανάλυσης doubleclick και άρα ότι πρέπει να καταργηθεί ο τρέχων κώδικας. Συνεπώς, δε θα εκτελεστεί ο κώδικας του εξυπηρετητή ιχνηλάτησης (π.χ. doubleclick) και δε θα γίνει ιχνηλάτηση. Αυτά τα προγράμματα πολλές φορές αναπτύσσονται ως επεκτάσεις για φυλλομετρητές, λόγω ευχρηστίας και άμεσης εκτέλεσης. Μπορούν να αποκλείουν όλους τους προς εκτέλεση κώδικες (μερικώς ή εξ ολοκλήρου) ή τους κώδικες που ιχνηλατούν με LSO, με LBS, με υπηρεσία ανάλυσης ή με άλλη συγκεκριμένη και ανιχνεύσιμη μέθοδο ιχνηλάτησης. [26]

Με την παραπάνω μέθοδο μπορεί να αποτραπεί και η δημιουργία αναγνωριστικών ID, περιορίζοντας την ικανότητα των ιστοτόπων για εξακρίβωση οντότητας. Ελέγχοντας τις ετικέτες `<canvas>`, τους συνδέσμους ή την ύπαρξη συγκεκριμένων όρων στο HTML αρχείο που λαμβάνεται, γίνεται να αποκλειστεί κάθε περιεχόμενο που θεωρείται σχετικό με δημιουργία αναγνωριστικού. Αυτό ίσως να μην αναιρεί την ικανότητα του διακομιστή να διακρίνει τον επισκέπτη του, καθώς μπορεί να εκτελεί διαδικασίες στο δικό του διακομιστή. Φερ' ειπείν, εάν ένας πελάτης επισκεφθεί έναν ιστότοπο, ο ιστότοπος λαμβάνει τη διεύθυνση IP του πελάτη. Παρόλο που είναι γνωστό, πως σε μια συγκεκριμένη διεύθυνση IP μπορούν να ανήκουν περισσότερα από ένα άτομα, αυτό αποτελεί μερική διάκριση. Διότι η διεύθυνση IP, που δέχθηκε ο διακομιστής, είναι διαφορετική από άλλες διευθύνσεις IP. Αυτή αποτελεί ένα αναγνωριστικό επισκέπτη, απλά δεν εξακριβώνει εξ ολοκλήρου την ταυτότητά του. [26]

3.4) Αποτροπή fingerprints

Τα αποτυπώματα από ιστοτόπους δημιουργούνται τόσο παθητικά όσο και ενεργητικά. Η παθητική δημιουργία αποτυπωμάτων χρησιμοποιεί εγγενείς ιδιότητες της επικοινωνίας (π.χ. τη σειρά των κεφαλίδων HTTP). Από την άλλη μεριά, τα ενεργητικά αποτυπώματα εκτελούν δέσμες ενεργειών συλλογής χαρακτηριστικών του μηχανήματος του πελάτη (π.χ. πόσο χρόνο διήρκεσε η εκτέλεση του κώδικα στο μηχανήμα του πελάτη). Συνεπώς, χρειάζεται αποκλεισμός του κώδικα ή τροποποίηση των κεφαλίδων HTTP για την αποτροπή δημιουργίας αποτυπωμάτων (π.χ. με Proxy ή καλύτερα με VPN). Καλό είναι αυτές οι δυο μέθοδοι να συνυπάρχουν. Εντούτοις, η τροποποίηση των κεφαλίδων είναι σημαντικό να γίνεται με καινούργιους και περίπλοκους τρόπους. Αυτό συμβαίνει, διότι πολλές φορές είναι εύκολα ανιχνεύσιμες οι ψεύτικες και τροποποιημένες κεφαλίδες HTTP και εύκολα ο ιχνηλάτης ανακτά το αποτύπωμα του χρήστη. Αυτή η ανάκτηση είναι εφικτή, όταν η διαδικασία τροποποίησης είναι γνωστή από τον ιχνηλάτη. Το φαινόμενο αυτό είναι συχνό. Άλλη μέθοδος αποτροπής είναι η εύρεση αποτυπωμάτων στο μηχανήμα του επισκέπτη. Εάν εντοπισθούν αποτυπώματα αποθηκευμένα στον υπολογιστή του χρήστη, η τροποποίησή τους ανά τακτά χρονικά διαστήματα (π.χ. με κάθε νέα εκτέλεση του browser) είναι ένα αποτελεσματικό μέσο κατά της εξακρίβωσης επισκέπτη. Με αυτή τη μέθοδο, όταν ο ιστότοπος θα ζητήσει και θα λάβει το αποτύπωμα από τον υπολογιστή του χρήστη (είτε μέσω browser, είτε από άλλη εφαρμογή), θα το συγκρίνει με τα αποτυπώματα στη βάση δεδομένων του. Είτε δε θα εντοπίσει το συγκεκριμένο χρήστη, είτε θα τον ταυτίσει με άλλο. Στην πρώτη περίπτωση θα δημιουργηθεί ένας νέος χρήστης στη βάση δεδομένων, αλλά την επόμενη φορά που θα συνδεθεί ο επισκέπτης στον ιστότοπο, θα έχει και πάλι νέο αποτύπωμα. Συνεπώς η διαδικασία θα επαναλαμβάνεται ξανά και ξανά, δυσχεραίνοντας τη σωστή εξακρίβωσή του. Το δεύτερο ενδεχόμενο είναι πιο απίθανο, αρκεί το αποτύπωμα να είναι μεγάλου μήκους ή να κρυπτογραφείται ή να κατακερματίζεται με μεγάλου μήκους κλειδί. [25, 26]

3.5) Άλλα συστήματα αποτροπής ιχνηλάτησης

Μια ακόμη μέθοδος για αποφυγή της παρακολούθησης διαδικτύου είναι το Tor (The Onion Routing). Το Tor είναι ένα ειδικό πρόγραμμα περιήγησης στο διαδίκτυο, το οποίο αναπτύσσει ανωνυμία στο IP στρώμα δικτύου με τη βοήθεια μιας σειράς από εικονικούς proxies. Πιο συγκεκριμένα, κατά την επικοινωνία του πελάτη με ένα διακομιστή προστίθενται επανειλημμένα πληροφορίες δρομολόγησης, οι οποίες κρυπτογραφούνται, σε μια σειρά από εικονικούς δρομολογητές (routers). Εξαιτίας των πολλών στρωμάτων κρυπτογραφήσεων, καθίσταται δύσκολη η ανίχνευση της διαδρομής του πακέτου που μεταδίδεται μέσω των δρομολογητών στο δίκτυο. Τούτο συνήθως εκτελείται στο τοπικό μηχανήμα του πελάτη και είναι προσβάσιμο από οποιαδήποτε εφαρμογή TCP με υποστήριξη SOCKS27, όπως προγράμματα περιήγησης ιστού ή πελάτες IRC. Με αυτό τον τρόπο επιτυγχάνεται ανωνυμία, προστασία της ιδιωτικότητας και δύσκολη ανάκτηση της αυθεντικής πληροφορίας κατά την υποκλοπή της. Αυτό, όμως, δεν αναιρεί εξ ολοκλήρου την ιχνηλάτηση από το διακομιστή που φιλοξενεί το χρήστη του διαδικτύου. Γιατί, εάν λόγου χάρη δε γνωρίζει τη διεύθυνση IP του επισκέπτη, δε θα γνωρίζει και πού να στείλει

την αιτούμενη ιστοσελίδα. Βέβαια, σωστά θα μπορούσε να επισημάνει κάποιος, πως το Tor μπορεί να τροποποιεί την πραγματική διεύθυνση IP χάρις τους proxies. Αυτό, όμως, δεν αναιρεί την ιχνηλάτηση από το διακομιστή που φιλοξενεί το χρήστη του διαδικτύου. Και αυτό, επειδή το Tor δεν εφαρμόζει κρυπτογράφηση από άκρη σε άκρη (End to end encryption), δηλαδή στέλνει μη κρυπτογραφημένα δεδομένα στον τελικό προορισμό. Ακόμα και αν αυτά είναι ψευδή, μπορούν να ανακτηθούν τα πραγματικά δεδομένα, αφού οι μεθοδολογίες κρυπτογράφησης του Tor είναι γνωστές. Ακόμα, μπορούν να ανακτηθούν τα πραγματικά δεδομένα συναρτήσει άλλων χαρακτηριστικών, όπως τα fingerprints. Και πράγματι, το Tor δε μπορεί από μόνο του να αποτρέψει τα fingerprints ή άλλα αναγνωριστικά. Ίσως η βασική ιχνηλάτηση από πρώτα μέρη να επιτρέπεται μέσω του Tor. Ίσως να μπορεί να αποκλείει τις τρίτες οντότητες ιχνηλάτησης, εάν αυτές δεν συλλέγουν πληροφορίες του χρήστη μέσω του πρώτου ιχνηλάτη, αλλά με δική τους πρωτοβουλία. Αυτό θα εξαρτηθεί σε μεγάλο βαθμό από την πιθανή συνεργασία του Tor με άλλες μεθόδους κατά της παρακολούθησης, όπως με χρήση του Privacy ή άλλου δημοφιλούς εργαλείου. [12, 26]

Η Deep packet inspection (βαθιά επιθεώρηση πακέτων) είναι ένας τύπος τεχνικής φιλτραρίσματος πακέτων στα δίκτυα υπολογιστών, που ελέγχει τα περιεχόμενα των πακέτων που αποστέλλονται μεταξύ πελάτη και διακομιστή, εξάγοντας τις πληροφορίες τους. Η λεπτομερής επιθεώρηση και το φιλτράρισμα των πακέτων με Deep packet inspection επιτρέπει την προηγμένη διαχείριση δικτύου, την εξυπηρέτηση των καταναλωτών, λειτουργίες ασφαλείας, καθώς και εξόρυξη δεδομένων στο διαδίκτυο. Αντί να αρκείται στην επιθεώρηση της κεφαλίδας HTTP των πακέτων, εξετάζει το τμήμα δεδομένων, ψάχνει για ιούς, ανεπιθύμητα μηνύματα, μη συμμόρφωση με κάποιο πρωτόκολλο, εισβολές ή καθορισμένα κριτήρια. Έπειτα, έχει τη δυνατότητα να αποφασίσει, εάν το πακέτο θα ληφθεί, αν θα παρακαμφθεί ή αν θα δρομολογηθεί σε διαφορετικό στόχο. Ως εκ τούτου, μπορεί να σταματάει την απόπειρα παρακολούθησης ενός ιστοτόπου, όταν κατά την αποστολή πακέτων ανιχνεύονται ιστοσελίδες με κώδικες που ανακατευθύνουν σε ιχνηλάτες, αναγνωριστικά αποτυπώματα, αιτήματα API και λοιπά. Βέβαια, μπορεί να χρησιμοποιηθεί και για ιχνηλάτηση, αλλά η χρήση κρυπτογράφησης μπορεί να την αποτρέψει. [26]

Τα Ad blockers είναι πολύ διαδεδομένες επεκτάσεις που μπλοκάρουν τις διαφημίσεις σε ιστοτόπους. Εκτός από την πιθανή ενόχληση, υπάρχει πράγματι μια παραβίαση της ιδιωτικότητας με τη χρήση διαφημίσεων. Αυτό συμβαίνει, καθώς οι διαφημίσεις συνοδεύονται από κώδικα, ώστε να γνωρίζουν οι υπηρεσίες ανάλυσης την επισκεψιμότητα σε αυτές τις διαφημίσεις. Ακόμα, εάν πατηθεί κλικ σε μια διαφήμιση, αντλείται και η πληροφορία, ότι το διαφημιζόμενο αγαθό (προϊόν ή υπηρεσία) είναι πιθανώς αρεστό στον επισκέπτη, σχηματίζοντας έτσι ένα προφίλ για αυτόν. Τα Ad blockers, όπως και τα script blockers, ανιχνεύουν τις ετικέτες που εμπεριέχουν κώδικα και τις διαγράφουν ή χρησιμοποιούν DNS. Παρόλα αυτά, επειδή η ιχνηλάτηση που επιτυγχάνεται μόνο από διαφημίσεις είναι μικρή, η αποτροπή αυτής από τις επεκτάσεις Ad blockers είναι περιορισμένη. [33]

Tracking Mechanisms \ Defense Mechanisms	Op-Out Cookies	Browser-based blocking / plugins	HTTP DNT Header	Proxy Servers	Tor	Privacy	End-to-end encryption	Private Browsing modes	FP Block	Clearing client-side state.	Blocking Pop ups	Share Me Not Defense	Open WPM	Tracking Observer	Disabling Script Execution
IP tracking				✓	✓										
Active Finger Printing					✓				✓				✓		✓
Passive finger Printing		✓			✓				✓				✓		✓
Canvas Finger Printing									✓				✓		✓
Audio Context Finger Printing															✓
Font Finger Printing		✓							✓				✓		✓
Battery API Finger Printing		✓													✓
Cross Browser Finger printing		✓							✓				✓		✓
Cookie Re-spawning					✓					✓					
Ever Cookies	✓				✓					✓					✓
Super Cookies/ Zombie Cookies		✓								✓					✓
HTTP Cookies	✓		✓			✓		✓		✓			✓		✓
Flash Cookies	✓		✓			✓		✓		✓			✓		✓
First-party Cookies	✓	✓	✓			✓		✓		✓			✓		✓
Third-party Cookies	✓		✓			✓		✓		✓			✓		✓
History Stealing										✓					
System-finger Printing Plugins		✓			✓			✓					✓		
Online Advertising		✓				✓									
Visited link tracking		✓						✓		✓					
Third-Party Analytics		✓										✓		✓	
Third-Party Advertising												✓		✓	
Third-Party Advertising with Pop ups											✓	✓		✓	
Third-Party Advertising Networks												✓		✓	
Cookie Syncing						✓				✓					✓
Deep Packet Inspection							✓								
Panopticklick Finger Printing															✓

Εικόνα 10: Πίνακας του άρθρου [26] με αμυντικές στρατηγικές έναντι μηχανισμών παρακολούθησης.

Κεφάλαιο 4 - Σύστημα ταξινόμησης (classification scheme)

Η ταξινόμηση των συστημάτων παρακολούθησης και καταγραφής αφορά το διαχωρισμό και τη διάκρισή τους βάσει των διαφορετικών τους χαρακτηριστικών. Μέχρι και το 2018 δεν έχει γίνει ιδιαίτερη ενασχόληση με την ταξινόμηση των συστημάτων αυτών. Όσες έχουν πραγματοποιηθεί ακολουθούν μια απλή διαδικασία ταξινόμησης. Αρχικά, εντοπίζονται όλες οι κατηγορίες συστημάτων παρακολούθησης από ένα σύνολο δεδομένων και στη συνέχεια ερευνώνται οι διάφοροι τρόποι παρακολούθησης, που απεικονίζουν τις κλάσεις. Οι τρόποι παρακολούθησης ταυτίζονται με την οπτική γωνία της έρευνας που πραγματοποιήθηκε στο σύνολο δεδομένων. Για παράδειγμα, κάποιες έρευνες επικεντρώνονται στη συμπεριφορά των συστημάτων καταγραφής και παρακολούθησης. Ως εκ τούτου, τα συστήματα αυτά ταξινομούνται σύμφωνα με την παρακολούθηση χρηστών εντός των τοποθεσιών του ιχνηλάτη, με την παρακολούθηση ως τρίτος σε όλους τους ιστοτόπους, με τη διαρροή μοναδικών αναγνωριστικών και με άλλες κλάσεις. Επίσης, άλλη ταξινόμηση είναι εφικτό να γίνει σύμφωνα με τον αντίκτυπο του ιδιωτικού απορρήτου του χρήστη, που δημιουργούν τα συστήματα αυτά. Εκεί μπορεί να γίνει, για παράδειγμα, διαχωρισμός των συστημάτων που καταγράφουν ευαίσθητες και απλές προσωπικές πληροφορίες ή που αποθηκεύουν ατομικά και συλλογικά δεδομένα χρηστών (π.χ. με συγχρονισμό cookies). [5]

Η παρούσα εργασία ερευνά τα συστήματα καταγραφής και παρακολούθησης κυρίως βάσει της τεχνολογικής αρχιτεκτονικής, δηλαδή του τρόπου υλοποίησης και λειτουργίας τους. Ως εκ τούτου, στα πλαίσια της παρούσας εργασίας η ταξινόμηση των συστημάτων αυτών αναπτύχθηκε με βάση την τεχνολογική αρχιτεκτονική τους. Αυτό το σύστημα ταξινόμησης δε στηρίζεται σε κάποια άλλη ερευνητική προσπάθεια (κατ' ελάχιστον από όσες πηγές βρέθηκαν ή χρησιμοποιήθηκαν) και από αυτή την άποψη είναι πρωτότυπη. Ανατρέχοντας στο 2ο κεφάλαιο θα παρατηρηθούν τα 17 εξής συστήματα ιχνηλάτησης: αυθεντικοποίηση φόρμας ιστού, αναγνωριστικά συνεδρίας σε κρυφά πεδία, HTTP Cookies, HTML, Web Beacons, Email Tracking, Flash cookies και local shared objects, Persistent cookies (ή αλλιώς zombie cookies ή evercookies), υπηρεσίες εντοπισμού θέσης LBS, HTML5, Taint tracking, Spyware, Canvas fingerprinting, συγχρονισμός cookies, ETags, υπηρεσίες ανάλυση, Web Privacy Measurements. Οι κύριοι τρόποι λειτουργίας στηρίζονται στη χρήση πρωτοκόλλου HTTP, στη χρήση API και σε συνδυαστικές μεθόδους. Βέβαια, παρατηρείται και μια άλλη σημαντική τεχνολογία που βοηθάει τα συστήματα αυτά. Αυτή είναι η εξακρίβωση ταυτότητας με αναγνωριστικά. Τοιουτοτρόπως, παρέχονται 4 κλάσεις, εκ των οποίων η 1η και η 2η κλάση χωρίζονται σε άλλες 2 υποκατηγορίες. Άρα, τελικά υφίστανται 6 κλάσεις κατηγοριοποίησης.

Κλάση Α - Αλληλεπίδραση με HTTP

Σε αυτήν την κλάση ταξινομούνται οι τεχνολογίες που αρκούνται στη χρήση του HTTP πρωτοκόλλου και για τη συλλογή των δεδομένων τους. Δηλαδή, με επικεφαλίδες παραπομπής (href), με χρήση URL, με αιτήματα HTTP GET αντλούν τα δεδομένα του χρήστη. Παράλληλα είναι αναγκαία η εκτέλεση κώδικα, ώστε να αποθηκευτούν τα δεδομένα σε αρχεία ή σε κάποια βάση δεδομένων. Η διεπαφή του χρήστη με αυτές τις τεχνολογίες γίνεται μέσω HTML και συχνά είναι σιωπηρή. Προκειμένου να αναπτυχθούν, χρειάζεται η συγγραφή μερικού κώδικα προγραμματισμού ή αξιοποίηση έτοιμων συστημάτων. Αξίζει να σημειωθεί ότι οι περισσότερες τεχνολογίες ιχνηλάτησης (αν όχι όλες) στηρίζονται στο HTTP, αφού αποτελεί το βασικότερο μέσο επικοινωνίας μεταξύ διακομιστή και πελάτη. Εντούτοις, στην παρούσα κλάση δεν κατατάσσονται οι τεχνολογίες που χρησιμοποιούν συνδυαστικές μεθόδους για την επίτευξη μεγαλύτερης και πολυπλοκότερης καταγραφής δεδομένων. Κατά συνέπεια στην κλάση Α εμπεριέχονται οι εξής τεχνολογίες:

- α) Αυθεντικοποίηση φόρμας ιστού
- β) Αναγνωριστικά συνεδρίας σε κρυφά πεδία
- γ) HTTP Cookies
- δ) HTML
- ε) Web Beacons
- στ) Email Tracking

Η κλάση Α χωρίζεται σε κλάση Α1 και Α2. Στην Α1 εμπεριέχονται οι τεχνολογίες που χρησιμοποιούν ως αποθηκευτικό χώρο τον υπολογιστή του πελάτη. Αντιθέτως, στην Α2 ανήκουν αυτές που αποθηκεύουν τα δεδομένα απευθείας στο δικό τους εξυπηρετητή. Οπότε, τα α), β), γ) εντάσσονται στην κλάση Α1 και τα δ), ε), στ) στην κλάση Α2.

Κλάση Β - Αλληλεπίδραση με API

Στις τεχνολογίες της κλάσης Β η συλλογή δεδομένων από το χρήστη μπορεί να γίνει με διάφορους τρόπους (π.χ. API, HTTP). Το σημαντικό χαρακτηριστικό αυτής της κλάσης είναι ο τρόπος καταγραφής δεδομένων και η γενικότερη υποστήριξη της μεθόδου ιχνηλάτησης. Η καταγραφή των συλλεγμένων δεδομένων σε αποθηκευτική δομή (βάσεις δεδομένων ή αρχεία) στηρίζεται στην αλληλεπίδραση με διεπαφές προγραμματισμού εφαρμογών (APIs) ή άλλων ειδών διεπαφές (interfaces). Άλλων ειδών διεπαφές είναι τα τμήματα που παρεμβάλλονται μεταξύ ενός αντικειμένου και του περιβάλλοντός του και διαμέσου των οποίων πραγματοποιείται κάθε επικοινωνία του αντικειμένου με το περιβάλλον του. Λόγου χάρη, οι τεχνολογίες αυτές μπορούν να χρησιμοποιούν APIs χαμηλού επιπέδου του λειτουργικού συστήματος (π.χ. τα Layered Service Providers στο API

δικτύου), APIs εφαρμογών (π.χ. της HTML5 ή του Adobe Flash), διεπαφές του προγράμματος περιήγησης (π.χ. IConnectionPoint ή COM (Component Object Model)). Η διεπαφή του χρήστη με αυτές γίνεται μέσω διαδικτυακών πλατφόρμων, εφαρμογών, ιστοσελίδων. Για να υλοποιηθούν αυτές οι μέθοδοι, απαιτείται αρκετή συγγραφή κώδικα προκειμένου να καλείται και να χρησιμοποιείται το API, καθώς και κατανόηση της αρχιτεκτονικής των APIs. Συμπερασματικά, στην κλάση Β εμπεριέχονται οι εξής τεχνολογίες:

- α) Flash cookies και local shared objects
- β) Persistent cookies, zombie cookies, evercookies
- γ) Υπηρεσίες εντοπισμού θέσης LBS
- δ) HTML5
- ε) Taint tracking
- στ) Spyware

Ένας επιπρόσθετος διαχωρισμός θα μπορούσε να γίνει ανάμεσα στις τεχνολογίες που χρησιμοποιούν δικό τους API, δηλαδή που ανήκει στην υπηρεσία ιχνηλάτησης, ως κλάση Β1 και στις τεχνολογίες που χρησιμοποιούν τρίτο API από πρόγραμμα ή λειτουργικό σύστημα ως κλάση Β2. Έτσι, διακρίνονται τα α), β), γ) στην κλάση Β1 και τα δ), ε), στ) στην Κλάση Β2.

Κλάση Γ - Εξακρίβωση ταυτότητας με αναγνωριστικά

Υπάρχουν τεχνολογίες, όπου δε συλλέγουν δεδομένα χρηστών, αλλά βοηθούν πολύ στην καλύτερη καταγραφή και παρακολούθηση. Αυτές είναι οι τεχνολογίες εξακρίβωσης ταυτότητας και κατηγοριοποιούνται στην κλάση Γ. Με τη χρήση κώδικα αναπαράγουν (με τυχαίο τρόπο) μοναδικά αναγνωριστικά για κάθε χρήστη ως επισκέπτη μιας ιστοσελίδας ή για κάθε χρήστη που περιηγείται πολλές ιστοσελίδες. Επιπρόσθετη πολυπλοκότητα επιτυγχάνεται με τη συλλογή πληροφοριών από το μηχάνημα του κάθε χρήστη. Εξάγοντας αναγνωριστικά με βάση αυτά τα στοιχεία, οι χρήστες του διαδικτύου διακρίνονται και παρακολουθούνται πιο αποτελεσματικά. Απαιτείται χρήση κώδικα ή και κλήση APIs για την υλοποίησή τους, αλλά η αποδοτικότητά τους εξαρτάται και από άλλα ερευνητικά πεδία (π.χ. τυχαιότητα αναπαραγωγής αριθμών). Ο χρήστης δεν έχει επαφή με αυτά τα συστήματα. Δεν έχει σημασία, εάν μπορεί ή δε μπορεί ένας χρήστης να βρει το μοναδικό του αναγνωριστικό. Ως εκ τούτου, στην κλάση Γ εμπεριέχονται οι εξής τεχνολογίες:

- α) Canvas fingerprinting
- β) Συγχρονισμός cookies
- γ) Etags

Κλάση Δ - Συνδυαστικές αλληλεπιδράσεις

Οι υπόλοιπες τεχνολογίες ιχνηλάτησης συνδυάζουν, εκτός από το HTTP και ίσως και το API, και άλλες τεχνολογίες διαδικτύου ή άλλες προγραμματιστικές εφαρμογές. Μπορούν να στηρίζονται στη χρήση των τεχνολογιών των προηγούμενων κλάσεων, όπως τα cookies με τα APIs της HTML5. Οι απαιτήσεις για τη δημιουργία αυτών των τεχνολογιών είναι πολλές και συμπεριλαμβάνονται κατ' ελάχιστον όλες οι προαναφερθείσες απαιτήσεις των προηγούμενων κλάσεων. Επιπλέον, σημειώνεται ότι απευθύνονται σε ερευνητικό ή εμπορικό κυρίως πλαίσιο. Τελικά, στην κλάση Δ εμπεριέχονται οι εξής τεχνολογίες:

α) Υπηρεσίες ανάλυσης

β) Web Privacy Measurements

Αυτή η ταξινόμηση βασίζεται εξ ολοκλήρου στην πηγή που προκαλεί τη συλλογή και καταγραφή των δεδομένων του χρήστη. Η τεχνολογική αρχιτεκτονική και λειτουργία ενός συστήματος είναι η παράμετρος που το εντάσσει σε μια κλάση. Έτσι, δεν εξετάζονται οι συμπεριφορές των τεχνολογιών σε υψηλό επίπεδο, όπως για παράδειγμα η συλλογή δεδομένων από έναν ιστότοπο ή από την ανταλλαγή δεδομένων μεταξύ των ιχνηλατών. Αυτά είναι τα πιο εμφανή χαρακτηριστικά, αφού είναι και τα αποτελέσματα της ιχνηλάτησης. Επιπρόσθετα, δε σχετίζεται η τωρινή ταξινόμηση με τη συμπεριφορά του χρήστη, δηλαδή, αν ο χρήστης επισκέπτεται απευθείας ή έμμεσα το διακομιστή του ιχνηλάτη. Ούτε σχετίζεται με τις επιπτώσεις στην ιδιωτική ζωή του χρήστη.

Φυσικά αυτή δεν είναι η μοναδική ταξινόμηση που μπορεί να υλοποιηθεί, ακόμα και με παράμετρο την τεχνολογική αρχιτεκτονική και λειτουργία ενός συστήματος ιχνηλάτησης. Αυτό συμβαίνει, επειδή η κάθε τεχνολογία καταγραφής και παρακολούθησης χρηστών διαδικτύου μπορεί να έχει πολλές και διαφορετικές παραλλαγές. Αυτό σημαίνει ότι η κάθε παραλλαγή τους μπορεί να ανήκει και σε 2η κλάση. Εντούτοις, στην προκείμενη ταξινόμηση λαμβάνεται υπόψη η γενική μορφή τής κάθε τεχνολογίας ιχνηλάτησης, η οποία καθίσταται συγκεκριμένη σύμφωνα με την ανάλυση και περιγραφή που έχει γίνει στο 2ο κεφάλαιο. Για αυτό το λόγο στην προκείμενη ταξινόμηση κάθε τεχνολογία αντιστοιχεί σε μια μόνο κλάση.

	Τεχνολογία Ιχνηλάτησης		
Κλάση A1 Ιχνηλάτηση με HTTP (αποθήκευση στον υπολογιστή του πελάτη)	Αυθεντικοποίηση φόρμας ιστού	Αναγνωριστικά συνεδρίας σε κρυφά πεδία	HTTP Cookies
Κλάση A2 Ιχνηλάτηση με HTTP (αποθήκευση στον υπολογιστή του εξυπηρετητή)	HTML	Web Beacons	Email Tracking
Κλάση B1 Ιχνηλάτηση με APIs της υπηρεσίας ιχνηλάτησης	Flash Cookies και LSO	Persistent Cookies	Υπηρεσίες εντοπισμού θέσης LBS
Κλάση B2 Ιχνηλάτηση με APIs του Λ.Σ. ή του Browser	HTML5	Taint Tracking	Spyware
Κλάση Γ Εξακρίβωση ταυτότητας	Canvas Fingerprinting	Συγχρονισμός Cookies	ETags
Κλάση Δ Συνδυαστική ιχνηλάτηση	Υπηρεσίες ανάλυσης	Web Privacy Measurements	

Εικόνα 11: Πίνακας τελικής ταξινόμησης των συστημάτων ιχνηλάτησης.

Κεφάλαιο 5 - Ενημερότητα χρηστών του διαδικτύου σχετικά με τις τεχνολογίες παρακολούθησης και συλλογής πληροφοριών χρηστών διαδικτύου

5.1) Σχεδίαση και ανάπτυξη ερωτηματολογίου

Υπάρχουν πολλές μέθοδοι για έρευνα και συλλογή απόψεων στον τομέα της πληροφορικής. Μεταξύ αυτών είναι τα ερωτηματολόγια (σε φυσική ή διαδικτυακή μορφή), οι συνεντεύξεις με καταγραφή ήχου ή βίντεο και οι τηλεφωνικές κλήσεις. Όλες αυτές οι μέθοδοι έχουν θετικά και αρνητικά στοιχεία για κάθε διαφορετική έρευνα. Στην παρούσα έρευνα χρησιμοποιούνται μόνο τα ερωτηματολόγια, διότι θεωρούνται πιο κατάλληλα στην προκείμενη περίπτωση. Αυτό συμβαίνει, επειδή η παρουσία του ερευνών στις συνεντεύξεις μπορεί να πιέζει χρονικά τον ερωτώμενο ή και ψυχολογικά με άθελο τρόπο και να δημιουργήσει ασάφεια σύμφωνα με τη γλώσσα του σώματός του. Αντιθέτως, το ερωτηματολόγιο δίνει χρόνο στον ερωτώμενο να διαβάσει τις ερωτήσεις όσες φορές χρειαστεί και να τις απαντήσει όποτε θέλει. Επιπλέον έχει συγκεκριμένες ερωτήσεις καθοδηγώντας τον στους στόχους της έρευνας, του δίνει περιθώριο να επιλέξει και δε χρειάζεται να σκεφτεί με πίεση, αφού έχει ήδη κάποιες απαντήσεις μπροστά του για να επιλέξει. Είναι πιο εύκολο να επιλέγει μια απάντηση παρά να τη δημιουργεί. Η καταγραφή βίντεο και ήχου είναι άβολη από τη μεριά του ερωτώμενου και πολύ χρονοβόρα για συλλογή, μετατροπή σε κατάλληλη μορφή και ανάλυση από τη μεριά του ρεπόρτερ. Ακόμα, η εύρεση τυχαίων τηλεφωνικών αριθμών είναι χρονοβόρα και η συλλογή πληροφοριών από τυχαία άτομα δεν είναι επιθυμητή, αλλά από άτομα που χρησιμοποιούν το διαδίκτυο. Για αυτό και το παρόν ερωτηματολόγιο διεξήχθη σε διαδικτυακή μορφή, ώστε να αποσταλεί σε κατάλληλους ιστοτόπους (π.χ. φοιτητική σελίδα στο facebook) και να συναντήσει το κατάλληλο κοινό. [36, 37, 38]

Τα ερωτηματολόγια χρησιμοποιούνται σε ένα ευρύ φάσμα ρυθμίσεων για τη συλλογή πληροφοριών σχετικά με τις απόψεις και τη συμπεριφορά των ατόμων. Η εγκυρότητα και η αξιοπιστία αυτού του εργαλείου μέτρησης οφείλει να δοκιμαστεί αυστηρά, με απώτερο σκοπό να διασφαλιστεί ότι τα συλλεχθέντα δεδομένα είναι αξιόπιστα και ορθά. Ο σχεδιασμός και ο τρόπος διαχείρισης ενός ερωτηματολογίου θα επηρεάσουν το ποσοστό ανταπόκρισης που επιτυγχάνεται και την ποιότητα των δεδομένων που συλλέγονται. Παρακάτω αναλύονται τα στάδια που ακολουθήθηκαν στην παρούσα εργασία για την υλοποίηση του παρόντος ερωτηματολογίου, υιοθετώντας την προσέγγιση που χρησιμοποιήθηκε από τις πηγές. [36, 37, 38]

1. Καθορισμός της ερευνητικής ερώτησης και του πληθυσμού της μελέτης

Οι ερευνητικές ερωτήσεις έχουν ως στόχο την άντληση πληροφοριών, σχετικά με τη γνώση των χρηστών του διαδικτύου για τις τεχνολογίες παρακολούθησης χρηστών διαδικτύου και συλλογής πληροφοριών. Ο πληθυσμός της μελέτης δεν έχει περιορισμούς σε συγκεκριμένα άτομα, αφού όλοι χρησιμοποιούν το διαδίκτυο, έστω και σε ένα μικρό βαθμό. Έπειτα, ο πληθυσμός κατηγοριοποιείται με βάση το φύλο, την ηλικία, τη συχνότητα

χρήσης του διαδικτύου και κατά κύριο λόγο το μορφωτικό τους επίπεδο στην επιστήμη της Πληροφορικής. Το τελευταίο έχει μεγαλύτερη βαρύτητα, καθώς ένα ερωτηματολόγιο δεν πρέπει να στέλνεται στον οποιοδήποτε, αλλά σε σχετικά άτομα. Ακόμα, εκτιμάται ότι το μορφωτικό επίπεδο θα είναι η κύρια αιτία για την άγνοια ή όχι στις τεχνολογίες καταγραφής και παρακολούθησης. Ωστόσο, η σφαιρική άποψη είναι εξίσου σπουδαία και συνεπώς το ερωτηματολόγιο απευθύνεται και σε άτομα μη σχετικά με την Πληροφορική. Με την ανάλυση των απαντήσεων επιδιώκονται 3 στόχοι:

α) να γίνουν αντιληπτές οι γνώσεις του μέσου χρήστη του διαδικτύου, σχετικά με τις τεχνολογίες παρακολούθησης χρηστών διαδικτύου και συλλογής πληροφοριών

β) να γίνουν αντιληπτές οι παράμετροι, που παίζουν ρόλο στις γνώσεις του ερωτώμενου (φύλο, ηλικία, χρόνος καθημερινής ενασχόλησης στο διαδίκτυο, σχέση με την επιστήμη της Πληροφορικής)

γ) να γίνει αντιληπτό κατά πόσο η άγνοια του ερωτώμενου επηρεάζει την καχυποψία και την επιφυλακτικότητα σε αυτές τις τεχνολογίες

2. Μέθοδος εφαρμογής του ερωτηματολογίου

Το ερωτηματολόγιο δημοσιεύθηκε σε ένα Google Form έγγραφο, επειδή αυτό είναι πολύ οικείο στους χρήστες του διαδικτύου. Ακόμη ένας λόγος είναι, διότι διακρίνεται για την ευχρηστία του, τόσο από τη μεριά του ερευνητή για την υλοποίηση, όσο και από τη μεριά του ερωτώμενου για τη συμπλήρωση. Θα μπορούσε να αποσταλεί και σε email, αλλά αυτή η μέθοδος δεν είναι αποτελεσματική, αφού λίγοι θα είναι πρόθυμοι να το συμπληρώσουν. [36, 37, 38]

3. Διατύπωση των ερωτήσεων

Οι ερωτήσεις χρειάζεται να είναι διατυπωμένες με απλή και κατανοητή γλώσσα, με σαφείς και ευνόητες διατυπώσεις, να είναι σύντομες και (όσο το δυνατό περισσότερο) ενδιαφέρουσες για τον ερωτώμενο. Το μήκος λέξεων κάθε ερώτησης χρειάζεται να είναι κάτω από 20 λέξεις. Υπάρχουν 2 ερωτήσεις που ξεπερνάνε για λίγο το όριο, αλλά κατά μέσον όρο το μήκος των ερωτήσεων του παρόντος ερωτηματολογίου είναι 12. Επίσης, αποφεύγεται το ειδικό και επιστημονικό λεξιλόγιο, η αμφισημία, τα διπλά αρνητικά, φορτωμένες λέξεις, οι υποθετικές ερωτήσεις, οι ερωτήσεις που παραπέμπουν σε μια συγκεκριμένη απάντηση. Εξυπακούεται η ομοιόμορφη δομή του ερωτηματολογίου, η σωστή ορθογραφία, η καλαισθησία, η ευχρηστία. Ένα άλλο πλεονέκτημα του παρόντος είναι ότι επιτυγχάνεται καλύτερος ρυθμός ανταπόκρισης από τους ερωτώμενους, επειδή τα γενικά ερωτήματα προηγούνται από τα πιο ειδικά. Παρακάτω σημειώνονται στα αριστερά οι ερωτήσεις και οι δυνατές απαντήσεις, ενώ στα δεξιά οι λόγοι που επιλέχθηκαν οι ερωτήσεις και οι απαντήσεις. [36, 37, 38]

<u>Ερωτήσεις και δυνατές απαντήσεις</u> <u>ερωτηματολογίου</u>	<u>Λόγοι υποβολής της ερώτησης και της</u> <u>απάντησης</u>
1) Τι φύλο είστε; ☐ Άνδρας. ☐ Γυναίκα. 2) Ποια είναι η ηλικία σας; ☐ 0 έως και 12. ☐ 13 έως και 18. ☐ 19 έως και 24. ☐ 25 έως και 38. ☐ 39 έως και 55. ☐ 56 και άνω. 3) Πόσο χρόνο αφιερώνετε στο διαδίκτυο; ☐ Σπάνια μπαίνω στο διαδίκτυο. ☐ Μερικές φορές την εβδομάδα μπαίνω στο διαδίκτυο. ☐ 0 έως 2 ώρες την ημέρα. ☐ 2 έως και 5 ώρες την ημέρα. ☐ Παραπάνω από 5 ώρες την ημέρα.	Οι ερωτήσεις 1, 2, 3 εξυπηρετούν το β σκοπό. Οι γνώσεις στο παρόν θέμα, αλλά και γενικότερα στον τομέα της Πληροφορικής, εξαρτώνται από το φύλο, την ηλικία και τον χρόνο περιήγησης στο διαδίκτυο.
4) Ποια η σχέση σας με την επιστήμη της Πληροφορικής;	Η ερώτηση 4 εξυπηρετεί το β σκοπό. Είναι αναγκαίο να διακριθούν τα άτομα, που κατά πάσα πιθανότητα έχουν γνώσεις στο συγκεκριμένο αντικείμενο. Εκτιμάται ότι

☐ Δεν έχω σπουδάσει Πληροφορική. ☐ Δεν έχω σπουδάσει Πληροφορική, αλλά το επάγγελμά μου σχετίζεται με αυτήν ή δραστηριοποιούμαι ερασιτεχνικά (π.χ. με προγραμματισμό). ☐ Σπουδάζω Πληροφορική σε ΤΕΙ, ΑΕΙ, ΙΕΚ ή κολλέγιο. ☐ Έχω τίτλο (πτυχίο, μεταπτυχιακό, κλπ.) στην Πληροφορική.	αυτή είναι η πιο σημαντική παράμετρος, που επηρεάζει τις επόμενες ερωτήσεις.
5) Γνωρίζετε ότι οι ιστοσελίδες που επισκέπτεστε στο διαδίκτυο καταγράφουν δεδομένα σχετικά με εσάς; (π.χ. τι ώρα μπήκατε σε μια ιστοσελίδα; μπήκατε από smartphone, υπολογιστή ή tablet;) ☐ Ναι. ☐ Όχι. 6) Γνωρίζετε με ποιους τρόπους γίνεται αυτό; ☐ Ναι, γνωρίζω. ☐ Ξέρω στο περίπου. ☐ Όχι, δε γνωρίζω. 7) Έχετε ακούσει για τα cookies; ☐ Ναι. ☐ Όχι.	Οι ερωτήσεις 5, 6, 7 εξυπηρετούν το β σκοπό. Αυτές οι ερωτήσεις είναι σημαντικές παράμετροι, όχι μόνο για την ορθή απάντηση των επόμενων ερωτήσεων, αλλά και για να αποδειχθεί ποιοι ερωτώμενοι απάντησαν τυχαία. Λόγου χάρη, δε θα ήταν λογικό ένα άτομο να μη γνωρίζει ότι οι ιστότοποι συλλέγουν πληροφορίες, αλλά να γνωρίζει τι είναι τα cookies. Στην ερώτηση 5 χρησιμοποιούνται και 2 παραδείγματα, οπότε γίνεται υπέρβαση των 20 λέξεων. Όμως, αυτό γίνεται για μεγαλύτερη σαφήνεια.
8) Τι νομίζετε ότι είναι τα cookies;	Η ερώτηση 8 εξυπηρετεί τον α στόχο του ερωτηματολογίου. Εκτιμάται ότι λόγω

☐ Είναι αρχεία κειμένου. ☐ Είναι προγράμματα. ☐ Είναι ιοί. ☐ Δε γνωρίζω. ☐ Άλλο. (περιγράψτε παρακαλώ με λίγα λόγια)	άγνοιας θα δοθούν διαφορετικές απαντήσεις για τον ορισμό των cookies. Για αυτό προτείνονται οι πιο πιθανές απόψεις και το πλαίσιο “Άλλο. (περιγράψτε παρακαλώ με λίγα λόγια)”.
9) Πιστεύετε πως είναι επικίνδυνα τα cookies; ☐ Ναι. ☐ Μάλλον. ☐ Όχι, δε νομίζω. ☐ Δε γνωρίζω. 10) Πιστεύετε ότι μπορεί ένας ιστότοπος να συλλέξει πληροφορίες σας, αν απλά δείτε μια εικόνα στο διαδίκτυο; ☐ Ναι. ☐ Μάλλον. ☐ Όχι, δε νομίζω. ☐ Δε γνωρίζω. 11) Πιστεύετε ότι ένας ιστότοπος μπορεί να συλλέξει πληροφορίες για εσάς με απλή HTML (χωρίς cookies); ☐ Ναι. ☐ Μάλλον. ☐ Όχι, δε νομίζω.	Οι ερωτήσεις 9, 10, 11, 12, 13, 14, 15 εξυπηρετούν τον α στόχο. Είναι σύντομες και σαφείς, επειδή χρειάζεται να αποφευχθούν οι ανοιχτές ερωτήσεις και η χρήση κλίμακας. Λόγου χάρη, “Πόσο επικίνδυνα θεωρείτε τα cookies; Βαθμολογήστε με μια κλίμακα από το 0 έως το 10, όπου το 0 σημαίνει εντελώς ακίνδυνα και το 10 σημαίνει εντελώς επικίνδυνα”. Αυτοί οι τρόποι ερώτησης πολλές φορές δεν ταιριάζουν σε θέματα ενημερότητας για τις τεχνολογίες πληροφορικής, διότι δε λειτουργούν με ντετερμινιστικό και συγκεκριμένο τρόπο. Στο τρέχον επιστημονικό πεδίο η χρήση μιας υποκειμενικής άποψης (π.χ. πιστεύω ότι τα cookies είναι λίγο επικίνδυνα) δυσχεραίνει την ανάλυση των απαντήσεων και δεν αποτελεί κατάλληλο μέτρο αξιολόγησης. Μια ακόμη αιτία αποφυγής είναι ο χρόνος σκέψης, που αφιερώνει ο ερωτώμενος προκειμένου να απαντήσει. Πρέπει πάντα να είναι ξεκούραστος για την καλύτερη απόδοσή του στις απαντήσεις, οι οποίες μάλιστα ενδέχεται να μην του είναι οικείες. Η οικειότητα μπορεί να μην αποτελεί πρόβλημα, εάν οι ερωτήσεις είναι ενδιαφέρουσες. Για αυτό ξεκινούν με πιο απλό και γενικό τρόπο. Μολαταύτα, είναι κλειστού τύπου και ενδέχεται να νιώσει ο ερωτώμενος ότι ελέγχονται και εξετάζονται οι γνώσεις του. Αυτό διορθώνεται με τη χρήση της λέξης “πιστεύετε”, με την απάντηση “Δε γνωρίζω”, με την προσθήκη ενός επιφυλακτικού “Μάλλον”, με την αποφυγή του σίγουρου “Όχι” και τη χρήση του “Όχι, δε νομίζω”. Τέλος, δε χρησιμοποιείται απάντηση “Άλλο”, γιατί

☐ Δε γνωρίζω. 12) Πιστεύετε ότι δυο ή περισσότεροι ιστότοποι μπορούν να μοιράζονται τα cookies μεταξύ τους; ☐ Ναι. ☐ Μάλλον. ☐ Όχι, δε νομίζω. ☐ Δε γνωρίζω. 13) Πιστεύετε ότι είναι εύκολο για έναν ιστότοπο να διακρίνει δυο ή περισσότερα άτομα; (π.χ. να ξεχωρίσει το Γιώργο από τη Μαρία) ☐ Ναι. ☐ Μάλλον. ☐ Όχι, δε νομίζω. ☐ Δε γνωρίζω. 14) Πιστεύετε ότι υπάρχουν τρόποι για να μη σας παρακολουθεί ένας ιστότοπος; ☐ Ναι. ☐ Μάλλον. ☐ Όχι, δε νομίζω. ☐ Δε γνωρίζω. 15) Πιστεύετε ότι τα Ad blocker ή η ανώνυμη περιήγηση (Incognito mode) δυσκολεύουν έναν ιστότοπο να σας παρακολουθεί;	βάζει σε σκέψη τον ερωτώμενο και χαρακτηρίζεται από υποκειμενισμό. Η σωστή απάντηση στο ερώτημα 8 είναι “αρχεία κειμένου”, όπως αναφέρεται και στο κεφαλαίο 2. Στο ερώτημα 9 η σωστή απάντηση είναι “όχι”, αφού δε μπορεί ένα αρχείο κειμένου να καταστρέψει αρχεία υπολογιστή. Επίσης, δεν αποτελεί κίνδυνο στην ιδιωτικότητα, εξαιτίας της υποχρεωτικής και ρητής συγκατάθεσης του υποκειμένου για τη χρήση cookies, βάσει του Κανονισμού ΕΕ/2016. Στα ερωτήματα 10, 11, 12, 13, 14, 16 η σωστή απάντηση είναι “ναι”, σύμφωνα με τα προηγούμενα κεφάλαια. Το ερώτημα 15 έχει ως σωστή απάντηση το “μάλλον”, επειδή υπάρχει πολύ μικρή αποτροπή παρακολούθησης από τους ιστοτόπους.
---	---

☐ Ναι. ☐ Μάλλον. ☐ Όχι, δε νομίζω. ☐ Δε γνωρίζω.	
16) Πιστεύετε ότι σας παρακολουθούν, όταν σας στέλνουν ένα email; Αν ναι πώς; (επιλέξτε ένα ή περισσότερα κουτάκια) ☐ Ναι. Χρησιμοποιούν προγράμματα ή κώδικα (π.χ. σε Javascript). ☐ Ναι. Χρησιμοποιούν ιούς. ☐ Ναι. Χρησιμοποιούν εικόνες. ☐ Μάλλον. ☐ Όχι, δε νομίζω πως γίνεται. ☐ Δε γνωρίζω.	Η ερώτηση 16 εξυπηρετεί τον α σκοπό. Επίσης εξυπηρετεί και το γ, εξαιτίας της απάντησης “χρησιμοποιούν ιούς”. Αυτή η ερώτηση μοιάζει και με τις προηγούμενες και με τις επόμενες ερωτήσεις. Είναι μια μετάβαση από τις γενικές ερωτήσεις στις πιο ειδικές. Εξετάζει την άποψη τού ερωτώμενου, βάζοντάς τον σε περαιτέρω σκέψη. Αυτή και οι επόμενες ερωτήσεις, αφού είναι πιο ανοιχτού τύπου, επιτρέπουν να δοθούν περισσότερες από μια απαντήσεις.
17) Ποιες πληροφορίες πιστεύετε ότι ένας ιστότοπος μπορεί να συλλέξει από εσάς; (επιλέξτε ένα ή περισσότερα κουτάκια) ☐ Ημερομηνία και ώρα που μπήκα σε έναν ιστότοπο. ☐ Ιστορικό περιήγησης (οι προηγούμενες επισκέψεις σε άλλους ιστοτόπους). ☐ Τοποθεσία (π.χ. πόλη, διεύθυνση κατοικίας). ☐ Πρόγραμμα περιήγησης (π.χ. Google Chrome, Mozilla Firefox). ☐ Λειτουργικό σύστημα (π.χ. Windows, IOS, Android). ☐ Διεύθυνση IP.	Η ειδική ερώτηση 17 εξυπηρετεί τον α, αλλά και το γ στόχο εξαιτίας της χρήσης των τριών λανθασμένων απαντήσεων “Κωδικούς πρόσβασης”, “Αριθμούς πιστωτικών καρτών” και “Αρχεία από τον υπολογιστή μου (π.χ. φωτογραφίες, βίντεο, μουσική)”. Οι δυο πρώτες πληροφορίες δε γίνεται να αντληθούν από έναν ιστότοπο, επειδή είναι ασφαλισμένες από το πρόγραμμα περιήγησης διαδικτύου, κρυπτογραφημένες και τέλος επειδή είναι κατά του νόμου. Η συλλογή αρχείων απαιτεί τη χρήση Spyware και καθίσταται πλέον ως τεχνολογία ιού που βλάπτει το χρήστη και όχι ως τεχνολογία ιχνηλάτησης.

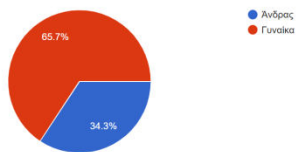
☐ Πληροφορίες από άλλα cookies στον υπολογιστή μου. ☐ Κωδικούς πρόσβασης. ☐ Αριθμούς πιστωτικών καρτών. ☐ Αρχεία από τον υπολογιστή μου (π.χ. φωτογραφίες, βίντεο, μουσική). ☐ Δε γνωρίζω.	
18) Ποια από τα παρακάτω συστήματα παρακολούθησης γνωρίζετε, έστω και λίγο; (επιλέξτε ένα ή περισσότερα κουτάκια) ☐ GPS ή άλλο Local Based Service. ☐ Flash cookies. ☐ Spyware. ☐ Email tracking. ☐ Web beacon (ή αλλιώς Web bug, tracking bug, pixel tag). ☐ HTML5 Tracking. ☐ Taint Tracking. ☐ Δε γνωρίζω κανένα από αυτά.	Η ειδική ερώτηση 18 είναι ίσως η πιο κρίσιμη για τον α σκοπό. Δε χρησιμοποιείται ως απάντηση το “άλλο”, γιατί δε χρειάζεται να σημειωθούν περαιτέρω συστήματα παρακολούθησης και δε χρειάζεται να ελεγχθεί, αν ο ερωτώμενος θα απαντήσει λάθος. Παρόλη τη μεγάλη σημασία της, αυτή η ερώτηση είναι από τις τελευταίες. Και αυτό, επειδή, εάν προηγηθεί από τις άλλες ερωτήσεις, ο ερωτώμενος μπορεί να υποψιαστεί τις επόμενες απαντήσεις.
19) Τέλος, για ποιους λόγους πιστεύετε ότι οι ιστότοποι συλλέγουν πληροφορίες για εσάς; (επιλέξτε ένα ή περισσότερα κουτάκια) ☐ Ίσως να θέλουν να μου κάνουν κακό. ☐ Για λόγους ασφάλειας και για να ξέρουν ότι δεν κάνω κάτι κακό/παράνομο στο διαδίκτυο.	Η ειδική ερώτηση 19 εξυπηρετεί τον α σκοπό. Επίσης, εμπεριέχει την απάντηση “Ίσως να θέλουν να μου κάνουν κακό” που εξυπηρετεί το γ σκοπό. Εκτιμάται ότι το ερωτηματολόγιο δεν είναι κουραστικό, επειδή ξεκινά με γενικές και εύκολες ερωτήσεις (1-5), συνεχίζει με πιο ενδιαφέρουσες ερωτήσεις (6-15) και με ίδιο

☐ Για να βελτιώσουν τις υπηρεσίες τους. ☐ Για να βγάλουν λεφτά, κυρίως μέσω διαφημίσεων. ☐ Για να πουλήσουν τις πληροφορίες μου σε τρίτους και να βγάλουν λεφτά. ☐ Δε γνωρίζω. ☐ Άλλο. (περιγράψτε παρακαλώ με λίγα λόγια)	είδος απαντήσεων (9-15) και τελειώνει με πιο ειδικές (16-19). Όμως, αφού οι αντοχές κάθε ατόμου ποικίλλουν, η τελευταία ερώτηση ξεκινά με τη λέξη “τέλος”. Συνεπώς, ανακουφίζει τον ερωτώμενο και θα βάλει την υπόλοιπη δύναμή του για το ερώτημα.
---	---

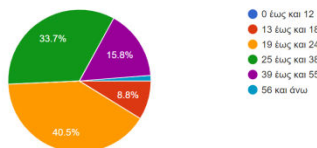
5.2) Εξαγωγή αποτελεσμάτων

Το ερωτηματολόγιο δέχθηκε περισσότερες από 1000 απαντήσεις, οι οποίες λήφθηκαν σε ένα αρχείο μορφής υπολογιστικού φύλλου. Το υπολογιστικό φύλλο δέχθηκε μερική τροποποίηση στο πρόγραμμα Microsoft Excel. Ειδικότερα, οι διάφορες επιλογές “άλλο” αντικαταστάθηκαν με τη λέξη “άλλο”, ώστε να είναι πιο εύκολη η διαχείριση των δεδομένων. Αξίζει να σημειωθεί, ότι πολλές από τις απαντήσεις στην επιλογή “άλλο” ήταν μερικώς σωστές. Στην ερώτηση “τι είναι τα cookies”, για παράδειγμα, μπορεί να είχε απαντηθεί ότι είναι δεδομένα που αποστέλλονται από τον ιστότοπο στον χρήστη ή πληροφορίες σχετικά με τη δραστηριότητα του χρήστη. Καμία, όμως, από τις προαναφερθείσες δεν είναι εντελώς σωστή. Διότι, πράγματι τα cookies εμπεριέχουν πληροφορίες ή δεδομένα του χρήστη, αλλά η απλή αποστολή αυτών από τον ιστότοπο στον υπολογιστή του χρήστη δεν αποτελεί μέθοδο παρακολούθησης. Απαιτείται και η καταγραφή αυτών σε αρχεία, τα οποία έχουν μορφή κειμένου για γρήγορη αποθήκευση, μεταφορά, επεξεργασία, ανάγνωση. Αυτό μπορεί να γίνει πιο κατανοητό με το εξής παράδειγμα. Δε θα είχε νόημα να λέγαμε σε ένα άτομο τον κωδικό πρόσβασής μας, χωρίς να τον καταγράψει σε ένα χαρτί, γιατί θα τον ξεχνούσε. Ομοίως και οι υπολογιστές, χωρίς την καταγραφή δεδομένων δε θυμούνται τίποτε. Κατά συνέπεια, αυτές οι μερικώς σωστές, καθώς και οι λανθασμένες απαντήσεις αντικαταστάθηκαν με τη λέξη “άλλο”. Μετά την τροποποίηση αυτή, το υπολογιστικό φύλλο ανέβηκε στην πλατφόρμα oracle apex, που αποτελεί μια εύχρηστη και ισχυρή βάση δεδομένων. Τοιουτοτρόπως, χρησιμοποιήθηκαν ερωτήματα (queries) σε γλώσσα SQL για τη λήψη στατιστικών δεδομένων.

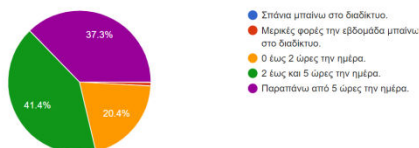
1) Τι φύλο είστε;



2) Ποια είναι η ηλικία σας;



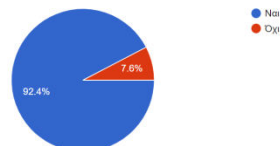
3) Πόσο χρόνο αφιερώνετε στο διαδίκτυο;



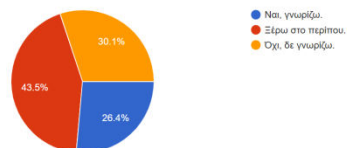
4) Ποια η σχέση σας με την επιστήμη της Πληροφορικής;



5) Γνωρίζετε ότι οι ιστοσελίδες που επισκέπτεστε στο διαδίκτυο καταγράφουν δεδομένα σχετικά με εσάς; (π.χ. τι ώρα μπήκατε σε μια ιστοσελίδα; μπήκατε από smartphone, υπολογιστή ή tablet);



6) Γνωρίζετε με ποιους τρόπους γίνεται αυτό;

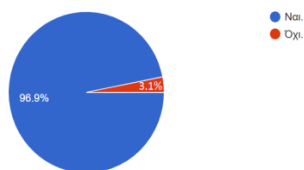


Τα παραπάνω γραφήματα πίτας εξήχθησαν αυτόματα από το google form. Αυτά αναφέρονται σε δημοσκοπικές κυρίως πληροφορίες του ερωτηματολογίου.

Σκοπός α

Ο σκοπός α αποσκοπεί στο να γίνουν αντιληπτές οι γνώσεις του μέσου χρήστη του διαδικτύου, σχετικά με τις τεχνολογίες παρακολούθησης χρηστών διαδικτύου και συλλογής πληροφοριών. Τα παρακάτω γραφήματα πίτας εξήχθησαν αυτόματα από το Google Forms. Σε αυτά φαίνονται οι γνώσεις των ερωτώμενων περί των τεχνολογιών ιχνηλάτησης.

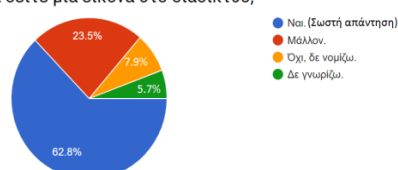
7) Έχετε ακούσει για τα cookies;



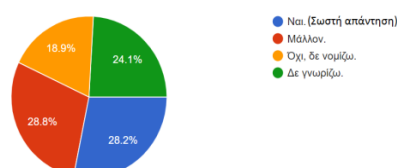
8) Τι νομίζετε ότι είναι τα cookies;



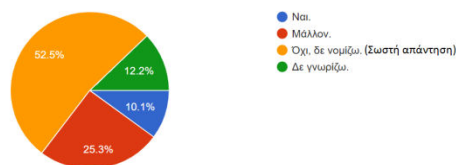
10) Πιστεύετε ότι μπορεί ένας ιστότοπος να συλλέξει πληροφορίες σας, αν απλά δείτε μια εικόνα στο διαδίκτυο;



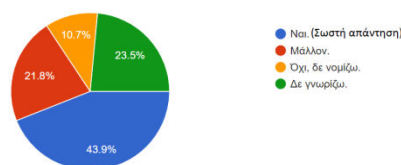
11) Πιστεύετε ότι ένας ιστότοπος μπορεί να συλλέξει πληροφορίες για εσάς με απλή HTML (χωρίς cookies);



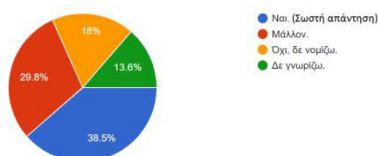
9) Πιστεύετε πως είναι επικίνδυνα τα cookies;



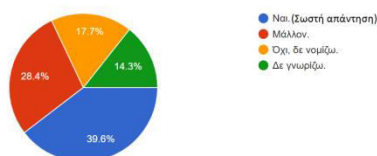
12) Πιστεύετε ότι δυο ή περισσότεροι ιστότοποι μπορούν να μοιράζονται τα cookies μεταξύ τους;



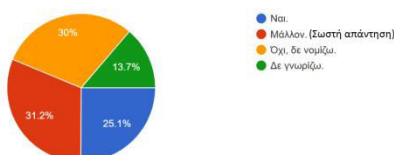
13) Πιστεύετε ότι είναι εύκολο για έναν ιστότοπο να διακρίνει δυο ή περισσότερα άτομα; (π.χ. να ξεχωρίσει το Γιώργο από τη Μαρία)



14) Πιστεύετε ότι υπάρχουν τρόποι για να μη σας παρακολουθεί ένας ιστότοπος;



15) Πιστεύετε ότι τα Ad blocker ή η ανώνυμη περιήγηση (Incognito mode) δυσκολεύουν έναν ιστότοπο να σας παρακολουθεί;

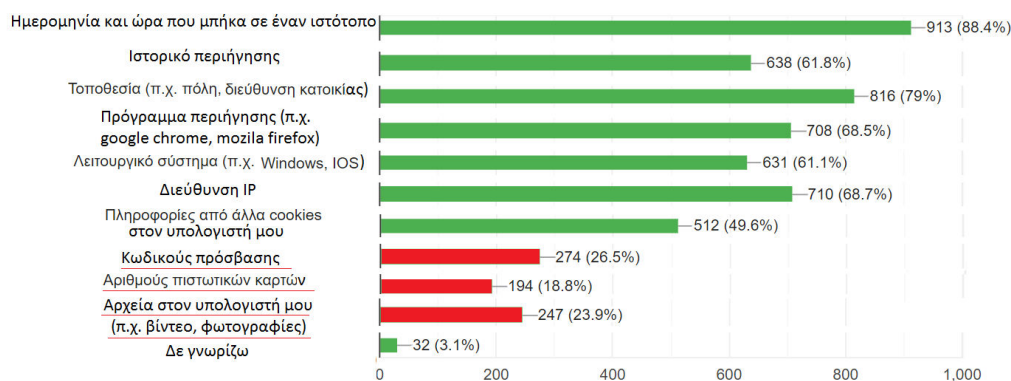


16) Πιστεύετε ότι σας παρακολουθούν, όταν σας στέλνουν ένα email; Αν ναι πώς; (επιλέξτε ένα ή περισσότερα κουτάκια)



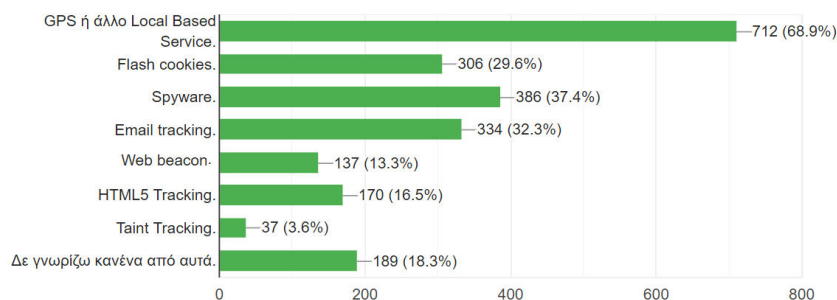
Ο μέσος όρος των σωστών απαντήσεων είναι 37.63%. Αυτό είναι θετικό, καθώς οι πιθανές απαντήσεις είναι τουλάχιστον 4 κάθε φορά (άρα 100% / αριθμός πιθανών απαντήσεων = 25%). Επομένως, επειδή το 37.53% είναι μεγαλύτερο του 25%, αυτό σημαίνει ότι ο αρκετός κόσμος γνωρίζει σε κάποιο βαθμό τις τεχνολογίες ιχνηλάτησης, αλλά όχι ο περισσότερος (αφού 37.53% < 50%). Το ποσοστό που απαντά “δεν γνωρίζω” είναι 17.9%. Ομοίως και εδώ, επειδή 17.9% < 25%, φαίνεται πως οι γνώσεις των ερωτώμενων είναι σε καλό επίπεδο. Στη συνέχεια προβάλλονται και τα διαγράμματα πίτας των τελευταίων ερωτήσεων και παρατηρούνται για ακόμα μια φορά παρόμοια με τα προηγούμενα αποτελέσματα.

17) Ποιες πληροφορίες πιστεύετε ότι ένας ιστότοπος μπορεί να συλλέξει από εσάς; (επιλέξτε ένα ή περισσότερα κουτάκια)

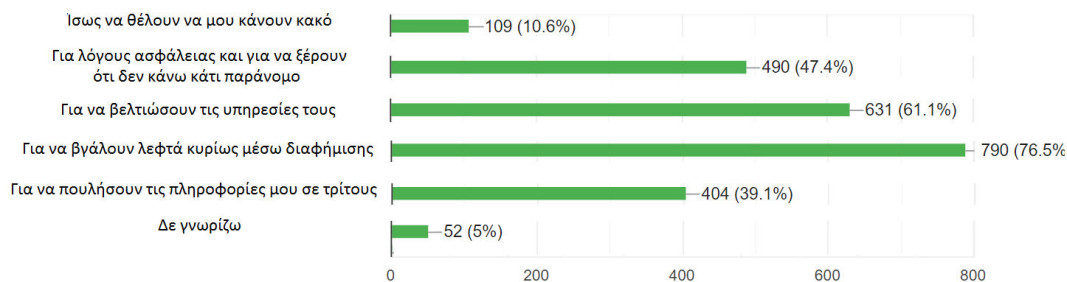


Οι σημειωμένες με κόκκινο πληροφορίες δεν είναι εφικτό να συλλεχθούν από έναν ιστότοπο.

18) Ποια από τα παρακάτω συστήματα παρακολούθησης γνωρίζετε, έστω και λίγο; (επιλέξτε ένα ή περισσότερα κουτάκια)



19) Τέλος, για ποιους λόγους πιστεύετε ότι οι ιστότοποι συλλέγουν πληροφορίες για εσάς; (επιλέξτε ένα ή περισσότερα κουτάκια)



Πολύ λίγα είναι τα άτομα που δε γνωρίζουν να απαντήσουν τις 17 και 19 ερωτήσεις. Οι περισσότεροι χρήστες του διαδικτύου απαντάνε σωστά στην ερώτηση 17, καθώς όλες οι σωστές απαντήσεις (δηλαδή οι πράσινες, εκτός του "δε γνωρίζω") έχουν επιλεγεί με ποσοστό 68.7% (που είναι μεγαλύτερο από 50%). Πολύ μικρή εξαίρεση αποτελεί η απάντηση "Πληροφορίες από cookies στον υπολογιστή μου", που απαντήθηκε με ποσοστό 49.6%. Θεωρείται λοιπόν ότι το 49.6% είναι σχεδόν ίσο με το 50% και έτσι συμπεραίνεται ότι οι περισσότεροι απάντησαν σωστά. Επίσης, οι κοκκινισμένες (και λανθασμένες) απαντήσεις στο ερώτημα 17 έχουν επιλεγεί από τα λιγότερα άτομα με ποσοστό μικρότερο του 27%. Ωστόσο, ο αριθμός των ατόμων δεν είναι ασήμαντος. Αυτό ίσως να αποτελεί αιτία, ώστε τα άτομα που επέλεξαν αυτές τις απαντήσεις να φοβούνται τις τεχνολογίες παρακολούθησης και καταγραφής δεδομένων. Αυτή είναι μια σημαντική πληροφορία/υπόθεση, που θα χρειαστεί στη συνέχεια. Επιπλέον, δε φαίνονται ιδιαίτερα γνωστά τα συστήματα ιχνηλάτησης στους ερωτώμενους, εφόσον το άθροισμα των ποσοστών είναι μόνο 25.2% χωρίς τα cookies, ενώ με τα cookies 33.6%. Εξαίρεση αποτελεί το gps, το οποίο δεν αποτελεί απαραίτητα σύστημα καταγραφής αλλά συνήθως είναι. Ακόμα, τα spyware ταυτίζεται πολλές φορές ως ιός, αλλά αποτελεί και αυτό μέθοδο ιχνηλάτησης. Παρόλο που αυτά είναι τα πιο γνωστά, δεν ήταν αναμενόμενο να ήταν γνωστά κατά μόνο 68.9% και 37.4% αντίστοιχα. Το πιο άγνωστο σύστημα είναι το taint tracking, κάτι το οποίο είναι λογικό. Διότι δεν είναι συχνά χρησιμοποιούμενη τεχνολογία και δε χρησιμοποιείται από ιστοτόπους. Τέλος, στο ερώτημα 19 φαίνεται ότι υπάρχει

μέτρια προς καλή γνώση των αιτίων που οι ιστότοποι συλλέγουν δεδομένα. Αυτό ισχύει, δεδομένου ότι η πρώτη απάντηση σωστά απαντήθηκε από λίγα άτομα, καθώς λίγοι είναι οι ιστότοποι που ιχνηλατούν με μόνο κακή πρόθεση (δηλαδή με μοναδικό σκοπό να βλάψουν). Ακόμα, μόνο 5% δηλώνει ότι δε γνωρίζει και ο μέσος όρος των υπολοίπων απαντήσεων (αμελώντας το "δε γνωρίζω") είναι 46.94% (λίγο μικρότερο από το 50%, δηλαδή τους μισούς ερωτώμενους).

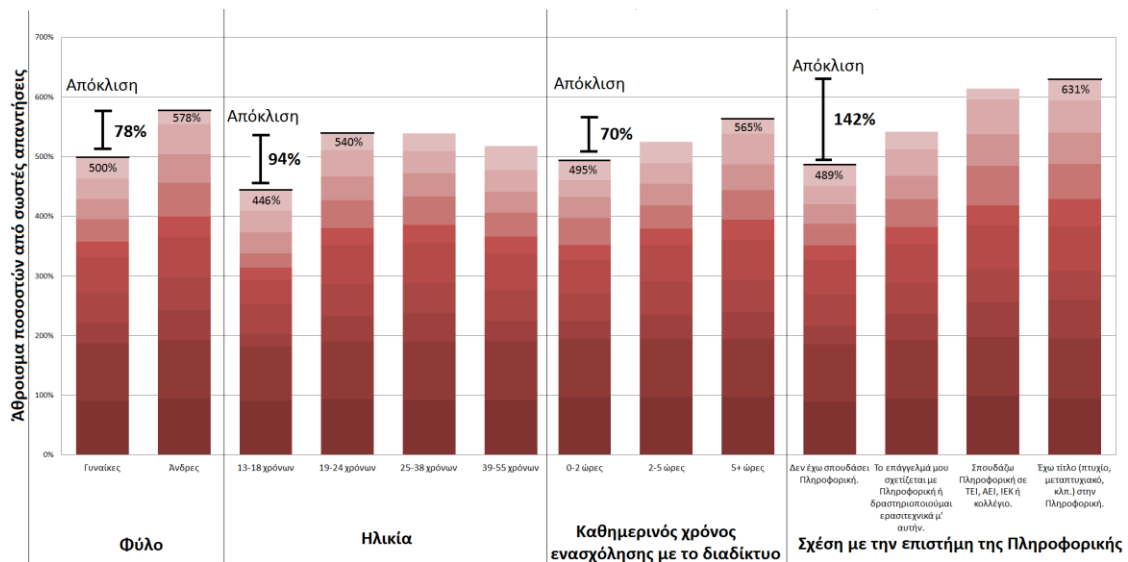
Συμπερασματικά, οι χρήστες του διαδικτύου δεν είναι ανίδεοι με το παρόν θέμα. Υπάρχει γνώση περί των συστημάτων καταγραφής και παρακολούθησης χρηστών, αλλά σε ένα πολύ βασικό επίπεδο.

Σκοπός β

Με το σκοπό β γίνεται αντιληπτός ο βαθμός σημασίας της κάθε παραμέτρου (φύλο, ηλικία, χρόνος καθημερινής ενασχόλησης στο διαδίκτυο, σχέση με την επιστήμη της Πληροφορικής), που επηρεάζει τις γνώσεις του ερωτώμενου. Για να συμβεί αυτό χρειάζεται η χρήση ερωτημάτων σε SQL, έτσι ώστε για κάθε παράμετρο να ληφθεί το ποσοστό των απαντήσεων ανά ερώτηση. Για παράδειγμα, να βρεθεί το ποσοστό της κάθε απάντησης στο ερώτημα "Τι νομίζετε ότι είναι τα cookies;" από τους άνδρες. Με άλλα λόγια να γνωστοποιηθεί ότι 51% των ανδρών απάντησαν ότι τα cookies είναι αρχεία κειμένου, 38% απάντησαν ότι είναι προγράμματα, 0% ότι είναι ιοί, 4% άλλο και 16% δε γνωρίζουν. Το συγκεκριμένο ερώτημα βρίσκεται με τη χρήση του:

```
SELECT COUNT (ID) / 354 FROM SURVEY WHERE COL002='Ανδρας' AND  
COL009='Ναι. Είναι αρχεία κειμένου.'
```

Ο πίνακας, που απεικονίζει το υπολογιστικό φύλλο, ονομάστηκε SURVEY. Η στήλη του πίνακα με όνομα COL002 εμπεριέχει τις απαντήσεις στο πρώτο ερώτημα, δηλαδή στο "Τι φύλο είστε;". Η στήλη του πίνακα με όνομα COL009 εμπεριέχει τις απαντήσεις στο 8ο ερώτημα, δηλαδή στο "Τι νομίζετε ότι είναι τα cookies;". Το ID είναι ένα αναγνωριστικό που διακρίνει τον κάθε ερωτώμενο. Το COUNT είναι μια συνάρτηση που μετράει τον αριθμό των αποτελεσμάτων από το περιεχόμενό του (το ID). Οι άνδρες που απάντησαν το ερωτηματολόγιο είναι συνολικά 354. Άρα με το παραπάνω ερώτημα επιλέγεται το ποσοστό των ανδρών που απάντησαν στο 8ο ερώτημα "Ναι. Είναι αρχεία κειμένου." από τον πίνακα SURVEY. Παρόμοια ερωτήματα οφείλουν να εκτελεστούν στη βάση δεδομένων για κάθε διαφορετική απάντηση σε κάθε διαφορετική ερώτηση για κάθε πιθανή παράμετρο. Αυτά είναι πάνω από 520 ερωτήματα τα οποία υλοποιήθηκαν και από τα οποία καταγράφηκαν οι απαντήσεις τους σε ένα νέο υπολογιστικό φύλλο. Έπειτα, επιλέγοντας τις σωστές απαντήσεις, εξάγεται το παρακάτω γράφημα με το Microsoft Excel (το οποίο επεξεργάστηκε ως εικόνα σε άλλο πρόγραμμα):



Κάθε κόκκινη στήλη αντιστοιχεί σε μια κατηγορία μιας παραμέτρου (π.χ. κατηγορία Άνδρας στην παράμετρο "Φύλο"). Κάθε κόκκινη στήλη εμπεριέχει το άθροισμα των ποσοστών από τις σωστές απαντήσεις αυτής της κατηγορίας (π.χ. αν οι άντρες απάντησαν 51% σωστά στην 8η ερώτηση και 54% στην 9η κ.ο.κ. τότε το άθροισμα είναι ίσο με $51\% + 54\% + \dots$). Στόχος είναι για κάθε παράμετρο να ελεγχθεί η τυπική απόκλιση του μικρότερου βαθμού σωστών απαντήσεων με το μεγαλύτερο. Έτσι, στην παράμετρο "Ηλικία" τα άτομα 13-18 χρονών έχουν το μικρότερο βαθμό σωστών απαντήσεων (446%), άρα αυτή είναι η ηλικία που κάνουν τα περισσότερα λάθη. Παράλληλα, τα άτομα ηλικίας 19-24 χρονών έχουν μεγαλύτερο βαθμό σωστών απαντήσεων (540%), άρα απαντούν πιο σωστά. Η τυπική απόκλιση είναι απλά η διαφορά αυτών των δυο ποσοστών, δηλαδή $540\% - 446\% = 94\%$. Επομένως, η "Ηλικία" επηρεάζει με 0.94 βαθμό το αν θα απαντήσει σωστά ένα άτομο σε μια ερώτηση. Με τον ίδιο τρόπο υπολογίζονται και οι υπόλοιπες αποκλίσεις. Η απόκλιση της παραμέτρου "Σχέση με την επιστήμη της Πληροφορικής" είναι η μεγαλύτερη (142%).

Ως εκ τούτου, η πιο σημαντική παράμετρος, που επηρεάζει τις γνώσεις ενός ατόμου στις τεχνολογίες ιχνηλάτησης, είναι η "Σχέση με την επιστήμη της Πληροφορικής". Με άλλα λόγια, το κατά πόσο γνωρίζει ένα άτομο τα συστήματα ιχνηλάτησης εξαρτάται περισσότερο από το αν έχει σπουδάσει ή όχι Πληροφορική, παρά με το πόσο χρόνο αφιερώνει καθημερινά στο διαδίκτυο. Η δεύτερη πιο σημαντική παράμετρος είναι η "Ηλικία", η τρίτη είναι το "Φύλο" και η 4η παράμετρος είναι ο "καθημερινός χρόνος ενασχόλησης με το διαδίκτυο".

Σκοπός γ

Με τον τρίτο σκοπό γίνεται αντιληπτό κατά πόσο η άγνοια του ερωτώμενου επηρεάζει την καχυποψία και την επιφυλακτικότητα του σε αυτές τις τεχνολογίες. Για την επίτευξή του χρησιμοποιείται ένα σύνολο ερωτημάτων SQL σαν το εξής:

SELECT COUNT(ID) FROM SURVEY WHERE COL017 LIKE '%Ναι. Χρησιμοποιούν ιούς.%' or COL020 LIKE 'Ίσως να θέλουν να μου κάνουν κακό%' OR COL009 LIKE '%ιοί%'

Αυτό το ερώτημα επιστρέφει τον αριθμό των ατόμων που έχουν απαντήσει τουλάχιστον μια φορά με τη λέξη "ιός". Υποτίθεται πως αυτά είναι τα άτομα που φοβούνται τις τεχνολογίες καταγραφής. Αυτά είναι 203 άτομα. Προσθέτοντας κάθε φορά τη συνθήκη "AND Ερώτηση X = σωστή απάντηση" (π.χ. SELECT COUNT(ID) FROM SURVEY WHERE (COL017 LIKE '%Ναι. Χρησιμοποιούν ιούς.%' or COL020 LIKE 'Ίσως να θέλουν να μου κάνουν κακό%' OR COL009 LIKE '%ιοί%') AND COL009 = 'Είναι αρχεία κειμένου.'), επιλέγεται ο αριθμός των ατόμων, που φοβούνται τις τεχνολογίες ιχνηλάτησης και ταυτόχρονα απαντάνε σωστά στις ερωτήσεις. Εάν αυτός ο αριθμός διαιρεθεί με το 203, βρίσκεται το ποσοστό των ατόμων αυτών που απάντησαν σωστά. Ο μέσος όρος αυτών των ποσοστών δίνει το ποσοστό επιτυχίας για όλες τις ερωτήσεις για τα άτομα που φοβούνται τις τεχνολογίες ιχνηλάτησης. Αυτό είναι 0.39 ή αλλιώς 39% σωστές απαντήσεις. Επαναλαμβάνοντας την ίδια διαδικασία για τα υπόλοιπα άτομα (που δε φοβούνται τις τεχνολογίες ιχνηλάτησης), εξάγεται 0.37 ποσοστό επιτυχίας.

Επειδή τα 0.37 και 0.39 είναι σχετικά κοντινά νούμερα, συμπεραίνεται ότι οι γνώσεις ενός ατόμου είναι ανεξάρτητες από την ενδεχόμενη φοβία του για τα συστήματα παρακολούθησης.

Κεφάλαιο 6 - Συμπεράσματα και μελλοντική έρευνα

Η παρούσα πτυχιακή εργασία επιτρέπει μια καλή γνωριμία με πολλές υπάρχουσες τεχνολογίες καταγραφής δεδομένων και παρακολούθησης χρηστών, εξετάζοντάς τες από τη σκοπιά του τρόπου λειτουργίας τους. Ο αναγνώστης ενημερώνεται σχετικά και με τα αντίμετρά τους, σχηματίζοντας μια πλήρη εικόνα της τεχνολογικής διαδικασίας που ακολουθούν τα συστήματα ιχνηλάτησης. Η ταξινόμηση αυτών των συστημάτων είναι σχετικά νέα διαδικασία και βοηθάει στην επιπρόσθετη κατανόηση των βασικών τεχνολογικών στοιχείων, που χτίζουν τα συστήματα αυτά (λ.χ. HTTP, APIs). Ακόμα, γνωστοποιούνται οι τρέχουσες γνώσεις των χρηστών του διαδικτύου σχετικά με τα συστήματα αυτά, οι οποίες βρίσκονται σε ένα πολύ βασικό επίπεδο. Οι γνώσεις αυτές εξαρτώνται περισσότερο από τη σχέση του ατόμου με την επιστήμη της Πληροφορικής. Πράγματι, η έλλειψη τεχνολογικής πληροφόρησης σημαίνει έλλειψη κατανόησης αρχικά του συμπεριφορισμού και έπειτα των συνεπειών που μπορούν να επιφέρουν τα συστήματα αυτά.

Ο εμπλουτισμός του παρόντος επιτυγχάνεται κυρίως με τη σφαιρική εξέταση και μελέτη των τεχνολογιών ιχνηλάτησης. Η τεχνολογική λειτουργία τους είναι ίσως η πιο σημαντική και λιγότερο διαδεδομένη οπτική γωνία ανάλυσης. Εντούτοις, οι τεχνολογίες ιχνηλάτησης μπορούν να ερευνηθούν από τις σκοπιές της ιδιωτικότητας και των νομικών πλαισίων, της προγραμματιστικής ανάπτυξης και της αλγοριθμικής αποδοτικότητας, της εμπορικής χρήσης και επιτυχίας στο χώρο των επιχειρήσεων και εταιρειών, καθώς και από άλλες επιστημονικές πλευρές. Διότι, κανένας κλάδος δεν είναι και δε θα έπρεπε να είναι ανεξάρτητος από τους υπόλοιπους, αλλά η συλλειτουργία τους οφείλει να φέρει την ανάπτυξη και πρόοδό τους. Είναι εξίσου σημαντική η μελέτη και έρευνα μεθόδων ενημέρωσης για τα συστήματα παρακολούθησης και καταγραφής δεδομένων προς τους χρήστες του διαδικτύου. Ενδεχομένως να απαιτείται αρκετός χρόνος για την εκμάθηση όρων πληροφορικής και τεχνολογίας, καθώς και των μεθόδων λειτουργίας τους. Παράλληλα τα συστήματα ιχνηλάτησης αναπτύσσονται γρήγορα και ποικίλλουν. Συμπερασματικά, χρειάζεται να βρεθεί ένας αποδοτικός και αποτελεσματικός τρόπος εκμάθησης ακόμα και σε άτομα που δεν έχουν ιδιαίτερη σχέση με την επιστήμη της πληροφορικής.

Βιβλιογραφικές και Διαδικτυακές Ερευνητικές Πηγές

- [1] Τριανταφυλλίδου Βικτώρια, Η επεξεργασία των προσωπικών δεδομένων στο χώρο των ηλεκτρονικών επικοινωνιών: νομικές πτυχές, Διπλωματική Μεταπτυχιακή Εργασία, 2018
- [2] Timothy Libert, Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on One Million Websites, International Journal of Communication, October 2015
- [3] Ivar Derksen, HTML5 Tracking Techniques in Practice, Bachelor thesis, 2016
- [4] Κανονισμός (ΕΕ) 2016/679 Του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016
- [5] Franziska Roesner, Tadayoshi Kohno, and David Wetherall, Detecting and Defending Against Third-Party Tracking on the Web, 2012
- [6] Σερεσιώτη Μαρία, Μηχανισμοί Συλλογής Στοιχείων Στο Internet (Cookies), Πτυχιακή Εργασία, 2006
- [7] Aaron Cahn, Scott Alfeld, Paul Barford, S. Muthukrishnan, An Empirical Study of Web Cookies, 2016
- [8] Janice C. Sipior , Burke T. Ward & Ruben A. Mendoza, Online Privacy Concerns Associated with Cookies, Flash Cookies, and Web Beacons, 2011
- [9] Marshall Brain, How Internet Cookies Work, 2007
- [10] Alex X. Liu, Jason M. Kovacs, Chin-Tser Huang, Mohamed G. Gouda, A Secure Cookie Protocol, 2009
- [11] Tomasz Bujlow, Valentín Carela-Español, Josep Solé-Pareta, Pere Barlet-Ros, Web Tracking: Mechanisms, Implications, and Defenses, 2015
- [12] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, Claudia Diaz, The Web Never Forgets: Persistent Tracking Mechanisms in the Wild, 2014
- [13] Vipul Patel, A 5 Minute Overview of All New JavaScript APIs in HTML5, 2013
- [14] David M. Martin Jr., Hailin Wu, Adil Alsaid, Hidden surveillance by Web sites: Web bugs in contemporary use, 2003
- [15] Thomas F. Stafford, Andrew Urbaczewski, Spyware: The Ghost in the Machine, Communications of the Association for Information Systems, Volume 14, 2004
- [16] Engin Kirda and Christopher Kruegel, Greg Banks, Giovanni Vigna, and Richard A. Kemmerer, Behavior-based Spyware Detection, 15th USENIX Security Symposium, Pp. 273–288 of the Proceedings

- [17] https://en.wikipedia.org/wiki/Keystroke_logging
- [18] Steven Englehardt, Jeffrey Han, and Arvind Narayanan, I never signed up for this! Privacy implications of email tracking, 2018
- [19] Johannes Haupta, Benedict Benderb, Benjamin Fabianc, Stefan Lessmann, Robust Identification of Email Tracking: A Machine Learning Approach, 2018
- [20] Michael G. Foulger, Thomas R. Chipperfield, Jeremy S. Cooper, Andrew C. Storms, System and method related to generating and tracking an email campaign, 2006
- [21] User Guide Tracking E-mail Campaigns with G-Lock Analytics, 2009
- [22] Marco Gruteser, Xuan Liu, Protecting Privacy in Continuous Location Tracking Applications, 2004
- [23] Manav Singhal, Anupam Shukla, Implementation of Location based Services in Android using GPS and Web Services, 2012
- [24] https://en.wikipedia.org/wiki/HTTP_ETag
- [25] Nick Nikiforakis, Wouter Joosen, Benjamin Livshits, PriVaricator: Deceiving Fingerprinters with Little White Lies, 2014
- [26] Salma Hussain Abbasi, Muhammad Aleem, Muhammad Arshad Islam, User tracking mechanisms and counter-measures Asra Ishtiaq, 2017
- [27] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung, Patrick Mcdaniel, Anmol N. Sheth, TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones, 2014
- [28] Jim Chow, Ben Pfaff, Tal Garfinkel, Kevin Christopher, Mendel Rosenblum, Understanding Data Lifetime via Whole System Simulation, 2004
- [29] Hassan Metwalley, Stefano Traverso, Marco Mellia, Unsupervised Detection of Web Trackers, 2015
- [30] Steven Englehardt, Chris Eubank, Peter Zimmerman, Dillon Reisman, Arvind Narayanan, OpenWPM: An automated platform for web privacy measurement Draft, 2015
- [31] Willem Boumans, Web tracking and current countermeasures, Bachelor Thesis, 2017
- [32] Lorrie Faith Cranor, Web Privacy with P3P, 2002
- [33] Georg Merzdovnik, Markus Huber, Damjan Buhov, Nick Nikiforakis, Sebastian Neuner, Martin Schmiedecker, Edgar Weippl, Block Me If You Can: A Large-Scale Study of Tracker Blocking Tools, 2017
- [34] Jeff Tyson, How Virtual Private Networks Work, 2013

[35] The Living Internet, How Anonymizers Work, 2000 (πρόσβαση στις 19/4/2019)

[36] Dr Thomas F Burgess, INFORMATION SYSTEMS SERVICES Guide to the Design of Questionnaires, 2001

[37] A. Williams University of Bristol, How to... Write and analyse a questionnaire, Journal of Orthodontics Vol. 30, 2003

[38] R. Andonie, I. Dzitac, Razvan Andonie, Ioan Dzitac, How to Write a Good Paper in Computer Science and How Will It Be Measured by ISI Web of Knowledge, Int. J. of Computers Communications & Control Vol. V, 2010