Πανεπιστήμιο Πειραιώς

Τμήμα
Ψηφιακών Συστημάτων

Σχολή
Πρόγραμμα Μεταπτυχιακών Σπουδών

Μεταπτυχιακό Δίπλωμα Ειδίκευσης
Τεχνοοικονομική Διοίκηση και Ασφάλεια Ψηφιακών Συστημάτων

Κατεύθυνση
Ασφάλεια Ψηφιακών Συστημάτων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Προστασία Εμπιστευτικών Πληροφοριών:
Η Τεχνολογία Πρόληψης Διαρροής Δεδομένων
- (Data Leakage Prevention- DLP)-
στην υπηρεσία της Ασφάλειας Πληροφοριών

ΟΝΟΜΑ ΦΟΙΤΗΤΗ

Τσιμπερδώνης Κωνσταντίνος

ΟΝΟΜΑ ΕΠΙΒΛΕΠΟΝΤΟΣ ΚΑΘΗΓΗΤΗ

Ξενάκης Χρήστος

ΑΘΗΝΑ, 2013

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Προστασία Εμπιστευτικών Πληροφοριών:
Η Τεχνολογία Πρόληψης Διαρροής Δεδομένων
- (Data Leakage Prevention- DLP)-
στην υπηρεσία της Ασφάλειας Πληροφοριών

ΟΝΟΜΑ ΦΟΙΤΗΤΗ

Τσιμπερδώνης Κωνσταντίνος

Επιβλέπων Διπλωματικής Εργασίας

Χρήστος Ξενάκης, Επίκουρος Καθηγητής

Περιεχόμενα

Πρόλογος

Στη σημερινή εποχή, οι πληροφορίες και η απώλειά τους αποτελούν σημαντική απειλή για εταιρίες και οργανισμούς, καθώς ο αριθμός των περιστατικών αλλά και το κόστος που επιφέρουν συνεχίζουν να αυξάνονται. Είτε πρόκειται για κακόβουλες είτε για ακούσιες πράξεις, η απώλεια δεδομένων μπορεί να βλάψει το κύρος και τη φήμη της εταιρίας και να μειώσει την αξία της μετοχής της. Για την αντιμετώπιση και καταστολή αυτών των φαινομένων μελετώνται και ερευνώνται τεχνολογίες πρόληψης απώλειας δεδομένων. Η παρούσα διπλωματική ασχολείται με την προστασία εμπιστευτικών πληροφοριών μέσω τεχνολογιών πρόληψης διαρροής δεδομένων.

Στο κεφάλαιο 1 γίνεται περιγραφή των βασικών ζητημάτων ασφάλειας πληροφοριών, δίνεται η βασική ορολογία των στοιχείων που απαρτίζουν αυτά τα ζητήματα και μια αρχική ταξινόμηση των περιστατικών απώλειας δεδομένων σε υπολογιστές και δίκτυα οργανισμών.

Στο κεφάλαιο 2 ορίζονται τα εμπιστευτικά δεδομένα και η πνευματική ιδιοκτησία τους, ενώ παράλληλα περιγράφονται οι νομοθεσίες σε ευρωπαϊκό και ελληνικό επίπεδο που διέπουν την προστασία τους.

Στο κεφάλαιο 3 ορίζεται η απώλεια δεδομένων ως περιστατικό, παρουσιάζονται μερικά χαρακτηριστικά παραδείγματα απώλειας δεδομένων, ενώ επίσης προτείνεται ένα πολυεπίπεδο μοντέλο προστασίας που μπορεί να συνδυαστεί με άλλα βασικά μέτρα προστασίας.

Στο κεφάλαιο 4 γίνεται μια ταξινόμηση των συστημάτων πρόληψης απώλειας δεδομένων με βάση την κατάσταση τους, τη θέση ανάπτυξης του συστήματος, ενώ περιγράφονται και δυο προσεγγίσεις διαχείρισης απώλειας δεδομένων.

Στο κεφάλαιο 5 γίνεται μια επισκόπηση των εμπορικά διαθέσιμων συστημάτων DLP και του τρόπου που προστατεύουν τις τρεις βασικές κατηγορίες δεδομένων, δηλαδή των σε κίνηση, ακινησία και σε χρήση.

Στο κεφάλαιο 6 παρουσιάζονται μερικές ακαδημαϊκές μελέτες στο πεδίο ανίχνευσης διαρροής δεδομένων και τις μεθόδους πρόληψής τους.

Στο κεφάλαιο 7 παρουσιάζονται οι βασικές κατηγορίες περιστατικών απώλειας δεδομένων με λεπτομερή ανάλυση των παραμέτρων και των στοιχείων που τις απαρτίζουν.

Στο κεφάλαιο 8 περιγράφεται ο όρος «ανωνυμία δεδομένων», οι βασικές τεχνικές ανωνυμίας και ορίζονται οι βασικές έννοιες των μοντέλων ιδιωτικότητας.

Στο κεφάλαιο 9 δίνονται τρεις χαρακτηριστικές μελέτες περιπτώσεων στον τομέα απώλειας δεδομένων και οι προτεινόμενες μέθοδοι για την ελαχιστοποίηση της απειλής.

Στο κεφάλαιο 10 γίνεται αναφορά για τις διάφορες περιπτώσεις κακόβουλων χρηστών που απειλούν άμεσα τους οργανισμούς με απώλεια ή διαρροή δεδομένων.

Στο κεφάλαιο 11 παρουσιάζονται οι βασικές ενέργειες πρόληψης απώλειας δεδομένων, ώστε η προστασία έναντι απώλειας δεδομένων να είναι πιο ολοκληρωμένη. Πιο συγκεκριμένα εξετάζονται τα βασικά βήματα που πρέπει να ακολουθηθούν για τη δόμηση ενός συστήματος DLP.

Στο κεφάλαιο 12 παρουσιάζονται συνοπτικά τα «πρέπει» που πρέπει να ληφθούν υπόψη για τη δημιουργία ενός αποτελεσματικού συστήματος πρόληψης απώλειας δεδομένων.

Περίληψη

Η λύση για την πρόληψη απώλειας/διαρροής δεδομένων, είναι ένα σύστημα που έχει σχεδιαστεί για την ανίχνευση παραβιάσεων ή μεταφοράς δεδομένων καθώς και την πρόληψη μέσω της παρακολούθησης, εντοπισμού και του αποκλεισμού ευαίσθητων δεδομένων, είτε σε χρήση (ενέργειες στο τελικό σημείο), είτε σε κίνηση (κίνηση στο δίκτυο) , είτε σε ακινησία (αποθήκευση δεδομένων). Σε περιστατικά απώλειας δεδομένων, ευαίσθητα δεδομένα αποκαλύπτονται σε μη εξουσιοδοτημένο προσωπικό, είτε από δόλο ή από ακούσιο σφάλμα. Τα ευαίσθητα δεδομένα μπορούν να περιέχουν ιδιωτικά ή εταιρικά δεδομένα, πνευματικής ιδιοκτησίας, οικονομικά ή ιατρικά καθώς και λοιπά, ανάλογα με τον οργανισμό, την επιχείρηση ή τη βιομηχανία.
Οι λύσεις DLP ταξινομούνται με βάση διάφορες παραμέτρους όπως η πηγή της διαρροής, η κατάσταση των δεδομένων, το κανάλι της διαρροής, τις προσεγγίσεις πρόληψης και εντοπισμού καθώς και τις ενέργειες που πρέπει να γίνουν με την εκδήλωση της διαρροής. Από την άλλη, τα εμπορικά διαθέσιμα προϊόντα DLP εξετάζονται από τη σκοπιά των επαγγελματικών ερευνητικών αναφορών αλλά και των πληροφοριών από επίσημους ιστοτόπους προμηθευτών - κατασκευαστών.
Εξάλλου οι διάφορες λύσεις και τεχνολογίες πρόληψης απώλειας δεδομένων αποτελούν αντικείμενο ακαδημαϊκής μελέτης και στη συνέχεια ομαδοποιούνται ανάλογα με τη φύση της διαρροής και της προστασίας που παρέχεται. Σύμφωνα με αυτά, παρουσιάζονται διάφορα περιστατικά απώλειας δεδομένων και οι προτεινόμενες λύσεις τους, είτε για τη μείωση της πιθανότητας εμφάνισης τους είτε για την ελαχιστοποίηση του αντίκτυπου που θα επιφέρει στον οργανισμό.

Λέξεις Κλειδιά
DLP, Πρόληψη απώλειας δεδομένων, Απώλεια Δεδομένων, Διαρροή, Ασφάλεια, Εμπιστευτικότητα, Ιδιωτικότητα

Summary

Information and data leakage pose a serious threat to companies and organizations as the number of leakage incidents and the cost they inflict continues to increase. Whether caused by malicious intent or by an inadvertent mistake, data loss can diminish a company's brand, reduce shareholder value, and damage the company's goodwill and reputation. Data Loss Prevention (DLP) has been studied both in academic research areas and in practical application domains.
Data loss prevention solution is a system that is designed to detect potential data breach / data ex-filtration transmissions and prevent them by monitoring, detecting & blocking sensitive data while in-use (endpoint actions), in-motion (network traffic), and at-rest (data storage). In data leakage incidents, sensitive data is disclosed to unauthorized personnel either by malicious intent or inadvertent mistake. Such sensitive data can come in the form of private or company information, intellectual property, financial or patient information, credit-card data, and other information depending on the business and the industry.
The terms "data loss" and "data leak" are closely related and are often used interchangeably, though they are somewhat different. Data loss incidents turn into data leak incidents in cases where media containing sensitive information is lost and subsequently acquired by unauthorized party.
This thesis provides a structural and comprehensive overview of current research and practical solutions in the DLP area. Existing solutions have been grouped into different categories based on a taxonomy described in the book. The taxonomy presented characterizes DLP solutions according to various aspects such as leakage source, data state, leakage channel, deployment scheme, prevention and detection approaches, and action taken upon leakage. In the commercial section solutions offered by the leading DLP market players are reviewed based on professional research reports and material obtained from vendor Web sites.
In the academic section available academic studies have been clustered into various categories according to the nature of the leakage and the protection provided. Next, the main data leakage scenarios are described, each with the most relevant and applicable solution or approach that will reduce the likelihood or impact of data leakage.
In addition, several case studies of data leakage and data misuse are presented. Finally, the related research areas of privacy, data anonymization, training employees and the general characteristics of an effective DLP system are discussed.

Keywords
DLP, data loss prevention, Data loss, Leakage, Security, Trustworthiness

Εισαγωγή

Κάθε οργανισμός έχει υποστεί το κόστος της απώλειας εμπιστευτικών πληροφοριών, απλά δεν γνωρίζει ποια δεδομένα, τον τρόπο και τη χρονική στιγμή απώλειάς τους. Ακόμα και όταν το ανακαλύπτει προσπαθεί να περιορίσει την έκταση του συμβάντος, ώστε να μην γνωστοποιηθεί σε πελάτες, σε συνεργάτες και στο ευρύ κοινό. Ανεξαρτήτως του αν οι πληροφορίες χάθηκαν λόγω λάθους ή κακόβουλων ενεργειών, το αποτέλεσμα για τον οργανισμό μπορεί να είναι καταστροφικό. Απώλεια οικονομικών πληροφοριών, στοιχείων πελατών, δεδομένων πνευματικής ιδιοκτησίας, επιχειρηματικών πλάνων ανάπτυξης, προσωπικών στοιχείων εργαζομένων, πληροφορίες σύνθεσης προϊόντων, πληροφορίες διαφημιστικής εκστρατείας, κυρώσεις μη συμμόρφωσης με σχετικές νομοθεσίες και κανονισμούς, δυσμενής δημοσιότητα καθώς και απώλεια μελλοντικών εσόδων είναι ορισμένα από τα αποτελέσματα που δεν μπορεί να μετρηθεί επακριβώς το κόστος τους.

Σε παγκόσμια κλίμακα, μεγάλοι κυβερνητικοί, στρατιωτικοί και ιδιωτικοί οργανισμοί έχουν υποστεί το πλήγμα της απώλειας πληροφοριών αλλά και της δυσφήμισης λόγω υποχρεωτικής κοινοποίησης των συμβάντων ή μη δυνατότητας απόκρυψης αυτών και αδιαμφισβήτητα χιλιάδες άλλοι που δεν το κοινοποίησαν ποτέ. Με αυτά τα δεδομένα, εύλογα δημιουργείται η απορία πως είναι δυνατόν όταν επενδύονται, και ειδικότερα σε μεγάλους οργανισμούς, τεράστια κονδύλια για την εφαρμογή μηχανισμών προστασίας υποδομών IT, πρόσληψη εξειδικευμένου ανθρώπινου δυναμικού διαχείρισης πληροφοριακών συστημάτων, εφαρμογή αυστηρών πολιτικών και διαδικασιών ασφάλειας πληροφοριών, να παρουσιάζονται με αυτή την συχνότητα τέτοια συμβάντα.

Βασική αιτία είναι ότι οι επιχειρήσεις έχουν εστιάσει σε μέτρα προστασίας έναντι εξωτερικών απειλών, αδυνατώντας να αναγνωρίσουν ότι ο πραγματικός κίνδυνος ευρίσκεται εντός των τειχών. Οι επιχειρήσεις παραδοσιακά επένδυαν και επενδύουν σε τεχνολογικές λύσεις που θα εμποδίσουν τους κακόβουλους εισβολείς να εισέλθουν στο εταιρικό δίκτυο και θα αποτρέψουν την εκτέλεση κακόβουλων εισερχόμενων προγραμμάτων, όπως viruses, Trojan horses, worms κτλ με τεχνολογίες όπως συστήματα anti-virus, anti-spam, firewalls κ.ά. Αυτή η προσέγγιση όμως στην προστασία των εταιρικών δεδομένων, έχει δημιουργήσει κενά ασφάλειας σε εσωτερικές απειλές, είτε αυτές είναι κακόβουλες ή από λάθος, με αποτέλεσμα σημαντικές απώλειες δεδομένων να οφείλονται σε εσωτερικούς χρήστες.

Για την προστασία των δεδομένων από εσωτερικές απειλές οι επιχειρήσεις υιοθετούν κυρίως τεχνολογίες όπως συστήματα διαχείρισης ταυτότητας, συστήματα ελέγχου πρόσβασης και κρυπτογράφηση δεδομένων, οι οποίες όμως παρέχουν προστασία μόνο έναντι μη εξουσιοδοτημένων εσωτερικών χρηστών. Τα μέτρα αυτά δεν αποτρέπουν την απώλεια πληροφοριών από πιστοποιημένους χρήστες στους οποίους έχει δοθεί πρόσβαση στα δεδομένα.

Για την αντιμετώπιση αυτών των περιστατικών έχουν αναπτυχθεί νέες τεχνολογίες προστασίας απώλειας δεδομένων (Data Loss Prevention - DLP) που επιτρέπουν στους διάφορους οργανισμούς την εφαρμογή εταιρικών πολιτικών ασφάλειας, την εποπτεία και αναφορά συμβάντων και την αποτροπή αυτών. Οι τεχνολογίες DLP βασίζονται σε αναγνώριση ακολουθίας δεδομένων (pattern analysis) όπως πιστωτικών καρτών, τραπεζικών λογαριασμών, αστυνομικών ταυτοτήτων κλπ, σε αναγνώριση περιεχομένου (content analysis) εξ αιτίας

χαρακτηρισμού δεδομένων ως εμπιστευτικού, απόρρητου, κοινού κλπ, ή και λόγω θέσης αποθήκευσης στο εταιρικό δίκτυο. Οι τεχνολογίες DLP παρέχουν προστασία σε πολλαπλά επίπεδα, όπως σε επίπεδο δικτύου και σε επίπεδο σταθμών εργασίας χρηστών, επιτρέποντας τον έλεγχο, την αναφορά και την αποτροπή απώλειας αυτών, όπως μέσω εκτύπωσης, πρόσβασης μη εξουσιοδοτημένων εφαρμογών, μεταφοράς μέσω δικτύων (email, web, p2p, wired & wireless), καθώς και θωράκισης των πληροφοριών έναντι αντιγραφής τους σε μεταφέρσιμα αποθηκευτικά μέσα, όπως μονάδες USB, CDs/DVDs, συσκευές Bluetooth κ.α.

1. Βασικά ζητήματα Ασφάλειας Πληροφοριών

Στην ενότητα αυτή επιχειρείται μια σύντομη αναφορά και αποσαφήνιση των όρων που σχετίζονται με ζητήματα της Ασφάλειας των Πληροφοριών τα οποία εμπλέκονται άμεσα με το προς ανάλυση θέμα δηλ. της Πρόληψης Διαρροής Δεδομένων.

Στην πλειοψηφία των περιπτώσεων, η πρώτη επαφή που έχει κάποιος με την ασφάλεια υπολογιστών είναι όταν προσπαθεί να συνδεθεί με έναν υπολογιστή, ο οποίος ζητά το όνομα χρήστη και το συνθηματικό του. Το πρώτο βήμα (δηλαδή η αίτηση για το όνομα χρήστη) στοχεύει στην αναγνώριση της ταυτότητας και το δεύτερο βήμα (δηλαδή η αίτηση για το συνθηματικό) στοχεύει στην επαλήθευση της ταυτότητας, όπου αποδεικνύεται ότι όντως είναι αυτός που ισχυρίζεται ότι είναι.

Η αναγνώριση και η επαλήθευση ταυτότητας μέσω ονόματος χρήστη και συνθηματικού παρέχουν μονομερή αυθεντικοποίηση. Ο χρήστης εισάγει ένα όνομα και ένα συνθηματικό και ο υπολογιστής επαληθεύει την ταυτότητα του χρήστη. Ο χρήστης όμως δεν έχει καμιά ένδειξη, πόσο μάλλον εγγύηση, για την ταυτότητα του μέρους που βρίσκεται στην άλλη άκρη της γραμμής. Αυτό είναι πραγματικό πρόβλημα και οδηγεί στη δεύτερη μέθοδο παραβίασης συνθηματικών. Σε μια επίθεση υποκλοπής, ο επιτιθέμενος, που μπορεί να είναι νόμιμος χρήστης, τρέχει ένα πρόγραμμα που παρουσιάζει μια ψεύτικη οθόνη σύνδεσης σε κάποιο τερματικό. Ένας ανυποψίαστος χρήστης έρχεται στο τερματικό αυτό και προσπαθεί να συνδεθεί. Το θύμα οδηγείται μέσω του κανονικού μενού σύνδεσης και του ζητείται το όνομα χρήστη και το συνθηματικό του. Αυτά αποθηκεύονται σε χώρο προσπελάσιμο από τον επιτιθέμενο. Στη συνέχεια, ο έλεγχος είτε περνά στο χρήστη ή εμφανίζεται ένα ψεύτικο μήνυμα λάθους σύνδεσης και το πρόγραμμα υποκλοπής τερματίζει. Ο έλεγχος μεταφέρεται στο λειτουργικό σύστημα, που τώρα παρουσιάζει στο χρήστη την αληθινή οθόνη σύνδεσης. Ο χρήστης ξαναπροσπαθεί, πετυχαίνει και μπορεί να μείνει με παντελή άγνοια του γεγονότος ότι το συνθηματικό του υποκλάπηκε.

Η ασφάλεια υπολογιστών αναφέρεται στον έλεγχο πρόσβασης σε πληροφορίες. Ωστόσο επειδή ο έλεγχος πρόσβασης σε πληροφορίες μπορεί να είναι δύσκολος ή πολλές φορές ακόμα και αδύνατος, αντικαθιστούμε το στόχο αυτό με έναν πιο εύκολο όπως αυτό του ελέγχου της πρόσβασης σε δεδομένα. Η διαφοροποίηση ανάμεσα στα δεδομένα και τις πληροφορίες είναι λεπτή, ενώ η έλλειψη κατανόησής της είναι μια από τις συχνότερες αιτίες παρεξηγήσεων. Δεδομένα δεν είναι παρά τα φυσικά φαινόμενα που έχουν επιλεγεί δια συμφωνίας για να αντιπροσωπεύσουν συγκεκριμένες πτυχές του πραγματικού και νοητού μας κόσμου. Οι ερμηνείες που αντιστοιχίζουμε στα δεδομένα ονομάζονται πληροφορίες. Τα δεδομένα χρησιμοποιούνται για την αποθήκευση και μετάδοση της πληροφορίας καθώς και για την παραγωγή νέας πληροφορίας διαμέσου της επεξεργασίας των δεδομένων σύμφωνα με αυστηρούς κανόνες [3].

Στο χώρο της πληροφορικής, ο όρος ασφάλεια πληροφοριών αποτελεί έναν όρο για τον οποίο έχει επιτευχθεί ευρεία συναίνεση ως προς την ερμηνεία του από το σύνολο της επιστημονικής κοινότητας. Έτσι λοιπόν ασφάλεια πληροφοριών είναι η διατήρηση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητάς τους, όπως και της διαθεσιμότητας του συστήματος που χειρίζεται τις πληροφορίες [3]. Συγκεκριμένα, εμπιστευτικότητα είναι η ιδιότητα της πληροφορίας να προσπελαύνεται από εξουσιοδοτημένες οντότητες. Αυτό μπορεί να επιτευχθεί με

διάφορους μηχανισμούς όπως η κρυπτογράφηση και ο έλεγχος προσπέλασης. Ακεραιότητα είναι η ιδιότητα της πληροφορίας να τροποποιείται μόνο από εξουσιοδοτημένες οντότητες. Αυτό μπορεί να επιτευχθεί με μηχανισμούς όπως οι ψηφιακές υπογραφές και κώδικες πιστοποίησης μηνυμάτων (message authentication code - MAC). Διαθεσιμότητα είναι η ιδιότητα της πληροφορίας να καθίσταται διαθέσιμη προς χρήση από εξουσιοδοτημένες οντότητες μέσα σε λογικό χρόνο από την υποβολή της σχετικής αίτησης. Αντιστοίχως διαθεσιμότητα του συστήματος είναι η ιδιότητα του συστήματος να μπορεί να διαθέτει τους πόρους του στις εξουσιοδοτημένες προς αυτό οντότητες. Συνήθως ένα σύστημα διαθέτει πολλούς μηχανισμούς για να εξασφαλίσει την διαθεσιμότητά του, όπως η χρήση ανεξάρτητων πηγών ενέργειας και πολλαπλών γραμμών επικοινωνίας.

Ο έλεγχος πρόσβασης, η αυθεντικότητα, το απόρρητο και η μη άρνηση ταυτότητας, θεωρούνται επίσης αναπόσπαστα μέρη της ασφάλειας πληροφοριών. Πιο συγκεκριμένα, η μη άρνηση της ταυτότητας εξασφαλίζει τη δυνατότητα απόδοσης πράξεων σε ένα χρήστη. Ο έλεγχος πρόσβασης σχετίζεται με το ποιες πληροφορίες και υπηρεσίες μπορεί να προσπελάσει ένας χρήστης αφού προηγουμένως έχει ταυτοποιηθεί από το σύστημα. Σε πολλά συστήματα, κάθε εξουσιοδοτημένη ενέργεια του χρήστη καταγράφεται από έναν μηχανισμό ελέγχου (audit). Η αυθεντικοποίηση σχετίζεται με τον έλεγχο της ταυτότητας του χρήστη για παροχή πρόσβασης στα αγαθά του συστήματος (πληροφορίες και πόρους). Αυτό μπορεί να επιτευχθεί με την εισαγωγή από το χρήστη ενός μοναδικού μυστικού, όπως έναν κωδικό πρόσβασης (ο χρήστης αποδεικνύει «κάτι που ξέρει»), ή χρησιμοποιώντας ένα μοναδικό δείγμα που κατέχει ο χρήστης (ο χρήστης αποδεικνύει «κάτι που έχει»), ή με τη χρήση βιομετρικών μηχανισμών αναγνώρισης όπως δακτυλικά αποτυπώματα (ο χρήστης αποδεικνύει «κάτι που είναι»). Το απόρρητο σχετίζεται με την εξασφάλιση ότι ο χρήστης έχει τον έλεγχο των πληροφοριών που συλλέγονται γι αυτόν και εκτίθενται και σε άλλους. Είναι δύσκολο να προσδιοριστούν οι συγκεκριμένοι μηχανισμοί που απαιτούνται για να διαφυλαχτούν τα προσωπικά δεδομένα των χρηστών. Το όλο σύστημα θα πρέπει να είναι σχεδιασμένο με τέτοιο τρόπο ώστε το απόρρητο του χρήστη να μην παραβιάζεται.

Υπάρχουν πολλές μέθοδοι αναγνώρισης της ταυτότητας ενός ατόμου. Για το σκοπό αυτό μπορεί να χρησιμοποιηθεί:

- Η εμφάνιση, δηλαδή το ύψος, το βάρος, το φύλο, η όψη κ.τ.λ.
- Η κοινωνική συμπεριφορά, δηλαδή ο τρόπος αλληλεπίδρασης με τους άλλους.
- Το όνομα.
- Κάποιος κωδικός, όπως για παράδειγμα ένας αριθμός μητρώου.
- Η γνώση που έχει το άτομο σχετικά με κάτι.
- Η κατοχή κάποιου αντικειμένου.
- Η βιοδυναμική, δηλαδή πώς εκτελεί κάτι το άτομο.
- Η φυσιολογία του, όπως αυτή προκύπτει για παράδειγμα από τα χαρακτηριστικά του.
- Ξένα επιβληθέντα χαρακτηριστικά, όπως, για παράδειγμα βραχιόλια, ταυτότητες κτλ.

Ωστόσο, ο συντριπτικά πιο διαδεδομένος τρόπος ταυτοποίησης ατόμων από υπολογιστικά συστήματα είναι με τη χρήση κάποιου κωδικού, που συνήθως αναφέρεται ως όνομα χρήστη (username).

Υπάρχουν τέσσερις μέθοδοι για να αποδείξει κάποιος την ταυτότητά του:

- Με κάτι που ξέρει.
- Με κάτι που έχει.
- Με κάτι που αποτελεί μοναδικό ατομικό χαρακτηριστικό του.
- Με το πού βρίσκεται, σε σχέση με το πού αναμένεται να βρίσκεται.

Η πρώτη μέθοδος είναι ίσως και η ευκολότερη για χρήση όταν στη διαδικασία αυθεντικοποίησης εμπλέκονται μηχανές. Παραδείγματα σχετικών μέσων αυθεντικοποίησης είναι τα συνθηματικά, τα PINs (Personal Identification Numbers), οι συνθηματικές φράσεις, και πληροφορίες σχετικές με το άτομο ή την οικογένεια κάποιου που δεν είναι ευρέως γνωστές.

Τα πλεονεκτήματα της μεθόδου αυτής είναι:

- Το μέσο αυθεντικοποίησης είναι πάντα στην κατοχή του χρήστη.
- Το μέσο αυθεντικοποίησης μπορεί να αλλάξει εύκολα.
- Η προστασία του μέσου αυθεντικοποίησης είναι σχετικά εύκολη.
- Το μέσο αυθεντικοποίησης εισάγεται εύκολα στο μηχανισμό αυθεντικοποίησης μέσω πληκτρολογίου, χωρίς να υπάρχει ανάγκη προσθήκης εξειδικευμένου υλικού.

Ωστόσο, η μέθοδος έχει και μειονεκτήματα. Το βασικότερο είναι ότι η φιλοσοφία της στηρίζεται στο ότι ο αυθεντικοποιούμενος γνωρίζει κάτι, το οποίο μπορεί να ξεχαστεί, να αντιγραφεί ή ακόμη και να εικαστεί από κάποιον άλλο, μη εξουσιοδοτημένο να το κάνει. Σε πολλές περιπτώσεις, με χρήση ενός τέτοιου μηχανισμού δεν είναι ιδιαίτερα δύσκολο για κάποιον επιτιθέμενο να μάθει το μέσο αυθεντικοποίησης, απλώς παρακολουθώντας τον εξουσιοδοτημένο χρήστη να το εισάγει στο σύστημα. Επιπλέον, δεν απαιτούνται ειδικά εργαλεία, γνώσεις ή μέθοδοι για να αντιγράψει κανείς το μέσο αυθεντικοποίησης. Εκτός αυτών, είναι δυνατόν συνήθως να αποκαλυφθούν με χρήση αυτοματοποιημένων μεθόδων. Παρ' όλο λοιπόν που η μέθοδος αυτή είναι σε ευρύτατη χρήση σήμερα σε υπολογιστές, αυτόματες ταμπελικές μηχανές, τηλεφωνικές κάρτες κτλ., εκτιμάται ως ατελής.

Η δεύτερη μέθοδος δεν είναι τόσο επιρρεπής σε αντιγραφή όσο η πρώτη. Η μέθοδος αυτή βασίζεται στην κατοχή από τον αυθεντικοποιούμενο ενός αντικειμένου, π.χ. μιας κάρτας (απλής, μαγνητικής, έξυπνης), ενός κλειδιού ή μιας γεννήτριας πρόκλησης – απάντησης. Είναι φανερό ότι ο ιδιοκτήτης των αντικειμένων αυτών πρέπει να καταβάλει προσπάθεια να προστατεύσει από κλοπή ή απώλεια το αντικείμενο που έχει. Ακριβώς αυτά τα δύο προβλήματα είναι και τα βασικά μειονεκτήματα αυτής της μεθόδου αυθεντικοποίησης. Από την άλλη πλευρά, τέτοια αντικείμενα δεν είναι εύκολο να αντιγραφούν, χωρίς ωστόσο να είναι και εντελώς αδύνατη η αντιγραφή τους. Για να αποφύγουμε εκτεταμένο κίνδυνο αντιγραφής, θα πρέπει το κόστος της να είναι αρκετά μεγαλύτερο από το αναμενόμενο όφελος. Αν και η πρακτική αυτή αποτρέπει τον ευκαιριακό επιτιθέμενο, δεν αποτελεί ωστόσο σοβαρό εμπόδιο για τον αποφασισμένο.

Η τρίτη μέθοδος βασίζεται στην αναγνώριση και επαλήθευση ατομικών χαρακτηριστικών του αυθεντικοποιούμενου. Χαρακτηριστικά που έχουν κατά καιρούς προταθεί ως πιθανά για τέτοια χρήση κατηγοριοποιούνται σε φυσιολογικά (ή ανθρωπομετρικά) χαρακτηριστικά και σε χαρακτηριστικά συμπεριφοράς [1,4]. Τα συστήματα αυτά παρέχουν μεγαλύτερη ασφάλεια από το ανωτέρω, αλλά υπάρχει πρόβλημα στην κατασκευή φθηνών και αξιόπιστων συσκευών αναγνώρισης. Οι

συσκευές αυτές έχουν ένα αρκετά μεγάλο κόστος, ενώ ταυτόχρονα δεν είναι αλάνθαστες. Επιπλέον, επιβάλει ελέγχους στο χρήστη οι οποίοι δεν είναι εύκολα αποδεκτοί

Η τέταρτη μέθοδος απαιτεί από τον χρήστη τη σύνδεσή του στο σύστημα με τις ανωτέρω μεθόδους, από συγκεκριμένο σημείο π.χ. σύνδεση του χρήστη σε συγκεκριμένο τερματικό ή σε επόμενη φάση μέσω συστήματος GPS με αναγνώριση της γεωγραφικής μας θέσεως.

Η διασφάλιση της ασφάλειας του υπολογιστή είναι ένα εξαιρετικά δύσκολο έργο [2]. Ωστόσο, δεν είναι σαφές το πώς να χρησιμοποιήσει κάποιος τους διάφορους μηχανισμούς για να καλύψει τις απαιτήσεις ασφαλείας. Ο μηχανισμός ασφάλειας, σε πολλές περιπτώσεις μπορεί να γίνει το επόμενο ευαίσθητο μέρος του συστήματος. Για παράδειγμα, αναγκάζοντας τον χρήστη να χρησιμοποιεί ένα σύνθετο κωδικό πρόσβασης μπορεί να οδηγήσει το χρήστη να γράφει ένα σημείωμα με τον κωδικό πρόσβασης δίπλα στην οθόνη του υπολογιστή. Η ασφάλεια υπολογιστών είναι μια συνεχής μάχη μεταξύ των επιτιθέμενων, οι οποίοι εντοπίζουν νέες τρύπες ασφαλείας και τα τρωτά σημεία των συστημάτων ασφαλείας και του τμήματος του οργανισμού που πρέπει να τους εμποδίσει.

Ένα αποτελεσματικό πλαίσιο προστασίας των πληροφοριών σε μια σύγχρονη κοινωνία πρέπει να έχει τα εξής, τουλάχιστον, βασικά χαρακτηριστικά:

- Να είναι κοινωνικά αποδεκτό, δηλαδή η ανάπτυξή του να έχει βασιστεί στις απόψεις του κοινωνικού συνόλου.
- Να είναι πολυδύναμο, δηλαδή η ανάπτυξή του να έχει βασιστεί στην ανάλογη διεθνή εμπειρία και πρακτική.
- Να είναι πολυδιάστατο, δηλαδή να συνδυάζει θεσμικές ρυθμίσεις, οργανωτικές ρυθμίσεις και κοινωνικές δράσεις.

Είναι αυτονόητο ότι ένα τέτοιο πλαίσιο δεν είναι δυνατόν να είναι ταυτόχρονα αρκετά χρήσιμο στην πράξη και αρκετά γενικό ώστε να μπορεί να καλύψει όλα τα είδη της πληροφορίας και όλα τα είδη των συστημάτων που συλλέγουν, αποθηκεύουν, επεξεργάζονται και μεταδίδουν πληροφορίες. Η κατάσταση, λοιπόν, αποτυπώνεται σε δύο άξονες, όπως φαίνεται και στον Πίνακα 1.1: Ο ένας άξονας αντιπροσωπεύει πληροφοριακά συστήματα κατά τομείς (π.χ. Πληροφοριακά Συστήματα Υγείας, Πληροφοριακά Συστήματα Δημόσιας Διοίκησης, Πληροφοριακά Συστήματα Κοινωνικής Ασφάλισης, Νομικά Πληροφοριακά Συστήματα κτλ.) και ο άλλος άξονας δράσεις που πρέπει να αναληφθούν ώστε να αναγνωριστούν και να ικανοποιηθούν οι ανάγκες και οι απαιτήσεις προστασίας των σχετικών πληροφοριών.

ΤΟΜΕΙΣ ΕΦΑΡΜΟΓΗΣ	**ΔΡΑΣΕΙΣ**		
	Θεσμικές Ρυθμίσεις	Οργανωσιακές Ρυθμίσεις	Κοινωνικές Δράσεις
Υγεία	✓	✓	✓
Δημόσια Διοίκηση	✓	✓	✓
Κοινωνική Ασφάλιση	✓	✓	✓
Δικαιοσύνη	✓	✓	✓

Πίνακας 1.1: Πλαίσιο προστασίας πληροφοριών [3]

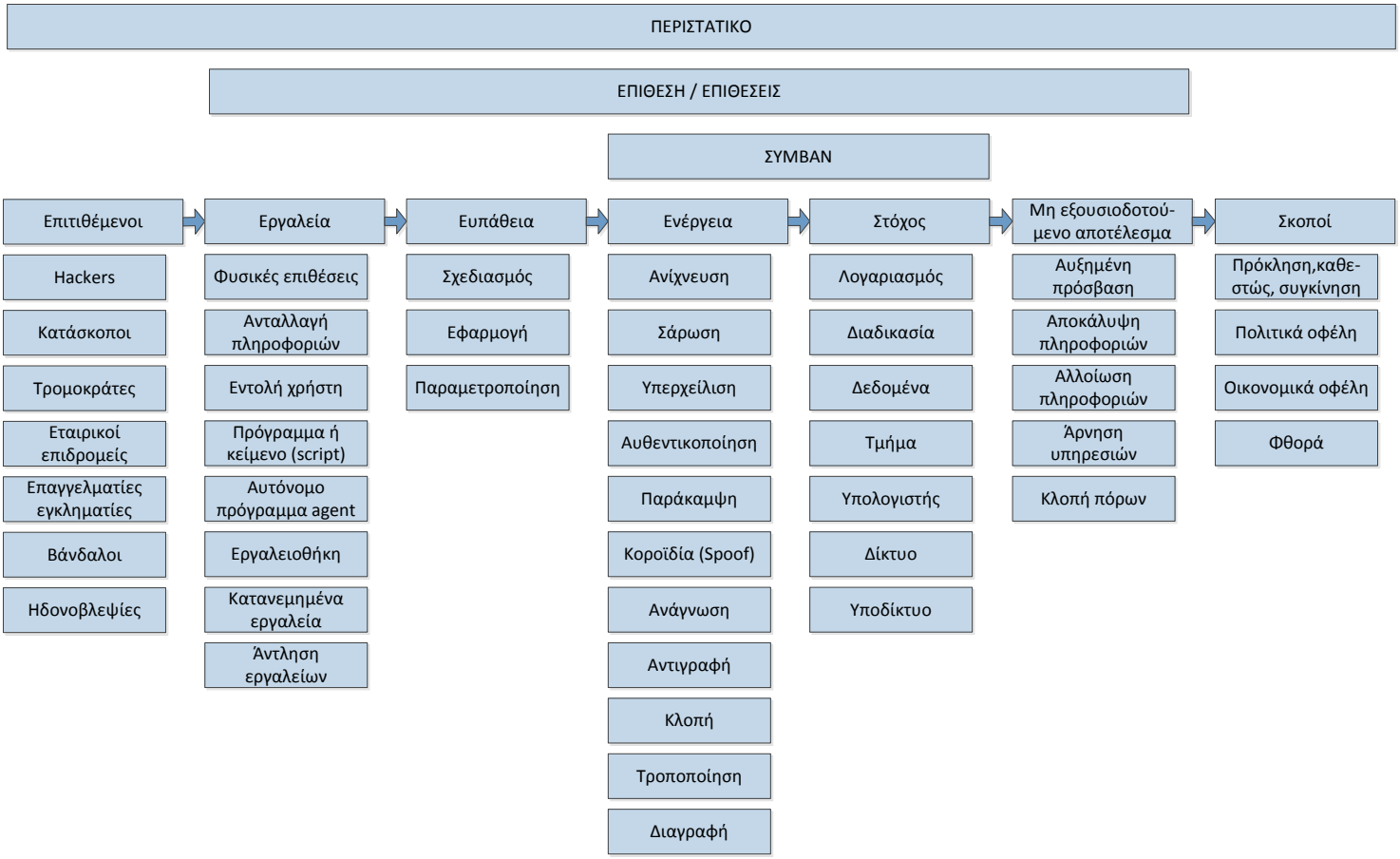
Κάποιες από τις δράσεις αυτές, καθώς και οι συνακόλουθες ρυθμίσεις μπορούν και πρέπει να είναι κοινές ανάμεσα στα διάφορα κατά τομείς πληροφοριακά συστήματα.

Η πολιτική που μπορεί να εφαρμόζει μια επιχείρηση είναι μια περιγραφή του συνόλου των κανόνων, των προτύπων και των διαδικασιών που καθορίζουν τα φυσικά, διαδικαστικά, τεχνικά και προσωπικά μέτρα ασφάλειας που λαμβάνονται στη διοίκηση, τη διανομή και την προστασία των περιουσιακών στοιχείων.

Μια πολιτική ασφάλειας, είτε εταιρική, είτε συστήματος, ακολουθεί μια ιεραρχική δομή που αποτελείται από κανόνες, πρότυπα και διαδικασίες. Μερικές φορές, κάποιες πολιτικές περιέχουν και οδηγίες. Οι κανόνες είναι ευρείας μορφής και με υψηλού επιπέδου διατυπώσεις που περιγράφουν τους σκοπούς της διοίκησης. Οι κανόνες απαντούν σε ερωτήσεις της μορφής «τι;», όπως, για παράδειγμα, «Τι θέλουμε να πετύχουμε;» ή «Με τι ρυθμιστικό πλαίσιο θέλουμε να λειτουργούμε;». Τα πρότυπα έχουν πιο συγκεκριμένες διατυπώσεις, που απαντούν σε ερωτήσεις της μορφής «πώς;», ενώ οι διαδικασίες απαιτούν ακόμη λεπτομερέστερες διατυπώσεις, που απαντούν σε ερωτήσεις της μορφής «πώς, με λεπτομέρεια;». Οι οδηγίες είναι συστάσεις για το πώς να γίνει κάτι και η συμμόρφωση μ' αυτές δεν είναι υποχρεωτική.

Η ιεραρχική αυτή δομή συνήθως υλοποιείται και παρουσιάζεται δημιουργώντας έναν κανόνα υψηλού επιπέδου που αναφέρεται σε πρότυπα, αλλά δεν τα συγκεκριμενοποιεί, περιγράφοντας τις λεπτομέρειες στα πρότυπα και στις διαδικασίες. Ένας τέτοιος κανόνας θα μπορούσε επίσης, εναλλακτικά, να οδηγεί σε ένα σύνολο υποκανόνων, που με τη σειρά τους, αναφέρονται σε πρότυπα και διαδικασίες. Πάντως είναι χρήσιμο να πούμε ότι δεν υπάρχει ένας και μοναδικός τρόπος σύνταξης πολιτικών. Ο συντάκτης πρέπει να προσπαθήσει να ισορροπήσει ανάμεσα στο πόση λεπτομέρεια και πόση γενικότητα θα δώσει στην πολιτική του, καθώς επίσης και να αποφασίσει πώς θα δομήσει την παρουσίασή του. Μια πολιτική ασφάλειας πρέπει τουλάχιστον να περιέχει σκοπό, πεδίο εφαρμογής, κανόνες, πρότυπα και διεργασίες. [1,3]

Στην ακόλουθη Εικόνα (1.1) παρουσιάζεται η πλειοψηφία των περιπτώσεων που μπορούν να εμφανιστούν σε περιστατικά Απώλειας Δεδομένων και εν γένει Ασφάλειας Πληροφοριών. Πιο συγκεκριμένα ταξινομήθηκαν οι πιθανοί επιτιθέμενοι, τα εργαλεία που συνήθως χρησιμοποιούν, το τρωτό σημείο του συστήματος που στοχεύουν, οι ενέργειες στις οποίες προβαίνουν, οι στόχοι τους, το αποτέλεσμα που επιφέρουν και οι σκοποί που έχουν εκ των προτέρων.



Εικόνα 1.1: Ταξινόμηση περιστατικών Απώλειας Δεδομένων σε δίκτυα και υπολογιστές [1]

2. Εμπιστευτικά Δεδομένα

2.1 Εισαγωγή

Εμπιστευτική ονομάζεται μια προσωπική πληροφορία όταν αφορά κάποιο πρόσωπο, τις αντιλήψεις του προσώπου, τον εκάστοτε «φορέα» της πληροφορίας, αλλά και το βαθμό εμπιστευτικότητας που χρήζουν οι εν λόγω πληροφορίες. Με τον τρόπο αυτό ορίζεται ένα πλαίσιο που καθορίζει την «ποιότητα της πληροφορίας». Υπό αυτήν την έννοια όλες οι προσωπικές πληροφορίες ποικίλλουν ανάλογα την ποιότητά τους, αλλά και το πώς θα τις διαχειριστούν τρίτοι. [5]
Λόγω της ραγδαίας εξέλιξης των νέων τεχνολογιών, οι πληροφορίες αποκτούν «υπόσταση» και άρα μπορούν να αποστέλλονται, να αναπαράγονται και να υφίστανται επεξεργασία. Η δυνατότητα επεξεργασίας κάθε προσωπικής πληροφορίας καθορίζεται κυρίως με τη σύνδεση της με κάποιο άτομο, που σημαίνει ότι έχει περιεχόμενο ιδιωτικού χαρακτήρα. [6]

Συνεπώς η νέα τεχνολογία, μέσω της αλλαγής των ρυθμίσεων και των αντιλήψεων, έχει επιφέρει αλλαγή και στο περιεχόμενο του όρου «προσωπική πληροφορία». Το «περιεχόμενο» φαίνεται να υπήρξε καταρχήν καθοριστικό για τη διατύπωση και την προστασία μιας πληροφορίας προσωπικού χαρακτήρα, υπό την έννοια ότι το ιδιαίτερο περιεχόμενο μιας πληροφορίας είναι εκείνο που δημιουργούσε την αναμονή ότι αυτή η πληροφορία θα θεωρηθεί ως αυστηρά προσωπική ή ευαίσθητη και για το λόγο αυτό θα έπρεπε να παρεμποδιστεί ή τουλάχιστον να περιοριστεί η συλλογή, η χρήση ή και η κυκλοφορία της.

Βάσει αυτής της αντίληψης γίνεται ο διαχωρισμός σε «αβλαβή δεδομένα», δεδομένα άνευ σημασίας και ταυτόχρονα «ευαίσθητα» δεδομένα που χρήζουν (ιδιαίτερης) προστασίας. [5,7]

Όσον αφορά τις κατηγορίες που υπάγονται στα ευαίσθητα δεδομένα, πρέπει να σημειωθεί ότι έχει διαμορφωθεί ένας στενός πυρήνας «ευαίσθητων δεδομένων» που αντανακλάται μάλιστα σε υπερεθνικά κείμενα, όπως η Σύμβαση 108 του Συμβουλίου της Ευρώπης για την προστασία του ατόμου έναντι της αυτοματοποιημένης προστασίας προσωπικών δεδομένων ή η Οδηγία 95/46/ΕΚ για την προστασία του ατόμου έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα. Λόγω του ότι οι αντιλήψεις περί ευαισθησίας ποικίλλουν από κοινωνία σε κοινωνία και από εποχή σε εποχή, υπάρχουν διαφοροποιήσεις στις επιμέρους εθνικές νομοθεσίες ιδίως στην φάση προ της έκδοσης της προαναφερόμενης κοινοτικής Οδηγίας και δεν εντοπίζεται ένα γενικότερο παραδεκτό κριτήριο κατάταξης των δεδομένων σε κατηγορίες.

Από τότε που καταγράφθηκε η ανάγκη νομοθετικής προστασίας της ιδιωτικότητας, άρχισαν να καταγράφονται και οι πρώτες αντιδράσεις στο πεδίο της προστασίας προσωπικών δεδομένων σε διεθνές επίπεδο. Η ανάγκη της ιδιωτικότητας διατυπώθηκε στη Σύμβαση της Ρώμης της 4ης Νοεμβρίου 1950 για την προστασία των ανθρωπίνων δικαιωμάτων και των θεμελιωδών ελευθεριών. Η Ευρωπαϊκή Σύμβαση των Δικαιωμάτων του Ανθρώπου (ΕΣΔΑ) του 1950 προστατεύει με το άρθρο 8 την ιδιωτική ζωή, στην οποία συγκαταλέγονται και τα προσωπικά δεδομένα. Ως προς τα ιατρικά δεδομένα το Δικαστήριο των Ανθρωπίνων Δικαιωμάτων όρισε

αυστηρές προϋποθέσεις για την ανακοίνωσή τους σε τρίτους. Οι πρώτες ανησυχίες για την ιδιωτικότητα τέθηκαν στον νόμο για την προστασία δεδομένων του 1970 (Hesse Data Protection Act 1970), στον Σουηδικό νόμο για την προστασία των δεδομένων του 1973 (Swedish Privacy Act 1973) και τον νόμο περί ιδιωτικότητας των ΗΠΑ του 1974 (US Privacy Act 1974), οι οποίοι έθεσαν τις απαιτήσεις χωρίς όμως να έχουν καμία εξουσία γύρω από την προστασία των δεδομένων. [6]

Ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (ΟΟΣΑ) ήταν ο δεύτερος διεθνής οργανισμός που το 1980 ασχολήθηκε με την προστασία προσωπικών δεδομένων, εκδίδοντας «Κατευθυντήριες Αρχές που διέπουν την προστασία της ιδιωτικότητας και τις διασυνοριακές ροές προσωπικών δεδομένων». Οι Αρχές αυτές περιλαμβάνουν την αρχή της περιορισμένης συγκέντρωσης και συλλογής δεδομένων, την αρχή της ποιότητας των δεδομένων, την αρχή του προσδιορισμένου σκοπού, της περιορισμένης χρήσης των προσωπικών δεδομένων, την αρχή μέτρων ασφαλείας και διαφάνειας των προσωπικών δεδομένων, την αρχή της συμμετοχής του ατόμου και την αρχή της ευθύνης. Είναι ένα πλαίσιο γενικών αρχών χωρίς δεσμευτικό χαρακτήρα που συγκέντρωσε για μεγάλο διάστημα τη συναίνεση πολλών χωρών και κυρίως εκείνων που στερούνταν ειδικής νομοθεσίας για την προστασία προσωπικών δεδομένων. [14]

2.2 Διεθνή και Ευρωπαϊκά νομικά εργαλεία προστασίας προσωπικών δεδομένων από την ηλεκτρονική τους διαχείριση.

Σύμβαση 108 του Συμβουλίου της Ευρώπης [16,19]

Η σύμβαση 108 του Συμβουλίου της Ευρώπης περί «Προστασίας των ατόμων από την αυτοματοποιημένη επεξεργασία προσωπικών δεδομένων» της 28 Ιανουαρίου 1981, ήταν η πρώτη σε ευρωπαϊκό επίπεδο ρύθμιση σχετική με την επεξεργασία προσωπικών δεδομένων.

Τα συμβαλλόμενα κράτη–μέλη δηλώνουν καταρχήν ότι επιθυμούν να ενισχύσουν τα μέτρα προστασίας των δικαιωμάτων και των θεμελιωδών ελευθεριών των πολιτών τους, ειδικότερα δε του δικαιώματος της προστασίας της προσωπικής ζωής τους ενόψει της αυξανόμενης διασυνοριακής ροής προσωπικών δεδομένων, τα οποία μάλιστα υφίστανται αυτοματοποιημένη επεξεργασία. Ταυτόχρονα επιβεβαιώνουν την προσήλωσή τους στην αρχή της ελεύθερης διακίνησης της πληροφορίας μεταξύ τους και διαπιστώνουν ότι υπάρχει ανάγκη υιοθέτησης κανόνων σεβασμού των αρχών αυτών.

Στη συνέχεια, η σύμβαση καθορίζει το σκοπό και το στόχο της. Ορίζει τις βασικές έννοιες (προσωπικά δεδομένα, αυτοματοποιημένο αρχείο, αυτοματοποιημένη επεξεργασία, ελεγκτής αρχείου) και καθορίζει το πεδίο εφαρμογής της, όπως αυτοματοποιημένα αρχεία προσωπικών δεδομένων του δημόσιου και του ιδιωτικού τομέα, επιτρέποντας έτσι στους συμβαλλόμενους να επεκτείνουν ή να περιορίσουν υπό όρους το πεδίο εφαρμογής και σε άλλες κατηγορίες αρχείων ή δεδομένων.

Δυο σημαντικές παρατηρήσεις μπορούν να γίνουν όσον αφορά τη συγκεκριμένη σύμβαση:

1. Η σύμβαση μιλά για δεδομένα και όχι για πληροφορίες. Έτσι, ο στόχος της προστασίας πληροφοριών συχνά αντικαθίσταται με το στόχο της προστασίας

δεδομένων.
2. Η σύμβαση αναφέρεται σε προστασία δεδομένων προσωπικού χαρακτήρα. Η ίδια η σύμβαση ορίζει τα δεδομένα αυτά ως κάθε πληροφορία που αναφέρεται σε φυσικό πρόσωπο, του οποίου η ταυτότητα είναι γνωστή ή μπορεί να εξακριβωθεί. Είναι φανερό ότι ανώνυμες, συγκεντρωτικές ή στατιστικές πληροφορίες δεν εντάσσονται στο πεδίο εφαρμογής της σύμβασης.

Επίσης η σύμβαση αυτή:
- αναφέρεται στην απαιτούμενη ποιότητα των δεδομένων,
- απαγορεύει την αυτοματοποιημένη επεξεργασία ειδικών κατηγοριών δεδομένων (φυλετικής προέλευσης, πολιτικών ή θρησκευτικών ή άλλων πεποιθήσεων, υγείας, σεξουαλικής ζωής, ποινικού μητρώου),
- απαιτεί τη λήψη κατάλληλων μέτρων ασφάλειας των αυτοματοποιημένων αρχείων,
- κατοχυρώνει το δικαίωμα του υποκειμένου των δεδομένων να γνωρίζει την ύπαρξη του αρχείου που περιέχει δικά του δεδομένα, να πληροφορείται τα καταχωρισμένα δεδομένα που την/τον αφορούν, να αιτείται τη διόρθωση ή διαγραφή λανθασμένων δεδομένων που την/τον αφορούν και να αποζημιώνεται αν αυτό δεν συμβεί,
- ορίζει συγκεκριμένες περιπτώσεις εξαίρεσης εφαρμογής των παραπάνω,
- θεμελιώνει την υποχρέωση των συμβαλλόμενων μερών να νομοθετήσουν ποινές για την παραβίαση των οριζομένων σ' αυτήν και τέλος
- επιτρέπει τη θεσμοθέτηση ισχυρότερων σχετικών ρυθμίσεων από τα συμβαλλόμενα μέρη.

Οδηγία της Ευρωπαϊκής Ένωσης 95/46/ΕΚ [16,20]

Η πιο γενική και πρόσφατη θεσμική ρύθμιση για την προστασία των πληροφοριών σε επίπεδο Ευρωπαϊκής Ένωσης είναι η οδηγία 95/46 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της Ευρωπαϊκής Ένωσης «για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων αυτών».
Η οδηγία διαπιστώνει καταρχήν ότι:
- Τα συστήματα επεξεργασίας δεδομένων υπηρετούν τον άνθρωπο και πρέπει να σέβονται τις θεμελιώδεις ελευθερίες και τα δικαιώματά του, ιδιαίτερα δε την ιδιωτική ζωή, συμβάλλοντας ταυτόχρονα στην ευημερία του.
- Απαιτείται μεν η εξασφάλιση της ελεύθερης διακίνησης των δεδομένων προσωπικού χαρακτήρα, με ταυτόχρονη όμως εξασφάλιση της προστασίας των θεμελιωδών δικαιωμάτων του ατόμου.
- Η επεξεργασία και η ανταλλαγή δεδομένων προσωπικού χαρακτήρα, ακόμη και η διασυνοριακή, προβλέπεται να αυξηθεί και πρέπει να ενθαρρυνθεί.
- Υπάρχει ανάγκη ευρωπαϊκής νομοθετικής ρύθμισης σχετικής με την προστασία δεδομένων προσωπικού χαρακτήρα, προκειμένου να διασφαλιστεί με ίσο βαθμό προστασίας η επεξεργασία τέτοιων πληροφοριών, που διακινούνται μεταξύ κρατών–μελών,
- Η οποιαδήποτε επεξεργασία δεδομένων προσωπικού χαρακτήρα πρέπει να εκτελείται κατά τρόπο θεμιτό και σύννομο και στη συνέχεια να προχωρεί στον ορισμό προϋποθέσεων σχετικά με τη θεμιτή επεξεργασία δεδομένων προσωπικού

χαρακτήρα, οι κυριότερες δε από τις οποίες είναι οι εξής:

- Τα δεδομένα πρέπει να συλλέγονται για καθορισμένους, σαφείς και νόμιμους σκοπούς και η μεταγενέστερη επεξεργασία τους να συμβιβάζεται με τους σκοπούς αυτούς. Πρέπει δε να είναι κατάλληλα, συναφή προς το θέμα, ακριβή, να διατηρούνται με μορφή που επιτρέπει τον προσδιορισμό του ατόμου στο οποίο αναφέρονται, όμως μόνο για όσο χρονικό διάστημα είναι απαραίτητο.
- Η επεξεργασία δεδομένων προσωπικού χαρακτήρα επιτρέπεται μόνο όταν υπάρχει η ρητή συγκατάθεση του υποκειμένου των δεδομένων.
- Απαγορεύεται η επεξεργασία δεδομένων προσωπικού χαρακτήρα σχετικών με τη φυλετική ή την εθνική καταγωγή, τα πολιτικά φρονήματα, τις θρησκευτικές ή φιλοσοφικές πεποιθήσεις, εκτός αν πληρούνται ρητά ορισμένες αναφερόμενες προϋποθέσεις.
- Προκειμένου να συλλεχθούν και καταχωριστούν ή μεταδοθούν δεδομένα προσωπικού χαρακτήρα, επιβάλλεται η πλήρης ενημέρωση του υποκειμένου σχετικά με την ταυτότητα του υπεύθυνου της επεξεργασίας, τους σκοπούς της επεξεργασίας, τους αποδέκτες των δεδομένων, την υποχρέωση ή όχι παροχής των δεδομένων και τις πιθανές συνέπειες σε περίπτωση άρνησης συμμόρφωσης στην ύπαρξη δικαιώματος πρόσβασης και πιθανής διόρθωσης των δεδομένων.
- Κατοχυρώνεται το δικαίωμα του υποκειμένου των δεδομένων να πληροφορείται αν υπάρχει επεξεργασία δεδομένων που το αφορούν, να λαμβάνει γνώση για τα δεδομένα που το αφορούν, να διορθώνει, διαγράφει ή δεσμεύει δεδομένα ελλιπή ή ανακριβή και να αντιτάσσεται σε αποφάσεις που βασίζονται αποκλειστικά σε αυτοματοποιημένη επεξεργασία δεδομένων που αξιολογεί ορισμένες πτυχές της προσωπικότητάς του.
- Καθορίζονται οι υποχρεώσεις του υπεύθυνου για την επεξεργασία των δεδομένων προσωπικού χαρακτήρα, προκειμένου να διασφαλιστεί το απόρρητο και η ασφάλεια της επεξεργασίας. Οι προϋποθέσεις αυτές περιλαμβάνουν την υποχρέωση του υπεύθυνου της επεξεργασίας να λάβει κατάλληλα τεχνικά και οργανωτικά μέτρα και να λαμβάνει άδεια επεξεργασίας από την αρμόδια εθνική αρχή ελέγχου.

2.3 Ελληνική πραγματικότητα για το θεσμικό πλαίσιο ασφαλείας.

Ο Νόμος 2472/97 για την προστασία του ατόμου από την επεξεργασία δεδομένων προσωπικού χαρακτήρα [3, 14]

Παρ' όλο που η Ελλάδα επικύρωσε τη σύμβαση 108 του Συμβουλίου της Ευρώπης με το νόμο 2068/1992, δεν προχώρησε ωστόσο άμεσα στη θέσπιση αντίστοιχης εθνικής νομοθεσίας. Αυτό έγινε στις 10 Απριλίου 1997 όταν δημοσιεύτηκε στην Εφημερίδα της Κυβερνήσεως ο νόμος 2472 περί «Προστασίας του ατόμου από την επεξεργασία δεδομένων προσωπικού χαρακτήρα». Αντικείμενο του νόμου αυτού είναι «η θέσπιση των προϋποθέσεων για την επεξεργασία δεδομένων προσωπικού

χαρακτήρα προς προστασία των δικαιωμάτων και των θεμελιωδών ελευθεριών των φυσικών προσώπων και ιδίως της ιδιωτικής ζωής».
Ο νόμος 2472 διαχωρίζει μια ειδική κατηγορία δεδομένων προσωπικού χαρακτήρα, τα οποία ονομάζει ευαίσθητα δεδομένα. Αυτά είναι «τα δεδομένα που αφορούν τη φυλετική ή εθνική προέλευση, τα πολιτικά φρονήματα, τις θρησκευτικές ή φιλοσοφικές πεποιθήσεις, τη συμμετοχή σε ένωση, σωματείο και συνδικαλιστική οργάνωση, την υγεία, την κοινωνική πρόνοια και την ερωτική ζωή, καθώς και τα σχετικά με ποινικές διώξεις ή καταδίκες». Η επεξεργασία δεδομένων προσωπικού χαρακτήρα επιτρέπεται υπό την προϋπόθεση ότι το υποκείμενο των δεδομένων έχει δώσει τη συγκατάθεσή του προς τούτο. Αντίθετα, η συλλογή και επεξεργασία των ευαίσθητων δεδομένων γενικά απαγορεύεται, εκτός εάν συντρέχουν πολύ συγκεκριμένες προϋποθέσεις και δοθεί η άδεια προς τούτο από την Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα.
Ο νόμος, επίσης, καθορίζει τους όρους υπό τους οποίους είναι νόμιμη η διασύνδεση αρχείων, επιτρέπει δε τη διασυνοριακή ροή δεδομένων προσωπικού χαρακτήρα προς χώρες της Ευρωπαϊκής Ένωσης ελεύθερα και προς τρίτες χώρες υπό προϋποθέσεις. Θεσπίζει ακόμη ρυθμίσεις που εξασφαλίζουν το απόρρητο και την ασφάλεια της επεξεργασίας, καθορίζει τα δικαιώματα του υποκειμένου των δεδομένων, θεσπίζει την ανεξάρτητη Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα και περιγράφει τις αρμοδιότητές της και, τέλος, ορίζει κυρώσεις σε περιπτώσεις μη συμμόρφωσης με τις διατάξεις του.
Ο νόμος 2472 αποτελεί ένα σχετικά πλήρες νομοθετικό πλαίσιο για την προστασία πληροφοριών στη χώρα μας. Πέρα από τις νομικές του ρυθμίσεις, περιέχει και σειρά διατάξεων οι οποίες είτε άμεσα είτε έμμεσα παραπέμπουν σε τεχνικές ρυθμίσεις που ο υπεύθυνος της επεξεργασίας πρέπει να τις έχει υπόψη του και να τις εφαρμόζει.

Κυρώσεις:

Οι κανόνες και οι διαδικασίες ελέγχου ακολουθούνται από την πρόβλεψη μηχανισμών καταστολής των παραβάσεων των διατάξεων που αφορούν την προστασία του προσώπου από την επεξεργασία δεδομένων προσωπικού χαρακτήρα. Πρόκειται για διοικητικές και ποινικές κυρώσεις καθώς και για την αστική ευθύνη του υπεύθυνου της επεξεργασίας.

Οι κυρώσεις επιβάλλονται από την Αρχή Προστασίας Προσωπικών Δεδομένων και είναι κυρίως διοικητικές. Οι κυρώσεις αυτές κλιμακώνονται από την σύσταση, την προειδοποίηση για άρση της παράβασης και την επιβολή προστίμων, ως και την οριστική ανάκληση της άδειας του αρχείου ή και την καταστροφή του αρχείου.

Ποινικές κυρώσεις προβλέπονται για δύο κατηγορίες συμπεριφορών που κρίνονται αξιόποινες: (α) για πράξεις και παραλείψεις για τις οποίες η Αρχή δεν έχει ακόμη αποφανθεί, όπως τη μη γνωστοποίηση αρχείου, τη λειτουργία αρχείου με ευαίσθητα δεδομένα χωρίς άδεια, τη διασύνδεση αρχείων χωρίς άδεια κλπ και (β) για την μη συμμόρφωση σε αποφάσεις της Αρχής. [8]

Προστασία του απορρήτου της επικοινωνίας:

Σκοπός του απορρήτου της επικοινωνίας, δεν είναι η προστασία του μηνύματος καθεαυτού, αλλά το απόρρητο του μηνύματος. Το ίδιο το μήνυμα προστατεύεται από την ελευθερία έκφρασης και διάδοσης της γνώμης. Η προστασία του απορρήτου αναφέρεται σε κάθε μορφή ιδιωτικής επικοινωνίας όχι μόνο το παραδοσιακό μέσο

των επιστολών και ανεξάρτητα εάν προνοείται για επικοινωνία με προσωπικό ή επαγγελματικό χαρακτήρα. Δεν υπάρχει διαβάθμιση στην προστασία. Το κρίσιμο στοιχείο είναι να πραγματώνεται η επικοινωνία υπό συνθήκες εμπιστευτικότητας. [9]

Η πρόβλεψη ανεξάρτητης αρχής σκοπεύει στην ενίσχυση της προστασίας του απορρήτου των επικοινωνιών (Άρθρο 19 παρ. 2 Σ 1975/86/01). Πρόκειται για τη μία από τις πέντε ανεξάρτητες αρχές που το Σύνταγμα, όχι απλώς κατοχυρώνει αλλά και περιβάλλει με ιδιαίτερες εγγυήσεις, όσον αφορά την επιλογή, τη θητεία, την προσωπική και λειτουργική ανεξαρτησία των μελών της (Άρθρο 101 Α' Σ 1975/86/01). Η αρχή αυτή συστήθηκε με τον Ν. 3115/03 (ΦΕΚ Α' 47/27.02.2003). Η Αρχή Διασφάλισης του Απορρήτου των Επικοινωνιών (ΑΔΑΕ) διενεργεί τακτικούς και έκτακτους ελέγχους σε εγκαταστάσεις και εξοπλισμό της Εθνικής Υπηρεσίας Πληροφοριών (ΕΥΠ) ή άλλων δημόσιων υπηρεσιών ή/και ζωτικών επιχειρήσεων που ασχολούνται με ταχυδρομικές, τηλεπικοινωνιακές ή άλλες υπηρεσίες σχετικές με την ανταπόκριση και την επικοινωνία, εξετάζει καταγγελίες σχετικά με την προστασία των δικαιωμάτων των αιτούντων και ελέγχει την τήρηση των όρων και της διαδικασίας άρσης του απορρήτου. [10]

2.4 Προβλήματα προστασίας προσωπικών δεδομένων στο Διαδίκτυο.

Το Διαδίκτυο αποτελεί ένα μέσο διάδοσης πληροφοριών, το οποίο θέτει νέα ζητήματα ως προς την επεξεργασία και την προστασία των προσωπικών δεδομένων. Έχει σημασία να εξετάσουμε τα χαρακτηριστικά του διαδικτύου, τα οποία είναι ιδιαίτερα κρίσιμα σε σχέση με τη λειτουργία του αλλά και την χρήση των προσωπικών δεδομένων: Τα τεχνικά εργαλεία είναι νέα (π.χ. browsing software) και εξελίσσονται με ραγδαίους ρυθμούς. Οι υποδομές αυτές παρέχουν νέους τρόπους και νέες δυνατότητες διανομής προσωπικών πληροφοριών (mailing lists, discussion groups, πρόσβαση σε βάσεις δεδομένων). Μέσω του Διαδικτύου τα δεδομένα διαβιβάζονται ταχύτατα σε κάθε σύστημα, τηρούνται σε διαφορετικούς χώρους, αλλά παράλληλα διασυνδέονται μεταξύ τους και παρέχονται για διαφορετικούς σκοπούς. Το Διαδίκτυο χρησιμοποιείται από πολλές και διαφορετικές κατηγορίες συμμετεχόντων, οι οποίοι είναι πολυάριθμοι.

Το Διαδίκτυο αποτελεί ένα διεθνή χώρο ο οποίος δεν μπορεί να υποστεί κρατικό ή ιδιωτικό έλεγχο. Όμως μπορεί να χρησιμοποιηθεί ως χώρος έκφρασης, ως χώρος εργασίας κλπ. Πολλά από τα στοιχεία του είναι ελκυστικά στο χρήστη, όπως η ταχύτητα, η ευκολία της πρόσβασης, η εργονομία της θέσης εργασίας, η (σχετικά) χαμηλή τιμή πρόσβασης, ο διάλογος και - τελευταίο αλλά όχι ύστερο - οι ποικίλες προσφορές. [11] [12].

Η χρήση προσωπικών δεδομένων είναι προϋπόθεση και δεν πρέπει να θεωρείται δευτερεύουσα συνέπεια της χρήσης των υπηρεσιών του Διαδικτύου. Αντίθετα συγκαταλέγεται στις προϋποθέσεις για την παροχή μιας υπηρεσίας, δημιουργείται δε από την χρήση της υπηρεσίας και αποτελεί τον σκοπό ή/και το αντικείμενο της υπηρεσίας.

Το Διαδίκτυο , όπως και όλα τα δίκτυα τηλεπικοινωνιών, παράγουν τεράστιες ποσότητες δεδομένων με σκοπό την εξασφάλιση των σωστών διασυνδέσεων. Για παράδειγμα απαιτείται ένας μεγάλος αριθμός δεδομένων για να είναι εφικτή η αποκατάσταση της σύνδεσης, η διαβίβαση των δεδομένων και η χρέωση. Τα

δεδομένα (δια)σύνδεσης αφορούν τις τεχνικές που χρησιμοποιούνται για την αποκατάσταση της σύνδεσης. Πρόκειται αφενός για τις διευθύνσεις των μηχανών του δικτύου (IP) και αφετέρου για τη σελίδα που θέλει να επισκεφθεί ο πλοηγός.

Οι μορφές προσβολής ιδιωτικότητας στο Διαδίκτυο θα μπορούσαν να κατανεμηθούν, για συστηματικούς και λειτουργικούς λόγους, σε δυο μεγάλες κατηγορίες: α) την εμφανή συλλογή δεδομένων και β) τη συλλογή δεδομένων εν αγνοία του υποκειμένου – χρήστη των υπηρεσιών του Διαδικτύου.

2.5 Ρύθμιση της προστασίας προσωπικών δεδομένων στον τηλεπικοινωνιακό τομέα.

Η προστασία των προσωπικών δεδομένων και της ιδιωτικής ζωής των χρηστών των δημόσιων τηλεπικοινωνιακών δικτύων δημιουργεί νέες απαιτήσεις, λόγω της εισαγωγής προηγμένων ψηφιακών τεχνολογιών. Η Ευρωπαϊκή Επιτροπή, παράλληλα με το σχέδιο γενικής οδηγίας για την προστασία προσωπικών δεδομένων, θεώρησε απαραίτητη τη θέσπιση ειδικών νομοθετικών, κανονιστικών και τεχνικών διατάξεων για την προστασία προσωπικών δεδομένων στα ψηφιακά δίκτυα, με σκοπό την αντιμετώπιση των προβλημάτων που εμφανίζονται. Η κοινοτική πρωτοβουλία ήταν απαραίτητη, ώστε να εναρμονισθούν οι εθνικές νομοθεσίες και να αρθούν τυχόν εμπόδια που θα δυσχέραιναν ή θα απέκλειαν «την προαγωγή και ανάπτυξη τηλεπικοινωνιακών υπηρεσιών και δικτύων μεταξύ των κρατών μελών» και κατ' αποτέλεσμα θα παρεμπόδιζαν τη διάδοση της κοινωνίας των πληροφοριών. Τα όρια της κοινοτικής πρωτοβουλίας τα διέγραψε ωστόσο εν πολλοίς η αρχή της επικουρικότητας. Η επίκλησή της, μετά τη Σύνοδο Κορυφής στο Μάαστριχτ το 1992, είχε ως συνέπεια να διαγραφούν πολλές από τις αρχικά προτεινόμενες ρυθμίσεις.

Ύστερα από τις πρωτοβουλίες αυτές, προέκυψε η Οδηγία 97/66 για την προστασία των προσωπικών δεδομένων στον τομέα των τηλεπικοινωνιών. Οι προτάσεις αφορούσαν αρχικά την επεξεργασία δεδομένων στα ψηφιακά δίκτυα ενοποιημένων υπηρεσιών (ISDN) και οι ρυθμίσεις ήταν προσανατολισμένες στις τεχνικές δυνατότητες και ευχέρειες που τα δίκτυα αυτά προσφέρουν. Στο τελικό στάδιο κατάρτισης της Οδηγίας αποφασίστηκε η διεύρυνση του ρυθμιστικού πεδίου στον τηλεπικοινωνιακό εν γένει τομέα.

Η Οδηγία κάλυπτε τα ζητήματα της ασφάλειας των δεδομένων, του απόρρητου των επικοινωνιών, των δεδομένων κίνησης και χρέωσης της αναλυτικής χρέωσης, της αναγραφής της ταυτότητας της καλούσας/συνδεδεμένης γραμμής, της αυτόματης προώθησης κλήσεων, των τηλεφωνικών καταλόγων των συνδρομητών καθώς και το πρόβλημα των αυτόματων συστημάτων κλήσης για εμπορικούς σκοπούς. Οι σχετικές ρυθμίσεις αφορούσαν τα δημόσια τηλεπικοινωνιακά δίκτυα, δηλαδή τα συστήματα μετάδοσης και ο εξοπλισμός μεταγωγής και τα λοιπά μέσα που επιτρέπουν την μεταφορά σημάτων μεταξύ συγκεκριμένων τερματικών σημείων με τη χρήση καλωδίου, ραδιοκυμάτων, οπτικών ή άλλων ηλεκτρομαγνητικών μέσων, τα οποία χρησιμοποιούνται, εν μέρει ή εν όλω, για την παροχή διαθέσιμων στο κοινό τηλεπικοινωνιακών υπηρεσιών. [13]

3. Απώλεια Δεδομένων

3.1 Εισαγωγή

Η απώλεια δεδομένων ορίζεται ως η τυχαία ή ακούσια διανομή προσωπικών ή ευαίσθητων δεδομένων σε μη εξουσιοδοτημένο φορέα. Τα ευαίσθητα δεδομένα εταιρειών ή οργανισμών περιλαμβάνουν πνευματική ιδιοκτησία (Intellectual Property IP), χρηματοπιστωτικές πληροφορίες, πληροφορίες ασθενών, προσωπικά δεδομένα πιστωτικών καρτών, καθώς και άλλες πληροφορίες ανάλογα με την επιχείρηση και τη βιομηχανία. Η απώλεια δεδομένων αποτελεί ένα σημαντικό ζήτημα για τις εταιρείες καθώς ο αριθμός των περιστατικών και το κόστος αυτών που τα βιώνουν συνεχίζεται να αυξάνεται. Η απώλεια δεδομένων ενισχύεται από το γεγονός ότι τα μεταφερόμενα δεδομένα (είτε εισερχόμενα, είτε εξερχόμενα) περιλαμβανομένης και της ηλεκτρονικής αλληλογραφίας, άμεσων μηνυμάτων, φορμών ιστοσελίδων, και μεταφοράς αρχείων μεταξύ ατόμων, δεν ελέγχονται ούτε παρακολουθούνται κατά τη διαδρομή προς τον προορισμό τους. Επιπροσθέτως, σε πολλές περιπτώσεις ευαίσθητα δεδομένα διαμοιράζονται μεταξύ διαφόρων ενδιαφερομένων, όπως υπάλληλοι που εργάζονται σε χώρο εκτός του οργανισμού (μέσω φορητού υπολογιστή), συνέταιροι και πελάτες. Αυξάνοντας παράλληλα το ρίσκο οι εμπιστευτικές πληροφορίες να υποπέσουν σε μη εξουσιοδοτημένα άτομα, είτε οφείλονται σε κακόβουλες προθέσεις είτε σε ανθρώπινο λάθος, εκ των έσω ή και όχι. Η αποκάλυψη ευαίσθητων πληροφοριών μπορεί να έχει αρνητικές συνέπειες στην εκάστοτε επιχείρηση.

Η πιθανή ζημία και οι δυσμενείς συνέπειες περιστατικών απώλειας δεδομένων μπορούν να χωριστούν σε δυο κατηγορίες: άμεσες και έμμεσες απώλειες. Οι άμεσες απώλειες αναφέρονται σε εμφανείς ζημιές που είναι εύκολο να μετρηθούν ή να εκτιμηθούν ποσοτικά. Πιο συγκεκριμένα περιλαμβάνουν παραβάσεις κανονισμών (όπως αυτούς για την προστασία των πελατών) που οδηγούν σε πρόστιμα, διακανονισμούς ή αποζημιώσεις πελατών, σε διαφορές που αφορούν αγωγές: Απώλεια μελλοντικών πωλήσεων, κόστη έρευνας, διόρθωσης και αποκατάστασης.

Οι έμμεσες απώλειες, από την άλλη, είναι δυσκολότερο να προσδιοριστούν και έχουν πολύ μεγαλύτερο αντίκτυπο όσον αφορά το κόστος, τον τόπο και το χρόνο [21]. Αυτές περιλαμβάνουν μειωμένη τιμή μετοχής λόγω αρνητικής δημοσιότητας, ζημιά στη φήμη και το κύρος της επιχείρησης, εγκατάλειψη πελατών και αποκάλυψη πνευματικής ιδιοκτησίας (επιχειρηματικά σχέδια, κώδικες, οικονομικές αναφορές κ.α.) σε ανταγωνιστές.

Η απώλεια δεδομένων μπορεί να συμβεί σε οποιαδήποτε μορφή και μέρος. Στον Πίνακα 3.1 παρουσιάζονται μερικά, επιλεγμένα από το www.datalossdb.org, περιστατικά διαρροής πληροφοριών υψηλού επιπέδου. Το δείγμα εστιάζεται σε πρόσφατα περιστατικά και στη δυσκολία εύρεσης ικανοποιητικής λύσης για την πρόληψη της απώλειας δεδομένων. Είναι εμφανές επίσης ότι οι επιχειρήσεις πρέπει να διευρύνουν τις προσπάθειες ασφάλειάς τους, πέραν από την εξασφάλιση της περιμέτρου του δικτύου και έναντι των απειλών από κλασικές απειλές (όπως εισβολές, ιοί, δούρειοι ίπποι, σκουλήκια, κτλ). Επιπρόσθετα οι οργανισμοί είναι υποχρεωμένοι να συμμορφώνονται με τους ομοσπονδιακούς και πολιτειακούς κανονισμούς, οι

οποίοι έχουν ως σκοπό την προστασία των οικονομικών και άλλων δεδομένων [22].

Πίνακας 3.1: Περιστατικά απώλειας δεδομένων.

Ημερομηνία	***Οργανισμός***	***Περιγραφή***
2003	Βρετανική Υπηρεσία Πληροφοριών	Μια αναφορά της Βρετανικής Υπηρεσίας Πληροφοριών υπό μορφή εγγράφου Word περιείχε τα ονόματα συγγραφέων μιας μελέτης, η οποία στη σύνοψη των μεταδιδόμενων είχε αναφορές από τις ΗΠΑ και παρουσιάστηκε σε ομιλία στα Ηνωμένα Έθνη. Τα μεταδεδομένα έδειξαν ότι η αναφορά στην πραγματικότητα γράφτηκε από Αμερικανούς ερευνητές
Ιούνιος 2004	AOL	Ένας υπάλληλος της AOL έκλεψε τους κωδικούς εξακρίβωσης ενός άλλου εργαζομένου της εταιρείας, ώστε να αποκτήσει πρόσβαση σε δεδομένα συνδρομητών. Στη συνέχεια υπεξαίρεσε 92 εκατομμύρια διευθύνσεις email, που ανήκαν σε 30 εκατομμύρια συνδρομητές και τις πούλησε σε spammers.
Αύγουστος 2007	Εργαστήριο Πυρηνικής φυσικής στο Los Alamos	Ένας υπάλληλος του Αμερικανικού εργαστηριού πυρηνικής φυσικής μετέδωσε εμπιστευτικές πληροφορίες μέσω email. Το περιστατικό χαρακτηρίστηκε ως σοβαρή απειλή για την πυρηνική ασφάλεια της χώρας.
Ιούλιος 2008	Google	Δεδομένα εκλάπησαν από τα κεντρικά γραφεία της θυγατρικής της εταιρείας, Colt Express. Οι κλέφτες εισέβαλλαν και έκλεψαν υπολογιστές της εταιρείας που περιείχαν αποκρυπτογραφημένα δεδομένα ονομάτων, διευθύνσεων και αριθμών Κοινωνικής Ασφάλισης υπαλλήλων της Google. Ως αποτέλεσμα, η Google τερμάτισε τη συνεργασία της με την Colt Express
Οκτώβριος 2008	UPS	Από φορητό υπολογιστή υπαλλήλου της UPS εκλάπησαν πληροφορίες μισθοδοσίας 9000 υπαλλήλων της Μ. Βρετανίας. Ως απάντηση η UPS ανακοίνωσε ότι θα κρυπτογραφεί όλα τα αποθηκευμένα δεδομένα σε όλες τις φορητές συσκευές της επιχείρησης.

Ιανουάριος 2009	Συστήματα πληρωμών Heartland	Κακόβουλο λογισμικό παραβίασε 10 εκατομμύρια συναλλαγές πιστωτικών και χρεωστικών καρτών.
Οκτώβριος 2009	Εθνικό Αρχείο των ΗΠΑ	Η διεύθυνση του Εθνικού Αρχείου των ΗΠΑ απέρριψε με ακατάλληλο τρόπο σκληρούς δίσκους που περιείχαν 76 εκατομμύρια ονόματα, διευθύνσεις και αριθμούς Κοινωνικής Ασφάλισης βετεράνων στρατιωτικών
Σεπτέμβριος 2011	Science Applications	Κασέτες αντιγράφων ασφαλείας εκλάπησαν από αυτοκίνητο και περιείχαν 5.117.799 ονόματα ασθενών, νούμερα τηλεφώνων, αριθμούς Κοινωνικής Ασφάλισης, και ιατρικές πληροφορίες
Μάρτιος 2012	Shanghai Roadway D&B Marketing Services Co. Ltd	Παράνομη αγοραπωλησία πληροφοριών 150.000.000 πελατών
Μάρτιος 2013	LivingSocial Inc (ΗΠΑ)	Hacker κατάφερε να υπεξαιρέσει ονόματα, emails και ημερομηνίες γεννήσεως 50.000.000 πελατών

3.2 Πολυεπίπεδο μοντέλο ασφάλειας.

Η απώλεια δεδομένων είναι ένα σύνθετο πολυπρόσωπο πρόβλημα που απαιτεί συστηματική ανάλυση για να μετριαστεί. Ο Clive Blackwell [23] σχεδίασε ένα μοντέλο ασφάλειας τριών επιπέδων καλούμενο Προβολέας (Searchlight) για τη διερεύνηση και εκτίμηση συστημάτων ασφαλείας, τα οποία εφαρμόζονται έναντι απώλειας δεδομένων. Η χρησιμοποίηση επιπέδων είναι μια κοινή μέθοδος για την αποσύνθεση και ανάλυση των συστημάτων. Το μοντέλο επηρεάστηκε από το σύστημα κατηγοριοποίησης επιθέσεων του Neumann που περιείχε οκτώ επίπεδα, τα οποία κατά φθίνουσα σειρά είναι: το εξωτερικό περιβάλλον, ο χρήστης, η εφαρμογή, το ενδιάμεσο λογισμικό, το δίκτυο, το λειτουργικό σύστημα, το υλικό και το εσωτερικό περιβάλλον.

Το μοντέλο Searchlight έχει τρία επίπεδα, τα οποία περιλαμβάνουν το κοινωνικό επίπεδο (άνθρωποι και οργανισμοί), το φυσικό επίπεδο και το μεσαίο λογικό επίπεδο, που περιλαμβάνει τους υπολογιστές και τα δίκτυα. Αυτό επιτρέπει μια ολιστική αναπαράσταση και ανάλυση σύνθετων συστημάτων, όπως οργανισμούς, συμπεριλαμβανομένου των ανθρώπινων και των φυσικών παραγόντων, καθώς και των τεχνικών συστημάτων.

Τα δεδομένα έχουν διαφορετική έκταση σε κάθε επίπεδο. Για παράδειγμα, ο αριθμός των ανθρώπων που γνωρίζουν κάποιες πληροφορίες είναι η έκτασή του στο κοινωνικό επίπεδο, το εύρος των λογικών δεδομένων είναι η διαθεσιμότητά του σε υπολογιστές και εφαρμογές που μπορεί να εκτείνεται παγκοσμίως στο Διαδίκτυο, ενώ τα φυσικά αποθηκευμένα δεδομένα μπορεί να περιορίζονται σε έγγραφα αρχείων ενός γραφείου.

Το κοινωνικό ή οργανωτικό επίπεδο περιέχει μια περιληπτική παρουσίαση των οργανισμών ανάλογα των χαρακτηριστικών τους, συμπεριλαμβανομένου των στόχων,

των πολιτικών και των διαδικασιών τους. Επίσης περιλαμβάνει τους ανθρώπους και τα χαρακτηριστικά τους, όπως τις γνώσεις τους, τα πιστεύω και τους στόχους τους. Οι εμπορικές και προσωπικές πληροφορίες ενυπάρχουν στο κοινωνικό επίπεδο, επειδή δεν έχουν νόημα για τα κατώτερα επίπεδα. Τα δεδομένα μπορούν να αποθηκευτούν, να επεξεργαστούν και να μεταφερθούν σε υπολογιστές και χαρτιά στα κατώτερα επίπεδα.

Το λογικό επίπεδο είναι το μεσαίο στρώμα που περιέχει άυλους υπολογιστικούς φορείς συμπεριλαμβανομένου των υπολογιστών, δικτύων, λογισμικού και δεδομένων. Τα λογικά δεδομένα είναι μια αναπαράσταση των πληροφοριών στο κοινωνικό επίπεδο, όπου είναι πιο δεκτικά σε επεξεργασία, διαμοιρασμό και αποθήκευση. Επίσης περιλαμβάνει δεδομένα που χρησιμοποιούνται για την παροχή λογικών υπηρεσιών στο κοινωνικό επίπεδο, συμπεριλαμβανομένου της αυθεντικοποίησης πληροφοριών, όπως κωδικών που συνδέουν ανθρώπους με τους λογαριασμούς τους. Το λογικό στρώμα είναι εσφαλμένα, το επίκεντρο της προσοχής όσον αφορά την ασφάλεια, επειδή όλα τα επίπεδα χρειάζονται προστασία για πιο αποτελεσματική ασφάλεια.

Το φυσικό επίπεδο είναι το κατώτερο στρώμα που περιέχει υλικά αντικείμενα, όπως κτίρια, εξοπλισμό, έγγραφα αρχείων και φυσικές πτυχές των υπολογιστών και άλλων σχετικών συσκευών. Επιπροσθέτως, περιέχει ηλεκτρομαγνητική ακτινοβολία, όπως ραδιοκύματα, ηλεκτρική και μαγνητική ενέργεια που χρησιμοποιούνται για την μεταφορά και αποθήκευση δεδομένων. Εξάλλου όλες οι οντότητες των υψηλότερων επιπέδων, συμπεριλαμβανομένου των ανθρώπων και των πληροφοριών, έχουν φυσική υπόσταση καθώς και αναπαράσταση υψηλότερου επιπέδου που πρέπει να ληφθεί υπόψη κατά την ανάλυση της ασφάλειας δεδομένων. [23]

3.3 Προστασία έναντι απώλειας δεδομένων.

3.3.1 Ισχυροποίηση του συστήματος.

Στόχος της ισχυροποίησης είναι η διακοπή της επίδρασης του κοινωνικού επιπέδου στον οργανισμό, ώστε να μπορούν να θεωρηθούν προστατευτικά μέτρα σε πολλαπλά στάδια πριν, κατά τη διάρκεια, αλλά και μετά από ένα περιστατικό απώλειας, το οποίο ισοδυναμεί με μείωση της επιφάνειας υπό επίθεση, ισχυροποίηση του στόχου, και περιορισμού των επιπτώσεων.

Τα ευαίσθητα δεδομένα πρέπει να είναι δύσκολο να καταχραστούν ή να αποκαλυφθούν. Αυτό απαιτεί αποτελεσματική προστασία σε όλα τα επίπεδα. Πρέπει να υπάρχει μια ολοκληρωμένη επιφάνεια επίθεσης που να περιορίζει την κίνηση και την πρόσβαση σε δεδομένα σε όλα τα στρώματα, ώστε να εμποδίσει μη εξουσιοδοτημένη πρόσβαση, αλλά και να ελέγξει σε ικανοποιητικό βαθμό και την εξουσιοδοτημένη χρήση.

Ο πιθανός αντίκτυπος μπορεί να περιοριστεί στο στόχο, γενικά στον οργανισμό ή να έχει εξωτερικό αντίκτυπο σε τρίτους. Επί πλέον μέτρα προστασίας μπορούν να παρέχουν άμυνα εις βάθος, παρέχοντας παράλληλα ανεξάρτητες ζώνες επιπτώσεων. Αυτό περιλαμβάνει αρχικούς ελέγχους χρήσης των δεδομένων εσωτερικά, και μεταγενέστερους ελέγχους κατά τη μεταφορά τους σε άλλους υπολογιστές, ή κατά την αντιγραφή σε συσκευές αποθήκευσης όπως CD, συσκευές USB ή εκτύπωση σε χαρτί.

Εξάλλου, οι επιπτώσεις μπορούν να περιοριστούν στα κατώτερα φυσικά και λογικά στρώματα ή να έχουν σημαντική επίδραση στο κοινωνικό στρώμα. Επιπρόσθετα μέτρα προστασίας μπορεί να εμποδίσουν παρεμβάσεις σε πηγές κατώτερων στρωμάτων να βλάψουν ουσιαστικές δραστηριότητες του οργανισμού. Για παράδειγμα, απώλεια πιστωτικής κάρτας μπορεί να ξεπεραστεί εντοπίζοντας «περίεργες» συναλλαγές ή αποζημιώνοντας τα θύματα. Δυστυχώς, η ανάκαμψη από αποκάλυψη δεδομένων είναι πολύ δύσκολη, μιας και τα δεδομένα ακολουθούν μεταγενέστερα δικό τους κύκλο ζωής, ειδικά στο Διαδίκτυο.

3.3.2 Στόχευση του δράστη

Ως επί το πλείστον τα περιστατικά απώλειας δεδομένων ξεκινούν από το κοινωνικό επίπεδο, από ένα άτομο το οποίο έχει ένα στόχο. Χρησιμοποιώντας τους στόχους του δράστη μπορεί να προσδιοριστούν τα απαραίτητα μέτρα προστασίας κοινωνικού επιπέδου για τη μείωση των κινήτρων των επιτιθέμενων και την αύξηση της εγρήγορσης των έντιμων υπαλλήλων.
Οι επιθέσεις συνήθως παρακινούνται από την ανάγκη επίλυσης προσωπικών και επαγγελματικών προβλημάτων. Τα προσωπικά θέματα περιλαμβάνουν διαζύγια, κατάχρηση ουσιών, οικονομικά προβλήματα και συναισθηματικές διαταραχές. Τα επαγγελματικά θέματα περιλαμβάνουν επαγγελματική δυσαρέσκεια, διαμάχες εργασιακού χώρου και πειθαρχικές κυρώσεις. Κάθε οργανισμός μπορεί να μειώσει την απειλή αντιμετωπίζοντας τα προσωπικά και οικονομικά προβλήματα και ενθαρρύνοντας μεγαλύτερη αφοσίωση μέσω δημιουργίας καλύτερων συνθηκών εργασίας, δίκαιης αντιμετώπισης και παρακολούθησης των παραπόνων. Μπορούν να αποτρέψουν επιθέσεις με ισχυρά μέτρα άμυνας που κάνουν τη σχέση κόστους/οφέλους λιγότερο ευνοϊκή, αυξάνοντας το ρίσκο ή μειώνοντας τα οφέλη κλοπής δεδομένων.
Όσον αφορά τα περιστατικά τυχαίας απώλειας δεδομένων, συνήθως από υπαλλήλους, ο δράστης δεν έχει επίγνωση των πιθανών συνεπειών, μιας και τα δεδομένα είναι άυλα και εύκολα υποτιμώνται. Οι υποχρεώσεις των υπαλλήλων πρέπει να είναι ξεκάθαρες, μέσω της ρητής κατανομής καθηκόντων και ευθυνών, με ορθά δομημένες και κατανοητές πολιτικές, ενισχυμένες με πειθαρχικές ενέργειες για παραβιάσεις οι οποίες αυξάνουν τα κίνητρα τους ώστε να είναι προσεκτικοί λόγω των επακόλουθων συνεπειών.
Τυχαία αποκάλυψη συχνά προκαλείται από υπαλλήλους που θέλουν να ολοκληρώσουν έγκαιρα τη δουλειά τους ή να βοηθήσουν κάποιο συνάδελφο τους. Συντρέχοντες παράγοντες είναι η άγνοια, το χαμηλό ηθικό, και η έλλειψη αφοσίωσης ή αποδοχής των κανόνων και των αξιών του οργανισμού. Μπορεί να υπάρχει χαλαρή εταιρική πολιτική και έτσι τα δεδομένα να εκτιμώνται ανεπαρκώς με ασυνεπείς διαδικασίες, οδηγώντας σε έλλειψη προσοχής. Τα στρεβλά κίνητρα μπορούν να αποφευχθούν, όπως σε κατάρρευση του χρηματοπιστωτικού συστήματος, όπου η δομή επιβράβευσης ωθεί τους υπαλλήλους σε επικίνδυνες συμπεριφορές. Τυχαία αποκάλυψη δεδομένων από υπαλλήλους μπορεί να μειωθεί με ευαισθητοποίηση και εκπαίδευση, έτσι ώστε να τους ενημερώσει για τα πιθανά ρίσκα, με επιπλέον κίνητρο την επιβράβευση καλής συμπεριφοράς.
Τα θέματα του συστήματος μπορούν να αντιμετωπιστούν με ανασχεδιασμό για την

αποφυγή των λεγόμενων «ατυχημάτων που πρόκειται να συμβούν». Οι οργανισμοί πρέπει να κάνουν τα συστήματα τους ώστε να διορθώνονται ευκολότερα και να δίνουν ξεκάθαρες ενδείξεις όταν πρόκειται να αποκαλυφθούν ευαίσθητα δεδομένα. [23]

3.3.3 Βασικά μέτρα προστασίας

Τα βασικά μέτρα προστασίας που χρησιμοποιούνται από πολλές οργανώσεις, περιλαμβάνουν κοινούς μηχανισμούς, όπως τείχος προστασίας (firewall), συστήματα εντοπισμού εισβολών (IDSs) και λογισμικά antivirus που παρέχουν προστασία είτε έναντι εξωτερικών επιθέσεων (όπως ένα τείχος προστασίας το οποίο περιορίζει την πρόσβαση στο εσωτερικό δίκτυο, καθώς και ένα σύστημα εντοπισμού εισβολών το οποίο ανιχνεύει τις προσπάθειες εισβολής) και εσωτερικών επιθέσεων (όπως σάρωση από το antivirus για τον εντοπισμό Trojan horse που μπορεί να έχει εγκατασταθεί σε έναν υπολογιστή και να αποστέλλει εμπιστευτικές πληροφορίες). Οι διάφορες πολιτικές και η εκπαίδευση του προσωπικού και των συνεταίρων αφορούν στο να είναι ενημερωμένοι με τις τελευταίες εξελίξεις, παρέχουν δε επιπρόσθετα μέτρα προστασίας.
Δημιουργώντας και εφαρμόζοντας πολιτικές στους οργανισμούς για ευρύ χειρισμό δεδομένων, με βάση τους κανονισμούς των βιομηχανιών και τις ειδικές απαιτήσεις του εκάστοτε οργανισμού, σημαντικό ζήτημα είναι ο έλεγχος όλων των πτυχών χειρισμού των προσωπικών δεδομένων. Αυτές οι πολιτικές εφαρμόζουν αυστηρούς κανόνες χειρισμού δεδομένων, όπως απόρριψη ή αρχειοθέτηση περιττών επιπλέον προσωπικών δεδομένων και δημιουργία μηχανισμών ελέγχου πρόσβασης, ώστε μόνο το εξουσιοδοτημένο προσωπικό να έχει πρόσβαση σε αυτά τα δεδομένα. Εξάλλου, η δημιουργία πολιτικής χειρισμού δεδομένων πρέπει να συνοδεύεται από την απαραίτητη εκπαίδευση, ώστε να πληροφορούνται οι υπάλληλοι τους κανόνες, αλλά και την απαίτηση να υπογράφουν οι υπάλληλοι δεσμευτικές δηλώσεις, όσον αφορά στις ευθύνες τους και τη δέσμευσή τους στην εργασία τους, ανάλογα με την πολιτική της επιχείρησης.
Τα προχωρημένα ή έξυπνα μέτρα προστασίας περιλαμβάνουν εκμάθηση μηχανής και χρονικούς αλγόριθμους συλλογιστικής για τον εντοπισμό ασυνήθιστης πρόσβασης στα δεδομένα (όπως βάσεις δεδομένων ή συστήματα ανάκτησης πληροφοριών), επιβεβαίωση με βάση τις ενέργειες των χρηστών (με βάση τις πληκτρολογήσεις και μοτίβα του ποντικιού), εντοπισμό ασυνήθιστων μοτίβων ανταλλαγής email, και εφαρμόζοντας δολώματα (honeypot) για τον εντοπισμό κακόβουλων εισβολέων.

Ο έλεγχος συσκευών πρόσβασης και η κρυπτογράφηση χρησιμοποιούνται για την αποτροπή πρόσβασης σε ένα μη εξουσιοδοτημένο χρήστη. Αυτά είναι τα πιο απλά μέτρα που μπορούν να παρθούν για την προστασία μεγάλου πλήθους προσωπικών δεδομένων έναντι κακόβουλων εξωτερικών και εσωτερικών επιθέσεων.
Καθορισμένες λύσεις DLP έχουν σκοπό την ανίχνευση και αποτροπή προσπαθειών αντιγραφής ή αποστολής ευαίσθητων δεδομένων, άθελα ή ηθελημένα, χωρίς εξουσιοδότηση, κυρίως από προσωπικό που έχει πρόσβαση σε αυτές τις ευαίσθητες πληροφορίες. Η κύρια δυνατότητα αυτών των λύσεων είναι η ικανότητα κατηγοριοποίησης του περιεχομένου ως ευαίσθητο. Οι προκαθορισμένες λύσεις

DLP υλοποιούνται συνήθως χρησιμοποιώντας μηχανισμούς, όπως ακριβή αντιστοίχηση δεδομένων, έλεγχος «δακτυλικού αποτυπώματος» δομημένων δεδομένων, στατιστικές μεθόδους, αντιστοίχηση κανόνων και τυπικών εκφράσεων, δημοσιοποιημένα λεξικά, εννοιολογικούς ορισμούς και λέξεις-κλειδιά [24].

Η κατηγορία των λύσεων Πρόληψης Διαρροής Δεδομένων (DLP), συνήθως αναφέρεται και ως Πρόληψη Απώλειας Δεδομένων (Data Loss/Leak/Leakage Protection/Prevention - DLP), Πρόληψη Διαρροής Πληροφοριών (Intrusion Loss/Leak Prevention/Protection - ILP), Συμμόρφωση Εξερχόμενου Περιεχομένου, Παρακολούθηση και Φιλτράρισμα Περιεχομένου, Παρακολούθηση και Προστασία Περιεχομένου (Content Monitor Filtering/Protection – CMF/CMP), ή Πρόληψη Εξώθησης (Extrusion Prevention) [25].

Πίνακας 3.2: Κατηγορίες τεχνολογικών προσεγγίσεων που χρησιμοποιούνται για τον εντοπισμό και την πρόληψη διαρροής δεδομένων

Βασικά μέτρα προστασίας	Τείχος Προστασίας, Antivirus, Σύστημα εντοπισμού εισβολών, Εξαρτημένα τερματικά, Πολιτικές.
Προχωρημένα/ Ευφυή μέτρα προστασίας	Εντοπισμός ανωμαλιών, Επιβεβαίωση βάσει δραστηριότητας, Παγίδες honeypots.
Έλεγχος πρόσβασης και Κρυπτογράφηση	Έλεγχος συσκευών, Κρυπτογράφηση, RMS (Rights Management Services)
Καθορισμένα Συστήματα DLP	Σαρώσεις δεδομένα σε κίνηση, δεδομένων σε χρήση και δεδομένων σε ακινησία.

4. Ταξινόμηση Συστημάτων Πρόληψης Απώλειας Δεδομένων

Οι λύσεις DLP μπορούν να χαρακτηριστούν σύμφωνα με την παρακάτω ταξινόμηση, που φαίνεται στην Εικόνα 4.1, η οποία έχει λάβει υπόψη της τα ακόλουθα χαρακτηριστικά: κατάσταση δεδομένων, πρόγραμμα ανάπτυξης, προσέγγιση διαχείρισης απώλειας, ενέργειες που γίνονται μετά την απώλεια και κανονιστική συμμόρφωση.

4.1 Κατάσταση Δεδομένων (Data State)

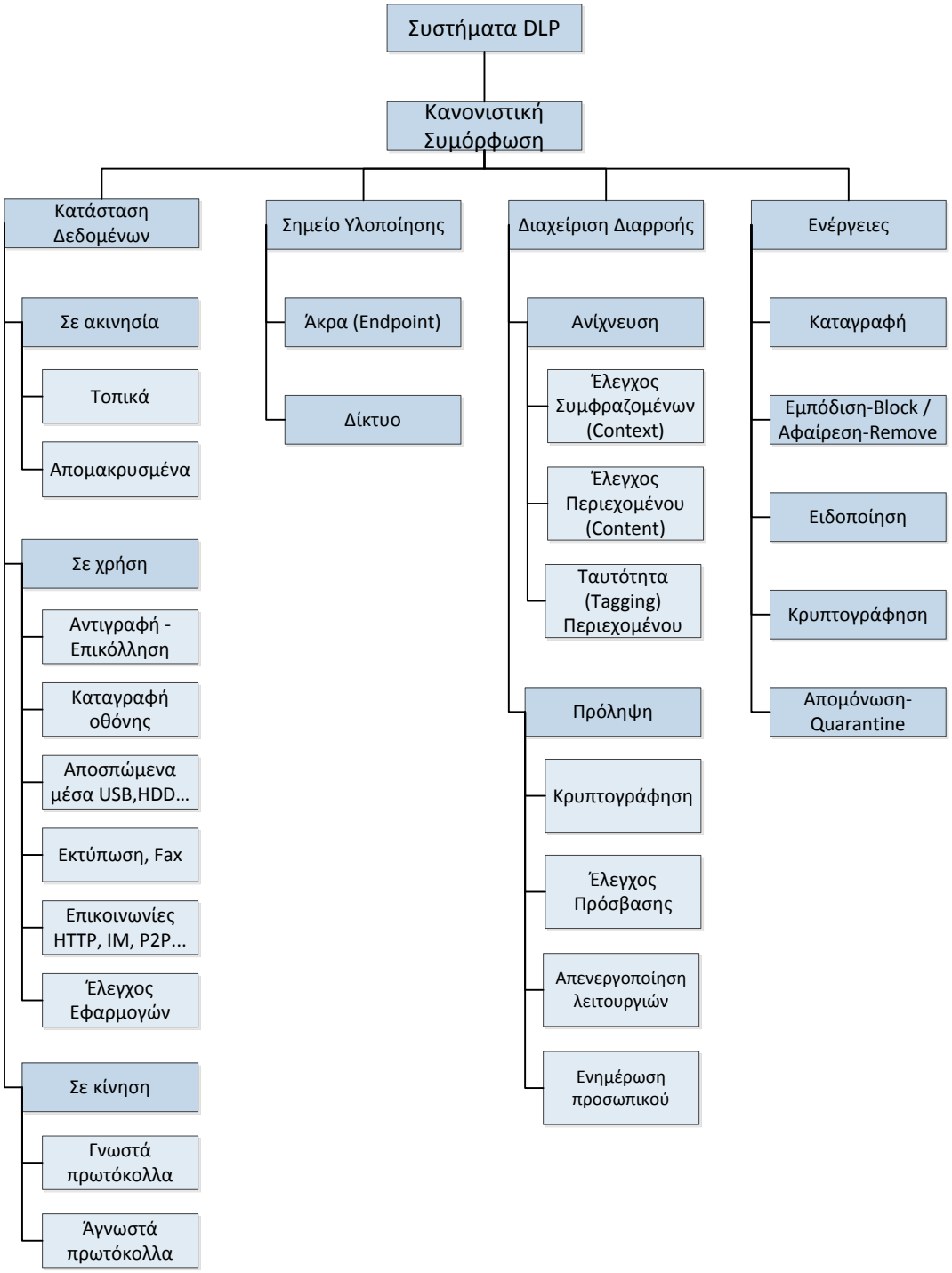
Οι λύσεις προστασίας απώλειας δεδομένων ταξινομούν τα δεδομένα σε τρεις κατηγορίες κατά τη διάρκεια του κύκλου ζωής τους: δεδομένα σε ακινησία (DAR), δεδομένα σε κίνηση (DIM) και δεδομένα σε χρήση (DIU).

4.1.1 Δεδομένα σε Ακινησία (Data at Rest - DAR)

Δεδομένα σε ακινησία ορίζονται όλα τα δεδομένα που είναι αποθηκευμένα στον υπολογιστή. Για να εμποδιστεί η πρόσβαση στα δεδομένα σε ακινησία, το να κλαπούν ή να αλλαχθούν από μη εξουσιοδοτημένα άτομα, χρησιμοποιούνται συνήθως μέτρα ασφαλείας, όπως η κρυπτογράφηση και ο έλεγχος πρόσβασης. Ένα προαπαιτούμενο για αυτά τα μέτρα ασφαλείας είναι η ανεύρεση του περιεχομένου, η οποία εξυπηρετεί στο να βρεθεί το σημείο που είναι όλα αυτά τα δεδομένα αποθηκευμένα. Ένας τρόπος για να επιτευχθεί αυτό είναι η χρήση των χαρακτηριστικών ανεύρεσης περιεχομένου των προϊόντων προστασίας διαρροής δεδομένων. Για παράδειγμα, μια πολιτική μπορεί να απαιτεί ότι όλοι οι αριθμοί πιστωτικών καρτών των πελατών μπορούν να αποθηκευτούν μόνο σε εγκεκριμένους εξυπηρετητές. Αν τα δεδομένα ανιχνευθούν σε μη εγκεκριμένο εξυπηρετητή, μπορούν να κρυπτογραφηθούν ή να απομακρυνθούν, ή να σταλεί μια προειδοποίηση στον ιδιοκτήτη των δεδομένων.

4.1.2 Δεδομένα σε Κίνηση (Data in Motion - DIM)

Τα δεδομένα σε κίνηση είναι δεδομένα που αποστέλλονται μέσω ενός δικτύου. Τα δεδομένα αυτά μπορούν να αποστέλλονται σε ένα εσωτερικό δίκτυο μιας επιχείρησης ή σε ένα εξωτερικό δίκτυο. Οι λύσεις προστασίας απώλειας δεδομένων χρησιμοποιούνται για να ανιχνεύσουν και να επιθεωρήσουν δεδομένα, τα οποία αποστέλλονται κατά μήκος των διαύλων επικοινωνίας σε ένα δίκτυο, το οποίο χρησιμοποιεί γνωστά πρωτόκολλα και εφαρμογές, συμπεριλαμβανομένων των email, http, άμεσου μηνύματος, αλλά και άγνωστων πρωτοκόλλων (επιθεωρώντας απλά το περιεχόμενο των πακέτων). Στην περίπτωση όμως που επιτρέπεται η κρυπτογράφηση ή οι κρυπτογραφημένες συνδέσεις χωρίς τη δυνατότητα αποκρυπτογράφησης των δεδομένων, μια τέτοια λύση προστασίας απώλειας δεδομένων δε θα είναι σε θέση να ανιχνεύσει την απώλεια ευαίσθητων κρυπτογραφημένων δεδομένων σε κίνηση. [26]



Εικόνα 4.1: Ταξινόμηση των λύσεων της Πρόληψης Απώλειας Δεδομένων [26]

4.1.3 Δεδομένα σε Χρήση (Data in Use - DIU)

Τα δεδομένα σε χρήση είναι οποιαδήποτε δεδομένα με το οποία ο χρήστης αλληλεπιδρά. Συστήματα τελικού σημείου (endpoint) χρησιμοποιούνται για να προστατέψουν τα δεδομένα σε χρήση και να παρακολουθούν τα δεδομένα καθώς ο χρήστης αλληλεπιδρά μαζί τους. Συνήθως χρησιμοποιείται ένα μέσο για να

παρακολουθεί τα δεδομένα κατά τη χρήση ή τη μεταφορά τους από τη συσκευή τελικού σημείου ή τον πελάτη μέσω διαφορετικών διαύλων εξόδου στις περιφερειακές συσκευές. Η υποκείμενη ιδέα είναι ότι αν γίνει μια προσπάθεια να αποσταλούν ευαίσθητα δεδομένα, η πιθανή απώλεια θα ανιχνευθεί αμέσως και θα εμποδιστεί (π.χ. θα μπλοκαριστεί ή θα κρυπτογραφηθεί) προτού σταλθούν τα δεδομένα. Τα εργαλεία των δεδομένων σε χρήση μπορούν να ελέγχουν τις ακόλουθες δραστηριότητες:

- Λειτουργίες αντιγραφής και επικόλλησης, λειτουργίες σύλληψης της οθόνης (screen capture).
- Μεταφορά ευαίσθητου περιεχομένου από ένα μέρος σε ένα άλλο με χρήση φορητών συσκευών, όπως USB δίσκων, CD/DVDs, έξυπνων τηλεφώνων και PDAs.
- Προσπάθειες να μεταφερθούν ευαίσθητα δεδομένα μέσω διαύλων επικοινωνίας, για παράδειγμα εσκεμμένη ή ακούσια αποστολή ευαίσθητων δεδομένων σε μορφή δακτυλογραφημένου περιεχομένου, συνημμένων αρχείων ή φωνητικών συνομιλιών μέσω μιας εφαρμογής άμεσων μηνυμάτων ή μίας ιστοσελίδας, ή αντιγράφοντας ευαίσθητο περιεχόμενο σε έναν μοιρασμένο φάκελο σε ένα LAN.
- Χρήση ευαίσθητων δεδομένων σε μη εγκεκριμένη εφαρμογή (όπως προσπάθεια κρυπτογράφησης δεδομένων με σκοπό να διαφύγει από τους αισθητήρες)

4.2 Επιλογή θέσης ανάπτυξης συστήματος Πρόληψης Απώλειας Δεδομένων

Υπάρχουν δύο βασικές επιλογές ανάπτυξης, που χρησιμοποιούνται για την εγκατάσταση προϊόντων προστασίας απώλειας δεδομένων:

Τελικό σημείο: Το λογισμικό DLP χρησιμοποιείται απευθείας σε συσκευές τελικού σημείου ή πελάτες. Παρακολουθεί και ελέγχει την πρόσβαση σε δεδομένα, ενώ ένας απομονωμένος εξυπηρετητής επιτήρησης αναλαμβάνει οργανωτικά καθήκοντα, κατανομή της πολιτικής και καταγραφή των συμβάντων. Το λογισμικό αυτό πρέπει να λειτουργεί εντός των περιορισμένων πόρων του τελικού σημείου, παρέχοντας ένα αποδεκτό επίπεδο προστασίας. Συνήθως προστατεύονται δεδομένα καθώς χρησιμοποιούνται από το χρήστη (δεδομένα σε χρήση) και ανιχνεύουν ευαίσθητα δεδομένα αποθηκευμένα στο τελικό σημείο (δεδομένα σε ακινησία). Ένα λογισμικό τελικού σημείου μπορεί να προστατεύσει δεδομένα, ακόμα και όταν είναι αποσυνδεδεμένο από το δίκτυο.

Δίκτυο: Μια λύση DLP μπορεί να αναπτυχθεί σε κάποια κομβικά σημεία του δικτύου. Αναλύοντας την κυκλοφορία του δικτύου, και βάσει μιας προκαθορισμένης πολιτικής, συμβάντα μπορούν να ενεργοποιούν τις πολιτικές αυτές και οι ύποπτες μεταφορές να μπλοκαριστούν. Ένα σύστημα DLP υλοποιημένο στο δίκτυο πρέπει να έχει τη δυνατότητα να υποστηρίζει πολλαπλά σημεία παρακολούθησης του δικτύου, ενώ ένας κεντρικός διοικητικός εξυπηρετητής να συλλέγει και να αναλύει όλα τα δεδομένα που λαμβάνονται από τα σημεία παρακολούθησης. [26]

4.3 Διαχείριση της Απώλειας Δεδομένων

Τα περιστατικά απώλειας δεδομένων μπορούν να αντιμετωπισθούν με βάση δύο

κύριες προσεγγίσεις: Προσέγγιση μέσω της ανίχνευσης και προσέγγιση μέσω της πρόληψης.

4.3.1 Προσέγγιση μέσω ανίχνευσης συμβάντων διαρροής (Detective Approach)

Στην προσέγγιση μέσω της ανίχνευσης, το σύστημα θα προσπαθήσει να ανιχνεύσει τα συμβάντα διαρροής και θα λάβει την κατάλληλη διορθωτική δράση για να αντιμετωπίσει οποιοδήποτε αναγνωρισμένο συμβάν διαρροής. Για παράδειγμα, όταν ένα τοπικό μέσο DLP ανιχνεύει ένα φάκελο που περιέχει ευαίσθητη πληροφορία σε έναν μη εγκεκριμένο εξυπηρετητή, μπορεί να μετακινήσει το φάκελο σε μια ασφαλή τοποθεσία. Η επιτήρηση βάσει του περιβάλλοντος, η επιτήρηση βάσει του περιεχομένου και η σήμανση του περιεχομένου είναι μορφές προσέγγισης της ανίχνευσης.

Η επιτήρηση βάσει του περιβάλλοντος (context based inspection), αξιοποιεί πληθώρα τεχνολογιών ασφαλείας, όπως τείχη προστασίας (firewalls), διακομιστές μεσολάβησης (proxies), συστήματα ανίχνευσης/πρόληψης (IDS/IPS) και φιλτράρισμα ενοχλητικής αλληλογραφίας. Ο όρος περιβάλλον (context) αναφέρεται από τη συμφραζόμενη πληροφορία που εξάγεται από τα παρακολουθούμενα δεδομένα, όπως πηγή, προορισμός, μέγεθος, παραλήπτες, αποστολέας, πληροφορία κεφαλίδας, metadata, χρονοσφραγίδες, τύπος αρχείου, τοποθεσία, μορφή, εφαρμογή και αναζητήσεις ή συναλλαγές. Ένα παράδειγμα συστήματος επιτήρησης μέσω περιβάλλοντος είναι το φιλτραρίσματος πακέτων στο firewall, το οποίο αποφασίζει αν σε ένα πακέτο δικτύου θα επιτραπεί να περάσει βάσει ρητών κριτηρίων φιλτραρίσματος, όπως διευθύνσεις IP πηγής/προορισμού, πύλη πηγής/προορισμού και άλλα χαρακτηριστικά του πακέτου. Μια λύση DLP βάσει περιβάλλοντος μπορεί να εμποδίσει την αποστολή αρχείων Java έξω από έναν οργανισμό, να μπλοκάρει όλα τα κρυπτογραφημένα αρχεία ή να εμποδίσει τη λειτουργία «αντιγραφή και επικόλληση» από ειδικές εφαρμογές.

Η επιτήρηση βάσει του περιεχομένου (content based inspection) ανιχνεύει τη διαρροή δεδομένων αναλύοντας το περιεχόμενο με χρήση ποικιλίας τεχνικών, όπως:

- Ένας συνδυασμός λεξικών που περιέχουν λέξεις κλειδιά, όπως «εμπιστευτικός», «οικονομική αναφορά», «πλάνο XYZ» και μοτίβα ή αντιστοίχιση τυπικών εκφράσεων (π.χ. ένα 16-ψηφιο μοτίβο για τον αριθμό μιας πιστωτικής κάρτας). Τα περισσότερα προϊόντα εμφανίζονται με κοινά λεξικά που αντιστοιχίζουν κανονισμούς και νόμους, όπως PCI-DSS, HIPAA, GLBA, και SOX. Αυτή η τεχνική ανίχνευσης δεδομένων είναι η πιο εύκολη και γρήγορη, αλλά παρέχει μικρή προστασία στην περίπτωση μη δομημένων δεδομένων.

- Ανάλυση φυσικής γλώσσας: Ανιχνεύει εάν τα ευαίσθητα δεδομένα και τα επιθεωρούμενα δεδομένα είναι παρόμοια, κάνοντας χρήση ανάλυσης της φυσικής γλώσσας. οι τεχνικές λεπτομέρειες τέτοιων μεθόδων δεν παρέχονται συνήθως από

τους προμηθευτές της λύσης.

- Στατιστική: Η προσέγγιση αυτή περιλαμβάνει στατιστική μετρική, η οποία λαμβάνεται από το υπό επιτήρηση περιεχόμενο (π.χ. συχνότητα όρων) και χρήση μεθόδων εκμάθησης μηχανής παρόμοιους με εκείνους που χρησιμοποιούνται για τον εντοπισμό ανεπιθύμητης αλληλογραφίας (spam). Η μέθοδος αυτή είναι αποτελεσματική για την ανίχνευση μη δομημένων δεδομένων σε περιπτώσεις όπου μια ντετερμινιστική τεχνική είναι δύσκολο να χρησιμοποιηθεί και η στατιστική μετρική είναι μια από τις καλύτερες διαθέσιμες προσεγγίσεις.

- Λήψη «δακτυλικών αποτυπωμάτων»: Μια μέθοδος που εξάγει τα «δακτυλικά αποτυπώματα» από ευαίσθητους φακέλους ή πεδία βάσεων δεδομένων και ψάχνει για ακριβή δακτυλικά αποτυπώματα για να ανιχνεύσει τη διαρροή. Ένα δακτυλικό αποτύπωμα είναι κατά προτίμηση η αντιστοίχιση μοναδικής τιμής τεμαχισμού (hash) με ένα σύνολο δεδομένων. Οι τιμές τεμαχισμού για όλους τους ευαίσθητους φακέλους αποθηκεύονται σε βάσεις δεδομένων ή τοπικά στην υπό επιτήρηση μηχανή. Το σύστημα συγκρίνει αυτές τις τιμές τεμαχισμού με τμήματα των επιτηρούμενων δεδομένων. Τα δακτυλικά αποτυπώματα μπορούν να δημιουργηθούν σε αρχεία βάσεων δεδομένων.

- Παρόλο που δεν είναι δύσκολο να κοιτάξει κάποιος ένα απλό κείμενο όπως ένα μήνυμα email, εμφανίζονται προβλήματα όταν πρόκειται να αντιμετωπιστούν δυαδικά αρχεία. Μια λύση DLP πρέπει να παρέχει δυνατότητες αναλύσεως των αρχείων (file-cracking), ώστε να ερμηνεύσει το κρυμμένο περιεχόμενο ενός αρχείου, το οποίο μπορεί να είναι αρκετά επίπεδα κάτω. Για παράδειγμα όταν ένα φύλο Excel βρίσκεται ενσωματωμένο σε ένα συμπιεσμένο φάκελο Word.

Επισήμανση περιεχομένου (content tagging): Στην προσέγγιση αυτή, μια επισήμανση αποδίδεται σε ένα αρχείο που περιέχει ευαίσθητα δεδομένα και επιβάλλεται μια πολιτική βάση της αποδιδόμενης σήμανσης. Το περιεχόμενο θα παραμείνει επισημασμένο ακόμα και όταν επεξεργάζεται από άλλες εφαρμογές. Για παράδειγμα, ένα αρχείο κειμένου Word που έχει επισημανθεί ως ευαίσθητο, θα παραμείνει επισημασμένο ακόμα και αν κρυπτογραφηθεί ή συμπιεστεί. Οι επισημάνσεις μπορούν να αποδοθούν σε αρχεία με διαφορετικούς τρόπους: χειροκίνητα από τον δημιουργό των ευαίσθητων δεδομένων, αυτόματα με χρήση ανάλυσης βάσει περιβάλλοντος ή περιεχομένου, αυτόματα σε όλα τα αρχεία που βρίσκονται σε μια συγκεκριμένη τοποθεσία ή αυτόματα σε όλα τα αρχεία που έχουν δημιουργηθεί από συγκεκριμένες εφαρμογές ή συγκεκριμένους χρήστες.

4.3.2 Προσέγγιση μέσω πρόληψης συμβάντων διαρροής (Preventive Approach)

Στην προληπτική προσέγγιση, τα συμβάντα διαρροής προλαμβάνονται, λαμβάνοντας κατάλληλα μέτρα. Οι λύσεις DLP υποστηρίζουν την πρόληψη διαρροής δεδομένων με χρήση των κάτωθι προληπτικών προσεγγίσεων:

Κρυπτογράφηση: Ορίζεται μια πολιτική που δηλώνει ποια ευαίσθητα δεδομένα πρέπει να κρυπτογραφηθούν και από ποιον επιτρέπεται η αίτηση αποκρυπτογράφησης των δεδομένων αυτών. Περιορίζονται επίσης οι εφαρμογές στις οποίες επιτρέπεται η χρήση ευαίσθητων δεδομένων, επιτρέποντας την κρυπτογράφηση μόνο με εγκεκριμένες λύσεις.

Έλεγχος πρόσβασης: Η προσέγγιση αυτή περιλαμβάνει τη δυνατότητα να επιτρέπεται ή να απαγορεύεται η χρήση συγκεκριμένων πόρων από μια συγκεκριμένη οντότητα. Μία λύση DLP, με την βοήθεια της ορισμένης πολιτικής της, μπορεί να απαγορεύσει τη χρήση ενός πόρου, ακόμα και αν έχει χορηγηθεί πρόσβαση π.χ. ανάγνωσης και εγγραφής σε κάποια ευαίσθητη πληροφορία. Ένας τρόπος επίτευξης ελέγχου πρόσβασης επιτυγχάνεται μέσω της συνεργασίας με τη διαχείριση ψηφιακών δικαιωμάτων της εταιρίας (Enterprise Digital Rights Management – (E-DRM)), η οποία εφαρμόζει έλεγχο πρόσβασης στα αρχεία αυτόματα.

Απενεργοποίηση λειτουργιών: Αυτή η προληπτική προσέγγιση περιλαμβάνει την απενεργοποίηση λειτουργιών, οι οποίες μπορούν να καταλήξουν σε ανάρμοστη χρήση ευαίσθητων δεδομένων. Αυτό μπορεί να γίνει, για παράδειγμα, περιορίζοντας τις λειτουργίες αντιγραφής και επικόλλησης σε ευαίσθητο περιεχόμενο, ή περιορίζοντας την αντιγραφή του περιεχομένου σε φορητά αποθηκευτικά μέσα.

Ενημέρωση προσωπικού: Περιλαμβάνει την αύξηση του επιπέδου συνειδητοποίησης των εργαζομένων, ενημερώνοντας τους ποιος έχει πρόσβαση σε τί, ποια δεδομένα είναι ιδιαιτέρως ευαίσθητα και τί πρέπει να γίνει για να διασφαλιστεί ότι τα δεδομένα δε θα υποστούν κακή χρήση.

4.4 Διορθωτικές Ενέργειες

Οι ενέργειες που μπορούν να γίνουν μετά από κάποια αναγνωρισμένη απόπειρα απώλειας δεδομένων, όπως αναφέρθηκε και στις ανωτέρω παραγράφους, μπορεί να είναι α) Η καταγραφή του συμβάντος, β) Το σταμάτημα της διαρροής, με τη αφαίρεση των ευαίσθητων δεδομένων ή της όλης ύποπτης κίνησης, γ) Την ειδοποίηση των υπευθύνων με διαφόρους τρόπους, δ) Την αυτόματη κρυπτογράφηση των ευαίσθητων δεδομένων, προ της αποστολής τους ε) Την καραντίνα των ευαίσθητων δεδομένων που προσπαθούν να διαφύγουν.

4.5 Κανονιστική Συμμόρφωση (Regulatory Compliance)

Κάθε οργανισμός οφείλει να υπακούει σε μια ή περισσότερες διεθνής, κοινοτικές, εθνικές, τοπικές ή εσωτερικές κανονιστικές ρυθμίσεις. Είτε πρόκειται για πρότυπα οργανισμών υγείας, ασφάλισης, οικονομικών, είτε για οδηγίες προστασίας δεδομένων από διάφορους φορείς (όπως η Ευρωπαϊκή Ένωση), είτε πρόκειται για άλλους κανονισμούς, οι εταιρίες οφείλουν να λάβουν μέτρα για την προστασία προσωπικών και ιδιωτικών πληροφοριών. Οι προσωπικές πληροφορίες ορίζονται ως οποιοδήποτε τμήμα πληροφοριών που μπορεί δυνητικά να χρησιμοποιηθεί για την ταυτοποίηση, την επικοινωνία, ή τον εντοπισμό ενός ατόμου, είτε πρόκειται για πελάτη, υπάλληλο,

φοιτητή, ασθενή, ή φορολογούμενο.

Η δημιουργία των κανονισμών είναι πολύπλοκη διαδικασία, και δεν έχουν γραφτεί όλες με γνώμονα τις πραγματικές δυνατότητες των σημερινών τεχνολογιών, ή τις εταιρικές πρακτικές που έχουν εξελιχθεί για να ανταποκρίνονται στις απαιτήσεις τους. Ακολούθως δίνονται τρία χαρακτηριστικά παραδείγματα κανονισμών που εφαρμόζονται σε μεγάλη πλειοψηφία οργανισμών.

Ο νόμος *Gramm-Leach-Biley* (GLB) του 1999 των ΗΠΑ, έχει σκοπό τη διασφάλιση της προστασίας των προσωπικών οικονομικών δεδομένων των πελατών, όπου αναφέρεται σε Μη Δημοσιοποιήσιμες Προσωπικές Πληροφορίες (Nonpublic Personal Information - NPI). Οι δυο περιοχές μεγαλύτερου ενδιαφέροντος για τις περισσότερες εταιρίες είναι ο Κανόνας Οικονομικής Εμπιστευτικότητας (Financial Privacy Rule), ο οποίος καλύπτει τη συλλογή, τη χρήση και την αποκάλυψη των NPI, και ο κανόνας Διασφάλισης, που περιγράφει τις διαδικασίες που πρέπει να υιοθετήσουν οι εταιρίες για την προστασία NPI. Έτσι, λοιπόν, ενώ η GLB δεν αναφέρεται σε συγκεκριμένες τεχνολογίες, στην πράξη, ο Κανόνας Διασφάλισης αφορά εφαρμογή πολιτικής που να μπορεί να κρυπτογραφήσει ή να εμποδίσει την κυκλοφορία ηλεκτρονικής αλληλογραφίας, ανάλογα τον αποστολέα, τον παραλήπτη και το περιεχόμενο. Επιπρόσθετα, οι εταιρίες θα πρέπει να υιοθετήσουν συστήματα που θα παρέχουν καταγραφή και αναφορά, επιτρέποντας να επιδεικνύουν τη συμμόρφωση τους στους κανόνες.

Ο νόμος *Φορητότητας και Λογοδοσίας της Ασφάλισης Υγείας* (Health Insurance Portability and Accountability Act - HIPAA) του 1996 των ΗΠΑ, θέτει έναν αριθμό απαιτήσεων όσον αφορά τις πρακτικές διαχείρισης πληροφοριών συστημάτων υγείας, και έχει άμεσο αντίκτυπο στη διαδικασία των συστημάτων μηνυμάτων. Υπό αυτόν τον κανονισμό, οι οργανισμοί πρέπει να διασφαλίζουν ότι ηλεκτρονική αλληλογραφία που περιέχει πληροφορίες ασθενών είναι προστατευμένη, ακόμη και όταν αποστέλλεται και από αποκρυπτογραφημένους συνδέσμους, αρκεί οι αποστολείς και οι παραλήπτες να είναι επαρκώς επαληθευμένοι και πιστοποιημένοι. Συνεπώς, ο κανονισμός HIPAA επηρεάζει πληροφορίες κατά τη μεταφορά τους αλλά και σε ηρεμία.

Τέλος, η *Οδηγία Προστασίας Δεδομένων της Ευρωπαϊκής Ένωσης* του 2002 είναι ένας κανονισμός που θέτει νέα νομικά πρότυπα για την επεξεργασία των προσωπικών δεδομένων και την προστασία της εμπιστευτικότητας τους. Ο νόμος θέτει αυστηρούς περιορισμούς όσον αφορά τη συλλογή και την αποθήκευση προσωπικών πληροφοριών και επίσης θέτει όρους για τη μεταφορά προσωπικών δεδομένων σε χώρες εκτός ΕΕ.

Η απώλεια δεδομένων δεν είναι σημαντικό πρόβλημα μόνο για εταιρίες με ευαίσθητα δεδομένα, αλλά για κάθε εμπορικό οργανισμό παγκοσμίως. Δυστυχώς, η συμμόρφωση των εταιριών επιτυγχάνεται με κανονιστικές παγίδες. Απλά καθημερινά σφάλματα, όπως αποστολή νόμιμης ηλεκτρονικής αλληλογραφίας (που περιέχει αποκρυπτογραφημένες πληροφορίες πιστωτικών καρτών) ή διαμοιρασμός κάποια αναφοράς που περιέχει ιατρικά δεδομένα υπαλλήλων σε μη εξουσιοδοτημένα άτομα, μπορεί να αποτελέσουν κανονιστικές παραβάσεις. [27, 14]

5. Επισκόπηση των ήδη υπαρχουσών εμπορικών DLP συστημάτων

5.1 Τεχνολογικές λύσεις των προϊόντων της αγοράς

Σύμφωνα με την έκθεση της Forrester Wave [28], οι πρώτες DLP λύσεις επικεντρώθηκαν στην αποκάλυψη ευαίσθητων δεδομένων, μέσω της παρακολούθησης των εν κινήσει δεδομένων σε διάφορα σημεία εξόδου του δικτύου, καθώς αυτά φεύγουν από το εταιρικό δίκτυο. Σε δεύτερο στάδιο, καθώς πολλαπλασιάστηκαν οι αποσπώμενες συσκευές αποθήκευσης όπως τα USB flash disks, οι εξωτερικοί σκληροί δίσκοι κ.τ.λ., οι λύσεις DLP άρχισαν να επικεντρώνονται στην ανίχνευση διαρροών δεδομένων στα άκρα (at the endpoint) και την παροχή δυνατοτήτων αποτροπής π.χ. αντιγραφής ευαίσθητων πληροφοριών σε USB συσκευές, CD ή DVD, ακόμη και αν ο εκάστοτε εξοπλισμός άκρου (Η/Υ γραφείου, φορητός Η/Υ, κινητό, κ.τ.λ.) δεν είναι συνδεδεμένος στο δίκτυο.
Η μεγαλύτερη πρόκληση για τα DLP συστήματα, έγκειται στην προστασία του μεγάλου αδόμητης μορφής όγκου δεδομένων ευαίσθητων δεδομένων όπως είναι οι διάφορες μορφές της πνευματικής ιδιοκτησίας, όπως ο πηγαίος κώδικας, οι κατάλογοι των πελατών ή τα σχέδια των νέων προϊόντων. Ως εκ τούτου, οι κατασκευαστές των DLP λύσεων συνεχώς βελτιώνουν τις μεθόδους αποκάλυψης ευαίσθητων δεδομένων, χρησιμοποιώντας προσεγγίσεις όπως η λήψη δακτυλικών αποτυπωμάτων και επεξεργασίας της φυσικής γλώσσας [22]. Σύμφωνα με τρεις κορυφαίες εκθέσεις ερευνών, μέσα στα επόμενα λίγα χρόνια, τα προϊόντα DLP αναμένεται να σταθεροποιηθούν και να είναι κοινός τόπος, όπως είναι οι υπάρχουσες λύσεις ασφαλείας των firewalls, IDS και IPS συστημάτων, καθώς και των συστημάτων ανίχνευσης κακόβουλου λογισμικού (anti-malware και antivirus) [22, 24, 28].

5.1.1 Τεχνολογικές προσφορές από ηγέτες της αγοράς

Με βάση την ταξινόμηση που περιγράφεται στο προηγούμενο κεφάλαιο, θα περιγράψουμε τις κύριες λειτουργίες των DLP συστημάτων που προσφέρονται από ηγέτες της αγοράς. Σε ένα αντιπροσωπευτικό παράδειγμα της αρχιτεκτονικής του συστήματος DLP έχει παρουσιαστεί από τον Lawton (2008). οι λειτουργίες αυτές μπορούν να χαρακτηριστούν, βασιζόμενοι στην κατάσταση των δεδομένων που προσπαθεί να προστατεύσει. Ο χαρακτηρισμός αυτός κατατάσσει τα δεδομένα σε δεδομένα σε ακινησία, σε χρήση, ή εν κινήσει.
Προστασία για δεδομένα σε ακινησία: Παρέχει λύσεις οι οποίες προσπαθούν να εντοπίσουν αλλά και να ελέγξουν το ευαίσθητο περιεχόμενο από σταθερούς Η/Υ, φορητούς, συσκευές μαζικής αποθήκευσης (π.χ. διακομιστές αρχείων), e-mail servers, συστήματα διαχείρισης εγγράφων, βάσεων δεδομένων κ.τ.λ., που είναι αποθηκευμένα σε μη εξουσιοδοτημένα σημεία. Υπάρχουν δύο βασικές τεχνικές για τον εντοπισμό του ευαίσθητου περιεχομένου:

- Τοπική σάρωση (Local scanning) (π.χ. agent-based): Σε αυτήν την τεχνική, ένα πρόγραμμα έχει εγκατασταθεί σε ένα μηχάνημα και σαρωτικά ελέγχει τοπικά το μηχάνημα για εύρεση περιεχομένου που παραβιάζει την πολιτική.

Αν βρεθούν τέτοιου είδους περιεχόμενα, γίνονται αυτόματα κινήσεις όπως η μετακίνηση, η κρυπτογράφηση ή η καραντίνα αυτών. Αν και η τεχνική αυτή έχει το μειονέκτημα ότι τα προγράμματα αυτά περιορίζονται από την επεξεργαστική ισχύ και τη μνήμη του μηχανήματος που είναι εγκατεστημένα και τρέχουν, το κύριο πλεονέκτημα της είναι ότι το πρόγραμμα μπορεί να είναι πάντα ενεργό και να επιβάλει την πολιτική, ακόμη και αν η τοπική συσκευή δεν είναι συνδεδεμένη στο δίκτυο.

- Απομακρυσμένη σάρωση: Η σάρωση γίνεται εξ αποστάσεως κάνοντας μια σύνδεση σε ένα διακομιστή ή μία συσκευή, χρησιμοποιώντας ένα πρωτόκολλο διαμοιρασμού αρχείων (file sharing) ή ένα πρωτόκολλο εφαρμογής. Το μειονέκτημα αυτής της τεχνικής είναι ότι αυξάνει την κίνηση του δικτύου και έχει περιορισμούς στην απόδοση με βάση το διαθέσιμο εύρος ζώνης του δικτύου και τους περιορισμούς του απομακρυσμένου μηχανήματος.

Για την προστασία των δεδομένων σε ακινησία, παρέχεται επίσης κρυπτογράφηση των δεδομένων στο τελικό σημείο. Αυτό μπορεί να γίνει με πλήρη κρυπτογράφηση δίσκου ή με κρυπτογράφηση σε επίπεδο αρχείου, αλλά και με τον έλεγχο της πρόσβασης. Η κρυπτογράφηση για παράδειγμα θα προστατεύσει τα ευαίσθητα δεδομένα, αν κλαπεί ή χαθεί ένας φορητός υπολογιστής.

Προστασία των δεδομένων σε χρήση: Παρέχεται από ένα τοπικά εγκατεστημένο πρόγραμμα (agent) το οποίο παρακολουθεί τοπικά και αποτρέπει ενέργειες που αφορούν τα ευαίσθητα δεδομένα. Τέτοιες ενέργειες μπορούν να είναι η αντιγραφή και επικόλληση, η εκτύπωσης οθόνη, η αντιγραφή σε ένα αποσπώμενο μέσο όπως είναι τα USB, CD ή DVD, η μετάδοση μέσω δικτύου χωρίς άδεια, ή η χρήση των δεδομένων σε μη εγκεκριμένες εφαρμογές.

Προστασία των δεδομένων εν κινήσει: Παρέχεται από στοιχεία μίας δικτυακής λύσης, η οποία παρακολουθεί τα δεδομένα που κινούνται στο δίκτυο, ελέγχει αν το περιεχόμενό τους παραβιάζει την πολιτική που έχει οριστεί και τα σταματάει αν όντως υπάρχει μία τέτοια παραβίαση. Τα στοιχεία παρακολούθησης του δικτύου συχνά τοποθετούνται στην δικτυακή πύλη (gateway) του οργανισμού ή κοντά σε αυτήν. Η λειτουργία τους καλύπτει πλήρης σύλληψη του κάθε πακέτου, επανασύσταση της συνόδου (session) και ανάλυση του περιεχομένου σε πραγματικό χρόνο. Ένας διακομιστής proxy συνήθως παρέχει βαθύ έλεγχο περιεχομένου σε διάφορα πρωτόκολλα, αλλά κυρίως σε HTTP, FTP, και IM. Για το έλεγχο του ηλεκτρονικού ταχυδρομείου ώστε να μπορούν να τεθούν κάποια μηνύματα σε καραντίνα, να κρυπτογραφηθούν ή να υπάρχει κάποιο φιλτράρισμα ύποπτων μηνυμάτων, ενσωματώνεται και μία υπηρεσία μεταφοράς ταχυδρομείου (mail transport agent (MTA)).

Ένα σημαντικό χαρακτηριστικό που παρέχεται από τα περισσότερα DLP προϊόντα είναι η κεντρική διαχείρισή τους, η οποία επιτρέπει:

α) Τον καθορισμό των ευαίσθητων δεδομένων (π.χ. μέσω μίας κανονικής έκφραση (regular expression) που ταιριάζει με έναν αριθμό πιστωτικής κάρτας)

β) Τον καθορισμό κανόνων που καθορίζουν τις ενέργειες που πρέπει να ληφθούν κατά τη διαρροή ευαίσθητων δεδομένων που ανιχνεύονται. Οι κανόνες αυτοί που μπορεί να καθορίζονται και να εφαρμόζονται ανά συγκεκριμένο χρήστη ή ομάδα χρηστών, να καθορίζονται σύμφωνα με το είδος των δεδομένων προς έλεγχο (π.χ., μόνο τα μηνύματα ηλεκτρονικού ταχυδρομείου, τις φόρμες, ή μόνο δεδομένα που

είναι αποθηκευμένα σε MS SharePoint), να καθορίζονται βάση της πληθικότητας (π.χ., περισσότεροι από 5 αριθμοί πιστωτικών καρτών στο ίδιο email) ή της γειτνίασης (π.χ., Όνομα και αριθμός ταυτότητας στο ίδιο e-mail είναι αποδεκτό, αλλά όνομα, αριθμός ταυτότητας, και ιατρικά δεδομένα δεν είναι αποδεκτό). Παραδείγματα τέτοιων κανόνων είναι τα κάτωθι:
i) Ένα μήνυμα ηλεκτρονικού ταχυδρομείου στο οποίο ανιχνεύθηκε ότι περιέχει ευαίσθητα δεδομένα, βασισμένο στον έλεγχο τύπου δακτυλικών αποτυπωμάτων, θα πρέπει να ανακατευθυνθεί προς κρυπτογράφηση.
ii) Ένα αρχείο που περιέχει τη λέξη-κλειδί «Έργο Χ», θα πρέπει να οριστούν συγκεκριμένα δικαιώματα πρόσβασης, ώστε να υπάρχει ελεγχόμενη πρόσβαση σε αυτό.
iii) Ένα αρχείο που περιέχει π.χ. αριθμούς ΑΦΜ ή ΑΜΚΑ σε δημόσια προσπελάσιμες τοποθεσίες, θα πρέπει να μεταφερθεί στον ορισμένο ασφαλή διακομιστή «Α».
iv) Κεντρική καταγραφή γεγονότων, μέσω της οποίας ο διαχειριστής μπορεί να αντιληφθεί σφαιρικότερα και ευκολότερα τα συμβάντα, μέσω της συνένωσης καταγραφών από διαφορετικά συστήματα, την ομαδοποίηση αυτών και την συσχέτισή τους.

Η διαχείριση ενός τέτοιου συστήματος παρέχει επίσης τη δυνατότητα διάδοσης των ενημερωμένων πολιτικών, σε όλες τις μονάδες του DLP συστήματος, παρέχοντας ταυτόχρονα την δυνατότητα αναφορών και ειδοποιήσεων σχετικά με τις παραβιάσεις της πολιτικής, όπως επίσης και τη διερεύνηση και αντιμετώπιση περιστατικών διαρροής. Ως εκ τούτου, ένα σημαντικό χαρακτηριστικό του συστήματος είναι η δυνατότητα αναπαράστασης περιστατικών διαρροής με όλες τις σχετικές πληροφορίες σχετικά με τις ενέργειες των χρηστών (όπως πληκτρολογήσεις, άνοιγμα αρχείων, και επισκεπτόμενων ιστοσελίδων) πριν ή μετά από κάποιο περιστατικό διαρροής.

Ο κάτωθι πίνακας, συνοψίζει λειτουργίες βασικών εμπορικών DLP προϊόντων, σύμφωνα με την ταξινόμηση που παρουσιάστηκε σε προηγούμενο κεφάλαιο.

Πίνακας 5.1: Λειτουργίες βασικών εμπορικών προϊόντων DLP [29]

Κατασκευαστές:	Α	Β	Γ	Δ	Ε	ΣΤ	Ζ	Η	Θ	Ι	ΙΑ	ΙΒ
Δεδομένα σε Ακινησία												
Τοπικά	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ
Απομακρυσμένα	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ
Δεδομένα σε Χρήση												
Αντιγραφή-Επικόλληση	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ
Καταγραφή Οθόνης	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	?
Αποσπώμενα μέσα (USB,	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ

HDD, …)												
Εκτύπωση, Fax	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ
Επικοινωνίες (HTTP, IM, FTP, P2P,…)	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ
Έλεγχος Εφαρμογών	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ
Δεδομένα εν Κινήση												
Γνωστά Πρωτόκολλα	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ
Άγνωστα Πρωτόκολλα	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	?	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ
Σημεία Υλοποίησης												
Στα Άκρα	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ
Στο Δίκτυο	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ
Διαχείριση Διαρροής												
Ανίχνευση												
Έλεγχος Συμφραζομένων(Context)	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ
Έλεγχος Περιεχομένου (Content)	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ
Ταυτότητα (Tagging) Περιεχομένου	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ
Πρόληψη												
Κρυπτογράφηση	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ
Έλεγχος Πρόσβασης	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ
Απενεργοποίηση λειτουργιών	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ
Ενέργειες												
Καταγραφή	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ
Εμπόδιση/Αφαίρεση	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ
Ειδοποίηση	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ
Κρυπτογράφηση	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ
Απομόνωση	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ

5.2 Περιορισμοί των Συστημάτων Πρόληψης Απώλειας Δεδομένων της αγοράς

Από την ανάλυση εμπορικών DLP προϊόντων, είναι προφανές ότι η DLP τεχνολογία είναι σχετικά νέα και η δομή των παρεχόμενων εμπορικών λύσεων χαρακτηρίζεται από μεγάλους εταιρίες ασφαλείας που έχουν ενσωματώσει λύσεις μικρών εταιριών, οι οποίες ειδικεύονται σε συγκεκριμένα σενάρια διαρροής δεδομένων.

Η σημασία των DLP λύσεων είναι κατανοητή από τους κατασκευαστές DLP λύσεων και συναφών λύσεων ασφαλείας. Ως εκ τούτου υπάρχει η τάση συγκέντρωσης DLP χαρακτηριστικών σε προϊόντα λύσεων ασφαλείας, σε μια προσπάθεια παροχής ολικής προστασίας των ευαίσθητων δεδομένων. Παραδείγματα αυτής της τάσης είναι η ενσωμάτωση DLP χαρακτηριστικών σε συστήματα διαχείρισης δικαιωμάτων, τοποθεσίες συνεργασίας και κοινής χρήσης τύπου Sharepoint, φιλτραρίσματος του Web και δυνατότητες φιλτραρίσματος εντός πλαισίου virtualization. Ωστόσο οι DLP λύσεις παρέχουν συνήθως ελάχιστη άμυνα κατά των ενδοεταιρικών διαρροών, και επικεντρώνονται στην παρούσα φάση τουλάχιστον, στην πρόληψη διαρροών εκτός της εταιρικής περιμέτρου. Η ενσωμάτωση όμως σε virtualization περιβάλλοντα, επιτρέπει στους οργανισμούς την παροχή εσωτερικής DLP ασφάλειας με μικρή προσπάθεια.

Το πιο σημαντικό βήμα κατά την εγκατάσταση ενός DLP προϊόντος, είναι ο καθορισμός της αρχικής πολιτικής και η παροχή του ευαίσθητου περιεχομένου που θα χρησιμοποιείται από το σύστημα για να εξάγεται η αναπαράσταση του ευαίσθητου περιεχομένου. Στις περισσότερες περιπτώσεις, ο οργανισμός υποχρεούται να παρέχει μια σειρά από ευαίσθητα έγγραφα, προκειμένου να παραμετροποιηθεί το σύστημα. Αυτό δεν είναι ένα ασήμαντο έργο, ειδικά για μεγάλους οργανισμούς, αλλά και ένα έργο που λόγω κόστους και κόπου συχνά παραλείπεται στις μικρές εταιρίες. Η προσθήκη νέων κανόνων, εξειδικευμένων για τον συγκεκριμένο οργανισμών είναι επίσης ένα δύσκολο έργο, γιατί πρέπει να ληφθεί υπόψη το ποσοστό θετικά ψευδών αποτελεσμάτων της μηχανής ανίχνευσης.

Επί του παρόντος, οι DLP λύσεις δεν περιλαμβάνουν την δυνατότητα συσχέτισης των ειδοποιήσεων. Τη συσχέτιση δηλαδή με άλλες ειδοποιήσεις που εμφανίστηκαν στο ίδιο DLP σύστημα ή σε άλλα συστήματα ασφαλείας όπως για παράδειγμα στα IDS, ή τα anti-malware συστήματα. Για παράδειγμα, το γεγονός ότι ένας χρήστης προσπάθησε και απέτυχε να στείλει εκτός οργανισμού ένα email με ευαίσθητο αρχείο, δεν έχει καμία σχέση για το σύστημα, με το ότι ο ίδιος χρήστης, εκτυπώνει ή κρυπτογραφεί το ίδιο αρχείο, ένα λεπτό αργότερα, το οποίο και επιτρέπεται. Αυτό καθιστά δύσκολη την ανάλυση των προβλεπόμενων π.χ. ημερήσιων αναφορών που προβλέπονται από το σύστημα. Ισχύει δε αυτό σε μεγαλύτερο βαθμό στους μεγάλους οργανισμούς οι οποίοι έχουν εκατοντάδες περιστατικά ανά ημέρα.

Τέλος, από την ανάλυση των μεθόδων ανίχνευσης που χρησιμοποιείται από τα προϊόντα αυτά, οι συγγραφείς κατέληξαν στο συμπέρασμα ότι οι σημερινές λύσεις DLP, προορίζονται κυρίως για τον μετριασμό των μη εσκεμμένων περιστατικών διαρροής. Το συμπέρασμα αυτό ενισχύεται από το γεγονός ότι οι μέθοδοι ανίχνευσης που χρησιμοποιούνται από τους περισσότερους κατασκευαστές DLP

λύσεων είναι πολύ απλές και περιορισμένοι, καθώς μπορούν εύκολα να παρακαμφθούν με την αντικατάσταση λέξεων-κλειδιών, φράσεων σε δομημένα δεδομένα για την αποφυγή του εντοπισμού κανονικών εκφράσεων καθώς και άλλων κοινών μηχανισμών. Η τεχνική των «δακτυλικών αποτυπωμάτων» μπορεί επίσης να παρακαμφθεί με ελάχιστη προσπάθεια, αντικαθιστώντας μίας μόνο λέξη ή χαρακτήρα δίνοντας ένα ψευδώς αρνητικό αποτέλεσμα. Επιπλέον κατά τη διάρκεια της διαδικασίας «λήψεως δακτυλικών αποτυπωμάτων» ενός κειμένου, πρέπει να ληφθεί μέριμνα ώστε να βεβαιωθεί ότι το εισερχόμενο κείμενο δεν επηρεάζει το αρχικό. Ο μετριασμός των περιστατικών σκόπιμης διαρροής δεδομένων παρέχεται κυρίως με την ενσωμάτωση κρυπτογράφηση του δίσκου, του αρχείου αλλά και με την υλοποίηση υπηρεσιών διαχείρισης ψηφιακών δικαιωμάτων (DRM - Digital Rights Management). [29]

6. Ακαδημαϊκές μελέτες στο πεδίο της Πρόληψης Απώλειας Δεδομένων

Κατά τα τελευταία χρόνια, έχει αναγνωρισθεί η πρόκληση της ενασχόλησης με την εκ των έσω κακόβουλη απειλή και έχουν προταθεί πολλές μέθοδοι για την επίλυση αυτού του προβλήματος. Η διαρροή δεδομένων είναι ένας από τους κύριους στόχους των εκ των έσω απειλών [29]. Ως εκ τούτου οι περισσότερες προτεινόμενοι μέθοδοι ανίχνευσης εσωτερικών απειλών έχουν ως σκοπό την πρόληψη και ανίχνευση διαρροών δεδομένων.

Αρχικά το 2005 [30] από τους Maybury και λοιπούς παρουσιάστηκε στην κοινότητα πληροφοριών των ΗΠΑ (U.S. intelligence community) συλλογική μελέτη που περιελάμβανε τον χαρακτηρισμό και την ανάλυση των μεθόδων που χρησιμοποιούνται για την αντιμετώπιση των εκ των έσω κακόβουλων απειλών. Η μελέτη προτείνει ένα γενικό μοντέλο κακόβουλης εκ των έσω συμπεριφοράς, διάκριση κινήτρων, δράσεις, καθώς και συναφή παρατηρήσιμα μεγέθη. Αρκετές πρότυπες τεχνικές αναπτύχθηκαν για την παροχή έγκαιρης προειδοποίησης της κακόβουλης δραστηριότητας, συμπεριλαμβανομένης της χρήσης δολωμάτων (honeytokens), σκιαγράφησης δικτυακής επικοινωνίας και γνωσιακών αλγορίθμων για δομημένη ανάλυση και συγχώνευση δεδομένων. Η απόδοση αυτών των τεχνικών μετρήθηκε σε ένα λειτουργικό δίκτυο από τρεις διακριτές κατηγορίες κατόχων εμπιστευτικών πληροφοριών (έναν αναλυτή, ένα διαχειριστή εφαρμογής, και το διαχειριστή του συστήματος). Το επόμενο έτος οι Gaonjur και Bokhoree σε μελέτη τους, επικεντρώθηκαν στους κινδύνους από εσωτερικές απειλές στον τομέα της πληροφορικής, μέσω εξωτερικών συνεργατών μιας εταιρίας. Η μελέτη αυτή πρότεινε μείωση αυτών των κινδύνων με τη χρήση μη παραπλανητικών τεχνικών, όπως συστήματα ανίχνευσης παραβίασης, αλλά και παραπλανητικές τεχνικών όπως τα honeypots. Ένα σωστά ασφαλισμένο δίκτυο με κατάλληλα συστήματα ανίχνευσης παραβίασης και ύπαρξη δολωμάτων, μπορεί να εξασφαλίσει τη συνολική ασφάλεια και την ακεραιότητα των ευαίσθητων πληροφοριών των χρηστών. Το 2008 και το 2010, [29, 31] έρευνα των Salem et al., και των Hong et al. αντίστοιχα, προτείνει μεθόδους για την ανίχνευση επιθέσεων εκ των έσω, μέσω σκιαγράφησης προφίλ χρηστών ανά υπολογιστή, παρατηρώντας χαρακτηριστικά όπως η χρήση βάσεων δεδομένων, η πρόσβαση σε αρχεία συστήματος, οι κλήσεις προς το σύστημα, οι εντολές προς το λειτουργικό σύστημα, ανίχνευσης της δικτυακής κίνησης καθώς και χρήσης δολωμάτων. Ένα παράδειγμα είναι ένα πλαίσιο [32], στο οποίο συνδυάζεται η σκιαγράφηση ενός προφίλ χρηστών ανά υπολογιστή με υπομονάδες δημιουργίας και παρακολούθησης εγγράφων - δολωμάτων. Ο αισθητήρας του συστήματος ελέγχει κάθε ανώμαλη συμπεριφορά στο αρχείο μητρώου (registry), σε αρχεία DLL, σε διεργασίες που εκτελούνται, και σε πρόσβαση του γραφικού περιβάλλοντος (GUI) [33].

Το 2010 οι Franqueira et al., διαχώρισαν τους εκ των έσω γνώστες εμπιστευτικών πληροφοριών σε δύο κατηγορίες: Στους εσωτερικούς - εταιρικούς και στους εξωτερικούς συνεργάτες π.χ., τους αναδόχους, τους επιχειρηματικούς εταίρους. Αυτοί εξετάστηκαν με διάφορες μεθόδους ανίχνευσης, συμπεριλαμβανομένου του λεπτομερούς ελέγχου (auditing), την ενημέρωση των δικαιωμάτων πρόσβασης, την

παρακολούθηση της συμπεριφοράς τους, τις πολιτικές ασφαλείας και τη διασφάλιση της ασφάλειας, αναφέροντας τις βασικές διαφορές εφαρμογής αυτών των μεθόδων μεταξύ των δύο αυτών κατηγοριών. Παρατηρήθηκε ότι οι συνεργάτες δεν υπόκεινται στα ίδια μέτρα ασφαλείας που χρησιμοποιούνται για την ανίχνευση εκ των έσω κακόβουλων. Ως εκ τούτου, η μετρίαση της απειλής κατά των εμπιστευτικών πληροφοριών από τους συνεργάτες θα πρέπει να ξεκινήσει με την αξιολόγηση της έκτασης των πιθανών απειλών και με τη χαρτογράφηση των δεδομένων που μοιράζονται όλοι οι συνεργάτες. Επιπλέον, είναι απαραίτητη η ύπαρξη συμβάσεων που προσδιορίζουν για παράδειγμα, τι επιτρέπεται και τι απαγορεύεται να κάνει ο κάθε συνεργάτης με τα στοιχεία που μοιράζονται οι συνεργάτες με τον οργανισμό – εταιρία.

Οι ακόλουθες ενότητες παρουσιάζουν τις ακαδημαϊκές μελέτες σχετικά με την ανίχνευση διαρροής δεδομένων και την πρόληψη. Οι μελέτες αυτές μπορούν να ομαδοποιηθούν στις ακόλουθες κατηγορίες.:

1. **Ανίχνευση κακής χρήσης σε Συστήματα Ανάκτησης Δεδομένων**. Επικεντρώνεται στην ανίχνευση εξουσιοδοτημένων χρηστών, οι οποίοι χρησιμοποιούν συστήματα ανάκτησης δεδομένων για να ανακτήσουν και να προβάλουν έγγραφα, τα οποία δεν πρέπει να αναγνωστούν από το χρήστη.

2. Ανίχνευση κακής χρήσης σε βάσεις δεδομένων.

Επικεντρώνεται στην ανίχνευση ανώμαλων μοτίβων πρόσβασης σε συστήματα βάσεων δεδομένων από μη εξουσιοδοτημένους χρήστες.

3. **Προστασία διαρροής ηλεκτρονικής αλληλογραφίας.**

Η τεχνολογία εξόρυξης δεδομένων έχει χρησιμοποιηθεί με επιτυχία για να παρέχει αυτόματα ανίχνευση κακής χρήσης υπογραφής και εντοπισμό της κακής χρήσης ανωμαλιών. Η εφαρμογή της τεχνολογίας αυτής μπορεί να περιλαμβάνει είτε την επιτήρηση του περιεχομένου της αλληλογραφίας είτε αναλύσεις συμπεριφοράς, όπως η ανίχνευση λογαριασμών ομάδων χρηστών που επικοινωνούν μεταξύ τους [34]. Οι δύο προσεγγίσεις μπορούν να αναμιχθούν για να προκύψει πιο αυστηρή επιβολή προστασίας.

4. Προστασία δικτύου.

Καθώς το Internet διευρύνεται και το εύρος ζώνης του δικτύου συνεχίζει να αυξάνεται, οι οργανισμοί αντιμετωπίζουν το πρόβλημα της απώλειας εμπιστευτικών πληροφοριών από το δίκτυο τους. Για να χειριστούν τον μεγάλο όγκο διαδικτυακής επικοινωνίας, οι ερευνητές έχουν προσπαθήσει να δημιουργήσουν συστήματα πρόληψης απώλειας πληροφοριών, τα οποία ελέγχουν την εξωτερική επικοινωνία για γνωστές εμπιστευτικές πληροφορίες. Τα συστήματα αυτά σταματούν τους αφελείς αντιπάλους από το να διαρρέουν δεδομένα, αλλά είναι πλήρως ανίκανα να αναγνωρίσουν διαρροές κωδικοποιημένης ή ασαφούς πληροφορίας [35].

5. Κρυπτογράφηση και έλεγχος πρόσβασης.

Η προστασία περιεχομένου απαιτεί τη διαβεβαίωση ότι το περιβάλλον στο οποίο το περιεχόμενο είναι προσβάσιμο, αποθηκεύεται και μεταφέρεται είναι ασφαλές και προστατευμένο. Το κύριο σημείο σε αυτό το σενάριο είναι να διασφαλιστεί ότι το περιεχόμενο μπορεί να προσπελαστεί μόνο από εξουσιοδοτημένες συσκευές και χρήστες [36].

6. Δεδομένα κρυμμένα σε αρχεία.

Η απόκρυψη ευαίσθητων δεδομένων σε ένα αρχείο δεν είναι επαρκής, διότι οι χρήστες μπορούν να χρησιμοποιήσουν την κοινή λογική (π.χ. Όλοι οι ασθενείς στο

ίδιο τμήμα έχουν την ίδια ασθένεια) για να συνάγουν περισσότερα δεδομένα, πράγμα το οποίο μπορεί να οδηγήσει σε διαρροή ευαίσθητων πληροφοριών [37]. Η ανάγκη απόκρυψης πληροφοριών από ένα ή περισσότερα έγγραφα είναι πολύ υψηλή σε έγγραφα XML που διαμοιράζονται στους εταίρους μιας εταιρίας μέσω διαδικτύου.

7. **Ανίχνευση κακόβουλης εσωτερικής απειλής μέσω ψηφιακών παγίδων (honeypots) και δολωμάτων(honeytokens).**

Στην προσέγγιση αυτή τα δολώματα ή τα honeyfiles τοποθετούνται σε εφαρμογές ή μηχανές του οργανισμού (π.χ. πλαστοί κωδικοί και ταυτότητες χρήστη ή ψευδή δεδομένα μισθών εργοδοτών). Οι πληροφορίες αυτές μπορούν στη συνέχεια να κατευθύνουν τον κάτοχο εμπιστευτικών πληροφοριών σε πιο εξελιγμένες παγίδες, όπου μπορούν να παρακολουθούνται οι κινήσεις του, για την αποκάλυψη των πραγματικών σκοπών.

6.1 Ανίχνευση κακής χρήσης σε Συστήματα Ανάκτησης Πληροφοριών (Information Retrieval – IR)

Η τυπική προσέγγιση που προτείνεται για την ανίχνευση διαρροών σε συστήματα ανάκτησης πληροφοριών είναι η ανίχνευση ανωμαλιών. Γενικά, το σύστημα μαθαίνει το κανονικό προφίλ συμπεριφοράς του χρήστη (ή μιας ομάδας χρηστών) και ανιχνεύει αποκλίσεις ή ανωμαλίες με αναφορά το προφίλ αυτό. Η ερώτηση κλειδί σε μια προσέγγιση ανίχνευσης ανωμαλιών είναι το πώς να μοντελοποιήσει κανείς τη συμπεριφορά του χρήστη. Οι Cathey et al. (2003) [38] πρότειναν τέσσερις μεθόδους μοντελοποίησης της συμπεριφοράς σε συστήματα ανάκτησης δεδομένων: α) την ομαδοποίηση εγγράφων, β) την ομαδοποίηση των αποτελεσμάτων αναζήτησης, γ) την ανάδραση συνάφειας και δ) τη μέθοδο της συγχώνευσης.

Η ομαδοποίηση των εγγράφων (Document clustering) λειτουργεί σε τρεις φάσεις. Πρώτα, όλα τα έγγραφα ενός συστήματος ανάκτησης πληροφοριών ομαδοποιούνται βάσει του περιεχομένου τους. Στη συνέχεια, κατά τη διάρκεια της φάσης κατάρτισης, το σύστημα ανίχνευσης κακής χρήσης χτίζει ένα προφίλ χρήστη με βάση την ομαδοποίηση εγγράφων που έχει ανακτηθεί από την αναζήτηση του χρήστη. Η υπόθεση στο μοντέλο αυτό είναι ότι ένας χρήστης που συνήθως αναζητά έγγραφα με μια δεδομένη ομαδοποίηση πρέπει να έχει την άδεια να ανακτήσει οποιοδήποτε άλλο έγγραφο στην ίδια ομάδα. Όμως, όταν ένας χρήστης προσπαθεί να έχει πρόσβαση σε μια ομάδα στην οποία συνήθως ο χρήστης δεν έχει κανονικά πρόσβαση και όταν η ομάδα αυτή δεν είναι παρόμοια με οποιαδήποτε από τις συστάδες «κανονικής πρόσβασης», ενεργοποιείται μια προειδοποίηση. Κατά συνέπεια, κατά τη φάση της ανίχνευσης, το σύστημα ανίχνευσης ανιχνεύει ένα περιστατικό κακής χρήσης συγκρίνοντας τα έγγραφα που πραγματικά ανακτήθηκαν με το προφίλ του χρήστη. Η σοβαρότητα της προειδοποίησης βασίζεται στην απόσταση μεταξύ του έγγραφου και της πλησιέστερης ομάδας του προφίλ.

Η ομαδοποίηση των εγγράφων της αναζήτησης (Clustering query results) είναι παρόμοια με την ομαδοποίηση των εγγράφων, μόνο που στη μέθοδο αυτή, τα έγγραφα που ομαδοποιούνται είναι μόνο όσα έχουν προσπελαστεί από το χρήστη (σε αντιδιαστολή με όλα τα υπόλοιπα έγγραφα του συστήματος ανάκτησης δεδομένων). Τα δεδομένα που έχει προσπελάσει ο χρήστης στο παρελθόν είναι μόνο ένα μικρό κλάσμα του πλήρους συνόλου δεδομένων, πράγμα το οποίο σημαίνει ότι

ομαδοποιούνται λιγότερα έγγραφα.

Στην προσέγγιση ανάδρασης συνάφειας (relevance feedback approach), η ακολουθία που έχει υποβληθεί αναλύεται και αναγνωρίζονται οι λέξεις κλειδιά (keysetQuery). Τα έγγραφα στα οποία υπάρχει συνήθως πρόσβαση από ένα χρήστη αναλύονται, εξάγονται από αυτά λέξεις κλειδιά καθώς και τυπικοί αλγόριθμοι ανάδρασης συνάφειας (keysetFeedback). Το προφίλ ενός χρήστη ορίζεται τέλος ως μια συνδυασμένη ομάδα λέξεων-κλειδιά τις οποίες ο χρήστης έχει αναζητήσει ή προσπελάσει (keysetQuery ∪ keysetFeedback). Όταν ο χρήστης υποβάλλει μια νέα αναζήτηση, εξάγονται οι λέξεις κλειδιά της αναζήτησης, και αν το πλήθος των λέξεων κλειδιά που δεν εμφανίζονται στο προφίλ χρήστη που έχει δημιουργηθεί υπερβαίνει ένα κατώφλι, ενεργοποιείται μια προειδοποίηση.

Τέλος, ***η μέθοδος της συγχώνευσης (fusion method)***, συνδυάζει απλώς όλες τις προηγούμενες μεθόδους σε ένα σταθμισμένο μέσο, ώστε να προσδιορίσει πότε να ενεργοποιηθεί η προειδοποίηση.

Πειραματικές αξιολογήσεις που έγιναν από τους Cathey et al. (2003) [38] έδειξαν ότι οι μέθοδοι της ανάδρασης συνάφειας και της συγχώνευσης παρείχαν τα καλύτερα συνολικά αποτελέσματα. Αναφέρθηκε επίσης μεγάλος βαθμός ακρίβειας (92%), όταν λαμβάνεται το πιο σημαντικό 10% των προειδοποιήσεων, αναφέροντάς τες ως εισβολές. Οι Ma and Gohrian (2005) [39, 40] έλεγξαν τη μέθοδο της ανάδρασης συνάφειας για ένα χρήστη ο οποίος υποβάλλει μια σύντομη (μέχρι τέσσερις όρους) ή μια μεγαλύτερη αναζήτηση (μέχρι 17 όρους). Τα αποτελέσματα της αξιολόγησης αυτής έχουν μια συνολική ακρίβεια του 83.9% και 82.2% για σύντομες και μεγαλύτερες αναζητήσεις, αντίστοιχα. Το ποσοστό της μη ανιχνεύσιμης κακής χρήσης για σύντομες αναζητήσεις ήταν λιγότερο από 2% και για μεγαλύτερες αναζητήσεις λιγότερο από 6%. Οι Gohrian et al. (2005) [40] επέκτειναν την εργασία του Ma (2005) [40] και αξιολόγησαν τους αλγόριθμους εκμάθησης μηχανής για την αυτόματη προσαρμογή των βαρών που αποδίδονται σε διάφορες συνιστώσες του προφίλ του χρήστη και στην αναζήτηση χρήση κατά τη διαδικασία ανίχνευσης (τα βάρη είχαν προηγουμένως προσαρμοσθεί χειροκίνητα). Τέσσερις ταξινομητές (SVM, ANN, BN, και C4.5) συγκρίθηκαν με την προσέγγιση χειροκίνητης προσαρμογής ως αναφορά. Αξιολογήθηκαν τρία σενάρια: (α) Σύντομες αναζητήσεις που χρησιμοποιούνται για την εξαγωγή των προφίλ και για την ανίχνευση, (β) Μεγαλύτερες αναζητήσεις, οι οποίες χρησιμοποιούνται για την εξαγωγή των προφίλ και για την ανίχνευση (γ) Μεγαλύτερες αναζητήσεις, οι οποίες χρησιμοποιούνται για την εξαγωγή των προφίλ και σύντομες αναζητήσεις για την ανίχνευση. Τα αποτελέσματα δείχνουν ότι για το καθένα από τα σενάρια που ελέγχτηκαν με ένα ή περισσότερους ταξινομητές, αποδίδουν εξίσου καλά ή καλύτερα από τη χειροκίνητη προσαρμογή.

Οι Mun et al. (2008) [41] πρότειναν τη χρήση ενός συστήματος IDS για την ανίχνευση εσωτερικών εισβολέων. Το προτεινόμενο σύστημα βασίζεται στην απόδοση βαθμών και επιπέδων δικαιωμάτων σε χρήστες και επίπεδα ασφαλείας έγγραφων, καθώς και στην παρακολούθηση της πρόσβασης του χρήστη στα έγγραφα.

6.2 Ανίχνευση κακής χρήσης σε Βάσεις Δεδομένων

Πολλές σχετικές μελέτες αναφέρονται στην ανίχνευση ανώμαλης πρόσβασης σε

βάσεις δεδομένων. Η σκοπός είναι η προστασία των δεδομένων των βάσεων δεδομένων σε διάφορα σενάρια. Παραδείγματα σεναρίων είναι η περίπτωση που ένας εσωτερικός υπάλληλος είναι εξουσιοδοτημένος να έχει πρόσβαση στη βάση δεδομένων (μέσω της διοίκησης ή άλλων τομέων), αλλά υποβάλλει περίεργες αναζητήσεις που δείχνουν πιθανή κακή χρήση δεδομένων (π.χ. συλλογή δεδομένων). Άλλη περίπτωση σεναρίου εξετάζει εσωτερικό υπάλληλο που κάνει κατάχρηση των νόμιμων προνομίων του ώστε να εμφανιστεί ως κάποιος άλλος χρήστης και να συλλέξει δεδομένα για κακόβουλους σκοπούς ή σε περίπτωση που κάποιος εξωτερικός εισβολέας προσπαθεί να εξάγει δεδομένα.

Οι περισσότερες από τις προτεινόμενες μεθόδους ανίχνευσης επικεντρώνονται αρχικά στην εκμάθηση του κανονικού μοτίβου αναζήτησης των χρηστών και αργότερα, κατά τη διάρκεια της φάσης παρακολούθησης, επικεντρώνονται στην αναγνώριση των περίεργων αναζητήσεων. Οι μελέτες διαφέρουν στο πώς αναπαρίστανται οι αναζητήσεις.

Δύο βασικοί τύποι χαρακτηριστικών χρησιμοποιούνται στη μοντελοποίηση μιας αναζήτησης: *με επίκεντρο τη σύνταξη και με επίκεντρο τα δεδομένα*. Η προσέγγιση με επίκεντρο τη σύνταξη, βασίζεται στη σύνταξη των εκφράσεων SQL της αναζήτησης ώστε να δημιουργηθεί το προφίλ του χρήστη. Η προσέγγιση με επίκεντρο τα δεδομένα επικεντρώνεται σε αυτό στο οποίο ο χρήστης προσπαθεί να έχει πρόσβαση αντί στο πώς εκφράζει το αίτημά του. Κατά συνέπεια, στην προσέγγιση με επίκεντρο τα δεδομένα, η αναζήτηση μοντελοποιείται εξάγοντας χαρακτηριστικά από το σύνολο των αποτελεσμάτων της αναζήτησης, όπως τον αριθμό των αναζητήσεων καθώς και την ελάχιστη, τη μέγιστη και τη μέση τιμή των χαρακτηριστικών των αναζητήσεων.

Και οι δύο προσεγγίσεις υποθέτουν μια αντιστοίχιση μεταξύ των χρηστών, των αναζητήσεων και των αποτελεσμάτων της αναζήτησης. Οι εφαρμογές όμως του διαδικτύου, στις οποίες γίνεται αυθεντικοποίηση σε δικές τους βάσεις δεδομένων χρηστών, δεν είναι μέρος των ανωτέρω προσεγγίσεων.

Οι Kamra et al. (2008) [42] αξιολόγησαν την προσέγγιση με επίκεντρο τη σύνταξη για την ανίχνευση κακής χρήσης δεδομένων σε συστήματα διαχείρισης βάσεων δεδομένων (database management systems - DBMS), οι οποίες διαχειρίζονται τις καταγραφές αναζητήσεων SQL για να δημιουργήσουν το προφίλ της «κανονικής» συμπεριφοράς πρόσβασης των χρηστών σε βάσεις δεδομένων. Αν κάθε χρήστης αντιστοιχισθεί σε ένα ρόλο, το σύστημα θα μάθει το κανονικό μοτίβο συμπεριφοράς κάθε ρόλου, και οι αποκλίσεις από τα μοτίβα αυτά θα ανιχνευθούν. Εναλλακτικά, οι χρήστες ομαδοποιούνται με βάσει το ρόλο τους. Κάθε ομάδα αντιμετωπίζεται στη συνέχεια ως ρόλος για την ανίχνευση ενός χρήστη, ο οποίος συμπεριφέρεται παρόμοια με εκείνους μιας άλλης ομάδας. Ένας αλγόριθμος ανίχνευσης ακραίων τιμών αναγνωρίζει τη συμπεριφορά του χρήστη, η οποία αποκλίνει από τα προφίλ.

Κάθε αναζήτηση αναπαρίσταται εξάγοντας χαρακτηριστικά από τη σύνταξη της αναζήτησης. Τα εξαγόμενα χαρακτηριστικά αναφέρονται στις εντολές SQL, τη σχετική πληροφορία στην οποία υπάρχει πρόσβαση και την πληροφορία χαρακτηριστικών στην οποία υπήρξε πρόσβαση Η πληροφορία μπορεί να εξαχθεί σε τρία επίπεδα διακριτότητας: (α) Το χονδρικό επίπεδο (coarse-grained) το οποίο είναι λιγότερο λεπτομερές, και περιέχει πληροφορίες όπως το πλήθος των διακριτών σχέσεων και χαρακτηριστικών, στα οποία υπάρχει πρόσβαση από την εντολή. (β) Το

ενδιάμεσου μεγέθους (medium-grained), το οποίο περιέχει περισσότερη λεπτομέρεια, όπως την καταμέτρηση του πλήθους των χαρακτηριστικών στα οποία υπήρξε πρόσβαση με την αναζήτηση και (γ) Το λεπτομερές (fine-grained), το οποίο καταγράφει λεπτομερώς πιο χαρακτηριστικό υφίσταται πρόσβαση σε ποια σχέση.

Η αξιολόγηση της προτεινόμενης μεθόδου διεξήχθη με χρήση του ταξινομητή Bayes και παρήχθη μια πραγματική βάση δεδομένων με 7583 εντολές SELECT, 213 εντολές INSERT και 572 εντολές UPDATE και ανώμαλες αναζητήσεις, οι οποίες εισήχθησαν στις ομάδες δεδομένων. Τα αποτελέσματα έδειξαν ένα λανθασμένο αρνητικό ποσοστό του 2.4% και ένα λανθασμένο θετικό ποσοστό του 17.9% , όταν χρησιμοποιήθηκαν τα λεπτομερή χαρακτηριστικά.

Η προτεινόμενη μέθοδος είναι απλή στην εφαρμογή και στην ενσωμάτωσή της σε συστήματα βάσεων δεδομένων. Είναι επίσης ελαφριά, γιατί τα δομικά χαρακτηριστικά μιας αναζήτησης μπορεί να εξαχθούν αρκετά γρήγορα. Όμως, η προσέγγιση με επίκεντρο τη σύνταξη είναι επιρρεπής σε λάθη. Για παράδειγμα, μπορεί να καταλήξει σε δύο αναζητήσεις οι οποίες διαφέρουν πολύ στη σύνταξη, αλλά παράγουν την ίδια «ομαλή» ή «ανώμαλη» απόφαση.

Οι Sunu et al. (2009) [43] πρότειναν μια μέθοδο δημιουργίας ενός στατιστικού προφίλ του μοτίβου πρόσβασης σε βάση δεδομένων του κανονικού χρήστη, ώστε να φανεί πότε ένας χρήστης αποκλίνει από τη ρουτίνα του. Οι συγγραφείς χρησιμοποίησαν την προσέγγιση βάσει δεδομένων και θεώρησαν τη σύνταξη της έκφρασης αναζήτησης μη σχετική για τη διάκριση του σκοπού του χρήστη, δίνοντας σημασία μόνο στα προκύπτοντα δεδομένα. Σύμφωνα με την προτεινόμενη προσέγγιση, υπολογίζεται ένα στατιστικό «άθροισμα» των αριθμών των αποτελεσμάτων της αναζήτησης. Στη συνέχεια, εξάγεται μια αναπαράσταση (στατιστική/αθροιστική) του διανύσματος S. Το διάνυσμα S περιέχει διάφορες στατιστικές παραμέτρους για κάθε χαρακτηριστικό της ομάδας αποτελεσμάτων, όπως το ελάχιστο, το μέγιστο και το μέσο ενός αριθμητικού χαρακτηριστικού ή το πλήθος των διακριτών τιμών ενός συμβολικού χαρακτηριστικού. Η προτεινόμενη μέθοδος αξιολογήθηκε σε ένα πραγματικό σύνολο δεδομένων, το οποίο λήφθηκε από μια εφαρμογή διαδικτύου με χρήση τριών διαφορετικών αλγορίθμων ταξινόμησης (naïve Bayes, δέντρα απόφασης και μηχανές διανυσμάτων υποστήριξης). Η προτεινόμενη προσέγγιση με επίκεντρο τα δεδομένα συγκρίνεται με την προσέγγιση με επίκεντρο τη σύνταξη του Kamra (2008) [42] με αναφερόμενα επίπεδα ακρίβειας που κυμαίνονται από 93% ως 96%.

Το μειονέκτημα της προσέγγισης επίκεντρο τα δεδομένα είναι ότι η αναζήτηση πρέπει να εκτελεστεί πριν ληφθεί η απόφαση. Επιπροσθέτως, τα δεδομένα θεωρούνται στατικά και αμετάβλητα. Αν τα δεδομένα μεταβάλλονται, απαιτείται επαναπροσδιορισμός.

Οι Fonseca et al. (2008) [44] πρότειναν τον ανιχνευτή κακόβουλης πρόσβασης δεδομένων (Malicious Data Access Detector MDAD), έναν μηχανισμό ανίχνευσης πρόσβασης κακόβουλης χρήσης, μέσω της ανάλυσης σε απευθείας σύνδεση, της διαδρομής ελέγχου των συστημάτων διαχείρισης βάσεων δεδομένων Ο προτεινόμενος μηχανισμός σκοπεύει στην προστασία των εφαρμογών βάσεων δεδομένων από επιθέσεις δεδομένων και των εφαρμογών διαδικτύου από εισβολές τύπου SQL-injection. Αυτό επιτυγχάνεται αναπαριστώντας το προφίλ των έγκυρων συναλλαγών από ένα κατευθυνόμενο γράφημα, το οποίο περιγράφει διαφορετικές ακολουθίες αναζήτησης SQL (SELECTs, INSERTs, UPDATEs, και DELETEs),

από την έναρξη της συναλλαγής μέχρι την εντολή δέσμευσης ή επαναφοράς. Οι κόμβοι του γραφήματος αναπαριστούν εντολές, ενώ τα τόξα αναπαριστούν έγκυρη εκτέλεση ακολουθιών. Στη φάση ανίχνευσης, ο μηχανισμός ανιχνεύει συναλλαγές που βρίσκονται εκτός του προφίλ εκμάθησης, ειδοποιείται ο διαχειριστής της βάσης και η κακόβουλη σύνοδος (session) τερματίζεται. Στα ανωτέρω θεωρήσαμε ότι η εφαρμογή που χρησιμοποιείται για να υποβάλει τις αναζητήσεις δεν αλλάζει. Για την αξιολόγηση του μηχανισμού ανίχνευσης και του αλγορίθμου εκμάθησης, χρησιμοποιήθηκε ως βασικό σημείο αναφοράς μετρήσεων η βάση δεδομένων αναφοράς TPC-C καθώς και η βάση δεδομένων σε παραγωγή. Η ακρίβεια δε που εμφανίστηκε, ήταν μεγαλύτερη από 99%.

Οι Yaseen and Panda (2009) [45] χρησιμοποίησαν επίσης την προσέγγιση με βάση τα δεδομένα και πρότειναν μια μέθοδο μοντελοποίησης της γνώσης που ένας κάτοχος εμπιστευτικών πληροφοριών μπορεί να συνάγει από ένα δεδομένο σύνολο εγγραφών. Με δεδομένο ότι ο κάτοχος εμπιστευτικών πληροφοριών έχει νόμιμη πρόσβαση σε πίνακες, χαρακτηριστικά και αρχεία, μπορεί να εφαρμόσει τη γνώση του για να δημιουργήσει νέα γνώση. Η μέθοδος χρησιμοποιεί γραφήματα εξάρτησης βασισμένα «στη γνώση του ειδικού στον τομέα» (domain-expert knowledge). Ο ειδικός του τομέα ορίζει: (1) Τη γνώση των εξαρτήσεων δεδομένων-αντικειμένων του οργανισμού, δηλαδή του τί επιπρόσθετη πληροφορία μπορεί να συναχθεί ή να υπολογισθεί από τα δεδομένα που εμφανίζονται στο χρήστη, (2) Τα δεδομένα που είναι περιορισμένα ή πολύ ευαίσθητα και (3) Τα άτομα εντός του οργανισμού που έχουν άδεια πρόσβασης στα δεδομένα αυτά. Τα γραφήματα αυτά προβλέπουν την ικανότητα του χρήστη να συνάγει ευαίσθητη πληροφορία, η χρήση της οποίας μπορεί να βλάψει την επιχείρηση. Η αποκτώμενη γνώση μπορεί να κατηγοριοποιηθεί σε τρεις τύπους: συναγόμενη, υπολογιστική και αθροιστική. Το αντικείμενο της βάσης δεδομένων (στο οποίο ο εσωτερικός υπάλληλος μπορεί να έχει ή να μην έχει πρόσβαση), η βασική γνώση και η αποκτώμενη γνώση μοντελοποιούνται ως ένα νευρωνικό γράφημα εξάρτησης και συναγωγής (neural dependency and inference graph – NDIG) , η έμπνευση για το οποίο έχει προέλθει από την ιδέα για τα τεχνητά νευρωνικά δίκτυα. Το NDIG παρουσιάζει τα αντικείμενα δεδομένων, τις εξαρτήσεις μεταξύ των αντικειμένων δεδομένων και την ποσότητα γνώσης την οποία μπορεί ένας εσωτερικός υπάλληλος να συνάγει με δεδομένη γνώση ως προς τα προηγούμενα αντικείμενα δεδομένων. Το μοντέλο NDIG μπορεί να βοηθήσει στην κατανόηση της ενδεχόμενης ζημιάς που μπορεί να προκαλέσει ένας εσωτερικός υπάλληλος, χρησιμοποιώντας τα νόμιμα δικαιώματα πρόσβασης και εφαρμόζοντας τη γνώση που κατέχει για το σύστημα και τον τομέα με μη εξουσιοδοτημένο τρόπο. Με χρήση του μοντέλου NDIG, το σύστημα μπορεί να εμποδίσει τους μη εξουσιοδοτημένους χρήστες από το να αποκτήσουν πληροφορία που τους επιτρέπει να συνάγουν ή να υπολογίσουν περιορισμένα δεδομένα, στα οποία δεν έχουν νόμιμη πρόσβαση.

Άλλες μελέτες αναφέρονται στην ανίχνευση εισβολής σε βάσεις δεδομένων, αντί να επικεντρώνονται στην απώλεια δεδομένων. Η έρευνα αυτή συνήθως γίνεται μέσω της προσέγγισης με επίκεντρο τη σύνταξη και στοχεύει στη διατήρηση της ακεραιότητας των δεδομένων, ανιχνεύοντας ανώμαλες συναλλαγές συμπεριλαμβανομένων των INSERTs και UPDATEs. Τα συστήματα αυτά δεν είναι τόσο κατάλληλα για την ανίχνευση κακής χρήσης ή απώλειας σε σχέση με όσα συζητήθηκαν παραπάνω, αλλά οι ιδέες που εφαρμόζονται στα συστήματα μπορεί να

είναι χρήσιμες για την εφαρμογή καινοτόμων λύσεων πρόληψης των απωλειών δεδομένων [46, 47, 48, 49, 50, 51, 52, 53, 54].

6.3 Προστασία διαρροής ηλεκτρονικής αλληλογραφίας.

Η έρευνα στον τομέα αυτό μπορεί να χωριστεί σε δύο βασικές κατηγορίες: *προσεγγίσεις βάσει περιεχομένου και προσεγγίσεις βάσει συμπεριφοράς*.

Η προσέγγιση βάσει περιεχομένου για την ανίχνευση και την πρόληψη διαρροής δεδομένων μπορεί να χωριστεί περαιτέρω σε:

- ***κανόνες βάσει λέξεων-κλειδιών***. Στην προσέγγιση αυτή, δημιουργούνται διάφοροι κανόνες από τις λέξεις κλειδιά που εμφανίζονται στο σώμα και την επικεφαλίδα ενός ηλεκτρονικού μηνύματος. Οι κανόνες αυτοί προσδιορίζουν το «επίπεδο εμπιστευτικότητας» του μηνύματος με βάση το πλήθος εμφάνισης συγκεκριμένων λέξεων κλειδιά [56, 57, 58].
- ***τεχνικές εκμάθησης μηχανής.*** Η βασική ιδέα της προσέγγισης αυτής είναι η χρήση τεχνικών εκμάθησης μηχανής όπως SVM ([56, 59] και naïve Bayes [60, 61, 62] για να προσδιοριστεί το «επίπεδο εμπιστευτικότητας» του ελεγμένου ηλεκτρονικού μηνύματος.

Δύο μέθοδοι χρησιμοποιούνται για την αναπαράσταση των δεδομένων κειμένου σε ηλεκτρονική αλληλογραφία. Η πρώτη μέθοδος είναι το μοντέλο διανυσματικού χώρου (vector space model). Τα διανύσματα αναπαριστούν έγγραφα και τα χαρακτηριστικά των διανυσμάτων αναπαριστούν όρους και τη συχνότητα εμφάνισής τους [63]. Τα διανύσματα χρησιμοποιούνται ως ομάδες εκμάθησης για την κατασκευή ενός πιθανολογικού μοντέλου, με βάση το οποίο λαμβάνονται αποφάσεις για το αν τα έγγραφα είναι εμπιστευτικά ή όχι.

Η δεύτερη μέθοδος για την αναπαράσταση των δεδομένων κειμένου είναι τα γραφήματα. Γενικά, οι λέξεις αναπαρίστανται στο γράφημα ως κόμβοι και συνδέονται με ακμές με λέξεις που εμφανίζονται στην κοντινή περιοχή. Ο Schenker (2003) [64] παρουσίασε έξι βασικές ομάδες αλγορίθμων δημιουργίας γραφημάτων από έγγραφα κειμένου: απλό, n-απλής απόστασης, τυπικό, n-απόστασης, σχετικής συχνότητας και απόλυτης συχνότητας. Οι αλγόριθμοι διαφέρουν στη χρήση των τεχνικών των όρων, για παράδειγμα στην αναπαράσταση της απόστασης μεταξύ δύο λέξεων (μέχρι μια προκαθορισμένη απόσταση) και στο αν οι συχνότητες; εμφάνισης δύο ή περισσότερων όρων μαζί πρέπει να υπολογιστούν. Οι μέθοδοι αναπαράστασης εγγράφων βάσει γραφημάτων μπορούν να συλλάβουν τη δομή του εγγράφου καθώς και το περιεχόμενό του, πράγμα το οποίο δε μπορεί να επιτευχθεί με τις μεθόδους αναπαράστασης βάσει διανυσμάτων.

Η προσέγγιση βάσει συμπεριφοράς επικεντρώνεται στα χαρακτηριστικά σχετικά με το περιβάλλον όπως η οργανωτική δομή και το ποιοι χρήστες στέλνουν ή λαμβάνουν ηλεκτρονικά μηνύματα. Για παράδειγμα, σύμφωνα με τον Kalyan (2007) [65] η πιθανοφάνεια ένα ηλεκτρονικό μήνυμα να έχει σταλεί κατά λάθος προσδιορίζεται με βάση την ανάλυση των περασμένων συνομιλιών αλληλογραφίας μεταξύ του αποστολέα και του παραλήπτη.

Στον Carvalho, (2007) [66], ένα αποστελλόμενο ηλεκτρονικό μήνυμα αναγνωρίζεται ως διαρροή, βασιζόμενοι στο περιεχόμενό του και την πιθανότητα του ότι ο αποδέκτης του μηνύματος θα το λάβει. Τα μηνύματα που αποστέλλονται σε προηγούμενους αποδέκτες μοντελοποιούνται ως ζευγάρια (μήνυμα, παραλήπτης). Ένα τέτοιο ζευγάρι θεωρείται μια πιθανή διαρροή, αν το μήνυμα είναι σημαντικά διαφορετικό από περασμένα μηνύματα που έχουν σταλθεί στον παραλήπτη. Για να βελτιωθεί η απόδοση, οι Carvalho και Cohen (2007) [56, 67] χρησιμοποίησαν διάφορα χαρακτηριστικά των κοινωνικών δικτύων.

Η λύση που πρότειναν χρησιμοποιεί δύο διαφορετικές τεχνικές ανίχνευσης. Η πρώτη τεχνική βασίζεται αυστηρά στο περιεχόμενο του κειμένου του μηνύματος. Μετρά την ομοιότητα μεταξύ δύο αναπαραστάσεων βασιζόμενες σε διανύσματα των ηλεκτρονικών μηνυμάτων. Το πρώτο διάνυσμα είναι μια αναπαράσταση TF-IDF όλων των προηγούμενων μηνυμάτων από τον τρέχοντα χρήστη u στον παραλήπτη r (ένα διαφορετικό διάνυσμα χρησιμοποιείται για κάθε παραλήπτη). Το δεύτερο διάνυσμα είναι μια αναπαράσταση TF-IDF του τρέχοντος μηνύματος που πρόκειται να σταλεί. Η απόσταση μεταξύ των δύο διανυσμάτων μετράται με χρήση ενός από τους προτεινόμενους αλγορίθμους: ομοιότητας συνημιτόνου ή των k κοντινότερων γειτόνων (k-nearest neighbors KNN). Αν η υπολογιζόμενη ομοιότητα είναι μικρότερη από ένα προκαθορισμένο κατώφλι, εμφανίζεται στο χρήστη που πρόκειται να στείλει το μήνυμα μια προειδοποίηση. Η σύγκριση αυτή γίνεται ξεχωριστά για κάθε παραλήπτη του μηνύματος που πρόκειται να σταλεί.

Η δεύτερη προτεινόμενη τεχνική ήταν μια μέθοδος βασιζόμενη στην κατηγοριοποίηση και εφαρμόσθηκε με χρήση της πληροφορίας των κοινωνικών δικτύων (όπως το πλήθος των ληφθέντων μηνυμάτων, το πλήθος των απεσταλμένων μηνυμάτων και το πλήθος των φορών που ένα συγκεκριμένο ζεύγος παραληπτών έχει αντιγραφεί στο ίδιο μήνυμα). Η ιδέα ήταν να γίνει πρόβλεψη της διαρροής σε δύο βήματα. Στο πρώτο βήμα, οι βαθμολογίες της ομοιότητας του κειμένου υπολογίστηκαν με χρήση της διαδικασίας διασταυρούμενης επικύρωσης. Στο δεύτερο βήμα, εξάγονται τα χαρακτηριστικά του δικτύου και στη συνέχεια υπολογίζεται μια συνάρτηση, η οποία συνδυάζει τα χαρακτηριστικά αυτά με τις βαθμολογίες κειμένου.

Για να ελεγχθεί η προτεινόμενη μέθοδος, οι διαρροές ηλεκτρονικής αλληλογραφίας προσομοιώθηκαν με βάση τα δεδομένα της ηλεκτρονικής αλληλογραφίας Enron (βάση δεδομένων με πάνω από 600.000 emails από 156 υπαλλήλους της εταιρίας Enron που κατέρρευσε) με χρήση διαφορετικών πιθανών κριτηρίων. Τα κριτήρια αυτά μιμούνται ρεαλιστικούς τύπους διαρροών, όπως ορθογραφικά λάθη σε ηλεκτρονικές διευθύνσεις, λάθη πληκτρολόγησης, παρόμοια μικρά ή κύρια ονόματα κλπ. Η μέθοδος αυτή ήταν σε θέση να ανιχνεύσει περίπου το 82% των περιπτώσεων ελέγχου ως διαρροές ηλεκτρονικής αλληλογραφίας. Το πλεονέκτημα της μεθόδου αυτής είναι ότι μπορεί να εφαρμοσθεί εύκολα σε ένα πελάτη ηλεκτρονικής αλληλογραφίας και δε χρησιμοποιεί πληροφορία που είναι διαθέσιμη μόνο στον εξυπηρετητή.

Σε πιο πρόσφατη μελέτη, οι Carvalho et al. (2009) [67] παρουσίασαν μια εφαρμογή της λύσης τους σε Mozilla Thunderbird. Επέκτειναν επίσης το σύστημά τους όχι μόνο για να ανιχνεύσουν ανεπιθύμητους παραλήπτες, αλλά και για να προτείνουν παραλήπτες τους οποίους ο χρήστης ενδεχομένως να έχει ξεχάσει να συμπεριλάβει. Η λύση αυτή απαιτεί την εγκατάσταση αρθρώματος

(plug-in) στην υπάρχουσα μηχανή Mozilla Thunderbird. Η αξιολόγηση της μελέτης αυτής δεν έδειξε υποσχόμενα αποτελέσματα. Μόνο το 15% των χρηστών ανέφερε ότι ο πελάτης πρόλαβε τα πραγματικά περιστατικά διαρροής ηλεκτρονικής αλληλογραφίας, και περισσότερο από το 47% δέχτηκαν συστάσεις από τεχνικές εξόρυξης δεδομένων. Από την άλλη πλευρά, περισσότερο από το 80% των υποκειμένων που συμμετείχαν στην περίπτωση ελέγχου ανέφεραν ότι θα χρησιμοποιούσαν τη λύση αυτή στην αλληλογραφία των πελατών σε εξελισσόμενη βάση, αν γίνονταν μερικές μεταβολές και βελτιώσεις.

6.4 Προστασία Δικτύου.

Οι Borders και Prakash (2008) [35] περιέγραψαν μια μέθοδο ποσοτικοποίησης των ενδεχόμενων διαρροών πληροφοριών δικτύου. Η προσέγγιση αυτή χρησιμοποιεί το γεγονός ότι ένα μεγάλο μέρος της κυκλοφορίας του δικτύου επαναλαμβάνεται ή περιορίζεται από προδιαγραφές πρωτοκόλλου. Αγνοώντας αυτά τα σταθερά δεδομένα, η πραγματική πληροφορία που οδηγείται από ένα πελάτη προς το διαδίκτυο μπορεί να απομονωθεί. Οι συγγραφείς επικεντρώθηκαν στο πρωτόκολλο μεταφοράς Hypertext (HTTP) και υπολόγισαν το περιεχόμενο των αναμενόμενων αιτημάτων HTTP με χρήση μόνο εξωτερικά διαθέσιμης πληροφορίας, συμπεριλαμβανομένων προηγούμενων αιτημάτων, προηγούμενων απαντήσεων του εξυπηρετητή και προδιαγραφών πρωτοκόλλου. Αυτό οδήγησε στη μέτρηση του πλήθους του μη επαναλαμβανόμενου και χωρίς περιορισμούς εύρους, το οποίο αναπαριστά το μέγιστο πλήθος πληροφοριών που θα μπορούσε να έχει υποστεί διαρροή από τον πελάτη. Αυτές οι τεχνικές ποσοτικοποίησης διαρροών αξιολογήθηκαν βάσει διαφόρων νόμιμων σεναρίων περιήγησης δικτύου. Τα αποτελέσματα της αξιολόγησης έδειξαν ότι ο νέος αλγόριθμος που παρήχθη απαιτεί μετρήσεις μεγέθους, οι οποίες ήταν 94% ως 99.7% μικρότερες από τις συνολικές τιμές της όλης πληροφορίας, πράγμα το οποίο έδειξε τη δυνατότητα αυτής της προσέγγισης να αποκλείει μεγάλο όγκο δεδομένων και να απομονώνει την πραγματική πληροφορία που θα έπρεπε να ελέγχετε, μειώνοντας κατά συνέπεια τον απαιτούμενο χρόνο ελέγχου της κίνησης του δικτύου. Η προσέγγιση όμως αυτή, σύμφωνα με τους συγγραφείς, δε μπορεί να αντιμετωπίσει κακόβουλες διαδικτυακές αιτήσεις από σελίδες με ενεργό κώδικα Javascript ή αντικείμενα Flash.

Οι Caputo et al. (2009) [68] παρουσίασαν το σύστημα Elicit το οποίο μπορεί να παρακολουθήσει την πρόσβαση των χρηστών σε πληροφορίες ενός εσωτερικού δικτύου. Το σύστημα χρησιμοποιεί δικτυακούς αισθητήρες, που επεξεργάζονται την κίνηση του δικτύου ώστε να παραγάγουν πληροφοριακές καταγραφές, όπως αναζήτηση, περιήγηση, ανάγνωση, διαγραφή και εκτύπωση. Τα συλλεγόμενα γεγονότα συνδυάζονται με συναφείς πληροφορίες και επεξεργάζονται από διάφορους στατιστικούς ανιχνευτές, και σύνολο κανόνων το οποίο μπορεί να σημάνει συναγερμό. Τέλος, οι συναγερμοί από τους δικτυακούς ανιχνευτές τροφοδοτούν ένα πιθανολογικό Bayesian δίκτυο, το οποίο παράγει την πιθανότητα για κακόβουλη δραστηριότητα του χρήστη.

6.5 Κρυπτογράφηση και έλεγχος πρόσβασης

Η κρυπτογράφηση και ο έλεγχος πρόσβασης είναι δύο από τους πιο κοινούς τρόπους πρόληψης διαρροής εμπιστευτικών δεδομένων μέσω περιορισμού πρόσβασης. Κάποια πλαίσια χρησιμοποιούν έλεγχο πρόσβασης και κρυπτογράφηση για να διασφαλιστούν ευαίσθητα δεδομένα σε ακινησία (π.χ. αποθηκευμένα σε φορητούς υπολογιστές, εξυπηρετητές , σταθερούς υπολογιστές κλπ), δεδομένα σε κίνηση (π.χ. μεταφερόμενα μέσω του τοπικού δικτύου ή του διαδικτύου) και δεδομένα σε χρήση (στα οποία υπάρχει πρόσβαση ή τροποποίηση).

Οι μηχανισμοί ελέγχου πρόσβασης μπορεί να μειώσουν το ρίσκο διαρροής δεδομένων, όμως, το ποσό της μείωσης περιορίζεται, διότι οι νόμιμοι χρήστες όπως οι εργοδότες και οι συνεργάτες συνεχίζουν να έχουν πρόσβαση σε ευαίσθητα δεδομένα.

Μια από τις ερωτήσεις κλειδιά στις λύσεις που παρέχουν κρυπτογράφηση δεδομένων ή και ολόκληρου του δίσκου είναι το πώς η κρυπτογράφηση θα επηρεάσει την ανάκτηση δεδομένων σε περιπτώσεις, όπου ο κωδικός πρόσβασης έχει ξεχασθεί ή σε περιπτώσεις έρευνας ή εγκληματολογίας [69].

Οι Abbadi and Alawneh (2008) [36] παρουσίασαν μια λύση για την πρόληψη διαρροής πληροφοριών όταν ο επιτιθέμενος είναι κάποιος που είναι εξουσιοδοτημένος να βλέπει τα δεδομένα. Γενικά, το προτεινόμενο πλαίσιο επιτρέπει σε εξουσιοδοτημένους χρήστες την πρόσβαση σε ευαίσθητα δεδομένα είτε από το εσωτερικό των εγκαταστάσεων μιας επιχείρησης είτε εξ' αποστάσεως μέσω VPN. Η βασική ιδέα είναι να επιτρέπεται η πρόσβαση σε ευαίσθητα δεδομένα μόνο από εξουσιοδοτημένες συσκευές και να προστατεύονται τα δεδομένα από μη εξουσιοδοτημένη αποκρυπτογράφηση. Αυτό επιτυγχάνεται με τη δημιουργία μιάς ομάδας συσκευών, από τις οποίες υπάρχει εξουσιοδοτημένη πρόσβαση στα δεδομένα. Κάθε ομάδα έχει το δικό του ελεγκτή, ο οποίος διαχειρίζεται την ασφάλεια αυθεντικοποίησης του διαχειριστή, την ασφαλή προσθήκη ή αφαίρεση συσκευών της συγκεκριμένης ομάδας και τη διανομή κλειδιών της ομάδας (key distribution - KD). Μόνο συσκευές εντός των εγκαταστάσεων της επιχείρησης είναι εξουσιοδοτημένες να συμμετέχουν στον ομάδα αυτή, διαφορετικά δε μπορούν να κατέχουν KD. Η συμμετέχουσα συσκευή πρέπει να είναι έμπιστη, δηλ. να αντιστοιχεί στην αναμενόμενη κατάσταση της συσκευής και πρέπει να προστεθεί φυσικώς από τον αυθεντικοποιημένο διαχειριστή ασφαλείας. Η μόνη οντότητα σε μια συσκευή που είναι εξουσιοδοτημένη να διαχειρίζεται κλειδιά κρυπτογράφησης είναι τα έμπιστα προγράμματα λογισμικού (trusted software agent), τα οποία θεωρείται ότι χρησιμοποιούν υλικό το οποίο παρέχει λειτουργίες κρυπτογράφησης.

Κατά τη μεταφορά ευαίσθητων δεδομένων μεταξύ των συσκευών της ομάδας, τα δεδομένα κρυπτογραφούνται με χρήση των κλειδιών της ομάδας. Λόγω του ότι το κλειδί μπορεί να μεταφερθεί μόνο από έμπιστο ελεγκτή σε μια εξουσιοδοτημένη συσκευή, αποθηκεύεται σε μια προστατευόμενη θέση αποθήκευσης και δε μπορεί να αντιγραφεί μεταξύ των συσκευών. Έτσι εξασφαλίζεται το ότι αν φθάσουν ευαίσθητα δεδομένα σε μια μη εξουσιοδοτημένη συσκευή, δε θα μπορέσουν να αποκρυπτογραφηθούν.

Ενώ τα ευαίσθητα δεδομένα βρίσκονται αποθηκευμένα σε μια συσκευή, κρυπτογραφούνται με χρήση ενός ειδικού κλειδιού συσκευής, το οποίο ονομάζεται KC (Ciphering Key). Το KC αποθηκεύεται σε μια προστατευμένη θέση

αποθήκευσης. Τα δεδομένα προτού μεταφερθούν, αποκρυπτογραφούνται με χρήση του KC και κρυπτογραφούνται ξανά με χρήση του KD.

Το πλαίσιο αυτό εμποδίζει τη μεταφορά μη προστατευμένων δεδομένων μέσω του διαδικτύου ή τη μαζική αποθήκευση δεδομένων (με την προϋπόθεση να προστατεύονται από τον έμπιστο λογισμικό). Εμποδίζει επίσης την πρόσβαση σε ευαίσθητα δεδομένα από μη εξουσιοδοτημένες συσκευές. Όμως, το προτεινόμενο πλαίσιο, δεν εμποδίζει έναν εξουσιοδοτημένο χρήστη να παρέχει το περιεχόμενο σε μια εξουσιοδοτημένη συσκευή, με τη φυσική παρουσία ενός μη εξουσιοδοτημένου χρήστη (υποθέτοντας ότι δεν υπάρχει επιτόπιος φυσικός έλεγχος). Δεν εμποδίζει επίσης έναν εξουσιοδοτημένο χρήστη να π.χ. διατηρήσει στη μνήμη του, να γράψει ή να φωτογραφίσει το περιεχόμενο και στη συνέχεια να το μεταφέρει σε άλλους.

Οι Alawneh και Abbadi (2008) [36] περιέγραψαν περαιτέρω την προστασία ευαίσθητων δεδομένων, τα οποία μοιράζονται μεταξύ συνεργαζόμενων οργανισμών. Στις περιπτώσεις αυτές, ένας οργανισμός ζητά ευαίσθητα δεδομένα από έναν άλλο οργανισμό, αλλά τα δεδομένα πρέπει να προστατευτούν από διαρροή προς μη εξουσιοδοτημένους χρήστες εντός ή εκτός του οργανισμού προορισμού. Η λύση που προτάθηκε βασίζεται στο «trusted computing», το οποίο παρέχει ιεράρχηση εμπιστοσύνης βασιζόμενη σε υλικό. Τα διαμοιραζόμενα δεδομένα αποστέλλονται στους συνεργαζόμενους οργανισμούς, με την εγκατάσταση VPN συνδέσεων. Οι ορισμοί των τομέων του οργανισμού προορισμού διασφαλίζουν ότι τα δεδομένα μπορούν να μοιρασθούν μεταξύ συσκευών των τομέων, ενώ ταυτόχρονα προστατεύονται από εξωτερική διαρροή. Η πλατφόρμα του trusted computing διασφαλίζει ότι τα δεδομένα παραμένουν κρυπτογραφημένα και ότι το κλειδί κρυπτογράφησης είναι προσβάσιμο μόνο από συσκευές του τομέα και δε μπορεί να μεταφερθεί σε συσκευές εκτός του οργανισμού ή της επιχείρησης. Ένας ειδικό πρόγραμμα (agent) εγκατεστημένο πάνω στη συσκευή θα αρνηθεί την απελευθέρωση ευαίσθητου περιεχομένου σε άλλες μη προστατευμένες συσκευές (ακόμα και αν είναι μέλη του συνολικού τομέα). Οι δυναμικοί τομείς χρησιμοποιούνται για να προσδιορίσουν υποομάδες συσκευών, οι οποίες θα πρέπει να είναι οι μόνες που μοιράζονται περιεχόμενο με χρήση του συγκεκριμένου κλειδιού του τομέα.

Ως μετασχηματισμό του υφιστάμενου συνδυασμού των λογισμικών ανοικτού κώδικα LAMP (Linux Apache MySQL PHP, Perl ή Python), οι Parno et al. (2009) [70] παρουσίασαν το CLAMP (Confidential Linux Apache MySQL PHP), το οποίο οδηγεί σε εφαρμογές πιο ασφαλείς έναντι των διαρροών. Ο μετασχηματισμός βασίζεται στην εξαγωγή της διαδικασίας αυθεντικοποίησης εκτός των ορίων της εφαρμογής, σε μία ξεχωριστή μονάδα αυθεντικοποιησης του χρήστη (User Authenticator - UA). Επίσης, κάθε χρήστης που συνδέεται με τον εξυπηρετητή θα παίρνει ένα καθαρό αντίγραφο του εξυπηρετητή (το οποίο ονομάζεται WebStack), το οποίο δημιουργείται από ένα προστατευμένο αμετάβλητο αντίγραφο. Το νέο WebStack λειτουργεί σε μια ξεχωριστή θέση εικονικής μνήμης, η οποία παρέχει πλήρη απομόνωση μεταξύ των εξυπηρετητών που εξυπηρετούν κάθε χρήστη, και το οποίο με τη σειρά του σημαίνει ότι μια ζημιά στο αντίγραφο του εξυπηρετητή ενός χρήστη, δεν θα επηρεάσει τους υπόλοιπους εξυπηρετητές και χρήστες. Η ταυτότητα (ID) του WebStack και ο μοναδικός UA ο οποίος είναι προσαρμοσμένος σε αυτό, χρησιμοποιούνται από τον περιοριστή αναζήτησης (query restrictor QR). Το QR είναι ένας διαμεσολαβητής (proxy) της βάση δεδομένων, ο οποίος δημιουργεί μια εικονική βάση δεδομένων, στην οποία χρησιμοποιεί μόνο τα δεδομένα τα οποία

επιτρέπεται να δει ο χρήστης και περιορίζει τις λειτουργίες SELECT, INSERT, και UPDATE, σύμφωνα με την προκαθορισμένη πολιτική. Είναι σχετικά εύκολο να τροποποιηθεί μια υπάρχουσα εφαρμογή διαδικτύου ώστε να λειτουργεί με το CLAMP. Από την άλλη πλευρά, η μέθοδος χρησιμοποιεί μεγάλο τμήμα της μνήμης και της CPU για τον εξυπηρετητή και δε μπορεί να εμποδίσει την εκ των έσω κακόβουλη επίθεση.

Οι Yasuhiro and Yoshik (2002) [71] παρουσίασαν ένα διαδικτυακό πλαίσιο με σκοπό την πρόληψη διαρροής εμπιστευτικής πληροφορίας. Αυτό επιτυγχάνεται με την κρυπτογράφηση των εμπιστευτικών δεδομένων και με την παροχή πρόσβασης μόνο σε εξουσιοδοτημένους χρήστες, καθώς και με τη χρήση εξειδικευμένου αναγνώστη ενσωματωμένου στον περιηγητή του διαδικτύου για την αποκρυπτογράφηση και την προβολή του περιεχομένου. Το σύστημα λειτουργεί σε δύο φάσεις: τη φάση της λήψης και τη φάση της προβολής. Η φάση της λήψης βασίζεται σε ένα έξυπνο διαμεσολαβητή, ο οποίος χρησιμοποιεί μια βάση δεδομένων εξουσιοδότησης για να προσδιορίσει αν ο τρέχων χρήστης μπορεί να λάβει το ζητούμενο περιεχόμενο και αν το περιεχόμενο πρέπει να κρυπτογραφηθεί πριν σταλεί. Στη φάση της προβολής, ένας αναγνώστης στον υπολογιστή του χρήστη παρουσιάζει το περιεχόμενο στο χρήστη, επιτρέποντας στο χρήστη να δει το περιεχόμενο μόνο μία φορά ανά λήψη του κλειδιού, και ο αναγνώστης έχει τη δυνατότητα να απενεργοποιήσει τις λειτουργίες αποθήκευσης, εκτύπωσης και εκτύπωσης οθόνης. Το προτεινόμενο πλαίσιο είναι διαφανές στο χρήστη και προστατεύει τα εμπιστευτικά δεδομένα ενώ βρίσκονται σε ηρεμία (κρυπτογραφημένα στη βάση δεδομένων), σε κίνηση (στέλνονται κρυπτογραφημένα στο δίκτυο) και σε χρήση (ο χρήστης μπορεί να παρακολουθεί το περιεχόμενο σε έναν εξειδικευμένο αναγνώστη που απαγορεύει την εκτύπωση ή την αποθήκευση). Όμως, η λειτουργία εκτύπωσης οθόνης δεν είναι πλήρως μπλοκαρισμένη και μπορεί να παρακαμφθεί.

Η έννοια του λεπτομερούς ελέγχου πρόσβασης για συστήματα βάσεων δεδομένων προτάθηκε αρχικά για να παρέχει καλύτερη προστασία δεδομένων, ελέγχοντας την πρόσβαση στο επίπεδο διακριτότητας των ανεξάρτητων γραμμών και στηλών[72]. Ο λεπτομερής έλεγχος πρόσβασης στο επίπεδο της βάσης δεδομένων, σε αντιδιαστολή με την εφαρμογή στο επίπεδο του προγράμματος, μπορεί να γίνει τροποποιώντας την αναζήτηση και προσαρτώντας κατηγορήματα στους όρους WHERE της αναζήτησης ή τροποποιώντας τον αρχικό προσβάσιμο πίνακα, εισάγοντας μία δυναμικά δημιουργημένη προσωρινή όψη μεταξύ της αναζήτησης και του πίνακα στόχου.

Οι De Capitani Di Vimercati et al. (2010) [73] πρότειναν την έννοια της επιλεκτικής κρυπτογράφησης ώστε να παρέχεται επιλεκτικός έλεγχος πρόσβασης σε εξωτερικά ευαίσθητα δεδομένα από τρίτους συνεργάτες. Σύμφωνα με την προτεινόμενη προσέγγιση, η πολιτική εξουσιοδότησης πρόσβασης δεδομένων υπολογίζει μια ιεραρχική δομή αδειοπλαισίων (tokens), τα οποία χρησιμοποιούνται για να εξάγουν μια ομάδα κρυπτογραφημένων κλειδιών. Αυτή η ομάδα κρυπτογραφημένων κλειδιών, αναφέρεται και ως πολιτική κρυπτογράφησης και επιτρέπει την επιλεκτική κρυπτογράφηση και την πρόσβαση στα δεδομένα. Οι συγγραφείς απέδειξαν ότι το πρόβλημα του υπολογισμού μιας ελάχιστης πολιτικής κρυπτογράφησης είναι NP-hard και παρουσίασαν έναν ευρετικό αλγόριθμο για την επίλυση του προβλήματος.

6.6 Δεδομένα κρυμμένα σε αρχεία

Τα έγγραφα που γράφονται και αποθηκεύονται σε μορφή εγγράφων π.χ. Microsoft Word μπορεί να περιέχουν κρυμμένα δεδομένα. Όμως, η επίγνωση του προβλήματος αυτού δεν είναι αρκετά διαδεδομένη, ειδικά μεταξύ μη εξειδικευμένων χρηστών των υπολογιστών [74]. Παραδείγματα κρυμμένων δεδομένων σε έγγραφα Word περιλαμβάνουν όνομα, όνομα χρήστη των συντακτών του εγγράφου και των συνεργατών τους και πληροφορίες της επιχείρησης των εμπλεκόμενων χρηστών.

Οι Yixiang et al. (2007) [37] αναφέρουν ότι η δημοσιοποίηση δεδομένων ενός εγγράφου XML με απαιτήσεις ασφαλείας θέτει πολλαπλές προκλήσεις, όταν οι χρήστες μπορούν να συμπεράνουν δεδομένα με χρήση της κοινής λογικής. Επιπλέον, όταν εμπλέκονται δύο ή περισσότερα έγγραφα, οι χρήστες μπορούν να συνάγουν ευαίσθητα δεδομένα, συνδυάζοντας τα έγγραφα. Ο πυρήνας του αλγορίθμου εξάλειψης εσωτερικών κόμβων (Eliminate Inner Nodes algorithm), ο οποίος χρησιμοποιείται όταν δημοσιοποιούνται αρκετά έγγραφα XML, είναι η εύρεση του μέγιστου μερικού εγγράφου, στο οποίο αποφεύγεται η διαρροή δεδομένων ενώ ταυτόχρονα επιτρέπει τη δημοσιοποίηση όσο το δυνατόν περισσότερων δεδομένων.

7. Περιστατικά Απώλειας Δεδομένων

7.1 Κατηγοριοποίηση περιστατικών απώλειας δεδομένων

Τα περιστατικά απώλειας δεδομένων μπορούν να χαρακτηριστούν με βάση τα ακόλουθα χαρακτηριστικά: Που συνέβη η απώλεια, ποιος την προκάλεσε, τι δεδομένα χάθηκαν, πως αποκτήθηκε πρόσβαση στα δεδομένα, και τελικά πως διέρρευσαν τα δεδομένα. Αυτοί οι παράμετροι επηρεάζουν τη λήψη αποφάσεων για τα μέτρα άμυνας που θα παρθούν έναντι της απώλειας δεδομένων.

Υπάρχουν τρείς πιθανές τοποθεσίες απώλειας δεομένων: (α) Εντός του οργανισμού, όπου δεδομένα απωλέσθησαν από μια πηγή εντός της φυσικής περιμέτρου του οργανισμού. (β) Εκτός του οργανισμού, όπου τα δεδομένα απωλέσθησαν από μια εξωτερική πηγή εκτός της περιμέτρου του οργανισμού π.χ. κλοπή φορητού υπολογιστή υπαλλήλου από το αυτοκίνητό του και (γ) Τοποθεσία τρίτων, όπου τα δεδομένα απωλέσθησαν από μια εμπιστευτική τοποθεσία τρίτου π.χ. παραβιάστηκε το δίκτυο συνεργάτη και τα πιστοποιητικά του χρησιμοποιήθηκαν για την απόκτηση πρόσβασης σε δεδομένα

Τα περιστατικά απώλειας δεδομένων μπορεί να προέλθουν από μια ή περισσότερες από τις ακόλουθες πηγές: Ένα πρόσωπο που κατέχει εμπιστευτική θέση, ένα πρόσωπο εκτός του οργανισμού, έναν προμηθευτή, έναν καταναλωτή ή πελάτη. Οι Franqueira et al. (2010) [75] κατηγοριοποίησαν της πηγές απώλειας ως άτομα εντός του οργανισμού, εκτός του οργανισμού και εξωτερικούς εμπιστευτικούς (εξωτερικοί συνεργάτες, εργολάβοι, προμηθευτές, πελάτες κ.α.) και μελέτησε τις προκλήσεις που προκύπτουν στον προσδιορισμό των εσωτερικών και των εξωτερικών συνεργατών του εκάστοτε οργανισμού.

Οι περισσότεροι εσωτερικοί χρήστες κάθε οργανισμού είναι αξιόπιστοι έως ένα βαθμό, και μερικοί άλλοι κατέχουν υψηλά δικαιώματα πρόσβασης (διευθυντές). Οι *εσωτερικοί χρήστες των επιχειρήσεων* μπορούν να διαφοροποιηθούν περαιτέρω με βάση τη φύση της πράξης που οδήγησε σε απώλεια δεδομένων, αν ήταν ακούσια ή εκούσια. Οι *εξωτερικοί κακόβουλοι χρήστες* (hackers, οργανωμένες εγκληματικές ομάδες, και κυβερνητικοί φορείς), είναι συνήθως αναξιόπιστα πρόσωπα, έτσι δεν τους παρέχονται δικαιώματα πρόσβασης. Εξωτερικοί συνεργάτες, όπως εργολάβοι και προμηθευτές έχουν εμπορικές σχέσεις με την εκάστοτε επιχείρηση και είναι γνωστοί, ως επέκταση της επιχείρησης. Η ανταλλαγή πληροφοριών είναι σημαντικό κομμάτι της επεκταθείσας επιχείρησης, και για αυτό συνήθως υπάρχει ένα επίπεδο εμπιστοσύνης μεταξύ των συνεργατών της επιχείρησης. Αυτές οι σχέσεις διευκολύνονται χρησιμοποιώντας τεχνολογίες όπως VPN, και κρυπτογράφηση. Οι καταναλωτές και οι πελάτες συχνά κατέχουν συγκεκριμένα δικαιώματα χρησιμοποίησης εφαρμογών ή υπηρεσιών.

Οι προμηθευτές λύσεων DLP και οι ακαδημαϊκοί μελετητές διακρίνουν τρείς φάσεις δεδομένων κατά τον κύκλο ζωής τους: δεδομένα σε ακινησία (DAR), δεδομένα σε κίνηση (DIM), και δεδομένα σε χρήση (DIU). Έτσι έχουν υιοθετηθεί

διάφορες προσεγγίσεις για την προστασία δεδομένων κατά τον κύκλο ζωής τους. Συνεπώς τα περιστατικά πρέπει να κατηγοριοποιούνται ανάλογα την κατάσταση των δεδομένων όταν χάθηκε ο έλεγχος σε αυτά. Για το λόγο αυτό κάθε περιστατικό μπορεί να κατηγοριοποιηθεί σε περισσότερες από μια καταστάσεις δεδομένων.

Το ερώτημα του «πως αποκτήθηκε πρόσβαση στα δεδομένα» εκτείνεται στο «ποιος προκάλεσε την απώλεια». Τα ερωτήματα αυτά δεν είναι εναλλάξιμα, αλλά συμπληρωματικά και συνεπώς οι διάφοροι τρόποι για την απόκτηση πρόσβασης σε ευαίσθητα δεδομένα μπορούν να ομαδοποιηθούν στις ακόλουθες κατηγορίες.

- *Ηλεκτρονική πειρατεία (Hacking)*: αυτός ο όρος περιλαμβάνει (α)Εκμετάλλευση κοινόχρηστων ή προεπιλεγμένων πιστοποιητικών π.χ. ο διαχειριστής της βάσης δεδομένων, η οποία δημιουργήθηκε από προεπιλογή δεν αλλάχθηκε ποτέ μετά την εγκατάσταση. (β) Εκμετάλλευση εσφαλμένων μηχανισμών ελέγχου πρόσβασης ή κερκόπορτων του συστήματος ώστε να παρακάμψει την ταυτοποίηση και να αποκτήσει πρόσβαση σε ευαίσθητα δεδομένα. (γ) Χρησιμοποιώντας κλεμμένα νόμιμα πιστοποιητικά (δ)Απόκτηση πρόσβασης σε ευαίσθητα δεδομένα ή πιστοποιητικά χρησιμοποιώντας SQL injection, προγραμματισμό (XSS), κλεμμένες συνεδριάσεις μεταβλητών και ρυθμιστικές επιθέσεις υπερχείλισης.
- *Κακόβουλο λογισμικό*: Μπορεί δυνητικά να οδηγήσει σε εκούσια ή άδολη απώλεια δεδομένων, για παράδειγμα με καταγραφή των πληκτρολογήσεων οι οποίες περιλαμβάνουν κωδικούς χρήστη και κωδικούς πρόσβασης, άνοιγμα κερκόπορτας σε έναν εισβολέα, ή απλά αποστολή προσβάσιμων δεδομένων στο δίκτυο. Κακόβουλο λογισμικό μπορεί να εγκατασταθεί εντός του οργανισμού από έναν εισβολέα, από έναν υπάλληλο ο οποίο περιηγείται σε επισφαλείς ιστότοπους στο διαδίκτυο, ή μέσω εκτέλεσης ενός αρχείου κακόβουλου λογισμικού όπως ένα συνημμένο ηλεκτρονικής αλληλογραφίας, απελευθερώνοντας ένα σκουλήκι το οποίο μεταδίδεται, εκμεταλλευόμενο τις αδυναμίες του συστήματος ή μέσω φυσικής διάδοσης χρησιμοποιώντας αφαιρούμενα αποθηκευτικά μέσα [76].
- *Κοινωνικές επιθέσεις*: εμφανίζονται υπό μορφή παρατήρησης ή περιήγησης, επίθεσης ή απειλής βλάβης και κοινωνικής μηχανικής.
- *Φυσικής πρόσβασης* στο μηχάνημα (ή στο μέσο) όπου βρίσκονται τα ευαίσθητα δεδομένα, παρακάμπτοντας οποιοδήποτε μηχανισμό προστασίας. Αυτό περιλαμβάνει κλοπή ή απώλεια του αντικειμένου π.χ. φορητός υπολογιστής, αλληλεπίδραση με το σύστημα μέσω πληκτρολογίου, και υποκλοπή (παρακολούθηση καλωδίου δικτύου, μετάδοση Wi-Fi, γραμμή τηλεφώνου ή οποιοδήποτε πρωτόκολλο μετάδοσης ευαίσθητων δεδομένων).
- *Ανθρώπινα σφάλματα*: περιλαμβάνουν σφάλματα προγραμματιστών, επαγγελματίες, ή κατόχους δεδομένων, όπως εσφαλμένες διαμορφώσεις, σφάλματα προγραμματισμού, αποθήκευση ευαίσθητων πληροφοριών σε εξυπηρετητή που εκτίθεται στο διαδίκτυο, και ακατάλληλη απόρριψη σημαντικών αρχείων ή ηλεκτρονικών μέσων όπως CD/DVD.

Η ταξινόμηση ανά απώλεια καναλιού είναι σημαντική για τον προσδιορισμό του

τρόπου πρόληψης των περιστατικών στο μέλλον.

- *Φυσική απώλεια καναλιού*: φυσικά μέσα (σκληροί δίσκοι, φορητοί υπολογιστές, σταθμοί εργασίας, CD/DVD, συσκευές USB) τα οποία περιλαμβάνουν ευαίσθητες πληροφορίες ή το ίδιο το αρχείο μετακινήθηκαν εκτός του οργανισμού. Αυτό όλο και συχνότερα σημαίνει ότι ο έλεγχος επί των δεδομένων είχε χαθεί πριν ακόμα βρεθούν εκτός του οργανισμού.
- *Λογική απώλεια καναλιού*: αναφέρεται σε σενάρια στα οποία δεδομένα απωλέσθησαν υπό μορφή ψηφιακών πληροφοριών, εκπομπών, μεταφόρτωσης, ή αποστολής εκτός του οργανισμού χρησιμοποιώντας εφαρμογές και δίκτυα υπολογιστών. Επίσης αυτό περιλαμβάνει μεταφορτώσεις διαδικτύου (πληροφορίες μεταφορτώθηκαν σε μια απομακρυσμένη τοποθεσία, όπως ένας εξυπηρετητής αρχείων, ιστότοπος Διαδικτύου, εξυπηρετητής ηλεκτρονικής αλληλογραφίας), κατάχρηση εφαρμογών, αποθήκευση σε εκτεθειμένες τοποθεσίες στο Διαδίκτυο, άμεση αλληλογραφία (Skype, ICQ, MS Messenger) εφαρμογές τρίτων (P2P), και κακόβουλο λογισμικό (άγνωστα πρωτόκολλα).

7.2 Περιγραφή βασικών περιστατικών απώλειας δεδομένων

Πίνακας 7.1: Κύρια σενάρια απώλειας δεδομένων

Α/Α	**Σενάριο**
1	Κλοπή ή απώλεια συσκευής μαζικής αποθήκευσης
2	Συνεργάτης ή τρίτη εταιρεία διαρρέει ευαίσθητα δεδομένα
3	Παράνομη αποθήκευση δεδομένων σε άλλα συστήματα/συσκευές/μέσα
4	Κλοπή ταυτότητας/διαμοιρασμός διαπιστευτηρίων
5	Κατάχρηση δικαιωμάτων
6	Δημοσιοποίηση δεδομένων
7	Διαρροή ηλεκτρονικής αλληλογραφίας
8	Hacker αποκτά πρόσβαση σε ευαίσθητα δεδομένα
9	Κακόβουλο λογισμικό υποκλέπτει δεδομένα
10	Κρυμμένα δεδομένα μέσα σε αρχεία
11	Παράνομη μεταφορά ευαίσθητων δεδομένων από ένα ελεγχόμενο σύστημα

Στον παραπάνω Πίνακα 7.1. παρουσιάζεται μια λίστα των κύριων σεναρίων απώλειας δεδομένων ονομαστικά. Παρακάτω θα ακολουθήσουν οι περιγραφές των σεναρίων. Κάθε σενάριο μπορεί να περιλαμβάνει είτε εκούσια είτε ακούσια περιστατικά. Για παράδειγμα, μια ακούσια περίπτωση είναι μπορεί να περιλαμβάνει κλοπή ή απώλεια κάποιας φυσικής συσκευής ή ακατάλληλη απόρριψή της (όπως απόρριψη σκληρού δίσκου με ευαίσθητα δεδομένα σε κάδο απορριμμάτων), ενώ μια εκούσια περίπτωση μπορεί να περιλαμβάνει κλοπή από υπάλληλο σκληρού δίσκου της εταιρείας με ευαίσθητα δεδομένα. Το τελικό αποτέλεσμα είναι το ίδιο και στις δύο περιπτώσεις: ο σκληρός δίσκος παραβιάζεται.

Παρόλα αυτά, όταν εκτιμώνται πιθανές λύσεις σε αυτά τα σενάρια, βρέθηκε ότι συνήθως οι λύσεις μπορεί να διαφέρουν σημαντικά μεταξύ δυο περιπτώσεων, ενώ και η εκούσια περίπτωση είναι συχνά δυσκολότερο να περιοριστεί. Ο Πίνακας 7.2

παρέχει μια συνοπτική επισκόπηση διαφόρων λύσεων πρόληψης απώλειας σε σχέση με τα σενάρια που παρουσιάστηκαν στον Πίνακα 7.1. Για κάθε σενάριο (που εμφανίζεται στον κάθετο άξονα), εκτιμάται ποσοτικά η αποτελεσματικότητά των διάφορων λύσεων (οι οποίες εμφανίζονται στον οριζόντιο άξονα), για τις εκούσιες (πρώτος αριθμός) και ακούσιες περιπτώσεις (δεύτερος αριθμός). Για παράδειγμα, το 4 σημαίνει άριστη λύση ενώ το 1 σημαίνει μία αναποτελεσματική λύση. Έτσι λοιπόν, παρακάτω παρουσιάζονται όλα τα σενάρια για τα διάφορα μέσα καθώς και η αντιμετώπιση τους.

Πίνακας 7.2: Χαρτογράφηση αποτελεσματικότητας εφαρμόσιμων λύσεων πρόληψης απώλειας δεδομένων για διάφορα σενάρια κατάχρησης/απώλειας

	Έλεγχος Συσκευών	Κρυπτο-γράφηση	DIM	DIU	DAR	RMS based access control	Activity based auth and verification	two factor auth	Ανίχνευση ανωμαλίας	Διαρροή email	Ψηφιακές παγίδες
1 Κλοπή	**1/2**	**2/4**		**1/2**	**1/2**						
2 Διαρροή μέσω συνεταίρου						**2/3**					**3/3**
3 Παράνομη αποθήκευση	**1/2**		**1/**	**1/3**		**2/3**					
4 Κλοπή ταυτότητας	**1/2**		**1/2**	**1/3**			**3/3**	**3/4**	**3/3**		
5 Κατάχρηση δικαιωμάτων	**1/0**		**1/0**	**1/0**					**3/0**		**3/0**
6 Δημοσιο-ποίηση		**2/3**	**2/2**	**1/3**	**1/3**	**2/3**					
7 email			**2/3**	**2/3**		**3/3**				**3/3**	
8 Hackers		**3/0**	**2/0**								**3/0**
9 Malware		**3/0**	**2/0**								
10 Κρυφά δεδομένα			**3/3**	**3/3**	**1/3**						
11 Παράνομη μεταφορά				**2/3**	**2/3**						

Σενάριο 1 (Κλοπή ή απώλεια συσκευής μαζικής αποθήκευσης): Ευαίσθητες πληροφορίες εταιρείας αποθηκευμένες σε μια συσκευή μαζικής αποθήκευσης που χάθηκε ή κλάπηκε ή έγινε μη σωστή απόρρηψή της, έχοντας ως αποτέλεσμα την έκθεση των ευαίσθητων πληροφοριών της σε μη εξουσιοδοτημένα πρόσωπα. Παραδείγματα τέτοιων μέσων περιλαμβάνουν: επιτραπέζιους υπολογιστές, φορητούς υπολογιστές, σκληροί δίσκοι, CD / DVD, δημιουργία αντιγράφων ασφαλείας σε ταινίες και άλλα ηλεκτρονικά μέσα. Αυτό το σενάριο μπορεί να προκληθεί από την εκ των έσω, από άτομα με νόμιμη φυσική πρόσβαση σε αυτά τα μέσα, οι οποίοι μπορούσαν να κλέψουν ή να αντιγράψουν τα ευαίσθητα δεδομένα που είναι αποθηκευμένα σε αυτά, ή από μια εξωτερικό εισβολέα ο οποίος έκλεψε τη συσκευή ενώ δεν είναι προσωρινά στις εγκαταστάσεις της εταιρίας.

Για το σενάριο 1, είναι δύσκολο να προληφθεί η αποθήκευση ευαίσθητων δεδομένων στη συσκευή, αλλά είναι εύκολη η κρυπτογράφηση ολόκληρου του δίσκου, με αποτέλεσμα η κρυπτογράφηση να παρέχει σχεδόν ολοκληρωτική προστασία. Παρόλα αυτά, η κρυπτογράφηση από μόνη της δεν είναι επαρκής και πρέπει να χρησιμοποιείται σε συνδυασμό με έλεγχο της συσκευής (για να διασφαλίζεται ότι τα δεδομένα μεταφέρονται μόνο σε κρυπτογραφημένες συσκευές) και με άλλα καθορισμένα προϊόντα DLP, τα οποία παρακολουθούν δεδομένα DIU και DAR. Επιπροσθέτως, η ικανότητα απόδειξης ότι τα ευαίσθητα δεδομένα ήταν

κρυπτογραφημένα και κατ' επέκταση προστατευμένα, σε μια κλεμμένη ή απολεσθείσα συσκευή, είναι σημαντική για τον οργανισμό από την άποψη της κανονιστικής συμμόρφωσης.

Σενάριο 2 (Συνεργάτης ή τρίτη εταιρεία διαρρέει ευαίσθητα δεδομένα): Ένας συνεργάτης πωλεί τα ευαίσθητα δεδομένα (π.χ., αριθμοί τηλεφώνου, ονόματα των πελατών, καθώς και λεπτομέρειες της σύμβασης), ενδεχομένως σε ανταγωνιστικές εταιρείες. Εναλλακτικά, ένας τρίτος προμηθευτή (δηλαδή ένας εξωτερικός συνεργάτης) μπορεί να εκθέσει κατά λάθος ευαίσθητα δεδομένα.

Το σενάριο 2 είναι δύσκολο να εντοπιστεί ή να προληφθεί, επειδή περιλαμβάνει τρίτους που δεν μπορούν να παρακολουθηθούν εύκολα. Ο έλεγχος πρόσβασης βάσει υπηρεσιών διαχείρισης δικαιωμάτων (Rights Management Services - RMS) προσφέρει ένα βαθμό ελέγχου στην ασφάλεια δεδομένων και είναι γενικά αποτελεσματικός έναντι ακούσιων διαρροών από τρίτους, επειδή ένα εργαλείο RMS αποτρέπει την αντιγραφή δεδομένων πέρα από την περίμετρο του πλαισίου του RMS, ενώ παράλληλα επιτρέπει την πρόσβαση μόνο σε εξουσιοδοτημένα άτομα. Ένας κακόβουλος χρήστης μπορεί να χρησιμοποιήσει λειτουργία print-screen (αποτύπωση οθόνης) ή εκτύπωση των δεδομένων για να παρακάμψει την προστασία του πλαισίου RMS. Οι ψηφιακές παγίδες (honeypots) είναι ένας αποτελεσματικός μηχανισμός εντοπισμού διαρροών από εταιρίες τρίτων, μιας και η διαρροή αποκαλύπτεται ακριβώς τη στιγμή που ενεργούν τα πλαστά δεδομένα. (όπως κλήση σε ένα εικονικό πελάτη). Σε κάθε περίπτωση οι ψηφιακές παγίδες που χρησιμοποιούνται για τον εντοπισμό εσωτερικών χρηστών της εταιρίας πρέπει να κατασκευάζονται και να χρησιμοποιούνται με σύνεση.

Σενάριο 3 (Παράνομη αποθήκευση δεδομένων σε άλλα συστήματα / συσκευές / μέσα): Ένας υπάλληλος αποθηκεύει ευαίσθητα δεδομένα σε άλλες συσκευές, π.χ., το προσωπικό του υπολογιστή στο σπίτι, φορητό υπολογιστή, συσκευή USB, CD / DVD, ή σε δικτυακό αποθηκευτικό χώρο που δεν ανήκουν στην εταιρεία (π.χ. Gmail).

Το σενάριο 3 έχει να κάνει με αποθήκευση ευαίσθητων δεδομένων σε μη εξουσιοδοτημένες συσκευές, όπως φορητοί υπολογιστές και συσκευές USB ή σε μη εξουσιοδοτημένες τοποθεσίες, όπως προσωπικά blog, χώροι συζητήσεων, Wikis, Gmail κτλ. Οι βασικές λύσεις DLP που παρακολουθούν δεδομένα DIM και DIU παρέχουν καλή προστασία (εντός του πλαισίου των ικανοτήτων εντοπισμού τους), επειδή παρακολουθούν ευαίσθητα δεδομένα και μπορούν να τα αποτρέψουν να μεταφορτωθούν στο Διαδίκτυο ή να μεταφερθούν σε εξωτερικές συσκευές. Τα βασικά μέτρα προστασίας DLP, παρόλα αυτά, είναι αναποτελεσματικά έναντι κακόβουλων χρηστών οι οποίοι εκούσια προσπαθούν να μεταφέρουν δεδομένα εκτός του οργανισμού, χρησιμοποιώντας το Διαδίκτυο ή φορητές συσκευές, επειδή σε τελική ανάλυση αυτός ο χρήστης θα βρει τρόπο να παρακάμψει το σύστημα εντοπισμού (π.χ. με κρυπτογράφηση του αρχείου). Ο έλεγχος συσκευών μπορεί να χρησιμοποιηθεί για τον περιορισμό της δυνατότητας του χρήστη να μεταφέρει δεδομένα σε μη εγκεκριμένες συσκευές, ενώ στη συνέχεια, εφαρμόζοντας κρυπτογράφηση και RMS στα ευαίσθητα δεδομένα μπορεί να διατηρήσει την προστασία, ακόμη και όταν τα δεδομένα έχουν διαρρεύσει και αποθηκευτεί σε άλλα συστήματα.

Σενάριο 4 (Κλοπή ταυτότητας/διαμοιρασμός διαπιστευτηρίων): Η ταυτότητα του εργαζομένου έχει κλαπεί ή έχει περάσει σε άλλο μη εξουσιοδοτημένο άτομο και στη συνέχεια χρησιμοποιείται για την πρόσβαση σε ευαίσθητες πληροφορίες. Περιλαμβάνει περιπτώσεις έξυπνων κάρτες, την ανταλλαγή κωδικών πρόσβασης, και του αφύλακτου ξεκλείδωτου υπολογιστή.

Για το σενάριο 4, τα βασικά μέτρα προστασίας DLP για παρακολούθηση των DIM και DIU προσφέρουν συγκεκριμένη προστασία, και ενεργούν για την πρόληψη οποιονδήποτε εντοπιζόμενων διαρροών ανεξαρτήτως του ατόμου που εμπλέκεται. Παρόλα αυτά, δεν προλαμβάνουν την πρόσβαση σε δεδομένα από κακόβουλο χρήστη, ο οποίος μπορεί εύκολα να παρακάμψει τους μηχανισμούς εντοπισμού αυτών των λύσεων. Λύσεις όπως αυθεντικοποίηση βάσει δραστηριότητας, η οποία μπορεί να εντοπίσει όταν ένας χρήστης συμπεριφέρεται ασυνήθιστα (διαφορετικά μοτίβα πληκτρολογήσεων), ή άλλους μηχανισμούς εντοπισμού ανωμαλιών (με βάση μοτίβα πρόσβασης σε βάσεις δεδομένων αρχείων/αναζητήσεων) που προσφέρουν προστασία και σε εκούσιες και σε ακούσιες περιπτώσεις. Δυστυχώς, αυτές οι λύσεις δεν είναι πάντα εφαρμόσιμες και μπορεί να προκαλέσουν πολλούς εσφαλμένους συναγερμούς. Χρησιμοποίηση διαδικασίας αυθεντικοποίησης δυο παραγόντων μπορεί να αποβεί αποτελεσματική σε εκούσια ή ακούσια σενάρια, ενώ παράλληλα δρα και ως σημαντικό φράγμα, το οποίο πρέπει να παρακάμψει ο επιτιθέμενος.

Σενάριο 5 (Κατάχρηση δικαιωμάτων): Ένας υπάλληλος (διαχειριστής DBA) καταχράται νόμιμα δικαιώματά του και της δυνατότητάς γιαι πρόσβαση σε ευαίσθητα δεδομένα πέρα από το πεδίο εφαρμογής του ή τις αναθέσεις του έργου του.

Το σενάριο 5 περιλαμβάνει ένα νόμιμο χρήστη, ο οποίος χρησιμοποιεί την πρόσβαση του σε ευαίσθητα δεδομένα με ακατάλληλο τρόπο. Αυτό αποτελεί ένα δύσκολο σενάριο να εντοπιστεί και να ληφθούν μέτρα προστασίας. Ο εντοπισμός ανωμαλιών παρέχει προστασία σε αυτό το σενάριο, μιας και μπορεί να εντοπίσει παράνομη πρόσβαση σε δεδομένα από ένα εξουσιοδοτημένο χρήστη (όπως εντοπισμός ασυνήθιστων αναζητήσεων σε βάσεις δεδομένων). Παρόλα αυτά, αν ο χρήστης είναι ενήμερος για αυτό το μηχανισμό, μπορεί να τον παρακάμψει διαμορφώνοντας «έξυπνες» ή «αόρατες» αναζητήσεις. Οι ψηφιακές παγίδες μπορούν να βοηθήσουν στον εντοπισμό τέτοιων περιστατικών όταν συμβαίνουν, αλλά προσφέρουν μικρό βαθμό πρόληψης. Ο έλεγχος πρόσβασης και παρακολούθησης δεδομένων DIU και DIM μπορεί να περιορίσει τη δυνατότητα μεταφοράς δεδομένων εκτός του οργανισμού, όμως ο εξουσιοδοτημένος χρήστης μπορεί να βρει τρόπους παράκαμψης αυτών των περιορισμών.

Σενάριο 6 (*δημοσιοποίηση δεδομένων*): Σε ένα υπάλληλο δόθηκε κατά λάθος πρόσβαση σε ευαίσθητα δεδομένα αποθηκευμένα στον υπολογιστή του μέσω κοινόχρηστων φακέλων ή εφαρμογών τύπου IM, P2P, ή κατά λάθος τοποθετήθηκαν ευαίσθητα δεδομένα σε δημόσιο προσβάσιμο χώρο (π.χ. web server)

Το σενάριο 6 αναφέρεται κυρίως σε εντοπισμό ακούσιων διαρροών με χρήση εκούσιων πράξεων κάποιου χρήστη, ο οποίος μπορεί να εκμεταλλευτεί αυτές τις λειτουργίες και να εξάγει ευαίσθητα δεδομένα. Οι βασικές λύσεις DLP και οι παρακολούθησης των DIM, DAR, και DIM προσφέρουν αποτελεσματική προστασία

σε αυτό το σενάριο. Αυτές οι λύσεις θα εντοπίσουν και θα εμποδίσουν μεταφορά ευαίσθητων δεδομένων που είναι αποθηκευμένα σε δημόσιες τοποθεσίες. Η κρυπτογράφηση και οι τεχνικές RMS αυξάνουν την ασφάλεια επειδή τα ευαίσθητα δεδομένα θα παραμείνουν προστατευμένα (κρυπτογραφημένα), ακόμη και όταν διαμοιράζονται.

Σενάριο 7 (Διαρροή ηλεκτρονικής αλληλογραφίας): Ένας εργαζόμενος στέλνει ένα μήνυμα ηλεκτρονικού ταχυδρομείου με λάθος συνημμένο ή σε λάθος παραλήπτη, διαρρέοντας ευαίσθητες πληροφορίες.

Το σενάριο 7 αναφέρεται σε μια ιδιαίτερα κρίσιμη περίπτωση, μιας και πληροφορίες ακούσια μπορούν να αποκαλυφθούν από ανασφαλείς επικοινωνίες ηλεκτρονικής αλληλογραφίας (εσφαλμένα συνημμένα, λάθος διευθύνσεις κτλ). Οι υπάρχουσες λύσεις DLP μπορούν να σαρώσουν ευαίσθητα δεδομένα υπό μορφή κειμένου ή γενικά των συνημμένων (αν είναι εντός των ορίων των δυνατοτήτων εντοπισμού τους) και στη συνέχεια να εμποδίσουν ή να ανακατευθύνουν τα ευαίσθητα δεδομένα προς ειδική διαχείριση μέσω εφαρμογών κρυπτογράφησης. Η κρυπτογράφηση RMS μπορεί να εφαρμοστεί σε ευαίσθητα δεδομένα, προτού επισυναφθούν σε ένα μήνυμα αλληλογραφίας για τη διατήρηση του ελέγχου επί των δεδομένων ακόμη και εκτός των ορίων του οργανισμού. Τα περιστατικά απώλειας μπορούν επίσης να προληφθούν μέσω εκμάθησης των μοτίβων ανταλλαγής ηλεκτρονικής αλληλογραφίας χρησιμοποιώντας τεχνικές εκμάθησης μηχανής (ML), όπως ταυτοποίησης ηλεκτρονικής αλληλογραφίας η οποία πρόκειται να αποσταλεί σε λάθος παραλήπτη.

Σενάριο 8 (Hackers αποκτά πρόσβαση σε ευαίσθητα δεδομένα): Hackers αποκτούν πρόσβαση σε θέσεις εργασίας ή σε εξυπηρετητή και κλέβουν ευαίσθητα δεδομένα. Εναλλακτικά, hackers λαμβάνουν πρόσβαση σε μια εσωτερική βάση δεδομένων μέσω διαδικτυακής εφαρμογής ή μέσω SQL injection

Σενάριο 9 (Κακόβουλο λογισμικό υποκλέπτει δεδομένα): Ένας ιός υπολογιστών ή κακόβουλο λογισμικό χρησιμοποιείται για να κλέψει ευαίσθητα δεδομένα τα οποία στη συνέχεια μεταδίδονται μέσω του διαδικτύου. Σε γενικές γραμμές, ένα ενημερωμένο αντι-υιικό θα εντοπίσει περιπτώσεις γνωστών malware και να τα αφαιρέσει. Αυτή δεν είναι η περίπτωση για τα «zero-day malware» ή προγράμματα απευθυνόμενα σε ειδικούς στόχους, μεμονωμένους Δούρειους ίππους, ή κακόβουλο λογισμικό το οποίο θα μπορούσε να αποφύγει την ανίχνευση.

Τα σενάρια 8 και 9 είναι εξίσου κακόβουλες και εκούσιες επιθέσεις, και οι υπάρχουσες λύσεις προσφέρουν μικρή προστασία έναντι τέτοιων σεναρίων. Η παρακολούθηση των DIM μπορεί να σαρώσει εξερχόμενες επικοινωνίες και να εμποδίσει ευαίσθητα δεδομένα να διαφύγουν από το χώρο της εταιρίας. Παρόλα αυτά, ένας ικανός κακόβουλος χρήστης μπορεί να σχεδιάσει ένα ειδικό κακόβουλο λογισμικό με δυνατότητα αποφυγής τέτοιου εντοπισμού (για παράδειγμα, μεταβάλλοντας τα δεδομένα ώστε να μην εμφανίζονται ως ευαίσθητα). Η κρυπτογράφηση των δεδομένων παρέχει ικανοποιητική προστασία, γιατί ακόμη και αν ένας επιτιθέμενος αποκτήσει πρόσβαση στα δεδομένα, αυτά είναι ακόμη προστατευμένα. Οι κακόβουλοι χρήστες μπορούν να εντοπιστούν με εφαρμογή ψηφιακών παγίδων σε συνδυασμό με τα άλλα γενικά μέτρα ασφάλειας, όπως τείχη προστασίας και αντι-υιικά προγράμματα.

Σενάριο 10 (Κρυμμένα δεδομένα μέσα σε αρχεία): Ένας εργαζόμενος χρησιμοποιεί ένα ευαίσθητο αρχείο ως πρότυπο (π.χ., ένα φύλλο του Excel που περιέχει προσωπικά στοιχεία πελατών) και διαγράφει το ευαίσθητο σημείο των δεδομένων. Ωστόσο, η χρήση του χαρακτηριστικού «track changes», σημαίνει ότι τα ευαίσθητα δεδομένα δεν έχουν πραγματικά διαγράφει.

Για το σενάριο 10, οι υπάρχουσες λύσεις DLP χρησιμοποιούν μηχανές για να εξάγουν όλα τα περιεχόμενα ενός αρχείου, συμπεριλαμβανομένων κρυμμένων δεδομένων, για βαθύτερο έλεγχο (όπως π.χ. έλεγχο των μετα-δεδομένων ενός εγγράφου Word). Αν η διαρροή είναι εντός των ορίων εντοπισμού των λύσεων μπορεί να προληφθεί.

Σενάριο 11 (Παράνομη μεταφορά ευαίσθητων δεδομένων από ένα ελεγχόμενο σύστημα): Ευαίσθητα δεδομένα (π.χ., τα δεδομένα των πελατών) μπορούν να εξαχθούν ή να αντιγραφούν σε ένα αρχείο από ένα σύστημα ελέγχου (π.χ., ένα CRM). Υποτίθεται ότι από τη στιγμή που τα δεδομένα αυτά φύγουν από το ελεγχόμενο σύστημα, η προσπάθεια προστασίας τους να φύγουν ακόμα ποιο μακριά, είναι δυσκολότερη. Ως εκ τούτου, πρέπει να υπάρχουν λύσεις που να εμποδίζουν στα δεδομένα να φύγουν παράνομα από τα ελεγχόμενα συστήματα. Σε ορισμένες περιπτώσεις, τα δεδομένα εξάγονται νόμιμα (για παράδειγμα, για να δημιουργηθεί μια αναφορά που περιέχει στατιστικά στοιχεία). Εάν η εξαγωγή των δεδομένων είναι νόμιμη και δεν θα πρέπει να εμποδίζεται, τα δεδομένα μπορεί να εξακολουθούν να μπορούν να διαρρεύσουν, για παράδειγμα, εάν η έκθεση στάλθεί σε λάθος παραλήπτη με ένα μήνυμα ηλεκτρονικού ταχυδρομείου ή αποθηκεύτεί σε ένα USB stick το οποίο χαθεί. Ωστόσο, οι περιπτώσεις αυτές είναι πέρα από το πεδίο αυτού του σεναρίου και έχουν αντιμετωπιστεί από τα σενάρια που αναφέρθηκαν.

Τέλος, το σενάριο 11 θεωρείται ως παραβίαση της ασφάλειας, μιας και παρακάμπτει την ενεργή ασφάλεια των δεδομένων, και κατ' επέκταση τα δεδομένα γίνονται πιο επιδεκτικά σε διαρροή. Γενικά ένα τέτοιο περιστατικό είναι δύσκολο να αντιμετωπιστεί με τις λύσεις που έχουν αναφερθεί σε αυτή την ενότητα, μιας και τα δεδομένα δεν βρίσκονται εντός του οργανισμού του ιδιοκτήτη. Τα υπάρχοντα προϊόντα DLP που παρακολουθούν τα DIU μπορούν να αλληλεπιδράσουν με ευαίσθητες εφαρμογές και να παρακολουθήσουν συγκεκριμένα τμήματα και παράθυρα αυτών των εφαρμογών, ώστε να αποτρέψουν επικίνδυνες ενέργειες, όπως αντιγραφή και επικόλληση ευαίσθητων δεδομένων. Η σάρωση των DAR μπορεί να εντοπίσει δεδομένα που έχουν «διαφύγει» του ελέγχου του συστήματος, αλλά πρέπει ακόμη να προστατευτούν.

7.3 Συμπεράσματα

Ο πίνακας 7.2 δείχνει ότι ο συνδυασμός των λειτουργιών που παρέχονται από εμπορικά διαθέσιμα προϊόντα DLP, όπως παρακολούθηση δεδομένων σε κίνηση, παρακολούθηση δεδομένων σε χρήση, και σάρωση δεδομένων σε ακινησία, μπορεί να μειώσει την πιθανότητα περιπτώσεων τυχαίων διαρροών. Παρόλα αυτά, οι υπάρχουσες λύσεις, δεν παρέχουν επαρκή προστασία έναντι σκόπιμων επιθέσεων. Η προστασία έναντι επιθέσεων εκ προθέσεως μπορεί να επιτευχθεί μερικώς,

χρησιμοποιώντας ένα βασικό πλαίσιο ελέγχου πρόσβασης RMS, αλλά και άλλες καινοτόμες λύσεις, όπως επιβεβαίωση και ταυτοποίηση βάσει δραστηριότητας, εντοπισμό ανωμαλιών (με στόχο τον εντοπισμό κατάχρησης δεδομένων), και ψηφιακές παγίδες. Εξάλλου, βελτιώνοντας τις μεθόδους εντοπισμού ευαίσθητων δεδομένων μπορεί να αυξηθεί η αποτελεσματικότητα των υπαρχόντων προϊόντων DLP της αγοράς.

Επιπρόσθετες μη-εξειδικευμένες DLP λύσεις είναι επίσης σχετικές και μπορεί να χρησιμοποιηθούν για τον εντοπισμό ή την πρόληψη διαρροών, ως μια από τις βοηθητικές λειτουργίες τους. Τα εξαρτημένα τερματικά μπορούν να χρησιμοποιηθούν για τον έλεγχο ευαίσθητων πληροφοριών και τη διασφάλισή τους μέσω αποθήκευσής τους σε μια κεντρική τοποθεσία και όχι σε τοπικά μη ελεγχόμενα συστήματα υπολογιστών. Οι τεχνικές διαχείρισης των σταθμών εργασίας μπορούν να χρησιμοποιηθούν ώστε να εφαρμοστούν πολιτικές χειρισμού δεδομένων συνοδευόμενες με τεχνολογικά μέτρα, τα οποία απαγορεύουν την εγκατάσταση τρίτων ή μη εγκεκριμένων εφαρμογών (όπως διαμοιρασμός αρχείων P2P) σε υπολογιστές της επιχείρησης. Τα κοινά εργαλεία ασφάλειας όπως λογισμικά antimalware, συστήματα εντοπισμού εισβολών, και τείχη προστασίας παρέχουν ισάξια βοήθεια στον εντοπισμό κακόβουλου λογισμικού ή προσπαθειών εισβολής.

Μια διαφορετική προσέγγιση για την ελαχιστοποίηση των ακούσιων ρίσκων διαρροής είναι μέσω προσδιορισμού και επιβολής πολιτικών, προτύπων, κανονισμών και οδηγιών. Προσδιορίζοντας ένα σχέδιο για την προστασία των δεδομένων, παράλληλα με την αύξηση του επιπέδου εγρήγορσης των υπαλλήλων (είτε με διαλέξεις για την ασφάλεια, είτε υπογράφοντας συμφωνητικά εμπιστευτικότητας κ.α.) μπορεί να αποβεί πολύ αποτελεσματικό στη μείωση του αριθμού των ακούσιων περιστατικών διαρροής. Μερικοί προμηθευτές DLP στέλνουν ένα ενημερωτικό μήνυμα στον χρήστη, όταν εντοπίζεται κάποιο περιστατικό διαρροής (όπως η προσπάθεια αντιγραφής ευαίσθητων δεδομένων σε μια συσκευή USB) με σκοπό την εκπαίδευση του χρήστη. Αυτές οι πολιτικές αποτελούν ταυτόχρονα τεχνικά και διοικητικά μέτρα. Μερικά παραδείγματα τεχνικών μέτρων είναι: Όλοι οι φορητοί υπολογιστές πρέπει να είναι εξοπλισμένοι με έξυπνες κάρτες και δυνατότητες κρυπτογράφησης, η ταυτοποίηση δύο παραγόντων πρέπει να χρησιμοποιείται σε συστήματα που φέρουν ευαίσθητα δεδομένα, ενώ πρέπει να είναι εγκατεστημένο σε κάθε υπολογιστή ενημερωμένο λογισμικό έναντι κακόβουλων ενεργειών, καθώς και αποτροπής εγκατάστασης μη εγκεκριμένων εφαρμογών. Τέλος, μερικά παραδείγματα διοικητικών μέτρων είναι: οι φορητοί υπολογιστές δεν πρέπει να αφήνονται χωρίς επιτήρηση σε αυτοκίνητα, δεδομένα που δεν χρησιμοποιούνται πρέπει να καταστρέφονται κατάλληλα, δεδομένα πελατών δεν πρέπει να μεταφορτώνονται σε απροστάτευτες ή δημόσιες τοποθεσίες αποθήκευσης, πρέπει να υιοθετούνται ισχυροί κωδικοί, καθώς επίσης και οι κωδικοί αυτοί δεν πρέπει να φυλάσσονται κοντά στους υπολογιστές και δεν πρέπει να δίνονται σε τρίτους.

8. Ανωνυμία Δεδομένων

8.1 Εισαγωγή στην ανωνυμία δεδομένων

Σκοπός της ανωνυμίας δεδομένων είναι να αμβλύνει την ανησυχία ως προς τα θέματα ιδιωτικότητας και ασφάλειας και να συμμορφώνεται με τις νομικές απαιτήσεις, αποκρύπτοντας προσωπικές λεπτομέρειες [77]. Η ανωνυμία δεδομένων εμποδίζει έναν αντίπαλο να χαρτογραφήσει ευαίσθητες πληροφορίες . Υπάρχουν τρεις βασικές περιπτώσεις, στις οποίες απαιτείται η ανωνυμία δεδομένων:

Κοινοποίηση δεδομένων. Η ανωνυμία απαιτείται πριν την έκθεση των ευαίσθητων δεδομένων (π.χ. προσωπικά δεδομένα καταναλωτών), έτσι ώστε τα δεδομένα να μπορούν να αναλυθούν χωρίς την αποκάλυψη προσωπικών πληροφοριών. Ο κάτοχος των δεδομένων μπορεί να επιθυμεί να αναλυθούν τα δεδομένα αυτά από τρίτους. Για παράδειγμα, μια φαρμακευτική εταιρεία μπορεί να επιθυμεί να μάθει τα μοτίβα κατανάλωσης των προϊόντων της για ερευνητικούς σκοπούς. Η ανάλυση αυτή μπορεί να πραγματοποιηθεί με πρωτογενή δεδομένα, χωρίς να παραβιάζεται όμως η ιδιωτικότητα των καταναλωτών. Είναι λοιπόν απαραίτητο να κοινοποιείται μόνο μια ανώνυμη έκδοση των δεδομένων.

Χρήση δεδομένων. Η ανωνυμία είναι απαραίτητη για τη χρήση συγκεκριμένων δεδομένων, αλλά επίσης και για τη συμμόρφωση με την παγκόσμια νομοθεσία προστασίας δεδομένων, η οποία απαγορεύει τη χρήση ευαίσθητων δεδομένων. Οι απαγορεύσεις αφορούν στον τρόπο με τον οποίο οι οργανισμοί μπορούν να χρησιμοποιήσουν νομικά τα ευαίσθητα δεδομένα εντός του οργανισμού. Η αρχή προστασίας δεδομένων (DPA) απαγορεύει τη χρήση δεδομένων αν ο κάτοχος μπορεί να αναγνωριστεί από αυτά. Απαιτεί επίσης από τις εταιρίες να χρησιμοποιούν το ελάχιστο ποσό δεδομένων για να επιτύχουν το στόχο τους, αλλά και να κάνουν το καλύτερο δυνατό για να διασφαλίσουν την ακρίβεια και την ασφάλεια των ευαίσθητων δεδομένων που χρησιμοποιούν.
Ένα σύνηθες παράδειγμα χρήσης δεδομένων είναι η εξαγωγή δεδομένων παραγωγής, τα οποία περιλαμβάνουν ευαίσθητα και προσωπικά δεδομένα για χρήση σε ελέγχους question answering. Η εναλλακτική της χρησιμοποίησης δεδομένων προσομοίωσης δεν είναι συνήθως μια εφικτή επιλογή, λόγω του ότι είναι χρονοβόρα και ακριβή. Είναι σημαντικό επίσης το ότι πρέπει να αναπαραχθεί ένα περιβάλλον βάσης δεδομένων για τον ακριβή έλεγχο και την ανάπτυξη, και στην περίπτωση αυτή η χρήση τεχνητών δεδομένων θα μπορούσε να περιορίσει την αποτελεσματικότητα και την αξιοπιστία του ελέγχου. Κατά συνέπεια, ο μοναδικός τρόπος που μπορεί να εφαρμοσθεί ώστε να εξασφαλίζεται ταυτόχρονα και η προστασία της ιδιωτικότητας και των ευαίσθητων πληροφοριών είναι η κάλυψη των δεδομένων παραγωγής, με παράλληλη διατήρηση της χρησιμότητας τους.

Διατήρηση δεδομένων. Η ανωνυμία απαιτείται από τις εταιρίες που επιθυμούν να διατηρήσουν τα δεδομένα τους. Η διατήρηση προσωπικών δεδομένων χωρίς ανωνυμία δεδομένων απαγορεύεται από το νόμο. Για παράδειγμα, η Google

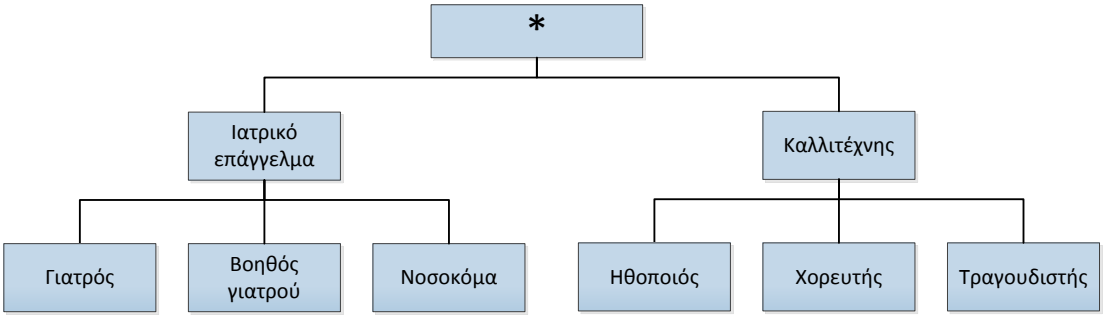
προοδευτικά κάνει ανώνυμες τις διευθύνσεις IP των αρχείων αναζήτησης, ώστε να συμμορφώνεται με τη συγκεκριμένη απαίτηση.
Πώς μπορούν όμως οι εταιρίες να προστατεύσουν τα προσωπικά τους δεδομένα ενώ παράλληλα επιτρέπουν τέτοιες κρίσιμες διαδικασίες; Μια προσέγγιση είναι η απαίτηση ότι οποιοσδήποτε διαχειρίζεται τα πραγματικά δεδομένα παραγωγής να υπογράφει μια εμπιστευτική συμφωνία, η οποία απαγορεύει την αποκάλυψη ευαίσθητων πληροφοριών. Όμως, ακόμα και με τέτοιες εμπιστευτικές συμφωνίες, δεν είναι εξασφαλισμένο ότι τα πραγματικά δεδομένα παραγωγής θα παραμείνουν εμπιστευτικά. Μια εφικτή λύση είναι το να γίνουν τα δεδομένα ανώνυμα πριν την έκθεσή τους.

8.2 Βασικές τεχνικές ανωνυμίας

Στην ενότητα αυτή περιγράφονται οι βασικές τεχνικές που μπορούν να εφαρμοσθούν σε δεδομένα, ώστε να γίνουν ανώνυμα. Τα διάφορα μοντέλα και οι αλγόριθμοι χρησιμοποιούν διάφορες παραλλαγές αυτών των τεχνικών. Τα διάφορα επίπεδα των τεχνικών ανωνυμίας έχουν διάφορες επιπτώσεις στην προστασία των δεδομένων, τη χρησιμότητα των δεδομένων και στο χώρο αναζήτησης. Όλες όμως καταλήγουν σε μια λιγότερο ακριβή αναπαράσταση των πρωτότυπων δεδομένων.

8.2.1 Γενίκευση

Η γενίκευση είναι ένα χαρακτηριστικό το οποίο αναφέρεται στην αντικατάσταση των τιμών από πιο γενικές (λιγότερο ακριβείς), αλλά σωστές τιμές. Αυτό, βέβαια, επηρεάζει τη δυνατότητα αναγνώρισης μοναδικών πλειάδων με συγκεκριμένες τιμές. Υποθέστε ότι ο τομέας δεδομένων έχει μια φυσική δομή ιεραρχίας. Για παράδειγμα, οι ταχυδρομικοί κώδικες μπορούν να παρασταθούν ως φύλλα μιας ιεραρχίας, όπου το 7522* είναι ο γονέας του 75221 και το 752* είναι ο πρόγονος του 7522*. Με χρήση της ιεραρχίας αυτής, η γενίκευση των χαρακτηριστικών μπορεί να επιτευχτεί αντικαθιστώντας την κάθε πρωτότυπη τιμή με αυτή του προγόνου της. Δηλαδή, οι τιμές του χαρακτηριστικού «φυλή», η οποία προέρχεται από την πρωτότυπη ομάδα {Μαύρη, Λευκή, Ασιατική} μπορεί να αντικατασταθεί από τη γενική τιμή {άτομο}. Η βαθμός εξειδίκευσης (ή γενίκευσης) μπορεί να μετρηθεί από το ύψος στην ιεραρχία της τιμής αντικατάστασης. Με άλλα λόγια, όσο ψηλότερα στην ιεραρχία είναι η θέση, τόσο πιο γενική είναι η τιμή. Μια τέτοια ιεραρχία αναφέρεται συνήθως ως δέντρο ταξινομίας. Στην Εικόνα8.1 παριστάνεται ένα δέντρο ταξινομίας τιμών επαγγελματικής συσχέτισης. Η τιμή *Γιατρός* είναι ένας κόμβος φύλλου. Ο γονικός κόμβος του είναι Ιατρικό Επάγγελμα και το *,το οποίο σημαίνει οποιοδήποτε επάγγελμα είναι ο γονικός κόμβος του ιατρικού επαγγέλματος (Ιατρικό Επάγγελμα).



Εικόνα 8.1: Δέντρο ταξινομίας

Γενίκευση πλήρους τομέα. Ένα δέντρο ταξινομίας, το οποίο σχετίζεται με το σχήμα γενίκευσης πλήρους τομέα εφαρμοσμένο σε ένα πίνακα, είναι ένα δέντρο στο οποίο όλες οι εμφανιζόμενες στον πίνακα τιμές, λαμβάνονται από το ίδιο επίπεδο ή ύψος του δέντρου. Έτσι, όταν γενικεύουμε ένα πίνακα σύμφωνα με τη γενίκευση πλήρους τομέα , όλες οι τιμές γενικεύονται στο ίδιο επίπεδο του δέντρου ταξινομίας. Κατά συνέπεια, όταν γενικεύουμε, για παράδειγμα, την τιμή φύλλου «Γιατρός» σε «Ιατρικό επάγγελμα» στην Εικόνα 8.1, όλες οι άλλες τιμές των φύλλων θα γενικευτούν στον αντίστοιχο πρόγονό τους του δέντρου ταξινομίας (τα «Νοσοκόμα» και «Βοηθός γιατρού» θα γενικευτούν σε «Ιατρικό επάγγελμα» και τα «Ηθοποιός», «Χορευτής» και «Τραγουδιστής» θα γενικευτούν σε «Καλλιτέχνης»). Αυτός ο τρόπος γενίκευσης έχει χρησιμοποιηθεί από πολλούς συγγραφείς όπως οι LeFevre et al. (2005) [78], Samarati (2001) [80], και Sweeney (2002b) [81].

Τομή/Γενίκευση υπόδεντρου.

Μια τομή σε ένα δέντρο ταξινομίας χαρακτηρίζεται από την ύπαρξη μίας μόνο τιμής σε κάθε μονοπάτι από τη ρίζα ως το φύλλο. Με διαφορετικά λόγια, είναι οποιαδήποτε επιλογή οποιωνδήποτε ασύνδετων κόμβων από το δέντρο, των οποίων η ένωση ισούται με τον τομέα του χαρακτηριστικού. Κατά την πραγματοποίηση γενικεύσεων, η δομή του τομής του δέντρου πρέπει να παραμείνει ανέπαφη. Κατά συνέπεια, όταν, για παράδειγμα, γενικεύεται μια τιμή στην γονική τιμή της, οι άλλες τιμές των «αδελφών» πρέπει να γενικευτούν στις αντίστοιχες γονικές τιμές τους.
Η γενίκευση του όρου «Γιατρός» σε «Ιατρικό επάγγελμα», απαιτεί επίσης τη γενίκευση των όρων «Νοσοκόμα» και «Βοηθός γιατρού» σε «Ιατρικό επάγγελμα». Όμως, οι τιμές των υπόλοιπων φύλλων, τα οποία δεν είναι απόγονοι του «Ιατρικό επάγγελμα», μπορούν να παραμείνουν μη γενικευμένες. Αυτός ο τρόπος γενίκευσης έχει χρησιμοποιηθεί σε διάφορα άρθρα, όπως των Bayardo and Agrawal (2005) [82], Fung et al. (2005) [83], Fung et al. (2007) [84], Iyengar (2002) [85], και LeFevre et al. (2005) [79].

Γενίκευση κελιών (Cell generalization). Η γενίκευση πραγματοποιείται σε συγκεκριμένα κελιά της συσχέτισης. Δεδομένου ότι μια συγκεκριμένη τιμή της συσχέτισης μπορεί να εμφανίζεται αρκετές φορές, η τιμή μπορεί να γενικευτεί σε μερικές από τις εμφανίσεις της, ενώ στις υπόλοιπες να παραμείνει μη γενικευμένη. Λόγω του ότι η γενίκευση μπορεί να περιοριστεί σε συγκεκριμένα κελιά, η γενίκευση σε επίπεδο κελιών θεωρείται πιο ευέλικτη σε σύγκριση με τα σχήματα γενίκευσης που

παρουσιάστηκαν παραπάνω, οδηγώντας έτσι σε εμφάνιση δεδομένων με πιο συγκεκριμένες τιμές, τα οποία όμως παράλληλα διατηρούν την ιδιωτικότητά τους. Αυτή η επιλογή γενίκευσης έχει χρησιμοποιηθεί σε αρκετά άρθρα, όπως για παράδειγμα, των LeFevre et al. (2005) [78], Wong et al. (2006) [86], και Xu et al. (2006) [87].

Πολυδιάστατη γενίκευση. Με δεδομένη μια συσχέτιση, η οποία περιέχει διάφορα χαρακτηριστικά και δέντρα ταξινομίας σχετικά με τα χαρακτηριστικά αυτά, μπορεί να πραγματοποιηθεί μια πολυδιάστατη γενίκευση, εφαρμόζοντας στη συσχέτιση μια συνάρτηση η οποία γενικεύει τα qid = $<v_1 , ... , v_n>$ σε qid' = $<u_1 , ... , u_n>$,όπου το κάθε v_i είτε είναι $v_i = u_i$ ή το v_i είναι ο κόμβος απόγονου του u_i στην ταξινομία του χαρακτηριστικού i.Αυτός ο τύπος γενίκευσης έχει χρησιμοποιηθεί σε αρκετά άρθρα, για παράδειγμα του [79].

Όσον αφορά τις γενικεύσεις αξίζει να αναφερθούν τα ακόλουθα σημεία:

- Είναι συχνή η αναπαράσταση γενικευμένων τιμών ενός συνεχούς χαρακτηριστικού (π.χ. μισθός ή ημερομηνία γέννησης) με διαστήματα (τα δέντρα ταξινομίας δεν χρησιμοποιούνται για τη γενίκευση τέτοιων συνεχών χαρακτηριστικών).
- Υπάρχουν περιπτώσεις ταξινόμησης, όπου δεν υπάρχει προκαθορισμένη ιεραρχική δομή.
- Όσον αφορά την επιλογή ενός συγκεκριμένου σχήματος ανωνυμίας, ώστε να εφαρμοσθεί στην εμφάνιση δεδομένων, πρέπει να λαμβάνονται υπόψη τα εξής:
  - Η πολυπλοκότητα της συνολικής διαδικασίας ανωνυμίας ή του αλγορίθμου ώστε να βρεθεί ένας ανώνυμος πίνακας.
  - Το ποσό της αλλοίωσης που δημιουργείται λόγω της ανωνυμίας.

Οι αλγόριθμοι γενίκευσης πλήρους τομέα έχουν τη μικρότερη πολυπλοκότητα, αλλά το μεγαλύτερο ποσό αλλοίωσης, ενώ οι αλγόριθμοι γενίκευσης κελιών έχουν μεγαλύτερη πολυπλοκότητα, αλλά λιγότερη αλλοίωση.

8.2.2 Καταστολή (Suppression)

Η καταστολή δεδομένων χρησιμοποιείται μαζί με τη γενίκευση και περιλαμβάνει την παράλειψη δεδομένων. Μπορεί να θεωρηθεί μια ειδική περίπτωση γενίκευσης: μια τιμή θεωρείται ότι έχει κατασταλεί, αν είναι γενικευμένη στην πιο γενική τιμή του τομέα. Για παράδειγμα, υποθέστε ότι μια ομάδα πλειάδων βάσεων δεδομένων έχει τιμές επαγγελματικού τομέα «Γιατρός» και «Ηθοποιός». Η ελάχιστη γενικευμένη τιμή, η οποία μπορεί να γενικεύσει όλες αυτές τις συγκεκριμένες τιμές είναι ολόκληρη η ομάδα * (σύμφωνα με την Εικόνα8.1). Αυτές οι συγκεκριμένες τιμές θεωρείται ότι έχουν κατασταλεί. Μερικές επιλογές καταστολής είναι οι εξής:

- Καταστολή πλειάδας: Καταστέλλει μια ολόκληρη γραμμή/πλειάδα (έχει χρησιμοποιηθεί σε αρκετές μελέτες όπως για παράδειγμα των Iyengar et al. [85]).
- Καταστολή χαρακτηριστικού: Η καταστολή γίνεται στο επίπεδο της στήλης,

αποκρύπτοντας όλες τις τιμές της στήλης (χρησιμοποιήθηκε για παράδειγμα από τον [Wang, 2005] [88])

- Καταστολή κελιού: Η καταστολή γίνεται στο επίπεδο του μεμονωμένου κελιού. Ως αποτέλεσμα σε έναν ανώνυμο πίνακα μπορεί να έχουν αφαιρεθεί δεδομένα από συγκεκριμένα κελιά μιας δεδομένης πλειάδας ή χαρακτηριστικού (χρησιμοποιήθηκε για παράδειγμα από τον [Meyerson, 2004]).

8.2.3 Μετάθεση

Η μετάθεση προτάθηκε από τους Zhang et al. (2007) [89]. Υποθέτοντας ότι ένας πίνακας περιλαμβάνει ευαίσθητα δεδομένα και αναγνωριστικά χαρακτηριστικά, προτάθηκε η απόκρυψη της προβολής του τμήματος του πίνακα, το οποίο αποτελείται από τα ευαίσθητα αυτά χαρακτηριστικά. Με τον τρόπο αυτό, καταστρέφονται οι συνδέσεις μεταξύ των ευαίσθητων και αναγνωριστικών χαρακτηριστικών, διατηρείται η ιδιωτικότητα και οι ιδιότητες συνάθροισης του πίνακα.

8.2.4 Διατάραξη

Η διατάραξη δεδομένων αναφέρεται στην αντικατάσταση των πρωτότυπων τιμών του πίνακα με συνθετικές τιμές. Οι συνθετικές τιμές επιλέγονται με τέτοιο τρόπο, ώστε η στατιστική ανάλυση στον πίνακα πριν και μετά την αντικατάσταση να μη διαφέρουν σημαντικά.
Σε σύγκριση με άλλες μεθόδους, μέσω της διατάραξης δεδομένων μπορεί να διατηρηθεί εύκολα η ανωνυμία. (τα δεδομένα δεν είναι πραγματικά). Όμως, τα δημοσιευμένα δεδομένα δεν είναι αξιόπιστα και κατά συνέπεια δεν είναι χρήσιμα. Αντιθέτως, οι μέθοδοι γενίκευσης καταλήγουν σε λιγότερο ακριβή, αλλά αξιόπιστα δεδομένα.
Δεδομένου ότι μόνο η στατιστική ανάλυση που πραγματοποιήθηκε για να κατευθύνει την αντικατάσταση παραμένει χρήσιμη και παρόμοια με την πρωτότυπη έκδοση, είναι λογικό ο εκδότης των δεδομένων να δημοσιοποιεί τις στατιστικές πληροφορίες ή τα αποτελέσματα εξόρυξης δεδομένων, αντί για τα διαστρεβλωμένα δεδομένα [90].

8.3 Βασικές έννοιες μοντέλων ιδιωτικότητας

Οι πληροφορίες που αφορούν τα άτομα περιέχου διάφορες κατηγορίες χαρακτηριστικών. Τα χαρακτηριστικά που περιγράφουν ένα άτομο μπορούν να χωριστούν στις ακόλουθες κατηγορίες:

Αναγνωριστικά (Identifiers - ID): Αυτοί οι τύπου χαρακτηριστικών περιγράφουν ένα άτομο με μονοσήμαντο τρόπο. Συνήθη παραδείγματα περιλαμβάνουν τον αριθμό κοινωνικής ασφάλειας το ΑΦΜ, το ΑΜΚΑ ή τον αριθμό άδειας οδήγησης.

Ημιαναγνωριστικά (Quasi-identifiers QID): Στην κατηγορία αυτή ανήκουν

χαρακτηριστικά τα οποία δεν παρέχουν μια μονοσήμαντη αναγνώριση, αλλά ο συνδυασμός τους μπορεί να αποδώσει μια μονοσήμαντη ταυτοποίηση μέσω σύνδεσης. Για παράδειγμα, αν υπάρχει η ακόλουθη πλειάδα σε ένα δημόσιο πίνακα ιατρικών πληροφοριών: «άντρας, ηλικία= 39, ταχυδρομικός κωδικός = 636363, χωρισμένος, γρίπη», και είναι γνωστό ότι υπάρχει μόνο ένας τριανταεννιάχρονος χωρισμένος άντρας με ταχυδρομικό κωδικό =636363, μπορεί να αποκαλυφθεί η ιατρική του κατάσταση.

Ευαίσθητα: Αυτά είναι χαρακτηριστικά τα οποία περιλαμβάνουν προσωπικές πληροφορίες ατόμων, όπως η υγεία τους, ο μισθός και οι αγοραστικές συνήθειες.

Μη ευαίσθητα: Τα χαρακτηριστικά αυτά δεν ταυτοποιούν ένα άτομο, ούτε το συσχετίζουν με ευαίσθητες προσωπικές πληροφορίες

9. Μελέτες περίπτωσης

Στην ενότητα αυτή παρουσιάζονται τρεις μελέτες περιπτώσεων στον τομές απώλειας δεδομένων και οι αντίστοιχες προτεινόμενοι μέθοδοι για την ελαχιστοποίηση της απειλής. Οι περιπτώσεις είναι οι εξής: 1)Εντοπισμός ενός εσωτερικού ατόμου που προσπαθεί να κάνει κακή χρήση και να διαρρεύσει αποθηκευμένα δεδομένα σε σύστημα βάσης δεδομένων. 2)Χρησιμοποίηση ψηφιακών παγίδων για τον εντοπισμό απειλών και 3)Εντοπισμός διαρροών μέσω ηλεκτρονικής αλληλογραφίας.

9.1 Εντοπισμός κακής χρήσης σε συστήματα βάσεων δεδομένων

Η προστασία ευαίσθητων δεδομένων (αρχεία πελατών ή ασθενών) από μη εξουσιοδοτημένη αποκάλυψη είναι πολύ σημαντικό ζήτημα για κάθε οργανισμό. Επειδή οι υπάλληλοι οργανισμών και οι συνέταιροι τους χρειάζονται πρόσβαση σε τέτοια δεδομένα για τη διεκπεραίωση της καθημερινής τους εργασίας, ο εντοπισμός και η πρόληψη απώλειας δεδομένων είναι σημαντικές και απαιτητικές εργασίες. Μια από τις μεγαλύτερες προκλήσεις είναι ο εντοπισμός ύποπτης πρόσβασης σε βάσεις δεδομένων από εσωτερικούς χρήστες της εταιρίας.

Σε αυτό το σενάριο, η υπόθεση είναι ότι οι χρήστες αλληλεπιδρούν με το σύστημα χρησιμοποιώντας μια εφαρμογή π.χ. φυλομετρητή ιστού (Web browser) και μπορούν να υποβάλλουν αιτήματα (για δεδομένα) ώστε να διεξάγουν διάφορες διεργασίες. Τα αιτήματα υποβάλλονται σε έναν εξυπηρετητή εφαρμογών που αλληλεπιδρά με μια βάση δεδομένων ώστε να συλλέξει τα απαιτούμενα δεδομένα και να αποστείλει τα αποτελέσματα στο χρήστη. Κάθε χρήστης εισέρχεται στο σύστημα με συγκεκριμένο ρόλο (π.χ. διευθυντής) και του έχουν επιτραπεί συγκεκριμένοι περιορισμοί για να εκτελεί ενέργειες. Αυτό όμως δημιουργεί ένα πρόβλημα, επειδή ένας χρήστης μπορεί να εκμεταλλευτεί τα νόμιμα δικαιώματα πρόσβασης του για διαρροή δεδομένων ή να κάνει ενέργειες που δεν συνάδουν με τους στόχους του οργανισμού. Παρακάτω παρουσιάζονται δυο μέθοδοι για τον εντοπισμό μη εξουσιοδοτημένης αποκάλυψης δεδομένων από έναν εσωτερικό χρήστη της εταιρίας.

9.1.1 Μη εποπτευόμενη εφαρμογή ανάλυσης με βάσει το περιβάλλον (unsupervised context-based analysis)

Η πρώτη μέθοδος που περιγράφεται είναι μια προσέγγιση χωρίς εποπτεία για τον προσδιορισμό ύποπτων προσβάσεων σε ευαίσθητα δεδομένα. Η προτεινόμενη μέθοδος δημιουργεί συνδέσμους μεταξύ των οντοτήτων χρησιμοποιώντας ένα *δέντρο ομαδοποίησης μιας κλάσης (one-class clustering tree - OCCT)*. Ένα δέντρο ομαδοποίησης είναι ένα δέντρο στο οποίο κάθε ένα από τα φύλλα περιέχει μια ομάδα αντί για μια και μόνο ταξινόμηση. Κάθε ομάδα γενικεύεται από ένα σύνολο κανόνων, η οποία αποθηκεύεται στο κατάλληλο φύλλο. Ο σκοπός χρησιμοποίησης του OCCT σε αυτόν τον τομέα είναι η δημιουργία ενός μοντέλου που θα συνοψίζει τα χαρακτηριστικά του αποτελέσματος (π.χ. δεδομένα), που συνήθως ο χρήστης επεξεργάζεται, με κάθε πιθανό περιεχόμενο. Έτσι, οι εσωτερικοί κόμβοι του OCCT

αντιπροσωπεύουν τα χαρακτηριστικά του περιεχομένου στα οποία συμβαίνει το αίτημα και το σύνολο κανόνων των φύλλων αναπαριστά τα δεδομένα που μπορούν νόμιμα να χρησιμοποιηθούν για το ειδικό πλαίσιο. Οι κανόνες εξάγονται χρησιμοποιώντας σύνολα αντικειμένων και προσδιορίζουν τι δεδομένα θα φαίνονται εντός του ειδικού πλαισίου με το σχετιζόμενο φύλλο.

Το σύνολο της εκπαίδευσης, που χρησιμοποιείται για τη δημιουργία του μοντέλου εντοπισμού, συντίθεται από σύνολα αποτελεσμάτων και το πλαίσιο από το οποίο ανακτώνται. Τα αιτήματα κατά την εκπαίδευση δεν χρειάζονται επισήμανση και γι' αυτό η εκμάθηση γίνεται χωρίς εποπτεία.

Κατά τη φάση εντοπισμού, το κατάλληλο σύνολο των κανόνων λαμβάνεται σύμφωνα με το πλαίσιο του αιτήματος. Μια καταγραφή στο σύνολο των αποτελεσμάτων, που είναι σύμφωνη με έναν από τους κανόνες θεωρείται κανονική. Αν το σύνολο των αποτελεσμάτων έχει βαθμολογία μεγαλύτερη από ένα προκαθορισμένο όριο, τότε η ενέργεια θεωρείται νόμιμη. Αναλύοντας και το πλαίσιο του αιτήματος και τα δεδομένα τα οποία ο χρήστης επεξεργάζεται, η μέθοδος αυξάνει την ακρίβεια εντοπισμού και την διάκριση μεταξύ κανονικών και ασυνήθιστων αιτημάτων.

9.1.2 Υπολογισμός βαθμού κακής χρήσης δεδομένων

Ο Harel (2010) πρότεινε μια μέθοδο υπολογισμού του βαθμού κακής χρήσης δεδομένων (Misusability score), η οποία εκτιμά την πιθανή ζημιά που προκαλείται μέσω μέτρησης της ευαισθησίας των δεδομένων που εκτέθηκαν στο χρήστη. Το μέτρο αυτό ενδείκνυται για δεδομένα σε μορφή πινάκων και είναι ανεξάρτητο του τομέα, που σημαίνει ότι βασίζεται σε ένα σύνολο ειδικών ορισμών που παρέχονται από τον εκάστοτε ειδικό του τομέα. Η συλλογή αυτών των δεδομένων είναι η κύρια πρόκληση για την εφαρμογή αυτού του μέτρου, ειδικά για τομείς με μεγάλο αριθμό χαρακτηριστικών και πολλές πιθανές τιμές. Το μέτρο ενσωματώνει τέσσερις διαφορετικούς παράγοντες:

1. *Αριθμός φορέων (Ποσότητα)*: Εδώ αναφέρεται στο μέγεθος των δεδομένων όσον αφορά τους διάφορους φορείς που εμφανίζονται στα δεδομένα. Η κατοχή περισσότερων δεδομένων, σημαίνει αυτόματα περισσότερες ζημιές ως αποτέλεσμα της κακής χρήσης των δεδομένων.
2. *Αριθμός ιδιοτήτων*: Τα δεδομένα περιλαμβάνουν ένα εύρος λεπτομερειών, ή ιδιοτήτων, για κάθε φορέα (όπως μισθοδοσία υπαλλήλου ή κάποια περιστατικό ασθενή). Επειδή κάθε επιπρόσθετη ιδιότητα μπορεί να αυξήσει την προκληθείσα ζημιά ως αποτέλεσμα κακής χρήσης, ο αριθμός των διαφορετικών ιδιοτήτων επηρεάζει το βαθμό κακής χρήσης.
3. *Τιμές των ιδιοτήτων (ποιότητα)*: Η τιμή μιας ιδιότητας ενός φορέα μπορεί να επηρεάσει σημαντικά το βαθμό κακής χρήσης των δεδομένων. Για παράδειγμα, το αρχείο για έναν ασθενή με HIV είναι πιο ευαίσθητο από ένα αρχείο ασθενή με ένα απλό κρυολόγημα.
4. *Επίπεδο ανωνυμίας (παράγοντας διαφοροποίησης)*: Αν και ο αριθμός των διαφορετικών φορέων στα δεδομένα μπορεί να αυξήσει το βαθμό κακής χρήσης, το επίπεδο ανωνυμίας μπορεί να το μειώσει. Το επίπεδο ανωνυμίας λαμβάνεται ως η προσπάθεια που πρέπει να καταβληθεί για τον προσδιορισμό του ειδικού φορέα κάθε

συνόλου δεδομένων. [91]

9.2 Χρησιμοποίηση δολωμάτων (honeytokens)

Τα δολώματα είναι τεχνητά αντικείμενα ψηφιακών δεδομένων τα οποία είναι σκοπίμως τοποθετημένα για τον εντοπισμό αποπειρών μη εξουσιοδοτημένης χρήσης των συστημάτων πληροφοριών. Χαρακτηρίζονται από ιδιότητες που τα κάνουν να φαίνονται ως αυθεντικά αντικείμενα δεδομένων και είναι προσβάσιμα από πιθανούς επιτιθέμενους οι οποίοι σκοπεύουν να παραβιάσουν την ασφάλεια του οργανισμού, προσπαθώντας να εκμαιεύσουν πληροφορίες με κακόβουλο σκοπό. Κάθε αλληλεπίδραση με αυτά είναι εξ ορισμού μια ανώμαλη κατάσταση που θα πρέπει να αναφερθεί και διερευνηθεί. Δημιουργούνται εγκληματολογικές πληροφορίες, καταγράφονται και αναλύονται για να αποκτηθεί γνώση για διάφορα σχέδια επίθεσης π.χ. ποιος είναι ο επιτιθέμενος, πού, πώς, και πότε ξεκίνησε μια επίθεση. Τα δεδομένα που συλλέγονται επιτρέπουν τον έλεγχο των επιθέσεων σε διάφορα επίπεδα, τα οποία κυμαίνονται από το χαμηλό επίπεδο διασύνδεσης του δικτύου και τα πρωτόκολλα δρομολόγησης ως το υψηλότερο επίπεδο εφαρμογής (Valli, 2005) [105]. Ανάλογα με τις απαιτήσεις ασφάλειας, τα δολώματα τοποθετούνται συνήθως μεταξύ κανονικών δεδομένων και παρακολουθούνται ώστε να εντοπίσουν οποιαδήποτε σχετιζόμενη δραστηριότητα.

Μια από τις κύριες προκλήσεις των δολωμάτων είναι η δημιουργία αντικειμένων δεδομένων τα οποία θα εμφανίζονται ως πραγματικά και θα είναι δύσκολο να διακριθούν από τα αληθινά δεδομένα. Ο Berkovitch et al. (2011) [92] παρουσίασε το «*HoneyGen*», μια μέθοδο αυτόματης δημιουργίας δολωμάτων. Το HoneyGen δημιουργεί δολώματα τα οποία είναι όμοια με τα αληθινά δεδομένα μέσω εξομοίωσης των χαρακτηριστικών και των ιδιοτήτων των αληθινών δεδομένων. Η είσοδος στο HoneyGen είναι σύνολα πραγματικών δεδομένων που αποθηκεύονται σε σχετική βάση δεδομένων που αποτελείται από έναν ή περισσότερους πίνακες. Το πρόγραμμα εξομοιώνει τους κανόνες που περιγράφουν την πραγματική δομή των δεδομένων, τα χαρακτηριστικά, τον τομέα, καθώς και οποιαδήποτε άλλα μετα-δεδομένα απαιτούνται. Στη συνέχεια, τα παραχθέντα δολώματα ταξινομούνται με βάση την ομοιότητα τους με τα πραγματικά δεδομένα, και τέλος επιλέγονται και χρησιμοποιούνται σε απλούς και αποτελεσματικούς μηχανισμούς ασφάλειας. Στις δυνατότητες τους είναι ο εντοπισμός εισβολών, της κακής χρήσης δεδομένων, και της απώλειας δεδομένων που διενεργούνται από εσωτερικές και εξωτερικές απειλές. Παράλληλα μπορεί να δημιουργήσει αυτόματα δολώματα για κάθε τύπο δεδομένων (αρχεία πελατών, προφίλ κοινωνικής δικτύωσης, δεδομένων που αποστέλλονται στο Διαδίκτυο, ιατρικών αρχείων ασθενών κ.α.), αρκεί τα δεδομένα να αποθηκεύονται υπό μορφή πινάκων.

Η διαδικασία δημιουργίας δολωμάτων από το HoneyGen αποτελείται από τρεις φάσεις: την εξόρυξη κανόνων, τη δημιουργία του δολώματος και το βαθμό ομοιότητας.

Κατά την φάση *εξόρυξης κανόνων*, η βάση δεδομένων, που αποτελείται από πραγματικά δεδομένα, χρησιμοποιείται για την εξαγωγή διαφόρων τύπων κανόνων για το χαρακτηρισμό των πραγματικών δεδομένων. Οι κανόνες αυτοί είναι διευθετημένοι σε πέντε κατηγορίες, όπως προσδιορίζονται από τους Duncan και

Wells (1999) [93]: Κανόνες ταυτότητας, αναφοράς, θεμελιώδεις, συνόλου τιμών και εξάρτησης χαρακτηριστικών.

Στη συνέχεια, κατά τη φάση *δημιουργίας του δολώματος*, χρησιμοποιείται μια μέθοδος που αναπτύχθηκε από τους Yahalom et al. (2010) [94] για τη δημιουργία τεχνητών βάσεων δεδομένων. Η μέθοδος αυτή δημιουργεί ανώνυμα ή τεχνητά δεδομένα για τη δοκιμή εφαρμογών με βάση πραγματικά δεδομένα, ώστε να δίνει τη δυνατότητα στους οργανισμούς να μοιράζονται τις διαδικασίες δοκιμής τους με τρίτους χωρίς να αποκαλύπτουν ευαίσθητα δεδομένα. Ως είσοδο, η μέθοδος λαμβάνει κανόνες που περιγράφουν περιορισμούς του συστήματος και τους μετατρέπει σε προβλήματα ικανοποίησης αυτών των περιορισμών (nstraint satisfaction problems - CSPs).

Τέλος, για το βήμα του *βαθμού ομοιότητας*, το HoneyGen βαθμολογεί και ταξινομεί τα δολώματα βάσει της τάξης ομοιότητας τους με πραγματικά δεδομένα. Η υψηλή βαθμολογία δείχνει ότι το δόλωμα αποτελείται από ένα συνδυασμό τιμών που είναι κοινές μεταξύ των πραγματικών δεδομένων, με συνέπεια και το δόλωμα να εμφανίζεται ως πραγματικό.

9.3 Διαρροή ηλεκτρονικής αλληλογραφίας

Οι σύγχρονες εμπορικές δραστηριότητες βασίζονται εκτενώς στην ανταλλαγή ηλεκτρονικής αλληλογραφίας (emails). Οι διάφορες λύσεις DLP επιχειρούν να αναλύσουν την ανταλλαγή ηλεκτρονικής αλληλογραφίας, ώστε να προλαμβάνεται η αποστολή αλληλογραφίας σε λάθος παραλήπτες. Ο εντοπισμός εσφαλμένων διευθύνσεων δεν είναι δυνατός, και σε ορισμένες περιπτώσεις, οι σωστοί παραλήπτες είναι εσφαλμένα επισημασμένοι ως πιθανά λάθη διεύθυνσης.

Οι περισσότερες ακαδημαϊκές λύσεις βασίζονται σε ανάλυση της κυκλοφορίας των κοινωνικών αλληλεπιδράσεων και εστιάζουν στην ανάλυση της ηλεκτρονικής αλληλογραφίας που αποστέλλεται και λαμβάνεται από ένα άτομο, όταν η αλληλογραφία προέρχεται από τον υπολογιστή του.

Η *ανάλυση ανταλλαγής ομάδων ηλεκτρονικής αλληλογραφίας* παρέχει επιπρόσθετες πληροφορίες για τις πιθανές συνδέσεις μεταξύ των χρηστών που συζητούν παρόμοια θέματα, αλλά δεν επικοινωνούν απαραίτητα μεταξύ τους. Με συνέπεια, να αντανακλά καλύτερα την πραγματική εικόνα των θεμάτων που είναι κοινά για διαφορετικούς χρήστες. Οι Zilberman et al. (2010) [95] παρουσίασαν μια νέα προσέγγιση που αναλύει την ανταλλαγή ηλεκτρονικής αλληλογραφίας μεταξύ όλων των μελών ενός οργανισμού, εξαγάγει τα θέματα συζήτησης και δημιουργεί ομάδες μελών που συζητούν τα ίδια θέματα. Με αυτόν τον τρόπο, κάθε μέλος μπορεί να ανήκει σε πολλαπλά θέματα, και μια ομάδα θεμάτων μπορεί να περιέχει μέλη που δεν έχουν επικοινωνήσει ποτέ στο παρελθόν. Κατά συνέπεια, όταν δημιουργείται ένα μήνυμα ηλεκτρονικού ταχυδρομείου, κάθε παραλήπτης κατηγοριοποιείται είτε ως πιθανός παραλήπτης διαρροής είτε ως νόμιμος, με βάση τον αποστολέα, τον παραλήπτη, αλλά και τις ομάδες θεμάτων που ανήκουν.

Η προτεινόμενη διαδικασία εντοπισμού διαρροής ηλεκτρονικής αλληλογραφίας αποτελείται από δυο φάσεις: μια φάση εκπαίδευσης και μια φάση ταξινόμησης. Η εκπαίδευση εφαρμόζεται σε ένα σύνολο μηνυμάτων που είναι γνωστό ότι δεν έχουν διαρρεύσει, και η ταξινόμηση εφαρμόζεται νέα μηνύματα που αναπαριστούν

αναζητήσεις.

Κατά τη φάση εκπαίδευσης, οι ομάδες των χρηστών ταυτοποιούνται βάσει της ηλεκτρονικής αλληλογραφίας που ανταλλάσσουν ανάλογα με το θέμα συζήτησης. Αυτή η φάση χωρίζεται σε δυο επιμέρους υποφάσεις: (1) ταυτοποίηση μέσω ομαδοποίησης των θεμάτων συζήτησης ενός οργανισμού και (2) προβολή των διαφόρων θεμάτων στους χρήστες. Αυτή η διαδικασία περιλαμβάνει την καταμέτρηση του αριθμού των μηνυμάτων ηλεκτρονικού ταχυδρομείου, που αποστέλλουν ή λαμβάνουν οι χρήστες, για το κάθε θέμα.

Κατά τη φάση ελέγχου, για κάθε παραλήπτη ενός μηνύματος ηλεκτρονικού ταχυδρομείου, επιβεβαιώνεται αν ο παραλήπτης και ο αποστολέας ανήκουν σε τουλάχιστον μια κοινή ομάδα θεμάτων. Αν δεν υπάρχει τέτοια ομάδα, μπορεί να εξαχθεί το συμπέρασμα ότι αφού δεν υπάρχει κοινό θέμα συζήτησης μεταξύ των δυο χρηστών, τότε ο παραλήπτης είναι λάθος. Σε αντίθετη περίπτωση, το περιεχόμενο της ηλεκτρονικής αλληλογραφίας συγκρίνεται με το περιεχόμενο των μηνυμάτων που έχουν ανταλλαχθεί μεταξύ των μελών κάθε κοινής ομάδας. Αν η βαθμολογία είναι αρκετά υψηλή, ο παραλήπτης μπορεί να θεωρηθεί ως νόμιμος.

10. Κακόβουλοι χρήστες

Στις μέρες μας, οι ευαίσθητες πληροφορίες έχουν γίνει πιο σημαντικές για τα οικονομικά τμήματα, τους ιατρικούς οργανισμούς, τους τομείς ασφαλείας και τις επιχειρήσεις παγκοσμίως. Από τη στιγμή που θα διαρρεύσουν οι πληροφορίες, οι κάτοχοι υφίστανται μεγάλη απώλεια. Μεταξύ των απειλών κατά της προστασίας ευαίσθητων πληροφοριών, όπως επιθέσεις από hackers, διεισδύσεις στο σύστημα, και ακούσιες λειτουργίες, το 59% δημιουργούνται από ενσυνείδητες διαρροές από κατόχους εμπιστευτικών πληροφοριών, και η συχνότητά τους συνεχώς αυξάνεται.
Υπάρχουν μελέτες επάνω στην προστασία διαρροής δεδομένων (DLP) σε διάφορες πλευρές. Οι μελέτες Bonatti P, Zhang X, Jajodia P, [97, 98, 99] αφορούν στους μηχανισμούς ελέγχου πρόσβασης, οι οποίοι επιτρέπουν μόνο σε εξουσιοδοτημένους χρήστες την πρόσβαση σε ευαίσθητα δεδομένα μέσω πολιτικών ελέγχου πρόσβασης. Όμως, οι πολιτικές αυτές είναι καταστολτικές.

Η πλειοψηφία των περιστατικών απώλειας δεδομένων αποδίδονται σε τωρινούς ή παλαιότερους υπαλλήλους. Η απώλεια δεδομένων μπορεί να προκύψει από ατύχημα, αλλά σε άλλες περιπτώσεις, από κακόβουλα άτομα του προσωπικού που διαρρέουν εμπιστευτικές πληροφορίες στο κοινό ή σε ανταγωνιστές.

Ο συνδυασμός του λογισμικού πρόληψης απώλειας δεδομένων και η τακτική εκπαίδευση των εργαζομένων στην καλύτερη πρακτική πρόληψης απώλειας δεδομένων μειώνουν τις πιθανότητες ένας εργαζόμενος να βγάλει ευαίσθητες πληροφορίες εκτός οργανισμού χωρίς άδεια.

Ένας έλεγχος του ιστορικού διαρροών δείχνει ότι όταν ένα άτομο καταφέρει μια φορά να κλέψει εμπιστευτικές πληροφορίες, θα το επαναλάβει. Οι διαχειριστές ασφάλειας των συστημάτων ΙΤ πρέπει να συνεργαστούν με το τμήμα ανθρωπίνων πόρων, ώστε να διασφαλιστεί ότι οι έλεγχοι ιστορικού περιλαμβάνουν την αξιολόγηση της πρότερης αντιμετώπισης εμπιστευτικών δεδομένων των εργαζομένων.

Ο έλεγχος είναι πολύ σημαντικός για θέσεις ατόμων με προνομιούχα δικαιώματα πρόσβασης, όπως ο διαχειριστής του συστήματος, οι DBAs, οι μηχανικοί του δικτύου, οι διαχειριστές ασφαλείας και το προσωπικό σε επίπεδα επίβλεψης και διοίκησης. Είναι σημαντικό να αναφερθεί ότι υπάρχουν νόμοι ανά τον κόσμο σχετικά με τον έλεγχο ιστορικού, όπως ο Foreign Contribution Regulation Act - FCRA στις Η.Π.Α. Για παράδειγμα, ο FCRA απαιτεί οι εργοδότες να έχουν την άδεια των εργαζομένων, για να ελέγξουν το ιστορικό τους. Δεν απαιτούν όμως όλοι οι έλεγχοι τόση γραφειοκρατία. Το διαδίκτυο αποτελεί έναν υποεκτιμημένο τρόπο πληροφόρησης σήμερα. Η απλή αναζήτηση του ονόματος ενός εργαζομένου και τα αποτελέσματα τα οποία θα εμφανίσει, μπορεί να αποκαλύψει ελλειπή προστασία δεδομένων. Τα μέσα κοινωνικής δικτύωσης αποτελούν έναν χώρο, όπου οι εργαζόμενοι έχουν τη συνήθεια να αποκαλύπτουν ευαίσθητες πληροφορίες στο φιλικό τους κύκλο μέσω του διαδικτύου, πράγμα το οποίο αποτελεί κίνδυνο για την εταιρία.

Ένας έλεγχος ιστορικού μπορεί να γίνει με την απλή παρατήρηση των δράσεων ενός ατόμου κατά τη διάρκεια της δοκιμαστικής φάσης. Αν ο νέος υπάλληλος δεν έχει ενδοιασμούς σχετικά με την κοινοποίηση λεπτομερούς εμπιστευτικής

ηλεκτρονικής πληροφορίας στο νέο εργοδότη του για προηγούμενους εργοδότες, πιθανότατα θα δράσει με τον ίδιο τρόπο όταν φύγει ή και κατά τη διάρκεια της θητείας του στην εταιρία.

Κατά την ανάπτυξη στρατηγικών πρόληψης απώλειας δεδομένων, η μεγαλύτερη έμφαση δίνεται κυρίως στις τεχνικές λύσεις. Αλλά οι διάφοροι οργανισμοί θα έπρεπε να δίνουν την αντίστοιχη έμφαση και σε μη τεχνικές λύσεις. Για παράδειγμα, μπορεί ένα λογισμικό απώλειας δεδομένων να ανιχνεύσει και να μπλοκάρει τη μετάδοση εμπιστευτικών δεδομένων με ηλεκτρονικό τρόπο, αλλά δεν προσφέρει τίποτα στην πρόληψη απώλειας δεδομένων μέσω τηλεφωνικής συνομιλίας. Είναι σημαντικό να διασφαλιστεί ότι το προσωπικό έχει κατανοήσει τις τεχνικές κοινωνικής μηχανικής εκμαίευσης εμπιστευτικών δεδομένων.

Ως κοινωνική μηχανική ορίζεται η χρήση εξαπάτησης και χειραγώγησης ώστε να αποσπαστούν εμπιστευτικές πληροφορίες. Ο χρήστης κοινωνικής μηχανικής κερδίζει την εμπιστοσύνη του υπαλλήλου μέσω της ευγένειας, της εξυπηρέτησης και της υποστήριξης. Ο επιτιθέμενος χρησιμοποιεί την εμπιστοσύνη αυτή για να εκμαιεύσει δεδομένα, τα οποία θα ήταν αδύνατο να αποσπάσει διαφορετικά. Λόγω του ότι είναι μη τεχνικός και μη δομημένος τρόπος, η απώλεια δεδομένων μέσω κοινωνικής μηχανικής είναι πολύ δύσκολο να προληφθεί.

Υπάρχουν όμως μερικές βασικές αρχές, οι οποίες εφαρμόζονται σχεδόν σε όλες τις τακτικές κοινωνικής μηχανικής. Πρώτον, ο επιτιθέμενος θα επιδιώξει να δείξει τουλάχιστον κάποια φυσικά χαρακτηριστικά ή χαρακτηριστικά συμπεριφοράς του ατόμου ή του ρόλου που υποδύεται. Όταν ένας ρόλος παρουσιάζεται με συγκεκριμένα χαρακτηριστικά, ο ανθρώπινος νους τείνει ασυναίσθητα να συμπληρώνει τα κενά και να κάνει υποθέσεις για την ταυτότητα του ατόμου. Για τον υποκλοπέα δεδομένων που επιλέγει να εκτελέσει την επίθεση κοινωνικής μηχανικής με φυσικό τρόπο, η προσοχή στο ρόλο είναι σημαντική. Για παράδειγμα, ένας άντρας που φορά ένα κομψό, ακριβό κοστούμι αποκτά έναν αέρα σοβαρότητας και πλούτου και ο κόσμος θα θεωρήσει ότι είναι αξιόπιστος και έξυπνος, πριν καν μιλήσουν μαζί του. Αν η υποκλοπή δεδομένων γίνεται εξ' αποστάσεως μέσω τηλεφώνου, μέρος του ρόλου αποτελεί η ομιλία με κατηγορηματική εμπιστοσύνη και ευγενική πειστικότητα. Η αναφορά ονομάτων ανθρώπων που είναι γνωστοί στον υπάλληλο, θα βοηθήσει τον χρήστη κοινωνικής μηχανικής να δημιουργήσει το απαραίτητο έρεισμα εμπιστοσύνης και να παρουσιαστεί ως κάποιος, τον οποίο θα έπρεπε να γνωρίζει ο υπάλληλος.

Η δεύτερη πτυχή της επίθεσης κοινωνικής μηχανικής είναι η εδραίωση της αξιοπιστίας. Ένα απλό παράδειγμα προσπάθειας εδραίωσης της αξιοπιστίας που θα μπορούσε να γίνει στο χώρο εργασίας είναι η εσκεμμένη υπονόμευσης της διαδικτυακής σύνδεσης ενός υπαλλήλου σε θέση κλειδί, τραβώντας το καλώδιο της παροχής και περιμένοντας να εκφράσει την απογοήτευσή του. Στη συνέχεια ο επιτιθέμενος προσφέρεται να «λύσει» το πρόβλημα, προτού ο υπάλληλος επικοινωνήσει με το τμήμα τεχνικής υποστήριξης, κερδίζοντας κατά συνέπεια αξιοπιστία ως «καλός άνθρωπος». Αφού εδραιωθεί η εμπιστοσύνη, ο χρήστης κοινωνικής μηχανικής χρησιμοποιεί πρόσθετες τεχνικές για να αποσπάσει δεδομένα από τον υπάλληλο. Στις τεχνικές αυτές περιλαμβάνονται το να πείσει ασυνείδητα το στόχο να αναλάβει ένα βοηθητικό ρόλο, υποβάλλοντας του διάφορες ασήμαντες διαδοχικές απαιτήσεις, κρύβοντας όμως καλά μέσα στην ακολουθία και την ζήτηση μιας εμπιστευτικής πληροφορίας, το να καλλιεργήσει το φόβο για ένα επικείμενο

γεγονός και να το χρησιμοποιήσει για να αποσπάσει «επειγόντως» ευαίσθητες πληροφορίες κλπ. Η πιο αποτελεσματική μέθοδος πρόληψης της απώλειας δεδομένων μέσω της κοινωνικής μηχανικής είναι η τακτική εκπαίδευση των εργαζομένων.

Πρόληψη απώλειας δεδομένων κατά τη διάρκεια των διακοπών

Η απώλεια δεδομένων είναι πολύ πιθανό να εμφανιστεί κατά τη διάρκεια των διακοπών, καθώς το μεγαλύτερο μέρος που προσωπικού βρίσκεται σε άδεια και ο οργανισμός υπολειτουργεί. Στη συνέχεια παρουσιάζονται κάποιες δράσεις τις οποίες μπορούν οι οργανισμοί να θέσουν σε εφαρμογή, ώστε να μειωθεί η πιθανότητα απώλειας δεδομένων.

- Δικαιώματα πρόσβασης στους μη μόνιμους υπαλλήλους και έλεγχος

Στους περισσότερους οργανισμούς το επίπεδο πρόσβασης που έχουν οι μη μόνιμοι υπάλληλοι σε πληροφορίες είναι περιορισμένο. ¨όμως, κατά τη διάρκεια των διακοπών, είναι συνηθισμένο να παραχωρούνται στους υπαλλήλους από τους διευθυντές περισσότερα προνόμια πρόσβασης, λόγω της προσπάθειας να διασφαλιστεί η εύρυθμη λειτουργία του οργανισμού, ακόμα και όταν το μεγαλύτερο μέρος του μόνιμου προσωπικού απουσιάζει. Αυτό αποτελεί ένα πολύ σοβαρό λάθος. Θα έπρεπε οι άδειες να είναι προγραμματισμένες, έτσι ώστε να παραμένει πάντα στον οργανισμό προσωπικό με προνομιούχα πρόσβαση και το μη μόνιμο προσωπικό να έχει περιορισμένη πρόσβαση μόνο σε εμπιστευτικά δεδομένα. Βέβαια, είναι σχεδόν βέβαιο ότι και το μη μόνιμο προσωπικό θα έρθει κάποια στιγμή σε επαφή με εμπιστευτικά δεδομένα. Για να διασφαλιστεί ότι το προσωπικό είναι άξιο εμπιστοσύνης, πρέπει να γίνεται στον καθένα ένας έλεγχος ιστορικού.

- Προειδοποίηση υπαλλήλων

Οι υποκλοπείς δεδομένων εφευρίσκουν πάντα νέος τρόπους απόσπασης εμπιστευτικών πληροφοριών. Όλο το προσωπικό πρέπει να λαμβάνει εκπαίδευση υπενθύμισης στις περιόδους υψηλού κινδύνου, τις εβδομάδες πριν την έναρξη των διακοπών.

- Προετοιμασία του προσωπικού και του συστήματος για περιόδους φόρτου εργασίας

Ορισμένα τμήματα του οργανισμού παρουσιάζουν άνοδο των συναλλαγών κατά τη διάρκεια των διακοπών. Οι υποκλοπείς δεδομένων το γνωρίζουν αυτό και μπορούν να χρησιμοποιήσουν το μαζικό όγκο ως συγκάλυψη για τις δραστηριότητές τους. Τα συστήματα του οργανισμού πρέπει να ελεγχθούν εκ των προτέρων ως προς τη μέγιστη δυνατότητα συναλλαγών. Αλλά είναι επίσης πιο σημαντικό να υπενθυμισθεί στους υπαλλήλους να μη βιάζονται να δράσουν ώστε να ανταποκριθούν στον φόρτο εργασίας, αν μια τέτοια δράση μπορεί να οδηγήσει σε απώλεια δεδομένων.

Συνήθως οι διευθυντές των οργανισμών αγνοούν τον κίνδυνο της «εσωτερικής απειλής». Οι διευθυντές πρέπει να είναι γνώστες των πιο κοινών τύπων συμπεριφοράς

υψηλού κινδύνου των υπαλλήλων , η οποία αυξάνει τον κίνδυνο απώλειας δεδομένων, κακόβουλης ή όχι. Οι τρεις πιο κοινές συμπεριφορές υψηλού κινδύνου είναι:

Η αποστολή ευαίσθητων πληροφοριών σε έναν προσωπικό λογαριασμό. Αυτή είναι μια πρακτική που εφαρμόζουν συνήθως πολλοί υπάλληλοι. Λόγω του ότι συνεχίζουν την εργασία τους από το σπίτι, είναι πιο βολική για αυτούς η αποστολή εγγράφων σε ένα λογαριασμό Gmail, Yahoo ή Hotmail, αντί να χρησιμοποιούν ένα εξουσιοδοτημένο από τον οργανισμό netbook ή laptop. Από τη στιγμή που τα έγγραφα αποστέλλονται σε ένα τέτοιο λογαριασμό, μπορούν να είναι προσβάσιμα από οποιοδήποτε μέρος και από οποιοδήποτε υπολογιστή που είναι συνδεδεμένος στο διαδίκτυο. Το πρόβλημα έγκειται στο ότι από τη στιγμή που τέτοιες πληροφορίες αποστέλλονται σε τέτοιου είδους λογαριασμού, βρίσκονται εκτός της υποδομής πρόληψης απώλειας δεδομένων.

Κοινή χρήση υπολογιστών χωρίς κατάλληλη εξουσιοδότηση. Η χρήση ενός εξουσιοδοτημένου από τον οργανισμό netbook ή laptop στο σπίτι για τη συνέχιση κάποιας επείγουσας εργασίας εισάγει επίσης άλλους κινδύνους για τα εμπιστευτικά δεδομένα. Δεν είναι ασυνήθιστο οι υπάλληλοι να επιτρέπουν σε άτομα της οικογένειας ή του φιλικού περιβάλλοντος τους να χρησιμοποιούν τα εξουσιοδοτημένα laptop. Επίσης όταν ένας υπάλληλος βρίσκεται σε δημόσιο χώρο και απομακρυνθεί από τον υπολογιστή του για λίγα λεπτά χωρίς να τον κλειδώσει, δίνει τη δυνατότητα σε έναν ευκαιριακό υποκλοπέα δεδομένων να κρυφοκοιτάξει.

Ανθρώπινο λάθος. Ο σκοπός κάθε υπαλλήλου είναι να εκπληρώσει τα καθήκοντά του όσο καλύτερα μπορεί, τηρώντας την πολιτική του οργανισμού. Όμως ακόμα και ο πιο ο σχολαστικός υπάλληλος μπορεί να κάνει λάθη, τα οποία μπορούν να οδηγήσουν σε αναπόφευκτη διαρροή δεδομένων. Για παράδειγμα, λόγου του αυτόματου βιβλίου διευθύνσεων που υπάρχει ως εφαρμογή στο ηλεκτρονικό ταχυδρομείο, μπορεί να γίνει κατά λάθος αποστολή ενός μηνύματος σε άλλο παραλήπτη, του οποίο το όνομα ξεκινά από το ίδιο γράμμα με τον πραγματικό παραλήπτη.

Ο μεγαλύτερος κίνδυνος για τους οργανισμού είναι οι υπάλληλοι που αποχωρούν. Ανεξάρτητα με το αν αποχωρούν από τον οργανισμό οικειοθελώς ή όχι, ένας υπάλληλος που αποχωρεί δεν παρουσιάζει την ίδια αφοσίωση στον οργανισμό όσο οι υπάλληλοι που συνεχίζουν να δουλεύουν. Παρά τη μη δεοντολογική φύση της συγκεκριμένης πρακτικής, είναι πολύ συνηθισμένο (σχεδόν αποδεκτό σε ορισμένες εταιρίες) οι υπάλληλοι που αποχωρούν να μεταφέρουν κάποια ευαίσθητα δεδομένα στο νέο εργοδότη τους. Οι επιχειρήσεις πρέπει κατά συνέπεια να εγκαταστήσουν δομές, οι οποίες δεν επιτρέπουν σε έναν υπάλληλο που αποχωρεί να γνωρίζει ευαίσθητες πληροφορίες.

Η πρόληψη απώλειας δεδομένων λόγω ενός υπαλλήλου που αποχωρεί από τον οργανισμό επιτυγχάνεται με χρήση λογισμικού DLP. Όμως, το λογισμικό DLP είναι τόσο αποτελεσματικό όσο είναι και η γενικότερη πολιτική του οργανισμού, όσον αφορά τους υπαλλήλους που αποχωρούν. Πρώτα από όλα, οι διευθυντές και οι προϊστάμενοι του οργανισμού πρέπει να έχουν τη στοιχειώδη υπευθυνότητα να ενημερώνουν τους διαχειριστές ασφάλειας ΙΤ για την οποιαδήποτε αλλαγή της

επαγγελματικής κατάστασης των ατόμων που έχουν υπό την ευθύνη τους. Η ενημέρωση πρέπει να γίνει όσο το δυνατόν γρηγορότερα, με άνω όριο τις 24 ώρες. Κάθε λεπτό που περνάει και τα δικαιώματα πρόσβασης ενός υπαλλήλου που έχει αποχωρήσει δεν έχουν απενεργοποιηθεί, αποτελεί ένα επιπλέον λεπτό κινδύνου για τον οργανισμό. Αν ο υπάλληλος που αποχωρεί δεν φεύγει αμέσως από την εταιρεία, αλλά του δίνεται περίοδος ενός μήνα να αποχωρήσει από τον οργανισμό, πρέπει η πληροφορία αυτή να δίνεται στους διαχειριστές. Σε πολλούς οργανισμούς δίνεται η δυνατότητα στους διαχειριστές ασφαλείας IT να κάνουν τις αλλαγές εντός 7 ημερών, ένα πολύ μεγάλο χρονικό διάστημα στο οποίο ένας κακόβουλος πρώην υπάλληλος μπορεί να ανακτήσει δεδομένα. Πρέπει επίσης να ελέγχεται περιοδικά, τουλάχιστον κάθε έξι μήνες, τί δικαιώματα πρόσβασης έχουν οι διάφοροι χρήστες του συστήματος.

Οι υπάρχουσες μελέτες των Cheng P C, Jason Program Office [100, 101] χρησιμοποιούν την εκτίμηση κινδύνου και επιβαρύνσεις κινδύνου για ροές πληροφορίας, ενώ ο κίνδυνος εκτιμάται βάσει της «αξιοπιστίας» του παραλήπτη και των πραγματικών πρακτικών ή παραδειγμάτων με χρήση στατιστικών διαπιστευτηρίων (π.χ. την εξουσιοδότηση ασφαλείας) του παραλήπτη, αντί για μια δυναμική μετρική αξιοπιστίας του παραλήπτη. Στη βιβλιογραφία [102] ο μηχανισμός ελέγχου πρόσβασης χρησιμοποιεί τις παρατηρημένες διαρροές ως ανάδραση για να εμποδίσει την πρόσβαση και το ρυθμό πρόσβασης σε πληροφορία. Αγνοεί όμως την επίδραση της ευπάθειας της πλατφόρμας του χρήστη για την αξιολόγηση κινδύνου. Στρατηγικές κατανομής δεδομένων μαζί με την εισαγωγή λανθασμένων αντικειμένων προτείνονται στη βιβλιογραφία [103], ώστε να βελτιωθεί η πιθανότητα αναγνώρισης διαρροών, αλλά δε λαμβάνουν υπόψη την «αξιοπιστία» των χρηστών, όταν υπολογίζεται η πιθανότητα ενοχής.

Σύμφωνα με μελέτες στη διαρροή πληροφοριών, οι δύο παράγοντες κλειδιά είναι το ιστορικό συμπεριφοράς των χρηστών και η πολιτική ασφάλειας της κάθε πλατφόρμας. Οι Fan et al [96] παρουσίασαν την εφαρμογή ενός διανεμητή, ως τρίτο μέρος, ο οποίος διαχειρίζεται και διανέμει αρχεία που περιέχουν ευαίσθητες πληροφορίες σε εξουσιοδοτημένους χρήστες, οι οποίοι τα απαιτούν. Επιπλέον, παρουσιάζεται ένα νέο μοντέλο διανομής αρχείων βασιζόμενο στην αξιοπιστία, λαμβάνοντας υπόψη τις ανεπάρκειες των προηγούμενων μελετητών. Εξετάζονται τόσο οι υποκειμενικοί όσο και αντικειμενικοί παράγοντες. Παράλληλα μπορούμε να εκτιμήσουμε τον υποκειμενικό κίνδυνο πιθανότητας ενοχής. Εφόσον υπολογιστεί η εκτίμηση αντικειμενικού κινδύνου σύμφωνα με την ευπάθεια πλατφόρμας του χρήστη, πράγμα το οποίο μπορεί να προκαλέσει ακούσια διαρροή δεδομένων, το μοντέλο είναι έτοιμο να εμποδίσει αποτελεσματικά τη διαρροή δεδομένων. Ο διανεμητής αξιολογεί την αξιοπιστία κάθε χρήστη, σύμφωνα με το αν στο ιστορικό συμπεριφοράς του χρήστη υπάρχει διαρροή ευαίσθητων δεδομένων ή όχι.

Συμπερασματικά, τα αποτελέσματα της προσομοίωσης δείχνουν ότι το μοντέλο είναι ικανό να διαχωρίσει τους διάφορους τύπους χρηστών και να ρυθμίσει τις πιθανότητες άρνησης των απαιτήσεων των χρηστών. Η πιθανότητα των τίμιων χρηστών είναι πολύ χαμηλή, των κοινών χρηστών μέτρια και των κακόβουλων χρηστών υψηλότατη. Κατά συνέπεια, το μοντέλο μπορεί να βοηθήσει στην πρόληψη της διανομής φακέλων σε κακόβουλους χρήστες.

11. Βασικές ενέργειες Πρόληψης Απώλειας Δεδομένων

Οι λύσεις DLP κοστίζουν ιδιαίτερα και σε μερικές περιπτώσεις, το κόστος είναι αποτρεπτικό για μικρές επιχειρήσεις. Επιπρόσθετα αυτές οι λύσεις πρέπει να συνοδεύονται με τεχνολογίες μεσολάβησης (proxy) ώστε η προστασία έναντι απώλειας δεδομένων να είναι πιο ολοκληρωμένη, με αποτέλεσμα το κόστος να αυξάνεται. Από την άλλη, το κόστος απώλειας δεδομένων μπορεί να είναι πολύ υψηλό λόγω κανονιστικών ποινών, κόστη κοινοποίησης πελατών και διατήρησης πελατών.

Στην ενότητα αυτή, θα επισημανθούν τα βήματα που μπορούν να ακολουθηθούν για τη δημιουργία στέρεων θεμελίων πάνω στα οποία μπορεί να χτιστεί ένα πλάνο πρόληψης απώλειας δεδομένων και εν γένει η επιβολή τεχνολογιών DLP. Ακολούθως, παρουσιάζεται μια λίστα με πέντε βασικά βήματα που πρέπει να υιοθετούνται για την προστασία των ευαίσθητων δεδομένων μιας επιχείρησης, έχοντας με το ελάχιστο δυνατό κόστος.

11.1 Αξιολόγηση υπάρχουσας κατάστασης πραγματοποιώντας Εκτίμηση κινδύνων.

Η εκτίμηση κινδύνου απώλειας δεδομένων είναι το πιο σημαντικό αντικείμενο, μιας και μέσω της αξιολόγησης δίνεται η δυνατότητα επίγνωσης του πλάνου που πρέπει να ακολουθηθεί για το σχεδιασμό της ασφάλειας του συστήματος. Σημαντικά ερωτήματα που πρέπει να απαντηθούν είναι τα εξής:

- Τι δεδομένα διαφεύγουν από το δίκτυο της εταιρίας, συμπεριλαμβανομένων των ευαίσθητων δεδομένων;
- Ποιοι υπάλληλοι (ή λειτουργικές ομάδες) αποτελούν τη μεγαλύτερη απειλή;
- Ποιο είναι το μέρος που πηγαίνουν οι πληροφορίες εκτός του δικτύου;
- Πως μεταδίδονται αυτά τα δεδομένα;
- Ποιες επιχειρησιακές διαδικασίες είναι προβληματικές και προκαλούν απώλεια δεδομένων;
- Τι δεδομένα βρίσκονται σε μη εξουσιοδοτημένους (και πιθανώς χωρίς προστασία) εξυπηρετητές δικτύου;

Η εκτίμηση ρίσκου θα δείξει ποια ευαίσθητα δεδομένα διαφεύγουν από το δίκτυο, ακόμη και για δεδομένα που δεν ήταν γνωστά ότι υπάρχουν, ή ότι χρειάζονταν προστασία. Για παράδειγμα, μια εταιρία βρήκε πολλές περιπτώσεις προσωπικών στοιχείων πελατών (personally identifiable information - PII) να διαφεύγουν από το δίκτυο. Επίσης, παράλληλα γινόταν μεταφορά μιας λίστας πρώην υπαλλήλων, η οποία περιείχε ονόματα, αριθμούς κοινωνικής ασφάλισης, και διευθύνσεις, σε έναν εξωτερικό προμηθευτή σε μορφή κειμένου.

Άλλα κοινά ευρήματα κατά τη φάση εκτίμησης των κινδύνων:

- Κύριες ομάδες ή υπάλληλοι είναι οι πιο συχνοί αποστολείς ευαίσθητων πληροφοριών, συνήθως λόγω της φύσης της εργασίας τους. Μερικές φορές όμως από

την εκτίμηση, μπορεί να βρεθεί ότι συγκεκριμένες λειτουργικές ομάδες που δεν έχουν την κατάλληλη εξουσιοδότηση, διαχειρίζονται ευαίσθητα δεδομένα σε καθημερινή βάση.

- Ευαίσθητα δεδομένα που αποστέλλονται σε διευθύνσεις email και φεύγουν από τα ασφαλή όρια του εσωτερικού δικτύου, όπως προσωπικές διευθύνσεις email ή και λογαριασμούς Hotmail και Gmail
- Ευαίσθητα δεδομένα μεταδίδονται εκτός του δικτύου μέσω μεθόδων που δεν είναι αναμενόμενο να υπάρχουν σε εργασιακά περιβάλλοντα, από δημοσιεύσεις blog έως άμεσο messaging και δίκτυα υπολογιστών P2P. Και ορισμένες φορές γενική κυκλοφορία TCP που δεν μπορεί να προσδιοριστεί.

11.2 Προσδιορισμός δεδομένων που χρήζουν προστασίας

Η κλοπή ταυτότητας αποτελεί έναν από τους μεγαλύτερους φόβους των πελατών της σημερινής εποχής. Οι εταιρίες που αποθηκεύουν ευαίσθητες πληροφορίες πελατών αποτελούν πρωταρχικούς στόχους για απώλεια προσωπικών πληροφοριών που μπορεί να οδηγήσει σε κλοπή ταυτότητας.

Το πιο προφανές σημείο που χρήζει προστασίας είναι η βάση δεδομένων των προσωπικών πληροφοριών των πελατών ή ιατρικών φακέλων των ασθενών. Εκτός από αυτά τα δεδομένα, υπάρχουν και άλλα που θα πρέπει να συμπεριληφθούν στο γενικό πλάνο προστασίας. Κάποια από αυτά μπορεί να επισημανθούν και από την εκτίμηση κινδύνου.

*Άλλα Προσωπικά στοιχεία (*PII*)*

Διάφορα άλλα προσωπικά στοιχεία μπορούν να εμφανιστούν σε πολλές μορφές εκτός από βάσεις δεδομένων πελατών.

- **Βάσεις δεδομένων υπαλλήλων Ανθρώπινου Δυναμικού**. Τα προσωπικά στοιχεία των υπαλλήλων είναι εξίσου ευαίσθητα με των πελατών και μπορούν να χρησιμοποιηθούν για κλοπή ταυτότητας αλλά και για άλλες απάτες εναντίον των υπαλλήλων. Μια τυπική βάση δεδομένων περιλαμβάνει ονόματα, διευθύνσεις, τηλέφωνα, αριθμούς κοινωνικής ασφάλισης, αριθμούς διαβατηρίου, ημερομηνίες γέννησης, διπλώματα οδήγησης, αποζημιώσεις και άλλα παρόμοια στοιχεία. Επίσης τα ίδια δικαιώματα εμπιστευτικότητας των προσωπικών τους στοιχείων έχουν και πρώην υπάλληλοι της εταιρίας.

- **Δεδομένα πρώην πελατών**. Τα δεδομένα των πρώην πελατών αποτελούν υποχρέωση της εκάστοτε εταιρίας να προστατευθούν εξίσου με των τωρινών, γι' αυτό και πρέπει να συμπεριληφθούν στο πλάνο προστασίας δεομένων.

Μη δημοσιοποιήσιμες πληροφορίες

Μη δημοσιοποιήσιμες πληροφορίες (NPI) περιλαμβάνουν οτιδήποτε μπορεί να έχει αρνητικό αντίκτυπο σε κάποιο οργανισμό, αν διαρρεύσουν πριν την

προγραμματισμένη δημόσια ανακοίνωση τους

- **Θεσμικές οικονομικές πληροφορίες**. Αποκάλυψη σημαντικών οικονομικών πληροφοριών που μπορεί να βλάψουν οποιονδήποτε οργανισμό ή να δώσουν σε ανταγωνιστές άδικο πλεονέκτημα. Αυτές μπορεί να είναι το μηνιαίο μισθολόγιο, έγγραφα πρόσφατων οικονομικών ελέγχων ή οικονομικά δεδομένα απευθείας από το λογιστήριο.

- **Συγχωνεύσεις ή σχέδια εξαγοράς**. Τα έγγραφα που περιλαμβάνουν τέτοιες δραστηριότητες πρέπει να προστατεύονται για την αποτροπή διαρροών που μπορούν να έχουν αρνητικό αντίκτυπο σε ενδεχόμενη συμφωνία και να προκαλέσουν περιττούς περισπασμούς του προσωπικού και των πελατών.

- **Άλλα δεδομένα Ανθρώπινου Δυναμικού**. Τα δεδομένα που χρήζουν προστασίας μπορεί να είναι αξιολογήσεις επιδόσεων υπαλλήλων, συμβουλευτικές σημειώσεις και λεπτομέρειες προσωπικής εξέλιξης όσον αφορά τα πλάνα συνταξιοδότησης των υπαλλήλων της εταιρίας.

Πνευματική Ιδιοκτησία

Στη σημερινή εποχή του ανταγωνιστικού εμπορίου, κάθε οργανισμός έχει κάποια πνευματική ιδιοκτησία (IP), η οποία δεν πρέπει να παραβλέπεται, σε σχέση με τα στοιχεία των πελατών.

- **Βάσεις δεδομένων πελατών**. Οι περισσότεροι χρηματοπιστωτικοί οργανισμοί θα μπορούσαν εκμεταλλευτούν τις βασικές πληροφορίες των ονομάτων και των τηλεφωνικών αριθμών των πελατών. Οι βάσεις δεδομένων των πρώην πελατών γίνονται πρωταρχικοί πελάτες και στόχοι του marketing ανταγωνιστικών εταιριών.

- **Marketing, σχέδια επωνυμίας και απόκτησης πελατών**. Και αυτές οι πληροφορίες σύμφωνα πάντα με τη σύγχρονη αγορά, πρέπει να λαμβάνονται υπόψη κατά το σχεδιασμό του πλάνου προστασίας των δεδομένων.

Μπορεί η πληθώρα αυτών των πληροφοριών να μοιάζει αποθαρρυντική για τον κατάλληλο σχεδιασμό προστασίας τους, όμως είναι σχετική απλή διαδικασία να συμπεριληφθούν στο πλάνο.

11.3 Καθορισμός Πολιτικής Αποτροπής Απώλειας Δεδομένων

Όλες οι εταιρίες που έχουν να κάνουν με ευαίσθητες πληροφορίες πελατών πρέπει να έχουν πολιτικές εμπιστευτικότητας/ασφάλειας που να περιγράφουν τη δέσμευση τους για την προστασία και τη διασφάλιση των προσωπικών στοιχείων των πελατών τους. Κατά τη δημιουργία εσωτερικής πολιτικής προστασίας δεδομένων, πρέπει η πολιτική εμπιστευτικότητας να επιτρέπει στο ίδρυμα να κάνει τις δουλειές του αποτελεσματικά. Η νέα αυτή εσωτερική πολιτική πρέπει να έχει θεσπισμένους

ξεκάθαρα τους γενικούς στόχους της επιχείρησης, ώστε να μη δίνει χώρο για αμφιβολίες μεταξύ των μελών του προσωπικού.

Οι κύριοι παράγοντες που θα ληφθούν υπόψη για την πολιτική ασφαλείας θα έχουν προσδιοριστεί από την εκτίμηση κινδύνου. Για τη θέσπιση αυτής της πολιτικής, πρέπει να γίνει σύσκεψη μεταξύ των ενδιαφερόμενων μερών ή έμπιστων μελών για τους οποίους είναι κύριας σημασίας η διαρροή δεδομένων. Εναλλακτικά ένα εξουσιοδοτημένη βασικό στέλεχος πρέπει να δημιουργήσει μία πολιτική, η οποία θα γίνει βάση εκκίνησης συζητήσεως.

Κύρια συστατικά πολιτικής

Τα κύρια συστατικά της θέσπισης πολιτικής προστασίας δεδομένων είναι τα ακόλουθα:

Ποιος; Καταγραφή όλων των σχετικών ομάδων χρηστών και ανεξάρτητων χρηστών που πρέπει να διαχειρίζονται ευαίσθητα δεδομένα.

Τι; Καταγραφή όλων των τύπων ευαίσθητων δεδομένων που πρέπει να προστατευτούν.

Που; Καταγραφή όλων των προμηθευτών, χρηστών και τοποθεσιών που αποστέλλονται ή αποθηκεύονται τα ευαίσθητα δεδομένα.

Πως; Καταγραφή όλων των αποδεκτών τρόπων αποστολής των ευαίσθητων δεδομένων.

Μονοπάτια ευαίσθητων δεδομένων

Εδώ αναλύεται η ροή των διαφόρων τύπων ευαίσθητων πληροφοριών μέσω του συστήματος και του οργανισμού. Πρέπει να γίνει προσδιορισμός των σημείων διαχείρισης των ευαίσθητων δεδομένων που μπορεί να υπάρξει διαρροή, είτε με ηλεκτρονικά είτε με φυσικά μέσα. Τα ερωτήματα που τίθενται κατά τη φάση σχεδιασμού είναι τα ακόλουθα:

Πως φτάνουν τα ευαίσθητα δεδομένα στον οργανισμό και με ποια μορφή;
Σε ποιο σημείο της διαδικασίας εισέρχονται τα δεδομένα στα εσωτερικά συστήματα;
Ποιος είναι εξουσιοδοτημένος να έχει πρόσβαση στα δεδομένα;
Σε ποια συστήματα εισέρχονται τα δεδομένα;
Σε ποιους εξουσιοδοτημένους εξυπηρετητές αποθηκεύονται αυτά τα δεδομένα;
Ποιες πολιτικές πρόσβασης έχουν υιοθετηθεί για αυτά τα δεδομένα;
Πως και για ποιο λόγο προσπελαύνονται αυτά τα δεδομένα;
Υπάρχουν περιπτώσεις που τα δεδομένα χρήζουν διαχείρισης;
Και αν ναι, από ποιον και με ποιον τρόπο;

Απαντώντας σε αυτά τα ερωτήματα τίθενται οι βάσεις για τη διαμόρφωση κατάλληλης πολιτικής για την εκάστοτε επιχείρηση. Εξάλλου, δεν υπάρχει ένας και μοναδικός δρόμος για τη σωστή πολιτική προστασίας. Σε πρώτη φάση αρκεί η θέσπιση αρχικής πολιτικής που θα αντιμετωπίζει τα πρωταρχικά ζητήματα ασφαλείας των δεδομένων, ενώ αναπτύσσεται η υπόλοιπη πολιτική. Τέλος, οι πολιτικές αυτές

μπορούν να ανανεωθούν και να αλλαχθούν ανάλογα με την εξέλιξη της τεχνολογίας και τις ανάγκες του οργανισμού.

11.4 Διόρθωση προβληματικών επιχειρησιακών διαδικασιών

Συνήθως, τη μεγαλύτερη συνεισφορά αποκάλυψης ευαίσθητων δεδομένων σε ένα οργανισμό έχουν οι διάφορες προβληματικές επιχειρησιακές διαδικασίες. Συχνά αυτές οι διαδικασίες λειτουργούν ικανοποιητικά χωρίς αποκάλυψη δεδομένων, όμως λόγω τροποποιήσεων που γίνονται στη δομή του δικτύου, ή λόγω ελλιπούς επικοινωνίας μεταξύ των μελών ή κάποιων άλλων αλλαγών, οι διαδικασίες αυτές μπορεί να εμφανίσουν προβλήματα. Σε άλλες περιπτώσεις, οι διαδικασίες είναι προβληματικές εξ αρχής, με αποτέλεσμα οι εταιρίες να αποκαλύπτουν μόνες τους δεδομένα λόγω του σχεδιασμού τους.

Η εκτίμηση κινδύνου, αν διεξαχθεί και αναλυθεί σωστά, μπορεί να παρέχει σαφείς ενδείξεις πιθανών προβληματικών επιχειρησιακών διαδικασιών. Κάποιες από αυτές τις διαδικασίες μπορούν να προσδιοριστούν και μέσω της επαναλαμβανόμενης μετάδοσης σε κάποια ή κάποιες δημόσιες ip διευθύνσεις. Παρακάτω παρουσιάζονται ορισμένα παραδείγματα προβληματικών επιχειρησιακών διαδικασιών που βρέθηκαν ότι αποκάλυπταν δεδομένα:

- Η αρχική και ασφαλής διαδικασία πήρε εισαχθέντα δεδομένα ενός πελάτη της εταιρίας από μια ασφαλή φόρμα εξυπηρέτησης στο διαδίκτυο και δημιούργησε ένα μήνυμα ηλεκτρονικού ταχυδρομείου που εστάλη σε επώνυμο λογαριασμό εντός του οργανισμού. Το δημιουργούμενο μήνυμα περιείχε όλα τα προσωπικά στοιχεία του λογαριασμού του πελάτη. Κάποια χρονική στιγμή, ο οργανισμός άλλαξε τη διαδικασία ώστε η ηλεκτρονική αλληλογραφία να προωθείται αυτόματα σε τρία διαφορετικά μέλη του προσωπικού. Όμως, μερικοί λογαριασμοί ηλεκτρονικής αλληλογραφίας βρέθηκαν εκτός του δικτύου του οργανισμού, με αποτέλεσμα τα δεδομένα να βρίσκονται σε κίνδυνο. Κατά τη διάρκεια ζωής αυτής της προβληματικής διαδικασίας εστάλησαν δεκάδες μηνύματα ηλεκτρονικού ταχυδρομείου υπό μορφή κειμένου.
- Η αρχική και ασφαλής διαδικασία δημιουργούσε στοίβες με τις νέες και ανανεωμένες πληροφορίες λογαριασμών που αποστέλλονταν μέσω ασφαλούς σύνδεσης δυο φορές την ημέρα. Κατά την εκτίμηση κινδύνου, καταγράφονταν δυο περιστατικά κάθε μέρα που περιείχαν εκατοντάδες καταχωρήσεις, κάθε μια με πλήρη προσωπικά στοιχεία πελατών. Οι πληροφορίες αποστέλλονταν σε δημόσιες διευθύνσεις IP που ανήκαν σε κέντρα δεδομένων εκτός του οργανισμού. Αποτέλεσμα ήταν η αποκάλυψη εκατοντάδων καταγραφών κάθε μέρα μέσω της προβληματικής διαδικασίας.

Μέσω του προσδιορισμού και της αντιμετώπισης οποιασδήποτε προβληματικής διαδικασίας, οι οργανισμοί μπορούν να μειώσουν σημαντικά τον κίνδυνο απώλειας δεδομένων. Και εδώ φαίνεται ότι κλειδί της διαδικασίας είναι η ουσιαστική και λεπτομερής εκτίμηση κινδύνου, γεγονός που ενισχύει την άποψη ότι είναι αναγκαία η χρησιμοποίηση μεγάλου μέρους του προϋπολογισμού της εταιρίας για το σκοπό αυτό. [104]

11.5 Εκπαίδευση χρηστών

Η πιο κοινή αιτία αποκάλυψης ευαίσθητων δεδομένων είναι και αυτή που είναι η πιο κοστοβόρα. Είναι η αντιμετώπιση των σφαλμάτων των χρηστών. Στις περισσότερες περιπτώσεις, υπάλληλοι αποκαλύπτουν ευαίσθητα δεδομένα επιτελώντας απλά τις καθημερινές τους εργασίες, χωρίς να έχουν κακόβουλες προθέσεις. Συνεπώς, το τελευταίο βήμα δημιουργίας στέρεων θεμελίων προστασίας DLP είναι η εκπαίδευση των τελικών χρηστών σε πολιτικές προστασίας δεδομένων, ώστε να ελαχιστοποιείται η εμφάνιση διαρροών πληροφοριών. Υπάρχει πληθώρα αποτελεσματικών τρόπων εκπαίδευσης του προσωπικού, τα κύρια βήματα των οποίων να είναι τα ακόλουθα:

α. Ενσωμάτωση εκπαίδευσης πρόληψης απώλειας δεδομένων στο εκπαιδευτικό πρόγραμμα του προσωπικού. Έτσι, ένα πρώτο πλεονέκτημα ενσωμάτωσης της εκπαίδευσης του προσωπικού είναι ότι ελέγχεται η συμμόρφωση του προσωπικού στις στρατηγικές εκπαίδευσης της εταιρίας. Τα περισσότερα συστήματα του Ανθρώπινου Δυναμικού κάθε εταιρίας έχουν προγράμματα που παρακολουθούν την πορεία των σεμιναρίων εκπαίδευσης και καταγράφουν ποιοι υπάλληλοι, τα έχουν παρακολουθήσει, ποιοι αναμένεται να τα παρακολουθήσουν και ποιοι ήταν ασυνεπείς. Το δεύτερο πλεονέκτημα είναι η οργάνωση, μιας και η αποτελεσματικότητα της εκπαίδευσης μπορεί να μειωθεί σε περίπτωση που είναι προγραμματισμένα σε σύντομα χρονικά διαστήματα με άλλα εκπαιδευτικά προγράμματα. Κατά συνέπεια είναι επιθυμητός ο σωστός χρονικός προγραμματισμός των εκπαιδευτικών σεμιναρίων. Τέλος, το τρίτο πλεονέκτημα είναι η συνοχή, μιας και οι οργανωτικές πολιτικές και διαδικασίες δεν είναι στατικές, με αποτέλεσμα τα θέματα διαφορετικών περιοχών να συμπίπτουν. Για παράδειγμα, η πολιτική πρόληψης απώλειας δεδομένων συμπίπτει σε πολλά σημεία με την πολιτική εμπιστευτικότητας του οργανισμού, του ανθρώπινου δυναμικού, της γενικής ασφάλειας, της τήρησης αρχείων και της νομιμότητας. Οποιεσδήποτε αλλαγές σε αυτές τις πολιτικές επηρεάζει την πολιτική DLP του οργανισμού και θα πρέπει να αποτυπωθεί και στην DLP εκπαίδευση του προσωπικού.

β. Χρησιμοποίηση όλων των διαθέσιμων καναλιών επικοινωνίας για την αύξηση της εγρήγορσης έναντι της απώλειας δεδομένων. Οι πρακτικές εκπαίδευσης απαιτούν οι περισσότεροι οργανισμοί να μπορούν να εκπαιδεύσουν όλους τους υπαλλήλους τους τουλάχιστον δυο φορές το χρόνο. Στην πραγματικότητα, ο κανόνας στις περισσότερες εταιρίες είναι εκπαίδευση μια φορά το χρόνο, με εξαίρεση προσωπικό με μεγάλο κίνδυνο απώλειας δεδομένων και σε θέσεις με υψηλά δικαιώματα πρόσβασης. Συνδυάζοντας εκπαίδευση με εγρήγορση, μέσω ηλεκτρονικής αλληλογραφίας, ενημερωτικών δελτίων και άρθρων σε πρωτοσέλιδα του εταιρικού ενδο-δικτύου, διασφαλίζεται ότι η πρόληψη απώλειας δεδομένων βρίσκεται στην κορυφή των υποχρεώσεων των υπαλλήλων.

γ. Εκπαίδευση βάσει του ρόλου του υπαλλήλου. Οι διαχειριστές ασφάλειας πρέπει να δημιουργούν προγράμματα εκπαίδευσης DLP ανάλογα τη θέση κάθε υπαλλήλου και τα εμπιστευτικά δεδομένα που διαχειρίζεται. Επίσης κάθε πρόγραμμα πρέπει να ανταποκρίνεται στο μέγεθος κάθε οργανισμού. Το ιδανικό θα ήταν η δημιουργία ενός βασικού προγράμματος εκπαίδευσης για όλους τους

υπαλλήλους και στη συνέχεια ειδικές ενότητες που θα αφορούν την αντιμετώπιση κινδύνων απώλειας δεδομένων, ανάλογα τη θέση κάθε υπαλλήλου. Για παράδειγμα, οι περιοχές που πρέπει να δοθεί έμφαση θα διαφέρουν για τους οικονομικούς επιθεωρητές ενός οργανισμού σε σχέση με τους υπαλλήλους ανάληψης δημοσίων συμβάσεων. Στο τέλος κάθε προγράμματος εκπαίδευσης, οι υπάλληλοι θα πρέπει να είναι σε θέση να αντιμετωπίσουν τα διάφορα περιστατικά που μπορεί να τους προκύψουν ανάλογα με τη φύση της εργασίας τους.

δ. Η εκπαίδευση θα πρέπει να είναι μια συνεχής διαδικασία. Το βήμα αυτό είναι σημαντικό για δύο λόγους: ο πρώτος είναι η εκ φύσεως περιορισμένη δυνατότητα συγκράτησης του ανθρώπινου μυαλού. Είναι ευρέως αποδεκτή η άποψη, όσον αφορά οποιαδήποτε μορφή εκπαίδευσης, ότι όσο συχνότερα ακούς κάτι τόσο πιο εύκολα μπορείς να το θυμάσαι μακροπρόθεσμα. Άρα όσο αποτελεσματικό και να είναι ένα πρόγραμμα εκπαίδευσης, δεν θα επιφέρει τα επιθυμητά αποτελέσματα αν θεωρείται ως ένα μοναδικό γεγονός που συμβαίνει μόνο στις περιπτώσεις πρόσληψης νέων υπαλλήλων. Ο δεύτερος λόγος είναι η μεταβαλλόμενη δυναμική του χώρου απώλειας των δεδομένων. Οι τεχνολογίες που χρησιμοποιούνται στα εργασιακά περιβάλλοντα μεταβάλλονται συνεχώς. Αν ένας οργανισμός δεν διοργανώνει συχνά προγράμματα εκπαίδευσης, οι υπάλληλοι είναι πιθανό να εφαρμόζουν τακτικές που είναι ξεπερασμένες και ανεπαρκείς για την αντιμετώπιση των καθημερινών περιστατικών που μπορεί να τους προκύψουν. Επιπροσθέτως, και οι πολιτικές και οι διαδικασίες DLP δεν πρέπει να είναι στατικές, μιας και αλλαγές δεν επιφέρονται μόνο λόγω της αλματώδους εξέλιξης της τεχνολογίας αλλά και των κανονισμών.
Οι περισσότεροι οργανισμοί πρέπει να διασφαλίζουν ότι όλοι υπάλληλοι επανεκπαιδεύονται στην προστασία των δεδομένων τουλάχιστον μια φορά το χρόνο, ενώ για κάποιες θέσεις υψηλού κινδύνου, πιο συχνά, όπως οι διαχειριστές συστημάτων/δικτύων και αντιπρόσωποι εξυπηρέτησης πελατών. [104]

12. Χαρακτηριστικά ενός αποτελεσματικού Συστήματος Πρόληψης Απώλειας Δεδομένων

Κάθε οργανισμός κατά την επιλογή του κατάλληλου λογισμικού προστασίας πρόληψης απώλειας δεδομένων πρέπει να εστιάσει σε έναν αριθμό σημαντικών χαρακτηριστικών και να συγκρίνει πόσο κοντά σε αυτά τα χαρακτηριστικά είναι το κάθε λογισμικό που του προσφέρεται.. Ακολούθως παρουσιάζονται τα «πρέπει» ενός αποτελεσματικού συστήματος πρόληψης απώλειας δεδομένων.

1. Οπτικοποίηση

Το λογισμικό πρέπει να παρέχει καθαρή εικόνα των υπαρχόντων ρυθμίσεων ελέγχου πρόσβασης για το συνολικό δίκτυο της επιχείρησης. Για ένα μεγάλο οργανισμό, γραφική απεικόνιση των δικαιωμάτων πρόσβασης θα ήταν ιδανική, μιας και επιτρέπει στους διαχειριστές να ανιχνεύουν διαμορφώσεις πρόσβασης υψηλού ρίσκου ταχύτατα και να κάνουν τις απαραίτητες διορθώσεις.
Οι διαχειριστές δεν είναι αναγκαίο μόνο να μπορούν να δουν τα δικαιώματα πρόσβασης των χρηστών αλλά και να έχουν τη δυνατότητα εντοπισμού των δικαιωμάτων μεμονωμένων χρηστών όπως και ομάδων χρηστών. Το λογισμικό DLP πρέπει να έχει τη δυνατότητα να συλλαμβάνει και να ανανεώνει τις αλλαγές στα δικαιώματα πρόσβασης, όσο πιο κοντά γίνεται, σε πραγματικό χρόνο.

2. Ρύθμιση

Το λογισμικό DLP πρέπει να επιτρέπει άμεση επέμβαση στις άδειες των αρχείων και των δικτύων. Το καλύτερο εργαλείο DLP είναι αυτό που επιτρέπει τον προσδιορισμό, τη δοκιμή, την εφαρμογή και την άρση της πρόσβασης σε δεδομένα, τη χρήση και την τροποποίηση των δικαιωμάτων. Αυτό σημαίνει ότι οι διαδικασίες που ξεκινούν από το λογισμικό DLP πρέπει να διευκολύνονται χρησιμοποιώντας ένα δίκτυο ή/και εξυπηρετητή λογαριασμού διαχειριστή ή έστω ενός λογαριασμού που να έχει τα απαιτούμενα δικαιώματα ελέγχου. Δεδομένης της ευαισθησίας των μεταβολών στους περιορισμούς του συστήματος, το λογισμικό DLP πρέπει να έχει ένα περιβάλλον δοκιμών, όπου θα μπορούν να δοκιμάζονται οποιεσδήποτε μεταβολές στα δικαιώματα πρόσβασης προτού υλοποιηθούν στο ενεργό περιβάλλον του συστήματος.

3. Ελεγκτική

Πρέπει να υπάρχουν διεξοδικές διαδικασίες ελέγχου που να καταγράφουν όλες τις σημαντικές δραστηριότητες που σχετίζονται με πρόσβαση, χρήση, μεταφορά και τροποποίηση δεδομένων. Πιο συγκεκριμένα, πρέπει να συλλαμβάνεται το άνοιγμα, η μετονομασία και η διαγραφή αρχείων κάθε συστήματος, με έμφαση στα πιο ευαίσθητα αρχεία και φακέλους.

Το λογισμικό DLP πρέπει επίσης να κρατά αρχείο για οποιεσδήποτε μεταβολές ρυθμίσεων ασφαλείας πραγματοποιούνται από διαχειριστές του συστήματος. Πρέπει

το ιστορικό της καταγραφής ελέγχου (audit trail) αυτής να παραμένει προσβάσιμο για αρκετό χρονικό διάστημα προτού αρχειοθετηθεί. Ένα ρεαλιστικό διάστημα θα ήταν 12 μήνες, ενώ το ιδανικό χρονικό πλαίσιο εξαρτάται από τη δυναμική κάθε οργανισμού. Η καταγραφή ελέγχου αυτή πρέπει να προστατεύεται από αλλαγές και η πρόσβαση σε αυτή πρέπει να περιορίζεται μόνο σε εξουσιοδοτημένα άτομα. Για το λόγο αυτό, πρέπει να είναι εξαγώγιμη σε πολλαπλές μορφές δεδομένων όπως txt, doc, csv και pdf, ώστε να έχουν πρόσβαση στις πληροφορίες αυτές και άτομα που κανονικά δεν έχουν τα απαραίτητα δικαιώματα (όπως εξωτερικοί ελεγκτές).

4. Απόδοση

Η χρήση του λογισμικού DLP δεν πρέπει να οδηγεί σε μείωση της απόδοσης του εξυπηρετητή αρχείων, της εμπειρίας του τελικού χρήστη ή στη ροή μεταφοράς των δεδομένων. Ενώ ο ρόλος του εργαλείου DLP σχεδόν αναπόφευκτα θα οδηγήσει σε υψηλότερο φορτίο στο δίκτυο και στους εξυπηρετητές, το φορτίο αυτό πρέπει να είναι μέσα σε αποδεκτά επίπεδα. Η λύση DLP πρέπει να διευκολύνει και όχι να αναστέλλει τις εξουσιοδοτημένες επιχειρησιακές διαδικασίες.

Τα ακόλουθα είναι συνέχεια των «πρέπει» που λαμβάνει υπόψη ο διαχειριστής όταν συγκρίνει τους διάφορους τύπους εργαλείων πρόληψης απώλειας δεδομένων

5. Επεκτασιμότητα

Στη σημερινή εποχή λίγοι οργανισμοί παραμένουν στατικοί σε μέγεθος. Οι επιχειρήσεις πάντα αποσκοπούν στην αύξηση των κερδών τους και αυτό οδηγεί σε αύξηση του μεγέθους του οργανισμού. Η ανάπτυξη αυτή μπορεί να είναι στον αριθμό των σταθμών εργασίας του δικτύου για να καλυφθεί ο μεγαλύτερος αριθμός υπαλλήλων. Η ανάπτυξη μπορεί να εμφανιστεί ως αύξηση του αριθμού των εξυπηρετητών ώστε να ανταπεξέλθουν στην αυξανόμενη ζήτηση για μεγαλύτερη χωρητικότητα δεδομένων, λόγω περισσότερων πελατών ή συναλλαγών.

Η καλύτερη επιλογή λογισμικού DLP είναι αυτή που μπορεί εύκολα να κλιμακωθεί και να επεκταθεί για να καλύψει την ανάπτυξη του οργανισμού. Κανονικά, οι διαχειριστές ασφάλειας πρέπει να έχουν υπόψη τους την πιθανότητα ο όγκος δεδομένων να διπλασιάζεται κάθε 12 μήνες επιχειρησιακών κύκλων.

6. Ευκολία εγκατάστασης

Όσο πιο εύκολο είναι να εγκατασταθεί ένα λογισμικό DLP, τόσο λιγότερο ενοχλητικό για τις διαδικασίες της επιχείρησης θα είναι. Ιδιαίτερο βολικό θα μπορούσε κάποιος να σημειώσει ότι θα ήταν η εγκατάσταση να γίνεται στο τέλος μιας μέρας ή σε μη εργάσιμες μέρες. Εξάλλου, λογισμικά που θέλουν μέρες για να εγκατασταθούν πλήρως, απαιτούν και ενεργή απασχόληση εξειδικευμένου τεχνικού προσωπικού, γεγονός που δεν τα καθιστά την καλύτερη επιλογή. Ενώ τέλος ένα λογισμικό που είναι δύσκολο να εγκατασταθεί, είναι πιθανό να παρουσιάσει προβλήματα στο μέλλον, σε σχέση με ένα που είναι πιο εύκολο στην εγκατάστασή του.

7. Ευκολία ενσωμάτωσης

Στενά συνδεδεμένη με την ευκολία εγκατάστασης (αν και διαφορετικό χαρακτηριστικό) είναι η ευκολία ενσωμάτωσης του συστήματος. Όσο πιο ποικίλα είναι τα λειτουργικά συστήματα, το υλικό των εξυπηρετητών, οι συσκευές του δικτύου, τα τελικά σημεία και οι εφαρμογές της επιχείρησης που μπορεί να υποστηρίξει το λογισμικό DLP, τόσο πιο αποτελεσματικό θα είναι. Συχνά, αν και δεν είναι πρακτικό, οι οργανισμοί επιλέγουν απλού τύπους συστημάτων και συσκευών.

Στην πραγματικότητα, τα περιβάλλοντα των επιχειρήσεων μπορεί να χρησιμοποιούν ταυτόχρονα, εξυπηρετητές Windows, Linux, Unix και Macintosh, ανάλογα της καταλληλότητας της κάθε εφαρμογής των επιχειρήσεων. Το λογισμικό DLP πρέπει να μπορεί απρόσκοπτα να καταγράφει δεδομένα από διάφορα περιβάλλοντα.

8. Ευκολία χρήσης

Κάθε σύστημα που εισάγεται σε ένα οργανισμό για πρώτη φορά απαιτεί κάποιας μορφής εκπαίδευση προτού αναπτυχθεί πλήρως. Η εκπαίδευση πρέπει να διεξάγεται από έναν εκπρόσωπο του προμηθευτή ή από ένα χρήστη που να έχει εμπειρία στο σύστημα. Φυσικά όλα τα συστήματα δεν απαιτούν το ίδιο επίπεδο εκπαίδευσης. Παρόλα αυτά, η ευκολία χρήσης του λογισμικού DLP είναι σημαντικός παράγοντας του βαθμού προστασίας των πληροφοριών του οργανισμού.

9. Συνολικό κόστος κτήσης

Το λογισμικό DLP πρέπει να έχει απτή αξία χρηματική, και εργατοωρών εγκαταστάσεως και συντηρίσεως.

13. Συμπεράσματα

Με την πάροδο του χρόνου, οι αναφορές ερευνητών για τη φύση των περιστατικών απώλειας δεδομένων, συγκλίνουν σε μια κοινή ιδέα: Η πλειοψηφία των περιστατικών δεν είναι κακόβουλα ή σκόπιμα. Στην πραγματικότητα, μερικές αναφορές δείχνουν ότι οι τυχαίες απώλειες δεδομένων είναι το 90% των συνολικών απωλειών.

Ας πάρουμε ένα πραγματικό περιστατικό, από τη ζωή ενός προγραμματιστή, του οποίου ο διαχειριστής βρήκε ότι έστελνε εμπιστευτικά αρχεία σε ανταγωνιστές. Υπό αυτό το πρίσμα, η ενέργεια θα κρινόταν κακόβουλη, ανήθικη και αθέμιτη. Ενώ και ο υποκλοπέας θα άξιζε σημαντική τιμωρία. Φυσικά βέβαια, αν η ενέργεια του ήταν σκόπιμα κακόβουλη. Από την εξακρίβωση όμως δεν ήταν, γιατί αυτό που συνέβαινε με τον προγραμματιστή αυτόν, ο οποίος δεν είχε την αγγλική ως μητρική του γλώσσα και συνεπώς δεν ήταν άριστα εξοικειωμένος με αυτή, έστελνε μέσω ηλεκτρονικού ταχυδρομείου στην σύζυγο του (η οποία γνώριζε καλύτερα Αγγλικά και δούλευε σε άλλη εταιρία) όλο το αρχειακό υλικό του πλάνου που είχε ετοιμάσει, ώστε να το αναγνώσει και να διορθώσει τυχόν γραμματικά ή συντακτικά λάθη, προτού αποστείλει τα αρχεία στον πελάτη. Με άλλα λόγια, θεωρούσε τον παραλήπτη της αλληλογραφίας ως τη σύζυγό του και όχι ως υπάλληλο ανταγωνιστικής εταιρείας. Αποδείχθηκε ένα σφάλμα χωρίς κακόβουλη πρόθεση.

Τέτοια σφάλματα είναι καθημερινότητα στο περιβάλλον των σημερινών επιχειρήσεων και επιβεβαιώνουν την πεποίθηση ότι κανένα σύστημα δεν είναι 100% ασφαλές. Η επιστήμη (εννοώντας την τεχνολογία) πίσω από τα περισσότερα προηγμένα συστήματα πρόληψης απώλειας δεδομένων δεν μπορεί να κάνει και πολλά σε αυτές τις περιπτώσεις. Αν ένας οργανισμός δεν έχει στρατηγικές που να εστιάζουν στην αντιμετώπιση διαρροής δεδομένων, τότε τα διάφορα συστήματα παρέχουν πολύ μικρή προστασία έναντι απώλειας δεδομένων.

Ακόμη και τα καλύτερα συστήματα απώλειας δεδομένων είναι αντιδραστικά σε κάποιο βαθμό. Εκτός αν υπάρχει μια στρατηγική πρόληψης απώλειας δεδομένων που να μπορεί να ανταπεξέλθει σε όλες τις περιπτώσεις, που όχι μόνο να περιλαμβάνει απόκτηση εξελιγμένων συστημάτων, αλλά και να διασφαλίζει ότι το προσωπικό κατανοεί τους διάφορους τρόπους που μπορεί να προκαλέσει τυχαία διαρροή δεδομένων, αλλιώς τα συστήματα θα παραμένουν ένα εργαλείο αντιμετώπισης αφού έχει συμβεί το γεγονός της απώλειας. Η ευαισθητοποίηση του προσωπικού περιλαμβάνει βασική καθοδήγηση και εκπαίδευση μέσα από σεμινάρια, μέχρι και αναδυόμενα παράθυρα που θα ειδοποιούν το προσωπικό για τις συνέπειες που θα έχουν αν εκτελέσουν ένα συγκεκριμένο σύστημα ενεργειών.

Μια αποτελεσματική στρατηγική DLP είναι αυτή που βοηθάει το προσωπικό να κατανοήσει το λόγο που η προστασία των εκάστοτε δεδομένων είναι σημαντική (κανονιστικές ρυθμίσεις, νόμοι εμπιστευτικότητας, ηθική, φήμη, εμπιστοσύνη πελατών, ανταγωνιστικά πλεονεκτήματα κ.α.) και τους ενθαρρύνει να μεταβάλλουν τη συμπεριφορά τους ώστε να κάνουν τις ενέργειες τους λιγότερο απειλητικές για την εμπιστευτικότητα των δεδομένων του οργανισμού.

14. Βιβλιογραφία

1. John D. Howard, Thomas A. Longstaff, A Common Language for Computer Security Incidents, Sandia National Laboratories, United States Department of Energy, California.
2. Elovici Y, Rokach, L., and Albayrak S. 2012. Special issue on Data Mining for Information Security, Information Sciences.
3. Σωκράτης Κάτσικας, Ασφάλεια Υπολογιστών, Σχολή Θετικών Επιστημών και Τεχνολογίας, Ελληνικό Ανοικτό Πανεπιστήμιο, Τόμος Α', Πάτρα 2001.
4. Stalling, W., and Brown, L. 2007. Computer Security: Principles and Practice, fi rst ed. Prentice Hall.
5. Wacks R., Personal information – Privacy and the Law, Oxford, 1983.
6. Druey J. – N., Information als
7. Σημίτης Σ., Οι ευαίσθητες πληροφορίες σε αναζήτηση νομικού καθεστώτος, σε Συμβούλιο της Ευρώπης, Νομοθετικά προβλήματα της προστασίας προσωπικών δεδομένων, Αθήνα 1992, σελ. 267 επ.
8. Αυγουστιανάκης Μ., Προστασία του ατόμου από την επεξεργασία προσωπικών δεδομένων. Προβλήματα και αντιμετώπιση από το δίκαιο, σε Δτα, 2011, σελ 673.
9. Μάνεσης Α., Συνταγματικά Δικαιώματα – Ατομικές Ελευθερίες, Αθήνα 1982.
10. Χρυσογόνος Κ., Ατομικά και Κοινωνικά Δικαιώματα, Αθήνα 2002.
11. Μήτρου Α., Το δίκαιο στην Κοινωνία της Πληροφορίας, Αθήνα 2002.
12. Μήτρου Α., «Προστασία Προσωπικών Δεδομένων στο Διαδίκτυο», στο συλλογικό έργο: Εθνική Επιτροπή Τηλεπικοινωνιών και Ταχυδρομείων, Η αγορά του Internet στην Ελλάδα, Αθήνα 1999, σελ 71.
13. Μήτρου Α., "Προστασία δεδομένων προσωπικού χαρακτήρα στα ψηφιακά δίκτυα", στο συλλογικό έργο: ΕΚΕΜΕ, "Κοινοτικές και εθνικές ρυθμίσεις στον τομέα των Τηλεπικοινωνιών - Εξελίξεις και Προοπτικές", Αθήνα 1998, σελ. 141 επ.
14. Μαρία Μαλλιαρού, άρθρο «Νομοθετική προστασία επεξεργασίας ευαίσθητων προσωπικών δεδομένων στον ηλεκτρονικό φάκελο υγείας».
15. Σωκράτης Κάτσικας , Δημήτρης Γκρίτζαλης, Ασφάλεια πληροφοριακών συστημάτων, Εκδόσεις Νέων Τεχνολογιών, 2004.
16. Barber B. Patient data and security: an overview. International journal of Medical Informatics 1998; 49(1):19-30.
17. Ελληνική Εταιρεία Επιστημόνων Ηλεκτρονικών Υπολογιστών και Πληροφορικής. Ασφάλεια Πληροφοριών: Τεχνικά, Νομικά και Κοινωνικά Θέματα. Αθήνα: Εκδόσεις Νέων τεχνολογιών, 1995.
18. Μαλλιαρού Μ. Πολιτική ασφαλείας και διασφάλιση ιατρικού απορρήτου ηλεκτρονικού φακέλου υγείας ασθενών. Μεταπτυχιακή Διπλωματική εργασία Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών Τμήμα Νοσηλευτικής

Διαπανεπιστημιακό Διατμηματικό πρόγραμμα μεταπτυχιακών σπουδών: Ειδίκευση Πληροφορική Υγείας, Αθήνα 2006.

19. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Council of Europe Convention 108, 1981 Jan, ISBN (1982) 92871-00225
20. Council of Europe, Directive 95/46/EC, On the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data (OJ L281/31 50, 24 October 1995), Strasbourg, 1995.
21. Bunker, G., and Gareth, F.K. 2009. Data Leaks for Dummies. Wiley.
22. Frost & Sullivan. 2008. World Data Leakage Prevention Market. Technical Report ND34D-74, Frost & Sullivan, United States.
23. Clive Blackwell, A Security Architecture to Protect Against Data Loss, Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, University of London, UK, 2010.
24. Ouellet, E., and Proctor, P.E. 2009. Magic Quadrant for Content-Aware Data Loss Prevention Technical Report, RA4 06242010, Gartner RAS Core Research.
25. Mogull, R. 2007. Understanding and Selecting a Data Loss Prevention Solution . Technical Report, SANS Institute, Securosis.
26. A. Shabtai, A Survey of Data Leakage Detection and Prevention Solutions, SpringerBriefs in Computer Science, 2012.
27. Bradley R. Hunter, Data Loss Prevention Best Practices - Managing Sensitive Data in the Enterprise, A messaging media publication, 2007.
28. Raschke, T. The Forrester Wave: Data Leak Prevention, Q2 2008 . Technical report, Forrester Research, Inc, 2008.
29. Salem, B.M., Heshkop, S., and Stolfo, S.J. A survey of insider attack detection research. Insider Attack and Cyber Security- Beyond the Hacker, Springer, 39, 23–27, 2008.
30. Maybury, M. Analysis and detection of malicious insiders. Proceedings, International Conference on Intelligence Analysis.
31. Hong, J., Kim, J., and Cho, J. The trend of the security research for the insider cyber threat, International Journal of Future Generation Communication and Networking , 3(2), 31–40, 2010.
32. Bowen, B.M., Salem M.B., Hershkop, S., Keromytis, A.D., and Stolfo, S.J.. Designing host and network sensors to mitigate the insider threat. IEEE Security and Privacy , 7(6), 22–29, 2009
33. . Salem, B.M., and Stolfo, S.J. Decoy document deployment for effective masquerade attack detection. Proceedings, 8th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA), pp. 35–54, 2011.

34. Stolfo, S.J., Hershkop, S., Hu, C.W., Li, W.J., Nimeskern, O., and Wang, K. Behavior-based modeling and its application to Email analysis. ACM Transactions Internet Technology, 6(2), 187–221, 2006.
35. Borders, K., and Prakash, A.. Towards quantification of network-based information leaks via HTTP. Proceedings, 3rd conference on Hot Topics in Security, 2008.
36. Abbadi, I.M., and Alawneh, M. Preventing insider information leakage for enterprises. Proceedings, International Conference on Emerging Security Information, Systems and Technologies , 99–106, 2008.
37. Yixiang, S., Tao, P., and Minghua, J. Secure multiple XML documents publishing without information leakage . Proceedings, International Conference on Convergence Information Technology , 2114–2119, 2007.
38. Cathey, R., Ma, L., Goharian, N., and Grossman, D. Misuse detection for information retrieval systems. Proceedings, 12th ACM Conference on Information and Knowledge Management (CIKM), 2003.
39. Goharian, N., Ma, L., and Meyers, C. Detecting misuse of information retrieval systems using data mining techniques (poster). Proceedings, IEEE International Conference on Intelligence and Security Informatics, 2005.
40. Ma, L., and Goharian, N. Query length impact on misuse detection in information retrieval systems. Proceedings, ACM Symposium on Applied Computing, 1070–1075, 2005.
41. Mun, H., Han, K., Yeun, C.Y., and Kim, K. Yet another intrusion detection system against Insider Attacks. Proceesings, Symposium on Cryptography and Information Security, 2008.
42. Kamra, A., Terzi, E., Evimaria, and Bertino, E. Detecting anomalous access patterns in relational databases. International Journal on Very Large Databases , 17(5), 1063–1077, 2008.
43. Sunu, M., Michalis, P., Hung, N., and Shambhu, U. A Data-Centric Approach to Insider Attack Detection in Database Systems. Technical Report, 2009.
44. Fonseca, J., Vieira, M., and Madeira, H. Online detection of malicious data access using DBMS auditing. Proceedings, 2008 ACM Symposium on Applied Computing , 1013–1020, 2008.
45. Yaseen, Q., and Panda, B. 2009. Knowledge acquisition and insider threat prediction in relational database systems. Proceedings, 12th International IEEE Conference on Computational Science and Engineering, 450–455, 2009.
46. Chung, C.Y., Gertz, M., and Levitt, K. DEMIDS: A misuse detection system for database systems. Proceedings, Conference on Integrity and Internal Control in Information Systems, 159–178, 1999.

47. Hu, Y., and Panda, B. A data mining approach for database intrusion detection. Proceedings, 2004 ACM symposium on Applied computing, pp. 711–716, 2004.
48. Hu, Y., and Panda, B. Identification of malicious transactions in database systems. Proceedings, 7th International Symposium on Database Engineering and Applications, pp. 329–335, 2003.
49. Lee, S.Y., Low, W.L. and Wong, P.Y. Learning fingerprints for a database intrusion detection system. ESORICS , 2502/2002, 264–279, 2002.
50. Lee, V., Stankovic, J.A., and Son, S.H. Intrusion detection in real-time database systems via time signatures. Proceedings, 6th IEEE Real Time Technology and Applications Symposium, 2000.
51. Valeur, F., Mutz, D., and Vigna, G . A learning-based approach to the detection of SQL attacks. Proceedings, 14th Conference on Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA), Vienna, Austria, 2005.
52. Spalka, E., and Lehnhardt, J. A comprehensive approach to anomaly detection in relational databases. Annual Working Conference on Data and Applications Security, 3654/2005, 207–211, 2005.
53. Srivastava, A., Sural, S., and Majumdar, A.K. Database intrusion detection using weighted sequence mining. Journal of Computers, 1(4), 8–17, 2006.
54. Wenhui, S., and Tan, D. A novel intrusion detection system model for securing web-based database systems. Proceedings, 25th IEEE International Computer Software and Applications Conference on Invigorating Software Development, 2001.
55. Cohen, W.W. Learning rules that classify e-mail. Proceedings, AAAI Symposium on Machine Learning in Information Access , 18–25, 1996.
56. Cohen, W.W., and Singer, Y. Context-sensitive learning methods for text categorization, 1999, ACM Transactions on Information Systems (TOIS), 17(2), 11–17.
57. Helfman, J., and Isbell, C. 1995. Ishmail: Immediate Identification of Important Information . Technical Report, AT&T Labs, 1995.
58. Rennie, J. ifile: an application of machine learning to e-mail filtering. Proceedings, KDDWorkshop on Text Mining, 2000.
59. Drucker, H., Wu, D., and Vapnik, V.N. Support vector machines for spam categorization. IEEE Transactions on Neural Networks, 10(5), 1048–1054, 1999.
60. Androutsopoulos, I., Koutsias, J., Cbandrinos, K.V., and Spyropoulos, C.D. An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages. Proceedings, 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval , 160–167, 2000.

61. Hovold, J. 2005. Naive Bayes span filtering using word-position-based attributes. Proceedings, 2nd Conference on Email and Anti-Spam, 2005.
62. Sahami, M., Dumais, S., Heckerman, D., and Horvitz, E. A Bayesian approach to filtering junk email. AAAI-98 Workshop on Learning for Text Categorization, 1998.
63. Salton, G., and McGill, M.J. Introduction to Modern Information Retrieval . McGraw-Hill, Inc., New York, USA, 1986.
64. Schenker, A. Graph-Theoretic Techniques for Web Content Mining . PhD thesis, University of South Florida, 2003.
65. Kalyan, C., and Chandrasekaran, K. Information leak detection in financial emails using mail pattern analysis under partial information. Proceedings, 7th Conference on 7th WSEAS International Conference on Applied Informatics and Communications , 104–109, 2007.
66. Carvalho, V.R., and Cohen, W. Preventing information leaks in email. Proceedings, SIAM International Conference on Data Mining, 2007.
67. Carvalho, V.R., and Balasubramanyan, R. Information leaks and suggestions: a case study using Mozilla Thunderbird. Proceedings, 6th Conference on Email and Anti-Spam, 2009.
68. Caputo D.D., Stephens G.D., and Maloof M.A. Detecting insider theft of trade secrets. IEEE Security and Privacy , 7(6), 14–21, 2009.
69. Forte, D. Do encrypted disks spell the end of forensics? Computer Fraud and Security, 18–20, 2009.
70. Parno, B., McCune, J.M., Wendlandt, D., Andersen, D.G., and Perrig, A. CLAMP: practical prevention of large-scale data leaks. Proceedings, IEEE Symposium on Security and Privacy, 2009.
71. Yasuhiro, K., and Yoshiki, S. A Web-based system for prevention of information leakage (poster). Proceedings, 11th International World Wide Web (WWW) Conference, 2002.
72. Kabra, G., Ramamurthy, R., and Sudarshan, S. Redundancy and information leakage in fine-grained access control. Proceedings, 2006 ACM SIGMOD international conference on Management of data, 2006.
73. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., and Samarati, P. Encryption policies for regulating access to outsourced data. ACM Transactions on Database Systems , 35(2), 2010, 12:2–12:46.
74. Byers, S. Information leakage caused by hidden data in published documents. IEEE Security and Privacy , 2(2), 2004, 23–27.
75. Franqueira,V., Cleeff, A., Eck, P., and Wieringa, R. External insider threat: a real security challenge in enterprise value webs. Proceedings, 5th International Conference on Availability, Reliability and Security, 446–453,.

76. Menahem, E., Shabtai, A., Rokach, L., and Elovici, Y. Improving malware detection by applying multi-inducer ensemble. Computational Statistics and Data Analysis , 53(4), 2009, 1483–1494.
77. Fung, B., Wang, K., Chen, R., and Yu, P. Privacy-preserving data publishing: A survey of recent developments. ACM Computing Surveys (CSUR) 42(4), 1–53, 2010.
78. Lefevre, K., Dewitt, D. J., and Ramakrishnan, R. Incognito: effi cient full-domain k -anonymity. Proceedings, of ACM SIGMOD . Baltimore, ML, 2005, 49–60
79. Lefevre, K., Dewitt, D. J., and Ramakrishnan, R.. Mondrian multidimensional k -anonymity. Proceedings, 22nd IEEE International Conference on Data Engineering (ICDE). Atlanta, GA, 2006.
80. Samarati, P. Protecting respondents' identities in microdata release. IEEE Transactions on Knowledge and Data Engineering (TKDE) 13(6), 2001, 1010–1027.
81. Sweeney, L. k -anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems . 10, 2002b, 557–570.
82. Agrawal, R., Imieliniski, T., Swami, A. Mining association rules between sets of items in large databases. SIGMOD Records 22(2), 1993, 207–216.
83. Fung, B.C.M., Wang, K., and Yu, P.S. Top-down specialization for information and privacy preservation. Proceedings, 21st IEEE International Conference on Data Engineering (ICDE), Tokyo, Japan, 2005, 205–216.
84. Fung, B.C.M., Wang, K., and Yu, P.S. Anonymizing classification data for privacy preservation. IEEE Transactions on Knowledge and Data Engineering (TKDE) 19(5), 2007, 711–725.
85. Iyengar, V.S. Transforming data to satisfy privacy constraints. Proceedings, 8th ACM SIGKDD . Edmonton, AB, Canada, 2002, 279–288.
86. Wong, R.C.W., Li., J., Fu, A.W.C., and Wang, K. (a ,k)-anonymity: An enhanced k –anonymity model for privacy preserving data publishing. Proceedings, 12th ACM SIGKDD . Philadelphia, PA, 2006, 754–759.
87. Xu, J., Wang, W., Pei, J., Wang, X., Shi, B., and Fu, A. W.C. Utility-based anonymization using local recoding. Proceedings, 12th ACM SIGKDD Conference . Philadelphia, PA, 2006.
88. Wang, K., Fung, B., and Yu, P. Template-based privacy preservation in classication problems. Proceedings, Fifth IEEE International Conference on Data Mining, 2005.
89. Zhang, Q., Koudas, N., Srivastava, D., and Yu, T. Aggregate query answering on anonymized tables. In Proceedings of the 23rd IEEE International Conference on Data Engineering (ICDE), 2007.

90. Domingo-Ferrer, J. Privacy-Preserving Data Mining: Models and Algorithms. Springer, Chapter A Survey of Inference Control Methods for Privacy-Preserving Data Mining, 53–80, 2008.
91. Harel, A., Shabtai, A., Rokach, L., and Elovici Y. M-score: estimating the potential damage of data leakage incident by assigning misuseability weight. Proceedings, 2010 ACM workshop on Insider threats, 2010.
92. Berkovich, M., Renford, M., Hansson, L., Shabtai, A., Rokach, L., and Elovici, Y. HoneyGen: an automated honeytokens generator, Proceeding, IEEE Intelligence and Security Informatics (ISI 2011), Beijing, China, July 10–12, 2011.
93. Duncan, K., and Wells, D.. Rule-based data cleansing. Journal of Data Warehousing , 4(3), 1999, 2–15.
94. Yahalom, R., Shmueli, E., and Zrihen, T. Constrained anonymization of production data: a constraint satisfaction problem approach. Secure Data Management, 2010, pp. 41–53.
95. Zilberman, P., Shabtai, A., Rokach, L. Analyzing group communication for preventing accidental data leakage via email. Proceedings, Workshop on Collaborative Methods for Security and Privacy (CollSec 2010), Washington DC, USA, August 10, 2010.
96. Yin Fan, Wang Yu, Wang Lina, Yu Rongwei, A Trustworthiness-Based Distribution Model for Data Leakage Prevention, Wuhan University Journal of Natural Sciences, Vol.15 No.3, 205-209, Hubei, China, 2010.
97. Bonatti P, Vimercati S D C D, Samarati P. An algebra for composing access control policies [J]. ACM Transactions on Information and System Security, 2002, 5(1): 1-35.
98. Zhang Xiaofei, Xu Fang, Shen Changxiang. Research on multilevel security model based on trustworthy state and its application [J]. Acta Electronica Sinica, 2007, 35(8): 1511-1515(Ch).
99. Jajodia P, Samarati P, Sapino M L, et al. Flexible support for multiple access control policies [J]. ACM Transactions on Database Systems, 2001, 26(2): 214-260.
100. Cheng P C, Rohatgi P, Keser C, et al. Fuzzy multi-level security: An experiment on quantified risk-adaptive access control [C] // 2007 IEEE Symposium on Security and Privacy. Oakland: The IEEE Computer Society Press, 2007, 222-230.
101. Jason Program Office. Horizontal Integration: Broader Access Models for Realizing Information Dominance [R]. Mclean: MITRE Corporation, 2004.
102. Srivatsa M, Balfe S, Paterson K G, et al. Trust management for secure information flows [C] // 15th ACM Conference on Computer and Communications Security. Alexandria: ACM Press, 2008: 175-188.

103.Papadimitriou P, Garcia-Molina H. A model for data leakage detection[C] // 25th International Conference on Data Engineering. Shanghai: IEEE Computer Society Press, 2009: 1307-1310.
104.Jared Thorkelson, Five Steps to Data Loss Prevention, DLP Experts, 2008.
105. Valli C., Honeypot technologies and their applicability as a strategic internal countermeasure. International Journal of Information and Computer Security, 2005: 1(4), 30–436