Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών *Ασφάλεια Υπολογιστών και Δικτύων*

Μεταπτυχιακή Διατριβή

Προηγμένες Τεχνικές Κρυπτογράφησης για Ψευδωνυμοποίηση Δεδομένων

Γεώργιος Κερμεζής

Επιβλέπων Καθηγητής
Κωνσταντίνος Λιμνιώτης

Μάιος 2020

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών *Ασφάλεια Υπολογιστών και Δικτύων*

Μεταπτυχιακή Διατριβή

Προηγμένες Τεχνικές Κρυπτογράφησης για Ψευδωνυμοποίηση Δεδομένων

Γεώργιος Κερμεζής

Επιβλέπων Καθηγητής
Κωνσταντίνος Λιμνιώτης

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε
προς μερική εκπλήρωση των απαιτήσεων για απόκτηση

μεταπτυχιακού τίτλου σπουδών
στην Ασφάλεια Υπολογιστών και Δικτύων

από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών
του Ανοικτού Πανεπιστημίου Κύπρου

Μάιος 2020

Περίληψη

Ένα από τα πιο σημαντικά ζητήματα της σύγχρονης εποχής είναι η προστασία των προσωπικών δεδομένων, ειδικά με την αύξηση της επιρροής και την επικράτηση της τεχνολογίας της πληροφορίας σε κάθε έκφανση του ανθρώπινου βίου. Αυτό το θέμα έχει αποτελέσει επίκεντρο πολλών και καίριων διενέξεων, από τη μία για την προστασία θεμελιωδών δικαιωμάτων του ανθρώπου και από την άλλη για τα πλεονεκτήματα σε πολλούς τομείς που παρέχει η συλλογή και επεξεργασία δεδομένων. Λόγω της σπουδαιότητας της σημασίας προσωπικών δεδομένων, στην Ευρωπαϊκή Ένωση υπάρχει στέρεο θεσμικό πλαίσιο, το οποίο συναντάται ιδίως στο Γενικό Κανονισμό για την Προστασία Δεδομένων. Ο Κανονισμός αυτός θέτει σύνολο προϋποθέσεων νόμιμης επεξεργασίας, ενώ σε διάφορα σημεία του προκρίνει τη χρήση μεθόδων ψευδωνυμοποίησης, ως εχέγγυο για την επεξεργασία δεδομένων για συγκεκριμένους σκοπούς. Ως άμεση απόρροια αυτού, η έρευνα τα τελευταία χρονιά έχει στραφεί σε εύρεση κατάλληλων τεχνικών ψευδωνυμοποίησης, ιδίως αξιοποιώντας κρυπτογραφικά εργαλεία.

Η παρούσα μεταπτυχιακή διατριβή εστιάζει κατ' αρχάς στη μελέτη και καταγραφή υπαρχουσών τεχνικών ψευδωνυμοποίησης. Στη συνέχεια επικεντρωθήκαμε στις πιο προηγμένες από αυτές και εντοπίσαμε ότι, αν και κλασικές συναρτήσεις κατακερματισμού εφαρμόζονται πολλές φορές ως εργαλεία ψευδωνυμοποίησης, αυτό δεν ισχύει για τα λεγόμενα δέντρα Merkle (μία ήδη γνωστή κρυπτογραφική τεχνική, εφαρμοζόμενη σε διάφορες άλλες περιπτώσεις), τα οποία αποτελούν κατά κάποιο τρόπο γενικεύσεις της συνάρτησης κατακερματισμού. Στη συνέχεια σχηματοποιήσαμε θεωρητικά τον τρόπο με τον οποίο θα μπορούσε να επιτευχθεί μία τέτοια τεχνική. Έπειτα, προσδιορίσαμε τα χαρακτηριστικά της και τις λεπτομέρειες του τρόπου λειτουργίας της. Έτσι, καταφέραμε να καταλήξουμε στο συμπέρασμα ότι η πρότασή μας υλοποιεί μία καινούρια μέθοδο ψευδωνυμοποίησης, με ιδιότητες τέτοιες οι οποίες δεν επιτυγχάνονται με κλασικές τεχνικές ψευδωνυμοποίησης (όπως, μεταξύ άλλων, ανταλλαγή δεδομένων μεταξύ δύο φορέων, μόνο μετά την συναίνεση του προσώπου στο οποίο αφορούν, χωρίς αποκάλυψη καμίας πλεονάζουσας πληροφορίας). Πέραν της πλήρους θεωρητικής ανάλυσης της νέας τεχνικής, αναπτύχθηκε αντίστοιχος αλγόριθμος που μεταφράστηκε σε μορφή κώδικα και εκτελέστηκε με σκοπό την πραγματοποίηση μετρήσεων, για να παρατηρηθεί η συμπεριφορά του σε πραγματικό περιβάλλον. Από τα αποτελέσματα αυτά συμπεράναμε ότι μπορεί, υπό συγκεκριμένες προϋποθέσεις, να λειτουργήσει, «ανοίγοντας» ταυτόχρονα νέο μονοπάτι για μελλοντική έρευνα.

Summary

One of the most important issues of modern times is the protection of personal data, especially with the rise of influence and the dominance of information technology in every aspect of human life. This issue has been the focus of many and key controversies, on the one hand for the protection of fundamental human rights and on the other hand for the benefits in many areas provided by data collection and processing. Due to the importance of personal data, in the European Union there is a solid legal framework, which is found in particular in the General Data Protection Regulation. This Regulation sets out all the requirements for legitimate processing, while in various respects it prefers the use of pseudonym methods, as a safeguard for the processing of data for specific purposes. As a direct result of this, research in recent years has focused on finding appropriate pseudonymization techniques, especially using cryptographic tools.

The present postgraduate Thesis focuses first on the study of all existing pseudonymization techniques. We then focus on the most advanced of them and found that, although classic hash functions are often used as pseudonymisation tools, this does not apply to the so-called Merkle trees (an already well-known cryptographic technique, used in several other cases), which are, in a way, generalizations of the hash function. Then we theoretically shaped the way in which such a technique could be achieved. Next, we identified its features and details of how it works. Thus, we have come to the conclusion that our proposal implements a new method of pseudonymization, with properties that are not achieved by classical pseudonymization techniques (such as, among other things, data exchange between two entities, only with the consent of the person concerned, without disclosing any surplus of information). In addition to the complete theoretical analysis of the new technique, a corresponding algorithm was developed that was translated into programming code and executed in order to perform measurements, to observe its behavior in a real environment. From these results we concluded that it can, under certain conditions, work, while "opening" a new path for future research.

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμότερες ευχαριστίες μου στον επιβλέποντα καθηγητή μου, για την παρούσα μεταπτυχιακή διατριβή, κ. Κωνσταντίνο Λιμνιώτη για την πολύτιμη υποστήριξή του, την εξαιρετική καθοδήγησή του, την υπομονή του και για την εμπιστοσύνη που μου επέδειξε αναλαμβάνοντας το ρόλο του επιβλέποντα.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

Στη σύγχρονη εποχή της πληροφορίας, όπως έχει χαρακτηριστεί, τα δεδομένα αποτελούν έναν πολύτιμο πόρο σχεδόν σε κάθε τομέα της ανθρώπινης δραστηριότητας, με αποτέλεσμα οι ραγδαίες τεχνολογικές εξελίξεις και η παγκοσμιοποίηση να έχουν δημιουργήσει σημαντική αύξηση στη συλλογή, επεξεργασία και διακίνηση δεδομένων.

1.1 Προστασία προσωπικών δεδομένων

Τμήμα αυτού του συνολικού όγκου πληροφορίας αποτελούν τα προσωπικά και ευαίσθητα προσωπικά δεδομένα, που απαιτούν την προστασία για τη διασφάλιση στοιχειωδών ανθρώπινων δικαιωμάτων.

Η σημασία των στοιχειωδών ανθρώπινων δικαιωμάτων και η ανάγκη προστασίας τους αποτελούν τη βάση για την ύπαρξη της ελευθερίας μας και της δυνατότητάς μας να καταφέρουμε να αναπτυχθούμε και να ευημερήσουμε και ένα από αυτά είναι η ιδιωτικότητα, η δυνατότητα, δηλαδή, το κάθε άτομο να έχει τον έλεγχο των πληροφοριών που το αφορούν και την επιλογή αν και με ποιον θα τις μοιραστεί. Η ανθρώπινη, όμως, δραστηριότητα σε οικονομικό, κοινωνικό, πολιτικό, επιστημονικό και τεχνολογικό επίπεδο αποτελεί επίσης ένα βασικό άξονα που συμβάλει στην βελτίωση της ποιότητας ζωής της ανθρωπότητας και κατ' επέκταση του κάθε ατόμου ξεχωριστά.

Η τεχνολογική ανάπτυξη, ευρισκόμενη σε μία συνεχή σχέση ανατροφοδότησης με τους υπόλοιπους τομείς, τους έχει επηρεάσει και συνεχίζει να τους επηρεάζει με καταλυτικό τρόπο, ώστε να θεωρείται αναπόσπαστο κομμάτι τους. Ειδικά ο τομέας των πληροφοριών έχει

γνωρίσει τέτοια ανάπτυξη και διάχυση προς κάθε έκφανση της ανθρώπινης δραστηριότητας που είναι εξαιρετικά δύσκολο να τον απομονώσουμε.

Αυτή η πραγματικότητα έχει οδηγήσει στη ζήτηση και επεξεργασία όλο και περισσότερων δεδομένων δημιουργώντας μία κατάσταση σύγκρουσης "αντίρροπων" δικαιωμάτων και ένα βασικό ερώτημα. Κατά πόσο είναι δυνατός ο συμβιβασμός ανάμεσα στην απαίτηση αξιοποίησης των δεδομένων και στο δικαίωμα προστασίας της ιδιωτικότητας. Πόσα και ποια δεδομένα είναι χρήσιμο να συλλεγούν και να επεξεργαστούν από ποιον και με τι τρόπο, χωρίς να μειώνεται η ανθρώπινη ελευθερία του ατόμου και χωρίς να περιορίζεται η δυνατότητα της ανάπτυξης της ανθρωπότητας.

Έτσι, δημιουργείται η ανάγκη διαχωρισμού της πληροφορίας σε δύο κύριες κατηγορίες. Σε αυτή που είναι ελεύθερη προς συλλογή και επεξεργασία και σε αυτή που πρέπει να προστατευθεί και να παραμείνει ιδιωτική. Ακόμα ωστόσο και για την πρώτη περίπτωση, το γεγονός ότι μία πληροφορία μπορεί να υποστεί επεξεργασία δεν συνεπάγεται ότι αυτό μπορεί να γίνεται ανεξέλεγκτα, χωρίς την τήρηση προϋποθέσεων. Από τα ανωτέρω προκύπτει νέο ερώτημα για το αν μπορεί η ιδιωτικότητα να περιορίσει την εξέλιξη και την ανάπτυξη του συνόλου. Καθώς και το τι μπορεί να γίνει στην περίπτωση που το άτομο είναι διατεθειμένο να γνωστοποιήσει τα δεδομένα του και ποια από αυτά θα πρέπει να παραμείνουν προστατευμένα ασχέτως των επιθυμιών του.

Το επόμενο ερώτημα που προκύπτει από όλη αυτή την αναζήτηση είναι το τι πλαίσιο απαιτείται για τη ρύθμιση όλων αυτών των απαιτήσεων και αν είναι δυνατό αυτό να βρει τη «χρυσή τομή». Αν είναι απαραίτητη η νομική προστασία για τη ρύθμιση αυτού του προβλήματος. Για να προσδιοριστούν ευκρινώς ποια από τα δεδομένα χρήζουν προστασίας, ποιος έχει πρόσβαση σε αυτά και κάτω από ποιες προϋποθέσεις είναι δυνατό να διαμοιραστούν.

Για τον προσδιορισμό αυτών των συνθηκών θα πρέπει να απαντηθεί το ερώτημα αν είναι δυνατό η διαχείριση των δεδομένων με τέτοιο τρόπο, ώστε να μην αποτελεί απειλή για την ελευθερία του ατόμου και του βασικού δικαιώματός του στην ιδιωτικότητα και την προστασία των προστασία των προσωπικών και ευαίσθητων προσωπικών δεδομένων.

Η Ευρωπαϊκή Ένωση αναγνωρίζει τη σημασία της προστασίας προσωπικών δεδομένων ως θεμελιώδες ανθρώπινο δικαίωμα. Συγκεκριμένα, το Συμβούλιο της Ευρώπης, το οποίο συστάθηκε μετά τον Β΄ Παγκόσμιο Πόλεμο με σκοπό να συνενώσει τα κράτη της Ευρώπης για

την προαγωγή του κράτους δικαίου και των ανθρωπίνων δικαιωμάτων, εξέδωσε το 1950 την ΕΣΔΑ (Ευρωπαϊκή Σύμβαση Δικαιωμάτων του Ανθρώπου), η οποία τέθηκε σε ισχύ το 1953. Το δικαίωμα στην προστασία των δεδομένων προσωπικού χαρακτήρα (ή, αλλιώς, προσωπικών δεδομένων) περιλαμβάνεται στα δικαιώματα που προστατεύονται βάσει του άρθρου 8 της ΕΣΔΑ (Union 2019:20). Έκτοτε ειδικότερη νομοθεσία έχει θεσπιστεί και συνεχίζει να θεσπίζεται για ειδικότερα θέματα. Σήμερα, το κύριο νομικό πλαίσιο στην Ευρωπαϊκή Ένωση είναι ο Γενικός Κανονισμός Προστασίας Δεδομένων (General Data Protection Regulation - GDPR), ο οποίος είναι σε εφαρμογή από τις 25 Μαϊου 2018 και ρυθμίζει προϋποθέσεις νόμιμης επεξεργασίας προσωπικών δεδομένων, αναγνωρίζοντας υποχρεώσεις σε όσους επεξεργάζονται προσωπικά δεδομένα αλλά και δικαιώματα στα πρόσωπα των οποίων τα δεδομένα υφίστανται επεξεργασία. Όπως αναφέρεται στη Σκέψη 4 του εν λόγω Κανονισμού, και σε σχέση με όσα ήδη έχουν αναφερθεί προηγούμενα, «*η επεξεργασία των δεδομένων προσωπικού χαρακτήρα θα πρέπει να προορίζεται να εξυπηρετεί τον άνθρωπο. Το δικαίωμα στην προστασία των δεδομένων προσωπικού χαρακτήρα δεν είναι απόλυτο δικαίωμα· πρέπει να εκτιμάται σε σχέση με τη λειτουργία του στην κοινωνία και να σταθμίζεται με άλλα θεμελιώδη δικαιώματα, σύμφωνα με την αρχή της αναλογικότητας*».

1.2 Η έννοια της ψευδωνυμοποίησης

Στο Γενικό Κανονισμό γίνεται συχνή αναφορά, σε διάφορα σημεία, στη χρήση ψευδωνύμων, με τρόπο ώστε να μην είναι αναγνωρίσιμη η ταυτότητα του προσώπου και έτσι να μην μπορούν να συνδεθούν τα δεδομένα που το αφορούν με αυτό. Αν και η έννοια της ψευδωνυμοποίησης ήταν ήδη χρησιμοποιούμενη σε τεχνικά κείμενα και πρότυπα, είναι η πρώτη φορά που εισάγεται σε αμιγώς νομικό κείμενο, ως εχέγγυο για την προστασία των δεδομένων. Για παράδειγμα, η χρήση της ψευδωνυμοποίησης μπορεί να «κρύψει» τις ταυτότητες φυσικών προσώπων, με τρόπο τέτοιο ώστε να μπορεί να γίνει άλλη επεξεργασία επ' αυτών (π.χ. στατιστική ανάλυση, επιστημονική έρευνα κτλ.). Εξάλλου, η ψευδωνυμοποίηση είναι απαραίτητη σε περιπτώσεις κατά τις οποίες αυτός ο οποίος επεξεργάζεται δεδομένα δεν χρειάζεται να γνωρίζει την ταυτότητα του προσώπου του οποίου τα δεδομένα επεξεργάζεται (αφού μία βασική προϋπόθεση νόμιμης επεξεργασίας προσωπικών δεδομένων είναι να γίνεται επεξεργασία των απολύτως απαραίτητων προσωπικών δεδομένων και όχι περισσότερων).

Με την εισαγωγή της ψευδωνυμοποίησης ως μέσο επίτευξης νόμιμης επεξεργασίας προσωπικών δεδομένων, γεννώνται αυτομάτως ερωτήματα ως προς τον τρόπο με τον οποίο

είναι δυνατό να επιτευχθεί αποτελεσματικά. Με άλλα λόγια, αν υπάρχουν τεχνικές που να μπορούν να διασφαλίσουν την ταυτότητα και άρα την ιδιωτικότητα του ατόμου τροποποιώντας ό,τι αποτελεί αναγνωριστικό στοιχείο, χωρίς, όμως, να καθιστά τα δεδομένα άχρηστα προς περαιτέρω επεξεργασία και θεμιτούς σκοπούς. Άρα η αναζήτηση ύπαρξης μεθόδων που να ικανοποιούν αυτόν τον «συμβιβασμό» (εύρεση χρυσής τομής) είναι πολύ μεγάλης σημασίας.

1.3 Η έννοια της κρυπτογραφίας

Η κρυπτογραφία είναι μία μέθοδος που χρησιμοποιείται από τον άνθρωπο για τη διασφάλιση της εμπιστευτικότητας, με την τροποποίηση της πληροφορίας με τέτοιο τρόπο έτσι ώστε να είναι αναγνώσιμη μόνο από τον επιθυμητό παραλήπτη της, εδώ και χιλιετίες. Αποτελεί τον κατ' εξοχήν τρόπο προστασίας της εμπιστευτικότητας της πληροφορίας τόσο κατά τη μετάδοσή της όσο και κατά την αποθήκευσή της. Ωστόσο, οι κρυπτογραφικοί μετασχηματισμοί μπορούν πλέον να αντιμετωπίζουν πολύ πιο σύνθετα προβλήματα (ψηφιακές υπογραφές, πρωτόκολλα μηδενικής γνώσης κτλ.) αφού υπάρχουν διαφόρων ειδών κρυπτογραφικά σχήματα με ειδικές, καλά θεμελιωμένες, μαθηματικές ιδιότητες.

Επισημαίνεται ότι και ο Γενικός Κανονισμός Προστασίας Δεδομένων αναγνωρίζει την αξία της κρυπτογραφίας ως μέσο ενίσχυσης της ασφάλειας των δεδομένων. Για παράδειγμα, σε περίπτωση «κλοπής» προσωπικών δεδομένων, ο Κανονισμός αναφέρει ότι αν τα δεδομένα είναι κρυπτογραφημένα με τρόπο τέτοιο ώστε να μην μπορεί ο υποκλοπέας να τα επανακτήσει στην αρχική τους μορφή (δηλαδή έχει διασφαλιστεί η εμπιστευτικότητα), τότε αυτός ο οποίος υπέστη την κλοπή δεν έχει κάποιες υποχρεώσεις σε σχέση με αυτή ούτε και κάποιες συνέπειες (αφού θεωρείται ότι, λόγω της κρυπτογράφησης, δεν θίγονται τα δικαιώματα των προσώπων).

Από τα ανωτέρω, συνδυάζοντας τις – διαφορετικές – έννοιες της κρυπτογραφίας και της ψευδωνυμοποίησης, είναι εύλογο το ερώτημα αν η πρώτη θα μπορούσε να χρησιμοποιηθεί για την κάλυψη των αναγνωριστικών δεδομένων του ατόμου (δηλαδή για ψευδωνυμοποίηση), επιτρέποντας την ανάγνωσή τους μόνο από όσους είναι επιθυμητό και επιτρέπεται. Η σύζευξη αυτή της κρυπτογραφίας με την ψευδωνυμοποίηση δεν είναι καινούρια: η ερώτηση αν έχει χρησιμοποιηθεί ή μπορεί να χρησιμοποιηθεί η κρυπτογραφία με αυτόν το τρόπο έχει καταφατική απάντηση και ήδη χρησιμοποιείται και κατ' αυτόν τον τρόπο. Ωστόσο, υπάρχουν περιπτώσεις όπου η επιθυμητή ψευδωνυμοποίηση είναι πολύ δύσκολο να επιτευχθεί με τις συμβατικές κρυπτογραφικές μεθόδους. Τρέχουσες ερευνητικές προσπάθειες εστιάζουν στην

αξιοποίηση προηγμένων κρυπτογραφικών τεχνικών, προκειμένου να δοθούν απαντήσεις σε περιπτώσεις όπου απαιτούνται ψευδωνυμοποιήσεις με πολύ ιδιαίτερα χαρακτηριστικά.

Για τη συνεισφορά στην απάντηση αυτών των ερωτημάτων εκπονήθηκε η παρούσα μεταπτυχιακή διατριβή.

1.4 Ερευνητικά ερωτήματα

Η παρούσα διατριβή μελετά την έννοια της ψευδωνυμοποίησης και εστιάζει στη μελέτη προηγμένων κρυπτογραφικών τεχνικών ως προς τη δυνατότητά τους να παρέχουν απαντήσεις σε περιπτώσεις όπου απαιτείται ψευδωνυμοποίηση με συγκεκριμένα χαρακτηριστικά, τέτοια που να καθιστούν δύσκολη (μη προφανή) την εύρεση μίας αποτελεσματικής τεχνικής. Συγκεκριμένα, μελετήθηκε η δυνατότητα εύρεσης τεχνικής ψευδωνυμοποίησης, η οποία να έχει τα ακόλουθα χαρακτηριστικά:

1) Τα ψευδώνυμα παράγονται από τους ίδιους τους χρήστες (και όχι από τους οργανισμούς/φορείς οι οποίοι θα επεξεργαστούν τα δεδομένα)

2) Οι οργανισμοί/φορείς είναι σε θέση να επιβεβαιώσουν ότι το ψευδώνυμο που δέχονται αντιστοιχεί πράγματι σε «έγκυρο» χρήστη

3) Ο κάθε χρήστης παράγει διαφορετικά ψευδώνυμα για διαφορετικό οργανισμό

4) Εάν ο χρήστης επιθυμεί να «αποδείξει» σε έναν οργανισμό Α ότι έχει ένα συγκεκριμένο ψευδώνυμο σε έναν άλλο οργανισμό Β, να μπορεί να το κάνει χωρίς να αποκαλύψει καμία πρόσθετη πληροφορία.

Το παραπάνω σχήμα είναι πολύ σημαντικό, γιατί φέρει μία πολύ σημαντική, από τη σκοπιά της προστασίας δεδομένων, ιδιότητα: το ψευδώνυμο παράγεται από τον ίδιο το χρήστη.

Ένα σχήμα ψευδωνυμοποίησης με τα παραπάνω χαρακτηριστικά θα μπορούσε να είναι εξαιρετικά χρήσιμο, για παράδειγμα, σε περίπτωση που ένας χρήστης δύο διαφορετικών ηλεκτρονικών υπηρεσιών θέλει να ζητήσει από τη μία να στείλει κάποια εκ των δεδομένων του στην άλλη (χωρίς να αποκαλυφθούν πρόσθετες πληροφορίες). Εναλλακτικά, θα μπορούσε να ισχύσει σε περίπτωση δύο δημόσιων φορέων, όπου ο ένας φορέας δεν πρέπει να μάθει (γιατί θα

προσέκρουε στη νομοθεσία προσωπικών δεδομένων) το κυρίως αναγνωριστικό που έχει ο χρήστης στον άλλο φορέα, αλλά προκύπτει η ανάγκη (π.χ. με τη συγκατάθεση του χρήστη) να μάθει κάποιες πληροφορίες για το χρήστη αυτόν.

1.5 Μεθοδολογία

Για την επίτευξη των στόχων που τέθηκαν από τα ερευνητικά ερωτήματα πραγματοποιήθηκε αρχικά βιβλιογραφική επισκόπηση για τον προσδιορισμό του πλαισίου μέσα στο οποίο θα κινούταν η έρευνα που πραγματοποιήθηκε. Αρχικά, κρίθηκε σκόπιμη η επικέντρωση της μελέτης στην προστασία των προσωπικών και ευαίσθητων προσωπικών δεδομένων από την υφιστάμενη νομοθεσία, που μας οδήγησε στη αναζήτηση του Γενικού Κανονισμού για την Προστασία Δεδομένων της Ευρωπαϊκής Ένωσης, μια και αυτός αποτελεί το πιο πρόσφατο ολοκληρωμένο νομικό κείμενο σχετικά με αυτό το θέμα, που οι απαιτήσεις του έχουν παγκόσμιες επιπτώσεις, ως προς τη συμμόρφωση με αυτόν. Επιπλέον, είναι το πρώτο νομικό κείμενο που αναφέρει ρητά την ψευδωνυμοποίηση, που αποτελεί έναν από τα κύρια στοιχεία του θέματος αυτής της μεταπτυχιακής διατριβής.

Στη συνέχεια αυτής της βιβλιογραφικής επισκόπησης αναζητήθηκαν πηγές που να επικεντρώνονται σε μεθόδους ψευδωνυμοποίησης που ικανοποιούν αυτό το νομικό πλαίσιο. Με δεδομένο ότι ο Γενικός Κανονισμός για την Προστασία Δεδομένων έχει χρόνο εφαρμογής μόνο δύο χρόνια, όπως ήταν αναμενόμενο, εντοπίστηκε η προσπάθεια της επιστημονικής κοινότητας για την αναθεώρηση και βελτίωση παλαιότερων μεθόδων και σχημάτων, την σύλληψη νέων και την έναρξη μελέτης τους, αλλά και τις προτάσεις για νέα πεδία αναζήτησης, με επίκεντρο την κρυπτογραφία.

Από αυτές επελέγη η χρήση δέντρων τύπου Merkle από τις προτάσεις του οργανισμού ENISA για τη διερεύνηση της δυνατότητας ανάπτυξης μιας νέας μεθόδου κρυπτογραφικής ψευδωνυμοποίησης. Τα δέντρα Merkle είναι επέκταση των κρυπτογραφικών συναρτήσεων κατακερματισμού – οι οποίες ήδη έχουν μελετηθεί ως τεχνική ψευδωνυμοποίησης. Έτσι, ακολούθησε αναζήτηση στη βιβλιογραφία για τον εντοπισμό υπαρχουσών ερευνών στο συγκεκριμένο θέμα και όταν διαπιστώθηκε ότι δεν έχει γίνει απόπειρα για αυτή τη χρήση, αποφασίστηκε η έναρξη μελέτης αυτής της δυνατότητας.

Αφού πραγματοποιήθηκε η συγκέντρωση βιβλιογραφικού υλικού για τη λειτουργία αυτής της μεθόδου, προχωρήσαμε στην επινόηση μιας νέας μεθόδου εφαρμογής ψευδωνυμοποίησης με τη χρήση των δέντρων τύπου Merkle και στη συνέχεια στη θεωρητικής ανάλυση για να προσδιοριστούν τα χαρακτηριστικά της και ο τρόπος λειτουργίας της. Σημειώνεται ότι η ασφάλεια των Merkle είναι από τη σκοπιά της κρυπτογραφίας πολύ καλά θεμελιωμένη – και, μάλιστα, σχήματα τύπου Merkle θεωρούνται ισχυρά ακόμα και στη μετα-κβαντική εποχή.

Μετά τη θεωρητική ανάλυση της νέας κατασκευής για σχήμα ψευδωνυμοποίησης βασισμένο σε δέντρα Merkle, πραγματοποιήθηκε πειραματική υλοποίηση, με ανάπτυξη κώδικα λογισμικού, της νέας αυτής τεχνικής.

Πραγματοποιήθηκαν μετρήσεις στο χρόνο εκτέλεσης σε προσωπικό υπολογιστικό σύστημα, για να διαπιστωθεί η συμπεριφορά του σε πραγματικές συνθήκες και με βάση αυτά καταλήξαμε στα συμπεράσματά μας.

1.6 Δομή της Μεταπτυχιακής Διατριβής

Η παρούσα μεταπτυχιακή διατριβή ακολουθεί στη δομή της την πορεία της μεθοδολογίας της έρευνας και της μελέτης του θέματος της πρότασης προηγμένων κρυπτογραφικών μεθόδων για την ψευδωνυμοποίηση δεδομένων.

Στο Κεφάλαιο 2 περιγράφονται τα βασικά στοιχεία και ορισμοί του Γενικού Κανονισμού Προστασίας Δεδομένων της Ευρωπαϊκής Ένωσης, όσον αφορά στα δεδομένα που απαιτούν προστασία, πώς επιτυγχάνεται και πώς προσδιορίζεται η ψευδωνυμοποίηση σε αυτό το πλαίσιο.

Στη συνέχεια, στο Κεφάλαιο 3 παρουσιάζονται οι μέθοδοι κρυπτογραφικής ψευδωνυμοποίησης που προτείνονται ως ικανές να συμμορφωθούν με τις απαιτήσεις του Γενικού Κανονισμού για την Προστασία Δεδομένων της Ευρωπαϊκής Ένωσης και πραγματοποιείται ένας προσδιορισμός της κρυπτογραφίας και των κυριότερων τρόπων υλοποίησής της. Επίσης, σε αυτό το σημείο αποφασίζεται και η διερεύνηση χρήσης των δέντρων τύπου Merkle, ανάμεσα σε άλλες νέες προτάσεις, για ψευδωνυμοποίηση.

Έπειτα, στο Κεφάλαιο 4 περιγράφεται αρχικά ο τρόπος λειτουργίας των δέντρων τύπου Merkle στις μέχρι τώρα υλοποιήσεις του και στη συνέχεια παρουσιάζεται με λεπτομέρεια ο τρόπος

λειτουργίας που προτείνουμε για την αντιμετώπιση του κυρίως ερευνητικού ερωτήματος της παρούσας διατριβής, και τα χαρακτηριστικά ασφάλειας που έχει.

Ακολούθως, στο Κεφάλαιο 5 περιγράφεται ο τρόπος λειτουργίας του κώδικα που παράχθηκε με βάση τον αλγόριθμο λειτουργίας του μηχανισμού που προτάθηκε στο προηγούμενο Κεφάλαιο, καθώς επίσης παρατίθενται μετρήσεις που πραγματοποιήθηκαν σε αυτόν, για να φανεί η απόδοσή του σε πραγματικό περιβάλλον.

Τέλος, στον Επίλογο παρουσιάζουμε τα συνολικά συμπεράσματά μας από την μελέτη που πραγματοποιήθηκε για την παρούσα μεταπτυχιακή διατριβή, συζητείται το πως βοηθάει στην έρευνα του τομέα της κρυπτογραφικής ψευδωνυμοποίησης, καθώς επίσης και οι σκέψεις μας για μελλοντική έρευνα στο συγκεκριμένο θέμα.

Κεφάλαιο 2

ΓΚΠΔ και Ψευδωνυμοποίηση

Στη σύγχρονη εποχή της πληροφορίας, όπως έχει χαρακτηριστεί, τα δεδομένα αποτελούν έναν πολύτιμο πόρο σχεδόν σε κάθε τομέα της ανθρώπινης δραστηριότητας, με αποτέλεσμα οι ραγδαίες τεχνολογικές εξελίξεις και η παγκοσμιοποίηση να έχουν δημιουργήσει σημαντική αύξηση στη συλλογή, επεξεργασία και διακίνηση δεδομένων.

Τμήμα αυτού του συνολικού όγκου πληροφορίας αποτελούν τα προσωπικά και ευαίσθητα προσωπικά δεδομένα, που απαιτούν την προστασία για τη διασφάλιση στοιχειωδών ανθρώπινων δικαιωμάτων. Για την ικανοποίηση αυτής της ανάγκης δημιουργήθηκε και εφαρμόστηκε ο Γενικός Κανονισμός για την Προστασία Δεδομένων (ΓΚΠΔ).

2.1 Γενικός Κανονισμός για την Προστασία Δεδομένων

Στις 25 Μαΐου 2018 τέθηκε σε ισχύ ο Γενικός Κανονισμός για την Προστασία Δεδομένων της Ευρωπαϊκής Ένωσης που κύριο στόχο έχει την εφαρμογή ενός κανονιστικού πλαισίου για τη συλλογή, επεξεργασία και διακίνηση των προσωπικών δεδομένων διασφαλίζοντας το θεμελιώδες δικαίωμα των φυσικών προσώπων στην προστασία των δεδομένων προσωπικού χαρακτήρα που το αφορούν (ΓΚΠΔ 2016: 1).

Ως δεδομένα προσωπικού χαρακτήρα σύμφωνα με τον ορισμό που δίνεται στο Άρθρο 4 νοείται (ΓΚΠΔ 2016: 33): «κάθε πληροφορία που αφορά ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο («υποκείμενο των δεδομένων»)· το ταυτοποιήσιμο φυσικό πρόσωπο είναι εκείνο του οποίου η ταυτότητα μπορεί να εξακριβωθεί, άμεσα ή έμμεσα, ιδίως μέσω αναφοράς σε αναγνωριστικό στοιχείο ταυτότητας, όπως όνομα, σε αριθμό ταυτότητας, σε δεδομένα θέσης, σε

επιγραμμικό αναγνωριστικό ταυτότητας ή σε έναν ή περισσότερους παράγοντες που προσιδιάζουν στη σωματική, φυσιολογική, γενετική, ψυχολογική, οικονομική, πολιτιστική ή κοινωνική ταυτότητα του εν λόγω φυσικού προσώπου.» Το πρόσωπο του οποίου τα δεδομένα υφίστανται επεξεργασία ονομάζεται, στο ΓΚΠΔ, «υποκείμενο των δεδομένων».

Στο Γενικό Κανονισμό για την Προστασία Δεδομένων γίνεται σαφής η αναγνώριση της σημασίας επεξεργασίας των δεδομένων για την οικονομική, κοινωνική και επιστημονική ανάπτυξη και για το λόγο αυτό είναι προσανατολισμένος στην περιγραφή και την εναπόθεση της ευθύνης της ασφάλειας των δεδομένων στους λεγόμενους υπεύθυνους επεξεργασίας και τους διενεργούντες την επεξεργασία, χωρίς να την αποτρέπει ή να την απαγορεύει.

Ως επεξεργασία σύμφωνα με τον ορισμό που δίνεται στο Άρθρο 4 νοείται (ΓΚΠΔ 2016: 33): «κάθε πράξη ή σειρά πράξεων που πραγματοποιείται με ή χωρίς τη χρήση αυτοματοποιημένων μέσων, σε δεδομένα προσωπικού χαρακτήρα ή σε σύνολα δεδομένων προσωπικού χαρακτήρα, όπως η συλλογή, η καταχώριση, η οργάνωση, η διάρθρωση, η αποθήκευση, η προσαρμογή ή η μεταβολή, η ανάκτηση, η αναζήτηση πληροφοριών, η χρήση, η κοινολόγηση με διαβίβαση, η διάδοση ή κάθε άλλη μορφή διάθεσης, η συσχέτιση ή ο συνδυασμός, ο περιορισμός, η διαγραφή ή η καταστροφή.»

Κύριο γνώρισμα του Γενικού Κανονισμού για την Προστασία Δεδομένων αποτελεί η υιοθέτηση γενικών αρχών και κανονισμών που αφορούν στην προστασία των προσωπικών δεδομένων επιτρέποντας με αυτόν τον τρόπο ευελιξία στη μέθοδο με την οποία μπορεί να επιτευχθεί από τους έχοντες την ευθύνη αυτή. Αυτό το γεγονός παρέχει τη δυνατότητα για τη εισήγηση νέων προτύπων στις μεθόδους προστασίας που μπορούν να χρησιμοποιηθούν.

2.2 Τρία Είδη Δεδομένων

Από το Γενικό Κανονισμό για την Προστασία Δεδομένων, με βάση την εργασία (Hu et al 2017:7), προκύπτουν τρεις κατηγορίες δεδομένων σχετικά με την προστασία από την προσωποποίηση, δηλαδή την ικανότητα ταυτοποίησης προσώπων και σύνδεσής τους με τα δεδομένα είτε με άμεσο ή με έμμεσο τρόπο.

2.2.1 Ανώνυμα Δεδομένα

Η πρώτη κατηγορία δεδομένων είναι οι, όπως αναφέρεται στη Σκέψη 26 του Γενικού Κανονισμού για την Προστασία Δεδομένων (ΓΚΠΔ 2016:5): «ανώνυμες πληροφορίες, δηλαδή πληροφορίες που δεν σχετίζονται προς ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο ή σε δεδομένα προσωπικού χαρακτήρα που έχουν καταστεί ανώνυμα κατά τρόπο ώστε η ταυτότητα του υποκειμένου των δεδομένων να μην μπορεί ή να μην μπορεί πλέον να εξακριβωθεί.»

Τα ανωνυμοποιημένα ή ανώνυμα δεδομένα, όπως θα μπορούσαν να χαρακτηριστούν, αποτελούν ένα είδος που η ίδια η φύση τους αποτρέπει τη σύνδεσή τους με την ταυτότητα κάποιου προσώπου. Έτσι, ακόμα και αν το περιεχόμενο της πληροφορίας έχει τα χαρακτηριστικά των προσωπικών δεδομένων, δεν μπορούν να αποτελέσουν πηγή απειλής για την παραβίαση του δικαιώματος προστασίας αυτών των δεδομένων. Άλλωστε, αναφέρεται ρητά και στην ίδια τη ίδιο Σκέψη 26 ότι (ΓΚΠΔ 2016:5): «Για να κριθεί κατά πόσον ένα φυσικό πρόσωπο είναι ταυτοποιήσιμο, θα πρέπει να λαμβάνονται υπόψη όλα τα μέσα τα οποία είναι ευλόγως πιθανό ότι θα χρησιμοποιηθούν, όπως για παράδειγμα ο διαχωρισμός του, είτε από τον υπεύθυνο επεξεργασίας είτε από τρίτο για την άμεση ή έμμεση εξακρίβωση της ταυτότητας του φυσικού προσώπου. Για να διαπιστωθεί κατά πόσον κάποια μέσα είναι ευλόγως πιθανό ότι θα χρησιμοποιηθούν για την εξακρίβωση της ταυτότητας του φυσικού προσώπου, θα πρέπει να λαμβάνονται υπόψη όλοι οι αντικειμενικοί παράγοντες, όπως τα έξοδα και ο χρόνος που απαιτούνται για την ταυτοποίηση, λαμβανομένων υπόψη της τεχνολογίας που είναι διαθέσιμη κατά τον χρόνο της επεξεργασίας και των εξελίξεων της τεχνολογίας. Οι αρχές της προστασίας δεδομένων δεν θα πρέπει συνεπώς να εφαρμόζονται σε ανώνυμες πληροφορίες».

Σημειώνεται ρητά ότι τα ανώνυμα δεδομένα δεν αποτελούν προσωπικά δεδομένα (και, συνεπώς, ο ΓΚΠΔ δεν εφαρμόζεται σε ανώνυμα δεδομένα). Ωστόσο, ελλοχεύει ο κίνδυνος να θεωρηθούν κάποια δεδομένα ανώνυμα χωρίς να είναι. Όπως είχε συμβεί με τη χρήστη No. 4417749 της AOL που αποδείχτηκε εύκολη διαδικασία η ταυτοποίησή της με βάση τα δεδομένα που είχαν δημοσιευθεί, από την εταιρεία, αν και θεωρούνταν ανωνυμοποιημένα, μια και δεν χρησιμοποιούνταν το όνομά της.(Barbaro and Jr 2006:1)

2.2.2 Ψευδωνυμοποιημένα Δεδομένα

Η δεύτερη κατηγορία δεδομένων είναι τα ψευδωνυμοποιημένα δεδομένα, δηλαδή αυτά που έχουν υποστεί διαδικασία ψευδωνυμοποίησης. Σύμφωνα με το Άρθρο 4 του Γενικού Κανονισμού για την Προστασία Δεδομένων (ΓΚΠΔ 2016:33): «ψευδωνυμοποίηση: η επεξεργασία δεδομένων προσωπικού χαρακτήρα κατά τρόπο ώστε τα δεδομένα να μην μπορούν πλέον να αποδοθούν σε

συγκεκριμένο υποκείμενο των δεδομένων χωρίς τη χρήση συμπληρωματικών πληροφοριών, εφόσον οι εν λόγω συμπληρωματικές πληροφορίες διατηρούνται χωριστά και υπόκεινται σε τεχνικά και οργανωτικά μέτρα προκειμένου να διασφαλιστεί ότι δεν μπορούν να αποδοθούν σε ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο.».

Σε αυτή την κατηγορία δεδομένων παρατηρείται η ρητή αναφορά σε μέτρα, τεχνικά και οργανωτικά, που μπορούν να ληφθούν προκειμένου να διασφαλιστεί η αποτροπή της ταυτοποίησης φυσικού προσώπου ή και η απόδοση του περιεχομένου των πληροφοριών σε αυτό.

Σημειώνεται ότι τα ψευδωνυμοποιημένα δεδομένα αποτελούν προσωπικά δεδομένα.

2.2.3 Δεδομένα του Άρθρου 11 του ΓΚΠΔ

Η τρίτη κατηγορία δεδομένων είναι αυτά που αναφέρονται στο Άρθρο 11 του Γενικού Κανονισμού για την Προστασία Δεδομένων και θα μπορούσαν να ονομαστούν Δεδομένα Άρθρου 11. Σύμφωνα με την Παράγραφο 1 του Άρθρου (ΓΚΠΔ 2016:39): «Εάν οι σκοποί για τους οποίους ο υπεύθυνος επεξεργασίας επεξεργάζεται δεδομένα προσωπικού χαρακτήρα δεν απαιτούν ή δεν απαιτούν πλέον την εξακρίβωση της ταυτότητας του υποκειμένου των δεδομένων από τον υπεύθυνο επεξεργασίας, ο υπεύθυνος επεξεργασίας δεν υποχρεούται να διατηρεί, να αποκτά ή να επεξεργάζεται συμπληρωματικές πληροφορίες για την εξακρίβωση της ταυτότητας του υποκειμένου των δεδομένων αποκλειστικά και μόνο για το σκοπό της συμμόρφωσης προς τον παρόντα κανονισμό.». Επιπλέον, στην Παράγραφο 2 του Άρθρου αναφέρεται (ΓΚΠΔ 2016:39): «Όταν, στις περιπτώσεις που αναφέρονται στην παράγραφο 1 του παρόντος άρθρου, ο υπεύθυνος επεξεργασίας μπορεί να αποδείξει ότι δεν είναι σε θέση να εξακριβώσει την ταυτότητα του υποκειμένου των δεδομένων, ο υπεύθυνος επεξεργασίας ενημερώνει σχετικά το υποκείμενο των δεδομένων, εάν είναι δυνατόν. Στις περιπτώσεις αυτές, τα άρθρα 15 ως 20 δεν εφαρμόζονται (τα άρθρα αυτά αναφέρονται στα δικαιώματα των προσώπων), εκτός εάν το υποκείμενο των δεδομένων, για τον σκοπό της άσκησης των δικαιωμάτων του που απορρέουν από τα εν λόγω άρθρα, παρέχει συμπληρωματικές πληροφορίες που επιτρέπουν την εξακρίβωση της ταυτότητάς του.».

Σε αυτή την κατηγορία δεδομένων υπεισέρχεται η έννοια των δεδομένων, χωρίς αναγνωριστικά, είτε γιατί δεν συγκεντρώθηκαν, είτε γιατί αφαιρέθηκαν εκ των υστέρων, που μπορούν να οδηγήσουν στην ταυτοποίηση φυσικών προσώπων ή στη συσχέτισή τους με τα εν λόγω

δεδομένα. Επίσης, πραγματοποιείται αναφορά και στη μη απόκτηση πρόσβασης σε συμπληρωματικά δεδομένα που δεν είναι απαραίτητα στην επεξεργασία και θα μπορούσαν να οδηγήσουν σε ταυτοποίηση φυσικού προσώπου και απόδοση σε αυτό χαρακτηριστικών ή και πληροφοριών που υπάρχουν συνολικά στα δεδομένα. Επιπλέον, σημαντικό σημείο του συγκεκριμένου Άρθρου είναι και αυτό που αναφέρεται στη δυνατότητα του υπεύθυνου επεξεργασία να αποδείξει ότι δεν είναι δυνατή η ταυτοποίηση ενός υποκειμένου των δεδομένων, χωρίς να προσδιορίζεται η διαδικασία ή η μέθοδος με την οποία αυτό επιτυγχάνεται.

Σε αυτό το σημείο είναι σημαντικό να αναφερθεί ότι, ενώ τα δεδομένα του Άρθρου 11 παρουσιάζονται ξεχωριστά, στην ουσία εμπίπτουν στις κατηγορίες των ανωνυμοποιημένων ή των ψευδωνυμοποιημένων ανάλογα με τον τρόπο που έχει επιλεγεί ο τρόπος που θα υλοποιηθούν οι περιορισμοί τους. Εάν υπάρχουν πρόσθετες πληροφορίες που θα επιτρέψουν την αναγνώριση (ακόμα και αν αυτές οι πληροφορίες είναι στα χέρια των ίδιων των προσώπων τους), τότε πρόκειται για ψευδωνυμοποιημένα δεδομένα.

2.3 Η Σημασία της Ψευδωνυμοποίησης

Μετά την παρουσίαση των τριών κατηγοριών δεδομένων σχετικά με την προστασία από την προσωποποίηση που προκύπτουν από το Γενικό Κανονισμό για την Προστασία Δεδομένων θα πρέπει να αναφερθεί ότι η χρήση της ψευδωνυμοποίησης προκρίνεται ως η μέθοδος για την περεταίρω διασφάλιση της επεξεργασίας των δεδομένων με την ταυτόχρονη προστασία των θεμελιωδών δικαιωμάτων, όπως φαίνεται στις Σκέψεις 28 και 29 (ΓΚΠΔ 2016:5): «*Η χρήση της ψευδωνυμοποίησης στα δεδομένα προσωπικού χαρακτήρα μπορεί να μειώσει τους κινδύνους για τα υποκείμενα των δεδομένων και να διευκολύνει τους υπευθύνους επεξεργασίας και τους εκτελούντες την επεξεργασία να τηρήσουν τις οικείες υποχρεώσεις περί προστασίας των δεδομένων. Η ρητή εισαγωγή της «ψευδωνυμοποίησης» του παρόντος κανονισμού δεν προορίζεται να αποκλείσει κάθε άλλο μέτρο προστασίας των δεδομένων.*» «*Για να δημιουργηθούν κίνητρα για την ψευδωνυμοποίηση κατά την επεξεργασία δεδομένων προσωπικού χαρακτήρα, θα πρέπει να είναι δυνατή η λήψη μέτρων ψευδωνυμοποίησης, με παράλληλη δυνατότητα μιας γενικής ανάλυσης, στο πλαίσιο του ίδιου υπευθύνου επεξεργασίας, όταν ο εν λόγω υπεύθυνος επεξεργασίας έχει λάβει τα τεχνικά και οργανωτικά μέτρα που είναι αναγκαία, ώστε να διασφαλιστεί, για τη σχετική επεξεργασία δεδομένων, η εφαρμογή του παρόντος κανονισμού και ότι οι συμπληρωματικές πληροφορίες για την απόδοση των δεδομένων προσωπικού χαρακτήρα σε συγκεκριμένο υποκείμενο των δεδομένων διατηρούνται χωριστά. Ο υπεύθυνος επεξεργασίας που*

επεξεργάζεται τα δεδομένα προσωπικού χαρακτήρα θα πρέπει να υποδεικνύει τα εξουσιοδοτημένα πρόσωπα εντός του ίδιου υπευθύνου επεξεργασίας.»

Επίσης, η ψευδωνυμοποίηση αναφέρεται ρητά σαν πιθανό μέσο εγγύησης της προστασίας των προσωπικών δεδομένων στις εξής περιπτώσεις:

- Άρθρο 6 παράγραφος 4 (ΓΚΠΔ 2016:36): «Όταν η επεξεργασία για σκοπό άλλο από αυτόν για τον οποίο έχουν συλλεγεί τα δεδομένα προσωπικού χαρακτήρα δεν βασίζεται στη συγκατάθεση του υποκειμένου των δεδομένων ή στο δίκαιο της Ένωσης ή το δίκαιο κράτους μέλους το οποίο αποτελεί αναγκαίο και αναλογικό μέτρο σε μια δημοκρατική κοινωνία για τη διασφάλιση των σκοπών που αναφέρονται στο άρθρο 23 παράγραφος 1, ο υπεύθυνος επεξεργασίας, προκειμένου να εξακριβωθεί κατά πόσο η επεξεργασία για άλλο σκοπό είναι συμβατή με τον σκοπό για τον οποίο συλλέγονται αρχικώς τα δεδομένα προσωπικού χαρακτήρα, λαμβάνει υπόψη, μεταξύ άλλων: (…)ε) την ύπαρξη κατάλληλων εγγυήσεων, που μπορεί να περιλαμβάνουν κρυπτογράφηση ή ψευδωνυμοποίηση.»

- Άρθρο 25 παράγραφος 1 (ΓΚΠΔ 2016:48): «Λαμβάνοντας υπόψη τις τελευταίες εξελίξεις, το κόστος εφαρμογής και τη φύση, το πεδίο εφαρμογής, το πλαίσιο και τους σκοπούς της επεξεργασίας, καθώς και τους κινδύνους διαφορετικής πιθανότητας επέλευσης και σοβαρότητας για τα δικαιώματα και τις ελευθερίες των φυσικών προσώπων από την επεξεργασία, ο υπεύθυνος επεξεργασίας εφαρμόζει αποτελεσματικά, τόσο κατά τη στιγμή του καθορισμού των μέσων επεξεργασίας όσο και κατά τη στιγμή της επεξεργασίας, κατάλληλα τεχνικά και οργανωτικά μέτρα, όπως η ψευδωνυμοποίηση, σχεδιασμένα για την εφαρμογή αρχών προστασίας των δεδομένων, όπως η ελαχιστοποίηση των δεδομένων, και την ενσωμάτωση των απαραίτητων εγγυήσεων στην επεξεργασία κατά τρόπο ώστε να πληρούνται οι απαιτήσεις του παρόντος κανονισμού και να προστατεύονται τα δικαιώματα των υποκειμένων των δεδομένων.»

- Άρθρο 32 παράγραφος 1 (ΓΚΠΔ 2016:51): «Λαμβάνοντας υπόψη τις τελευταίες εξελίξεις, το κόστος εφαρμογής και τη φύση, το πεδίο εφαρμογής, το πλαίσιο και τους σκοπούς της επεξεργασίας, καθώς και τους κινδύνους διαφορετικής πιθανότητας επέλευσης και σοβαρότητας για τα δικαιώματα και τις ελευθερίες των φυσικών προσώπων, ο υπεύθυνος επεξεργασίας και ο εκτελών την επεξεργασία εφαρμόζουν κατάλληλα τεχνικά και οργανωτικά μέτρα προκειμένου να διασφαλίζεται το κατάλληλο επίπεδο ασφάλειας

έναντι των κινδύνων., περιλαμβανομένων, μεταξύ άλλων, κατά περίπτωση: α) της ψευδωνυμοποίησης και της κρυπτογράφησης δεδομένων προσωπικού χαρακτήρα, (…)»

- Άρθρο 89 παράγραφος 1 (ΓΚΠΔ 2016:84): «Η επεξεργασία για σκοπούς αρχειοθέτησης για το δημόσιο συμφέρον ή για σκοπούς επιστημονικής ή ιστορικής έρευνας ή για στατιστικούς σκοπούς υπόκειταισε κατάλληλες εγγυήσεις, σύμφωνα με τον παρόντα κανονισμό, ως προς τα δικαιώματα και τις ελευθερίες του υποκειμένου των δεδομένων, σύμφωνα με τον παρόντα κανονισμό. Οι εν λόγω εγγυήσεις διασφαλίζουν ότι έχουν θεσπιστεί τα τεχνικά και οργανωτικά μέτρα, ιδίως για να διασφαλίζουν την τήρηση της αρχής της ελαχιστοποίησης των δεδομένων. Τα εν λόγω μέτρα μπορούν να περιλαμβάνουν τη χρήση ψευδωνύμων, εφόσον οι εν λόγω σκοποί μπορούν να εκπληρωθούν κατ' αυτόν τον τρόπο. Εφόσον οι εν λόγω σκοποί μπορούν να εκπληρωθούν από περαιτέρω επεξεργασία η οποία δεν επιτρέπει ή δεν επιτρέπει πλέον την ταυτοποίηση των υποκειμένων των δεδομένων, οι εν λόγω σκοποί εκπληρώνονται κατ' αυτόν τον τρόπο.»

- Άρθρο 40 παράγραφος 2 (ΓΚΠΔ 2016:56): «Ενώσεις και άλλοι φορείς που εκπροσωπούν κατηγορίες υπευθύνων επεξεργασίας ή εκτελούντων την επεξεργασία μπορούν να εκπονούν κώδικες δεοντολογίας ή να τροποποιούν ή να επεκτείνουν υφιστάμενους κώδικες δεοντολογίας, προκειμένου να προσδιορίσουν την εφαρμογή του παρόντος κανονισμού, όπως όσον αφορά (μεταξύ άλλων) την ψευδωνυμοποίηση δεδομένων προσωπικού χαρακτήρα».

Είναι φανερή, λοιπόν, η σημασία που δίνεται στην ψευδωνυμοποίηση στο Γενικό Κανονισμό για την Προστασία Δεδομένων και για αυτό το λόγο η παρούσα μεταπτυχιακή διατριβή θα ασχοληθεί με τη μελέτη μεθόδων εφαρμογής της.

Κεφάλαιο 3

Ψευδωνυμοποίηση και Κρυπτογραφία

Η σημασία της ψευδωνυμοποίησης για την προστασία των δεδομένων είναι φανερή εδώ και πολλά χρόνια, όπου και εμφανίζεται ως έννοια σε πάρα πολλά πρότυπα ασφαλείας. Για παράδειγμα, ήδη από το 2008 στο πρότυπο ISO/TS 25237:2008 για τα δεδομένα υγείας εμφανίζεται η έννοια της ψευδωνυμοποίησης ως μέσο ενίσχυσης της προστασίας των δεδομένων αυτών. Μετά την έναρξη ισχύος του ΓΚΠΔ, του πρώτου αμιγώς νομικού κειμένου στο οποίο υπεισέρχεται ο ορισμός της ψευδωνυμοποίησης, επιτείνονται οι προσπάθειες για την περιγραφή και την, κατά κάποιο τρόπο, «προτυποποίηση» των μεθόδων που μπορούν να ακολουθηθούν, ώστε οι υλοποιήσεις να ανταποκρίνονται στο ισχύον κανονιστικό πλαίσιο.

Στο παρόν Κεφάλαιο θα παρουσιάσουμε τις πλέον γνωστές τεχνικές ψευδωνυμοποίησης που έχουν ως κύρια βάση την κρυπτογραφία. Από την πληθώρα πηγών και ορισμών, θα επικεντρωθούμε σε αυτές που προέρχονται από κείμενα οργάνων της Ευρωπαϊκής Ένωσης που σχετίζονται με την προστασία προσωπικών δεδομένων και της ασφάλειας επικοινωνιών και δικτύων.

3.1 Κρυπτογραφία

Σε αυτό το σημείο κρίνεται δόκιμη μία επεξηγηματική αναφορά στις μεθόδους κρυπτογραφίας, μια και όπως θα φανεί στις επόμενες Ενότητες αυτού του Κεφαλαίου οι περισσότερες μέθοδοι που προτείνονται για ψευδωνυμοποίηση είναι κρυπτογραφικές.

Αρχικά θα πρέπει να δοθεί ο ορισμός της κρυπτογραφίας σύμφωνα με τον οποίο είναι η μελέτη μαθηματικών τεχνικών σχετικών από πλευράς ασφάλειας πληροφορίας, όπως η εμπιστευτικότητα, η ακεραιότητα δεδομένων, ταυτοποίηση οντότητας και επιβεβαίωση προέλευσης δεδομένων (Menezes et al 1996:4).

Και έχει κύριους στόχους:

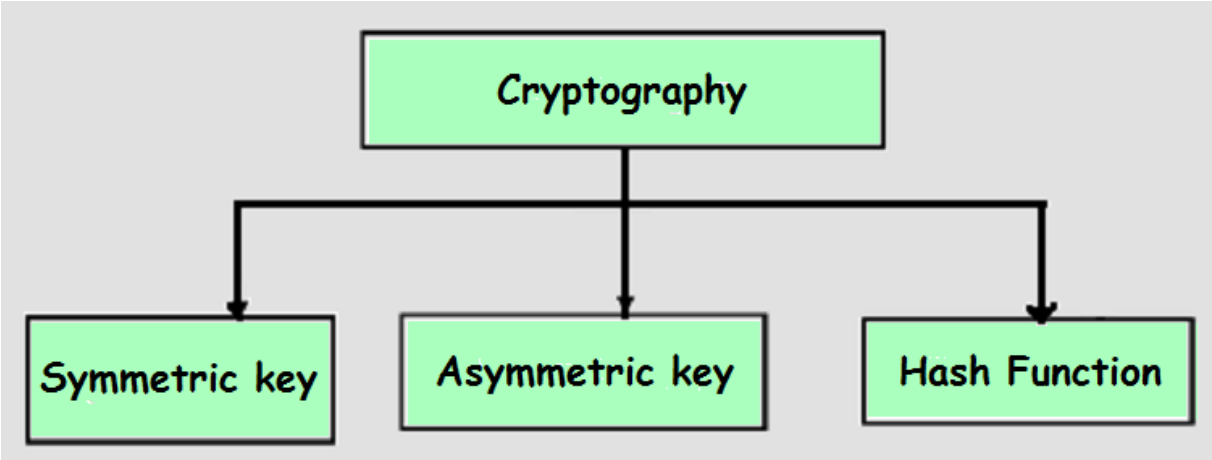
- Την εμπιστευτικότητα, δηλαδή την διασφάλιση ότι πρόσβαση σε μία πληροφορία θα έχουν μόνο όσοι έχουν την απαραίτητη εξουσιοδότηση και σε όλους τους υπόλοιπους θα είναι μη αναγνώσιμη.

- Την ακεραιότητα δεδομένων, δηλαδή τη διαβεβαίωση ότι τα δεδομένα δεν έχουν τροποποιηθεί από μη εξουσιοδοτημένες οντότητες.

- Την ταυτοποίηση τόσο οντοτήτων όσο και δεδομένων, δηλαδή όταν δύο οντότητες επικοινωνούν να είναι σε θέση να επιβεβαιώσουν την ταυτότητά τους μεταξύ τους, και την αυθεντικότητα των δεδομένων που ανταλλάσσονται.

- Την απόδοση ευθύνης, δηλαδή να μην μπορεί να αρνηθεί μία οντότητα την ευθύνη των πράξεών της ή των δεσμεύσεών της.

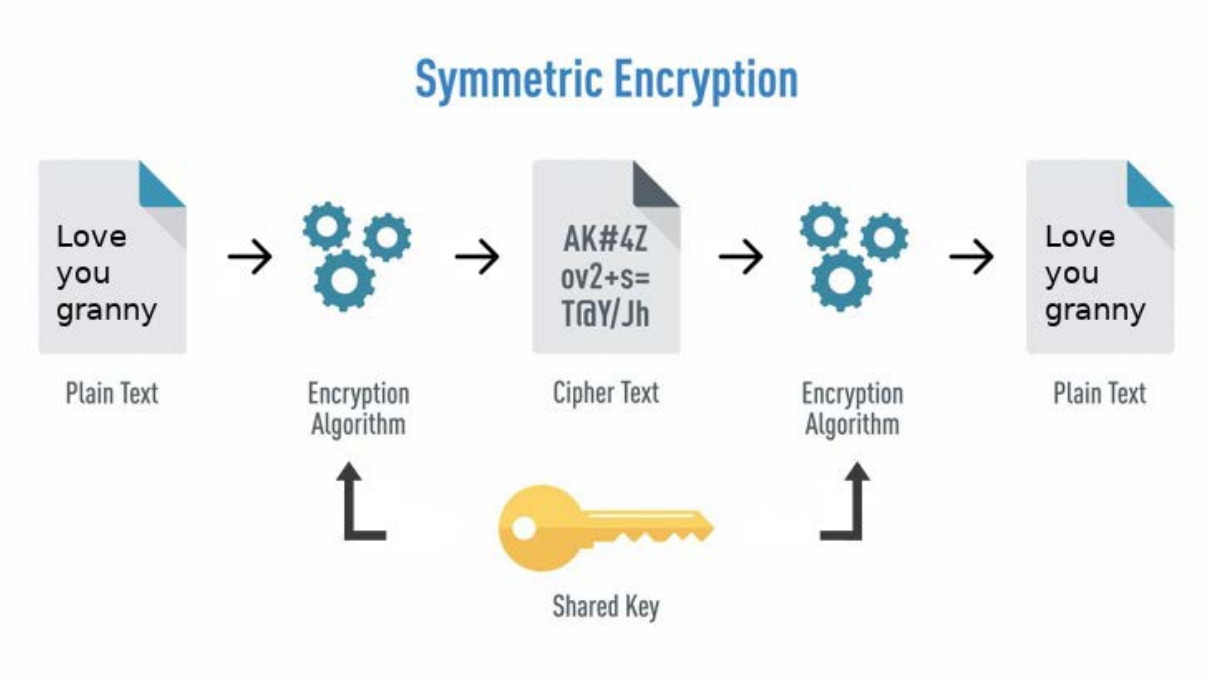
Για την επίτευξη αυτών των στόχων έχουν δημιουργηθεί πολλές τεχνικές που μπορούν ομαδοποιηθούν σε τρεις κύριες κατηγορίες όπως φαίνεται και από την Εικόνα 1 παρακάτω:



Εικόνα 1. Οι βασικές κατηγορίες της κρυπτογραφίας.

3.1.1 Κρυπτογραφία Συμμετρικού Κλειδιού

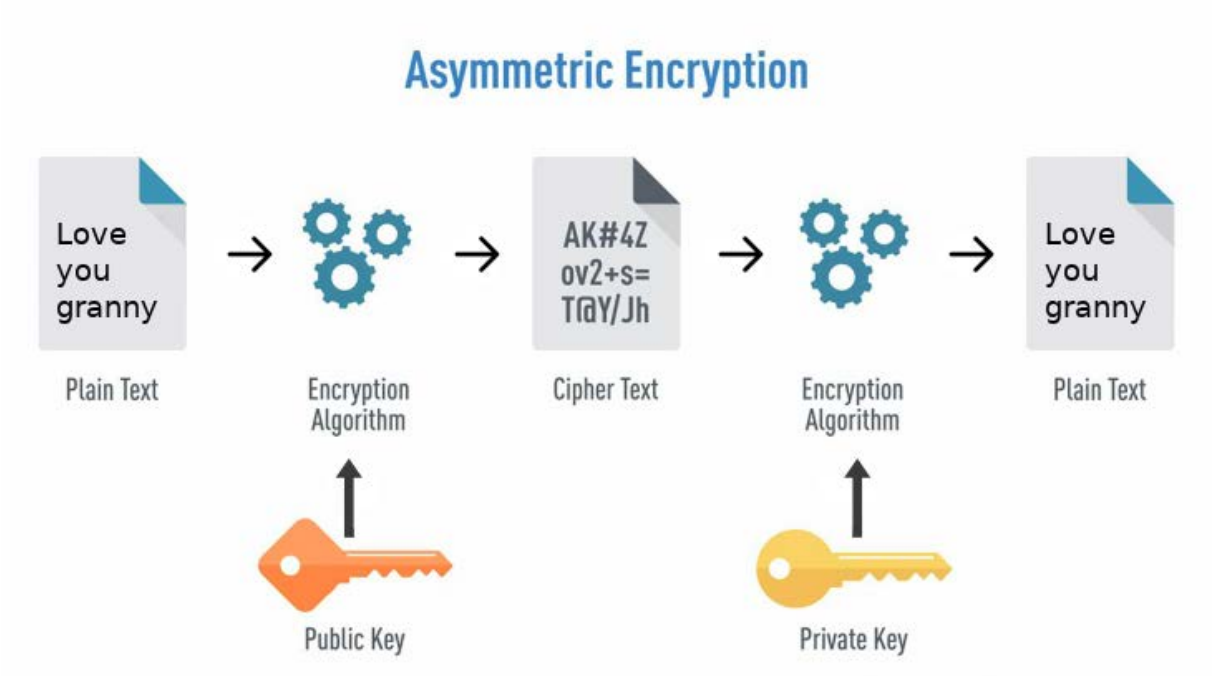
Αποτελεί μία μέθοδο κρυπτογραφίας κατά την οποία η πληροφορία που πρέπει να προφυλαχθεί τροποποιείται με τρόπο που την καθιστά μη αναγνώσιμη με τη χρήση ενός αλγορίθμου και μιας μυστικής τιμής, που συνήθως ονομάζεται κλειδί. Ο μόνος τρόπος που μπορεί να επιστρέψει από αυτή την κατάσταση στην αρχική της και να είναι πάλι αναγνώσιμη, είναι απαραίτητη η γνώση τόσο της συνάρτησης που εφαρμόστηκε όσο και του κλειδιού, όπως φαίνεται και στη Εικόνα 2 παρακάτω:



Εικόνα 2. Απεικόνιση της λειτουργίας της κρυπτογραφίας συμμετρικού κλειδιού. (Verma 2019:1)

3.1.2 Κρυπτογραφία Ασύμμετρου Κλειδιού

Αυτή η τεχνοτροπία κρυπτογράφησης βασίζεται σε έναν αλγόριθμο και την ύπαρξη δύο κλειδιών. Όταν η πληροφορία τροποποιηθεί από αυτόν με τη χρήση του ενός κλειδιού, που συνήθως έχει την ονομασία δημόσιο κλειδί, μετατρέπεται σε μη αναγνώσιμη και ο μόνος τρόπος να επανέλθει στην προηγούμενη κατάσταση, είναι η εφαρμογή του ίδιου αλγορίθμου με τη χρήση αυτή τη φορά του δεύτερου κλειδιού, που συνήθως έχει την ονομασία ιδιωτικό κλειδί, όπως φαίνεται και στην Εικόνα 3 παρακάτω:

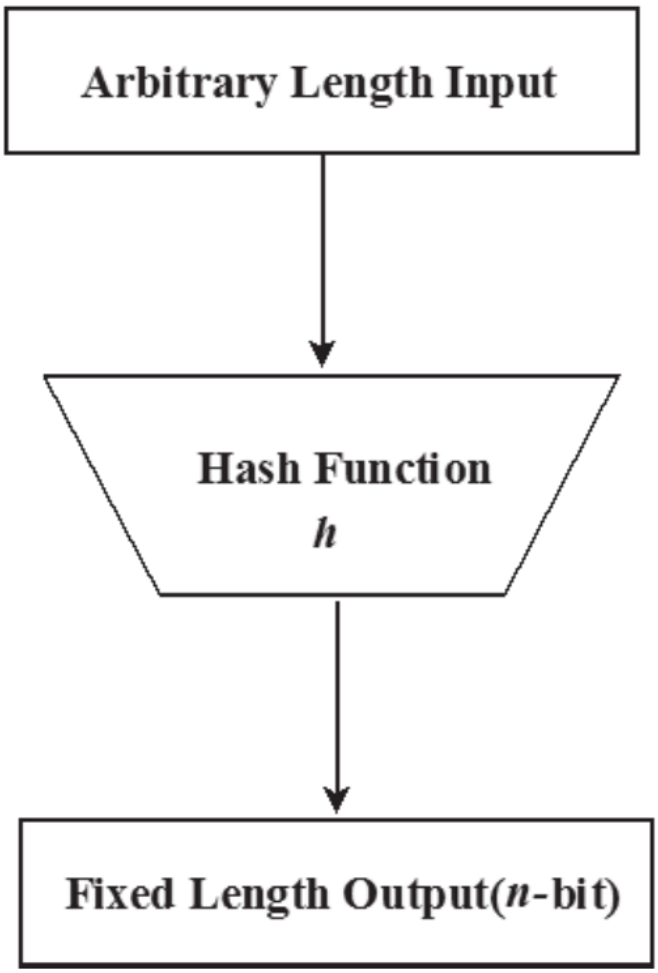


Εικόνα 3. Απεικόνιση της λειτουργίας της κρυπτογραφίας ασύμμετρου κλειδιού. (Verma 2019:1)

Ένας βασικός περιορισμός που πρέπει να ισχύει, για να έχει αποτελεσματικότητα αυτή η μέθοδος, είναι ο πρακτικά μη δυνατός προσδιορισμός του ιδιωτικού κλειδιού από τη γνώση του δημόσιου.

3.1.3 Συναρτήσεις Κατακερματισμού

Πρόκειται για συναρτήσεις που για μία οσοδήποτε μεγέθους είσοδο πληροφορίας σε αυτές επιστρέφεται μία σταθερού μεγέθους έξοδος, χωρίς να είναι μαθηματικά αντιστρέψιμη η διαδικασία, γεγονός που σημαίνει ότι τα δεδομένα παύουν να ενυπάρχουν μέσα στα τροποποιημένα δεδομένα, σε αντίθεση με τις προηγούμενες μεθόδους που αναφέρθηκαν. Μια απεικόνιση αυτής τη διαδικασίας φαίνεται στην Εικόνα 4 παρακάτω:



Εικόνα 4. Απεικόνιση της λειτουργίας της συνάρτησης κατακερματισμού.

Για την αποτελεσματικότητα ως προς την ασφάλεια αυτών των συναρτήσεων θα πρέπει ικανοποιούνται οι εξής ιδιότητες (Menezes et al 1996:323):

- Για κάθε έξοδο h(m) της συνάρτησης κατακερματισμού να είναι πρακτικά αδύνατος ο υπολογισμός του στοιχείου εισόδου της m.

- Για ένα δοθέν στοιχείο εισόδου m να είναι πρακτικά αδύνατος ο υπολογισμός ενός διαφορετικού στοιχείου εισόδου m', που να δίνει την ίδια έξοδο με το m, δηλαδή h(m)= h(m').

- Να είναι πρακτικά αδύνατος ο υπολογισμός δύο διαφορετικών στοιχείων εισόδου m≠m', για τα οποία να παράγεται η ίδια έξοδος από τη συνάρτηση κατακερματισμού, δηλαδή h(m)= h(m').

.

3.2 Προτεινόμενες Μέθοδοι Ψευδωνυμοποίησης της Ομάδας Εργασίας του Άρθρου 29

Η ομάδα εργασίας του άρθρου 29 υπήρξε η ανεξάρτητη ευρωπαϊκή ομάδα εργασίας που χειριζόταν θέματα σχετικά με την προστασία της ιδιωτικής ζωής και των δεδομένων προσωπικού χαρακτήρα έως τις 25 Μαΐου 2018. Τα μέλη της αποτελούνταν από εκπροσώπους των Αρχών Προστασίας Προσωπικών Δεδομένων της Ευρωπαϊκής Ένωσης. Η ομάδα θεσπίστηκε με το άρθρο 29 της Οδηγίας 95/46/ΕΚ. Με την κατάργηση της εν λόγω Οδηγίας από το ΓΚΠΔ, η εν λόγω ομάδα άλλαξε νομική υπόσταση και ρόλο (πλέον πρόκειται για το Ευρωπαϊκό Συμβούλιο Προστασίας Δεδομένων).

Στην εργασία της ομάδας σχετικά με ανωνυμοποίηση αλλά και ψευδωνυμοποίηση, αναφέρεται ότι η ψευδωνυμοποίηση είναι μία μέθοδος που μειώνει τη συνδεσιμότητα ενός συνόλου δεδομένων με την αρχική ταυτότητα του υποκειμένου αυτών των δεδομένων και για αυτό το λόγο αποτελεί ένα χρήσιμο εργαλείο ασφάλειας.

Στη συνέχεια η Ομάδα Εργασίας του Άρθρου 29 προχωρά στην απαρίθμηση των πιο συχνά χρησιμοποιούμενων μεθόδων που είναι οι εξής (Article 29 Working Party, 2014: 20) :

- Κρυπτογράφηση με μυστικό κλειδί: σε αυτή την περίπτωση ο κάτοχος του κλειδιού είναι σε θέση να αναγνώσει τα αρχικά δεδομένα (δηλαδή να αντιστρέψει την ψευδωνυμοποίηση) με μία απλή διαδικασία αποκρυπτογράφησης, μια και αυτά εξακολουθούν να υπάρχουν, απλά είναι τροποποιημένα μέσω της κρυπτογράφησης. Εφόσον χρησιμοποιείται ισχυρός αλγόριθμος κρυπτογράφησης, τότε η αποκρυπτογράφηση είναι πρακτικά εφικτή μόνο με τη γνώση του κλειδιού.

- Συνάρτηση κατακερματισμού: σε αυτή την περίπτωση για μία οσοδήποτε μεγέθους είσοδο της συνάρτησης επιστρέφεται μία σταθερού μεγέθους έξοδος, χωρίς να είναι μαθηματικά αντιστρέψιμη η διαδικασία, γεγονός που σημαίνει ότι τα δεδομένα παύουν να ενυπάρχουν μέσα στα τροποποιημένα δεδομένα, όπως συμβαίνει στην κρυπτογράφηση. Όμως, αν το εύρος των δυνατών εισόδων είναι γνωστό, τότε είναι δυνατό με συνεχόμενες δοκιμές των πιθανών τιμών να μπορεί να εντοπιστεί η αντιστοίχιση ανάμεσα στην έξοδο της συνάρτησης κατακερματισμού και της εισόδου. Οι συναρτήσεις αυτές είναι σχεδιασμένες ώστε να υπολογίζονται σχετικά γρήγορα και είναι

ευάλωτες σε επιθέσεις εξαντλητικών δοκιμών. Επίσης προϋπολογισμένοι πίνακες μπορούν να δημιουργηθούν για την αντιστροφή μεγάλων συλλογών από δεδομένα που έχουν τροποποιηθεί με τη χρήση συνάρτησης κατακερματισμού. Η χρήση μιας τροποποιημένης συνάρτησης κατακερματισμού, όπου στο στοιχείο εισόδου προστίθεται μια τυχαία τιμή (salted-hash function) είναι δυνατό να μειώσει, χωρίς να την εξαλείψει εντελώς, την πιθανότητα να μπορέσει να υπολογιστεί η αρχική τιμή που αντιστοιχεί στην έξοδο.

- Συνάρτηση κατακερματισμού με αποθηκευμένο κρυφό κλειδί: σε αυτή την περίπτωση μπορεί να χρησιμοποιηθεί μία συγκεκριμένου τύπου συνάρτηση κατακερματισμού η οποία χρησιμοποιεί ένα κρυφό κλειδί σαν ένα επιπλέον στοιχείο εισόδου της. Ο υπεύθυνος επεξεργασίας των δεδομένων έχει τη δυνατότητα έτσι να επιβεβαιώσει την αντιστοίχιση ενός στοιχείου εισόδου με το αποτέλεσμα του κατακερματισμού, αλλά για κάποιον που δεν έχει γνώση του κρυφού κλειδιού η πιθανότητα του εντοπισμού της αντιστοίχισης είναι τόσο μικρή που καθίσταται πρακτικά μη εφικτός.

- Ντετερμινιστική κρυπτογράφηση ή συνάρτηση κατακερματισμού με διαγραφή του κλειδιού: αυτή η τεχνική μπορεί να εξισωθεί με το να επιλεγεί ένας τυχαίος αριθμός για κάθε πεδίο μιας βάσης δεδομένων ως ψευδώνυμο και μετά να διαγραφεί ο πίνακας αντιστοίχισης. Αυτή η λύση μειώνει τον κίνδυνο συσχέτισης των προσωπικών δεδομένων ενός συνόλου δεδομένων με αυτά που αντιστοιχούν στο ίδιο πρόσωπο σε ένα άλλο σύνολο δεδομένων, αφού τα ψευδώνυμα θα είναι διαφορετικά. Δεδομένου, λοιπόν, ότι χρησιμοποιείται ισχυρός αλγόριθμος για την κρυπτογράφηση, θα είναι υπολογιστικά δύσκολο (πρακτικά, ανέφικτο) για έναν επιτιθέμενο να αποκρυπτογραφήσει ή να αντιστρέψει τη συνάρτηση, αφού το κλειδί δεν είναι διαθέσιμο.

- Χρήση ετικετών μιας χρήσης (tokenisation): αυτή η τεχνική χρησιμοποιείται κυρίως, αλλά όχι μόνο, στον οικονομικό τομέα, για να αντικατασταθούν αναγνωριστικοί αριθμοί από τιμές που έχουν μειωμένη αξία σε ένα επιτιθέμενο (δηλαδή δεν του αποκαλύπτουν κάποια πληροφορία). Αυτή η μέθοδος προέρχεται από τις προηγούμενες βασιζόμενη στην εφαρμογή μονόδρομης κρυπτογράφησης ή στην ανάθεση ενός σειριακού αριθμού ή μιας τυχαίας τιμής που δημιουργήθηκε μέσω μιας συνάρτησης, που δεν μπορεί να υπολογιστεί μαθηματικά από τα αρχικά δεδομένα.

3.3 Προτεινόμενες μέθοδοι ψευδωνυμοποίησης από τον ENISA

Ένας από τους κύριους ρόλους του Οργανισμού της Ευρωπαϊκής Ένωσης για την Κυβερνοασφάλεια (πρώην Οργανισμός για την Ασφάλεια Δικτύων και Πληροφοριών - ENISA) είναι η ανάπτυξη και παροχή συμβουλών και προτροπών για μεθόδους καλής πρακτικής σε θέματα ασφάλειας πληροφοριών και όπως ήταν φυσικό, με την εφαρμογή του Γενικού Κανονισμού για την Προστασία Δεδομένων, ξεκίνησε την παραγωγή έργου πάνω σε αυτόν το ρόλο. Σχετικά με την ψευδωνυμοποίηση και τις πρακτικές και τεχνικές που μπορούν να χρησιμοποιηθούν, ο ENISA έχει προχωρήσει και δημοσιεύσει δύο αναφορές: μία το Νοέμβριο του 2018 και μία το Νοέμβριο του 2019.

3.3.1 Αναφορά Νοεμβρίου 2018

Όσον αφορά στην ψευδωνυμοποίηση σε αυτή την αναφορά ο ENISA παραθέτει ότι ο τρόπος σχεδίασης της μεθόδου ψευδωνυμοποίησης που θα ακολουθηθεί θα πρέπει να ικανοποιεί τις ακόλουθες ιδιότητες (ENISA, 2018:19):

- Τα ψευδώνυμα δε θα πρέπει να επιτρέπουν εύκολη επαναταυτοποίηση από τρίτους (π.χ. οποιονδήποτε πέρα από τον υπεύθυνο των δεδομένων) μέσα σε ένα πλαίσιο επεξεργασίας δεδομένων.

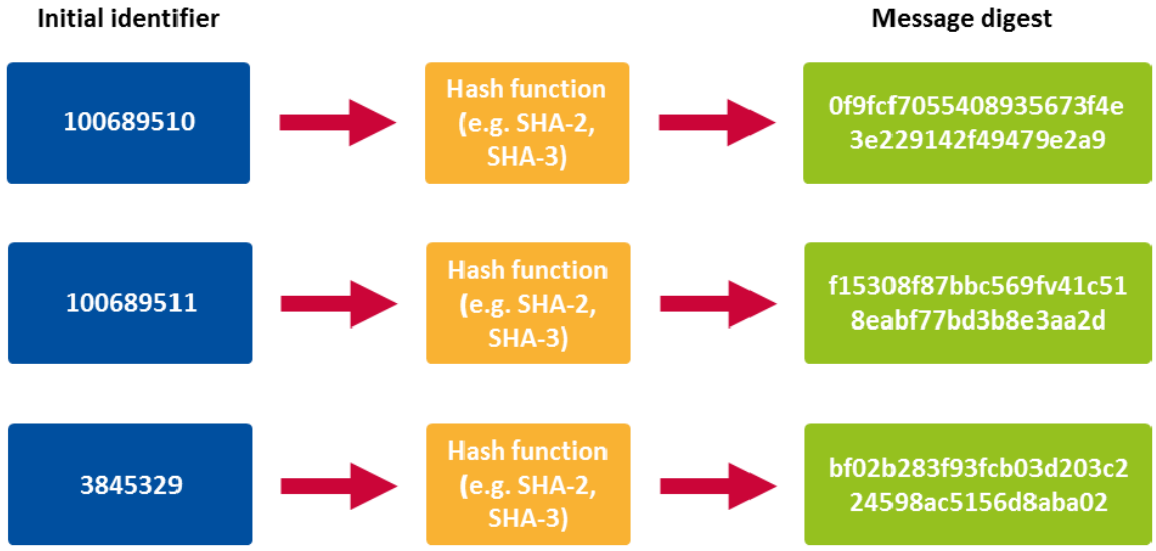
- Δε θα πρέπει να είναι απλό σε κάποιον τρίτο η αναπαραγωγή των ψευδωνύμων, έτσι ώστε να μην είναι δυνατή η χρήση του ίδιου ψευδώνυμου για διάφορα δεδομένα (μη δυνατότητα διασύνδεσης).

Επίσης γίνεται ιδιαίτερη αναφορά στην προτροπή ότι στο σχεδιασμό της φύλαξης των δεδομένων θα πρέπει να επιτυγχάνεται διαχωρισμός των ψευδωνυμοποιημένων δεδομένων από τα υπόλοιπα δεδομένα και ότι, σε κάποιες περιπτώσεις, προτείνεται η χρήση κάποιου πίνακα αντιστοίχισης ανάμεσά τους.

Στη συνέχεια πραγματοποιείται μία σχετικά σύντομη περιγραφή των μεθόδων ψευδωνυμοποίησης με προτάσεις για τις ιδιότητες που πρέπει να ικανοποιούνται ώστε να είναι

αποτελεσματικά ασφαλείς και αξιολόγησή τους με βάση τους δύο παραπάνω κανόνες που αναφέρθηκαν σύμφωνα με την αναφορά αυτή (ENISA, 2018:20):

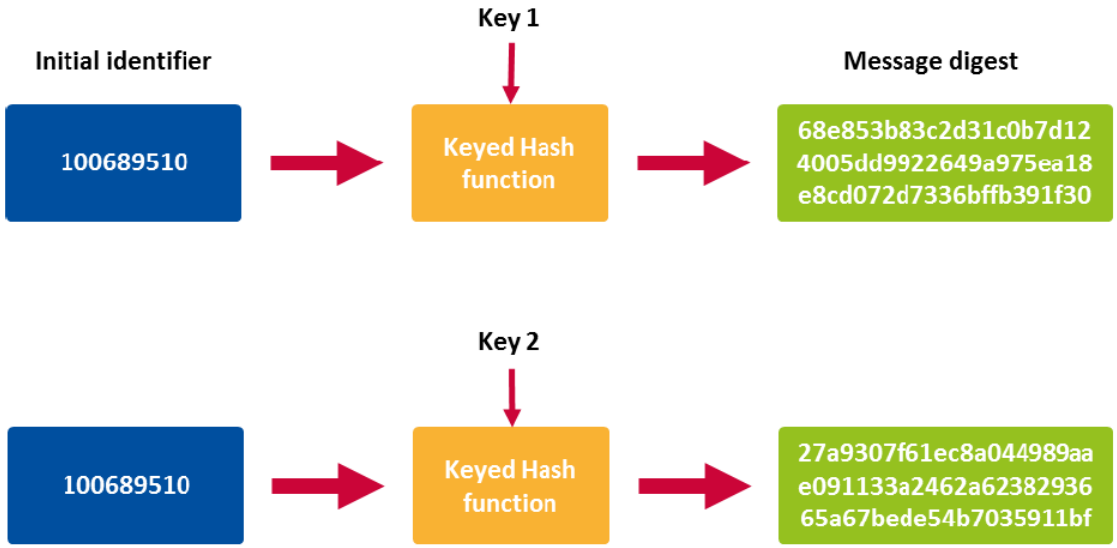
- Συνάρτηση κατακερματισμού χωρίς κλειδί: Πρόκειται για κρυπτογραφικές συναρτήσεις κατακερματισμού, όπως φαίνεται στο παράδειγμα της Εικόνας 5, που αναφέρθηκε ανωτέρω και από την Ομάδα Εργασίας του Άρθρου 29 που όμως δεν μπορούν να ικανοποιήσουν τους δύο κανόνες και έτσι εμφανίζουν σημαντικά μειονεκτήματα στη χρήση της για ψευδωνυμοποίηση (ENISA, 2018:20).



Εικόνα 5. Παράδειγμα χρήσης συνάρτησης κατακερματισμού. (ENISA, 2018:21)

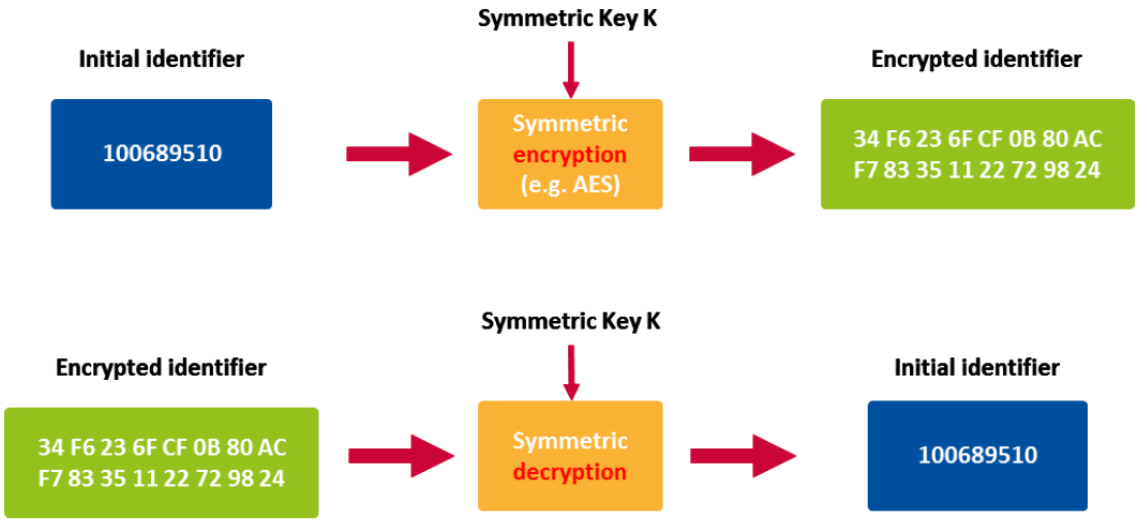
- Συνάρτηση κατακερματισμού με χρήση κλειδιού ή salt (ENISA, 2018:22): Σε αυτές τις συναρτήσεις κατακερματισμού εισάγεται και ένα δεύτερο στοιχείο εισόδου, όπως φαίνεται στο παράδειγμα της Εικόνας 6, που προσδίδει στη διαδικασία επιπλέον πολυπλοκότητα, καθιστώντας τη πιο ανθεκτική στις επιθέσεις, καθώς και το χαρακτηριστικό ότι για ίδιο στοιχείο εισόδου, αν χρησιμοποιηθεί διαφορετικό κλειδί, θα παραχθεί διαφορετική έξοδος. Αυτή η ιδιότητα είναι που, σε αντίθεση με την απλή συνάρτηση κατακερματισμού, την κάνει να ικανοποιεί τους δύο κύριους κανόνες που προαναφέρθηκαν. Σε αυτό το σημείο θα πρέπει να αναφερθεί ότι είναι πολύ σημαντική η πολιτική που θα ακολουθηθεί για τη φύλαξη του κλειδιού ή του salt, ώστε να παραμένουν κρυφά. Ένα γνωστό πρότυπο κρυπτογραφικής συνάρτησης κατακερματισμού με κλειδί (γνωστές στην κρυπτογραφία ως κώδικες αυθεντικοποίησης μηνύματος – Message Authentication Codes) είναι ο HMAC (NIST 2008:4). Ο HMAC είναι

ασφαλής εφόσον οι υποκείμενες απλές κρυπτογραφικές συναρτήσεις κατακερματισμού στις οποίες βασίζεται είναι ασφαλείς – για παράδειγμα, οι SHA-2 και SHA-3 (NIST 2012:3, NIST 2015:3).

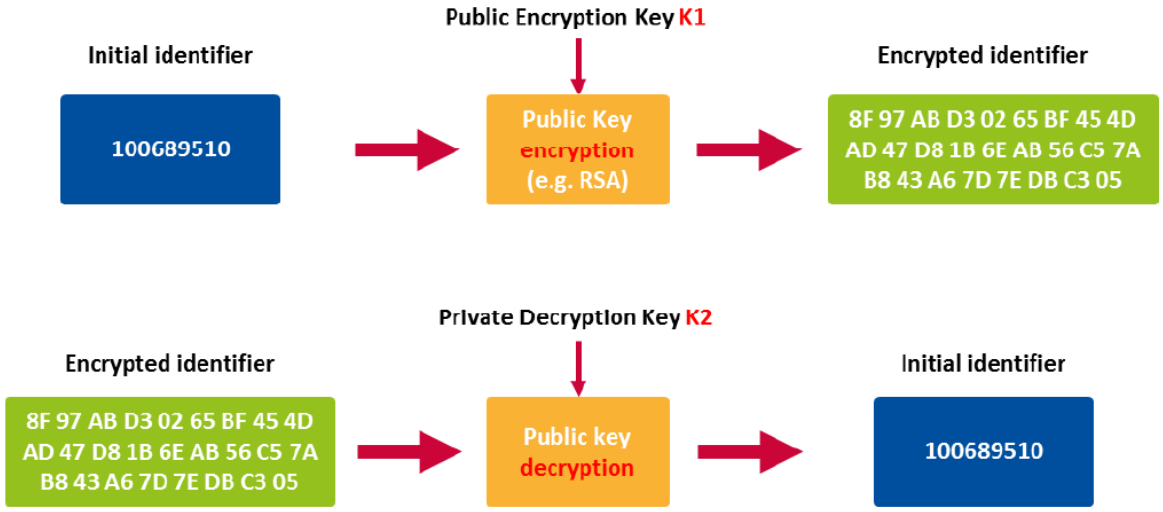


Εικόνα 6. Παράδειγμα χρήσης συνάρτησης κατακερματισμού. (ENISA, 2018:23)

- Κρυπτογράφηση (ENISA, 2018:25): Σε αυτή την περίπτωση η αναφορά περιγράφει τις μεθόδους συμμετρικής κρυπτογραφίας, όπως φαίνεται στο παράδειγμα της Εικόνας 7, δεδομένου ότι θα πρέπει να χρησιμοποιείται αλγόριθμος ικανοποιητικής ασφάλειας, όπως ο AES, και αποφαίνεται ότι ικανοποιούνται οι δύο κανόνες που τέθηκαν, για την υλοποίηση ασφαλούς ψευδωνυμοποίησης, προτείνοντας κλειδί μεγέθους τουλάχιστον 256 bits, ώστε να καλύπτονται ικανοποιητικά οι απαιτήσεις και στην μετακβαντική περίοδο (Bernstein and Lange 2017:3). Στη συνέχεια αναφέρεται ότι και οι μέθοδοι ασύμμετρης κρυπτογραφίας, όπως φαίνεται στο παράδειγμα της Εικόνας 8, θα μπορούσε σε ορισμένες περιπτώσεις να χρησιμοποιηθούν για την ψευδωνυμοποίηση των προσωπικών δεδομένων και μάλιστα ότι η χρήση πιθανοτικής ασύμμετρης κρυπτογράφησης θα μπορούσε να προσδώσει τη δυνατότητα παραγωγής διαφορετικών ψευδωνύμων για το ίδιο αρχικό αναγνωριστικό. Επίσης, όμως, θα πρέπει να ληφθεί υπόψιν ότι οι μέθοδοι αυτές απαιτούν πολύ μεγάλο μέγεθος κλειδιού, για να παρέχουν ασφάλεια, και ότι τόσο ο RSA όσο και οι αλγόριθμοι κρυπτογράφησης ελλειπτικών καμπυλών (οι πλέον γνωστές τεχνικές ασύμμετρης κρυπτογράφησης) δεν θα είναι ασφαλείς σε μία μετακβαντική περίοδο (Bernstein and Lange 2017:3).



Εικόνα 7. Παράδειγμα χρήσης συμμετρικής κρυπτογραφίας. (ENISA, 2018:25)

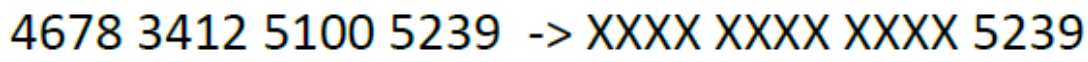


Εικόνα 8. Παράδειγμα χρήσης ασύμμετρης κρυπτογραφίας. (ENISA, 2018:27)

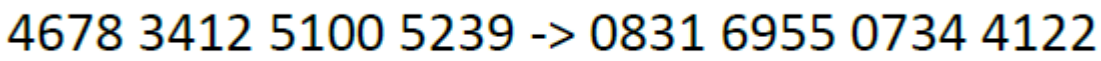
- Άλλες κρυπτογραφικές μέθοδοι: Σε αυτό το σημείο η αναφορά επικεντρώνεται σε μεθόδους που είναι υπό μελέτη και αναφέρει ως πιθανές προτάσεις την ψευδωνυμοποίηση πολυμορφικής κρυπτογραφίας (Eric Verheul et al 2016:6) και την προσέγγιση μιας αποκεντρωμένης παραγωγής ψευδωνύμων (Lehnhardt and Spalka 2011:117) στην οποία το ίδιο το πρόσωπο στο οποίο αφορούν τα δεδομένα που πρέπει να προστατευθούν παράγει τα ψευδώνυμα που θα χρησιμοποιηθούν και ο υπεύθυνος των δεδομένων μπορεί να αναγνωρίσει το πρόσωπο μόνο με την άδεια αυτού.

- Ετικέτα μιας χρήσης (tokenisation): Η αναφορά του Οργανισμού της Ευρωπαϊκής Ένωσης για την Ασφάλεια Δικτύων και Πληροφοριών επαναφέρει την πρόταση της Ομάδας Εργασίας του Άρθρου 29 και κρίνει ότι ικανοποιούν τους δύο κανόνες που τέθηκαν, αλλά εξαιτίας του περιορισμού της χρήσης τους σε συγκεκριμένες περιπτώσεις αρκετά δύσκολο, αποφαίνεται ότι είναι προτιμότερο να χρησιμοποιηθούν οι μέθοδοι που αναφέρθηκαν ανωτέρω.

- Άλλες προσεγγίσεις: Σε αυτή την περίπτωση αναφέρεται η χρήση της κάλυψης τμήματος του αναγνωριστικού όπως φαίνεται στο παράδειγμα της Εικόνας 9, της αλλαγής της θέσης των χαρακτήρων που απαρτίζουν το αναγνωριστικό, όπως φαίνεται στο παράδειγμα της Εικόνας 10 και της προσεγγυστικής μεθόδου των δεδομένων, έτσι ώστε να μειώνεται η ακρίβεια των δεδομένων, τεχνική που συχνά συναντάται σε εικόνες, όπως φαίνεται στο παράδειγμα της Εικόνας 11. Όμως, όλες θεωρούνται αδύναμες στον να προσφέρουν ικανοποιητική ασφάλεια.



Εικόνα 9. Παράδειγμα κάλυψης τμήματος αναγνωριστικού. (ENISA, 2018:29)



Εικόνα 10. Παράδειγμα αλλαγής θέσης των χαρακτήρων του αναγνωριστικού. (ENISA, 2018:29)

Εικόνα 11. Παράδειγμα προσεγγιστικής μεθόδου σε εικόνα.

3.3.2 Αναφορά Νοεμβρίου 2019

Σε συνέχεια της προηγούμενης αναφοράς ο ENISA επανέρχεται στο θέμα των προτεινόμενων μεθόδων για την προστασία των προσωπικών δεδομένων με βάση το Γενικό Κανονισμό για την Προστασία Δεδομένων δημοσιεύει την αναφορά (ENISA, 2019: 5), μελετά περεταίρω την έννοια της ψευδωνυμοποίησης και τους τρόπους με τους οποίους είναι δυνατόν να υλοποιηθεί, επικεντρώνοντας σε συγκεκριμένες εφαρμογές και παραδείγματα.

Ιδιαίτερη σημασία δίνεται, σε αυτή την αναφορά, στις μεθόδους ψευδωνυμοποίησης ενός μοναδικού αναγνωριστικού, δηλαδή στην παραγωγή ψευδωνύμου από ένα αναγνωριστικό. Ως μοναδικό αναγνωριστικό μπορεί να θεωρηθεί οποιοδήποτε στοιχείο πληροφορίας που φανερώνει άμεσα την ταυτότητα ενός προσώπου μέσα σε ένα πλαίσιο επεξεργασίας δεδομένων και είναι διαφορετικό από τα αντίστοιχα όλων των υπολοίπων προσώπων. Αυτό, μπορεί να χρησιμοποιηθεί σαν μέσω ταυτοποίησης και συνδέεται με τα υπόλοιπα δεδομένα του προσώπου. Ένα παράδειγμα τέτοιου αναγνωριστικού θα μπορούσε να είναι ο αριθμός ταυτότητας ή ο αριθμός φοιτητικής ταυτότητας (το Πανεπιστήμιο αναγνωρίζει έναν φοιτητή από αυτόν).

Οι μέθοδοι που αναφέρονται είναι οι εξής:

- Αύξων αριθμός: Σε αυτή, το αναγνωριστικό αντικαθίσταται από έναν αριθμό που εξαρτάται από τη θέση του αναγνωριστικού στη λίστα με τα υπόλοιπα αναγνωριστικά και απαιτεί την ύπαρξη πίνακα αντιστοίχισης που να επιτρέπει την αντιστροφή της διαδικασίας. Η συγκεκριμένη μέθοδος θεωρείται πολύ απλή στην υλοποίηση, δεν σχετίζεται το αρχικό αναγνωριστικό με τη νέα τιμή που το αντικαθιστά, αλλά μπορεί να εμφανίζει πρόβλημα σε πολύπλοκα σύνολα δεδομένων με μεγάλο μέγεθος, καθώς ο πίνακας αντιστοίχισης απαιτείται να αποθηκευτεί.

- Τυχαίος αριθμός: Αυτή η μέθοδος είναι όμοια με αυτή του αύξοντος αριθμού με τη διαφορά ότι ο αριθμός που αντικαθιστά το αναγνωριστικό προέρχεται από μία μηχανή παραγωγής τυχαίων αριθμών. Σε αυτή την περίπτωση εκτός από το μειονέκτημα που εμφανίζεται και στην προηγούμενη μέθοδο, ειδική μέριμνα θα πρέπει να εφαρμοστεί, έτσι ώστε να αποφευχθούν οι «συγκρούσεις», η ανάθεση ίδιου τυχαίου αριθμού σε διαφορετικά αναγνωριστικά.

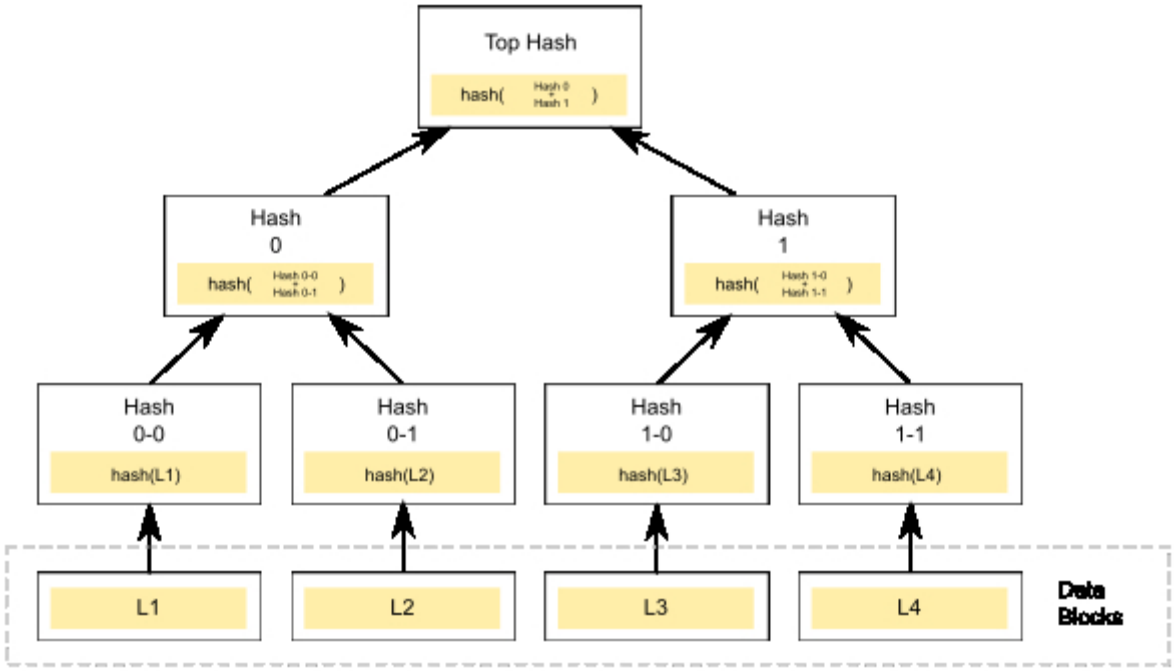
- Κρυπτογραφική συνάρτηση κατακερματισμού: Είναι η ίδια μέθοδος που είχε περιγραφεί και στην προηγούμενη αναφορά του Νοεμβρίου 2018 και αναφέρεται η αδυναμία που μπορεί να παρουσιάσει απέναντι σε επιθέσεις εξαντλητικών δοκιμών, με αποτέλεσμα να μην προτείνεται σαν ικανοποιητική μέθοδος ψευδωνυμοποίησης.

- Κώδικας αυθεντικοποίησης μηνύματος (MAC): Είναι η ίδια μέθοδος με την συνάρτηση κατακερματισμού με χρήση κλειδιού, που είχε περιγραφεί και αυτή στην προηγούμενη αναφορά του Νοεμβρίου 2018 και θεωρείται αξιόπιστη μέθοδος κρυπτογράφησης, όσο το κλειδί παραμένει κρυφό.

- Συμμετρική κρυπτογραφία: Και σε αυτή τη μέθοδο είχε αναφερθεί η προηγούμενη αναφορά του Νοεμβρίου 2018 και είχε καταλήξει ότι με κατάλληλο μέγεθος του κλειδιού αποτελεί μία αξιόπιστη μέθοδο ψευδωνυμοποίησης. Το νέο στοιχείο που προστίθεται είναι ότι για τη χρήση της στην κρυπτογράφηση του αναγνωριστικού, αν το μέγεθός του είναι μικρότερο του κλειδιού απαιτείται να συμπληρωθεί καταλλήλως (van Tilborg and Jajodia 2011:416).

- Προηγμένες μέθοδοι ψευδωνυμοποίησης: Σε αυτή την περίπτωση προτείνεται η μελέτη της δυνατότητα χρήσης των δέντρων τύπου Merkle για την ψευδωνυμοποίηση και των αλυσιδωτών συναρτήσεων κατακερματισμού, καθώς η πολυπλοκότητα που εισάγουν οι

συνεχείς χρήσεις τους καθιστούν εξαιρετικά δύσκολη έως αδύνατη την αντιστροφή της διαδικασίας, όπως φαίνεται στο παράδειγμα της Εικόνας 12. Ωστόσο, αναφέρονται ως πιθανές μελλοντικές ερευνητικές κατευθύνσεις, χωρίς να παρέχεται κάποιο παράδειγμα χρήσης τους.



Εικόνα 12. Παράδειγμα δέντρου τύπου Merkle.

3.4 Συμπερασματικά

Η εφαρμογή του Γενικού Κανονισμού για την Προστασία Δεδομένων απαιτεί την εφαρμογή ενός αυστηρού πλαισίου προστασία των προσωπικών δεδομένων με αποτέλεσμα ο ENISA να προχωρεί σε προτάσεις τόσο όσον αφορά στις υφιστάμενες μεθόδους που μπορούν να χρησιμοποιηθούν για την επίτευξη αυτής της ασφάλειας, αλλά και σε πιθανές κατευθύνσεις που μπορούν να ακολουθηθούν, για την υλοποίηση νέων.

Από τις προτεινόμενες μεθόδους αυτές που έχουν χαρακτηριστεί ικανοποιητικές είναι οι συναρτήσεις κατακερματισμού με χρήση κλειδιού, οι συμμετρικής και μη συμμετρικής κρυπτογραφίας. Οι αλγόριθμοι που τις υλοποιούν είναι αρκετά γνωστοί και έχουν μελετηθεί διεξοδικά.

Για το λόγο αυτό η παρούσα μεταπτυχιακή διατριβή θα ασχοληθεί με τη μελέτη προηγμένων τεχνικών κρυπτογράφησης, κάτι που αποτελεί ήδη έναν ανοιχτό ερευνητικό κλάδο και εμπίπτει και στις προτάσεις του ENISA, προκειμένου να επιτευχθεί ψευδωνυμοποίηση η οποία να έχει ιδιότητες που επιβάλλονται από το νομικό πλαίσιο προστασίας προσωπικών δεδομένων, αλλά που όμως δεν μπορούν να επιτευχθούν με τις κλασικές τεχνικές ψευδωνυμοποίησης. Έτσι, θα επικεντρωθεί στην πρόταση της χρήσης δέντρων τύπου Merkle, μια και η προαναφερθείσα ψευδωνυμοποίηση πολυμορφικής κρυπτογραφίας είναι ακόμα στη φάση της ανάπτυξης και της έρευνας και αναμένεται να έχει οριστικά αποτελέσματα το 2021 (E. Verheul and Jacobs 2017:7).

Κεφάλαιο 4

Τεχνική ψευδωνυμοποίησης βασισμένη σε δέντρα Merkle

Τα δέντρα τύπου Merkle αποτέλεσαν την πρόταση βελτιστοποίησης της μεθόδου ψηφιακής υπογραφής μιας χρήσης των Lamport-Diffie (Diffie and Hellman 1976:649) από την πρόταση του Ralph Merkle (Ralph Charles Merkle 1979:48, Ralph C. Merkle 1988:373) και ο βασικός μηχανισμός τους χρησιμοποιείται κατά κύριο λόγω σε κρυπτονομίσματα (Bitcoin) (Koblitz and Menezes 2016:96, Nakamoto2009:4).

Στο παρόν Κεφάλαιο θα επιχειρηθεί η περιγραφή μιας μεθόδου ψευδωνυμοποίησης μοναδικού αναγνωριστικού για πολλαπλές οντότητες χρησιμοποιώντας το βασικό μηχανισμό των δέντρων τύπου Merkle. Απώτερος σκοπός είναι η εύρεση μίας τεχνικής η οποία θα επιτρέπει σε ένα χρήστη (υποκείμενο των δεδομένων) να δημιουργεί με δικά του μέσα ένα ψευδώνυμο, διαφορετικό ανά οργανισμό, με τρόπο τέτοιο ώστε ο κάθε οργανισμός να μπορεί επιβεβαιώνει ότι το ψευδώνυμο ανήκει πράγματι σε αυτόν χωρίς αποκάλυψη καμίας πρόσθετης προσωπικής πληροφορίας.

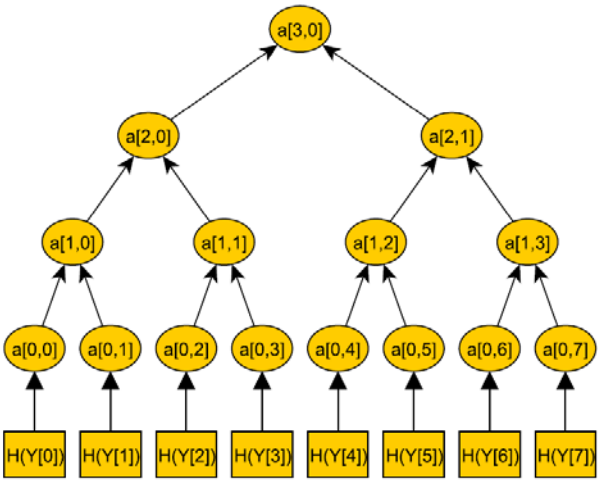
4.1 Βασικός Τρόπος Λειτουργίας Υπογραφής Σχήματος Merkle

Κρίνεται, λοιπόν, σκόπιμο να μελετηθεί η βασική λειτουργία της υπογραφής σχήματος Merkle σε αυτή την Ενότητα, πριν τη μέθοδο που θα προτείνουμε.

4.1.1 Δημιουργία Δέντρου Τύπου Merkle και Δημόσιου Κλειδιού

Το σχήμα ψηφιακής υπογραφής Merkle μπορεί να χρησιμοποιηθεί για περιορισμένο αριθμό μηνυμάτων με τη χρήση ενός δημόσιου κλειδιού. Ο αριθμός αυτός είναι δύναμη του δύο και άρα το πλήθος των πιθανών μηνυμάτων μπορεί να είναι $N=2^n$. Σε πρώτο βήμα δημιουργούνται X_i δημόσια κλειδιά και Y_i ιδιωτικά κλειδιά από N ψηφιακές υπογραφές μιας χρήσης με τη μέθοδο Winternitz One-time Signature Scheme (Ralph C. Merkle 1988:372) που είναι βελτίωση Lamport One-Time Signature Scheme (Diffie and Hellman 1976:649). Για κάθε Y_i, όπου $1 \leq i \leq 2^n$, υπολογίζεται και μία τιμή συνάρτησης κατακερματισμού $h_i = H(Y_i)$. Με το σύνολο αυτών των τιμών δημιουργείται το δέντρο τύπου Merkle, όπου η κάθε μία αποτελεί φύλλο του δέντρου. Σε αυτό $\alpha_{i,j}$ είναι οι κόμβοι του και το i υποδεικνύει το επίπεδο του δέντρου και η τιμή του εξαρτάται από την απόστασή του από τα φύλλα. Όπως είναι φυσικό, ένα τέτοιο δέντρο θα έχει n+1 επίπεδα και για το επίπεδο των φύλλων θα ισχύει i=0, ενώ για τη ρίζα του i=n. Η τιμή του δείκτη j ξεκινάει από την τιμή 0 και αυξάνεται από τα αριστερά προς τα δεξιά, σε κάθε επίπεδο j. Αρχικά, στο πρώτο επίπεδο του δέντρου αποδίδονται οι τιμές της συνάρτησης κατακερματισμού και ισχύει $\alpha_{0,i}= h_i = H(Y_i)$. Ο κάθε εσωτερικός κόμβος του δέντρου τύπου Merkle, μια και είναι δυαδικό, παράγεται από την εφαρμογή της συνάρτησης κατακερματισμού στη συνένωση των δύο παιδιών του και άρα $\alpha_{i,j}=h(\alpha_{i-1,2*j} || \alpha_{i-1,2*j+1,})$. Έτσι, το δέντρο αποτελείται από 2^n φύλλα, περιλαμβάνει n+1 επίπεδα και 2^{n+1}-1 κόμβους συνολικά με $\alpha_{n,o}$ τη ρίζα του δέντρου που αποτελεί το δημόσιο κλειδί του σχήματος υπογραφής Merkle.

Ένα παράδειγμα αυτής της λειτουργίας για αριθμό φύλλων 8 παρουσιάζεται στην Εικόνα 13 που ακολουθεί:

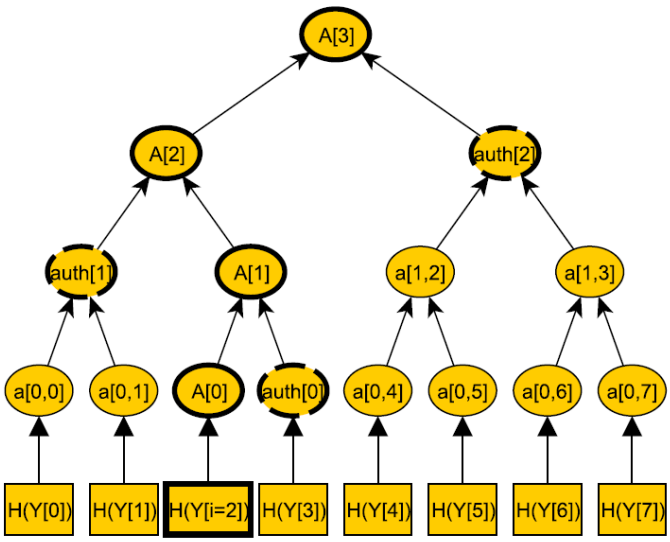


Εικόνα 13. Δέντρο τύπου Merkle με 8 φύλλα.

4.1.2 Δημιουργία της Υπογραφής

Για την υπογραφή ενός μηνύματος M με το σχήμα Merkle, αυτό υπογράφεται με τη χρήση ενός ζεύγους εκ των δημόσιων και ιδιωτικών κλειδιών (X_i, Y_i) δημιουργώντας την υπογραφή υ'. Το αντίστοιχο φύλλο του δέντρου με το δημόσιο κλειδί μιας χρήσης Y_i είναι $\alpha_{i,0}= H(Y_i)$. Για να μπορέσει ο παραλήπτης του μηνύματος να επιβεβαιώσει ότι η υπογραφή μιας χρήσης που χρησιμοποιήθηκε προέρχεται από το σύνολο υπογραφών του αποστολέα θα πρέπει να μπορέσει να το συσχετίσει με τη δημόσια υπογραφή που είναι η ρίζα του δέντρου. Για να επιτευχθεί αυτό απαραίτητο είναι να του παρασχεθεί η ελάχιστη πληροφορία για τους κόμβους του δέντρου, ώστε χρησιμοποιώντας τους να καταλήξει στο δημόσιο κλειδί του αποστολέα. Άρα θα χρειαστεί το $A = \{A_0, A_1, \ldots, A_n\}$ μονοπάτι με τους ενδιάμεσους κόμβους. Αυτό αποτελείται από n+1 κόμβους ξεκινώντας από το φύλλο $A_0= \alpha_{0,i}$ και να καταλήξει στο $A_n= \alpha_{n,0}$. Κάθε εσωτερικός κόμβος A_{i+1} παρασκευάζεται από τα παιδιά του, θα απαιτηθεί η γνώση του δυαδικού «αδερφού» του A_i που θα ονομαστεί $auth_i$. Έτσι, $A_{i+1}=H(A_i|| auth_i)$ και άρα θα χρειαστούν n κόμβοι, για να μπορέσει ο παραλήπτης του μηνύματος να υπολογίσει το A_n. Το σύνολο αυτών των κόμβων μαζί με την υπογραφή μιας χρήσης υ' αποτελούν την υπογραφή σχήματος Merkle $\upsilon= \{\upsilon', auth_0, auth_1, ..., auth_{n-1}\}$.

Ένα παράδειγμα της διαδικασίας υπολογισμού των απαραίτητων κόμβων για το δέντρο τύπου Merkle των 8 φύλλων για την υπογραφή Y_2 φαίνεται στην Εικόνα 14 παρακάτω, όπου οι κόμβοι του μονοπατιού επαλήθευσης της υπογραφής – η οποία περιγράφεται στην αμέσως επόμενη Ενότητα - φαίνονται με διακεκομμένο έντονο περίγραμμα:



Εικόνα 14. Διαδρομή σε δέντρο τύπου Merkle 8 φύλλων.

4.1.3 Επιβεβαίωση της Υπογραφής

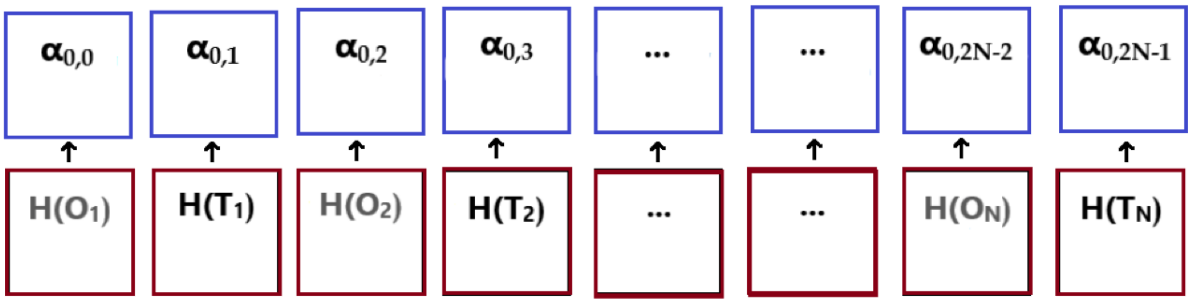
Ο παραλήπτης γνωρίζει το μήνυμα M, το δημόσιο κλειδί A_n και την υπογραφή υ= {υ', $auth_0$,$auth_1$, ..., $auth_{n-1}$}. Αρχικά, επιβεβαιώνει την υπογραφή μιας χρήσης υ' του μηνύματος M και σε περίπτωση που ταιριάζει συνεχίζει για την επιβεβαίωση της υπογραφής με το δημόσιο κλειδί. Στην αρχή, υπολογίζει από το $A_0 = H(Y_i)$ και στη συνέχεια για κάθε $A_i = H(A_{i-1} || auth_{i-1})$. Έτσι, για το παράδειγμα της Εικόνας 10 ο παραλήπτης είναι σε θέση να υπολογίσει το A_1, στη συνέχεια μαζί με το $auth_1$ να υπολογίσει το A_2= H($auth_1$|| A_1) και με παρόμοιο τρόπο να καταλήξει στο A_3= H(A_2 || $auth_2$). Μόλις υπολογίσει το A_n ελέγχει την εγκυρότητά του με το δημόσιο κλειδί.

4.2 Πρόταση Ψευδωνυμοποίησης με Δέντρα Τύπου Merkle

Στη συνέχεια θα επιχειρηθεί να προσδιοριστεί ο τρόπος με τον οποίο θα μπορούσε να χρησιμοποιηθεί η τεχνοτροπία των δέντρων τύπου Merkle, για την ψευδωνυμοποίηση μοναδικών αναγνωριστικών, που συνδέονται με σύνολα δεδομένων που πρέπει να προστατευθούν. Για την ανάγκη αυτής της μελέτης θα οριστεί το όνομα οντότητα, για να περιγραφεί κάθε υπηρεσία, οργανισμός, επιχείρηση κλπ, που συγκεντρώνει νόμιμα δεδομένα προσώπων και αποτελεί σύμφωνα με το Γενικό Κανονισμό για την Προστασία Δεδομένων τον υπεύθυνο επεξεργασίας δεδομένων. Ως υποκείμενο θα οριστεί το πρόσωπο που παρέχει τα δεδομένα του στην κάθε οντότητα. Ως μοναδικό αναγνωριστικό ορίζεται μία τιμή, όχι απαραίτητα αριθμητική, που αποδίδει η κάθε οντότητα σε κάθε υποκείμενο και είναι διαφορετικές μεταξύ τους. Τέλος θα πρέπει να αναφερθεί ότι στη μέθοδο αυτή το υποκείμενο δημιουργεί το ψευδώνυμό του σε κάθε οντότητα και συνδέεται με το μοναδικό αναγνωριστικό που του έχει αποδοθεί από αυτή. Στο σενάριό μας θεωρούμε ότι το μοναδικό αναγνωριστικό που αντιστοιχεί για το πρόσωπο (υποκείμενο των δεδομένων) σε κάθε οντότητα (υπεύθυνο επεξεργασίας) δεν πρέπει – εκ της νομοθεσίας περί προσωπικών δεδομένων – να κοινοποιηθεί σε άλλη οντότητα (παράδειγμα: η μία οντότητα μπορεί να είναι μία Υπηρεσία για φορολογικούς σκοπούς και το αναγνωριστικό να είναι ο μοναδικός αριθμός φορολογικού μητρώου, ο οποίος όμως δεν πρέπει να καταστεί γνωστός σε άλλη οντότητα).

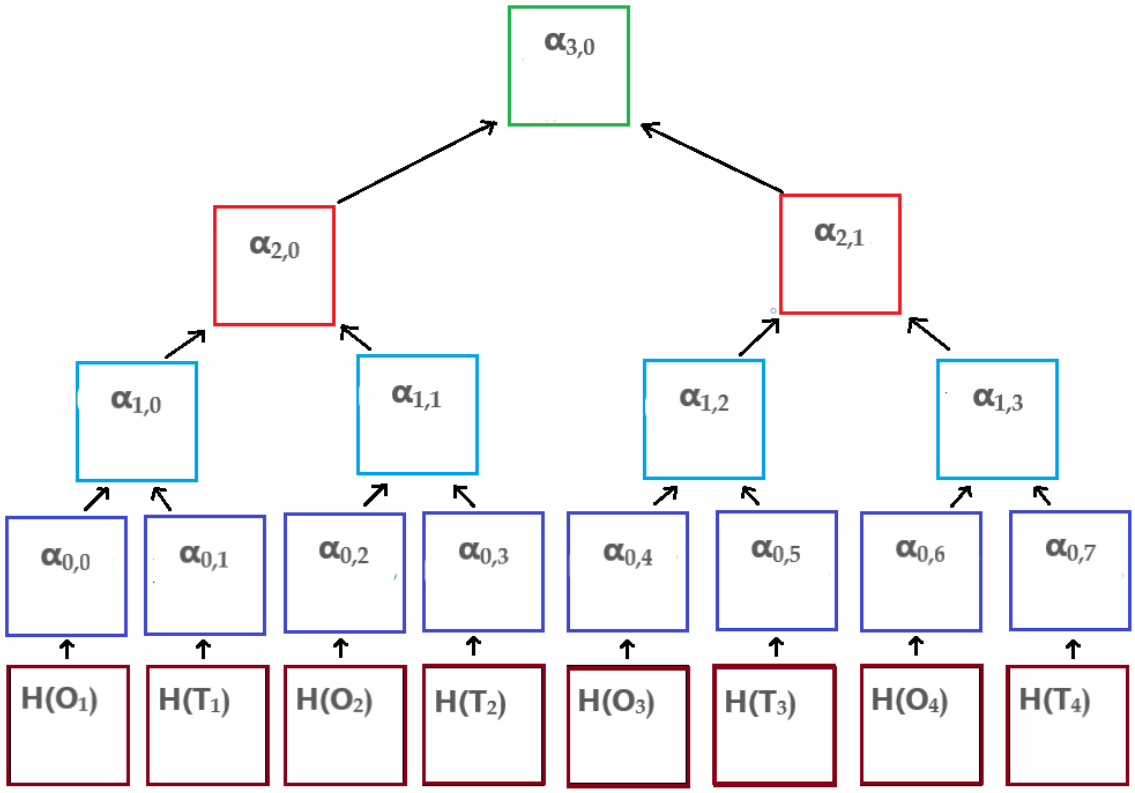
4.2.1 Δημιουργία Ψευδώνυμου

Ένα υποκείμενο μπορεί να έχει δώσει δεδομένα του, όχι τα ίδια, σε N οντότητες. Έτσι, θα έχει λάβει από την κάθε μία ένα μοναδικό αναγνωριστικό και άρα έχει στην κατοχή του ένα σύνολο από N αναγνωριστικά O_i, i=1,2,…,N, που είναι μοναδικά για αυτό και αντιστοιχεί το καθένα σε μία οντότητα και τα δεδομένα που της έχει παραχωρήσει $\{O_1, O_2, \ldots, O_N\}$. Στη συνέχεια παράγει N ιδιωτικά κλειδιά T, διαφορετικά μεταξύ τους $\{T_1, T_2, \ldots, T_N\}$, και αντιστοιχεί ένα σε κάθε μοναδικό αναγνωριστικό, έτσι ώστε να δημιουργούν μοναδικά ζεύγη $\{(O_1, T_1) (O_2, T_2), \ldots, (O_N, T_N)\}$. Αυτή η διάταξη που δημιουργείται θα αποτελέσει τη βάση για τη δημιουργία των φύλλων του δέντρου τύπου Merkle, το οποίο, όπως είναι προφανές απαιτείται να έχει ν=2*N φύλλα και θα είναι $\{H(O_1), H(T_1), H(O_2), H(T_2), \ldots, H(O_N), H(T_N)\}$, όπου H η εφαρμογή συνάρτησης κατακερματισμού σε κάθε στοιχείο του συνόλου δημιουργώντας $\{h_0, h_1, h_2, \ldots, h_{v-1}\}$.



Εικόνα 15. Δημιουργία των φύλλων

Στη συνέχεια ακολουθείται η διαδικασία που περιγράφηκε ανωτέρω και έτσι παράγονται οι κόμβοι του κατώτερου επιπέδου του δέντρου (τα φύλλα) με $\alpha_{0,i}=h_i$, όπως φαίνεται στην Εικόνα 15. Οι κόμβοι των ανώτερων επιπέδων του δέντρου δημιουργούνται από την εφαρμογή της συνάρτησης κατακερματισμού στη συνένωση των δύο παιδιών τους στο δυαδικό δέντρο $\alpha_{i,j}=H(\alpha_{i-1,2*j} || \alpha_{i-1,2*j+1,})$, παράγοντας με αυτόν τον τρόπο ένα δέντρο τύπου Merkle με 2*N φύλλα, $\log_2 2N+1$ επίπεδα. Ένα παράδειγμα αυτής της διαδικασίας για την περίπτωση τεσσάρων οντοτήτων φαίνεται στην Εικόνα 16 παρακάτω:



Εικόνα 16. Παράδειγμα δημιουργίας δέντρου τύπου Merkle από τέσσερις οντότητες.

Με την ολοκλήρωση αυτής της διαδικασίας παράγεται η ρίζα του δέντρου που θα αποτελεί το ψευδώνυμο. Για κάθε οντότητα, όπως θα αναλυθεί στη συνέχεια, θα παραχθεί διαφορετικό ψευδώνυμο Ψ_i, αφού θα δημιουργηθεί διαφορετικό δέντρο Merkle.

4.2.2 Γνωστοποίηση και Επιβεβαίωση Ψευδώνυμου

Όπως έχει αναφερθεί η παραγωγή του ψευδωνύμου είναι διαδικασία την οποία εκτελεί το υποκείμενο, ενώ η κάθε οντότητα έχει μόνο τη γνώση του μοναδικού αναγνωριστικού O_k με το οποίο συνδέει την ταυτότητα του υποκειμένου με τα υπόλοιπα δεδομένα. Σε αυτό το σημείο η ψευδωνυμοποίηση δεν έχει ολοκληρωθεί.

Η γνωστοποίηση του ψευδωνύμου από το υποκείμενο προς την οντότητα πραγματοποιείται με παρόμοιο τρόπο με αυτόν που παράγεται η υπογραφή στο σχήμα υπογραφής Merkle. Με δεδομένη τη γνώση του μοναδικού αναγνωριστικού από την οντότητα, το υποκείμενο, από το δέντρο τύπου Merkle, που δημιούργησε, παράγει το μονοπάτι των ενδιάμεσων κόμβων που απαιτείται για την παραγωγή της ρίζας του δέντρου από το φύλλο στο οποίο έχει γνώση η

οντότητα. Αυτό είναι, για μοναδικό αναγνωριστικό O_k, το $\alpha_{0,2k-2}$, το οποίο παράγεται από μία απλή εφαρμογή της συνάρτησης κατακερματισμού $H(O_k)$.

Με αυτόν τον τρόπο το υποκείμενο θα αποστείλει στην οντότητα μία σειρά από κόμβους, εκ των οποίων ο τελευταίος θα είναι η ρίζα του δέντρου και το ψευδώνυμο που δημιούργησε. Άρα θα χρειαστεί το Α μονοπάτι με τους ενδιάμεσους κόμβους. Αυτό αποτελείται από, $n= \log_2 2N +1$ κόμβους ξεκινώντας από το φύλλο $A_0= \alpha_{0,2k-1}$ και να καταλήξει στο $A_n= \alpha_{n,0}$, στη ρίζα του δέντρου. Κάθε εσωτερικός κόμβος A_{i+1} παρασκευάζεται από τα παιδιά του, θα απαιτηθεί η γνώση του δυαδικού «αδερφού» του A_i που θα ονομαστεί $node_i$. Έτσι, ο κάθε κόμβος $A_{i+1}=H(A_i|| node_i)$ και άρα θα χρειαστούν n κόμβοι, για να μπορέσει ο παραλήπτης του μηνύματος να υπολογίσει το A_n. Το σύνολο της πληροφορίας που αποστέλλεται είναι $\{ \alpha_{0,2k-1}, node_1, node_2, \ldots, node_{n-1}, A_n \}$.

Στην συνέχεια, η οντότητα υπολογίζει το $H(O_k)$, που είναι ίσο με $\alpha_{0,2k-2}$, ώστε να εκτελέσει την πράξη $A_1 = (H(O_k)|| \alpha_{0,2k-1})$ και σταδιακά για κάθε $A_i = H(A_{i-1} || node_{i-1})$, όπου $i\leq n$. Μόλις υπολογίσει το A_n ελέγχει αν αυτό που έχει υπολογίσει είναι ίσο με το τελευταίο στοιχείο του συνόλου της πληροφορίας που έχει λάβει. Αν είναι ίσο τότε έχει γνωρίσει το ψευδώνυμο του υποκειμένου και επιβεβαιώσει ότι προήλθε από το μοναδικό αναγνωριστικό που αντιστοιχεί στο υποκείμενο. Αν δεν είναι ίσο τότε αγνοεί την πληροφορία. Με άλλα λόγια, η i-ιοστή οντότητα είναι σε θέση να επικυρώσει ότι το ψευδώνυμο Ψ_i που της παρέχει ο χρήστης προέρχεται πράγματι από το αναγνωριστικό Oi με το οποίο η ίδια «αναγνωρίζει» το χρήστη. Είναι σημαντικό να τονιστεί ότι το ψευδώνυμο «προέρχεται» (εξαρτάται) από όλα τα αναγνωριστικά που έχει ο χρήστης σε όλες τις οντότητες – ωστόσο αυτά δεν διαρρέουν (δηλαδή δεν αποκαλύπτονται), ούτε μπορούν να υπολογιστούν από το ψευδώνυμο αυτό καθ' αυτό (ως άμεση απόρροια των ιδιοτήτων των συναρτήσεων κατακερματισμού).

4.2.3 Πολλαπλά Δέντρα

Στην Υποενότητα 4.2.2 περιγράφηκε ο βασικός τρόπος με τον οποίο παράγεται ένα ψευδώνυμο με ένα δέντρο τύπου Merkle από τη συνολική πληροφορία του υποκειμένου των μοναδικών αναγνωριστικών του σε N οντότητες $\{O_1, O_2, \ldots, O_N\}$ και από τα ιδιωτικά κλειδιά του $\{ T_1, T_2, \ldots, T_N\}$, με την κάθε οντότητα να γνωρίζει μόνο το δικό της O και καμία από τις υπόλοιπες πληροφορίες.

Αν η διαδικασία αυτή ακολουθηθεί για κάθε οντότητα με ένα μόνο δέντρο, κάθε φορά το ίδιο υποκείμενο θα καταλήγει να έχει το ίδιο ψευδώνυμο σε όλες τις οντότητες. Αυτό αποτελεί μία

ευκολία, αλλά είναι απειλή καθώς η κάθε οντότητα θα είναι σε θέση να γνωρίζει το ψευδώνυμο του υποκειμένου, που συνδέεται με τα δεδομένα που έχει παραχωρήσει και πρέπει να είναι προστατευμένα, και στις υπόλοιπες οντότητες. Με άλλα λόγια, θα είναι εύκολη η «σύζευξη» (linking) των δύο βάσεων δεδομένων που τηρούν δύο οποιεσδήποτε οντότητες, βάσει του κοινού ψευδωνύμου.

Για να αποφευχθεί αυτή η συνδεσιμότητα, η οποία σε ορισμένες περιπτώσεις δεν είναι επιτρεπτή σύμφωνα με το Γενικό Κανονισμό για την Προστασία Δεδομένων (και το σενάριό μας θεωρούμε ότι εμπίπτει σε μία τέτοια περίπτωση), προκρίνεται η χρήση διαφορετικών δέντρων για την παραγωγή ψευδωνύμων για κάθε οντότητα, που θα προέρθουν από την χρήση διαφορετικών φύλλων. Με δεδομένο ότι το σύνολο των μοναδικών αναγνωριστικών είναι μοναδικό για κάθε υποκείμενο $\{O_1, O_2, \ldots, O_N\}$ και αποτελεί χαρακτηριστικό του, η αλλαγή θα πρέπει να προέλθει από διαφορετικό σύνολο ιδιωτικών κλειδιών $\{ T_1, T_2, \ldots, T_N\}$. Αφού το δέντρο τύπου Merkle προέρχεται από το συνδυασμό των δύο αυτών συνόλων $\{ O_1, T_1, O_2, T_2, \ldots, O_N, T_N \}$, η κάθε αλλαγή στο $\{ T_1, T_2, \ldots, T_N\}$ θα προκαλεί αλλαγή και στο ψευδώνυμο που παράγεται. Έτσι, για κάθε οντότητα το υποκείμενο θα χρησιμοποιεί διαφορετικό σύνολο ιδιωτικών κλειδιών και άρα θα παράγει διαφορετικά δέντρα τύπου Merkle και ψευδώνυμα.

4.2.4 Επικοινωνία Οντοτήτων

Στο μηχανισμό που έχει περιγραφεί μέχρι στιγμής δεν είναι δυνατό οι οντότητες να γνωρίζουν το ψευδώνυμο του υποκειμένου στις υπόλοιπες και έτσι δεν είναι δυνατό να έχουν πρόσβαση στα δεδομένα αυτά. Στην πράξη όμως, δεν αποκλείεται η ανταλλαγή πληροφοριών να είναι επιτρεπτή (για παράδειγμα, θα μπορούσε να είναι επιτρεπτή με τη συναίνεση του ατόμου). Για το λόγο αυτό, θα παρουσιαστεί ένας τρόπος επίτευξης της επικοινωνίας των οντοτήτων για ανταλλαγή δεδομένων, χωρίς να παραβιάζεται η ιδιωτικότητα του υποκειμένου, με τη χρήση των ήδη υπαρχόντων μηχανισμών.

Ένα βασικό χαρακτηριστικό του ανωτέρου μηχανισμού είναι ότι οι οντότητες έχουν ελάχιστη γνώση για τα δεδομένα που δημιουργούν τα ψευδώνυμα, ενώ όλη η απαραίτητη πληροφορία βρίσκεται στα χέρια του υποκειμένου. Έτσι, αν κάποια από αυτές επιθυμεί να αποκτήσει πρόσβαση στα δεδομένα μιας άλλης είναι απαραίτητο να λάβει πληροφορία από το υποκείμενο και άρα να επικοινωνήσει μαζί του γνωστοποιώντας την επιθυμία της και ζητώντας την έγκρισή του.

Έστω Χ, Ζ δύο οντότητες στις οποίες το υποκείμενο υ των δεδομένων έχει μοναδικό χαρακτηριστικό O_X και O_Z αντίστοιχα, τα οποία είμαι στοιχεία του συνόλου $\{O_1, O_2, \ldots, O_N\}$. Με τη διαδικασία που περιγράφηκε ανωτέρω στην Χ το O_X έχει ψευδωνυμοποιηθεί μέσω δέντρου τύπου Merkle σε Ψ_X με είσοδο του συνόλου $\{O_1, O_2, \ldots, O_N\}$ και του συνόλου ιδιωτικών κλειδιών του υ $\{ T_1, T_2, \ldots, T_N\}$. Ομοίως, και για την Ζ του O_Z έχει ψευδωνυμοποιηθεί σε Ψ_Z με χρήση των $\{O_1, O_2, \ldots, O_N\}$ και $\{ T_1', T_2', \ldots, T_N'\}$. Επειδή ισχύει $\{ T_1, T_2, \ldots, T_N\} \neq \{ T_1', T_2', \ldots, T_N'\}$ ισχύει και $\Psi_X \neq \Psi_Z$. Έτσι, η γνώση που έχει ο κάθε εμπλεκόμενος στη συγκεκριμένη περίπτωση είναι:

- υ: $\{O_1, O_2, \ldots, O_N\}$, $\{ T_1, T_2, \ldots, T_N\}$, $\{ T_1', T_2', \ldots, T_N'\}$, Ψ_X, Ψ_Z
- Χ: O_X, Ψ_X
- Ζ: O_Z, Ψ_Z

Οι Χ και Ζ ξέρουν και το «μονοπάτι» επικύρωσης, του κάθε δέντρου, για το αντίστοιχο ψευδώνυμο (διαφορετικό δέντρο για την κάθε οντότητα). Για λόγους απλότητας του σεναρίου θεωρούμε ότι διαγράφουν το μονοπάτι και δεν το τηρούν, χωρίς ωστόσο να αλλάζει η ουσία και τα οφέλη της περιγραφόμενης τεχνικής ψευδωνυμοποίησης ακόμα και στην περίπτωση που θεωρηθεί ότι το τηρούν. Αν η Χ χρειάζεται δεδομένα που κατέχει η Ζ θα χρειαστεί το Ψ_Z, το οποίο δεν έχει τρόπο να το παράξει και το ζητάει από το υ. Αν το υ συμφωνεί να γίνει η γνωστοποίηση των δεδομένων που κατέχει η Ζ, με τη μέθοδο που περιγράφηκε στην Υποενότητα 4.2.2 ο υ μπορεί να αποστείλει στη Χ το μονοπάτι του δέντρου τύπου Merkle που δημιούργησε το Ψ_Z το οποίο όμως σε αυτή την περίπτωση ταιριάζει με το μοναδικό αναγνωριστικό O_X αντί του O_Z.

Αν ο Χ ακολουθήσει τη διαδικασία επιβεβαίωσης θα καταλήξει στο Ψ_Z και έτσι θα μάθει την ψευδωνυμοποιημένη μορφή του O_Z. Επίσης, με αυτή την διαδικασία επιβεβαιώνεται και η ταυτότητα του υ από την Χ, αφού για να οδηγηθεί σε έγκυρη επιβεβαίωση θα πρέπει να χρησιμοποιηθούν τα σύνολα $\{O_1, O_2, \ldots, O_N\}$ και $\{ T_1', T_2', \ldots, T_N'\}$, των οποίων τη γνώση τους κατέχει μόνο ο υ, όπου τα O_X και O_Z είναι στοιχεία που αφορούν στο ίδιο υ.

Μετά την απόκτηση του Ψ_Z από το υ, η Χ μπορεί να το χρησιμοποιήσει για να γνωστοποιήσει στην Ψ ότι επιθυμεί τα δεδομένα του υ. Η γνώση του Ψ_Z από την Χ, για την Ζ αποτελεί απόδειξη συμφωνίας του υ να παραχωρηθούν τα δεδομένα του στη Χ καθώς και το απαραίτητο αναγνωριστικό για τον εντοπισμό τους ανάμεσα σε άλλα. Με αυτόν τον τρόπο επιτυγχάνεται η ανταλλαγή δεδομένων.

4.2.5 Προσθαφαίρεση Οντοτήτων

Όπως, είναι φυσικό ένα υποκείμενο δεν έχει αλληλεπίδραση με ένα σταθερό αριθμό οντοτήτων. Έτσι, στην ανωτέρω διαδικασία ψευδωνυμοποίησης θα πρέπει να υπάρχει μηχανισμός αύξησης ή μείωσης του συνόλου $\{O_1, O_2, \ldots, O_N\}$ και άρα και των συνόλων ιδιωτικών κλειδιών.

Η πρόσθεση μπορεί να πραγματοποιηθεί με την προσθήκη του νέου Ο από τη νέα οντότητα στο σύνολο των $\{O_1, O_2, \ldots, O_N\}$ και ταυτόχρονα η δημιουργία ενός νέου συνόλου ιδιωτικών κλειδιών $\{T_1, T_2, \ldots, T_N\}$ που θα αντιστοιχεί στο Ο. Η αφαίρεση με όμοιο τρόπο πραγματοποιείται με την αφαίρεση του αντίστοιχου Ο και του αντίστοιχου $\{T_1, T_2, \ldots, T_N\}$.

Όπως είναι προφανές, η αλλαγή των συνόλων αυτών προκαλεί και τροποποίηση των δέντρων τύπου Merkle που δημιουργούν τις ψευδωνυμοποιήσεις των μοναδικών αναγνωριστικών σε όλες τις οντότητες. Για το λόγο αυτό, κάθε φορά που πραγματοποιείται μία τέτοια αλλαγή είναι απαραίτητο να γίνεται ενημέρωση των οντοτήτων για τα νέα ψευδώνυμα που προέρχονται από τον εκ νέο υπολογισμό των δέντρων με τα νέα σύνολα.

4.3 Προβλήματα και λύσεις

Σε αυτό το σημείο θα προσδιοριστούν κάποιες επιπλέον λεπτομέρειες του τρόπου λειτουργίας της μεθόδου ψευδωνυμοποίησης με χρήση δέντρων τύπου Merkle.

4.3.1 Πλήθος Οντοτήτων

Ένα από τα χαρακτηριστικά των δέντρων τύπου Merkle είναι το γεγονός ότι το πλήθος των φύλλων τους είναι δύναμη του 2, δηλαδή 2^v. Στην προηγούμενη παρουσίαση του προτεινόμενου μηχανισμού για λόγους απλότητας δεν έγινε αναφορά σε αυτό το θέμα.

Στο μηχανισμό ψευδωνυμοποίησης που προτείνεται τα φύλλα ενός δέντρου τύπου Merkle προέρχονται από το σύνολο των μοναδικών αναγνωριστικών του υποκειμένου που έχει για κάθε μία από τις διάφορες οντότητες $\{O_1, O_2, \ldots, O_N\}$ και από το σύνολο των ιδιωτικών κλειδιών $\{T_1, T_2, \ldots, T_N\}$ που χρησιμοποιούνται για την παραγωγή του κάθε δέντρου $\{ O_1, T_1, O_2, T_2, \ldots, O_N, T_N \}$. Αυτό σημαίνει ότι για Ν οντότητες το δέντρο δημιουργείται από 2*Ν φύλλα.

Έτσι, αν $2N = 2^{v}$ όπου v>0 τότε το δέντρο τύπου Merkle μπορεί να δημιουργηθεί κανονικά και ο μηχανισμός που περιεγράφηκε ανωτέρω να λειτουργήσει κανονικά. Στην περίπτωση, όμως, που ισχύει $2^{v} < 2N < 2^{v+1}$ θα πρέπει να συμπληρωθεί το κενό που δημιουργείται, έτσι ώστε $2N+\chi = 2^{v+1}$. Η συμπλήρωση αυτή μπορεί να πραγματοποιηθεί με την εισαγωγή εικονικών οντοτήτων πλήθους $\psi = 2^{v} - N$. Αυτή η αναγκαία αύξηση θα οδηγήσει αναπόφευκτα και στην απαραίτητη αύξηση των ιδιωτικών κλειδιών (και σε μείωση απόδοσης του μηχανισμού, αφού παράγεται μεγαλύτερο δέντρο) αφού η ύπαρξή τους εξαρτάται από το πλήθος των οντοτήτων με σχέση ένα προς ένα για κάθε δέντρο.

Επίσης, θα πρέπει να αναφερθεί ότι, όπως είναι προφανές η ύπαρξη αυτών των εικονικών οντοτήτων δεν έχει καμία άλλη χρησιμότητα πέρα από τη συμπλήρωση των φύλλων, ώστε να δημιουργηθεί ένα δέντρο τύπου Merkle. Γι' αυτό το λόγο, η διατήρησή τους δεν είναι απαραίτητη και έτσι στην επόμενη εισαγωγή μιας πραγματικής οντότητας, στο σύστημα, αντί απλώς να προστεθεί θα αντικαταστήσει μία εικονική. Αυτή η διαδικασία έχει σαν αποτέλεσμα η μεταβολή του αριθμού των πραγματικών οντοτήτων να μην προκαλεί και αντίστοιχη μεταβολή στις διαστάσεις των νέων δέντρων που παράγονται. Για παράδειγμα, τα δέντρα που θα δημιουργηθούν από 8 πραγματικές οντότητες θα έχουν τις ίδιες διαστάσεις με αυτά που προέρχονται από 5, 6 ή 7 πραγματικές οντότητες, όπως φαίνεται και στον Πίνακα 1 παρακάτω:

Πλήθος Πραγματικών οντοτήτων	Πλήθος φύλλων	Πλήθος επιπέδων δέντρου	Πλήθος κόμβων δέντρου	Μήκος μονοπατιού ψευδωνύμου
2	4	3	7	3
2-4	8	4	15	4
5-8	16	5	31	5
9-16	32	6	63	6
17-32	64	7	127	7
33-64	128	8	255	8

65-128	256	9	511	9
129-256	512	10	1023	10

Πίνακας 1. Χαρακτηριστικών του δέντρου τύπου Merkle, ανάλογα με το πλήθος των οντοτήτων.

4.3.2 Κατεύθυνση στη Δημιουργία του Μονοπατιού

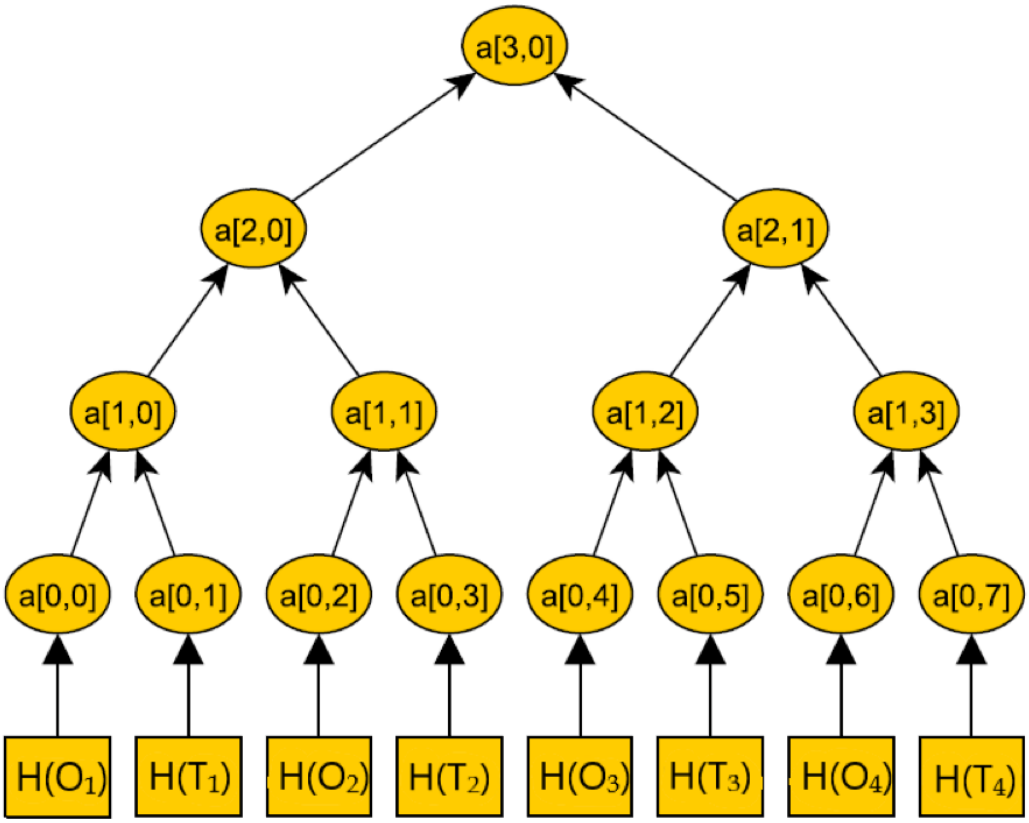
Όπως και προηγουμένως, για λόγους απλότητας, εσκεμμένα παραλείφθηκε, από την περιγραφή του βασικού μηχανισμού ψευδωνυμοποίησης με τη χρήση δέντρου τύπου Merkle και ένα ακόμα στοιχείο της λειτουργίας του που θα περιγραφεί ακολούθως.

Η δημιουργία του μονοπατιού, προς τη ρίζα του δέντρου, ξεκινάει με δεδομένο ότι το στοιχείο που θα μπορεί η οντότητα να γνωρίζει είναι πάντα το φύλλο που προέρχεται από την εφαρμογή συνάρτησης κατακερματισμού στο αντίστοιχο μοναδικό αναγνωριστικό $\alpha_{0,i}$. Έτσι, το πρώτο στοιχείο που προστίθεται για τη δημιουργία του μονοπατιού είναι το διπλανό του φύλλο, δηλαδή το $\alpha_{0,i+1}$ και με αυτόν τον τρόπο θα μπορέσει η οντότητα να παράξει τον επόμενο, γνωστό σε αυτή κόμβου του αμέσως ανώτερου επιπέδου με την πράξη $\alpha_{1,j}=H(\alpha_{0,i}|| \alpha_{0,i+1})$.

Στη συνέχεια θα πρέπει να προστεθεί στο μονοπάτι ο κατάλληλος κόμβος για το $node_{1,j}$, ώστε να παραχθεί ο $\alpha_{2,z}$. Σε αντίθεση, όμως, με την προηγούμενη περίπτωση δεν είναι απαραίτητα ο $\alpha_{1,j+1}$, μπορεί να είναι ο $\alpha_{1,j-1}$. Και έτσι να αλλάζει η θέση τους κατά τη συνένωσή τους, πριν την εφαρμογή της συνάρτησης κατακερματισμού.

Για να ολοκληρωθεί, λοιπόν, η λειτουργία της παραγωγής του μονοπατιού, πριν από κάθε επόμενο ενδιάμεσο κόμβο θα πρέπει να παρέχεται και η οδηγία αν θα πρέπει η συνένωσή του με τον γνωστό στην οντότητα κόμβο $\alpha_{i,j}$ να πραγματοποιηθεί από αριστερά ή από δεξιά. Αυτό μπορεί να αποφασιστεί με βάση τη θέση του $\alpha_{i,j}$. Αν $j \bmod 2 \neq 0$ τότε ο κόμβος που θα πρέπει να αποκαλυφθεί στο μονοπάτι είναι ο $\alpha_{i,j-1}$ με την ένδειξη ότι πρέπει να συνενωθεί από δεξιά με τον $\alpha_{i,j}$ για να παραχθεί ο επόμενος γνωστός κόμβος $\alpha_{i+1,z} = H(\alpha_{i,j-1}|| \alpha_{i,j})$, ενώ αν $j \bmod 2 = 0$ $\alpha_{i+1,z} = H(\alpha_{i,j}|| \alpha_{i,j+1})$.

Για παράδειγμα σε ένα δέντρο που προήλθε από 4 οντότητες όπως φαίνεται στην Εικόνα 17 παρακάτω:



Εικόνα 17. Παράδειγμα δέντρου τύπου Merkle από τέσσερις οντότητες.

Το μονοπάτι που αφορά στην οντότητα O_2 θα είναι {$a_{0,3}$, "right", $a_{1,0}$, "left", $a_{2,1}$, $a_{3,0}$}.

4.3.3 Πλήθος Ιδιωτικών Κλειδιών

Με δεδομένο ότι η δημιουργία των ψευδωνύμων βασίζεται στο υποκείμενο και κυρίως στην παροχή από αυτό των ιδιωτικών κλειδιών που απαιτούνται για την παραγωγή των δέντρων τύπου Merkle είναι σημαντικό να μελετηθεί αυτό το χαρακτηριστικό.

Όπως έχει περιγραφεί ανωτέρω ο μηχανισμός για ένα πλήθος πραγματικών οντοτήτων N απαιτείται N πλήθος ιδιωτικών κλειδιών για τη δημιουργία των φύλλων του κάθε δέντρου. Επιπλέον, τα ιδιωτικά αυτά κλειδιά για κάθε δέντρο πρέπει να είναι διαφορετικά μεταξύ τους αλλιώς θα είναι ευάλωτα σε επιθέσεις εξαντλητικής αναζήτησης, αφού το πρώτο στοιχείο του μονοπατιού είναι πάντα το αποτέλεσμα της εφαρμογής συνάρτησης κατακερματισμού στο ιδιωτικό κλειδί T_i που αντιστοιχίζεται με το μοναδικό αναγνωριστικό O_i. Επίσης, για κάθε δέντρο θα πρέπει το σύνολο των ιδιωτικών κλειδιών που χρησιμοποιείται να είναι διαφορετικό με το

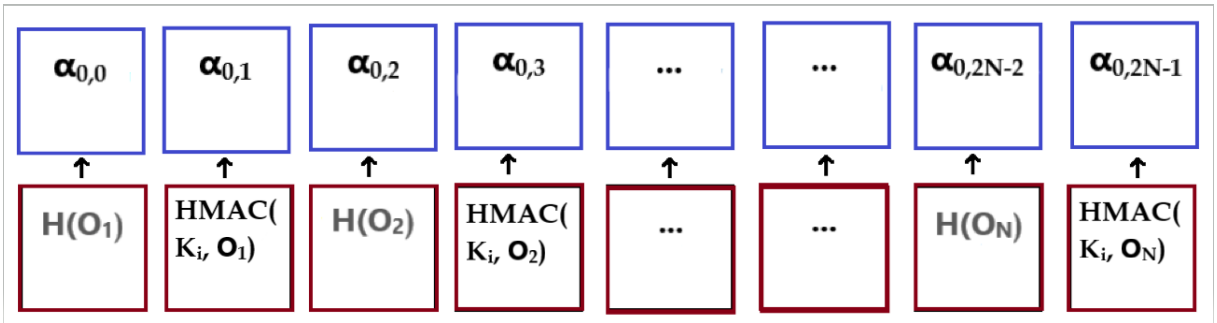
αντίστοιχο που χρησιμοποιείται για κάθε άλλο δέντρο, αλλιώς θα παράγεται το ίδιο ψευδώνυμο, και άρα κρίνονται απαραίτητα N^2 διαφορετικά κλειδιά. Ένα ακόμη στοιχείο που πρέπει να ληφθεί υπόψιν είναι και η ύπαρξη των εικονικών οντοτήτων, που αυξάνουν ακόμα περισσότερο τον αριθμό των απαιτούμενων ιδιωτικών κλειδιών και καθιστούν το N^2 το ελάχιστο πλήθος των διαφορετικών ιδιωτικών κλειδιών όπως φαίνεται και στον Πίνακα 2 παρακάτω:

Πλήθος Πραγματικών οντοτήτων	Πλήθος φύλλων	Πλήθος απαιτούμενων ιδιωτικών κλειδιών
2	4	4
2-4	8	16
5-8	16	64
9-16	32	256
17-32	64	1024
33-64	128	4096
65-128	256	16384
129-256	512	65536

Πίνακας 2. Πίνακας συσχέτισης πλήθους οντοτήτων με πλήθος απαιτούμενων ιδιωτικών κλειδιών.

Όπως είναι προφανές ο αριθμός των ιδιωτικών κλειδιών που απαιτούνται αυξάνεται με πολύ μεγάλο ρυθμό με την αύξηση των πραγματικών οντοτήτων δημιουργώντας πρόβλημα στην παραγωγή τους και τη διασφάλιση της διαφορετικότητάς τους. Στο συγκεκριμένο πρόβλημα θα μπορούσαν να χρησιμοποιηθούν μία από δύο λύσεις, είτε η παραγωγή των κλειδιών μέσω μιας μηχανής παραγωγής τυχαίων αριθμών, με πρόβλεψη για τη μη επανάληψη τους, είτε τη χρήση ενός συνόλου κλειδιών που για κάθε δέντρο πραγματοποιείται εναλλαγή της θέσης τους κατά την παραγωγή των φύλλων.

Οι δύο, όμως, αυτές λύσεις δεν περιορίζουν το μεγάλο πλήθος δεδομένων που θα πρέπει να αποθηκευτούν και να ανακληθούν σε περίπτωση που χρειαστούν. Εξάλλου, στην περίπτωση της μηχανής παραγωγής τυχαίων αριθμών, ακόμα και αν αυτή είναι ντετερμινιστική και τηρούμε μόνο την αρχική της κατάσταση (οπότε ξέρουμε ότι θα παράγονται κάθε φορά οι ίδιες ψευδοτυχαίες ακολουθίες, τις οποίες δεν χρειάζεται να τις τηρούμε), θα πρέπει κάθε φορά να εκτελείται Για το λόγο αυτό προτείνεται μία διαφορετική λύση. Η δημιουργία των ιδιωτικών κλειδιών, που θα παράγουν τα φύλλα του δέντρου, από την εφαρμογή του HMAC στα μοναδικά αναγνωριστικά, με τη χρήση ενός κλειδιού, διαφορετικό για κάθε δέντρο, που θα αντιστοιχεί στο ψευδώνυμο που παρέχεται στην κάθε οντότητα, όπως φαίνεται στην Εικόνα 18 παρακάτω:



Εικόνα 18. Απεικόνιση παραγωγής των φύλλων του δέντρου, με χρήση της HMAC για την παραγωγή ιδιωτικών κλειδιών.

Με αυτόν τον τρόπο σε κάθε οντότητα θα αντιστοιχίζεται ένα ιδιωτικό κλειδί και η παραγωγή όλων των δέντρων τύπου Merkle θα προέρχεται από δύο σύνολα δεδομένων πλήθους N το καθένα, που θα αντιστοιχίζονται μεταξύ τους, αυτό των μοναδικών αναγνωριστικών του υποκειμένου σε κάθε οντότητα $\{O_1, O_2, \ldots, O_N\}$ και αυτό με τα ιδιωτικά κλειδιά του υποκειμένου $\{K_1, K_2, \ldots, K_N\}$. Έτσι, τα φύλα του δέντρου που θα παράγουν το ψευδώνυμο θα παράγονται από $\alpha_{0,2i-1} = H(HMAC(K_i, O_i))$ και $\alpha_{0,2i-2} = H(O_i)$.

4.3.4 Απόδειξη Ασφάλειας

Ο κύριος λόγος της ψευδωνυμοποίησης είναι η προστασία των αρχικών δεδομένων. Είναι σημαντικό, λοιπόν, να ελεγχθεί αν με αυτόν τον τρόπο παραγωγής ψευδωνύμων, είναι δυνατή η εύρεση από αυτά των μοναδικών αναγνωριστικών που συνδέονται με λοιπά δεδομένα που πρέπει να προστατευτούν.

Θα πρέπει να αναφερθεί ότι ο μηχανισμός που περιγράφηκε ανωτέρω προϋποθέτει την ύπαρξη ασφαλούς διαύλου επικοινωνίας ανάμεσα στις οντότητες και τα υποκείμενα και γι' αυτό το λόγο δεν θα εξεταστεί η περίπτωση της υποκλοπής, από τρίτους, των μονοπατιών και η χρήση τους για την εξαπάτηση ως προς την πραγματική τους ταυτότητα.

Στη μέχρι τώρα μελέτη του μηχανισμού ψευδωνυμοποίησης με χρήση δέντρου τύπου Merkle, δεν έχει γίνει αναφορά στη χρήση κάποιας συγκεκριμένης συνάρτησης κατακερματισμού, όπου αυτή χρησιμοποιείται. Αυτή η παράλειψη έχει γίνει εσκεμμένα, αφού το όλο σχήμα δεν εξαρτάται από αυτή και μπορεί να χρησιμοποιηθεί οποιαδήποτε, εφόσον βέβαια η ίδια, ως συνάρτηση κατακερματισμού, είναι ασφαλής. Έτσι, ακόμα και αν οι προτεινόμενες λύσεις σταματήσουν να θεωρούνται ασφαλείς μπορούν εύκολα να αντικατασταθούν από τις νέες.

Για το λόγο αυτό προτείνεται για τη χρήση συνάρτησης κατακερματισμού ο SHA-2 ή ο SHA-3 (NIST 2012:3, NIST 2015:3) ως οι προτεινόμενες πιο ανθεκτικές συναρτήσεις κατακερματισμού σύμφωνα με τον ENISA (ENISA 2018:23).

Με αυτόν τον τρόπο η γνώση κάποιου κόμβου δεν μπορεί να οδηγήσει στον υπολογισμών των τιμών από τις οποίες προέκυψε, ειδικά αν ληφθεί υπόψιν ότι τα ιδιωτικά κλειδιά που χρησιμοποιούνται για την κατασκευή του δέντρου προέρχονται από την εφαρμογή HMAC για τον οποίον επίσης προτείνεται η χρήση κάποιας εκ των ανωτέρων συναρτήσεων κατακερματισμού.

Επιπλέον, θα πρέπει να αναφερθεί ότι ασφάλεια ενός δέντρου τύπου Merkle, δεν εξαρτάται μόνο από την ασφάλεια που παρέχει η συνάρτηση κατακερματισμού, αλλά και από τον τρόπο λειτουργίας του ως δυαδικό δέντρο, καθιστώντας πρακτικά αδύνατη την τροποποίηση κάποιου κόμβου, ενώ η ρίζα του δέντρου παραμένει η ίδια (Coronado García 2005:3).

Επίσης, είναι σημαντικό να αναφερθεί ότι η χρήση και η μελέτη των δέντρων τύπου Merkle έχει προχωρήσει σε αρκετά μεγάλο βαθμό και η ανθεκτικότητά τους σε μετακβαντικούς αλγόριθμους είναι καλά θεμελιωμένη (Chen et al 2016:4, Buchmann et al 2009:106, Butin et al 2015:43).

Τέλος, στο μηχανισμό ψευδωνυμοποίησης που περιεγράφηκε ανωτέρω είναι ξεκάθαρο ότι τα δέντρα που παράγονται τροποποιούνται με κάθε μεταβολή στο πλήθος των οντοτήτων που συμμετέχουν. Αυτό έχει σαν άμεσο αποτέλεσμα και την τροποποίηση των ψευδωνύμων,

γεγονός που καθιστά τα προηγούμενα ανίκανα να χρησιμοποιηθούν και αυτό αποτελεί ένα επιπλέον στοιχείο ασφάλειας.

Κεφάλαιο 5

Υλοποίηση και Μετρήσεις

Αφού μελετήθηκε σε θεωρητικό επίπεδο η δυνατότητα ψευδωνυμοποίησης με τη χρήση δέντρων τύπου Merkle, κρίνεται σκόπιμη η υλοποίηση του μηχανισμού, που περιγράφηκε στο προηγούμενο Κεφάλαιο, και η διενέργεια μετρήσεων με κύριο σκοπό την πρακτική επιβεβαίωση της λειτουργίας του αλλά και την αναγνώριση των επιδόσεών του σε ένα περιβάλλον απλού προσωπικού υπολογιστή.

5.1 Υλοποίηση

Για την υλοποίηση του αλγόριθμου χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python έκδοσης 3.8.2 σε περιβάλλον win32.

Σκοπός της υλοποίησης είναι η διενέργεια μετρήσεων σε ένα περιβάλλον απλού προσωπικού υπολογιστή των διαφόρων διεργασιών που αποτελούν το μηχανισμό και η συμπεριφορά τους για διάφορα πλήθη δεδομένων, αφού, όπως έχει ειπωθεί, πρόκειται για μέθοδο που την παραγωγή των ψευδωνύμων την εκτελεί το υποκείμενο και τα διαμοιράζει στις αντίστοιχες οντότητες. Για το λόγο αυτό, δεν κρίθηκε σκόπιμη η υλοποίηση της λειτουργίας της επικοινωνίας του υποκειμένου με τις οντότητες, ειδικά από το γεγονός ότι αυτή προτείνεται να πραγματοποιείται μέσω ενός ασφαλούς διαύλου επικοινωνίας. Έτσι, στην υλοποίηση που ακολουθεί όλα οι λειτουργίες πραγματοποιούνται από την ίδια κλάση.

5.1.1 Η Κλάση Person

Η κλάση Person, που φαίνεται στο Παράρτημα Α.1.1, δημιουργήθηκε, για να προσομοιάσει τη λειτουργία ενός υποκειμένου, το οποίο θα αποκτά μοναδικό αναγνωριστικό από κάθε οντότητα, η οποία διατηρεί δεδομένα του που πρέπει να προστατευτούν, και πρέπει να ψευδωνυμοποιηθεί, με τον τρόπο που περιγράφηκε στο προηγούμενο Κεφάλαιο. Έτσι, της παρέχεται η δυνατότητα να προσθέτει και να διαγράφει οντότητες, να κατασκευάζει δέντρα τύπου Merkle, να υπολογίζει το μονοπάτι, όταν αυτό απαιτείται, και τέλος να επιβεβαιώνει ότι από το μονοπάτι παράγεται το σωστό ψευδώνυμο.

Για την επίτευξη όλων αυτών των ενεργειών θα πρέπει να έχει προσβαση στα απαιτούμενα δεδομένα, που είναι άλλωστε και η συνολική γνώση που έχει το υποκείμενο στο μηχανισμό που περιγράψαμε. Έτσι, χρησιμοποιούνται οι εξής μεταβλητές:

- organisation_list: Είναι μία λίστα από αλφαριθμητικά όπου αποθηκεύονται οι ονομασίες των οντοτήτων με τις οποίες το υποκείμενο αλληλεπιδρά.

- private_key_list: Είναι μία λίστα όπου αποθηκεύονται αλφαριθμητικά που επιλέγει το υποκείμενο, για την παραγωγή των ιδιωτικών κλειδιών του κάθε δέντρου.

- organisation_identifiers_list: Είναι μία λίστα από αλφαριθμητικά όπου αποθηκεύονται τα μοναδικά αναγνωριστικά που λαμβάνει το υποκείμενο από κάθε οντοτητα.

- dummy_organisations: Είναι το πλήθος των εικονικών οντοτήτων που έχουν προστεθεί σε κάθε φάση παραγωγή των απαιτούμενων δέντρων.

Ένα γνώρισμα που έχουν οι ανωτέρω δομές είναι ότι υπάρχει άμεση αντιστοίχιση μεταξύ τους, με εξαίρεση τη μεταβλητή dummy_organisations. Έτσι, το organisation_list[i] περιέχει το όνομα της οντότητας που έχει δώσει στο υποκείμενο το μοναδικό αναγνωριστικό που είναι στο organisation_identifiers_list[i] και για την παραγωγή του δέντρου που θα δώσει το ψευδώνυμο αυτού του αναγνωριστικού θα χρησιμοποιηθεί το private_key_list[i] στην HMAC.

5.1.2 Δημιουργία Αντικειμένου της Κλάσης

Η δημιουργία ενός αντικειμένου της κλάσης Person είναι αντίστοιχη με τη δημιουργία ενός νέου υποκειμένου που θα αλληλεπιδράσει με κάποιες οντότητες και θα είναι ξεχωριστό για κάθε ένα

που δημιουργείται. Η δημιουργία ενός αντικειμένου πραγματοποιείται μέσω της μεθόδου του κατασκευαστή (constructor) def __init__(self).

Όπως φαίνεται και στο Παράρτημα Α.1.1 δεν έχει ορίσματα αλλά πραγματοποιεί έλεγχο στο φάκελο που υπάρχει το αρχείο του πηγαίου κώδικα για αρχείο με όνομα «myfile.txt». Αν το εντοπίσει θεωρεί πως υπάρχουν δεδομένα αποθηκευμένα σε αυτό από προηγούμενη δραστηριότητα του υποκειμένου και με αυτά αρχικοποιεί τις μεταβλητές του αντικειμένου, ώστε να μπορεί να συνεχίζεται η λειτουργία από εκεί που σταμάτησε την προηγούμενη φορά. Αν δεν εντοπίσει το κατάλληλο αρχείο θεωρεί ότι είναι η πρώτη φορά που το υποκείμενο αλληλεπιδρά με κάποια οντότητα και έτσι ζητά από αυτό να εισάγει τα απαραίτητα δεδομένα, δηλαδή το όνομα της οντότητας, το μοναδικό αναγνωριστικό που του ανέθεσε η οντότητα, τέλος το μοναδικό ιδιωτικό κλειδί που θα χρησιμοποιηθεί για τη δημιουργία του ψευδωνύμου σε αυτή. Στη συνέχεια δημιουργεί το δέντρο τύπου Merkle και αποκτά από τη ρίζα του το ψευδώνυμο και κατασκευάζει το κατάλληλο μονοπάτι, το οποίο αποστέλλει στην οντότητα για επιβεβαίωση και σύνδεση με το μοναδικό αναγνωριστικό. Όπως έχει αναφερθεί ανωτέρω, στη συγκεκριμένη υλοποίηση, ο έλεγχος αυτός πραγματοποιείται εντός του αντικειμένου.

5.1.3 Αποθήκευση Δεδομένων σε Αρχείο

Όπως είναι φυσικό για μεγάλο αριθμό εγγραφών το πλήθος των στοιχείων που απαιτούνται να θυμάται το υποκείμενο γίνεται πολύ μεγάλο και γι' αυτό το λόγο προστέθηκε η δυνατότητα αποθήκευσης των δεδομένων σε αρχείο.

Αυτή η διαδικασία πραγματοποιείται από τη μέθοδο def save_data(self), όπως φαίνεται στο Παράρτημα Α.1.2. Αυτή η μέθοδος δεν απαιτεί ορίσματα αλλά αποθηκεύει τα δεδομένα των μεταβλητών του αντικειμένου σε ένα αρχείο κειμένου με όνομα «myfile.txt», το οποίο μπορεί να χρησιμοποιηθεί από τη μέθοδο κατασκευαστή (constructor) def __init__(self), που περιεγράφηκε ανωτέρω. Ο τρόπος που αποθηκεύει τα δεδομένα είναι μία κατακόρυφη διάταξη, όπου στην αρχή αποθηκεύονται με τη σειρά τα ονόματα των οντοτήτων, στη συνέχεια τα μοναδικά αναγνωριστικά που αντιστοιχούν σε αυτές, έπειτα τα ιδιωτικά κλειδιά που ανέθεσε σε κάθε μία το υποκείμενο και τέλος ο αριθμός των εικονικών οντοτήτων που χρειάστηκαν για την παραγωγή των δέντρων τύπου Merkle.

5.1.4 Προσθαφαίρεση Οντοτήτων

Η κλάση Person δημιουργήθηκε εσκεμμένα δυναμική, για να μπορεί να ακολουθήσει τη αλληλεπίδραση ενός υποκειμένου, το οποίο μπορεί να προστίθεται σε οντότητες, παραχωρώντας δεδομένα του, ή και να διαγράφεται από αυτές. Στη συγκεκριμένη υλοποίηση έχουν πραγματοποιηθεί δύο διαδικασίες πρόσθεσης οντοτήτων, όπως φαίνεται και στο Παράρτημα Α.1.3:

- def add_organisation(self,orgname,orgid,pkey): Αυτή η μέθοδος είναι η κύρια διαδικασία, που πραγματοποιεί την δυναμική εισαγωγή οντοτήτων. Στα ορίσματα που χρησιμοποιεί εισάγονται τα απαραίτητα δεδομένα και είναι orgname,orgid,pkey, όπου αντίστοιχα περιέχουν το όνομα της οντότητας, το μοναδικό αναγνωριστικό, που ανέθεσε αυτή στο υποκείμενο, και το ιδιωτικό κλειδί του υποκειμένου για τη δημιουργία του δέντρου. Αφού πραγματοποιηθεί έλεγχος για την ύπαρξη εικονικών οντοτήτων, αν εντοπιστούν αφαιρεί τις καταγραφές τους και εισάγει τη νέα καταγραφή. Τέλος, αυτή η νέα εγγραφή προκαλεί, όπως περιεγράφηκε στο προηγούμενο Κεφάλαιο αλλαγή σε όλα τα δέντρα όλων των οντοτήτων και άρα στα ψευδώνυμά τους. Για το λόγο αυτό ακολουθεί ο επαναϋπολογισμός όλων και η επιβεβαίωσή τους, όχι όμως για τις εικονικές οντότητες, αφού δεν είναι υπαρκτές και άρα δεν απαιτείται η ενημέρωσή τους..

- def add_organisation_without_checking(self,orgname,orgid,pkey): Αυτή η μέθοδος υλοποιήθηκε μόνο για τη χρήση της στις δοκιμές και δεν αποτελεί μέρος του τελικού κώδικα. Είναι ίδια με την def add_organisation(self,orgname,orgid,pkey), χωρίς τον έλεγχο της τροποποίησης των δέντρων, ώστε να μην καθυστερείται η διαδικασία κατά την συνεχόμενη εισαγωγή δεδομένων από τον επαναϋπολογισμό.

Η αφαίρεση οντοτήτων πραγματοποιείται από τη μέθοδο def remove_organisation(self,orgname), που σα μοναδικό όρισμα λαμβάνει το όνομα της οντότητα που πρέπει να διαγραφεί, εντοπίζει τη θέση της στην λίστα organisation_list την αφαιρεί και αφαιρεί από τις λίστες organisation_identifiers_list και private_key_list τα αντίστοιχα δεδομένα, επαναϋπολογίζει τα δέντρα για όλες τις εναπομείνουσες οντότητες και τα νέα μονοπάτια, για να τις ενημερώσει με αυτά για τα νέα ψευδώνυμα. Η συγκεκριμένη μέθοδος προσομοιάζει τη διαδικασία διαγραφής ενός υποκειμένου από μία οντότητα και την κατάργηση της καταγραφής του από αυτή.

Τέλος, η μέθοδος def add_dummy_organisation(self) είναι αυτή που δημιουργεί εικονικές οντότητες και καλείται από τις προηγούμενες κάθε φορά που πραγματοποιείται αλλαγή στη

λίστα organisation_list, έτσι ώστε όποτε χρειάζεται να συμπληρώνεται και το πλήθος των φύλλων των δέντρων να είναι πάντα δύναμη του δύο.

5.1.5 Δημιουργία Δέντρου και Μονοπατιού

Το κάθε δέντρο που δημιουργεί το ψευδώνυμο παράγεται δυναμικά, κάθε φορά που απαιτείται, και δεν παραμένει αποθηκευμένο, αφού κάτι τέτοιο, για μεγάλο αριθμό οντοτήτων, θα οδηγούσε σε μεγάλη απαίτηση χώρου. Έτσι, το δέντρο τύπου Merkle σχηματίζεται κάθε φορά με την κλήση της μεθόδου def create_merkle_tree(self,org_place), με την εισαγωγή ενός ορίσματος που είναι η θέση στη λίστα organisation_list της οντότητας στην οποία αφορά και παράγει το ψευδώνυμο για το μοναδικό αναγνωριστικό της, για το συγκεκριμένο υποκείμενο. Στο τέλος επιστρέφεται μία λίστα από λίστες που η κάθε μία αποτελείται από τους κόμβους του κάθε επιπέδου του δέντρου.

Για παράδειγμα ένα δέντρο που δημιουργείται από 3 οντότητες και φυσικά συμπληρώνεται με μία εικονική φαίνεται στην Εικόνα 19 παρακάτω:

```
[['b6c4c0199b73490ad05a969a2b1fbc28762d04f2e050d407313eab045df79a14', '
877b10db8e0b76463c1295158a4fadf1a84a96794a37a60ad4c6e02f2d558229', 'c2a
cd1852a5239be6977219026d9f36ae4fd627a9e4fdaf62d6ca9e615f6ed7e', '8a463d
2fe796fdf5abbd23697ea9919e64ee9dbcf0515f290fadb30f1ae603f0', '05a18cc51
d17f94f486df9eb51909c2613b26659762dedd460147a3eb033771e', '385a8a209716
690fcaf4fcc352e29cb93185d9eaf10e91bb4e30dcc5dc70b0eb', '37a729124f9634f
8f5047531b88c839e5aa19c97d941f88b8eac135bcd00d377', '8b49d66d3d1f6bc7af
36b8d0b407b773844e4e02bf5fcc86fd09de6013c2d17c'], ['d1260e4037c02bb383c
9bd4661f4d8859f988d4314a3d763861fdd1dcb513e2e', '1e3643d23d84faafc8f13b
11c91d643f8d65950146873047dd1c6096c0ce88b6', '77d2a2a216cdfef619c71d32b
b51bc53dc638f703c5fc722ab78b89bcbd86872', '8683d3626e9a1a9b8a27b89368cb
6712818f7c33649ef11342b25d646fd6787e'], ['57b4617aeb30739732aae1d02e72f
caedba689ca5fde907f6e294de0d79e4f5f', '62456cd04c896569a11aabae10471e0d
7c08f57bfc8fe5e4cfd23ce11f2d1b46'], ['73e6d8633db7a82361458bcf1323b3f77
feb6f72df7742c0652312c96a452303']]
```

Εικόνα 19. Παράδειγμα δέντρου που δημιουργείται από τρεις οντότητες.

Το τελευταίο στοιχείο της λίστας αποτελεί και το ψευδώνυμο που παράχθηκε.

Η ανωτέρω διαδικασία καλείται κατά τη δημιουργία του μονοπατιού, που θα χρειαστεί είτε για την ανάθεση ψευδωνύμου είτε για την επιβεβαίωσή του. Αυτό πραγματοποιείται από τη μέθοδο def get_verification_path(self,organisation,target_organisation), που φαίνεται στο Παράρτημα Α.1.4, και δέχεται σαν ορίσματα τα ονόματα της οντότητας organisation που επιθυμεί να αποκτήσει πρόσβαση στα δεδομένα του υποκειμένου που κατέχει η οντότητα target_organisation. Αφού εντοπιστούν τα απαραίτητα δεδομένα της target_organisation δημιουργείται το δέντρο που παράγει το ψευδώνυμο του υποκειμένου σε αυτή. Στη συνέχεια

υπολογίζεται το μονοπάτι των κόμβων που χρειάζεται να γνωρίζει η organisation σε αυτό το δέντρο και επιστρέφεται. Ένα παράδειγμα μονοπατιού σε ένα δέντρο που δημιουργείται από 3 οντότητες και φυσικά συμπληρώνεται με μία εικονική φαίνεται στην Εικόνα 20 παρακάτω:

```
['6e551aad7c6e043ba69e690c0afb5e824b0418f53505c414e0f32306610c5047', 'right', '0
5f5ae3a27a927485aea3f348b08e4992a5c7159b4d44ecbe29c7e91b2e4b016', 'left', '27d02
32d5c3db665bad93f9753e5442cc4f07e3530380edbb5086369940f79a1', 'ceeb8b8fa91a0057b
38f2f09aebffa3be7e288ee763b1308e9852a6dba6a1dd3']
```

Εικόνα 20. Παράδειγμα μονοπατιού σε ένα δέντρο που δημιουργείται από τρεις οντότητες

5.1.6 Επιβεβαίωση Μονοπατιού

Η κλήση της def get_verification_path(self,organisation,target_organisation), που φαίνεται στο Παράρτημα Α.1.5 πραγματοποιείται από την def pseudonym_verification(self,organisation,target_organisation), που παίρνει τα ίδια ορίσματα. Στη συνέχεια «διαβάζει» τα δεδομένα που υπάρχουν σε αυτή τη λίστα και ακολουθεί τις οδηγίες «left» ή «right», που φαίνονται χαρακτηριστικά στην Εικόνα 20, για τον τρόπο που θα πρέπει να ενωθεί ο κόμβος που δίνεται από το μονοπάτι με αυτόν που έχει παράξει η οντότητα με τα δεδομένα που έχει. Το τελευταίο στοιχείο της λίστας είναι και το ψευδώνυμο του υποκειμένου στο target_organisation στο οποίο πρέπει να καταλήξει η organisation ακολουθώντας το μονοπάτι ξεκινώντας από την εφαρμογή συνάρτησης κατακερματισμού στο δικό της μοναδικό αναγνωριστικό για το υποκείμενο. Από αυτή τη λειτουργία, είναι προφανές, ότι αν στα ορίσματα δοθεί η ίδια τιμή, δηλαδή organisation = target_organisation, παράγεται το μονοπάτι που επιβεβαιώνει το ψευδώνυμο αυτής της μίας οντότητας. Αυτή η τεχνική χρησιμοποιείται, άλλωστε, και για την ενημέρωση των οντοτήτων κατά την αλλαγή των ψευδωνύμων.

Επιπλέον, στην υλοποίηση αυτή πραγματοποιείται και ένα επιπλέον έλεγχος για την επιβεβαίωση ότι το ψευδώνυμο που παράγεται από τις οδηγίες του μονοπατιού είναι ίδιο με αυτό του δέντρου του target_organisation με απευθείας σύγκριση.

Ένα παράδειγμα ολοκληρωμένης εκτέλεσης φαίνεται στην Εικόνα 21, όπου εισάγονται 3 οντότητες και ζητείται από τη δεύτερη το ψευδώνυμο του υποκειμένου στην πρώτη:

```
Give Organisation name:  amka
Give organisation identifier:  6548984161
Give private key:  iam@heaven.com
Give Organisation name:  eforia
Give organisation identifier:  123456789
Give private key:  anything
Give Organisation name:  astynomia
Give organisation identifier:  AA456789
Give private key:  something
Give 2 organisation names
Give Organisation name requesting access:  eforia
Give target Organisation's name:  amka
The pseudonym created by the path for the organisation eforia matches with the pseudonym in the path
The pseudonym created by the path for the organisation eforia matches with the pseudonym in the tree of amka
206d15e9a8d952c12c03cf938b4b1b7e0680e61558f7538f2412b87a262a2536
 206d15e9a8d952c12c03cf938b4b1b7e0680e61558f7538f2412b87a262a2536
```

Εικόνα 21. Παράδειγμα ολοκληρωμένης εκτέλεσης.

Το συγκεκριμένο αποτέλεσμα παράχθηκε από την εκτέλεση των εντολών στο Παράρτημα Α.2.

5.1.7 Βιβλιοθήκες

Για την υλοποίηση και λειτουργία όλων των προηγουμένων είναι απαραίτητη η χρήση κάποιων βιβλιοθηκών της γλώσσας προγραμματισμού python. Αυτές είναι οι εξής:

- hashlib: Χρησιμοποιείται για τη λειτουργία της συνάρτησης hashlib.sha256(x), που πραγματοποιεί την εφαρμογή της αντίστοιχης συνάρτησης κατακερματισμού στο όρισμα που δέχεται και είναι ένα αντικείμενο σε μορφή byte. Με τη χρήση της ίδιας βιβλιοθήκης μπορεί πολύ εύκολα η υλοποίηση να μετατραπεί με τη χρήση hashlib.sha512(x).
- hmac: Χρησιμοποιείται για τη λειτουργία της συνάρτησης hmac.new(key, msg, digestmod=''), όπου τα ορίσματα key και msg είναι σειρές από bytes, ενώ στο όρισμα digestmod δηλώνεται η συνάρτηση κατακερματισμού που θα χρησιμοποιηθεί. Στη συγκεκριμένη υλοποίηση χρησιμοποιείται η sha256, που και πάλι μπορεί να αντικατασταθεί πολύ εύκολα από την sha512.
- os.path: Χρησιμοποιείται για τη λειτουργία ανοίγματος, διαβάσματος και γραψίματος σε αρχείο.
- random: Χρησιμοποιείται για τη χρήση της συνάρτησης random() που επιστρέφει μία τυχαία τιμή χ, όπου 0<χ<1, στην παραγωγή στοιχείων των εικονικών οντοτήτων και κατά τη διαδικασία των μετρήσεων, για την παραγωγή στοιχείων μεγάλου αριθμού οντοτήτων.
- time: Χρησιμοποιείται για τη χρήση της κλάσης time, για την χρονομέτρηση λειτουργιών στις διάφορες μετρήσεις.

5.2 Μετρήσεις

Μετά την περιγραφή της υλοποίησης της μεθόδου ψευδωνυμοποίησης, που προτείνεται, πραγματοποιήθηκαν μετρήσεις στους χρόνους εκτέλεσης, για την απεικόνιση των επιδόσεων του μηχανισμού. Τα χαρακτηριστικά του φορητού υπολογιστή που χρησιμοποιήθηκε για την εκτέλεση του κώδικα που παράχθηκε είναι:

- Επεξεργαστής: AMD Ryzen 3 2200U with Radeon Vega Mobile Gfx 2.5 GHz
- Μνήμη RAM: 8 GB
- Λειτουργικό σύστημα: Windows 10 Pro

5.2.1 Επιβεβαίωση Μονοπατιού

Στην υλοποίηση που πραγματοποιήθηκε, όπως έχει αναφερθεί, επελέγη η δυναμική δημιουργία των δέντρων και των μονοπατιών και όχι ο εξαρχής υπολογισμός τους και αποθήκευσή τους, για εξοικονόμηση χώρου. Έτσι, ένα βασικό στοιχείο είναι ο χρόνος που απαιτείται για των υπολογισμό αυτών.

Η δημιουργία κάποιου δέντρου καλείται μόνο κατά τη δημιουργία μονοπατιού, που με τη σειρά της καλείται στην περίπτωση που απαιτείται η δημιουργία ψευδωνύμου, είτε για την ενημέρωση κάποια οντότητας, είτε για την πληροφόρησή της για το ψευδώνυμο που χρησιμοποιείται για το υποκείμενο σε κάποια άλλη. Η διαδικασία είναι ίδια και στις δύο περιπτώσεις, χωρίς διαφορές μεταξύ τους.

Για το λόγο αυτό πραγματοποιήθηκαν μετρήσεις για το χρόνο που απαιτείται από τον κώδικα να πραγματοποιήσει επιβεβαιώσεις ψευδωνύμων για διάφορα μεγέθη δέντρων. Σε αυτό το σημείο θα πρέπει να αναφερθεί ότι, όπως περιγράφηκε στην Υποενότητα 4.3.1 και όπως φαίνεται στον Πίνακα 1 το μέγεθος των δέντρων για διάφορα πλήθη οντοτήτων N παραμένει σταθερό για τιμές $2^{\nu} \leq 2N < 2^{\nu+1}$. Έτσι, οι μετρήσεις πραγματοποιήθηκαν για $N = 2^{\mu}$ όπου $\mu>0$, όπου κάθε επιβεβαίωση μονοπατιού πραγματοποιήθηκε δέκα φορές και υπολογίστηκε ο μέσος όρος των χρόνων, με τη χρήση του κώδικα που φαίνεται στο Παράρτημα Α.3. Τα αποτελέσματα των μετρήσεων αυτών φαίνονται στον Πίνακα 3:

Πλήθος οντοτήτων	Μέσος Χρόνος Ολοκληρωμένης	Πλήθος Φύλων του Δέντρου

	Επιβεβαίωσης (σε msec)	
2	31.212	4
4	31.383	8
8	32.341	16
16	34.385	32
32	34.387	64
64	35.947	128
128	37.512	256
256	42.200	512
512	51.577	1024
1024	70.335	2048
2048	103.158	4096
4096	289.150	8192
8192	304.778	16384

Πίνακα 3. Αποτελέσματα μέσου όρου μετρήσεων χρόνου εκτέλεσης ολοκληρωμένης επιβεβαίωσης μονοπατιού.

Όπως φαίνεται από τις μετρήσεις και όπως ήταν άλλωστε λογικό για μεγάλο μέγεθος δέντρων αυξάνεται ο χρόνος επιβεβαίωσης, αλλά παραμένει μέσα σε ανεκτά επίπεδα.

5.2.2 Ενημέρωση Οντοτήτων

Κάθε φορά που προστίθεται ή αφαιρείται μία οντότητα από το σύστημα, όπως έχει αναφερθεί, προκαλεί αλλαγή στα ψευδώνυμα όλων των οντοτήτων που αφορούν στο ίδιο υποκείμενο. Έτσι, για κάθε τέτοια μεταβολή πραγματοποιείται εκ νέου ο υπολογισμός που μετρήθηκε στην προηγούμενη Υποενότητα για κάθε έγκυρη οντότητα.

Είναι προφανές ο χρόνος υπολογισμού και επιβεβαίωσης θα είναι πολλαπλάσιος από αυτόν που μετρήθηκε ανωτέρω, αφού αυτή η διαδικασία θα πραγματοποιείται για κάθε οντότητα ξεχωριστά και για μεγάλο αριθμό είναι πολύ πιθανό να αυξάνεται σημαντικά.

Για το σκοπό αυτό αναπτύχθηκε ο κώδικας που φαίνεται στο Παράρτημα Α.4, όπου μετρά το χρόνο εκτέλεσης της διαδικασίας ενημέρωσης των οντοτήτων για διάφορα πλήθη οντοτήτων. Σε αυτό το σημείο πρέπει να αναφερθεί ότι και σε αυτή την περίπτωση λήφθηκε υπόψιν το γεγονός ότι το μέγεθος του δέντρου τροποποιείται σημαντικά σε συγκεκριμένα μόνο πλήθη οντοτήτων και όχι συνεχόμενα. Η διαφορά σε αυτή την περίπτωση είναι ότι για ίδιο μέγεθος δέντρου κάθε φορά υπάρχει διαφορετικό πλήθος έγκυρων οντοτήτων, μια και για τις εικονικές δεν πραγματοποιείται ενημέρωση. Για το λόγο αυτό οι μετρήσεις πραγματοποιήθηκαν για δέντρα με μηδενικό αριθμό εικονικών οντοτήτων και άρα μέγιστο πραγματικών και έτσι ο χρόνος που μετράται είναι και ο ενδεικτικά μέγιστος για το συγκεκριμένο δέντρο. Τα αποτελέσματα των μετρήσεων αυτών φαίνονται στον Πίνακα 4:

Πλήθος οντοτήτων	Χρόνος Ολοκληρωμένης Ενημέρωσης (σε msec)	Πλήθος Φύλων του Δέντρου
2	15.350	4
4	15.618	8
8	31.258	16
16	93.790	32
32	234.435	64
64	453.281	128

128	953.411	256
256	2063.127	512
512	5204.681	1024
1024	14973.290	2048
2048	47241.006	4096
4096	268850.044	8192
8192	582600.016	16384

Πίνακα 4. Αποτελέσματα μετρήσεων χρόνου εκτέλεσης ολοκληρωμένης ενημέρωσης οντοτήτων.

Από αυτές τις μετρήσεις φαίνεται άμεσα ότι για μικρό πλήθος οντοτήτων ο χρόνος εκτέλεσης και ολοκλήρωσης της ενημέρωσής τους είναι σχετικά μικρός και ανεκτός. Αντίθετα, για μεγάλο αριθμό οντοτήτων ο χρόνος, που απαιτείται για την προσθήκη ή την αφαίρεση νέων, γίνεται αρκετά μεγάλος και μπορεί να δημιουργήσει προβλήματα.

5.3 Σκέψεις

Η ψευδωνυμοποίηση με τη χρήση των δέντρων τύπου Merkle είναι προφανές ότι μπορεί να υλοποιηθεί με τον τρόπο που περιεγράφηκε ανωτέρω. Αυτή αποτελεί μία πρώτη προσπάθεια για τη μελέτη της συμπεριφοράς του θεωρητικού μηχανισμού σε πραγματικό περιβάλλον.

Όπως φαίνεται, από τα αποτελέσματα των μετρήσεων για μικρό πλήθος οντοτήτων έχει ικανοποιητική απόδοση και θα μπορούσε να λειτουργήσει στην πράξη. Όταν, όμως, ο αριθμός των οντοτήτων, που εμπλέκονται, μεγαλώνει σε σημαντικό βαθμό ο χρόνος ολοκλήρωσης της διαδικασίας της ενημέρωσής τους με τα νέα ψευδώνυμα.

Αφού το πρόβλημα δημιουργεί το πλήθος των οντοτήτων που πρέπει να ενημερωθούν, ο περιορισμός αυτών θα μπορούσε να μειώσει το χρόνο ολοκλήρωσης της διαδικασίας. Κάτι τέτοιο θα μπορούσε να επιτευχθεί αν οι οντότητες ομαδοποιούνταν και δημιουργούσαν μια

ενότητα, η οποία θα λειτουργεί με τον τρόπο που έχει περιγραφεί και θα είναι ανεξάρτητες μεταξύ τους. Κάτι τέτοιο προφανώς δημιουργεί πρόβλημα σχετικά με την δυνατότητα επικοινωνίας οντοτήτων διαφορετικών ενοτήτων, όπως περιεγράφηκε στην Υποενότητα 4.2.4 και χρειάζεται περαιτέρω μελέτης για τον τρόπο που μπορεί αυτή να επιτευχθεί.

Επίσης, βελτιώσεις στο μηχανισμό που αναλύθηκε θα μπορούσαν να προσφέρουν τόσο η πρόταση για ένα διαφορετικό και πιο αποδοτικό τρόπο διάσχισης ενός δέντρου Merkle (Jakobsson et al 2003:315, Szydlo 2004:542), αλλά θα πρέπει να εξεταστεί στον αν και με πιο τρόπο θα μπορούσαν να εφαρμοστούν στην περίπτωση του τρόπου ψευδωνυμοποίησης που προτείναμε.

Κεφάλαιο 6

Επίλογος

Η παρούσα μεταπτυχιακή διατριβή εκπονήθηκε με σκοπό να συνεισφέρει στο διάλογο σχετικά με την προστασία των προσωπικών και ευαίσθητων προσωπικών δεδομένων, με τη μέθοδο της ψευδωνυμοποίησης μέσω της χρήσης καινοτόμων κρυπτογραφικών μεθόδων.

Για την επίτευξη αυτού του σκοπού ξεκινήσαμε την έρευνά μας από τη μελέτη του Γενικού Κανονισμού για την Προστασία Δεδομένων της Ευρωπαϊκής Ένωσης, όπου διαπιστώσαμε την έντονη απαίτηση για προστασία των δεδομένων ειδικά των προσωπικών και ευαίσθητων προσωπικών. Επίσης, η παρατήρηση ότι για πρώτη φορά εμφανίζεται ο ορισμός του όρου της ψευδωνυμοποίησης σε νομικό πλαίσιο, καθώς και η πρόταση από αυτό ως μέθοδος προστασίας των δεδομένων και η επαναλαμβανόμενη αναφορά σε αυτή επιβεβαιώνει τη σημασία της και το ρόλο που κατέχει. Άρα καθίσταται απαραίτητη η προώθηση της έρευνας για την εξέλιξή της είτε βελτιώνοντας παλαιότερες τεχνικές είτε αναπτύσσοντας καινούριες. Επιπλέον, μέσα από αυτό το νομοθετικό πλαίσιο εντοπίστηκαν οι δυνατότητες που πρέπει να παρέχουν οι μέθοδοι ψευδωνυμοποίησης, δηλαδή να επιτρέπεται η ανταλλαγή δεδομένων, με την σύμφωνη γνώμη του προσώπου στο οποίο αφορούν, και να είναι μόνο οι απολύτως απαραίτητες, για την πραγματοποίηση του έργου για το οποίο απαιτούνται.

Στη συνέχεια εντοπίσαμε τις μεθόδους που προτείνονται ως λύσεις στο ερώτημα για τους τρόπους που μπορεί να επιτευχθεί η ψευδωνυμοποίηση με τη χρήση της κρυπτογραφίας. Με κύρια αναφορά στις σχετικές εργασίες του ENISA, που είναι το επίσημο συμβουλευτικό όργανο της Ευρωπαϊκής Ένωσης σε θέματα ασφάλειας πληροφοριών καταγράφηκαν οι τρόποι αυτοί και επικεντρωθήκαμε στις κατευθύνσεις που δίνει για καινοτόμες και προηγμένες λύσεις κρυπτογραφικής ψευδωνυμοποίησης.

Επειδή, όμως, η ανάπτυξη της μεθόδου της πολυμορφικής κρυπτογραφικής ψευδωνυμοποίησης είναι ακόμη στο στάδιο της έρευνας και των δοκιμών προσανατολιστήκαμε στη δεύτερη καινοτόμα μέθοδο που προτείνεται και αυτή είναι η χρήση των δέντρων τύπου Merkle. Στην προσπάθειά μας να εντοπιστεί βιβλιογραφία που να αναφέρεται στην εφαρμογή αυτού του σχήματος παρατηρήσαμε ότι έχει πραγματοποιηθεί εκτενής μελέτη του όσον αφορά στις ψηφιακές υπογραφές και στα κρυπτονομίσματα, όπου έχει και ήδη εφαρμοστεί, αλλά για την επίτευξη ψευδωνυμοποίησης υπάρχει κενό.

Έτσι, αποφασίσαμε να προτείνουμε έναν τρόπο ψευδωνυμοποίησης με τη χρήση των δέντρων τύπου Merkle, για να συνεισφέρουμε επιπλέον γνώση με αυτή τη νέα πρόταση. Δημιουργήσαμε, λοιπόν, αυτή τη μέθοδο που περιγράφηκε στην παρούσα μεταπτυχιακή διατριβή και μελετήσαμε τη συμπεριφορά της και τις δυνατότητες που παρέχει. Με αυτόν τον τρόπο καταλήξαμε ότι είναι δυνατή η ψευδωνυμοποίηση με τη χρήση ενός σχήματος δέντρων τύπου Merkle, η οποία επιτρέπει την ασφαλή τρόπο, ακόμα και με μετακβαντικά κριτήρια, την τροποποίηση των αναγνωριστικών των προσώπων και ότι μπορεί εύκολα να ανανεωθεί η ασφάλειά της με την εύκολη αλλαγή των συναρτήσεων κατακερματισμού, που χρησιμοποιεί, ώστε να ακολουθεί και στο μέλλον τις απαιτήσεις ασφάλειας, όταν αυτές αλλάζουν. Επιπλέον, επιτρέπει την ανταλλαγή δεδομένων, όταν αυτό απαιτείται, παρά μόνο με τη συναίνεση του προσώπου στο οποίο αφορούν, ακολουθώντας τις επιταγές του Γενικού Κανονισμού για την Προστασία Δεδομένων.

Με αυτό το θεωρητικό υπόβαθρο, στη συνέχεια, προχωρήσαμε στην υλοποίηση κώδικα που να εφαρμόζει τη μέθοδο που προτείναμε, για να προσομοιώσουμε τη συμπεριφορά της σε πραγματικό περιβάλλον. Με τα αποτελέσματα των μετρήσεων που πραγματοποιήθηκαν καταλήξαμε στο συμπέρασμα ότι θα μπορούσε να χρησιμοποιηθεί, για την ψευδωνυμοποίηση των αναγνωριστικών ενός προσώπου σε διαφορετικές οντότητες, στις οποίες έχει παραχωρήσει τα δεδομένα του και ότι υπάρχει περιθώριο για περαιτέρω μελέτη της με σκοπό τη βελτίωσή της και μείωση του χρόνου εκτέλεσης του κώδικα που χρησιμοποιήσαμε.

Έτσι, προτείνουμε για μελλοντική έρευνα, για τη δυνατότητα εφαρμογής στη μέθοδο που περιγράψαμε, τεχνικών που έχουν προταθεί και μειώνουν το χρόνο διάσχισης των δέντρων τύπου Merkle, καθώς και τη δυνατότητα ομαδοποίησης δέντρων με τρόπο που να επιτρέπει την επικοινωνία τους. Σε κάθε περίπτωση, η αξιοποίηση άλλων κρυπτογραφικών μοντέλων (π.χ. πρωτόκολλα μηδενικής γνώσης) στο πλαίσιο της ψευδωνυμοποίησης δεδομένων αποτελούν σαφώς ένα πολύ ενδιαφέρον και ανοιχτό ερευνητικό πεδίο.

Βιβλιογραφία

Article 29 Working Party, 2014 "Opinion 05/2014 on anonymisation techniques". (WP29, 2014)

Barbaro M and Jr TZ (2006) A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times*, 9 August. Available at: https://www.nytimes.com/2006/08/09/technology/09aol.html (accessed 01/05/20).

Bernstein DJ and Lange T (2017) Post-quantum cryptography : dealing with the fallout of physics success. IACR. Available at: https://research.tue.nl/en/publications/post-quantum-cryptography-dealing-with-the-fallout-of-physics-suc (accessed 29/04/20).

Buchmann J, Lindner R, Rückert M and Schneider M (2009) Post-quantum cryptography: lattice signatures. *Computing* 85(1–2): 105–125.

Butin D, Gazdag S-L and Buchmann J (2015) Real-World Post-Quantum Digital Signatures. In: Cleary F and Felici M (eds) *Cyber Security and Privacy*. Cham: Springer International Publishing, 41–52. Available at: http://link.springer.com/10.1007/978-3-319-25360-2_4 (accessed 25/04/20).

Chen L, Jordan S, Liu Y-K, Moody D, Peralta R, Perlner R and Smith-Tone D (2016) *Report on Post-Quantum Cryptography*. National Institute of Standards and Technology, NIST IR 8105. Available at: https://nvlpubs.nist.gov/nistpubs/ir/2016/NIST.IR.8105.pdf (accessed 26/04/20).

Coronado García L (2005) On the security and the eciency of the Merkle signature scheme. *IACR Cryptology ePrint Archive* 2005: 192.

Diffie W and Hellman M (1976) New directions in cryptography. *IEEE Transactions on Information Theory*. paper presented at the IEEE Transactions on Information Theory 22(6): 644–654.

ENISA, (2018) "Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation".

ENISA. (2019) "Pseudonymisation Techniques and Best Practices : Recommendations on Shaping Technology According to Data Protection and Privacy Provisions." Website. Publications Office of the European Union.

Hu R, Stalla-Bourdillon S, Yang M, Schiavo V and Sassone V (2017) *Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=3034261 (accessed 13/01/20).

Jakobsson M, Leighton T, Micali S and Szydlo M (2003) Fractal Merkle Tree Representation and Traversal. In: Joye M (ed) *Topics in Cryptology — CT-RSA 2003*. Berlin, Heidelberg: Springer Berlin Heidelberg, 314–326. Available at: http://link.springer.com/10.1007/3-540-36563-X_21 (accessed 21/04/20).

Koblitz N and Menezes AJ (2016) Cryptocash, cryptocurrencies, and cryptocontracts. *Designs, Codes and Cryptography* 78(1): 87–102.

Lehnhardt J and Spalka A (2011) Decentralized generation of multiple, uncorrelatable pseudonyms without trusted third parties. *Proceedings of the 8th international conference on Trust, privacy and security in digital business*. Toulouse, France: Springer-Verlag, 113–124.

Menezes AJ, Vanstone SA and Oorschot PCV (1996) *Handbook of Applied Cryptography* (1st edition). USA: CRC Press, Inc.

Merkle Ralph C. (1988) A Digital Signature Based on a Conventional Encryption Function. In: Pomerance C (ed) *Advances in Cryptology — CRYPTO '87*. Berlin, Heidelberg: Springer, 369–378.

Merkle Ralph Charles (1979) Secrecy, authentication, and public key systems. phd, Stanford, CA, USA, Stanford University.

Nakamoto, Satoshi. (2009). Bitcoin: A Peer-to-Peer Electronic Cash System. Cryptography Mailing list at https://metzdowd.com.

NIST (2008) *The Keyed-Hash Message Authentication Code (HMAC)*. U.S. Department of Commerce. Available at: https://csrc.nist.gov/publications/detail/fips/198/1/final (accessed 17/04/20).

NIST (2012) *Secure Hash Standard (SHS)*. U.S. Department of Commerce. Available at: https://csrc.nist.gov/publications/detail/fips/180/4/archive/2012-03-06 (accessed 18/04/20).

NIST (2015) *Secure Hash Standard (SHS)*. U.S. Department of Commerce. Available at: https://csrc.nist.gov/publications/detail/fips/180/4/final (accessed 18/04/20).

Szydlo M (2004) Merkle Tree Traversal in Log Space and Time. In: Cachin C and Camenisch JL (eds) *Advances in Cryptology - EUROCRYPT 2004*. Berlin, Heidelberg: Springer, 541–554.

van Tilborg HCA and Jajodia S (2011) *Encyclopedia of Cryptography and Security*. Boston, MA: Springer US. Available at: http://link.springer.com/10.1007/978-1-4419-5906-5 (accessed 01/05/20).

Union PO of the E (2019) *Εγχειρίδιο σχετικά με την ευρωπαϊκή νομοθεσία για την προστασία των προσωπικών δεδομένων : έκδοση 2018.* Website. Publications Office of the European Union. Available at: http://op.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1/language-el (accessed 01/05/20).

Verheul E. and Jacobs B (2017) Polymorphic Encryption and Pseudonymisation in Identity Management and Medical Research. *172*. Available at: https://repository.ubn.ru.nl/handle/2066/178461 (accessed 29/04/20).

Verheul Eric, Jacobs B, Meijer C, Hildebrandt M and Ruiter J de (2016) *Polymorphic Encryption and Pseudonymisation for Personalised Healthcare*. . Available at: https://eprint.iacr.org/2016/411 (accessed 18/04/20).

Verma K (2019) *Symmetric, Asymmetric and Hybrid Encryption*. *Medium*. Available at: https://medium.com/@kapilvermarbl/symmetric-asymmetric-and-hybrid-encryption-25d57c1c327b (accessed 30/04/20).

ΓΚΠΔ (2016) Κανονισμός (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 27ης Απριλίου 2016, για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων αυτών και την κατάργηση της οδηγίας 95/46/ΕΚ (Γενικός Κανονισμός για την Προστασία

Δεδομένων) (Κείμενο που παρουσιάζει ενδιαφέρον για τον ΕΟΧ). *OJ L*. Available at: http://data.europa.eu/eli/reg/2016/679/oj/ell (accessed 02/01/20).

Παράρτημα Α

Κώδικας σε Python

Στο Παράρτημα αυτό παρουσιάζεται ο κώδικας σε γλώσσα προγραμματισμού Python που χρησιμοποιήθηκε για την υλοποίηση στο Κεφάλαιο 5.

Α.1 Η Κλάση Person

Σε αυτή την Ενότητα του Παραρτήματος αυτού παρουσιάζεται ο κώδικας υλοποίησης της κλάσης Person.

Α.1.1 Βιβλιοθήκες, Μεταβλητές και Μέθοδος Δημιουργού (Constructor)

```
import hashlib
import hmac
import time
from random import random
import os.path
from os import path

class Person:
    organisation_list=[]          #list with the names of the organisations
    private_key_list=[]           #list with the personal identifiers (contains lists with the personal identifiers)
```

```
    organisation_identifiers_list=[]  #list with the organisation identifiers in accordance with the organisations in the the organisation_list


    dummy_organisations=0           #number of dummy organisations created to fill the merkle tree



    #initialisation function that reads the data froma a file or requests the name of the organisation, the id that is used for the person and
    #the private key that the person will use for the creation of the private id in one Merkle tree
    def __init__(self):
        if path.exists("myfile.txt"):
            file1 = open("myfile.txt","r")


            f=file1.readlines()
            self.dummy_organisations=int(f.copy()[len(f)-1])
            f.pop()
            list1=[]


            for x in f:
              list1.append(x[0:-1])


            self.organisation_list=list1.copy()[0:int((len(list1.copy())/3))]

self.organisation_identifiers_list=list1.copy()[int((len(list1.copy())/3)):int(((len(list1.copy())/3)*2))]
            self.private_key_list=list1.copy()[int(((len(list1.copy())/3)*2)):]
            file1.close()
        else:
```

```
        self.organisation_list.append(str(input("Give Organisation name:  ")))
        self.organisation_identifiers_list.append(str(input("Give organisation identifier:
")))
        self.private_key_list.append(str(input("Give private key:  ")))
        self.dummy_organisations=1
```

A.1.2 Μέθοδος Αποθήκευσης σε Αρχείο

```
#function to save in a file the current data
   def save_data(self):
      list1=[]
      for i in self.organisation_list.copy():
         list1.append(i)
      for i in self.organisation_identifiers_list.copy():
         list1.append(i)
      for i in self.private_key_list.copy():
         list1.append(i)
      list1.append(str(self.dummy_organisations))
      file1 = open("myfile.txt","w")

      for i in list1:
         file1.writelines(i+"\n")
      file1.close()
```

A.1.3 Μέθοδοι Προσθαφαίρεσης Οντοτήτων

```
#function to be used only for testing
   def add_organisation_without_checking(self,orgname,orgid,pkey):
      if self.dummy_organisations!=0:
         for k in range(self.dummy_organisations):
```

```
            self.organisation_list.pop()
            self.organisation_identifiers_list.pop()
            self.private_key_list.pop()
            self.dummy_organisations-=1

        self.organisation_list.append(orgname)
        self.organisation_identifiers_list.append(orgid)
        self.private_key_list.append(pkey)
        k=1
        while(len(self.organisation_list)>(2**k)):
            k+=1
        difference=(2**k)-(len(self.organisation_list))
        for i in range (difference):
            self.add_dummy_organisation()
            self.dummy_organisations+=1

    #function to add one extra organisation.Needs the name of the organisation, the id that is used for the person and
    #the private key that the person will use for the creation of the private id in one Merkle tree
    def add_organisation(self,orgname,orgid,pkey):
        if self.dummy_organisations!=0:
            for k in range(self.dummy_organisations):
                self.organisation_list.pop()
                self.organisation_identifiers_list.pop()
                self.private_key_list.pop()
                self.dummy_organisations-=1

        self.organisation_list.append(orgname)
        self.organisation_identifiers_list.append(orgid)
```

```
    self.private_key_list.append(pkey)
    k=1
    while(len(self.organisation_list)>(2**k)):
        k+=1
    difference=(2**k)-(len(self.organisation_list))
    for i in range (difference):
        self.add_dummy_organisation()
        self.dummy_organisations+=1

    for i in range (len(self.organisation_list)-self.dummy_organisations):
        self.pseudonym_verification(self.organisation_list[i],self.organisation_list[i])

#function to remove one organisation
def remove_organisation(self,orgname):
    if self.dummy_organisations!=0:
        for k in range(self.dummy_organisations):
            self.organisation_list.pop()
            self.organisation_identifiers_list.pop()
            self.private_key_list.pop()
            self.dummy_organisations-=1

    x=self.organisation_list.index(orgname)
    self.organisation_list.remove(self.organisation_list[x])
    self.organisation_identifiers_list.remove(self.organisation_identifiers_list[x])
    self.private_key_list.remove(self.private_key_list[x])

    k=1
    while(len(self.organisation_list)>(2**k)):
        k+=1
```

```
        difference=(2**k)-(len(self.organisation_list))
        for i in range (difference):
            self.add_dummy_organisation()
            self.dummy_organisations+=1

        for i in range (len(self.organisation_list)-self.dummy_organisations):
            self.pseudonym_verification(self.organisation_list[i],self.organisation_list[i])

#function to add one extra dummy organisation
    def add_dummy_organisation(self):
        self.organisation_list.append(str(random()))
        self.organisation_identifiers_list.append(str(random()))
        self.private_key_list.append(str(random()))
```

A.1.4 Μέθοδος Παραγωγής Δέντρου Merkle και Μονοπατιού

```
#function to create a Merkle tree for the organisation in the given position in the
organisation_list
    def create_merkle_tree(self,org_place):
        list1=[]
        list3=[]

        for k in range(len(self.organisation_list)):

list1.append(hashlib.sha256(self.organisation_identifiers_list[k].encode()).hexdigest())
            list1.append(hashlib.sha256(hmac.new(self.private_key_list[org_place].encode(),
self.organisation_identifiers_list[k].encode(), hashlib.sha256).digest()).hexdigest())

        list3.append(list1.copy())
        list1.clear()
```

```
        for k in range (len(list3)):
            list1.append(list3[k].copy())
            length=len(list3[k])

            list4=list3[k].copy()
            while length>1:
                list2=[]
                for i in range(0,length,2):

                    list2.append(hashlib.sha256((list4[i]+list4[i+1]).encode()).hexdigest())



                length=len(list2)
                list1.append(list2)
                list4=list2.copy()
        return list1.copy()

#function to return the path of hashes for the current organisation in the merkle tree of the target organisation
    def get_verification_path(self,organisation,target_organisation):
        if (organisation in self.organisation_list) and (target_organisation in self.organisation_list):
            org=self.organisation_list.index(organisation)

            torg=self.organisation_list.index(target_organisation)
            list1=self.create_merkle_tree(torg).copy()
            list2=[]
            list2.append(list1[0][((org+1)*2)-1])
            i=org+1
            for k in range(1,len(list1)-1,1):
```

```
            if i%2!=0:
                list2.append("right")
                list2.append(list1[k][i])
                i=int((i+1)/2)
            else:
                list2.append("left")
                list2.append(list1[k][i-2])
                i=int((i/2))
        list2.append(list1.copy()[len(list1)-1][0])

        return list2.copy()


    else:
        print("There is no ", organisation, "or", target_organisation)
        return 0
```

A.1.5 Μέθοδος Επιβεβαίωσης Μονοπατιού

```
#function that takes the path from the get_verification_path function constructs the
pseudonym using that path checks it and checks it with the root
   #of target organisation's merkle tree
   def pseudonym_verification(self,organisation,target_organisation):
      if (organisation in self.organisation_list) and (target_organisation in
self.organisation_list):
         org=self.organisation_list.index(organisation)
         torg=self.organisation_list.index(target_organisation)
         l=self.get_verification_path(organisation,target_organisation)

pseudo=hashlib.sha256((hashlib.sha256(self.organisation_identifiers_list[org].encode()
).hexdigest()+l[0]).encode()).hexdigest()
```

```
        for k in range(2,len(l),2):
            if l[k-1]=="right":
                pseudo=hashlib.sha256((pseudo+l[k]).encode()).hexdigest()
            elif l[k-1]=="left":
                pseudo=hashlib.sha256((l[k]+pseudo).encode()).hexdigest()

        if l[len(l)-1]==pseudo:
            print("The pseudonym created by the path for the organisation
"+self.organisation_list[org]+" matches with the pseudonym in the path")
        else:
            print("failure")

        k=0

        list1=self.create_merkle_tree(torg).copy()

        while len(list1[k])!=1:
            k=k+1

        if list1[k][0]==pseudo:
            print("The pseudonym created by the path for the organisation
"+self.organisation_list[org]+" matches with the pseudonym in the tree of
"+self.organisation_list[torg])
        else:
            print("failure")
        print (list1[k][0],"\n",pseudo)
      else:
        print("There is no ", organisation, "or", target_organisation)
```

A.2 Κώδικας Παρουσίασης Λειτουργίας

```
p=Person()

p.add_organisation_without_checking(str(input("Give Organisation name:
")),str(input("Give organisation identifier:  ")), str(input("Give private key:  ")))

p.add_organisation_without_checking(str(input("Give Organisation name:
")),str(input("Give organisation identifier:  ")),str(input("Give private key:  ")))

print("Give 2 organisation names")

p.pseudonym_verification(str(input("Give Organisation name requesting access:
")),str(input("Give target Organisation's name:  ")))
```

A.3 Κώδικας για τις Μετρήσεις Επιβεβαίωσης Μονοπατιού

```
p=Person()

file1 = open("testpath.txt","a")

for i in range (8191):

   p.add_organisation_without_checking(str(random()),str(random()), str(random()))

   if (len(p.organisation_list)-p.dummy_organisations) in
[2,4,8,16,32,64,128,256,512,1024,2048,4096,8192]:

      test=0

      for k in range (10):

         t1=time.time()

         p.pseudonym_verification("amka","amka")

         test=test+(time.time()-t1)

      file1.writelines(str(len(p.organisation_list))+"\t"+ str(test/10)+"\n")


file1.close()
```

A.4 Κώδικας για τις Μετρήσεις Ενημέρωσης Οντοτήτων

```
p=Person()
file1 = open("testaddition.txt","a")
for i in range (8191):


   if ((len(p.organisation_list)-(p.dummy_organisations+1)) in
[2,4,8,16,32,64,128,256,512,1024,2048,4096,8192]) or(len(p.organisation_list))==1:
      test=0
      t1=time.time()
      p.add_organisation(str(random()),str(random()), str(random()))
      test=test+(time.time()-t1)
      file1.writelines(str(len(p.organisation_list))+"\t"+ str(test)+"\n")
   else:
      p.add_organisation_without_checking(str(random()),str(random()), str(random()))


file1.close()
```