ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΔΥΝΑΜΙΕΣ ΑΣΦΑΛΕΙΑΣ ΣΕ IOT ΚΑΙ ΑΝΤΙΜΕΤΩΠΙΣΗ ΤΟΥΣ

Της φοιτήτριας

Μπέλλου Παρασκευή

Αρ.Μητρώου: 123905

Επιβλέπων Καθηγητής

Γιακουμής Άγγελος

Θεσσαλονίκη 2023

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνές Πανεπιστημίου της Ελλάδος δεν υποδηλώνει απαραιτήτως και αποδοχή των απόψεων της δημιουργού, εκ μέρους του Τμήματος.

Ευχαριστίες

Η παρούσα πτυχιακή εργασία πραγματοποιήθηκε στο Διεθνές Πανεπιστήμιο της Ελλάδος, στο τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων.

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κύριο Άγγελο Γιακουμή για την επίβλεψη, την άριστη συνεργασία και εμπιστοσύνη που μου έδειξε.

Θα ήθελα επίσης να ευχαριστήσω τους γονείς μου για την στήριξη τους όλα αυτά τα χρόνια.

Πίνακας περιεχομένων

Κατάλογος εικόνων

Περίληψη

Η παρούσα εργασία καλύπτει μια σειρά εργαλείων και τεχνικών που μπορούν να εφαρμοστούν σε εφαρμογές και υπηρεσίες IoT, ιδιαίτερα όπου χρησιμοποιούνται αποθήκευση μεγάλων δεδομένων, αναλυτικά στοιχεία και μηχανική μάθηση με σκοπό την αντιμετώπιση των αδυναμιών ασφάλειας.

Το IoT παρέχει ένα ευρύ φάσμα χαρακτηριστικών που είναι σημαντικό να κατανοήσουμε κατά την αντιμετώπιση του απορρήτου. Αυτά αντικατοπτρίζουν γενικά τη φύση των μεγάλων δεδομένων (όγκος, ποικιλία, μεταβλητότητα, αλήθεια, ακρίβεια).

Η παροχή διαφάνειας (openness και transparency) και ειδοποίησης αποτελεί πρόκληση, όταν οι συσκευές IoT που χρησιμοποιούνται δεν διαθέτουν κατάλληλους μηχανισμούς εμφάνισης/εισόδου.

Στην βιβλιογραφική ανασκόπηση της εργασίας μελετάται η κατηγοριοποίηση των προσωπικών δεδομένων, τα οφέλη των μεγάλων δεδομένων και της τεχνητής νοημοσύνη καθώς και η παρουσίαση τεχνικών και συστάσεων βέλτιστων πρακτικών από την σύγχρονη διεθνή βιβλιογραφία που παρέχουν διάφορες προστασίες σχετικά με την απόκτηση, αποθήκευση και χρήση προσωπικών πληροφοριών (Κρυπτογράφηση, Κατακερματισμός, Salting, Ανωνυμοποίηση, Ψευδωνυμοποίηση). Για τα παραπάνω παρατίθενται ενδεικτικά παραδείγματα από την σύγχρονη διεθνή βιβλιογραφία.

Abstract

This paper covers a range of tools and techniques that can be applied to IoT applications and services, especially where big data storage, analytics, and machine learning are used to address security vulnerabilities.

The IoT provides a wide range of features that are important to understand when dealing with privacy. These generally reflect the nature of big data (volume, variety, variability, truth, accuracy)

Providing openness and transparency and alert is a challenge when the IoT devices in use do not have proper display / input mechanisms.

The bibliographic review of the work studies the categorization of personal data, the benefits of big data and artificial intelligence as well as the presentation of techniques and best practice recommendations from the modern international literature that provide various protections regarding the acquisition, storage and use of personal information (Encryption, Fragmentation, Salting, Anonymization, Nickname). Indicative examples from the modern international bibliography are given for the above.

1. Εισαγωγή

Η ταχεία ανάπτυξη του Διαδικτύου των πραγμάτων δημιουργεί σημαντικές ευκαιρίες στους χρήστες να επωφεληθούν από υπηρεσίες που βασίζονται στην απόκτηση και αποθήκευση δεδομένων, στην ανάλυση και στη μηχανική μάθηση. Είναι σαφές ότι οι σχεδιαστές υπηρεσιών πρέπει να ακολουθήσουν μια υπεύθυνη προσέγγιση για να διασφαλίσουν ότι τα προσωπικά δεδομένα και το απόρρητο προστατεύονται για τους χρήστες, ιδιαίτερα καθώς το IoT επεκτείνεται περαιτέρω στην καθημερινή ζωή. Αυτό το έγγραφο καλύπτει μια σειρά εργαλείων και τεχνικών που μπορούν να εφαρμοστούν σε εφαρμογές και υπηρεσίες IoT, ιδιαίτερα όπου χρησιμοποιούνται αποθήκευση μεγάλων δεδομένων, αναλυτικά στοιχεία και μηχανική μάθηση. Αναμφίβολα, ένα από τα μεγαλύτερα πλεονεκτήματα του IoT είναι η δυνατότητα συσχέτισης μιας ποικιλίας διαφορετικών πηγών δεδομένων για τη δημιουργία νέων υπηρεσιών [1].

Η χρήση διαφορετικών πηγών δεδομένων επιτρέπει τόσο στους χρήστες όσο και στους παρόχους υπηρεσιών να αποκτήσουν καλύτερες γνώσεις σε σχέση με μια συγκεκριμένη κατάσταση. Για παράδειγμα, οι υπηρεσίες αντιμετώπισης καταστροφών που χρησιμοποιούν IoT και τεχνολογίες μεγάλων δεδομένων, όπως πληροφορίες σε πραγματικό χρόνο, ιστορικές και προβλέψεις για τον καιρό, τη στάθμη του νερού, το πλήθος, την κυκλοφορία και άλλα σχετικά σημεία δεδομένων, μπορούν να βελτιώσουν σημαντικά την ικανότητα των ομάδων έρευνας και διάσωσης για καλύτερη προετοιμασία και αντίδραση στα γεγονότα. Αυτό μπορεί να φανεί στην τρέχουσα αντιμετώπιση καταστροφών και έκτακτης ανάγκης, όπου οι απώλειες ζωών έχουν μειωθεί σημαντικά από τις βελτιώσεις στις επικοινωνίες και τη διανομή πληροφοριών. Η συλλογή και αποθήκευση δεδομένων μέσω συσκευών IoT θα πρέπει φυσικά να πληροί τις νομικές απαιτήσεις [2].

Το έγγραφο GSMA «Αξιολόγηση κανονιστικών απαιτήσεων της διαχείρισης απορρήτου για μέλη που προσφέρουν υπηρεσίες IoT με χρήση προσωπικών δεδομένων» εξετάζει τις απαιτήσεις από ρυθμιστικές αρχές στην ΕΕ, τις ΗΠΑ, την Ιαπωνία, την Ινδία και τη Βραζιλία σχετικά με το απόρρητο των

προσωπικών δεδομένων, με αυτές που ενδέχεται να διαμορφώσουν άλλα ρυθμιστικά καθεστώτα σε όλο τον κόσμο. Ο σεβασμός και η προστασία της ιδιωτικής ζωής είναι μια ευκαιρία για την οικοδόμηση εμπιστοσύνης των καταναλωτών. Στις «Αρχές απορρήτου για φορητές συσκευές»[2], η GSMA έχει ήδη προσδιορίσει ένα σύνολο οκτώ αρχών που συμβάλλουν στην προώθηση του απορρήτου των καταναλωτών στο οικοσύστημα κινητής τηλεφωνίας. Ορισμένες από αυτές τις αρχές, όπως το άνοιγμα και η διαφάνεια, η ελαχιστοποίηση δεδομένων και ο σεβασμός της επιλογής και του ελέγχου των χρηστών είναι ιδιαίτερα σχετικές με το IoT.

Η έρευνα GSMA [3] που διεξήχθη στο παρελθόν σχετικά με τις υπηρεσίες κινητής τηλεφωνίας υπογραμμίζει τον τρόπο με τον οποίο οι χρήστες κινητών τηλεφώνων θέλουν πραγματικά να διαχειρίζονται το απόρρητό τους με σαφή, απλό και διακριτικό τρόπο. Οι φόβοι για το απόρρητο μπορούν να εμποδίσουν την ανάπτυξη των υπηρεσιών κινητής τηλεφωνίας και η ίδια λογική ισχύει και για το IoT [5].

Ουσιαστικά, οι τελικοί χρήστες εκτιμούν πραγματικά τον τρόπο με τον οποίο αντιμετωπίζεται το απόρρητο και η οικοδόμηση εμπιστοσύνης μαζί τους είναι win-win: οδηγεί στην υιοθέτηση και επιτρέπει την καινοτομία. Αντίθετα, η αποτυχία αντιμετώπισης ζητημάτων απορρήτου οδηγεί σε συνέπειες όπως βλάβη στη φήμη ή δαπανηρές αγωγές. Ως εκ τούτου, οι επιχειρήσεις θα επωφεληθούν από τη διενέργεια ανάλυσης επιπτώσεων στο απόρρητο πριν από την ανάπτυξη και την ανάπτυξη μιας νέας υπηρεσίας. Ένα παράδειγμα τέτοιας έλλειψης προσοχής στο απόρρητο οδήγησε την ολλανδική κυβέρνηση να αναγκαστεί να ανακαλέσει το νομοσχέδιο για τις έξυπνες μετρήσεις το 2009 λόγω ανησυχιών για την προστασία της ιδιωτικής ζωής[4].

Πρόσφατα, τα μεγάλα δεδομένα έχουν συγκεντρώσει μεγάλη προσοχή από τη βιομηχανία, τις επιστημονικές και τεχνολογικές κοινότητες, τα μέσα ενημέρωσης και πολλά κυβερνητικά τμήματα. Πολλές χώρες χρησιμοποιούν επίσης μεγάλα δεδομένα για να παρέχουν υπηρεσίες σε διάφορους τομείς όπως η υγειονομική περίθαλψη, η ιατρική, οι επιχειρήσεις του δημόσιου τομέα, η διανομή, το μάρκετινγκ και η μεταποίηση. Τα μεγάλα δεδομένα είναι ουσιαστικά μια τεχνολογία που βασίζεται σε πληροφορίες που αναλύει

μεγάλες ποσότητες δεδομένων για να εξάγει πολύτιμες πληροφορίες και προβλέπει αλλαγές με βάση την εξαγόμενη γνώση. Θεωρείται μια νέα πηγή ενέργειας που οδηγεί τις επιχειρηματικές και τεχνολογικές καινοτομίες καθώς και την οικονομική ανάπτυξη [1].

Πολλά οικονομικά και πολιτικά συμφέροντα οδηγούν τα μεγάλα δεδομένα, ειδικά τις διαδικασίες ενοποίησης δεδομένων, ανάλυσης και εξόρυξης δεδομένων. Συγκεκριμένα, τα οργανωμένα μεγάλα δεδομένα που συλλέγονται από διάφορες πηγές, όπως πλατφόρμες μέσων κοινωνικής δικτύωσης, ιστότοποι και συστήματα παγκόσμιας τοποθέτησης θα βοηθήσουν στον εντοπισμό διαφόρων κοινωνικοοικονομικών προβλημάτων και θα βοηθήσουν επίσης στην παροχή αποτελεσματικών λύσεων και μέτρων. Η χρήση μεγάλων δεδομένων σε διάφορους τομείς έχει οδηγήσει σε ταχεία αύξηση σε μεγάλη ποικιλία πόρων δεδομένων και διάφορες τεχνολογίες ανάλυσης δεδομένων, όπως η τυποποιημένη εξόρυξη δεδομένων και οι τεχνικές στατιστικής ανάλυσης, επιταχύνουν τη συνεχή επέκταση της αγοράς μεγάλων δεδομένων. Ένα σημαντικό χαρακτηριστικό των μεγάλων δεδομένων είναι ότι τα δεδομένα από διάφορες πηγές έχουν κύκλους ζωής από τη συλλογή έως την καταστροφή και νέες πληροφορίες μπορούν να προκύψουν μέσω ανάλυσης, συνδυασμού και χρήσης [3].

Όπως αναφέρθηκε προηγουμένως, τα μεγάλα δεδομένα προσφέρουν πολλά πλεονεκτήματα και δυνατότητες για καινοτομία σε διάφορους τομείς, αλλά επίσης παρουσιάζουν πολλά ζητήματα και προκλήσεις. Πρώτον, η ασφάλεια των δεδομένων, η διατήρηση της ιδιωτικής ζωής και τα ηθικά ζητήματα αποτελούν σημαντικές ανοιχτές προκλήσεις στο οικοσύστημα καινοτομίας μεγάλων δεδομένων και περιλαμβάνουν μεθόδους διαχείρισης πληροφοριών, προστασία προσωπικών ή θανατηφόρων πληροφοριών και κακή χρήση αναλύσεων δεδομένων. Συγκεκριμένα, ένας μεγάλος όγκος κοινών πληροφοριών, συμπεριλαμβανομένου του απορρήτου, μπορεί να αξιοποιηθεί σε ένα διασυνδεδεμένο ανοιχτό περιβάλλον [5].

Ως εκ τούτου, διάφοροι οργανισμοί τυποποίησης έχουν δημοσιεύσει σχετικά πρότυπα για την ασφάλεια και τη διατήρηση του απορρήτου των μεγάλων δεδομένων και έχουν θεσπιστεί νόμοι για την προστασία της ιδιωτικής ζωής,

όπως ο γενικός κανονισμός προστασίας δεδομένων (GDPR) στην Ευρώπη και ο νόμος περί απορρήτου των καταναλωτών της Καλιφόρνια (CCPA) στις Ηνωμένες Πολιτείες θεσπίστηκε. Ωστόσο, τα πρότυπα που σχετίζονται με την ασφάλεια μεγάλων δεδομένων εξηγούν μόνο τις απαιτήσεις ασφαλείας και δεν έχουν καμία περιγραφή που σχετίζεται με τεχνικές ασφαλείας.

Εικόνα 1.Ο κύκλος ζωής των μεγάλων δεδομένων

Επιπλέον, δεδομένου ότι ο GDPR και ο CCPA στοχεύουν συγκεκριμένες περιοχές, δεν γενικεύονται σε διάφορους οργανισμούς και ερευνητές που χρησιμοποιούν μεγάλα δεδομένα. Δεύτερον, κάθε φάση του κύκλου ζωής έχει ζητήματα ασφάλειας και αξιοπιστίας δεδομένων και η προστασία των προσωπικών πληροφοριών είναι ζωτικής σημασίας. Συγκεκριμένα, οι τάσεις των χρηστών μπορούν να αναλυθούν με τη χρήση διαφόρων αναλυτικών στοιχείων μεγάλων δεδομένων, οδηγώντας σε παραβίαση του προσωπικού απορρήτου.

Διάφορες τεχνολογίες για τη διατήρηση της ασφάλειας και του απορρήτου σε ένα περιβάλλον μεγάλων δεδομένων έχουν προταθεί και ήταν υπό ανάπτυξη μέχρι πρόσφατα. Αυτά μπορούν να χωριστούν και να ομαδοποιηθούν σύμφωνα με τις φάσεις του κύκλου ζωής των μεγάλων δεδομένων [6].

Σκοπός και συνεισφορά της εργασίας

Η παρούσα εργασία εντοπίζει απειλές και ζητήματα ασφάλειας που εμφανίζονται στον κύκλο ζωής των μεγάλων δεδομένων επιβεβαιώνοντας τα τρέχοντα πρότυπα που έχουν αναπτυχθεί από διεθνείς οργανισμούς τυποποίησης και αναλύοντας σχετικές μελέτες. Ακόμα καλύπτει μια σειρά εργαλείων και τεχνικών που μπορούν να εφαρμοστούν σε εφαρμογές και

υπηρεσίες IoT, ιδιαίτερα όπου χρησιμοποιείται η αποθήκευση μεγάλων δεδομένων, η ανάλυση και η μηχανική μάθηση.

2. Θεωρήσεις απορρήτου για το IoT

Το IoT περιλαμβάνει μια τεράστια ποικιλία συσκευών, πραγμάτων, εφαρμογών και λύσεων και επομένως δεν είναι δυνατό να επιλεγεί ένα οριστικό σύνολο προβληματισμών που μοιράζονται ή καλύπτουν όλα τα πράγματα στο IoT. Τα θέματα αυτής της ενότητας έχουν επιλεγεί ως συχνά επαναλαμβανόμενα θέματα στο IoT που ενδέχεται να επηρεάσουν το απόρρητο. Επομένως, οι προγραμματιστές εφαρμογών και υπηρεσιών θα πρέπει να λαμβάνουν υπόψη αυτά τα ζητήματα κατά το σχεδιασμό λύσεων [7].

2.1. Δεδομένα στο IoT

Το IoT παρέχει ένα ευρύ φάσμα χαρακτηριστικών που είναι σημαντικό να κατανοήσουμε κατά την αντιμετώπιση του απορρήτου. Αυτά αντικατοπτρίζουν γενικά τη φύση των μεγάλων δεδομένων: [4]

- Όγκος – ο αριθμός των συσκευών IoT και των πραγμάτων που θα συνδεθούν στο Διαδίκτυο θα ανέλθει σε πολλά δισεκατομμύρια, επομένως ο αριθμός των πηγών δεδομένων θα είναι τεράστιος.
- Ποικιλία – ο τύπος και η φύση των δεδομένων θα διαφέρουν σημαντικά μεταξύ των τύπων συσκευών, των προσεγγίσεων του κατασκευαστή και ακόμη και των μοντέλων συσκευών. Για παράδειγμα, ένας οικιακός θερμοστάτης «χαμηλού επιπέδου» από έναν κατασκευαστή μπορεί απλώς να αναφέρει μια ένδειξη θερμοκρασίας σε μια υπηρεσία cloud και να λαμβάνει εντολές ελέγχου για τη λειτουργία της θέρμανσης, ενώ μια συσκευή ανώτερης τεχνολογίας από διαφορετικό κατασκευαστή μπορεί να χρησιμοποιεί τοποθεσία GPS για βελτιστοποίηση της λειτουργίας της θέρμανσης & συστήματα ψύξης ανάλογα με τις καιρικές συνθήκες και έχουν μεγαλύτερο βαθμό αυτονομίας στη διαδικασία ελέγχου που απαιτεί τοπική αποθήκευση και επεξεργασία.

- Velocity - ένα σημαντικό ποσοστό των συσκευών IoT θα έχουν την ικανότητα να δημιουργούν μεγάλους όγκους δεδομένων, ειδικά όταν παρέχουν υπηρεσίες παρακολούθησης και ελέγχου σε πραγματικό ή σχεδόν πραγματικό χρόνο. Σε συνδυασμό με τον τεράστιο όγκο των συσκευών IoT και των πραγμάτων, θα προκύψουν εξαιρετικά μεγάλοι όγκοι δεδομένων. Αντίθετα, ωστόσο, θα υπάρχει επίσης σημαντικός αριθμός συσκευών που χρειάζεται να επικοινωνούν μόνο περιστασιακά κατά τη διάρκεια ζωής της συσκευής. Λόγω αυτών των διαφορών, είναι πιθανό τα δεδομένα να υποβάλλονται σε επεξεργασία και να αποθηκευτούν με αρκετά κατανεμημένο και ιεραρχικό τρόπο και, επομένως, τυχόν δεδομένα που τυχαίνει να είναι προσωπικά δεδομένα ενδέχεται να μην αποθηκευτούν σε ένα μόνο μέρος. Επίσης, δεν θα είναι πρακτικό να ελέγχεται με μη αυτόματο τρόπο κάθε στοιχείο δεδομένων, πηγή ή εφαρμογή.
- Μεταβλητότητα – οι κατασκευαστές ενδέχεται να αποθηκεύουν και να επεξεργάζονται δεδομένα με διαφορετικό τρόπο ακόμη και για παρόμοιες τιμές δεδομένων. Για παράδειγμα, ένας κατασκευαστής μπορεί να αποθηκεύσει τη διεύθυνση email του κατόχου στη συσκευή και ένας άλλος στο cloud. Ή τα ονόματα μεταβλητών μπορεί να περιέχουν πολύ διαφορετικούς τύπους δεδομένων μεταξύ δύο κατασκευαστών π.χ. ένα πεδίο «τοποθεσίας» από έναν κατασκευαστή μπορεί να περιέχει δεδομένα GPS όταν ένας άλλος κατασκευαστής τα χρησιμοποιεί μόνο για τη χώρα του πελάτη. Μια πλατφόρμα που σχεδιάστηκε αρχικά με βάση τη λήψη δεδομένων χαμηλού κινδύνου από τη συσκευή IoT ενός κατασκευαστή θα μπορούσε αργότερα να λάβει δεδομένα υψηλότερου κινδύνου από έναν άλλο κατασκευαστή, εάν ο δεύτερος κατασκευαστής «μιμηθεί» την υπάρχουσα διεπαφή αλλά περιλαμβάνει δεδομένα υψηλού κινδύνου χωρίς τη γνώση των προγραμματιστών της πλατφόρμας υπάρχει ένα ζήτημα που ο προγραμματιστής μπορεί να αγνοεί τελείως.
- Ειλικρίνεια – η αλήθεια ή η ακρίβεια των καταγεγραμμένων δεδομένων μπορεί να διαφέρει μεταξύ συσκευών και κατασκευαστών και επομένως η «ακρίβεια» μπορεί να έχει αντίκτυπο στο απόρρητο. Για

παράδειγμα, εάν μια συσκευή IoT στέλνει σε πραγματικό χρόνο ροή δεδομένων GPS με ακρίβεια 1 μέτρου, έχει τη δυνατότητα να αποκαλύψει πολλά σχετικά με τα πρότυπα ζωής του ιδιοκτήτη, αλλά εάν η ακρίβεια είναι 1 km και αναφέρεται σε ωριαία βάση, θα έχει μικρότερες ανησυχίες;

Εκτός από τους παραπάνω παράγοντες, η μακροζωία των δεδομένων και ο κύκλος ζωής της υπηρεσίας που συλλέγει τα δεδομένα είναι σημαντικά για το IoT. Ορισμένες συσκευές IoT μπορεί να παράγουν πολύ λίγα δεδομένα, αλλά η διάρκεια ζωής της συσκευής μπορεί να είναι πολύ μεγάλη, της τάξης των 10-20 ετών. Για αυτές τις συσκευές ενδέχεται να υπάρχουν αλλαγές τόσο στην ιδιοκτησία όσο και στους χρήστες, με τυπικά παραδείγματα τη μέτρηση. Αυτές οι αλλαγές θα πρέπει να ληφθούν υπόψη κατά την αξιολόγηση των επιπτώσεων στο απόρρητο. Για παράδειγμα, η μάρκα ρούχων Benetton [7] το 2003 αντιμετώπισε διαμαρτυρίες για παραβίαση απορρήτου όταν αποφάσισαν να προσθέσουν ετικέτες RFID σε όλα τα ρούχα, καθώς δεν είχαν λάβει υπόψη την πλήρη διάρκεια ζωής του αντικειμένου [10].

Ένα κοινό επιχείρημα ήταν ότι «η αποθήκευση είναι φθηνή», επομένως είναι πιο εύκολο να αποθηκεύσετε τα πάντα αντί να αφιερώσετε πολύ χρόνο στο να σκεφτείτε ποια δεδομένα πρέπει να αποθηκευτούν. Η αρχή της «ελαχιστοποίησης και διατήρησης δεδομένων» προκαλεί τους προγραμματιστές να σκεφτούν ολιστικά για την απόκτηση και την αποθήκευση δεδομένων [9] και αυτό είναι ιδιαίτερα σχετικό με το IoT, ωστόσο, με τους παραπάνω παράγοντες είναι επίσης πολύ πιο δύσκολο να αντιμετωπιστεί. Στην πράξη, η ελαχιστοποίηση των δεδομένων θα πρέπει να βασίζεται στην υπηρεσία που λαμβάνουν οι χρήστες την τρέχουσα στιγμή, καθώς και να διατηρεί ανοιχτές τις εύλογες επιλογές για μελλοντικές εξελίξεις των υπηρεσιών, αποφεύγοντας ταυτόχρονα τη συλλογή δεδομένων που είναι απίθανο να σχετίζονται με τις απαιτήσεις των πελατών.

2.2. Περιορισμένες διεπαφές χρήστη ή εμπειρία χρήστη

Οι συσκευές συνδεδεμένες στο Διαδίκτυο είχαν ιστορικά οθόνες και μηχανισμούς εισόδου, όπως πληκτρολόγια, ποντίκια και οθόνες αφής. Είτε ένας χρήστης είχε υπολογιστή, είτε είχε έξυπνο τηλέφωνο υπήρχε η δυνατότητα να ενημερώνεται ο χρήστης μέσω οθόνης και να λαμβάνει κάποια σχόλια από τον χρήστη π.χ. κάνοντας κλικ στην επιβεβαίωση όρων και προϋποθέσεων, εγγραφή της συσκευής, εγκατάσταση σχετικής εφαρμογής ή έλεγχος ορισμένων πτυχών εφαρμογών και υπηρεσιών. Οι δυνατότητες των συσκευών IoT θα ποικίλλουν σημαντικά και, ενώ ορισμένες από τις πιο εξελιγμένες συσκευές θα έχουν οθόνες και μηχανισμούς εισόδου, αυτό δεν θα ισχύει για όλους. Πολλές συσκευές IoT έχουν σχεδιαστεί για «ανάπτυξη και λήθη», με σκοπό να μην απαιτείται αλληλεπίδραση με τους ανθρώπους και επομένως δεν έχουν καμία διεπαφή χρήστη. Η αλληλεπίδραση με τέτοιες συσκευές μπορεί να συμβεί μέσω ενός συστήματος διαχείρισης που μπορεί να είναι είτε κεντρικό είτε μέσω εφαρμογής σε παραδοσιακές συσκευές όπως smartphone, tablet ή υπολογιστές. Οι δυνατότητες θα κυμαίνονται, όπως φαίνεται στην παρακάτω εικόνα [12]

Εικόνα 2. Δυνατότητες αλληλεπίδρασης συσκευών IOT

Είναι πιθανό ότι οι όγκοι των συσκευών IoT θα είναι σημαντικά υψηλότεροι στο χαμηλότερο άκρο της καμπύλης τιμολόγησης και ικανότητας, και αυτό θα σημαίνει ότι αυτές οι συσκευές είναι λιγότερο πιθανό να έχουν οθόνη ή άμεσο μηχανισμό για την είσοδο του χρήστη. Επίσης, δεν μπορεί να θεωρηθεί ότι ο πελάτης για μια υπηρεσία IoT έχει μια «συνοδευτική» συσκευή (όπως μια συσκευή smartphone) με διεπαφή για την παροχή επιλογών επιλογής και ελέγχου ή επίβλεψης υπηρεσιών. Επίσης, όλο και περισσότερο, οι συσκευές και οι υπηρεσίες IoT θα αγοράζονται από άτομα που μπορεί να μην έχουν καν μια κατάλληλα ικανή φορητή συσκευή για να παρέχουν επιλογή και έλεγχο ή μπορεί να αγοράζονται από άτομα μεταξύ των 750 εκατομμυρίων ενηλίκων που εκτιμάται ότι είναι αναλφάβητοι και δεν μπορούν να ασκούν άμεσα τις προτιμήσεις τους. Επομένως, η παροχή διαφάνειας, και ειδοποίησης μπορεί να είναι πρόκληση, όταν οι συσκευές IoT που χρησιμοποιούνται δεν διαθέτουν κατάλληλους μηχανισμούς εμφάνισης/εισόδου. Οι συγκεκριμένες προκλήσεις περιλαμβάνουν: [11]

- Προβολή και αποδοχή Όρων & Προϋποθέσεων – εάν δεν υπάρχει μηχανισμός εισαγωγής οθόνης/χρήστη ή η εμφάνιση της συσκευής IoT δεν είναι πρακτική για το σκοπό αυτό.
- Παροχή ενημερωμένης συγκατάθεσης για επεξεργασία ή αποθήκευση δεδομένων ή συναφείς σκοπούς.
- Παροχή οποιασδήποτε ειδοποίησης σε σχέση με μια αλλαγή κατάστασης, π.χ. λήξη της συγκατάθεσης, τροποποιημένοι όροι προϋποθέσεων, νέα λειτουργικότητα ή σκοπός επεξεργασίας που προστέθηκε που απαιτεί συναίνεση κ.λπ.
- Αίτηση από τους χρήστες να αναβαθμίσουν τις συσκευές τους για να λάβουν ενημερώσεις ασφαλείας που βελτιώνουν το απόρρητο. Ως εκ τούτου, οι σχεδιαστές λύσεων θα πρέπει να εξετάσουν πώς να επιτύχουν τη διαφάνεια και την ενημερωμένη συναίνεση για υπηρεσίες που βασίζονται στο IoT.

2.3. Κύκλος ζωής ιδιοκτησίας και χρήσης

Πολλές κινητές συσκευές, συμπεριλαμβανομένων των έξυπνων τηλεφώνων και των tablet, έχουν έναν χρήστη που χρησιμοποιεί τη συσκευή όλη την ώρα. Αυτές είναι συχνά προσωπικές συσκευές και ο χρήστης θα εξατομικεύσει τις υπηρεσίες του ανάλογα, συμπεριλαμβανομένης της εγγραφής ή της διαμόρφωσης των στοιχείων του λογαριασμού του, της αποδοχής των Όρων & Προϋποθέσεων και της διαμόρφωσης λειτουργιών που σχετίζονται με το απόρρητο, όπως η κοινή χρήση τοποθεσίας. Ωστόσο, οι συσκευές IoT μπορεί να είναι αρκετά διαφορετικές. Αν και μια συσκευή IoT, όπως μια προσωπική οθόνη φυσικής κατάστασης, θα μπορούσε να χρησιμοποιηθεί από έναν μόνο «κάτοχο», υπάρχουν πολλές περιπτώσεις χρήσης στο IoT όπου υπάρχουν πιο περίπλοκα σενάρια ιδιοκτησίας και χρήσης όπως παρουσιάζονται στην παρακάτω εικόνα [9].

Εικόνα 3. Χρήσεις στο IOT με περίπλοκα σενάρια ιδιοκτησίας και χρήσης

Επομένως, ο κάτοχος/αγοραστής της συσκευής μπορεί να μην είναι ο χρήστης αυτής της συσκευής, μια συσκευή μπορεί με την πάροδο του χρόνου να χρησιμοποιηθεί από διαφορετικά άτομα. Επομένως, προκύπτουν ερωτήματα σχετικά με το εάν ο χρήστης πρέπει να παράσχει τη συγκατάθεσή του και εάν υπάρχει τρόπος για έναν νέο ή πρόσθετο χρήστη να παράσχει τη συγκατάθεσή του. Αυτές οι ερωτήσεις επιδεινώνονται εάν ο κατασκευαστής δεν μπορεί να διακρίνει εάν έχει μεταβιβαστεί η ιδιοκτησία ή εάν η συσκευή χρησιμοποιείται από κάποιον νέο. Επομένως, ο κύκλος ζωής της συσκευής θα πρέπει να λαμβάνεται υπόψη κατά το σχεδιασμό τόσο της συσκευής όσο και οποιωνδήποτε υπηρεσιών που βασίζονται στη συσκευή. Ο σχεδιασμός της υπηρεσίας θα πρέπει να λαμβάνει υπόψη θέματα όπως οι αλλαγές

ιδιοκτησίας της συσκευής και η επίδραση στο απόρρητο, και να εξετάζει τη δυνατότητα εκ νέου αναγνώρισης της συναίνεσης ή διαγραφής δεδομένων στα σημεία όπου αλλάζει η ιδιοκτησία.

Πολυπλοκότητες αγοράς και οικοσυστήματος όπως περιγράφηκε παραπάνω, υπάρχουν διάφορα σενάρια όπου δεν ισχύει η παραδοσιακή αρχή «ένα άτομο που ελέγχει την εγγραφή ή τη διαμόρφωση συσκευών και εφαρμογών». Αυτό καθιστά ακόμη πιο δύσκολες διαδικασίες όπως η απόκτηση συναίνεσης για κοινή χρήση δεδομένων, επιτρέποντας ακόμη και περιορισμούς διεπαφής χρήστη για το IoT. Οι παρακάτω είναι ρόλοι που μπορούν να συμμετάσχουν στην παράδοση και τη χρήση μιας πλούσιας υπηρεσίας IoT, όπως μια υπηρεσία ελέγχου πρόσβασης κτιρίου: [12]

- Ο ιδιοκτήτης του «πράγματος» στο οποίο έχει εγκατασταθεί μια συσκευή IoT – όπως ο ιδιοκτήτης ενός κτιρίου γραφείων.
- Ο κάτοχος της(των) συσκευής(ών) IoT – όπως ένας πάροχος υπηρεσιών εξειδικευμένων συστημάτων ασφαλείας.
- Ο πάροχος της υπηρεσίας επικοινωνιών που χρησιμοποιεί η συσκευή IoT για την αποστολή/λήψη δεδομένων π.χ. το δίκτυο MobileIoT.
- Ο πάροχος της υπηρεσίας που συλλέγει και επεξεργάζεται τα δεδομένα της συσκευής IoT και αποστέλλει οποιαδήποτε πληροφορία/ έλεγχο στη συσκευή IoT – συμπεριλαμβανομένων:

    ✚ Ο πάροχος της υποδομής αποθήκευσης/επεξεργασίας δεδομένων π.χ. Υπηρεσίες Ιστού της Amazon που χρησιμοποιεί αυτή η κύρια υπηρεσία.

    ✚ Η συνεργαζόμενη εταιρεία που ανέπτυξε την κύρια εφαρμογή/υπηρεσία και παραμένει υπεύθυνη για τη συντήρησή της.

- Το άτομο, τα άτομα ή ο οργανισμός που συνάπτει σύμβαση με τον ιδιοκτήτη του «πράγματος» στο οποίο έχουν εγκατασταθεί οι συσκευές IoT, π.χ., μια επιχείρηση που συνάπτει σύμβαση για χώρους γραφείων, ο οποίος μπορεί να περιλαμβάνει: [8]

✚ Οι εργαζόμενοι που διαμένουν στο γραφείο και τους επιτρέπεται η πρόσβαση σε γενικά μέρη του γραφείου.

✚ Εργαζόμενοι στους οποίους επιτρέπεται η πρόσβαση σε εξειδικευμένα μέρη του γραφείου, π.χ. πρόσβαση τμήματος πληροφορικής σε μηχανοστάσιο.

✚ Καθαριστικά γραφείων που παρέχονται από υπεργολάβο στην επιχείρηση.

✚ Επισκέπτες του γραφείου στους οποίους επιτρέπεται η πρόσβαση σε περιοχές επισκεπτών.

✚ Μια ομάδα διαχείρισης γραφείου που μπορεί να διαμορφώσει τις λεπτομέρειες πρόσβασης.

- Άλλοι οργανισμοί εταίροι που ενδέχεται να επεξεργάζονται πληροφορίες πρόσβασης, για παράδειγμα «διαστημικός σχεδιασμός» για τη βελτίωση της χωρητικότητας ή της αποτελεσματικότητας της διάταξης του γραφείου

Σε ένα τέτοιο σενάριο, οι διάφορες εξωτερικές συμβατικές ρυθμίσεις και οι εσωτερικές υποχρεώσεις ενδέχεται να επιτρέπουν την κοινή χρήση δεδομένων, αλλά ενδέχεται να μην υπάρχει ένα ενιαίο έργο όπου όλα τα μέρη εγγράφονται αρχικά για να καθορίσουν τις διάφορες άδειες, συναινέσεις και περιορισμούς σχετικά με την κοινή χρήση δεδομένων. Ομοίως, ενδέχεται να μην υπάρχει σύμβαση έργου ορισμένου χρόνου μετά την οποία όλα τα δεδομένα διαγράφονται, καθώς ορισμένα από αυτά τα δεδομένα μπορεί εύλογα να χρειάζονται για σκοπούς συμμόρφωσης για πολλά χρόνια μετά τη συλλογή τους. Μπορεί να υπάρχουν μεμονωμένες συμφωνίες για συγκεκριμένους σκοπούς - π.χ. δέσμευση του οργανισμού «διαστημικού σχεδιασμού» – και αυτό το έργο και η συμφωνία θα μπορούσαν να καλύπτουν θέματα όπως η κρυπτογράφηση δεδομένων και επίσης η διαγραφή προσωπικών δεδομένων κατά την ολοκλήρωση του έργου. Επομένως, οι λύσεις θα πρέπει να σχεδιάζονται έτσι ώστε να υπάρχει

υποστήριξη για τα διάφορα άτομα ή οργανισμούς που χρησιμοποιούν τη λύση για την προστασία του απορρήτου τους σύμφωνα με τον ρόλο τους, τις συναινέσεις και τα ζητήματα όπως η συμμόρφωση με τη σύμβαση [13].

2.4. Προσωπικά δεδομένα στο IoT

Τα προσωπικά δεδομένα μπορεί να είναι αρκετά δύσκολο να κατηγοριοποιηθούν για το IoT. Γενικά, όσο περισσότερα δεδομένα και περισσότεροι τύποι δεδομένων συλλέγονται, τόσο πιο πιθανό είναι να υπάρχουν αναγνωρίσιμα προσωπικά δεδομένα ως μέρος αυτού του συνόλου δεδομένων. Γενικά, μπορεί να θεωρηθεί ότι προσωπικά δεδομένα είναι οποιαδήποτε πληροφορία προσδιορίζει σαφώς ή είναι εύλογα αναγνωρίσιμη για ένα συγκεκριμένο άτομο. Ωστόσο, ο ορισμός των προσωπικών δεδομένων ποικίλλει ανάλογα με το πλαίσιο στο οποίο συλλέγονται. Δεδομένα που μπορεί να μην θεωρούνται προσωπικά από τη φύση τους μπορεί να γίνουν προσωπικά όταν συνδυάζονται με άλλες πηγές δεδομένων που τους επιτρέπουν να ταυτοποιήσει μοναδικά ένα άτομο. Μερικά παραδείγματα: [7]

- Το φυσικό όνομα ενός ατόμου χωρίς άλλα στοιχεία ταυτοποίησης μπορεί να μην είναι προσωπικά δεδομένα εκτός εάν υπάρχει ένα ιδιαίτερα ασυνήθιστο όνομα. Το «JohnSmith» είναι αρκετά κοινό ώστε το όνομα από μόνο του δεν πρέπει να ανησυχεί σε ορισμένες χώρες, αλλά το «BarackObama» είναι κάπως πιο αναγνωρίσιμο.
- Μια διεύθυνση email συχνά προσδιορίζει ένα άτομο με το όνομά του και μερικές φορές τον οργανισμό στον οποίο εργάζεται. Μια διεύθυνση email είναι μοναδική, σχετίζεται με ένα μεμονωμένο άτομο και ως εκ τούτου γενικά μπορεί να θεωρηθεί ότι είναι προσωπικά αναγνωρίσιμα δεδομένα.
- Οι αριθμοί κινητών τηλεφώνων είναι συνήθως προσωπικά αναγνωριστικά επειδή συνήθως χρησιμοποιούνται από ένα μόνο άτομο. Επιπλέον, εάν διατηρούνται άλλα σχετικά αναγνωριστικά, όπως συσκευή IMEI ή IMSI, αυτά συνήθως συσχετίζονται με ένα μόνο άτομο.

- Μια φυσική διεύθυνση που σχετίζεται με μεγάλο αριθμό ατόμων (όπως ένα μεγάλο μπλοκ γραφείου) δεν είναι από μόνη της προσωπικά δεδομένα, αλλά εάν προστεθούν άλλες πληροφορίες, όπως το όνομα ενός ατόμου που εργάζεται εκεί, είναι πολύ πιθανό να θιχτούν προσωπικά δεδομένα. Αντίθετα, μια φυσική διεύθυνση που προσδιορίζει ένα μεμονωμένο σπίτι ή διαμέρισμα διατρέχει μεγαλύτερο κίνδυνο να είναι προσωπικά στοιχεία [3].
- Μια μεμονωμένη θέση τοποθεσίας για μια συσκευή δεν αποτελεί γενικά πρόβλημα εάν δεν είναι γνωστές άλλες πληροφορίες χρήστη, αλλά εάν η τοποθέτηση είναι εξαιρετικά ακριβής και επιλύεται σε μια τοποθεσία που είναι γνωστό ότι σχετίζεται με ένα ή πολύ λίγα άτομα, ή η συσκευή αναφέρει συνεχώς τη θέση του GPS για μεγάλα χρονικά διαστήματα, είναι πιο πιθανό να ταυτοποιεί ένα άτομο ή ευαίσθητα δεδομένα.
- Ταχυδρομικοί κώδικες που καλύπτουν μεγάλες περιοχές μιας χώρας ή πολλές χιλιάδες άτομα δεν είναι άμεσα προσωπικά δεδομένα, αλλά σε αραιές περιοχές ή με πολύ ακριβείς κωδικούς όπως χρησιμοποιούνται στο Ηνωμένο Βασίλειο υπάρχει αυξανόμενος κίνδυνος να συσχετιστούν με μικρούς αριθμούς των ατόμων·
- Μια «διεύθυνση υλικού» κάποιου είδους, όπως η διεύθυνση «MAC» μιας συσκευής παρακολούθησης φυσικής κατάστασης Bluetooth χαμηλής ενέργειας από μόνη της δεν είναι χρησιμοποιήσιμα προσωπικά δεδομένα, αλλά εάν υπάρχει ευρεία συλλογή αυτής της διεύθυνσης σε μια πόλη, τότε μπορεί να είναι δυνατή η αναγνωρίση ενός ατόμου από τα πρότυπα του τρόπου ζωής του.
- Η ηλικία ενός ατόμου σε χρόνια είναι σχετικά χαμηλός κίνδυνος να είναι προσωπικά δεδομένα, αλλά η ημερομηνία γέννησης είναι πολύ υψηλότερος κίνδυνος, ιδιαίτερα εάν αυτό συνδέεται με άλλες πληροφορίες όπως το όνομά του.
- Η θερμοκρασία που αναφέρεται από έναν απλό αισθητήρα δεν είναι, όταν χρησιμοποιείται μεμονωμένα, προσωπικά δεδομένα, αλλά εάν είναι γνωστό ότι ο αισθητήρας είναι γνωστό ότι μετρά τη θερμοκρασία ενός συγκεκριμένου ατόμου, είναι προφανέστερα προσωπικά δεδομένα. Εναλλακτικά, εάν ένας αισθητήρας θερμοκρασίας είναι

εγκατεστημένος σε μια ακριβή τοποθεσία και είναι γνωστό ποιος βρίσκεται στη θέση του αισθητήρα, τα ίδια δεδομένα θερμοκρασίας με την πάροδο του χρόνου μπορεί να αποκαλύψουν εάν οι κάτοικοι είναι σπίτι [10].

- Οι διευθύνσεις IP (των συνδεδεμένων συσκευών) μπορεί να θεωρηθούν προσωπικά δεδομένα σε ορισμένες περιπτώσεις, ιδιαίτερα εάν υπάρχει σύνδεση άλλων δεδομένων, όπως το όνομα ή η διεύθυνση email του εγγεγραμμένου χρήστη με την ίδια διεύθυνση IP.
- Ο σεξουαλικός προσανατολισμός, η θρησκεία, η εθνικότητα, τα ιατρικά αρχεία και τα γενετικά δεδομένα, τα οικονομικά δεδομένα είναι μεταξύ των κατηγοριών δεδομένων που θεωρούνται ευαίσθητα προσωπικά δεδομένα σε ορισμένα νομικά πλαίσια και αυτό θα ισχύει γενικά στις περισσότερες περιπτώσεις.

Ως εκ τούτου, είναι σημαντικό να αναγνωριστεί ότι ένα μεμονωμένο στοιχείο δεδομένων γενικά επηρεάζεται από άλλα στοιχεία και ότι επομένως είναι η σύνδεση πολλών στοιχείων δεδομένων μεταξύ τους είτε σε μια σύνθετη εγγραφή είτε ως μέρος μιας ακολουθίας δεδομένων (χρονική ή γεωγραφική) που μπορεί να οδηγήσει σε στοιχεία προσωπικής ταυτοποίησης - ακόμη και από μεμονωμένα στοιχεία δεδομένων που ξεκίνησαν ως «χαμηλού κινδύνου». Για παράδειγμα, ενώ ένα όνομα γενικά δεν είναι προσωπικά δεδομένα, εάν προστεθεί η ηλικία του ατόμου και η πόλη κατοικίας, υπάρχει πολύ μεγαλύτερη πιθανότητα να ταυτοποιήσει ένα άτομο (π.χ. όνομα Elizabeth, επώνυμο Windsor, πόλη καταγωγής Λονδίνο, ηλικία 92). Αυτή η «σύνδεση δεδομένων» έχει άμεσο αντίκτυπο στις διαδικασίες που περιγράφονται παρακάτω, όπου ανόμοια σύνολα δεδομένων ενώνονται μεταξύ τους για σκοπούς ανάλυσης. Επομένως, συνιστάται να καταβληθούν προσπάθειες για τη μείωση των δεδομένων που συλλέγονται και αποθηκεύονται με βάση το τι απαιτείται για την υποστήριξη των τρεχουσών και προγραμματισμένων λύσεων. Ιδιαίτερη προσοχή πρέπει να δοθεί στα δεδομένα προσωπικού χαρακτήρα και θα πρέπει να χρησιμοποιούνται μέθοδοι όπως κρυπτογράφηση, ανωνυμοποίηση, ψευδωνυμοποίηση και συγκέντρωση για την προστασία των προσωπικών δεδομένων [14].

2.5. Σκέψεις σχετικά με μεγάλα δεδομένα και τεχνητή νοημοσύνη

Τα οφέλη των μεγάλων δεδομένων και της τεχνητής νοημοσύνης/μηχανικής νοημοσύνης επιτυγχάνονται γενικά καλύτερα όταν το σύνολο δεδομένων που αναλύεται είναι μεγαλύτερο (από άποψη αριθμού εγγραφών) και ευρύτερο (όσον αφορά τον αριθμό των διαφορετικών στοιχείων που είναι διαθέσιμα σε κάθε εγγραφή). Αυτό σε επιφανειακό επίπεδο έρχεται σε σύγκρουση με τον στόχο της ελαχιστοποίησης των δεδομένων που αντικατοπτρίζεται στις ευρέως αποδεκτές αρχές απορρήτου, αν και όταν οι πληροφορίες δεν είναι προσωπικά αναγνωρίσιμες, η σύγκρουση μπορεί να επιλυθεί. Στην πράξη, η ανάλυση μεγάλων δεδομένων και η τεχνητή νοημοσύνη είναι πιο πιθανό να χρησιμοποιηθούν για την εύρεση κοινών μοτίβων σε μεγάλα σύνολα δεδομένων μέσω στατιστικών τεχνικών που συγκεντρώνονται σε μεγάλο αριθμό χρηστών, συσκευών και δεδομένων. Επομένως, σε μεγάλο βαθμό, αυτές οι στατιστικές τεχνικές μπορούν να θεωρηθούν ως τεχνικές ενίσχυσης της ιδιωτικής ζωής όταν εφαρμόζονται σωστά. Για παράδειγμα: [15]

- Η τεχνητή νοημοσύνη θα μπορούσε να εκπαιδευτεί να προβράζει έναν συνδεδεμένο ηλεκτρικό βραστήρα σε ώρες της ημέρας που η ιστορική ανάλυση υποδηλώνει ότι είναι συνηθισμένες φορές που χρησιμοποιείται συσκευή IoT. Για μια απλή υπηρεσία όπως αυτή, δεν είναι πραγματικά απαραίτητο να υπάρχει γνώση σε ποιον ανήκει ο βραστήρας ή πού βρίσκεται, αλλά είναι απαραίτητο να μπορεί να δημιουργηθεί το προσαρμοσμένο πρόγραμμα για κάθε συγκεκριμένο βραστήρα.
- Τα δεδομένα ανάπτυξης των καλλιεργειών, τα δεδομένα καιρού, τα δεδομένα βροχοπτώσεων και τα δεδομένα εφαρμογής λιπασμάτων μπορούν να σχηματίσουν ένα εξαιρετικά χρήσιμο σύνολο δεδομένων για να βοηθήσουν τους αγρότες να ποτίζουν και να λιπάνουν σωστά τις καλλιέργειές τους. Η ανάλυση μεγάλων δεδομένων και η τεχνητή νοημοσύνη μπορούν να χρησιμοποιηθούν για τον καθορισμό των βέλτιστων στρατηγικών άρδευσης και λιπασμάτων για μεμονωμένα

χωράφια με βάση μακροπρόθεσμα ιστορικά δεδομένα. Δεδομένου ότι το μοντέλο πρόβλεψης έχει εκπαιδευτεί χρησιμοποιώντας μεγάλο αριθμό δειγμάτων από διαφορετικά αγροκτήματα και χωράφια σε ένα σύνολο ιστορικών δεδομένων, θα πρέπει να μαθαίνονται οι «κανόνες» αντί να χρειάζεται να διανεμηθούν τυχόν ακατέργαστα δεδομένα [16].

- Σημαντικές πρόοδοι στη διάγνωση καταστάσεων υγείας επιτυγχάνονται από την ανάλυση των δεδομένων που ελήφθησαν σε πολλούς ασθενείς και σε πολλές περιπτώσεις από μακροχρόνιες μελέτες που συχνά δημιουργήθηκαν προτού ήταν εύλογα προβλέψιμα ποια θέματα υγείας θα μπορούσαν να εντοπιστούν ή να αντιμετωπιστούν. Η τεχνητή νοημοσύνη μπορεί να αναλύσει τα δεδομένα από μεγάλους αριθμούς προσωπικών οθονών φυσικής κατάστασης IoT, και παρόλο που δεν χρειάζεται να γνωρίζουμε την προσωπική ταυτότητα των χρηστών για αυτόν τον σκοπό, υπάρχει όφελος από τη γνώση άλλων δεδομένων που μπορεί να είναι σχετικά, όπως η ηλικία και το φύλο. Η τεχνητή νοημοσύνη μπορεί επίσης να εκπαιδευτεί για τον εντοπισμό σπάνιων ή «απομακρυσμένων» καταστάσεων, όπως ένα ασυνήθιστο αλλά θεραπεύσιμο «βάδισμα» που θα μπορούσε να αναγνωριστεί από τα δεδομένα του βηματόμετρου. Αυτός ο τύπος δεδομένων θα μπορούσε επίσης να αποκλειστεί ως ακραία δεδομένα για τη διατήρηση του απορρήτου. Για μακροπρόθεσμες μελέτες τα δεδομένα μπορεί να καταγράφονται χρήσιμα για χρόνια ή και δεκαετίες, και για ορισμένες συνθήκες μπορεί να είναι χρήσιμα τα δεδομένα υψηλής συχνότητας, π.χ. δεδομένα καρδιακού παλμού ανά λεπτό.

Μερικά από τα τεχνικά ζητήματα είναι: [15]

- Συχνά τα δεδομένα από πολλά σύνολα δεδομένων θα πρέπει να συνδέονται μεταξύ τους και για να γίνει αυτό είναι απαραίτητο να υπάρχουν πεδία που είναι μοναδικά για έναν χρήστη ή μια συσκευή IoT. Η αναγνώριση του χρήστη ή της συσκευής IoT συνήθως δεν απαιτείται στο προκύπτον σύνολο δεδομένων, αλλά πρέπει να υπάρχει μια αξιόπιστη και αναπαραγώγιμη μέθοδος για τη σύνδεση των

δεδομένων και αυτή μπορεί να προέρχεται από μια συσκευή ή ένα προσωπικό αναγνωριστικό που χρησιμοποιεί μια μέθοδο όπως κατακερματισμός, με αποτέλεσμα ψευδωνυμοποιημένα δεδομένα·

- Η απόδοση των αναλυτικών στοιχείων μεγάλων δεδομένων και της τεχνητής νοημοσύνης βελτιώνονται μέσω της διαθεσιμότητας μεγάλων και κατάλληλα αντιπροσωπευτικών συνόλων δεδομένων που αριθμούν χιλιάδες ή εκατομμύρια εγγραφές. Για παράδειγμα, ένα σύστημα διαχείρισης κινητήρα που εκπαιδεύτηκε χρησιμοποιώντας δεδομένα που συλλέγονται αποκλειστικά σε χώρες με ζεστά κλίματα δεν θα έχει καλή απόδοση για οχήματα που χρησιμοποιούνται σε ψυχρά κλίματα.
- Η αναγνώριση των πραγματικών χρηστών ή οι υπερβολικά ακριβείς πληροφορίες σχετικά με τους χρήστες ή τις συσκευές σε ένα σύνολο δεδομένων είναι συχνά αντιπαραγωγικές στη χρήση της τεχνητής νοημοσύνης/μηχανικής εκμάθησης, καθώς αυτό μπορεί να οδηγήσει στην πιθανότητα «υπερπροσαρμογής» [11] – δηλαδή όταν η μηχανική μάθηση και οι αλγόριθμοι συντονίζονται υπερβολικά σε συγκεκριμένα δεδομένα εκπαίδευσης και αποδίδουν ελάχιστα όταν εκτίθενται σε νέα δεδομένα. Ενώ αναπτύσσεται το μοντέλο μηχανικής εκμάθησης, ίσως είναι χρήσιμο αρχικά να συμπεριληφθούν όσο το δυνατόν περισσότερα χαρακτηριστικά – τα οποία στη συνέχεια κλαδεύονται επιλεκτικά κατά τη διάρκεια της δοκιμής και της βελτιστοποίησης για να αποφευχθεί η υπερβολική προσαρμογή. Η εξαίρεση προσωπικών αναγνωριστικών ή αναγνωριστικών συσκευής από τα δεδομένα εκπαίδευσης θα βοηθήσει στην αποφυγή υπερβολικής τοποθέτησης.
- Ωστόσο, η δυνατότητα εκτέλεσης ενός εκπαιδευμένου μοντέλου μηχανικής εκμάθησης σε δεδομένα για έναν συγκεκριμένο χρήστη ή συσκευή προσφέρει τη δυνατότητα εντοπισμού μιας πολύ συγκεκριμένης ενέργειας που μπορεί να γίνει, π.χ., λέγοντας στον χρήστη ότι διατρέχει άμεσο κίνδυνο καρδιακής προσβολής ή κλείσιμο μιας πύλης λόγω της ανόδου της στάθμης του νερού. Ωστόσο, η ευρέως αποδεκτή αρχή της ελαχιστοποίησης δεδομένων βάσει σχεδιασμού απορρήτου μας λέει ότι μπορεί να υπάρχουν όρια στην αποθήκευση ιστορικών δεδομένων που συνδέονται με έναν χρήστη ή

μια συσκευή. Ως εκ τούτου, τα οφέλη που μπορούν να δημιουργήσουν τα μεγάλα δεδομένα και η τεχνητή νοημοσύνη δεν δικαιολογούν άδεια αορίστου χρόνου αποθήκευσης ιστορικών προσωπικών δεδομένων και πρέπει να εξισορροπούνται με τις απαιτήσεις απορρήτου [17].

- Το να ζητείται από τους χρήστες να αποδεχτούν ή να αποδεχτούν εκ νέου τους μακροχρόνιους και περίπλοκους Όρους και Προϋποθέσεις για νέες περιπτώσεις χρήσης μεγάλων δεδομένων και τεχνητής νοημοσύνης ή συλλογή δεδομένων δεν είναι πραγματικά μια λύση για ενημερωμένη συγκατάθεση και θα έχει ως αποτέλεσμα οι περισσότεροι χρήστες απλώς να κάνουν κλικ χωρίς να διαβάσουν πραγματικά τους Όρους και Προϋποθέσεις ή αλλιώς αποβολή από τη χρήση της υπηρεσίας. Ούτε είναι μια καλή λύση που βασίζεται στον εξαναγκασμό των χρηστών να επεξεργάζονται σελίδα-σελίδα αδειών με πλαίσιο ελέγχου ή στο να ζητούν τακτικά από τους χρήστες να ελέγχουν κάθε σελίδα των νέων στοιχείων ελέγχου απορρήτου.

  Αυτό επιτρέπει ακόμη και το ζήτημα της καταλληλότητας της συσκευής IoT για παρουσίαση πληροφοριών και επιτρέποντας εισόδους. Κατά την εφαρμογή λύσεων αποθήκευσης μεγάλων δεδομένων και τεχνητής νοημοσύνης, οι σχεδιαστές θα πρέπει να επικεντρωθούν στη μείωση του όγκου των προσωπικών δεδομένων που αποθηκεύονται ή υποβάλλονται σε επεξεργασία. Οποιαδήποτε δεδομένα θα πρέπει επίσης να αποθηκεύονται με ασφάλεια - ιδανικά χρησιμοποιώντας πλατφόρμες αποθήκευσης με κρυπτογράφηση υλικού. Η ανωνυμοποίηση, η ψευδωνυμοποίηση και η συγκέντρωση μπορούν επίσης να εφαρμοστούν για την προστασία των δεδομένων σε όλα τα στάδια αποθήκευσης και επεξεργασίας [10].

2.6. Όφελος από τη χρήση περιπτώσεων από «παγκόσμια» σύνολα δεδομένων

Ορισμένες χώρες έχουν εισαγάγει περιορισμούς στη διασυνοριακή ροή δεδομένων, που απορρέουν από ανησυχίες για την εθνική ασφάλεια,

ανησυχίες σχετικά με το απόρρητο δεδομένων ή την επιθυμία προστασίας των εγχώριων αγορών. Αυτοί οι περιορισμοί έχουν διάφορες μορφές, όπως η απαίτηση ρητής συναίνεσης από τους πολίτες ή προηγούμενη εξουσιοδότηση από τις αρχές προστασίας δεδομένων. Πιο απαγορευτικοί κανόνες εμποδίζουν τους οργανισμούς να μεταφέρουν προσωπικά δεδομένα ή μεταδεδομένα. Τα αποτελέσματα τέτοιων περιορισμών είναι πολλά. Για παράδειγμα, η απαίτηση από τους οργανισμούς να διατηρούν ένα πρόσθετο αντίγραφο των δεδομένων που δημιουργούνται από τις δραστηριότητές τους σε μια χώρα αυξάνει το κόστος παραγωγής φυσικών και ψηφιακών αγαθών και υπηρεσιών σε αυτήν την αγορά. Το κόστος αυξάνεται περαιτέρω όταν η ανάλυση και η επεξεργασία των δεδομένων πρέπει να διεξάγονται στο εσωτερικό εκτός από την αποθήκευση. Στο πλαίσιο του IoT, οι περιορισμοί στις διασυνοριακές ροές δεδομένων μπορεί να οδηγήσουν σε σημαντικά εμπόδια στην ανάπτυξη και την παράδοση προϊόντων και υπηρεσιών: [3]

- Τα άτομα που διαθέτουν προσωπικές συσκευές παρακολούθησης υγείας μπορούν επίσης να ταξιδεύουν για επαγγελματικούς λόγους ή αργίες και η συσκευή τους θα αναφέρει δεδομένα σε διακομιστές σε μια περιοχή, με επεξεργασία σε μια άλλη, και πρόσθετα δεδομένα πιθανώς αποθηκευμένα σε επιπλέον περιοχές. Οι υπηρεσίες απλώς δεν θα λειτουργούσαν εάν δεν υπήρχε η δυνατότητα επεξεργασίας δεδομένων που συλλέγονται σε αυτές τις πολλαπλές περιοχές.
- Πολλά προϊόντα μεταφέρονται διεθνώς και σε διάφορες ηπείρους και μια συσκευή IoT που αναπτύσσεται στην αλυσίδα εφοδιασμού για την παρακολούθηση προϊόντων και την περιβαλλοντική παρακολούθηση μπορεί επομένως να δημιουργήσει δεδομένα σε πολλές χώρες και περιοχές. Σε ορισμένα σημεία του ταξιδιού ενδέχεται να υπάρχουν προσωπικά δεδομένα, π.χ. το όνομα και η διεύθυνση ηλεκτρονικού ταχυδρομείου ενός τελωνειακού επιθεωρητή. Αυτή η παρακολούθηση και πάλι δεν θα μπορούσε να λειτουργήσει εάν δεν ήταν δυνατή η διασυνοριακή μεταφορά δεδομένων.
- Η επεξεργασία ενός συνόλου παγκόσμιων δεδομένων που αποκτήθηκε από μια παγκόσμια βάση πελατών χρησιμοποιώντας προσωπικές οθόνες υγείας IoT μπορεί να εντοπίσει καταστάσεις που επηρεάζουν

συγκεκριμένες εθνοτικές ομάδες που διαφορετικά δε θα μπορούσαν να εντοπιστούν μεμονωμένα σε μία μόνο χώρα ή περιοχή.

- Τα δεδομένα ατυχημάτων και σφαλμάτων που συλλέγονται και επεξεργάζονται από ένα κατασκευαστή αυτοκινήτων από τις παγκόσμιες πωλήσεις του είναι πιο χρήσιμα από ό,τι εάν οι πλατφόρμες συλλογής και ανάλυσης δεδομένων που βασίζονται σε τοπικό επίπεδο περιορίζονται στη συλλογή δεδομένων από τη συγκεκριμένη περιοχή. Για παράδειγμα, η ευαισθησία των φρένων αυτοκινήτων με κακή απόδοση σε εξαιρετικά ψυχρές συνθήκες μπορεί να γίνει κατανοητή και να αντιμετωπιστεί μόνο με τη συλλογή και επεξεργασία δεδομένων από τον Καναδά, τη Βόρεια Ευρώπη και τη Ρωσία. Τεχνικές όπως η ανωνυμοποίηση και η ψευδωνυμοποίηση μπορούν να εφαρμοστούν για την αποθήκευση και την επεξεργασία δεδομένων, έτσι ώστε να είναι δυνατή η παροχή λύσεων με χρήση παγκόσμιων συνόλων δεδομένων, ελαχιστοποιώντας παράλληλα τον κίνδυνο για τα προσωπικά δεδομένα του χρήστη. Τα δεδομένα θα πρέπει επίσης να μεταφέρονται μεταξύ περιοχών χρησιμοποιώντας ισχυρή κρυπτογράφηση και τα δεδομένα να αποθηκεύονται σε πλατφόρμες ιδανικά χρησιμοποιώντας κρυπτογράφηση σε επίπεδο υλικού [4].

3. Τεχνική υλοποίηση

Τα παρακάτω είναι ένα σύνολο τεχνικών και συστάσεων βέλτιστων πρακτικών που παρέχουν διάφορες προστασίες σχετικά με την απόκτηση, αποθήκευση και χρήση προσωπικών πληροφοριών. Αυτές είναι σχετικές τεχνικές λύσεις που μπορούν να χρησιμοποιηθούν για την υλοποίηση του «PrivacybyDesign». Πολλές από τις τεχνικές μπορούν να χρησιμοποιηθούν σε συνδυασμό για την υποστήριξη βέλτιστων πρακτικών για εφαρμογές μεγάλων δεδομένων και μηχανικής μάθησης [2].

3.1. Κρυπτογράφηση

Η κρυπτογράφηση δεδομένων, έτσι ώστε να μπορεί να αποκρυπτογραφηθεί μόνο από ένα άτομο ή ένα σύστημα με αντίστοιχο κλειδί αποκρυπτογράφησης, είναι μια σημαντική τεχνική για την προστασία των δεδομένων, ιδιαίτερα εάν περιέχουν προσωπικά ή εμπορικά εμπιστευτικά δεδομένα. Σε αντίθεση με το «κατακερματισμό» (παρακάτω), υπάρχει μια καθορισμένη διαδικασία για την ανάκτηση των αρχικών δεδομένων χρησιμοποιώντας ένα κλειδί αποκρυπτογράφησης. Ωστόσο, τα κρυπτογραφημένα δεδομένα δεν μπορούν να υποβληθούν σε επεξεργασία σε αναλυτικά στοιχεία ή μηχανική εκμάθηση χωρίς πρώτα να αποκρυπτογραφηθούν. Τόσο τα κλειδιά αποκρυπτογράφησης υψηλής αντοχής όσο και η ισχυρή ασφάλεια του κλειδιού αποκρυπτογράφησης είναι σημαντικά για τη διασφάλιση στο σχεδιασμό και τη λειτουργία του συστήματος. Οι βασικοί τομείς των βέλτιστων πρακτικών κρυπτογράφησης περιλαμβάνουν: [9]

- Χρήση «ασφαλούς στοιχείου» [12] σε συσκευές για την αποθήκευση ορισμένων ευαίσθητων π.χ. προσωπικά δεδομένα χρήστη ή σύνολα κλειδιών χρήστη. Αυτό είναι συχνά ένα αποκλειστικό στοιχείο υλικού στη συσκευή που είναι ανθεκτικό στην παραβίαση για συσκευές που

χρησιμοποιούν το δίκτυο κινητής τηλεφωνίας, η SIM είναι το καταλληλότερο μέρος για την αποθήκευση τέτοιων πληροφοριών.

- Κρυπτογράφηση τελικού σημείου, δηλαδή η κρυπτογράφηση πληροφοριών σε μια συσκευή ή σε αφαιρούμενα μέσα, έτσι ώστε τυχόν αποθηκευμένα δεδομένα να μην μπορούν να διαβαστούν από τρίτους, εκτός εάν έχουν το κλειδί αποκρυπτογράφησης.
- Κρυπτογράφηση μεταφοράς π.χ. χρησιμοποιώντας την «ασφαλή» φόρμα του πρωτοκόλλου «HTTP», «HTTPS», για την αποτροπή επιθέσεων από άνθρωπο στη μέση, για παράδειγμα κατά τη μεταφορά δεδομένων μεταξύ μιας συσκευής και συστημάτων cloud·
- Κρυπτογράφηση σε κατάσταση ηρεμίας – γενικά σε οποιοδήποτε σύστημα που αποθηκεύει δεδομένα από μια συσκευή IoT ή σχετικά με τον χρήστη ή για την αποθήκευση συνόλων δεδομένων πριν ή μετά την ανάλυση ή τη μηχανική εκμάθηση. Αυτό βοηθά επίσης στην προστασία των δεδομένων από την ανάκτηση από τρίτο μέρος σε περίπτωση μεταπώλησης πλεονάζοντος εξοπλισμού υπολογιστών, αν και αυτό δεν σημαίνει ότι θα είναι απαραίτητα ασφαλές στο μέλλον.
- Χρήση ισχυρών μέτρων κρυπτογράφησης π.χ. AES-256 για προστασία δεδομένων και επικοινωνιών από επίθεση.
- Τήρηση του απορρήτου των κλειδιών κρυπτογράφησης, ιδιαίτερα των ιδιωτικών κλειδιών, ειδικά εάν οι συνεργαζόμενοι οργανισμοί θα χρειαστούν πρόσβαση σε αυτά τα κλειδιά για οποιονδήποτε λόγο [17].

Η λύση δημιουργεί προστασία για τους συνοριακούς κόμβους μέσω της συνεργασίας μεταξύ των οριακών κόμβων και της πλατφόρμας cloud με δυνατότητες ασφάλειας. Αυτό επεκτείνει το φάσμα των μηχανισμών άμυνας που είναι δυνατοί για το IoT και άλλες εφαρμογές.

3.2. Κατακερματισμός

Υπάρχουν περιπτώσεις όπου είναι χρήσιμο να μπορούμε να κρατάμε «μυστικές» πληροφορίες με κωδικοποιημένο τρόπο χωρίς να χρειάζεται ποτέ να αποκωδικοποιήσουμε τις κωδικοποιημένες πληροφορίες. Ένα συνηθισμένο παράδειγμα είναι ο τρόπος με τον οποίο οι κωδικοί πρόσβασης αποθηκεύονται συχνά στα συστήματα. Για παράδειγμα, εάν ένας κωδικός πρόσβασης περιλαμβάνει τα ψηφία 65349811, το αποτέλεσμα της συνάρτησης κατακερματισμού[13] αυτού είναι 4A12751EE967410334811ECE84FB16FA και έτσι οποιοσδήποτε κωδικός πρόσβασης που έχει εισαχθεί και δεν έχει την ίδια τιμή κατακερματισμού θεωρείται λανθασμένος χωρίς να προσθέτει στους αυξημένους κινδύνους που εμπεριέχει ο κωδικός πρόσβασης [11] .

Οι τιμές κατακερματισμού είναι χρήσιμες επειδή είναι εξαιρετικά ανθεκτικές στην αντίστροφη μηχανική, ειδικά όταν χρησιμοποιούνται ισχυροί αλγόριθμοι κατακερματισμού όπως ο SHA-25614. Για παράδειγμα, εάν δημιουργηθεί μια τιμή κατακερματισμού για μια διεύθυνση email, ο εισβολέας θα πρέπει να επιχειρήσει τη διαδικασία κατακερματισμού για όλες τις γνωστές ή πιθανές διευθύνσεις ηλεκτρονικού ταχυδρομείου, προκειμένου να δει εάν ο χρήστης υπήρχε σε ένα συγκεκριμένο σύνολο δεδομένων και γενικά δεν αξίζει τον κόπο ή προσπάθεια [13].

Οι τιμές κατακερματισμού είναι επομένως χρήσιμες για την απόκρυψη προσωπικών πληροφοριών σε σύνολα δεδομένων, αλλά με τρόπο που είναι ντετερμινιστικό, δηλαδή η τιμή κατακερματισμού μιας δεδομένης εισόδου θα είναι η ίδια για έναν δεδομένο αλγόριθμο κατακερματισμού. Ως εκ τούτου, οι τιμές κατακερματισμού μπορούν να χρησιμοποιηθούν για να επιβεβαιώσουν κάτι που γνωρίζει ένα άλλο σύστημα ή ένας χρήστης και στη μάθηση μεγάλων δεδομένων/μηχανής παρέχουν ένα χρήσιμο εργαλείο για την απόκρυψη προσωπικών δεδομένων με τρόπο που προστατεύει το απόρρητο. Η συνάρτηση κατακερματισμού είναι επίσης μια χρήσιμη τεχνική για την εφαρμογή ψευδωνυμοποίησης.

Οι προτάσεις βέλτιστης πρακτικής για τον κατακερματισμό είναι ότι χρησιμοποιούνται ασφαλείς αλγόριθμοι κατακερματισμού (π.χ. SHA-256 ή νεότεροι), μαζί με salting για την αποτροπή επιθέσεων λεξικού σε κατακερματισμένα δεδομένα [15].

Για παράδειγμα, ο προσδιορισμός μιας ημερομηνίας γέννησης από μια κατακερματισμένη τιμή απαιτεί μόνο την επεξεργασία περίπου 43.800 κατακερματισμένων τιμών (120 έτη x 365 ημέρες του έτους [17]) – αυτή είναι μια μέθοδος γνωστή ως «επίθεση ωμής δύναμης». Οι τιμές έχουν δημιουργηθεί εκ των προτέρων. Αυτή είναι μια επίθεση λεξικού. Το salting απαιτεί την επιλογή της τυχαίας τιμής και την εφαρμογή στα δεδομένα εισόδου που κατακερματίζονται. Για παράδειγμα, στην περίπτωση ημερομηνίας γέννησης ένας τυχαία επιλεγμένος ακέραιος αριθμός θα μπορούσε να προστεθεί στο μέρος του έτους της ημερομηνίας, με διαφορετική τυχαία μετατόπιση (το «αλάτι») που δημιουργείται για κάθε εγγραφή στα δεδομένα εισόδου. Η χρήση διαφορετικής τιμής αλατιού για κάθε εγγραφή προστατεύει περαιτέρω από προσπάθειες διάρρηξης του κατακερματισμού χρησιμοποιώντας μεθόδους επίθεσης ωμής βίας και λεξικού. Στην «περίπτωση ημερομηνίας γέννησης» η προσθήκη μιας τιμής άλατος στο εύρος -500.000 έως 500.000 στο μέρος του έτους θα σήμαινε ότι ο αριθμός των τιμών κατακερματισμού που θα επιχειρηθεί θα ήταν 365.043.800 [11].

Αυτή η διαδικασία διασφαλίζει ότι για οποιοδήποτε στοιχείο δεδομένων που υπόκειται σε κατακερματισμό, υπάρχει επαναληψιμότητα στη δημιουργία του κατακερματισμού για δεδομένα εισόδου για σκοπούς όπως η ανάλυση ιστορικών τάσεων, αλλά υπάρχει προστασία των προσωπικών δεδομένων από βιώσιμες μεθόδους επίθεσης. Στην ιδανική περίπτωση, οι τιμές αλατιού θα πρέπει να αποθηκεύονται χωριστά από τις πραγματικές εγγραφές δεδομένων και να αποθηκεύονται σε έναν τόμο χρησιμοποιώντας ισχυρή κρυπτογράφηση.

3.3. Ανωνυμοποίηση

Πρόκειται για τη διαδικασία μετατροπής δεδομένων, συχνά προσωπικών δεδομένων ή δεδομένων που ενδέχεται να γίνουν προσωπικά δεδομένα κατά τη μεταγενέστερη επεξεργασία, σε μορφή που είναι μη αναστρέψιμη και αδύνατο να συσχετιστεί με τα αρχικά προσωπικά δεδομένα. Το Future of Privacy Forum ορίζει τέτοια «αποπροσδιορισμένα δεδομένα» με τους εξής τρόπους: [8]

i. δεδομένα από τα οποία έχουν αφαιρεθεί οριστικά τα άμεσα και έμμεσα αναγνωριστικά· ή
ii. δεδομένα που έχουν διαταραχθεί σε βαθμό που ο κίνδυνος επαναπροσδιορισμού είναι μικρός, δεδομένου του πλαισίου του συνόλου δεδομένων· ή
iii. Τα δεδομένα που έχει επιβεβαιώσει ένας εμπειρογνώμονας ενέχουν πολύ μικρό κίνδυνο οι πληροφορίες να μπορούν να χρησιμοποιηθούν από έναν αναμενόμενο παραλήπτη για την ταυτοποίηση ενός ατόμου. Ο σκοπός της ανωνυμοποίησης είναι να καταστεί δυνατή η αποθήκευση ή η κοινοποίηση δεδομένων σε άλλα συστήματα ή οργανισμούς για αποθήκευση ή επεξεργασία, προστατεύοντας παράλληλα το απόρρητο των χρηστών ή άλλο απόρρητο (π.χ. εμπορικό απόρρητο).

Υπάρχουν διάφορες μέθοδοι ανωνυμοποίησης: [6]

- Απλώς αφαιρώντας ένα συγκεκριμένο πεδίο από τα δεδομένα προέλευσης, π.χ. Εάν είναι γνωστό ότι η τοποθεσία της συσκευής θα είναι πάντα άσχετη με τα αναλυτικά στοιχεία, τότε μπορεί απλώς να αφαιρεθεί -είτε αφαιρώντας ολόκληρο το πεδίο είτε αντικαθιστώντας από μια μηδενική τιμή σε όλες τις εγγραφές.
- Αφαίρεση ολόκληρων αρχείων που περιέχουν προσωπικά δεδομένα, π.χ. οποιοδήποτε αρχείο όπου η τοποθεσία βρίσκεται κοντά σε θρησκευτικό χώρο λατρείας ·
- Αντικατάσταση ενός συγκεκριμένου πεδίου με μια τυπική υποκατάστατη τιμή που σημαίνει ότι έχει ανωνυμοποιηθεί. Π.χ. «*» για

ένα πεδίο κειμένου, -999 για ένα αριθμητικό πεδίο (με την προϋπόθεση ότι αυτό το πεδίο δεν αναμένεται ποτέ να έχει τέτοια τιμή).

- Αντικατάσταση ενός συγκεκριμένου πεδίου με μια εντελώς τυχαία αντικατάσταση, π.χ. μια τιμή τυχαίας συμβολοσειράς κειμένου ή μια τιμή τυχαίας αριθμητικής συμβολοσειράς [5].
- Η «απόκρυψη χαρακτήρων» που εφαρμόζεται σε τιμές κειμένου μπορεί να εφαρμοστεί έτσι ώστε τα ευαίσθητα δεδομένα να καλύπτονται επαρκώς ώστε να μην είναι πλέον προσωπικά. Για παράδειγμα, το όνομα «AliceSmith» θα μπορούσε να αλλάξει σε «A* S*». Αυτό χρησιμοποιείται συνήθως με αριθμούς πιστωτικών καρτών όπου ο αριθμός αποθηκεύεται αποκρύπτοντας έναν αριθμό ψηφίων π.χ. «1234 **** **** 8901» ή «**** **** **** 8901». Αυτή η τεχνική μπορεί να είναι χρήσιμη εάν υπάρχει απαίτηση για την εμφάνιση του ανωνυμοποιημένου πεδίου στον χρήστη σε μεταγενέστερο σημείο, αλλά υπάρχει επίσης ο κίνδυνος να χρησιμοποιηθεί ένα καλυμμένο πεδίο για συσχέτιση [2].

3.4. Ψευδωνυμοποίηση

Για την προστασία του ατομικού απορρήτου, μια βασική λύση που μπορεί να χρησιμοποιηθεί για τη μετατροπή των προσωπικών δεδομένων σε δεδομένα «αποαναγνωρισμένα» είναι η ψευδωνυμοποίηση. Σε αυτό, ένα αρχικό κομμάτι προσωπικών δεδομένων μετατρέπεται, συνήθως με έναν πίνακα ή συνάρτηση μονόδρομης αντιστοίχισης, σε μια τιμή που είναι μοναδική για τα αρχικά δεδομένα αλλά μη αναστρέψιμη σε τρίτο μέρος. Η ψευδωνυμοποίηση γενικά επιτυγχάνεται με την αντικατάσταση ενός εναλλακτικού αναγνωριστικού που δημιουργείται από το σύστημα για να αντικαταστήσει ένα ή περισσότερα πεδία προσωπικών δεδομένων του χρήστη [4].

Εάν το ψευδώνυμο αναγνωριστικό είναι μέρος ενός συνόλου δεδομένων που προκύπτει ή κοινοποιείται σε εξωτερικά μέρη, θα πρέπει να είναι αδύνατο για τους παραλήπτες να αναγνωρίσουν προσωπικές πληροφορίες σχετικά με τον χρήστη από αυτό το σύνολο δεδομένων, ενώ άλλα αναλυτικά στοιχεία

παραμένουν δυνατά, π.χ. ανάλυση τάσεων για ένα σύστημα IoT. Όπως και με τις βέλτιστες πρακτικές για την ανωνυμοποίηση, συνιστάται τα «προσωπικά δεδομένα» να ονομάζονται ψευδώνυμα το συντομότερο δυνατό κατά την αποθήκευση και την επεξεργασία δεδομένων.

Η ψευδωνυμοποίηση έχει ορισμένα πλεονεκτήματα σε σχέση με μεθόδους όπως η ανωνυμοποίηση και η συγκέντρωση: [1]

- Είναι δυνατή η διατήρηση ενός συνόλου δεδομένων ανά χρήστη ή ανά συσκευή, απλώς χωρίς το σύνολο δεδομένων να αποκαλύπτει την ταυτότητα του χρήστη ή της συσκευής.
- Είναι δυνατός ο συσχετισμός πολλαπλών συνόλων δεδομένων μαζί εάν υπάρχει ένα κοινό ψευδώνυμο αναγνωριστικό που εμφανίζεται σε αυτά τα σύνολα δεδομένων, χωρίς να χρειάζονται πρωτότυπα προσωπικά αναγνωριστικά.
- Εάν τα αναλυτικά στοιχεία ή η τεχνητή νοημοσύνη εντοπίσουν ένα ζήτημα για αυτόν τον χρήστη ή τη συσκευή, εξακολουθεί να είναι δυνατή η υλοποίηση μιας ενέργειας ή απάντησης που μπορεί να παρασχεθεί σε αυτόν τον χρήστη ή τη συσκευή - εάν υπάρχει υποστήριξη από μια οντότητα υψηλότερου επιπέδου που γνωρίζει την αντιστοίχιση μεταξύ αρχικών αναγνωριστικών και ψευδώνυμων αναγνωριστικών·
- Είναι δυνατό να διατηρηθεί ένα συνεχές νήμα που καλύπτει τόσο τα ιστορικά όσο και τα τρέχοντα δεδομένα, έτσι ώστε οι αναλύσεις να μπορούν να διεξαχθούν σε ολόκληρο τον τομέα του χρόνου, π.χ. ανάλυση τάσεων.
- Μπορούν να παραδοθούν πιο εξατομικευμένες υπηρεσίες IoT, παρόλο που η ταυτότητα του χρήστη παραμένει καλύτερα προστατευμένη.

Οι προτάσεις βέλτιστης πρακτικής για την ψευδωνυμοποίηση περιλαμβάνουν τη χρήση ενός ή περισσότερων από τα ακόλουθα: [17]

- Χρήση μηχανισμών όπως ένα μεγάλο, τυχαίο UUID (Universal Unique Identifier) για την αντικατάσταση ενός ή περισσότερων προσωπικών

δεδομένων. Συνιστώνται τουλάχιστον τυχαία UUID τύπου 4 128-bit, π.χ. «cb3eb75f-cbbc-414c-a146-624518ed7537» (δεκαεξαδικό)

- Χρήση ισχυρών κρυπτογραφικών λειτουργιών κατακερματισμού21 (π.χ. SHA-256 ή καλύτερη) με «αλάτισμα» για μονόδρομη κωδικοποίηση προσωπικών δεδομένων με τρόπο που πρακτικά δεν μπορεί να αντιστραφεί. [11]
- Δημιουργία διαφορετικών ψευδώνυμων αναγνωριστικών για προσωπικά δεδομένα για κάθε ξεχωριστή εφαρμογή, συμπεριλαμβανομένων των αναλυτικών στοιχείων ή της διαδικασίας μηχανικής μάθησης, που χρησιμοποιεί τα δεδομένα.
- Δημιουργία διαφορετικών χρονικά δεσμευμένων ψευδώνυμων αναγνωριστικών για προσωπικά δεδομένα σε καθορισμένες χρονικές περιόδους π.χ. Κάθε μέρα δημιουργείται ένα νέο ψευδώνυμο αναγνωριστικό για οποιοδήποτε δεδομένο πεδίο προσωπικών δεδομένων και κάθε δεδομένη εφαρμογή, έτσι ώστε οποιαδήποτε μεταγενέστερη εφαρμογή να μην μπορεί να παρακολουθεί τις μακροπρόθεσμες ενέργειες των χρηστών.

Σε περίπτωση που μια υπηρεσία μπορεί να χρειαστεί να χρησιμοποιήσει το αποτέλεσμα των αναλυτικών στοιχείων για να δημιουργήσει ένα αποτέλεσμα ή μια ειδοποίηση για έναν χρήστη ή μια συσκευή που αναγνωρίζεται χρησιμοποιώντας ένα ψευδώνυμο αναγνωριστικό, μια πιο ασφαλής υλοποίηση μπορεί να παραχθεί με διαχωρισμό της πλατφόρμας που αποθηκεύει τα προσωπικά δεδομένα από την πλατφόρμα που διεξάγει τις αναλύσεις: [12]

1. Η πλατφόρμα που διατηρεί τα δεδομένα χρήστη ή συσκευής (το σύστημα «κύριος δεδομένων») διατηρεί τη γνώση του τρόπου «αντιστροφής χάρτη» από ένα ψευδώνυμο αναγνωριστικό στον αρχικό χρήστη ή συσκευή.
2. Η πλατφόρμα αναλυτικών στοιχείων μεταβιβάζει το ψευδώνυμο αναγνωριστικό στο κύριο σύστημα δεδομένων μαζί με οποιοδήποτε

αποτέλεσμα ή ενέργεια, και το κύριο σύστημα δεδομένων διαχειρίζεται στη συνέχεια τη δρομολόγηση στον χρήστη.

3. Αυτή η αποθήκευση δεδομένων χαρτογράφησης πρέπει να είναι εξαιρετικά ασφαλής, να έχει την πιο περιορισμένη δυνατή πρόσβαση τόσο στο διοικητικό προσωπικό όσο και στα συνδεδεμένα συστήματα και να χρησιμοποιεί ισχυρή κρυπτογράφηση για τυχόν αποθηκευμένα δεδομένα.
4. Εάν δεν υπάρχει ανάγκη να προσφερθούν τέτοιες υπηρεσίες, δεν υπάρχει ανάγκη να αναπτυχθεί ένας τέτοιος τρόπος τρόπος αντίστροφης χαρτογράφησης.
5. Εάν μια εφαρμογή ή υπηρεσία αποσυρθεί, το κατάστημα χαρτογράφησης θα πρέπει επίσης να απαλειφθεί από τα δεδομένα που σχετίζονται με αυτήν την εφαρμογή ή υπηρεσία.
6. Εάν ένας χρήστης τερματίσει την υπηρεσία του, το κατάστημα χαρτογράφησης θα πρέπει επίσης να καθαριστεί από τα δεδομένα που σχετίζονται με αυτόν τον χρήστη.

4. Case studies από τη διεθνή βιβλιογραφία

4.1. Κατακερματισμός - Salting

Αυτή είναι μια διαδικασία όπου τυχαία δεδομένα προστίθενται σε άλλα δεδομένα πριν εφαρμοστούν διεργασίες όπως ο κρυπτογραφικός κατακερματισμός. Είναι μια τεχνική που χρησιμοποιείται ευρέως στην αποθήκευση τιμών κωδικού πρόσβασης και βοηθά στην αποφυγή συσχέτισης κωδικών πρόσβασης που έχουν σπάσει σε ένα σύστημα που χρησιμοποιείται για την αντιστροφή κωδικών πρόσβασης που είναι αποθηκευμένοι σε άλλο σύστημα. Το συγκεκριμένο πρόβλημα που αντιμετωπίζεται είναι ότι οι συναρτήσεις κατακερματισμού, εξ ανάγκης και σχεδίασης, παράγουν τα ίδια byte εξόδου για τα ίδια byte εισόδου. Επομένως, εάν ένα σύστημα χρησιμοποιεί μια συγκεκριμένη συνάρτηση κατακερματισμού χωρίς να προσθέσει άλλα δεδομένα (το «αλάτι») στην τιμή εισόδου, είναι δυνατή η σύγκριση των τιμών εξόδου κατακερματισμού για ισότητα. Υπάρχει μια συγκρίσιμη ευπάθεια για τις διαδικασίες κρυπτογράφησης που αντιμετωπίζεται χρησιμοποιώντας ένα «Διάνυσμα εκκίνησης» [11].

Το "Salting" έχει ευρύτερη εφαρμογή στο απόρρητο επειδή, για παράδειγμα, εάν μια συγκεκριμένη συνάρτηση κατακερματισμού εφαρμόζεται σε μια διεύθυνση ηλεκτρονικού ταχυδρομείου που θα έχει μια ξεχωριστή υπογραφή κατακερματισμού που θα ήταν ίδια σε οποιοδήποτε άλλο σύστημα ή σε οποιοδήποτε άλλο σύνολο δεδομένων που χρησιμοποιεί τον ίδιο κατακερματισμό λειτουργούν στην ίδια διεύθυνση email. Θα ήταν δυνατό να συσχετιστούν τέτοια δεδομένα αναζητώντας τις ίδιες τιμές κατακερματισμού. Επιπλέον, για ορισμένα δεδομένα με σχετικά περιορισμένες δυνατότητες είναι εφικτό να πραγματοποιηθεί μια επίθεση, όπως μια επίθεση λεξικού[16], χρησιμοποιώντας ένα σύνολο από τις πιο πιθανές τιμές για ένα κατακερματισμένο πεδίο.

4.2. Κρυπτογράφηση - China Mobile SAFE link

Το IoT εισάγει ένα διευρυμένο φάσμα κινδύνων ασφαλείας σε ένα ευρύ φάσμα τεχνολογιών, συμπεριλαμβανομένου του φυσικού επιπέδου (αισθητήρες και τερματικά), του επιπέδου μεταφοράς (κινητά ή άλλα δίκτυα επικοινωνιών) και του επιπέδου εφαρμογής (παραδοσιακή ασφάλεια Διαδικτύου). Καθώς ένα αυξανόμενο φάσμα βιομηχανιών, συμπεριλαμβανομένης της μεταποίησης, της γεωργίας, της ιατρικής περίθαλψης και των μεταφορών και της εφοδιαστικής χρησιμοποιούν το IoT, ο αντίκτυπος των κινδύνων για την ασφάλεια γίνεται ολοένα και μεγαλύτερος. Για την αντιμετώπιση αυτών των κινδύνων, η China Mobile έχει εισαγάγει μια αποκλειστική λύση «SAFE link» που συνδυάζει μια «πλατφόρμα cloud με δυνατότητες ασφαλείας», πρόσβαση σε λειτουργίες ασφάλειας της συσκευής και μια εφαρμογή διαχείρισης ασφάλειας. Αυτό ενσωματώνει φιλτράρισμα κυκλοφορίας, μετάδοση κρυπτογράφησης δεδομένων (υποστηρίζει κβαντική κρυπτογράφηση), κατανεμημένη κοινή χρήση και απομακρυσμένη αποθήκευση και άλλες δυνατότητες ασφάλειας. Η ασφάλεια αντιμετωπίζεται μέσω δυνατοτήτων όπως η ασφαλής πρόσβαση στο δίκτυο, η παρακολούθηση της ασφάλειας και η ασφαλής μετάδοση. Οι συσκευές IoT γίνονται ουσιαστικά αξιόπιστοι κόμβοι, το όριο αξιόπιστου δικτύου ωθείται στην πλευρά της κοντινής πηγής δεδομένων και επομένως κοντά σε χρήστες και συσκευές [14].

4.2.1. Λύσεις

Η λύση China Mobile επιτρέπει στη συσκευή να γίνει ένα ασφαλές σημείο πρόσβασης δικτύου που συνδέεται στο Διαδίκτυο ή σε ένα Intranet. Ο συνδυασμός δυνατοτήτων σε επίπεδο συσκευής και πλατφόρμας cloud ενισχύει την ασφάλεια του αιτήματος πρόσβασης στο δίκτυο του χρήστη, συμπεριλαμβανομένης της παρακολούθησης κακόβουλων διευθύνσεων URL,

ιστοτόπων ηλεκτρονικού ψαρέματος κ.λπ., και επιτρέπει ειδοποιήσεις χρηστών. Η λειτουργικότητα περιλαμβάνει [16]

✚ Πλατφόρμα cloud με δυνατότητα ασφάλειας: Αυτή η πλατφόρμα ενσωματώνει μια εθνική έγκυρη βάση δεδομένων ασφαλείας, σύστημα «Αρχή πιστοποιητικών», μια βάση δεδομένων ασφαλείας China Mobile, βάση δεδομένων ασφαλείας τρίτων και άλλους πόρους ασφαλείας. Η πλατφόρμα διατηρεί επίσης ενιαία τεχνικά πρότυπα ασφαλείας και προδιαγραφές προϊόντων. Οι τεχνολογίες επεξεργασίας και ανάλυσης μεγάλων δεδομένων επιτυγχάνουν υπηρεσίες ασφαλείας χωρίς την επίγνωση των χρηστών. Η πλατφόρμα cloud έχει σχεδιαστεί για να αντιμετωπίζει τους κινδύνους ασφαλείας παρέχοντας μια αξιόπιστη υπολογιστική πλατφόρμα, η οποία δημιουργεί από μόνη της μια συστηματική πλατφόρμα για μέτρα προστασίας της ασφάλειας. Υπάρχουν τρία επίπεδα διαχείρισης συστήματος, διαχείρισης ασφάλειας και διαχείρισης ελέγχου που υποστηρίζονται από την πλατφόρμα. Η πλατφόρμα διαθέτει επίσης προληπτικά αμυντικά μέτρα για να εγγυηθεί την ασφάλεια της ίδιας της πλατφόρμας.

✚ Εξοπλισμός πρόσβασης ασφαλείας: Αυτός ο εξειδικευμένος εξοπλισμός συνεργάζεται με τη συσκευή και την πλατφόρμα cloud με δυνατότητα ασφάλειας για τη διατήρηση της ασφαλούς λειτουργίας της συσκευής τελικού χρήστη. Σε συνεργασία με τους πόρους ασφαλείας που διαχειρίζεται η πλατφόρμα cloud με δυνατότητες ασφαλείας, παρακολουθεί τις διευθύνσεις URL που ζητούνται από τη συσκευή και χειρίζεται κατάλληλα αιτήματα για μη συμβατές διευθύνσεις URL. Ο εξοπλισμός πρόσβασης ασφαλείας υποστηρίζει επίσης εμπορικό VPN/κβαντική κρυπτογράφηση VPN που επιτρέπει ένα ασφαλές κανάλι μετάδοσης δεδομένων, υποστηρίζει κρυπτογράφηση αρχείων και κοινή χρήση αποθήκευσης, παρέχει στους χρήστες και τις εφαρμογές περιβάλλον αποθήκευσης και μετάδοσης δεδομένων υψηλής ασφάλειας και άνεσης και επεκτείνει την αξιόπιστη ασφάλεια κινητής τηλεφωνίας όριο για ασφαλείς συσκευές πρόσβασης. [12]

✚ Εφαρμογή διαχείρισης ασφάλειας: Η εφαρμογή διαχείρισης ασφάλειας μπορεί να εκτελέσει διάφορες λειτουργίες, όπως διαχείριση συσκευής, διαμόρφωση δικτύου, προβολή δυνατοτήτων ασφαλείας και διαχείριση απομακρυσμένης αποθήκευσης δεδομένων στη συσκευή πρόσβασης σύμφωνα με τις δυνατότητες της συσκευής ασφαλούς πρόσβασης. Η επικοινωνία με την πλατφόρμα cloud και τη συσκευή ασφαλούς πρόσβασης βασίζεται στην κρυπτογράφηση μεταφοράς που διασφαλίζει την ασφάλεια της μετάδοσης δεδομένων μεταξύ του τερματικού, της πλατφόρμας cloud με δυνατότητα ασφαλείας και της συσκευής πρόσβασης.

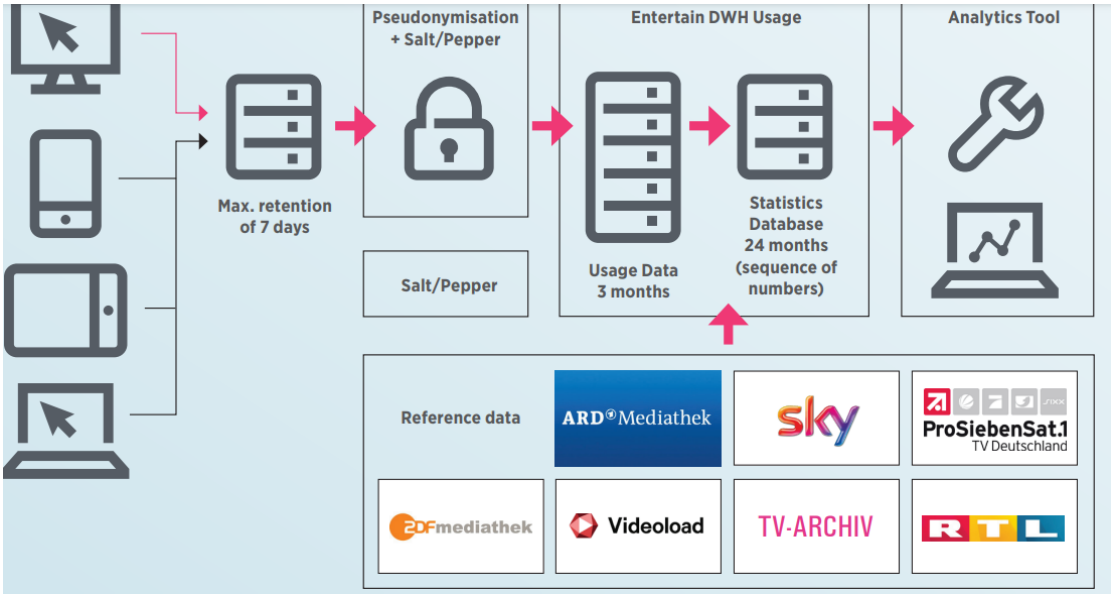
4.2.2. Συμπεράσματα

Η λύση δημιουργεί προστασία για τους συνοριακούς κόμβους μέσω της συνεργασίας μεταξύ των οριακών κόμβων και της πλατφόρμας cloud με δυνατότητες ασφάλειας. Αυτό επεκτείνει το φάσμα των μηχανισμών άμυνας που είναι δυνατοί για το IoT και άλλες εφαρμογές.

4.3. Ψευδωνυμοποίηση

Η Deutsche Telekom AG. εμπορεύεται πρόσβαση σε τηλεοπτικά προγράμματα και ταινίες μέσω του Διαδικτύου με το όνομα προϊόντος Entertain TV. Η εταιρεία παρέχει στους πελάτες έναν αποκωδικοποιητή για τη χρήση του προϊόντος. Στατιστικά στοιχεία σχετικά με τις συνήθειες των τηλεθεατών διατηρούνται για διάφορους σκοπούς, συμπεριλαμβανομένων των υποχρεώσεων έναντι των ραδιοτηλεοπτικών φορέων. [4]

Το παρακάτω διάγραμμα δείχνει τη ροή δεδομένων μέσω των επιμέρους συστημάτων, από τον αποκωδικοποιητή έως τα στατιστικά στοιχεία. Η ενημέρωση και οι σχετικές πληροφορίες διαμόρφωσης επισημαίνονται ως μη λειτουργικές.



Εικόνα 4. Ροή δεδομένων μέσω των επιμέρους συστημάτων, από τον αποκωδικοποιητή έως τα στατιστικά στοιχεία. Πηγή [13]

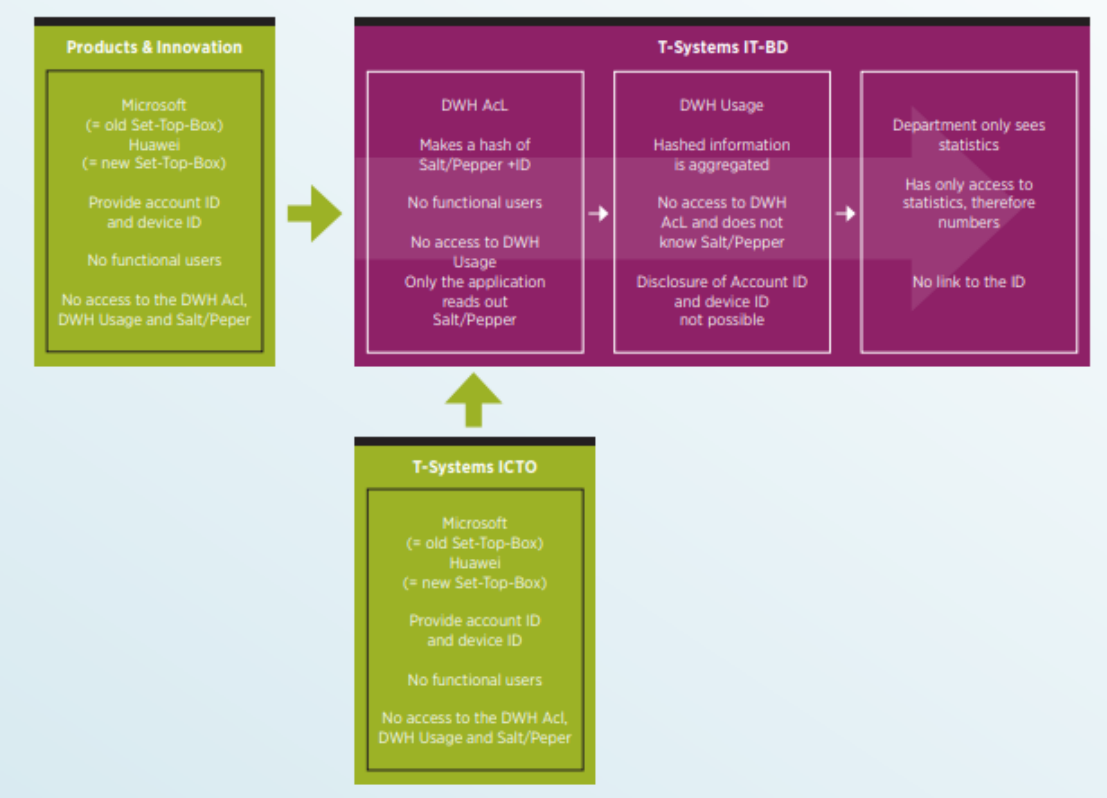
4.3.1. Παραγωγή δεδομένων

Χρήση του αποκωδικοποιητή, π.χ. όταν ο καταναλωτής χρησιμοποιεί το τηλεχειριστήριο του συστήματος, δημιουργεί μια σειρά συμβάντων ανάλογα με το κουμπί που πατήθηκε και το σχετικό πλαίσιο. Αυτά τα συμβάντα αποτελούν τη βάση διαφόρων αναλύσεων, οι οποίες τεκμηριώνουν δραστηριότητες όπως ενεργοποίηση/απενεργοποίηση, αλλαγές καναλιών, πληροφορίες σχετικά με τα προγράμματα που παρακολουθούν, πληροφορίες σχετικά με δραστηριότητες εγγραφής χρηστών ή πληροφορίες σχετικά με χρήστες που παρακολουθούν εγγεγραμμένα προγράμματα. Τα αντίστοιχα σύνολα δεδομένων συμβάντων περιέχουν μια σειρά πληροφοριών, π.χ. σχετικά με τον αποκωδικοποιητή (αναγνωριστικό συσκευής), το αναγνωριστικό λογαριασμού του πελάτη, την ημερομηνία/ώρα και άλλα συγκεκριμένα θέματα. [7]

4.3.2. Ψευδωνυμοποίηση

Το αναγνωριστικό λογαριασμού είναι ψευδώνυμο για τον πελάτη και το αναγνωριστικό συσκευής είναι ψευδώνυμο για το σχετικό αποκωδικοποιητή. Τα σύνολα δεδομένων συμβάντων που απαιτούνται για σκοπούς ανάλυσης δεν περιέχουν χαρακτηριστικά που να παρουσιάζουν προσωπικά δεδομένα (άμεσου είδους). Με χωριστή διαχείριση από οργανωτική άποψη, υπάρχουν πίνακες κατανομής που επιτρέπουν τη σύνδεση των ψευδωνύμων (αναγνωριστικά λογαριασμού και συσκευών) με πελάτες ή αποκωδικοποιητές. Η πρόσβαση σε αυτούς τους πίνακες θα καθιστούσε δυνατή την τελική ιχνηλάτηση των αναγνωριστικών της συσκευής και των λογαριασμών για τον εντοπισμό στους πελάτες. Η ανίχνευση είναι μερικές φορές απαραίτητη, δηλαδή για υπηρεσίες τιμολόγησης σύμφωνα με τη σύμβαση. Ωστόσο, καθώς

κανένα μέρος δεν θέλει να ανιχνεύσει λεπτομέρειες σε "απλά δεδομένα" με σκοπό τη δημιουργία στατιστικών στοιχείων, τα αναγνωριστικά λογαριασμού και συσκευών υπόκεινται σε πρόσθετη ψευδωνυμοποίηση πριν από την επεξεργασία. Σε αυτήν την περίπτωση, η ψευδωνυμοποίηση λαμβάνει χώρα εντός της μονάδας Επιπέδου Απόκτησης Αποθήκης Δεδομένων (DWH ACL) και η επεξεργασία (δημιουργία στατιστικών στοιχείων) πραγματοποιείται στη χρήση της αποθήκης ημερομηνίας (Χρήση DWH), μια ξεχωριστή μονάδα (όπως φαίνεται παρακάτω). [8]



Εικόνα 5. Οργανωτικό concept. Πηγή: [11]

4.3.3. Δημιουργία στατιστικών στοιχείων

Όλα τα χαρακτηριστικά που μπορούν να ανιχνευθούν στον τελικό πελάτη ονομάζονται ψευδώνυμα χρησιμοποιώντας ένα αναγνωριστικό λογαριασμού ή

συσκευής. Οι πληρωμές είναι δυνατές καθώς χρησιμοποιούνται συνδεόμενα ψευδώνυμα. Για παράδειγμα, αυτό σημαίνει ότι μπορείτε να απαντήσετε σε μια ερώτηση σχετικά με το πόσα νοικοκυριά ή αποκωδικοποιητές παρακολούθησαν ένα συγκεκριμένο κανάλι σε μια συγκεκριμένη στιγμή. Τα ανώνυμα στατιστικά στοιχεία δεν περιέχουν πλέον αναγνωριστικά λογαριασμών και συσκευών ή τα ψευδώνυμα που δημιουργούνται, γεγονός που εμποδίζει τον εντοπισμό των στατιστικών στοιχείων στα κατακερματισμένα αναγνωριστικά. Η Deutsche Telekom πρέπει να εκπληρώσει ορισμένες υποχρεώσεις έναντι των ραδιοτηλεοπτικών φορέων, επομένως μεταφέρει μόνο ανώνυμα στατιστικά στοιχεία σχετικά με τη συμπεριφορά των χρηστών των τηλεθεατών, π.χ. μερίδιο αγοράς χρησιμοποιώντας σχετικά στοιχεία. [5]

4.3.4. Εξαίρεση

Οι ειδοποιήσεις προστασίας δεδομένων ενημερώνουν κάθε πελάτη του Entertain TV ότι τα δεδομένα συλλέγονται για στατιστικούς σκοπούς. Η Deutsche Telekom χρησιμοποιεί e-mail και αναδυόμενα παράθυρα στην ίδια την Entertain TV για να ενημερώσει τους πελάτες σχετικά πριν παρουσιάσει αυτήν τη λύση ανάλυσης. Κάθε πελάτης έχει τη δυνατότητα να αντιταχθεί (εξαιρεθεί) για τη συλλογή και ανάλυση των ψευδώνυμων δεδομένων χρήσης του. Ο πελάτης μπορεί να χρησιμοποιήσει τον αποκωδικοποιητή του για να πραγματοποιήσει αυτήν την εξαίρεση. Προηγουμένως, αυτό συνεπαγόταν την εισαγωγή ενός αριθμού PIN, αλλά με νεότερους αποκωδικοποιητές δεν απαιτείται εισαγωγή PIN. Με την εξαίρεση, τα δεδομένα χρήσης του πελάτη δεν χρησιμοποιούνται ούτε για ψευδώνυμο προφίλ χρήσης ούτε για ανώνυμα στατιστικά στοιχεία. Οι πελάτες μπορούν επίσης να χρησιμοποιήσουν συμβατικά κανάλια επικοινωνίας για να ενημερώσουν την Deutsche Telekom ότι θέλουν να ασκήσουν το δικαίωμα εξαίρεσης. [10]

Συμπεράσματα

Τα μεγάλα δεδομένα προσφέρουν πολλά πλεονεκτήματα και πολλά υποσχόμενες δυνατότητες για καινοτομία σε διάφορους τομείς, αλλά παρουσιάζουν επίσης πολλά ζητήματα και προκλήσεις. Συγκεκριμένα, κάθε φάση του κύκλου ζωής των μεγάλων δεδομένων έχει ζητήματα ασφάλειας και αξιοπιστίας δεδομένων και μπορεί να αποτελέσει απειλή για την παραβίαση του απορρήτου μέσω διαφόρων αναλύσεων μεγάλων δεδομένων.

Από την παρούσα εργασία προκύπτει ότι η Κρυπτογράφηση ως λύση δημιουργεί προστασία για τους συνοριακούς κόμβους μέσω της συνεργασίας μεταξύ των οριακών κόμβων και της πλατφόρμας cloud με δυνατότητες ασφάλειας. Αυτό επεκτείνει το φάσμα των μηχανισμών άμυνας που είναι δυνατοί για το IoT και άλλες εφαρμογές. Ωστόσο, επειδή όλες οι φάσεις στον κύκλο ζωής των μεγάλων δεδομένων είναι αλληλένδετες και έχουν μεγάλη επιρροή, θα πρέπει να αντιμετωπιστούν ζητήματα ασφάλειας και απορρήτου σε όλες τις φάσεις. Σε μελλοντικές εργασίες, θα πρέπει να αποσαφηνιστεί η προτεινόμενη ταξινόμηση ασφάλειας και να σχεδιαστεί μια αρχιτεκτονική ασφάλειας σύμφωνα με τον κύκλο ζωής των μεγάλων δεδομένων.

Η υποκείμενη διαδικασία ψευδωνυμοποίησης σημαίνει ότι τα στατιστικά στοιχεία δημιουργούνται χρησιμοποιώντας ψευδώνυμα που μπορούν να συνδεθούν αλλά δεν μπορούν να αποκαλυφθούν για τη χρήση Data warehouse, τη μονάδα που εμπλέκεται στην επεξεργασία. Τα ψευδώνυμα δημιουργούνται χρησιμοποιώντας μια ντετερμινιστική κρυπτογραφική διαδικασία κατακερματισμού (SHA-512) και ένα στοιχείο ομοιόμορφης ψευδωνυμοποίησης αλάτι/πιπέρι. Η ικανότητα σύνδεσης ψευδωνύμου καθιερώνεται επειδή οι ντετερμινιστικές διεργασίες μεταφέρουν τα ίδια απλά κείμενα σε ίδιες τιμές αποτελέσματος (ψευδώνυμα) όταν το αλάτι/πιπέρι είναι πανομοιότυπα. Το μήκος εξόδου του SHA-512 σημαίνει ότι ο κίνδυνος συγκρούσεων είναι αμελητέος (< 1070). Καθώς το DWH Usage δεν έχει πρόσβαση στην ψευδωνυμοποίηση αλάτι/πιπέρι, δεν μπορεί πρακτικά να εντοπίσει τα ψευδώνυμα πίσω σε πραγματικούς ανθρώπους και έτσι να αποκαλύψει απλά κείμενα. Τελικά, η διαδικασία ψευδωνυμοποίησης που

χρησιμοποιείται σημαίνει ότι κανένας υπάλληλος του Ομίλου Deutsche Telekom δεν μπορεί να δει, να αναλύσει ή να μεταφέρει σε οποιονδήποτε άλλο πληροφορίες σχετικά με τη συμπεριφορά χρήσης συγκεκριμένου πελάτη.

Βιβλιογραφία

1. Adams, M. (2017). Big data and individual privacy in the age of the internet of things. *Technology Innovation Management Review*, *7*(4).
2. Fabiano, N. (2017, June). Internet of Things and blockchain: Legal issues and privacy. The challenge for a privacy standard. In *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (pp. 727-734). IEEE.
3. Azmoodeh, A., & Dehghantanha, A. (2020). Big data and privacy: Challenges and opportunities. In *Handbook of Big Data Privacy* (pp. 1-5). Springer, Cham.
4. Sollins, K. R. (2019). IoT big data security and privacy versus innovation. *IEEE Internet of Things Journal*, *6*(2), 1628-1635.
5. Hajli, N., Shirazi, F., Tajvidi, M., & Huda, N. (2021). Towards an understanding of privacy management architecture in big data: an experimental research. *British Journal of Management*, *32*(2), 548-565.
6. Bertino, E., & Ferrari, E. (2018). Big data security and privacy. In *A comprehensive guide through the Italian database research over the last 25 years* (pp. 425-439). Springer, Cham.
7. Lundqvist, B. (2018). Big data, open data, privacy regulations, intellectual property and competition law in an internet-of-things world: The issue of accessing data. In *Personal Data in Competition, Consumer Protection and Intellectual Property Law* (pp. 191-214). Springer, Berlin, Heidelberg.
8. Oh, H., Park, S., Lee, G. M., Heo, H., & Choi, J. K. (2019). Personal data trading scheme for data brokers in IoT data marketplaces. *IEEE Access*, *7*, 40120-40132.
9. Becerril, A. A. (2018). The value of our personal data in the Big Data and the Internet of all Things Era.
10. Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. *Technovation*, *98*, 102173.
11. Xie, W., & Karan, K. (2019). Consumers' privacy concern and privacy protection on social network sites in the era of big data: Empirical evidence from college students. *Journal of Interactive Advertising*, *19*(3), 187-201.

12. Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of cybersecurity*, *4*(1), tyy001.
13. Abi Sen, A. A., & Basahel, A. M. (2019, March). A comparative study between security and privacy. In *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1282-1286). IEEE.
14. Casanovas, P., De Koker, L., Mendelson, D., & Watts, D. (2017). Regulation of Big Data: Perspectives on strategy, policy, law and privacy. *Health and Technology*, *7*(4), 335-349.
15. Keller, B., Eling, M., Schmeiser, H., Christen, M., & Loi, M. (2018). *Big data and insurance: implications for innovation, competition and privacy*. Geneva Association-International Association for the Study of Insurance Economics.
16. Line, N. D., Dogru, T., El-Manstrly, D., Buoye, A., Malthouse, E., & Kandampully, J. (2020). Control, use and ownership of big data: A reciprocal view of customer big data value in the hospitality and tourism industry. *Tourism Management*, *80*, 104106.
17. Poyner, I. K., & Sherratt, R. S. (2018, March). Privacy and security of consumer IoT devices for the pervasive monitoring of vulnerable people. In *Living in the Internet of Things: Cybersecurity of the IoT-2018* (pp. 1-5). IET.