ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Π.Μ.Σ.: «Τεχνοοικονομική Διοίκηση και Ασφάλεια Ψηφιακών Συστημάτων»

ΚΑΤΕΥΘΥΝΣΗ: «Ασφάλεια Ψηφιακών Συστημάτων»

(ΨΣ-ΑΦ-888) – Μεταπτυχιακή Διπλωματική Εργασία

ΘΕΜΑ: «BIG DATA SECURITY»

Μεταπτυχιακός Φοιτητής:

Παυλόπουλος Δημήτριος ΜΤΕ 1530 jim.7.pavlop@ssl-unipi.gr

Πειραιάς, 22 Μαΐου 2017

Σελίδα Σκόπιμα Κενή

Επιβλέπων Καθηγητής:

Λαμπρινουδάκης Κωνσταντίνος

Αναπληρωτής Καθηγητής

Εξεταστική Επιτροπή:

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22 Μαΐου 2017.

(Υπογραφή)

Λαμπρινουδάκης Κωνσταντίνος
Αναπληρωτής Καθηγητής

(Υπογραφή)

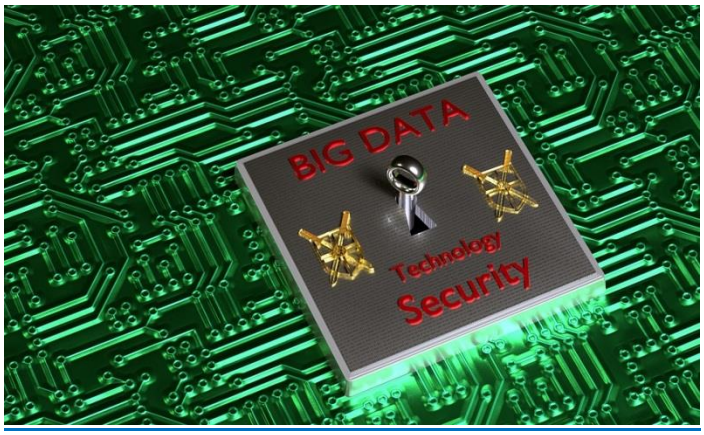
Ξενάκης Χρήστος
Αναπληρωτής Καθηγητής

(Υπογραφή)

Νταντογιάν Χριστόφορος
Αναπληρωτής Καθηγητής

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Π.Μ.Σ.: «Τεχνοοικονομική Διοίκηση και Ασφάλεια Ψηφιακών Συστημάτων»

ΚΑΤΕΥΘΥΝΣΗ: «Ασφάλεια Ψηφιακών Συστημάτων»

Μεταπτυχιακός Φοιτητής:

Παυλόπουλος Δημήτριος ΜΤΕ 1530 jim.7.pavlop@ssl-unipi.gr

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα μεταπτυχιακή διπλωματική εργασία εκπονήθηκε στο πλαίσιο του ΠΜΣ «Τεχνοοικονομική Διοίκηση και Ασφάλεια Ψηφιακών Συστημάτων» Κατεύθυνση «Ασφάλεια Ψηφιακών Συστημάτων» του τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιά.

Αισθάνομαι την υποχρέωση να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα Αναπληρωτή Καθηγητή **κ. Κωνσταντίνο Λαμπρινουδάκη**, τόσο για την εμπιστοσύνη που μου έδειξε αναθέτοντας μου την εκπόνηση της παρούσας μεταπτυχιακής διπλωματικής εργασίας, όσο και για την πολύτιμη καθοδήγησή του καθ' όλη τη διάρκεια της εκπόνησή της.

Παράλληλα θα ήθελα να ευχαριστήσω όλους τους καθηγητές του μεταπτυχιακού προγράμματος σπουδών, αλλά και τους λοιπούς προσκεκλημένους καθηγητές, για το σύνολο των πολύτιμων γνώσεων που μου μεταλαμπάδευσαν καθ' όλη τη διάρκεια των μεταπτυχιακών μου σπουδών μέσα από τις διαλέξεις των διδασκόμενων μαθημάτων τους στα πλαίσια του ΠΜΣ «Τεχνοοικονομική Διοίκηση και Ασφάλεια Ψηφιακών Συστημάτων» Κατεύθυνση «Ασφάλεια Ψηφιακών Συστημάτων».

Επιπλέον θα ήθελα να ευχαριστήσω τους συναδέλφους μου, τους συμφοιτητές μου και τους φίλους μου, για την οιανδήποτε βοήθειά τους και συμβολή τους κατά την εκπόνηση της παρούσας μεταπτυχιακής διπλωματικής εργασίας μου.

Αφιερώνω την παρούσα μεταπτυχιακή διπλωματική εργασία **στους γονείς μου** για την αδιάλειπτη βοήθεια τους σε κάθε βήμα της ζωής μου και τους ευχαριστώ γιατί, μεταξύ πολλών άλλων, μου εμφύσησαν τη «δίψα» για μάθηση, πρόοδο και μου έμαθαν πώς να είμαι σωστός άνθρωπος και χρήσιμο μέλος του κοινωνικού συνόλου. Επίσης, την αφιερώνω στα αδέρφια μου Ιωάννη και Αγγελική, αλλά και στην σύζυγό μου Κωνσταντίνα, για την συνεχή ενθάρρυνσή τους κατά τη διάρκεια των μεταπτυχιακών σπουδών μου.

Παυλόπουλος Δημήτριος

Πειραιάς, 22 Μαΐου 2017

ΠΕΡΙΛΗΨΗ

Η τεχνολογική επανάσταση, που έχει συντελεστεί στον τομέα της πληροφορικής, έχει προκαλέσει ραγδαίες εξελίξεις σε ποικίλους τομείς της σύγχρονης ζωής. Είναι γεγονός ότι ένας τεράστιος όγκος πληροφοριών παράγεται καθημερινά από πλήθος πηγών, όπως τα κινητά τηλέφωνα, τις έξυπνες συσκευές, τους φορητούς και σταθερούς υπολογιστές, τους διάφορους αισθητήρες, τις επιστημονικές συσκευές, τους δορυφόρους, τις κάμερες και τα μέσα κοινωνικής δικτύωσης. Έτσι, ο όρος «Big Data» χρησιμοποιείται για να περιγράψει τα μεγάλα σε όγκο ποικιλόμορφα δεδομένα που παράγονται, συνήθως σε υψηλούς ρυθμούς, από διαφορετικές και μεταξύ τους ανεξάρτητες πηγές. Μάλιστα, τα δεδομένα αυτά θεωρούνται υψηλής αξίας, καθώς από την ανάλυσή τους με τα «Big Data Analytics» εξάγεται πολύτιμη γνώση, η οποία μπορεί να αξιοποιηθεί σε διάφορους τομείς της σύγχρονης ζωής, όπως το εμπόριο, την επιστήμη, την υγεία, την έρευνα, τον επιχειρηματικό τομέα και την εσωτερική ασφάλεια. Αποτέλεσμα αυτού είναι τα «Big Data» να χρησιμοποιούνται ευρέως στην διαδικασία λήψης τεκμηριωμένων αποφάσεων. Ως εκ τούτου, προκύπτει η ανάγκη της διαχείρισης και της επεξεργασίας του ποικιλόμορφου αυτού όγκου δεδομένων μέσω της συλλογής, της αποθήκευσης, του διαμοιρασμού, της ανάλυσης και της ερμηνείας τους. Εντούτοις, παρά τα οφέλη, στο ιδιαίτερο αυτό περιβάλλον των «Big Data» υποβόσκουν και κίνδυνοι για την ιδιωτική ζωή των πολιτών και τα προσωπικά τους δεδομένα.

Σκοπός λοιπόν της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι, αφενός η διερεύνηση των ζητημάτων που άπτονται της ασφάλειας στα συστήματα «Big Data» και αφετέρου η μελέτη των θεμάτων που σχετίζονται με την προστασία της «Ιδιωτικότητας» (Privacy) στο περιβάλλον αυτό. Συγκεκριμένα, αναλύεται κατ' αρχάς η έννοια των «Big Data» και στην συνέχεια εξετάζονται τα ζητήματα ασφαλείας, τόσο της υποδομής και των τεχνολογιών που χρησιμοποιούνται σε αυτή, όσο και της διαδικασίας που ακολουθείται για την επεξεργασία των δεδομένων. Έπειτα, αφού πρώτα οριστεί η έννοια της «Ιδιωτικότητας» και παρουσιαστεί η ευρωπαϊκή και ελληνική νομοθεσία που την διέπει, διερευνώνται τα ζητήματα που σχετίζονται με την προστασία της στο περιβάλλον αυτό. Τέλος, εξετάζεται το επίπεδο προστασίας των προσωπικών δεδομένων των Ελλήνων πολιτών από την εφαρμογή

της «Ανοιχτής Διακυβέρνησης» (Open Government) στην Ελληνική Δημόσια Διοίκηση.

Συνεπώς, η παρούσα μεταπτυχιακή διπλωματική εργασία συνεισφέρει στην σφαιρική μελέτη των ζητημάτων που σχετίζονται, τόσο με την «Ασφάλεια» των «Big Data», όσο και με την προστασία της «Ιδιωτικότητας» στο περιβάλλον αυτό. Μάλιστα, στις μελλοντικές προεκτάσεις της είναι, αφενός η σε βάθος εξέταση των διαφορετικών υφιστάμενων λύσεων που έχουν προταθεί για την επίλυση των προβλημάτων που έχουν προκύψει, έτσι ώστε αυτές να καταστούν αποτελεσματικές στην προσπάθειά τους να τα αντιμετωπίσουν, και αφετέρου στην εύρεση νέων, προκειμένου να επιλυθούν τα μέχρι σήμερα ανοιχτά ζητήματα.

Λέξεις – Κλειδιά: Μαζικά Δεδομένα, Ασφάλεια, Ασφάλεια εκ Σχεδιασμού, Ιδιωτικότητα, Ιδιωτικότητα εκ Σχεδιασμού, Απειλές, Προκλήσεις, Ευπάθειες, Ζητήματα Ασφαλείας, Μελέτη Περίπτωσης Ιδιωτικότητας.

ABSTRACT

Various fields of our modern life have been developed at an imprecedent rate, due to the technological revolution in the domain of «Information Technology». In fact, huge amounts of information are generated daily at an incredible speed from a variety of sources, such as mobile phones, smart devices, laptops, desktops, various sensors, scientific equipment, satellites, cameras and the social media. As a result, the term «Big Data» is used to describe the massive, fast and diverse data produced by diverse and independent sources, which in some cases may be generated in high rate. Furthermore, their analysis with the «Big Data Analytics» procedure transforms them into valuable knowledge, which is considered to be of high value in several domains of our modern life, such as commerce, science, health, research, business and internal security. Consequently, they are used widely in the informed decision making process. Hence, the need for the management and the processing of this diverse volume of data arises. This can be achieved through the collection, the storage, the sharing, the analysis and the interpretation of these data. However, apart from its benefits, several risks concerning citizen's privacy underlie in this special environment of «Big Data».

Therefore, the scope of the MSc thesis is the examination of «Big Data» technologies from a security perspective and the conduction of a privacy risk analysis in this environment. Specifically, after having analyzed the concept of «Big Data», the security issues related to both the infrastructure and the technologies used in it and the procedure followed for the processing of these data are examined. Then, after having established the concept of «Privacy» and having presented the European and the Greek legislation protecting it, the issues related to its protection in the environment of «Big Data» are being investigated. Finally, the possibility that a leakage of a Greek citizen's personal data may occur due to «Open Government» in Greek Public Administration is under research.

Consequently, this MSc thesis contributes to the comprehensive study of the issues related to both the field of «Big Data Security» and the field of «Privacy Security» in this environment. Furthermore, its future extensions are at first the in depth examination and improvement of the various existing solutions, so as to make

them effective in their effort to minimize the abovementioned security and privacy issues, and secondly finding new ones to address those issues that still remain open and unsolvable.

Keywords: Big Data, Security, Security by Design, Privacy, Privacy by Design, Threats, Challenges, Vulnerabilities, Security Issues, Privacy Case Study.

1. ΚΕΦΑΛΑΙΟ 1ο: ΕΙΣΑΓΩΓΗ

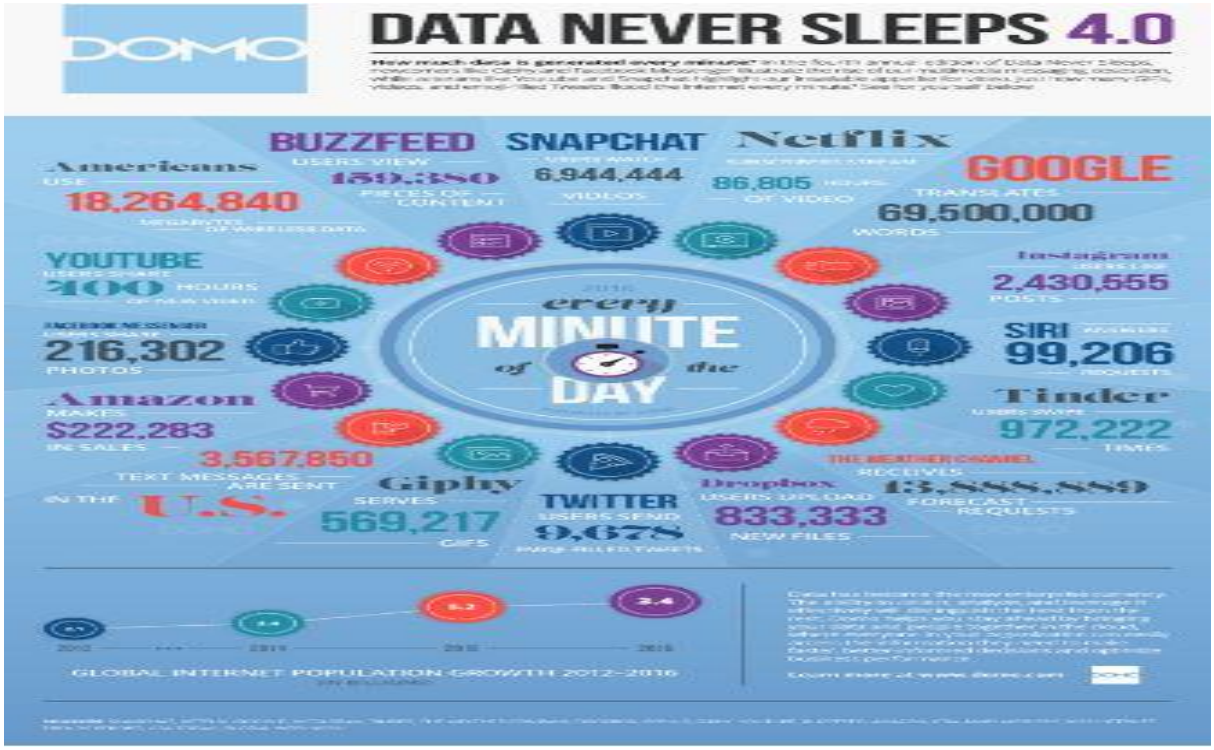
1.1 ΠΡΟΛΟΓΟΣ

Κατά την διάρκεια των τελευταίων ετών, τα «Big Data» (Μαζικά Δεδομένα) αποτελούν ένα ραγδαία εξελισσόμενο πεδίο εφαρμογής. Στο πεδίο αυτό καινοτόμες τεχνολογίες έχουν εισάγει νέους τρόπους για την αποτελεσματική διαχείριση του τεράστιου όγκου δεδομένων που παράγεται, στην πλειονότητα των περιπτώσεων σχεδόν σε πραγματικό χρόνο, από μια πληθώρα πηγών, όπως τους διάφορους αισθητήρες του «Internet of Things» (IoT), τα μέσα κοινωνικής δικτύωσης, τις κινητές ή/και σταθερές συσκευές, τις διαφόρων ειδών κάμερες, τους δορυφόρους, τις συσκευές καταγραφής κίνησης, τις συσκευές εντοπισμού θέσης (Global Position Services - GPS) κλπ. Αποτέλεσμα αυτού είναι και η εξέλιξη του πεδίου των «Big Data Analytics» (Αναλυτική Μαζικών Δεδομένων), το οποίο προσφέρει καινοτόμες μεθοδολογίες για την επαναχρησιμοποίηση αυτού του μεγάλου όγκου δεδομένων που παράγεται καθημερινά, προκειμένου να εξαχθεί πολύτιμη πληροφορία από το «πληροφοριακό χάος». Έτσι, ξεφεύγοντας από τις περιορισμένες δυνατότητες των σχεσιακών μοντέλων δεδομένων που αξιοποιούντο μέχρι σήμερα, απώτερος στόχος είναι να εντοπιστούν νέοι συσχετισμοί ή να συλληφθούν νέες και απροσδόκητες χρήσεις των δεδομένων αυτών. Μάλιστα, αυτά τα πεδία εφαρμογής προσφέρουν μια εντελώς νέα εποχή ευκαιριών και δυνατοτήτων σε ποικίλους τομείς της ζωής του σύγχρονου ανθρώπου, από το εμπόριο και τις «online» συναλλαγές, μέχρι την επιστήμη, την υγεία, την έρευνα και την ασφάλεια, αλλάζοντας σημαντικά την ποιότητα της καθημερινής ζωής.

Οπότε, ο όρος «Big Data» αναφέρεται γενικά στον τεράστιο όγκο ψηφιακών πληροφοριών που παράγεται καθημερινά από τους ανθρώπους και το περιβάλλον τους και ο οποίος αποθηκεύεται από τις διάφορες εταιρείες και τις κυβερνήσεις για περαιτέρω επεξεργασία και εκμετάλλευση. Ενδεικτικά, προκειμένου να γίνει αντιληπτός ο όγκος των δεδομένων αυτών, παρατίθενται κάποια στατιστικά στοιχεία από διάφορες πηγές δεδομένων. Έτσι, σύμφωνα με τον [160], η ποσότητα των δεδομένων που ήταν αποθηκευμένη στο «Internet» το 2010 ανερχόταν στα 500 δισεκατομμύρια «Gigabytes» και αναμενόταν να διπλασιαστεί μέσα στο επόμενο έτος. Ενώ ο [215] υπολόγισε ότι το 2012 παρήχθησαν 2500 «Exabytes» δεδομένων

και θεωρεί ότι το 2020 η ποσότητα αυτή θα ανέλθει στα 40.000 «Exabytes». Μάλιστα εκτιμά ότι η ποσότητα των δεδομένων που παράγεται θα διπλασιάζεται κάθε δύο χρόνια. Συμπληρωματικά, ο [216] υπολογίζει ότι η ποσότητα που θα τυγχάνει επεξεργασίας από τα συστήματα «Big Data» θα διπλασιάζεται κάθε δυο χρόνια.

Επιπλέον, πρόσφατες στατιστικές έρευνες μαρτυρούν την ραγδαία παραγωγή δεδομένων στα διάφορα πληροφοριακά συστήματα. Έτσι, σύμφωνα με τους [158] και [159], κάθε λεπτό η «Google» δέχεται 4 εκατομμύρια επερωτήσεις (Queries), οι χρήστες του ηλεκτρονικού ταχυδρομείου στέλνουν περισσότερα από 200 εκατομμύρια μηνύματα, οι χρήστες του «YouTube» ανεβάζουν 72 ώρες νέων βίντεο, οι χρήστες του «Facebook» μοιράζονται περισσότερα από 2,4 εκατομμύρια περιεχομένου και οι χρήστες του «Twitter» παράγουν 277 χιλιάδες σχόλια (Εικόνα 1).



ΕΙΚΟΝΑ 1: DOMO – DATA NEVER SLEEPS 4.0

Παράλληλα, στον [6] αναφέρεται ότι, σύμφωνα με την «IBM», 2.5 πεντάκις εκατομμύρια «Bytes» (2.5 Quintillion) δεδομένων παράγονται ημερησίως. Επίσης, άλλες στατιστικές αναφέρουν ότι το «YouTube» έχει πάνω από 1 δισεκατομμύριο χρήστες, οι οποίοι αντιστοιχούν στο 1/3 των χρηστών του «Internet» και οι οποίοι παρακολουθούν ημερησίως περίπου 5 δισεκατομμύρια βίντεο, ανεβάζουν περί τις 300 ώρες βίντεο κάθε λεπτό και παράγουν δισεκατομμύρια σχόλια [161], [162].

Αντίστοιχα στο «Facebook» κάθε 20 λεπτά αποστέλλονται 3 εκατομμύρια μηνύματα, διαμοιράζονται περίπου 1 εκατομμύριο σύνδεσμοι και ο συνολικός αριθμός των φωτογραφιών που είναι αποθηκευμένος ανέρχεται στις 10 εκατομμύρια [163]. Ενώ στο «Twitter» παράγονται ημερησίως κατά μέσο όρο 58 εκατομμύρια «tweets» [164] και στο «Instagram» οι χρήστες έχουν δημοσιεύσει συνολικά περί τις 35 δισεκατομμύρια φωτογραφίες, αναρτώντας ημερησίως κατά μέσο όρο περίπου 52 εκατομμύρια φωτογραφίες [165]. Επιπρόσθετα, το «Χρηματιστήριο της Νέας Υόρκης» παράγει ημερησίως 1 «Terabyte» χρηματοοικονομικών δεδομένων, το πρόγραμμα «Ancestry.com» έχει αποθηκευμένα δεδομένα όγκου περίπου 2.5 «Petabytes», στο «Internet» είναι αποθηκευμένα περίπου 2 «Petabytes» δεδομένων με τον ρυθμό αποθήκευσης να αυξάνει κατά 20 «Terabytes» ανά μήνα και το πρόγραμμα «Large Hadron Collider» στη Γένοβα της Ελβετίας παράγει ετησίως 15 «Petabytes» δεδομένων [7].

Από τα παραπάνω εξάγεται το συμπέρασμα ότι τα «Big Data» έχουν εισβάλει σε όλους τους τομείς της σύγχρονης ζωής για τα καλά και ότι αναμένεται να την επηρεάσουν σε πολύ μεγάλο βαθμό. Παρόλα αυτά όμως, ενώ τα αποτελέσματα από την εφαρμογή των «Big Data» και «Big Data Analytics» είναι στην πλειονότητα των περιπτώσεων καλοδεχούμενα από το σύνολο των ανθρώπων, όπως για παράδειγμα όταν υποστηρίζουν την πρόβλεψη της κλιματικής αλλαγής, ή την εξάπλωση κάποιας επιδημίας, ή ακόμα και των παρενεργειών κάποιου φαρμάκου, εντούτοις συχνά εγείρονται σοβαρές ανησυχίες που σχετίζονται με την προστασία, τόσο της ιδιωτικής ζωής, όσο και των προσωπικών δεδομένων των ατόμων. Καθώς λοιπόν τα «Big Data» και «Big Data Analytics» ήρθαν για να μείνουν, δεν μπορεί να γίνει αποδεκτό ότι αυτό θα γίνει εις βάρος της ιδιωτικής ζωής. Έτσι, με δεδομένο ότι οι τεχνολογίες αυτές και οι καινοτομίες που εισάγουν δεν θα πρέπει να παρεμποδιστούν και να μπλοκαριστούν, κρίνεται υψίστης σημασίας να βρεθεί η χρυσή τομή μεταξύ της χρήσης αυτών των τεχνολογιών, έτσι ώστε να απολαμβάνουμε τα πλεονεκτήματα που μας παρέχουν, και της απαίτησης για την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων των ατόμων. Συνεπώς, προκειμένου να επιτευχθεί η σύγκλιση των δυο αυτών αντιτιθέμενων καταστάσεων, θα πρέπει να διενεργηθεί έρευνα σε βάθος, έτσι ώστε να καταστούν οι έννοιες της «Ασφάλειας» (Security) και της «Προστασίας της Ιδιωτικότητας» (Privacy Preserving) βασικές αρχές κατά την σχεδίαση και ανάπτυξη των συστημάτων «Big Data».

1.2 ΟΡΙΣΜΟΣ «ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ» (BIG DATA) ΚΑΙ «ΑΝΑΛΥΤΙΚΗΣ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ» (BIG DATA ANALYTICS)

Στην ενότητα αυτή λοιπόν κρίνεται σκόπιμο, πριν την διερεύνηση των εννοιών της «Ασφάλειας» και της «Προστασίας της Ιδιωτικότητας» στα συστήματα «Big Data», να παρουσιαστεί σφαιρικά το νέο αυτό περιβάλλον, καθώς η περαιτέρω μελέτη του προϋποθέτει την πλήρη κατανόησή του. Αρχικά παρατίθενται οι ορισμοί των «Big Data» και «Big Data Analytics», ακολουθεί η περιγραφή της διαδικασίας επεξεργασίας των δεδομένων αυτών και τέλος παρουσιάζεται η αρχιτεκτονική των συστημάτων αυτών. Η ενότητα ολοκληρώνεται με τον σκοπό και την διάρθρωση της παρούσας μεταπτυχιακής διπλωματικής εργασίας.

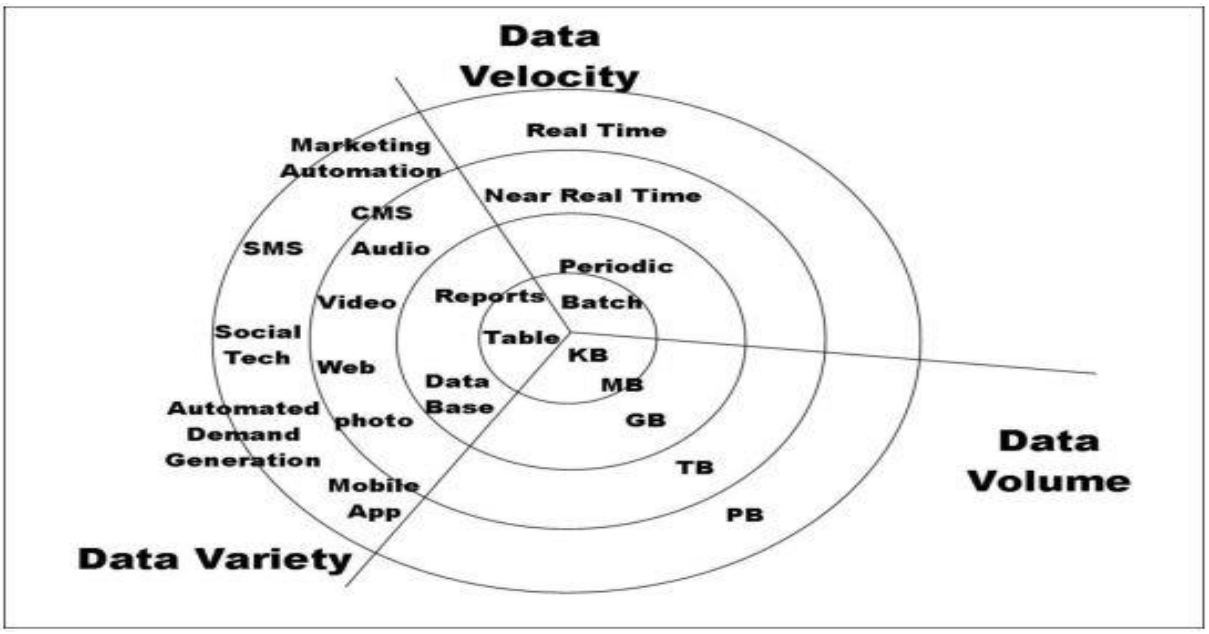
1.2.1 ΟΡΙΣΜΟΣ «ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ» (BIG DATA)

Ο όρος «Big Data» αρχικά χρησιμοποιήθηκε για να περιγράψει τον τεράστιο όγκο δεδομένων που παραγόταν από τις διάφορες ηλεκτρονικές συσκευές την εκάστοτε ψηφιακή εποχή [166]. Ήδη χρησιμοποιείται από το 1990 και ο John Mashey ήταν αυτός που συνέβαλλε σημαντικά στο να γίνει ευρέως γνωστός [14]. Αν και δεν αμφισβητείται το γεγονός ότι η ποσότητα των δεδομένων που διαχειρίζονται τα διάφορα πληροφοριακά συστήματα είναι τεράστια στην εποχή μας, το χαρακτηριστικό του όγκου τους (Volume) δεν αποτελεί και το σημαντικότερο χαρακτηριστικό τους, καθώς η τεχνολογική εξέλιξη που έχει συντελεστεί αντιμετωπίζει αποτελεσματικά τις προκλήσεις που σχετίζονται με την αποθήκευση και την επεξεργασία τους. Έτσι, το μέγεθος των δεδομένων των «Big Data» διαρκώς αυξάνεται σε κλίμακα, με αποτέλεσμα, ενώ το 2012 το μέγεθος να αναφερόταν στην τάξη των «Terabytes», τώρα να έχει φθάσει να αναφέρεται στην τάξη των «Petabytes», ή και ακόμα των «Zettabytes».

Οπότε το 2001 ο [8] όρισε τα χαρακτηριστικά των «Big Data» (Εικόνα 2), τα οποία είναι τα εξής:

- **«Increasing Volume»** (Όγκος), που αφορά στον όγκο των δεδομένων.
- **«Velocity»** (Ταχύτητα), που αφορά στην ταχύτητα με την οποία τα δεδομένα εισέρχονται και εξέρχονται από ένα σύστημα, αλλά και την ταχύτητα με την οποία αυτά τυγχάνουν επεξεργασίας και ανάλυσης από αυτό.

➢ **«Variety»** (Ποικιλομορφία), που αφορά την ποικιλία των τύπων (Structured, Semi-Structured, Unstructured) και της φύσης των δεδομένων (Documents, Databases, Images, Videos, Logs, κλπ.), αλλά και το πλήθος των διαφόρων πηγών τους.



ΕΙΚΟΝΑ 2: CHARACTERIZATION OF BIG-DATA (3V)

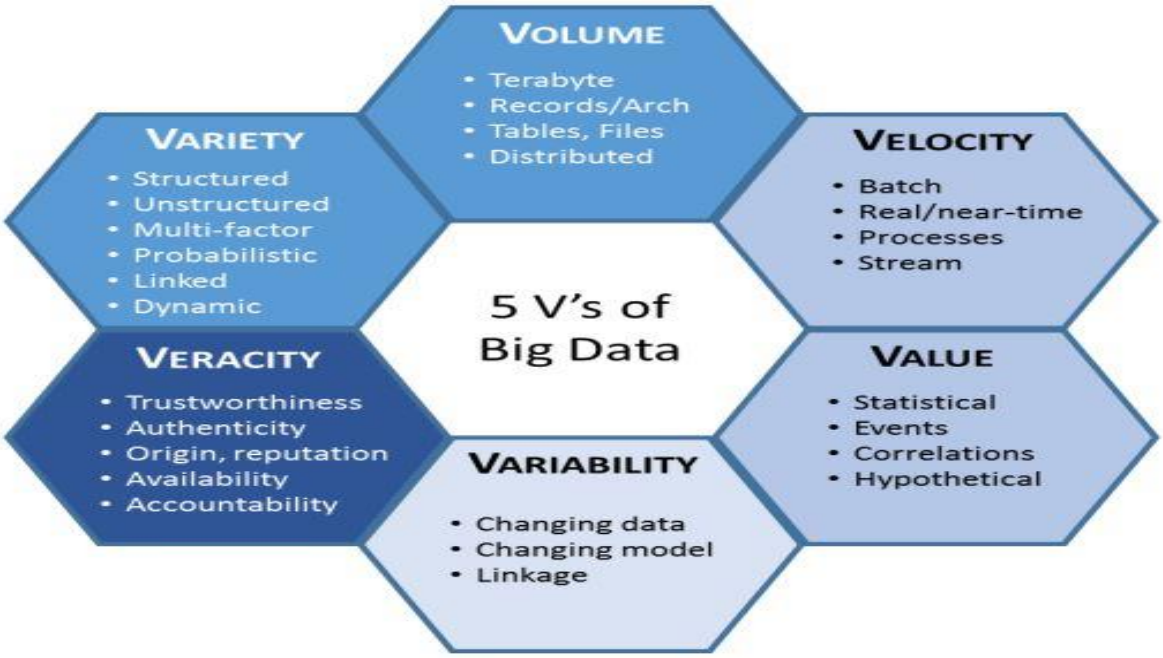
Έκτοτε οι [8], [15], [168], αλλά και ο επιχειρηματικός τομέας, χρησιμοποιούν αυτά τα τρία χαρακτηριστικά «3Vs» για να περιγράψουν τα «Big Data». Έτσι ορίζονται ως «Big Data» τα δεδομένα εκείνα τα οποία, λόγω του μεγάλου όγκου, της υψηλής ταχύτητας ή/και της ποικιλομορφίας τους, απαιτούν νέες και αποδοτικές μεθόδους και τεχνικές επεξεργασίας, προκειμένου να καταστεί δυνατή η επαύξηση των δυνατοτήτων στους τομείς της διαδικασίας ενισχυμένης λήψης απόφασης (Enhanced Decision Making), της διορατικότητας (Insight Discovery) και της βελτιστοποίησης μιας διαδικασίας (Process Optimization) [13], [167].

Επιπρόσθετα, κάποια επιπλέον «Vs» έχουν προστεθεί στην περιγραφή των «Big Data» [16], [168] (Εικόνα 3). Αυτά είναι:

➢ **«Veracity»** (Εμπιστοσύνη / Φιλαλήθεια), που αφορά την αυθεντικότητα των δεδομένων και στην εμπιστοσύνη που πρέπει να τους δοθεί, αφού η ποιότητα των δεδομένων που έχει καταγραφεί μπορεί να ποικίλλει σε μεγάλο βαθμό και να επηρεάσει την ακρίβεια μιας ανάλυσης, η οποία εξαρτάται σημαντικά από την ορθότητα, την συνέπεια και την ακρίβεια των δεδομένων μιας πηγής.

➢ **«Variability»** (Μεταβλητότητα), που αφορά το κατά πόσο οι έννοιες των δεδομένων μπορεί να αλλάξουν με την πάροδο του χρόνου, καθιστώντας τα ασυνεπή και δυσχεραίνοντας έτσι τον αποτελεσματικό χειρισμό και διαχείρισή τους.

➢ **«Value»** (Αξία), που αφορά στα δυνητικά έσοδα, οικονομικά και κοινωνικά, που μπορεί να προκύψουν από την εκμετάλλευσή τους.



ΕΙΚΟΝΑ 3: IBM - 5V OF BIG DATA

Επίσης χρησιμοποιούνται από τον επιχειρηματικό τομέα επιπλέον δυο «Vs» [11]:

➢ **«Volatility»** (Μεταβλητότητα), που αφορά στην τάση που έχουν οι δομές δεδομένων και τα δεδομένα να μεταβάλλονται με την πάροδο του χρόνου, γεγονός που καθορίζει για πόσο χρόνο αυτά θα είναι έγκυρα και μέχρι πότε θα πρέπει αυτά να τηρούνται στο σύστημα.

➢ **«Validity»** (Καταλληλότητας), που αφορά στο κατά πόσο τα δεδομένα που έχουν συλλεχθεί είναι αφενός χρήσιμα και αφετέρου αρκετά για την επίτευξη του σκοπού για τον οποίο συλλέχτηκαν.

Ενώ ο [12], πλέον των «5Vs», εισάγει και τα παρακάτω χαρακτηριστικά:

➢ **«Quality»** (Ποιότητα), που αφορά στην χρησιμότητα των δεδομένων με την ποιοτική μέτρηση των εξής χαρακτηριστικών: της Πληρότητας (Completeness), της Ακρίβειας (Accuracy), της Διαθεσιμότητας (Availability) και του κατά πόσο είναι Ενημερωμένα (Timeliness).

- **«Discovery»** (Ανακαλυψιμότητα), που αφορά στο διαχωρισμό των υψηλής αξίας και χρήσιμων δεδομένων για την εκτέλεση κάποιας εργασίας από το σύνολο των δεδομένων.
- **«Dogmatism»** (Δογματισμό), που αναφέρεται στην ανάγκη να μην υπάρχει αυστηρή προσήλωση μόνο στα αποτελέσματα, αλλά αυτά να ερμηνεύονται και να αναλύονται μέσα στο ευρύτερο πλαίσιο του εκάστοτε προβλήματος.

Καθώς το πεδίο των «Big Data» ακόμα εξελίσσεται και ερευνάται, διάφοροι εναλλακτικοί ή/και συμπληρωματικοί ορισμοί έχουν προταθεί σε μια προσπάθεια να συμπεριληφθούν οι διάφορες εκφάνσεις του. Έτσι ο [9] προκειμένου να τονίσει την εξελικτική τους φύση, όρισε ως «Big Data» τα σύνολα δεδομένων (Datasets), των οποίων το μέγεθος είναι υπεράνω της ικανότητας των τυπικών βάσεων δεδομένων και εργαλείων στο να τα συλλέξουν, να τα αποθηκεύσουν, να τα διαχειριστούν και να τα αναλύσουν. Το 2014 ο [11] όρισε τα «Big Data» ως τη νέα γενιά τεχνολογιών και αρχιτεκτονικών που σχεδιάστηκαν για να εξάγουν οικονομικό όφελος από ένα μεγάλο ποικιλόμορφο όγκο δεδομένων, χρησιμοποιώντας υψηλής ταχύτητας τεχνικές και εφαρμογές για την συλλογή, διαχείριση ή/και ανάλυσή τους.

Επιπλέον, στον [11] καταγράφονται τα χαρακτηριστικά τα οποία, στο σύνολό τους ή και εν μέρει, περιγράφουν ένα σύστημα «Big Data». Αυτά είναι:

1) Η γρήγορη εισροή δεδομένων (Fast Data Insertion) τα οποία αρχικά αποθηκεύονται και στη συνέχεια αναλύονται.

2) Το ανθεκτικό κατανεμημένο σύστημα αποθήκευσης (Distributed Redundant Data Storage), το οποίο προσδίδει ανθεκτικότητα στο σύστημα αποθήκευσης και εγγυάται την διαθεσιμότητα των δεδομένων του.

3) Η εκτέλεση παράλληλων διεργασιών (Parallel Task Processing), τόσο για την αποτελεσματική διαχείριση του όγκου των δεδομένων, όσο και για κέρδος χρόνου.

4) Οι διάφοροι τύποι δεδομένων (Different Types of Data), ήτοι δομημένων (Structured), μη δομημένων (Unstructured) και ημι-δομημένων (Semi-structured).

5) Η ευελιξία του συστήματος και η κλιμάκωσή του (Scalable), το οποίο σημαίνει ότι πολλαπλά υποσυστήματα υποστηρίζουν την λειτουργία του.

6) Η μεγάλης κλίμακας «Αναλυτική» (Large Scale Analytics), που σχετίζεται με την ανάλυση του όγκου των δεδομένων.

7) Η ανεξαρτησία από το υλικό (Hardware Agnostic), δηλαδή ότι το σύστημα λειτουργεί ανεξάρτητα από την υποδομή που το υποστηρίζει.

8) Η προσβασιμότητα (Accessible), η οποία αφορά στην δυνατότητα ενσωμάτωσης και νέων πηγών στο σύστημα, αλλά και νέων τύπων δεδομένων.

9) Το ικανοποιητικό κόστος (Cost Effective) για την υλοποίηση και την λειτουργία του.

Γενικά, θα πρέπει να επισημανθεί ότι τα «Big Data» δεν αποτελούν μια μεμονωμένη τεχνολογία, αντιθέτως είναι ένας συνδυασμός παλαιών και πολλών καινούριων τεχνολογιών που αποσκοπούν στην αποδοτική και αποτελεσματική διαχείριση των δεδομένων με τα χαρακτηριστικά που περιγράφηκαν στην ενότητα αυτή [15].

1.2.2 ΟΡΙΣΜΟΣ «ΑΝΑΛΥΤΙΚΗΣ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ» (BIG DATA ANALYTICS)

Αντίστοιχα, ο όρος «Big Data Analytics» αναφέρεται σε ολόκληρο τον κύκλο ζωής της διαχείρισης των δεδομένων. Περιλαμβάνει τις φάσεις της συλλογής, της οργάνωσης και της ανάλυσή τους και αποσκοπεί στο να ανακαλυφθούν πρότυπα και τάσεις, να συναχθεί η ολοκληρωμένη εικόνα μιας κατάστασης, να γίνουν προβλέψεις και να κατανοηθούν συμπεριφορές [3].

Συγκεκριμένα, με τον όρο «Analytics» (Αναλυτική) εννοείται η εύρεση, επεξεργασία, ανάλυση και παρουσίαση των διαφόρων μοτίβων που υπάρχουν μέσα σε ένα πλήθος δεδομένων με σκοπό να εξαχθεί νόημα και να παραχθεί πληροφορία [169]. Έτσι, η έννοια των «Analytics» (Αναλυτική) αποκτά νόημα σε τομείς όπου υπάρχουν αποθηκευμένα σύνολα δεδομένων και για τα οποία χρησιμοποιούνται εκτεταμένα οι εφαρμογές διαφόρων επιστημών, όπως των μαθηματικών, της στατιστικής, της πληροφορικής, της επιχειρησιακής έρευνας, των προγνωστικών μοντέλων και άλλων, ανάλογα με το εκάστοτε πρόβλημα που έχει τεθεί, για την αποτελεσματική επεξεργασία τους. Επιπλέον, συμπεριλαμβάνονται και οι διάφορες μέθοδοι για την οπτικοποίηση των αποτελεσμάτων της.

Από την άλλη πλευρά, ο όρος «Analysis» (Ανάλυση) αναφέρεται στην διαδικασία της μετατροπής των ακατέργαστων δεδομένων (Raw Data) σε χρήσιμες πληροφορίες (Information) [170] και περιλαμβάνει τις εξής φάσεις:

1) Του καθορισμού των απαιτούμενων δεδομένων (Data Requirements).

2) Της συλλογής τους (Data Collection).

3) Της επεξεργασίας τους (Data Processing).

4) Του καθαρισμού τους (Data Cleaning).

5) Της διερευνητικής ανάλυσής τους (Exploratory Data Analysis).

6) Του καθορισμού των μοντέλων και των αλγορίθμων που θα χρησιμοποιηθούν (Modeling and Algorithms).

7) Της εξαγωγής του αποτελέσματος (Data Product).

8) Της παρουσίασης του αποτελέσματος (Communication).

Ωστόσο, θα πρέπει να τονιστεί ότι ο όρος «Analytics» (Αναλυτική) αναφέρεται σε όλη τη μεθοδολογία που ακολουθείται για να εξαχθεί κάποια πληροφορία, ενώ ο όρος «Analysis» (Ανάλυση) αποτελεί μια διεργασία αυτής της μεθοδολογίας. Έτσι η «Data Analysis» (Ανάλυση των Δεδομένων) είναι μια διαδικασία, η οποία περιλαμβάνει την έρευνα, την εύρεση, τον καθαρισμό, τη μετατροπή και τη μοντελοποίηση των δεδομένων, με στόχο την ανακάλυψη χρήσιμων πληροφοριών για την περαιτέρω εξαγωγή συμπερασμάτων και λήψη αποφάσεων. Η έννοια της ανάλυσης των δεδομένων έχει πολλαπλές όψεις και προσεγγίσεις, περιλαμβάνει διάφορες τεχνικές και εφαρμόζεται σε διάφορους τομείς της επιστήμης, των επιχειρήσεων και στις κοινωνικές επιστήμες. Ενδεικτικά κάποιες από αυτές είναι το «Data Mining», το «Business Intelligence», το «Predictive Analytics», το «Text Analytics», κλπ (Εικόνα 4).

ΕΙΚΟΝΑ 4: EXAMPLES OF TYPES OF ANALYTICS

Συνεπώς, ο όρος «Big Data Analytics» αναφέρεται στη μεθοδολογία της ανάλυσης του μεγάλου όγκου δεδομένων που αποθηκεύονται και τυγχάνουν επεξεργασίας από τα συστήματα «Big Data» [171]. Αυτά τα δεδομένα συλλέγονται από μια μεγάλη ποικιλία πηγών, όπως τα μέσα κοινωνικής δικτύωσης, τους αισθητήρες, τα αρχεία καταγραφής ενεργειών, τις διάφορες συσκευές κλπ, και ο απώτερος στόχος των «Big Data Analytics» είναι να ανακαλυφθούν τα μοτίβα και οι διασυνδέσεις που μπορεί να υποκρύπτονται σε αυτά, προκειμένου να εξαχθούν συμπεράσματα και να ληφθούν αποφάσεις.

1.3 ΑΝΑΛΥΣΗ ΤΩΝ ΦΑΣΕΩΝ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ (BIG DATA MANAGEMENT)

Έτσι η διαδικασία για την επεξεργασία των δεδομένων στο περιβάλλον των «Big Data» (Big Data Management), σύμφωνα με τους [3], [12] και [17], περιλαμβάνει διάφορες διακριτές φάσεις (Εικόνα 5), οι οποίες είναι:

1. Λήψη Δεδομένων και Καταγραφή (Data Acquisition and Recording).

Στην φάση αυτή περιλαμβάνονται οι διαδικασίες της συλλογής των δεδομένων από τις διάφορες πηγές δεδομένων, όπως είναι οι αισθητήρες, οι κινητές συσκευές, τα μέσα κοινωνικής δικτύωσης κλπ, και της αποθήκευσή τους στις διάφορες αποθήκες δεδομένων, που μπορεί να είναι είτε σχεσιακές (Relational DBs), είτε μη-σχεσιακές (NoSQL DBs) βάσεις δεδομένων. Ζητήματα που προκύπτουν σε αυτή την φάση αφορούν, αφενός το φιλτράρισμα των εισερχόμενων δεδομένων και την συμπίεση τους, προκειμένου να καταστεί πιο αποδοτική η επεξεργασία τους, και αφετέρου την αυτόματη δημιουργία κατάλληλων μεταδεδομένων. Επίσης, ένα άλλο ζήτημα σχετίζεται με την ύπαρξη κατάλληλων μηχανισμών, που θα μπορούν να μεταφέρουν την προέλευση των δεδομένων και των μεταδεδομένων καθ' όλη την διαδικασία της ανάλυσής τους.

2. Εξαγωγή Πληροφορίας, Καθαρισμός και Σχολιασμός (Information Extraction, Cleaning and Annotation).

Στη φάση αυτή, καθώς τα δεδομένα που έχουν συλλεχθεί δεν είναι έτοιμα για ανάλυση, απαιτείται η χρήση κατάλληλων τεχνικών προκειμένου από αυτά να εξαχθεί η ζητούμενη πληροφορία η οποία θα είναι κατάλληλη για περαιτέρω ανάλυση. Ωστόσο αυτό εξαρτάται κάθε φορά από την εκάστοτε περίπτωση και επηρεάζεται σημαντικά από τα χαρακτηριστικά των «Big Data». Επιπλέον απαιτείται αυτά να καθαριστούν από τυχόν διπλές εγγραφές, ή κενές τιμές, ή λανθασμένες τιμές,

διαδικασία που είναι εξαιρετικά επίπονη, και στη συνέχεια να εισαχθούν σχόλια, έτσι ώστε να διευκολυνθεί η περαιτέρω ανάλυσή τους.

3. Συγκέντρωση Δεδομένων, Ομαδοποίηση και Αναπαράστασή τους (Data Integration, Aggregation and Representation).

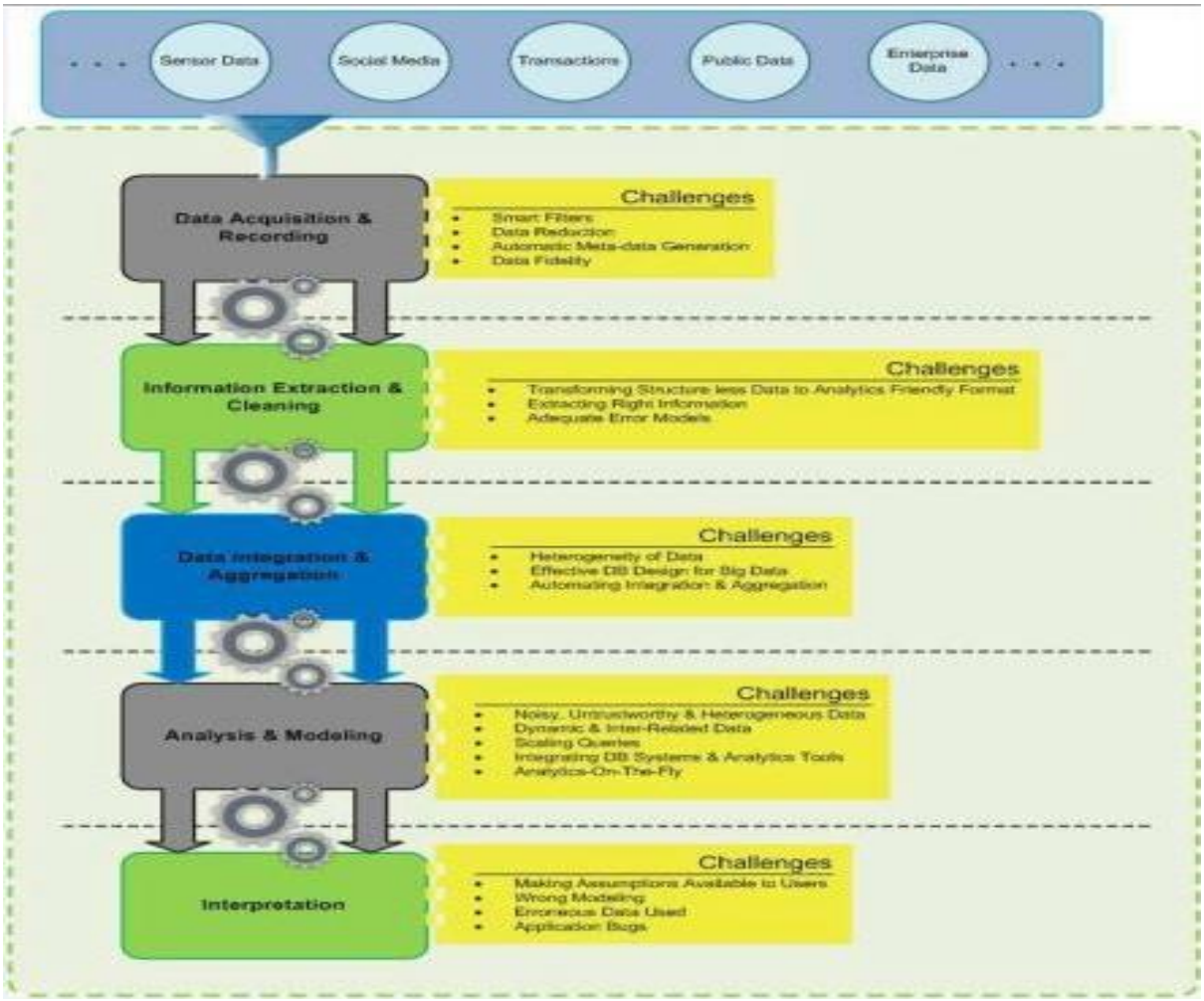
Στη φάση αυτή, τα δεδομένα του προηγούμενου βήματος συγκεντρώνονται, ομαδοποιούνται και αναπαρίστανται, έτσι ώστε να καταστούν έτοιμα για επεξεργασία. Τα ζητήματα που προκύπτουν αφορούν στην αυτοματοποιημένη και αποδοτική συγκέντρωση και ομαδοποίηση των ανομοιογενών αυτών δεδομένων, καθώς και στην αναπαράστασή τους με τρόπο που να τα καθιστά αναγνώσιμα από τον υπολογιστή που θα τα επεξεργαστεί. Επιπλέον πρόκληση είναι ο σχεδιασμός της κατάλληλης κάθε φορά βάσης δεδομένων για την αποθήκευση των δεδομένων αυτών, έτσι ώστε να μπορεί να υποστηρίξει την διαδικασία της επεξεργασίας και της ανάλυσής τους με τον αποδοτικότερο τρόπο.

4. Ανάλυση και Μοντελοποίηση (Analysis and Modeling).

Στη φάση αυτή τα «Big Data» αναλύονται με την χρήση των τεχνικών των επερωτήσεων (Querying) και της εξόρυξης δεδομένων (Data Mining), προκειμένου να ανακαλυφθούν αρχικά τα διάφορα μοτίβα και οι συσχετίσεις που μπορεί να υποκρύπτονται μέσα σε αυτά. Σκοπός είναι στην συνέχεια να εξαχθεί πολύτιμη γνώση, να διασταυρωθούν πληροφορίες, να επιβεβαιωθούν συσχετίσεις και να εντοπιστούν κρυμμένα μοντέλα ή τάσεις. Τα ζητήματα στη φάση αυτή είναι ποικίλα. Κατ' αρχάς τα δεδομένα για να αναλυθούν σωστά θα πρέπει να είναι ομοιογενή, συσχετισμένα μεταξύ τους, καθαρά, έμπιστα και εύκολα προσβάσιμα. Ωστόσο τα «Big Data» είναι ανομοιογενή, δυναμικά, αναξιόπιστα, ασυσχέτιστα και με θόρυβο. Επομένως θα πρέπει να χρησιμοποιηθούν αποτελεσματικές διαδικασίες για να μετατραπούν σε διαχειρίσιμη μορφή προκειμένου να αναλυθούν. Επίσης, για την ανάλυσή τους απαιτούνται, αφενός οι κατάλληλοι και ευέλικτοι αλγόριθμοι και αφετέρου το απαραίτητο υπολογιστικό σύστημα, που θα υποστηρίξουν τις τεχνικές των επερωτήσεων και της εξόρυξης στο νέο αυτό περιβάλλον των «Big Data». Παράλληλα με το ζήτημα αυτό, το πρόβλημα της έλλειψης αποτελεσματικής συνεργασίας των διαφόρων βάσεων δεδομένων με τις διάφορες τεχνικές ανάλυσης και επεξεργασίας των δεδομένων αυτών επιδεινώνει την κατάσταση. Τέλος, ένα άλλο ζήτημα αφορά στην υποστήριξη αναλύσεων σε πραγματικό χρόνο (On-The-Fly Analysis).

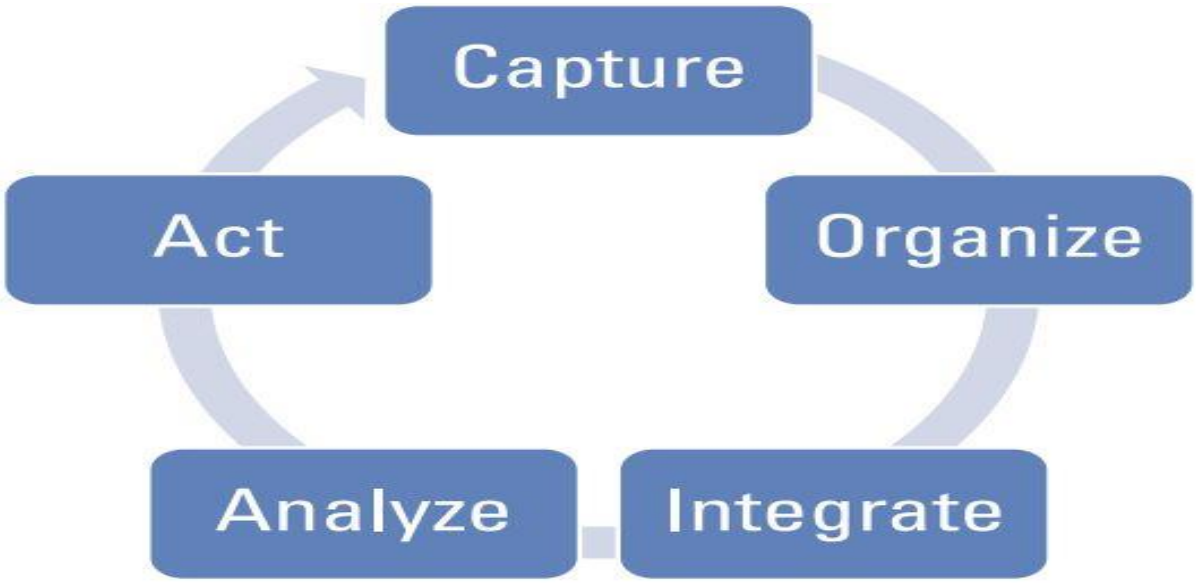
5. Ερμηνεία (Interpretation).

Στην φάση αυτή τα αποτελέσματα του προηγούμενου βήματος θα πρέπει αρχικά να ερμηνευτούν, μέσα στο πλαίσιο της εκάστοτε περίπτωσης που επέβαλε αυτή την ανάλυση, και εν συνεχεία τα αποτελέσματα της ερμηνείας τους να παρουσιαστούν στους τελικούς χρήστες μαζί τις παραδοχές υπό τις οποίες αυτή η διαδικασία εκτελέστηκε. Ένα βασικό ζήτημα που προκύπτει είναι η εξαγωγή εσφαλμένων αποτελεσμάτων που μπορεί να οφείλεται, είτε σε σφάλματα στις εφαρμογές και το σύστημα, είτε στις παραδοχές των μοντέλων ανάλυσης που χρησιμοποιήθηκαν, είτε σε εσφαλμένα δεδομένα από τις πηγές. Ένα άλλο ζήτημα αφορά στη διατήρηση και παρουσίαση στον τελικό χρήστη όλων των παραδοχών που έγιναν κατά την διαδικασία της ανάλυσης, έτσι ώστε να γνωρίζει παράλληλα με την ερμηνεία και τις συνθήκες διεξαγωγής της. Τέλος, οι τρόποι οπτικής αναπαράστασης των αποτελεσμάτων, ώστε αυτά να είναι πιο εύκολα αντιληπτά, είναι άλλο ένα ζήτημα που πρέπει να εξεταστεί.



ΕΙΚΟΝΑ 5: BIG DATA PROCESSING PIPELINE AND CHALLENGES

Μια άλλη προσέγγιση της ανωτέρω διαδικασίας είναι αυτή που ακολουθεί την λογική του **«Capture -> Organize -> Integrate -> Analyze -> Act»** (Εικόνα 6), η οποία όμως δεν διαφοροποιείται από την προηγούμενη διαδικασία [15].



ΕΙΚΟΝΑ 6: BIG DATA CYCLE PROCESS

1.4 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ (BIG DATA INFRASTRUCTURE)

Όσον αφορά την αρχιτεκτονική των συστημάτων «Big Data» έχουν προταθεί διάφορα σχήματα και απεικονίσεις. Για τις ανάγκες της παρούσας μεταπτυχιακής διπλωματικής εργασίας παρουσιάζονται στην ενότητα αυτή ενδεικτικά κάποιες από αυτές, οι οποίες θα χρησιμοποιηθούν μετέπειτα για την περαιτέρω διερεύνηση και ανάλυση των ζητημάτων που σχετίζονται με την «Ασφάλεια» και την «Προστασία της Ιδιωτικής Ζωής» στο περιβάλλον των «Big Data». Οι αρχιτεκτονικές αυτές αποτελούν ένα υψηλού επιπέδου μοντέλο το οποίο, αφενός θα βοηθήσει στο να γίνουν κατανοητά τα διάφορα μέρη από τα οποία απαρτίζονται τα συστήματα αυτά και αφετέρου θα συμβάλει στην εύρεση των απαιτήσεων ασφαλείας στο περιβάλλον αυτό.

Συγκεκριμένα, στο [2] παρουσιάζεται η αρχιτεκτονική του συστήματος «Big Data» ως μια γενική απεικόνιση πέντε επιπέδων στα οποία περιγράφονται τα διάφορα τμήματα που την απαρτίζουν (Εικόνα 7). Αναλυτικότερα, η αρχιτεκτονική αυτή περιλαμβάνει τα εξής επίπεδα:

❖ Πηγές Δεδομένων (Data Sources).

Στο επίπεδο αυτό περιλαμβάνονται οι ποικίλες πηγές δεδομένων οι οποίες παρέχουν στο σύστημα διάφορους τύπους δεδομένων, ήτοι δομημένα (Structured Data), ημι-δομημένα (Semi-Structured Data), μη-δομημένα (Unstructured Data) ή/και πραγματικού χρόνου δεδομένα (Streaming Data). Ενδεικτικά, στις πηγές του συστήματος μπορεί να συμπεριλαμβάνονται οι αισθητήρες, οι κινητές συσκευές, τα μέσα κοινωνικής δικτύωσης, τα αρχεία καταγραφής κίνησης (logs), οι κάμερες, τα διάφορα μηχανήματα στις επιχειρήσεις και την υγεία, οι δορυφόροι κλπ. Ωστόσο, θα πρέπει να σημειωθεί ότι αυτές οι πηγές δεδομένων μπορεί, είτε να είναι στην ιδιοκτησία του κατόχου του «Big Data» συστήματος, είτε όχι. Τα δεδομένα που παρέχονται σε ένα τέτοιο σύστημα είναι ποικίλα, από έγγραφα και σχεσιακά δομημένα δεδομένα, μέχρι φωτογραφίες, αρχεία ήχου, βίντεο, δεδομένα αισθητήρων κλπ.

❖ Εργαλεία και Τεχνικές για την Συγκέντρωση, Ομαδοποίηση και Αναπαράσταση των Δεδομένων των Πηγών (Integration Process).

Στο επίπεδο αυτό περιλαμβάνονται όλα τα εργαλεία και οι τεχνικές προ-επεξεργασίας των δεδομένων των πηγών του συστήματος, που ως στόχο έχουν τη συλλογή των δεδομένων από τις πηγές τους, τον καθαρισμό τους, το σχολιασμό τους και εν συνεχεία τη συγκέντρωσή τους, την ομαδοποίησή τους και την αναπαράστασή τους, προκειμένου αυτά να καταστούν ενιαία και ομοιογενή σύνολα έτοιμα να τύχουν επεξεργασίας από την διαδικασία των «Big Data Analytics» (Αναλυτικής) που θα ακολουθήσει. Στο επίπεδο αυτό βρίσκονται τεχνικές όπως η «Extract-Transform-Load» (ETL), η «Message-Based», η «Complex-Event-Processing» (CEP) για «Streaming Data», τα διάφορα «Application-Programming-Interfaces» (APIs) κλπ, και εργαλεία, όπως το «Hadoop» το οποίο χρησιμοποιεί τον αλγόριθμο «MapReduce», το «Apache Spark» για «Streaming Data» κλπ.

❖ Αποθήκες Δεδομένων (Data Storage).

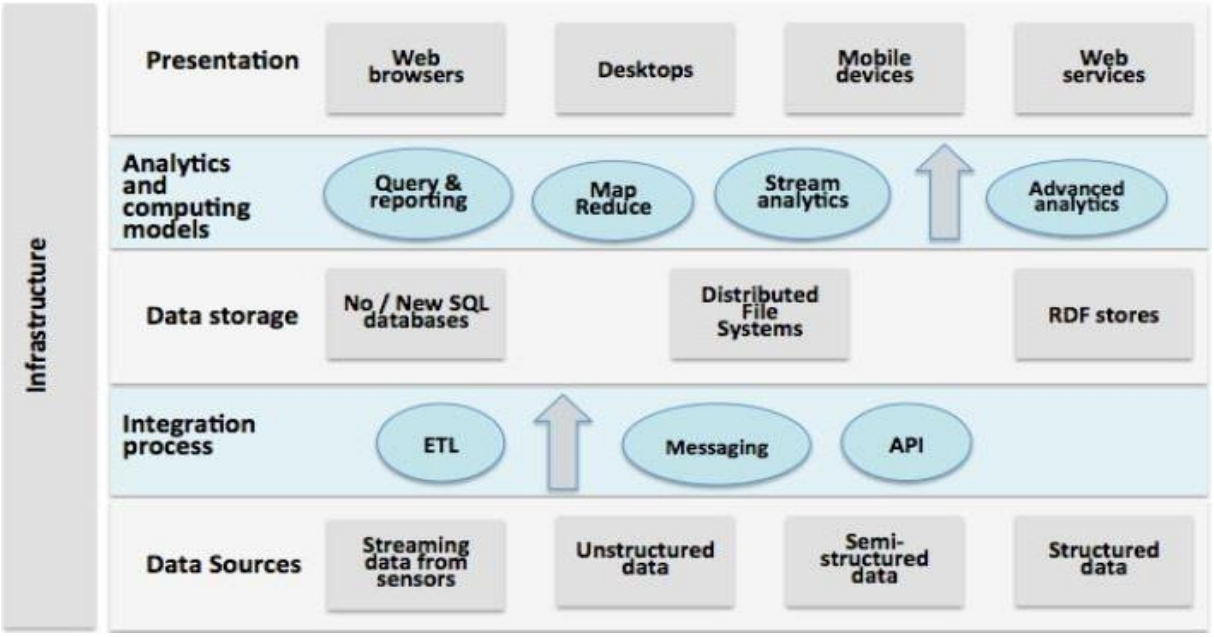
Στο επίπεδο αυτό περιλαμβάνονται οι διάφορες αποθήκες δεδομένων (Data Stores), αλλά και τα συστήματα αποθήκευσης αυτών των δεδομένων (File Systems). Οπότε εδώ βρίσκονται τα διάφορα κατανεμημένα συστήματα αρχείων (Distributed File System) με αντιπροσωπευτικότερο όλων το «Hadoop Distributed File System» (HDFS), αλλά και οι διάφορες αποθήκες δεδομένων, όπως οι «No/New SQL» βάσεις δεδομένων και οι «Resource Description Framework» αποθήκες δεδομένων (RDF-Stores) ή αλλιώς « Triplestores».

❖ Εργαλεία Εκτέλεσης Υπολογισμών και Τεχνικές Αναλυτικής (Analytics and Computing Models).

Στο επίπεδο αυτό περιλαμβάνονται οι διάφορες τεχνικές και τα εργαλεία που χρησιμοποιούνται κατά την διαδικασία των «Big Data Analytics». Σκοπός αυτών είναι να ανακαλυφθούν τα μοτίβα και οι διασυνδέσεις που μπορεί να κρύβονται μέσα σε αυτά, προκειμένου να εξαχθούν συμπεράσματα και να ληφθούν αποφάσεις. Ενδεικτικά, κάποια από αυτά είναι το «Data Mining», το «Business Intelligence», το «Predictive Analytics», το «Prescriptive Analytics», το «Text Analytics», το «Query and Reporting», το «Stream Data Analytics», το «Advanced Analytics», το «Event Processing Analytics», το «Web Analytics», το «Behavioral Analytics» κλπ.

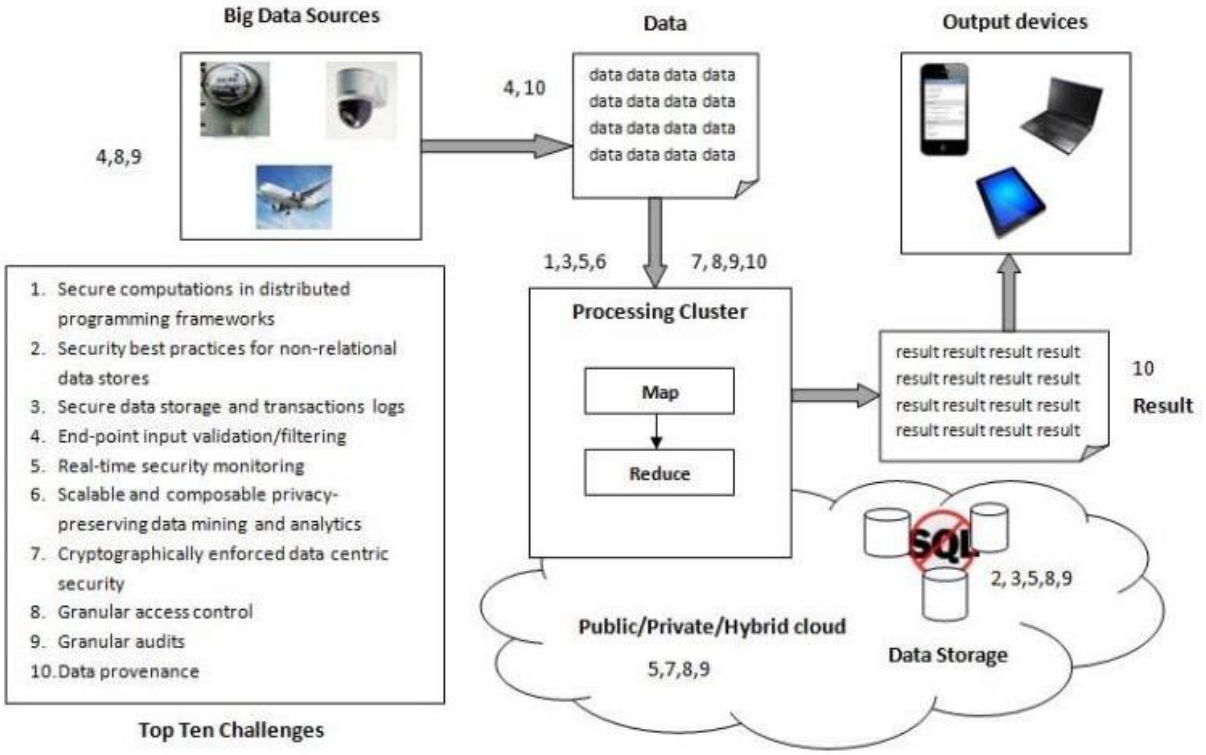
❖ Εργαλεία και Τεχνικές Παρουσίασης των Αποτελεσμάτων (Presentation).

Στο επίπεδο αυτό περιλαμβάνονται οι τεχνολογίες για την απεικόνιση και την παρουσίαση των αποτελεσμάτων της διαδικασίας των «Big Data Analytics», καθώς και οι διάφορες εφαρμογές που υποστηρίζουν την υπηρεσία αυτή. Τέτοιες τεχνολογίες απεικόνισης είναι τα διάφορα εργαλεία παραγωγής ειδοποιήσεων (Alerts), γραφημάτων (Dashboards) ή/και αναφορών (Reports). Ενώ στις εφαρμογές που υποστηρίζουν την δυνατότητα αυτή συμπεριλαμβάνονται, τόσο οι διάφορες εφαρμογές των κινητών ή σταθερών συσκευών και του διαδικτύου, όσο και αυτές που έχουν σχεδιαστεί και υλοποιηθεί αποκλειστικά για κάποια συγκεκριμένη περίπτωση (Custom Applications).



ΕΙΚΟΝΑ 7: ENISA – LAYERED ARCHITECTURE OF BIG DATA SYSTEMS

Το [1] παρουσιάζει μια εναλλακτική απεικόνιση της αρχιτεκτονικής των συστημάτων «Big Data» (Εικόνα 8), η οποία όμως δεν διαφέρει σημαντικά ως προς τα τμήματα που απαρτίζουν ένα τέτοιο σύστημα συγκρινόμενη με την αρχιτεκτονική που περιγράφηκε από τον [2]. Η ουσιαστική διαφορά των δυο αυτών απεικονίσεων έγκειται στο γεγονός ότι το [1] προβαίνει στην απαρίθμηση των ζητημάτων «Ασφαλείας» και «Προστασίας της Ιδιωτικής Ζωής» στα συστήματα αυτά και μάλιστα τα απεικονίζει στο ίδιο σχήμα με αριθμούς, προσδιορίζοντας έτσι και σε ποιά φάση αυτά συναντώνται, αλλά και ποιά τμήματα του συστήματος εμπλέκονται κάθε φορά.



ΕΙΚΟΝΑ 8: CSA - TOP 10 SECURITY AND PRIVACY CHALLENGES IN BIG DATA ECOSYSTEM

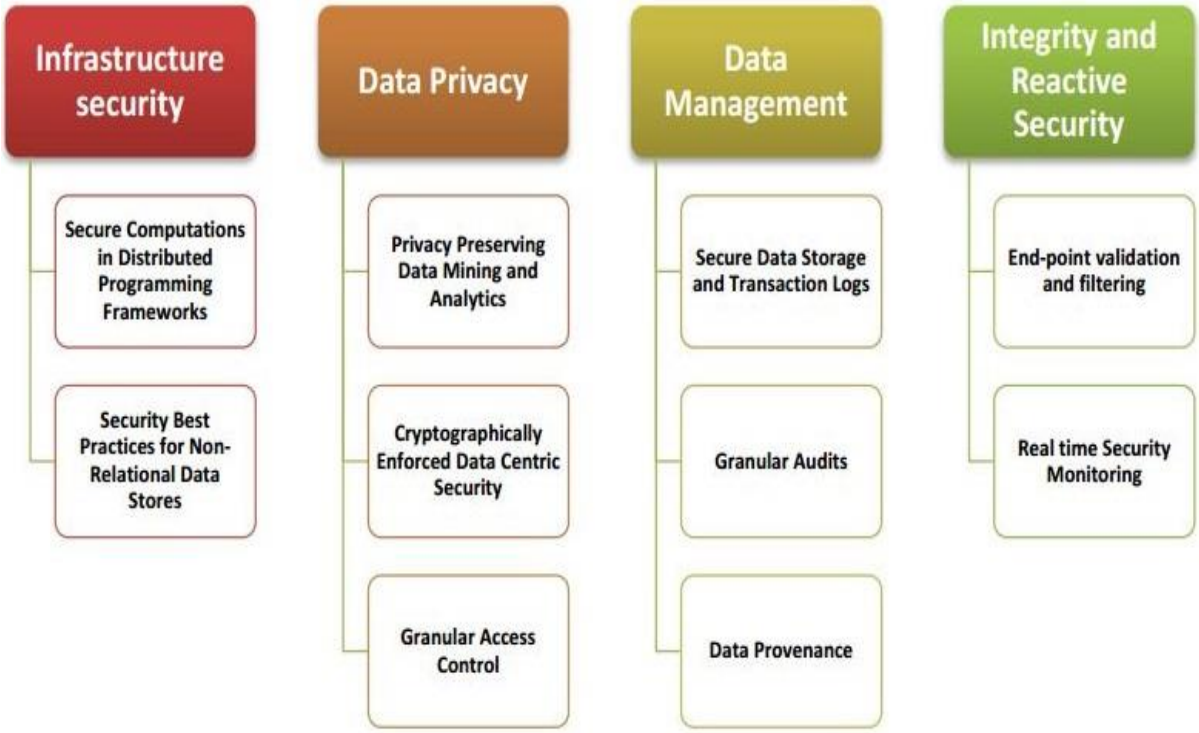
Μάλιστα προβαίνει και στην περαιτέρω ταξινόμηση των ζητημάτων αυτών σε τέσσερις κατηγορίες (Εικόνα 9), οι οποίες είναι οι παρακάτω:

✓ Ασφάλεια της Υποδομής των «Big Data» (Infrastructure Security).

✓ Ασφάλεια της Διαδικασίας Επεξεργασίας και Διαχείρισης των Δεδομένων (Data Management).

✓ Έλεγχος της Ακεραιότητας των Δεδομένων Εισόδου και Επιτήρηση του Συστήματος σε Πραγματικό Χρόνο (Integrity and Reactive Security).

✓ Προστασία της «Ιδιωτικότητας» (Data Privacy).



ΕΙΚΟΝΑ 9: CSA – CLASSIFICATION OF THE TOP 10 CHALLENGES

Εναλλακτικά, οι εργασίες [12] και [15] υιοθετούν ένα άλλο μοντέλο για την αρχιτεκτονική των συστημάτων «Big Data». Σε αυτό το μοντέλο αναπτύσσεται ένα ολοκληρωμένο περιβάλλον, στο οποίο συμπεριλαμβάνονται το υλικό (Hardware), το λογισμικό για την λειτουργία της υποδομής (Infrastructure Software), το επιχειρησιακό λογισμικό (Operational Software), το λογισμικό διαχείρισης (Management Software) και οι διεπαφές του χρήστη με την υποδομή και τις εφαρμογές του συστήματος (Application Programming Interfaces - APIs). Η εννοιολογική αναπαράσταση αυτής της αρχιτεκτονικής σε πολλαπλά επίπεδα ονομάζεται «Big Data Technology Stack» (Εικόνα 10). Η αναπαράσταση αυτή βοηθά στον καθορισμό των τμημάτων που απαρτίζουν τα συστήματα αυτά, καθώς και στην εύρεση των αλληλεξαρτήσεων μεταξύ τους. Επιπλέον, καθιστά εύκολο να εντοπιστεί κάθε φορά η κατάλληλη τεχνολογία που θα πρέπει να χρησιμοποιηθεί ανά επίπεδο, προκειμένου να αντιμετωπιστεί η οποιαδήποτε ανάγκη ή πρόβλημα που θα προκύψει ανάλογα με την εκάστοτε περίπτωση.

Αναλυτικότερα, τα επίπεδα αυτής της αρχιτεκτονικής (Εικόνα 10) είναι τα εξής:

- Διεπαφές και Ανταλλαγή Δεδομένων (Interfaces and Feeds).

Αρχικά θα πρέπει να καταστεί σαφές ότι σε ένα σύστημα «Big Data» υπάρχουν πολλαπλές διεπαφές μέσω των οποίων γίνεται η ανταλλαγή δεδομένων από και

προς αυτό το σύστημα. Τα δεδομένα που ανταλλάσσονται περιλαμβάνουν, τόσο τα εσωτερικά δεδομένα που αυτό το σύστημα διαχειρίζεται, όσο και τα δεδομένα με τα οποία αυτό τροφοδοτείται από τις διάφορες εξωτερικές πηγές. Στην πραγματικότητα ένα σύστημα «Big Data» διαχειρίζεται πολλαπλά δεδομένα από πολλές και διαφορετικές πηγές. Επομένως σε ένα τέτοιο σύστημα οι διεπαφές αποτελούν το θεμελιώδες δομικό στοιχείο του. Επιπρόσθετα, θα πρέπει να επισημανθεί ότι διεπαφές υπάρχουν και μεταξύ του κάθε επιπέδου με τον εξωτερικό κόσμο, αλλά και μεταξύ των επιπέδων της στοίβας.

❖ Ανθεκτική Υλική Υποδομή (Redundant Physical Infrastructure).

Η υλική υποδομή καθίσταται κρίσιμη για την λειτουργία και την ευελιξία του συστήματος «Big Data». Για να αντιμετωπιστούν αποτελεσματικά τα χαρακτηριστικά του μεγάλου όγκου, της υψηλής ταχύτητας και της ποικιλομορφίας των δεδομένων απαιτούνται ανθεκτικές υλικές υποδομές. Επιπλέον, το γεγονός ότι τα συστήματα αυτά στηρίζονται στο κατανεμημένο υπολογιστικό μοντέλο, συνεπάγεται ότι τα δεδομένα τους βρίσκονται αποθηκευμένα σε πολλές διαφορετικές τοποθεσίες. Έτσι, αυτά διασυνδέονται μεταξύ τους μέσω δικτύων, χρησιμοποιώντας τα κατανεμημένα συστήματα αρχείων. Συνεπώς, οι υλικές υποδομές θα είναι αυτές που θα υποστηρίξουν την ασφαλή και αποτελεσματική αποθήκευση, επεξεργασία και δικτύωση των διαφόρων τμημάτων του συστήματος αυτού. Ωστόσο, για την επιλογή της κατάλληλης κάθε φορά υλικής υποδομής θα πρέπει να λαμβάνονται υπόψη οι απαιτήσεις που τίθενται από την φύση της εκάστοτε υλοποίησης, λαμβάνοντας υπόψη τα κριτήρια της «Απόδοσης» (Performance), της «Διαθεσιμότητας» (Availability), της «Επεκτασιμότητας» (Scalability) και της «Ευελιξίας» (Flexibility). Συγκεκριμένα, η «Απόδοση» μετρά τον βαθμό απόκρισης του συστήματος και την ταχύτητά του, ενώ η «Διαθεσιμότητα» μετρά τον χρόνο λειτουργίας του συστήματος. Η «Επεκτασιμότητα» μετρά το κατά πόσο εύκολα και αποτελεσματικά μπορεί να επεκταθεί το σύστημα, όταν προκύψουν τροποποιήσεις όπως η ανάγκη για περισσότερο χώρο αποθήκευσης, ή μεγαλύτερης υπολογιστικής ισχύος. Τέλος, η «Ευελιξία» σχετίζεται με την ταχύτητα με την οποία είναι δυνατό, είτε να προστεθούν επιπλέον πηγές στο σύστημα, είτε αυτό να ανακάμψει σε περίπτωση αποτυχίας του.

❖ Υποδομή Ασφαλείας (Security Infrastructure).

Στο επίπεδο αυτό ορίζονται, αφενός όλες οι απαιτήσεις ασφαλείας που θα πρέπει να πληροί το σύστημα κατά τον σχεδιασμό και την υλοποίησή του και αφετέρου υλοποιούνται όλα εκείνα τα μέτρα που θα τις διασφαλίζουν κατά την φάση της λειτουργίας του, έτσι ώστε να μην προκύψει η ανάγκη να αντιμετωπιστούν αυτές εκ

των υστέρων. Οπότε σε αυτό το επίπεδο βρίσκονται όλες οι τεχνολογίες και οι μηχανισμοί ασφάλειας, όπως η κρυπτογραφία, οι μηχανισμοί ελέγχου πρόσβασης και εξουσιοδότησης, οι τεχνικές ανωνυμοποίησης κλπ, που απαιτούνται προκειμένου να προστατευτούν τα δεδομένα, η υποδομή, οι επικοινωνίες της και η «Ιδιωτικότητα» στα περιβάλλοντα αυτά.

❖ Επιχειρησιακές Πηγές Δεδομένων και Βάσεις Δεδομένων (Operational Data Sources and Databases).

Σε αυτό το επίπεδο βρίσκονται, αφενός οι διάφορες πηγές δεδομένων που τροφοδοτούν το σύστημα με δεδομένα και αφετέρου οι διάφορες βάσεις δεδομένων που θα χρησιμοποιηθούν από το σύστημα «Big Data» για την αποθήκευση και την διαχείριση των δεδομένων αυτών. Επομένως, στις πηγές δεδομένων συμπεριλαμβάνονται όλες εκείνες οι πηγές που θα πρέπει να ενσωματωθούν στο σύστημα «Big Data», με γνώμονα ότι τα δεδομένα τους κρίνονται χρήσιμα για την ολοκληρωμένη και σφαιρική επίλυση του εκάστοτε προβλήματος το οποίο και επέβαλε την υλοποίηση του συγκεκριμένου συστήματος. Ενώ, ανάλογα με τον τύπο των δεδομένων που παρέχονται στο σύστημα από τις διάφορες πηγές, ήτοι δομημένα (Structured), ημι-δομημένα (Semi-Structured) ή/και μη-δομημένα (Unstructured), θα πρέπει να χρησιμοποιηθούν και οι αντίστοιχες βάσεις δεδομένων, έτσι ώστε να βελτιστοποιείται το αποτέλεσμα ανά περίπτωση. Ως εκ τούτου, εδώ συμπεριλαμβάνονται οι διάφορες βάσεις δεδομένων με τις αντίστοιχες γλώσσες επερωτήσεων, ήτοι σχεσιακές (Relational or SQL) και μη (Non-Relational or NoSQL), που θα διαχειρίζονται αυτή την ποικιλία των δεδομένων, ανάλογα με την εκάστοτε περίπτωση και το πρόβλημα που τίθεται.

❖ Οργάνωση Υπηρεσιών και Εργαλείων (Organizing Data Services and Tools).

Στο επίπεδο αυτό συμπεριλαμβάνονται τα ποικίλα εργαλεία και οι υπηρεσίες που θα χρησιμοποιηθούν για την συλλογή, την αποθήκευση και την επεξεργασία των ποικιλόμορφων και μεγάλων σε όγκο και ταχύτητα δεδομένων του συστήματος «Big Data». Δηλαδή εδώ βρίσκεται όλο εκείνο το κατάλληλο λογισμικό, που θα επεξεργαστεί τα ακατέργαστα δεδομένα, που συγκεντρώνονται από τις διάφορες πηγές του συστήματος, και που θα τα μετατρέψει σε κατάλληλη μορφή για την περαιτέρω επεξεργασία τους από τις διάφορες τεχνικές της «Αναλυτικής» (Analytics). Οπότε εδώ βρίσκονται τεχνολογίες όπως τα εργαλεία «Apache Hadoop» και «Apache Spark», ο αλγόριθμος «MapReduce», το σύστημα αρχείων «Big Table» ή «HDFS» κλπ.

❖ Αποθήκες Δεδομένων για την Διεξαγωγή των Μεθόδων της Αναλυτικής (Analytical Data Warehouses and Data Marts).

Τα δεδομένα που παράγονται από τα εργαλεία και τις υπηρεσίες του προηγούμενου επιπέδου θα πρέπει να αποθηκευτούν σε κατάλληλες αποθήκες δεδομένων, οι οποίες θα υποστηρίζουν την εκτέλεση επερωτήσεων επ' αυτών μέσα από τις διάφορες τεχνικές της «Αναλυτικής». Σκοπός είναι οι τεχνικές αυτές να διεξάγονται αποτελεσματικά, έτσι ώστε να εξαχθούν πολύτιμα συμπεράσματα χωρίς να υπάρχει επιβάρυνση στο σύστημα. Οπότε εδώ συμπεριλαμβάνονται οι αποθήκες δεδομένων «Data Warehouses» και «Data Marts» οι οποίες υποστηρίζουν συμπίεση, πολυεπίπεδη κατανομή και μαζική παράλληλη επεξεργασία.

❖ Παραδοσιακές και Εξελιγμένες Μέθοδοι Αναλυτικής (Traditional and Advanced Analytics).

Στο επίπεδο αυτό βρίσκονται όλες εκείνες οι παραδοσιακές αλλά και εξελιγμένες τεχνικές και μεθοδολογίες, οι οποίες θα χρησιμοποιηθούν για τις ανάγκες της «Αναλυτικής» επί των δεδομένων του συστήματος «Big Data», προκειμένου να εξαχθούν πολύτιμα συμπεράσματα που θα συμβάλουν στην αποτελεσματική επίλυση κάποιου προβλήματος. Οπότε εδώ βρίσκονται διάφορες τεχνικές της «Αναλυτικής» όπως η «Predictive Analytics», η «Social Media Analytics», η «Text Analytics», το «Business Intelligence», η «Simulation Analytics», η «Optimization Analytics», κλπ.

❖ Σύστημα Αναφορών και Οπτικοποίησης (Reporting and Visualization).

Στο επίπεδο αυτό βρίσκονται οι διάφορες εφαρμογές και τεχνολογίες που θα χρησιμοποιηθούν για να μετατρέψουν τα συμπεράσματα και τα αποτελέσματα, που εξήχθησαν από τις διαδικασίες της «Αναλυτικής» στο προηγούμενο επίπεδο, σε μορφή που θα είναι εύκολο να γίνει κατανοητή και αντιληπτή από τον άνθρωπο. Οπότε εδώ βρίσκονται εργαλεία για την παραγωγή αναφορών, καθώς και εργαλεία που αναπαριστούν τα αποτελέσματα αυτά γραφικά.

❖ Εφαρμογές (Big Data Applications).

Στο επίπεδο αυτό συμπεριλαμβάνονται οι εφαρμογές εκείνες που θα εκμεταλλεύονται «έξυπνα» τα αποτελέσματα από την επεξεργασία των «Big Data» και είτε θα λαμβάνουν προληπτικά μέτρα για την αντιμετώπιση κάποιου προβλήματος, πριν αυτό λάβει χώρα, είτε θα παρέχουν έγκαιρα τις απαραίτητες πληροφορίες για την επίτευξη βέλτιστου κέρδους, ή την λήψη της βέλτιστης απόφασης σε κάποιο ζήτημα. Οι εφαρμογές του επιπέδου αυτού, οι οποίες τυγχάνουν εφαρμογής σε διάφορους τομείς όπως στον τομέα της υγείας, στον κατασκευαστικό τομέα, στον επιστημονικό

τομέα κλπ, απευθύνονται στους τελικούς χρήστες και μπορεί, είτε να είναι σχεδιασμένες και ενσωματωμένες μέσα στο ίδιο σύστημα, είτε να παρέχονται από ανεξάρτητα τρίτα μέρη.

ΕΙΚΟΝΑ 10: BIG DATA FOR DUMMIES – THE BIG DATA TECHNOLOGY STACK

Οπότε, οι προαναφερθείσες αρχιτεκτονικές θα αποτελέσουν την βάση για την περαιτέρω διερεύνηση και ανάλυση των ζητημάτων που σχετίζονται με την «Ασφάλεια» και την «Προστασία της Ιδιωτικής Ζωής» στα συστήματα «Big Data». Μάλιστα θα υιοθετηθεί η κατηγοριοποίηση των ζητημάτων αυτών σύμφωνα με το [1] για δυο βασικούς λόγους. Αφενός γιατί καταγράφει και κατηγοριοποιεί το σύνολο των ζητημάτων που εντοπίζονται στο περιβάλλον «Big Data» σε δύο κατηγορίες, ήτοι σε ζητήματα ασφαλείας και σε ζητήματα προστασίας της «Ιδιωτικότητας», και αφετέρου γιατί η απεικόνιση των ζητημάτων αυτών πάνω στην αρχιτεκτονική που προτείνει, θα συμβάλλει στην καλύτερη κατανόησή τους, καθώς για κάθε ζήτημα που θα διερευνάται θα είναι δυνατό παράλληλα να εντοπιστούν, τόσο η αντίστοιχη φάση επεξεργασίας που εκτελείται, όσο και τα αντίστοιχα τμήματα που εμπλέκονται και συνεπώς θα πρέπει να προστατευτούν.

1.5 ΣΚΟΠΟΣ-ΔΙΑΡΘΡΩΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Συνεπώς, η παρούσα μεταπτυχιακή διπλωματική εργασία έχει διττό σκοπό:

α) Αποσκοπεί στη διερεύνηση και τον εντοπισμό των ζητημάτων εκείνων που άπτονται της ασφάλειας στις διάφορες υλοποιήσεις των «Big Data», και

β) Μελετά τα θέματα εκείνα που σχετίζονται με την προστασία της «Ιδιωτικότητας» (Privacy) στα περιβάλλοντα αυτά.

Μάλιστα, η σφαιρική μελέτη των ζητημάτων αυτών και η παρουσίαση των διαφόρων προτεινόμενων λύσεων για την αντιμετώπιση των επιμέρους προβλημάτων, που δημιουργούνται ανά κατηγορία, θα αποτελέσουν την βάση για μελλοντική έρευνα, καθώς, είτε οι περισσότερες από αυτές τις προτεινόμενες λύσεις βρίσκονται σε ερευνητικό στάδιο ή αποτελούν απλά προτάσεις, είτε γιατί κάποια επιμέρους προβλήματα παραμένουν ανοιχτά και άλυτα. Απώτερος στόχος λοιπόν της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι να καταστούν η ασφάλεια και η προστασία της «Ιδιωτικότητας» αναπόσπαστο μέρος κατά τον σχεδιασμό και την υλοποίηση των συστημάτων «Big Data». Έτσι, με γνώμονα τα ανωτέρω, η παρούσα μεταπτυχιακή διπλωματική εργασία διαρθρώνεται όπως παρακάτω.

Στο πρώτο κεφάλαιο εισάγεται ο όρος των «Big Data» (Μαζικά Δεδομένα) και ο όρος των «Big Data Analytics» (Αναλυτική των Μαζικών Δεδομένων). Στη συνέχεια αναλύονται οι διάφορες φάσεις επεξεργασίας των δεδομένων και παρουσιάζεται η αρχιτεκτονική που συνήθως έχουν αυτά τα συστήματα. Έτσι θα καταστεί πιο εύκολη η διερεύνηση των ζητημάτων ασφαλείας, αλλά και προστασίας της «Ιδιωτικότητας» στο περιβάλλον των «Big Data», καθώς αυτά θα διερευνηθούν σε βάθος, τόσο κατά τις διάφορες φάσεις επεξεργασίας, όσο και κατά την διαχείριση των δεδομένων από τα διάφορα τμήματα που απαρτίζουν τα συστήματα αυτά. Τα επόμενα δυο κεφάλαια υιοθετούν την αρχιτεκτονική και την κατηγοριοποίηση του [1] για την περαιτέρω μελέτη των ζητημάτων αυτών.

Αναλυτικότερα, στο δεύτερο κεφάλαιο παρουσιάζονται και μελετώνται τα ζητήματα ασφαλείας των συστημάτων «Big Data». Αρχικά εξετάζεται η ασφάλεια της υποδομής (Infrastructure Security) που υποστηρίζει το περιβάλλον «Big Data» και συγκεκριμένα διερευνώνται ζητήματα που σχετίζονται με την εκτέλεση ασφαλών υπολογισμών σε κατανεμημένα συστήματα (Secure Computations in Distributed

Programming Frameworks) ,αλλά και με την ασφάλεια των μη-σχεσιακών αποθηκών δεδομένων που χρησιμοποιούνται στα συστήματα αυτά (Security Best Practices for Non-Relational Data Stores). Έπειτα, εξετάζεται η ασφάλεια της διαδικασίας επεξεργασίας των δεδομένων (Data Management) που διαχειρίζονται τα συστήματα «Big Data». Έτσι μελετώνται τα θέματα που αφορούν στον έλεγχο της προέλευσης των δεδομένων από τις διάφορες πηγές του συστήματος (Data Provenance), στην ασφάλεια κατά την αποθήκευση των δεδομένων αυτών και στην τήρηση ιστορικού ενεργειών (Secure Data Storage and Transaction Logs), καθώς και στην διεξαγωγή αποτελεσματικών επιθεωρήσεων επί της διαδικασίας επεξεργασίας των δεδομένων (Granular Audits). Τέλος, εξετάζονται τα ζητήματα που σχετίζονται με την ακεραιότητα των δεδομένων εισόδου (Integrity – End Point Validation and Filtering) και την επιτήρηση του συστήματος «Big Data» σε πραγματικό χρόνο (Reactive Security – Real Time Security Monitoring).

Στο τρίτο κεφάλαιο παρουσιάζονται και μελετώνται τα ζητήματα που σχετίζονται με την προστασία της «Ιδιωτικότητας» (Data Privacy) στα συστήματα «Big Data». Αρχικά παρατίθενται οι ορισμοί του «Privacy» και του «Privacy By Design». Στη συνέχεια παρατίθεται συνοπτικά το ρυθμιστικό και κανονιστικό πλαίσιο που διέπει την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων στην Ευρωπαϊκή Ένωση και την Ελλάδα. Τέλος, εξετάζονται τα ζητήματα που σχετίζονται με την προστασία της «Ιδιωτικότητας». Αυτά συγκεκριμένα είναι η εφαρμογή της μεθόδου για την «Αξιολόγηση των Επιπτώσεων επί της Ιδιωτικότητας» (Privacy Impact Assessment - PIA) κατά τον σχεδιασμό των συστημάτων, η εφαρμογή τεχνικών Ανωνυμοποίησης (Anonymization Techniques), η υλοποίηση των Κρυπτογραφικών αλγορίθμων και μηχανισμών ανάλογα με την εκάστοτε περίπτωση (Cryptography Mechanisms), η υλοποίηση μηχανισμών Ελέγχου Πρόσβασης (Access Control Mechanisms), η υιοθέτηση μεθόδων Ενημέρωσης και Διαφάνειας (Information and Transparency Methods) και η υλοποίηση μηχανισμών Συναίνεσης, Δήλωσης Κατοχής και Ελέγχου (Consent, Ownership and Control Mechanisms).

Στο τέταρτο κεφάλαιο εξετάζεται σε πρακτικό επίπεδο, κατά πόσο είναι εφικτό, στα πλαίσια της «Ανοιχτής Διακυβέρνησης» (Open Government) στην Ελληνική Δημόσια Διοίκηση, να διαρρεύσουν τα προσωπικά δεδομένα των Ελλήνων πολιτών, ύστερα από την συγκέντρωση και επεξεργασία των διαφόρων πληροφοριών που αποθηκεύονται ή/και παρέχονται, τόσο από τα διάφορα πληροφοριακά συστήματα

της Δημόσιας Διοίκησης, όσο και από άλλες ευρέως διαθέσιμες μη κυβερνητικές πηγές δεδομένων. Για τις ανάγκες της παρούσας έρευνας στις κυβερνητικές πηγές δεδομένων συμπεριλαμβάνονται το «Υπουργείο Οικονομικών» (ΑΦΜ), το «Υπουργείο Προστασίας του Πολίτη» (ΑΔΤ), το «Υπουργείο Εργασίας και Κοινωνικής Ασφάλισης» (ΑΜΚΑ), το «Υπουργείο Εσωτερικών» (Εκλογές, Δημοτολόγιο) και το «Υπουργείο Διοικητικής Ανασυγκρότησης» (Διαύγεια). Ενώ αντίστοιχα από τις μη κυβερνητικές πηγές χρησιμοποιείται η «Υπηρεσία Τηλεφωνικού Καταλόγου του ΟΤΕ» (Τηλέφωνο). Στόχος της μελέτης της περίπτωσης αυτής είναι η διερεύνηση του κατά πόσο είναι εφικτό να παραβιαστεί η «Ιδιωτικότητα» σε ένα τέτοιο περιβάλλον και να συλλεχθούν ευαίσθητα και μη προσωπικά δεδομένα για κάποιο Έλληνα πολίτη, μέσα από την επεξεργασία των διαφορετικών συνόλων δεδομένων των πηγών αυτών.

Τέλος, στο πέμπτο κεφάλαιο συνοψίζονται τα συμπεράσματα της παρούσας μεταπτυχιακής διπλωματικής εργασίας, που αφορούν τα ζητήματα ασφαλείας αλλά και προστασίας της «Ιδιωτικότητας» στα συστήματα «Big Data». Παράλληλα, εκτίθενται και τα αποτελέσματα από την μελέτη της περίπτωσης του 4^{ου} κεφαλαίου. Επιπρόσθετα παρέχονται οι κατευθύνσεις εκείνες, οι οποίες αφορούν στην διεξαγωγή μελλοντικής διερεύνησης σε βάθος, για την αποτελεσματική επίλυση του κάθε ενός ζητήματος που παρουσιάστηκε στην παρούσα μεταπτυχιακή διπλωματική εργασία με στόχο, είτε να ανευρεθούν νέες λύσεις, είτε να τροποποιηθούν και να υιοθετηθούν οι υπάρχουσες, προκειμένου να καταστούν τα περιβάλλοντα «Big Data» ασφαλή.

2. ΚΕΦΑΛΑΙΟ 2ο: ΕΞΕΤΑΣΗ ΤΗΣ ΑΣΦΑΛΕΙΑΣ ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ (BIG DATA SECURITY)

2.1 ΠΡΟΛΟΓΟΣ

Στο παρόν κεφάλαιο παρουσιάζονται και μελετώνται σε βάθος τα ζητήματα ασφαλείας των συστημάτων «Big Data». Είναι γεγονός ότι τα θέματα που σχετίζονται με την ασφάλεια και την προστασία της ιδιωτικής ζωής μεγεθύνονται από τον όγκο, την ποικιλία και την ταχύτητα των «Big Data». Η χρήση υποδομών «Νεφοϋπολογιστικής» (Cloud Infrastructures), η προέλευσή τους από διάφορες πηγές δεδομένων, η ποικιλομορφία των δεδομένων, ο υψηλός ρυθμός ανταλλαγής δεδομένων μεταξύ των διαφόρων υπολογιστικών συστημάτων, η ταχύτητα με την οποία εισέρχονται αυτές οι ροές δεδομένων στο σύστημα για επεξεργασία και η ταχύτητα με την οποία αυτά θα πρέπει να αναλυθούν για να εξαχθούν συμπεράσματα, δημιουργούν ένα μοναδικό περιβάλλον για τα «Big Data» στο οποίο τα ζητήματα ασφαλείας, που σχετίζονται με την εύρεση των ευπαθειών, την αξιολόγηση των απειλών και την αντιμετώπισή τους, θα πρέπει να εξεταστούν ενδελεχώς.

Παράλληλα θα πρέπει να ληφθεί υπόψη ότι δεν είναι μόνο ο όγκος των δεδομένων που δημιουργεί νέες προκλήσεις στην ασφάλεια των «Big Data». Αυτά συλλέγονταν και χρησιμοποιούνταν από πολλές μεγάλες οργανώσεις, όπως κυβερνήσεις και μεγάλες επιχειρήσεις, για αρκετές δεκαετίες, καθώς μπορούσαν να αντέξουν οικονομικά τη δημιουργία και τη συντήρηση των αναγκαίων υποδομών για την αποθήκευση και την εκμετάλλευσή τους. Οι υποδομές αυτές ήταν συνήθως ιδιόκτητες και απομονωμένες από άλλα δίκτυα, γεγονός που καθιστούσε πιο εύκολη την προστασία τους από διάφορες απειλές. Ωστόσο σήμερα, οι ραγδαίες τεχνολογικές εξελίξεις που έχουν πραγματοποιηθεί συνέβαλλαν στην έλευση της εποχής των «Big Data». Οι οργανισμοί όλων των μεγεθών έχουν πλέον πρόσβαση στα «Big Data» μέσα από τις δημόσιες υποδομές αλλά και την διασύνδεση των δικτύων. Εργαλεία για την παράλληλη επεξεργασία δεδομένων με την αξιοποίηση χιλιάδων υπολογιστικών κόμβων, όπως το «Hadoop», αλλά και για την περαιτέρω ανάλυσή τους και αξιοποίησή τους πλέον είναι διαθέσιμα με σχετικά μικρό κόστος και με μεγάλη αποτελεσματικότητα. Τα προβλήματα, που αφορούσαν τους περιορισμούς

σε μέσα αποθήκευσης αλλά και υπολογιστικής ισχύος, πλέον αντιμετωπίζονται αποτελεσματικά με τη βοήθεια, αφενός της εξέλιξης της τεχνολογίας στους τομείς αυτούς και αφετέρου χάρη στην «Νεφοϋπολογιστική» (Cloud), η οποία μπορεί να παρέχει και τα δυο αποτελεσματικά κατά βούληση και με σχετικά μικρό κόστος. Συνεπώς, οι παραδοσιακοί μηχανισμοί ασφαλείας, που είχαν υλοποιηθεί για την προστασία των δεδομένων που ήταν αποθηκευμένα σε δίκτυα απομονωμένα ή ημι-απομονωμένα, πλέον κρίνονται ανεπαρκείς και περαιτέρω έρευνα στον τομέα της ασφάλειας των «Big Data» απαιτείται.

2.2 ΕΞΕΤΑΣΗ ΤΗΣ ΑΣΦΑΛΕΙΑΣ ΤΗΣ ΥΠΟΔΟΜΗΣ ΤΩΝ ΣΥΣΤΗΜΑΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ (BIG DATA INFRASTRUCTURE SECURITY)

Έτσι για να είναι ολοκληρωμένη η εξέταση των ζητημάτων «Ασφαλείας» στα συστήματα «Big Data», σε πρώτη φάση εξετάζεται στην ενότητα αυτή η ασφάλεια της υποδομής που υποστηρίζει τα συστήματα αυτά. Αναλυτικότερα μελετάται, αφενός η προστασία των υπολογισμών που εκτελούνται στο κατανεμημένο υπολογιστικό περιβάλλον και αφετέρου η προστασία των διαφόρων αποθηκών δεδομένων ολόκληρου του συστήματος «Big Data», συμπεριλαμβανομένων και αυτών που χρησιμοποιούνται στις διάφορες πηγές δεδομένων.

2.2.1 ΑΣΦΑΛΕΙΣ ΥΠΟΛΟΓΙΣΜΟΙ ΣΕ ΚΑΤΑΝΕΜΗΜΕΝΟ ΥΠΟΛΟΓΙΣΤΙΚΟ ΠΕΡΙΒΑΛΛΟΝ (DISTRIBUTED PROGRAMMING FRAMEWORK)

Τα κατανεμημένα υπολογιστικά περιβάλλοντα χρησιμοποιούνται ευρέως στα συστήματα «Big Data», γιατί παρέχουν την δυνατότητα της παράλληλης επεξεργασίας των δεδομένων, καθώς και της ασφαλούς αποθήκευσης αυτών, συμβάλλοντας στην αποδοτική επεξεργασία του μεγάλου όγκου τους. Έτσι για παράδειγμα, στην περίπτωση της παράλληλης επεξεργασίας των δεδομένων, ο αλγόριθμος «MapReduce» χωρίζει αρχικά την είσοδο των δεδομένων σε πολλαπλά τμήματα. Σε πρώτη φάση αυτά τα τμήματα τυγχάνουν επεξεργασίας το καθένα από τον εκάστοτε «Mapper» και το αποτέλεσμα που παράγεται στην έξοδο είναι μια λίστα με ζεύγη τιμών της μορφής («ΠΕΔΙΟ»/«ΤΙΜΗ»). Στην επόμενη φάση οι «Reducers» συνδυάζουν τα αποτελέσματα των «Mappers», παράγοντας ως έξοδο ο καθένας το σύνολο των τιμών που αντιστοιχούν σε συγκεκριμένο πεδίο [15]. Οπότε στο παράδειγμα αυτό θα πρέπει να ληφθούν υπόψη τα εξής δυο σημαντικά ζητήματα

ασφαλείας, προκειμένου να προληφθούν. Το πρώτο είναι να προστατευθούν οι ίδιοι οι «Mappers» και οι «Reducers» και το δεύτερο είναι να εξασφαλιστεί ότι τα δεδομένα που παράγονται και διοχετεύονται στους «Reducers» για περαιτέρω επεξεργασία θα παραμένουν, αφενός αξιόπιστα και αφετέρου ασφαλή από ενδεχόμενη διαρροή τους, ακόμα κι αν στο σύστημά μας υπάρχει/ουν κάποιος/οι «Mappers» που δεν πρέπει να εμπιστευόμαστε [52].

Γενικά, οι διάφοροι κόμβοι του συστήματος που είτε αναλαμβάνουν το έργο του «Mapper», είτε το έργο του «Reducer» μπορεί να τροποποιηθούν από κάποιον επιτιθέμενο, είτε για να υποκλέπτουν αιτήματα, είτε για να αλλοιωθεί ο κώδικάς τους και η λειτουργικότητά τους, είτε για να αλλοιώσουν και να αλλάξουν τα παραγόμενα αποτελέσματα [52]. Από τις τρείς περιπτώσεις αυτή που είναι πιο δύσκολο να ανιχνευθεί είναι ο εντοπισμός για παράδειγμα του «Mapper» που παράγει αλλοιωμένα αποτελέσματα μέσα στον τεράστιο όγκο των δεδομένων που τυγχάνουν επεξεργασίας από τους διαφόρους «Mappers» του συστήματος και τα οποία με τη σειρά τους θα μεταβάλλουν το τελικό αποτέλεσμα. Παράλληλα, θα πρέπει να εξετάζεται κάθε φορά η εσκεμμένη ή μη διαρροή πληροφοριών από κάποιο «Mapper» κατά την λειτουργία του, η οποία θα υπονομεύσει την «Ιδιωτικότητα» του «Υποκειμένου» των δεδομένων αυτών.

Συνεπώς το μοντέλο των απειλών για το κατανεμημένο υπολογιστικό περιβάλλον [1] έχει ως εξής:

1) Υπολογιστικοί Κόμβοι που δεν Λειτουργούν Σωστά (Malfunctioning Compute Worker Nodes).

Οι κόμβοι του κατανεμημένου υπολογιστικού συστήματος μπορεί να μη λειτουργούν σωστά, είτε λόγω κακής παραμετροποίησης τους, είτε λόγω βλάβης. Αυτό συνεπάγεται ότι το αποτέλεσμα που θα παραχθεί θα είναι αλλοιωμένο με απώτερη συνέπεια να αλλοιωθεί και η ακεραιότητα του τελικού αποτελέσματος της διαδικασίας «MapReduce». Επιπρόσθετα, ένας τέτοιος κόμβος που δεν λειτουργεί σωστά μπορεί να προκαλέσει, εσκεμμένα ή μη, διαρροή πληροφοριών κατά την επεξεργασία τους.

2) Κόμβοι του Κατανεμημένου Υπολογιστικού Συστήματος που έχουν Καταληφθεί από τον Επιτιθέμενο (Compromised Worker Nodes).

Οι κόμβοι, οι οποίοι έχουν καταληφθεί και βρίσκονται στον έλεγχο του επιτιθέμενου, μπορεί να χρησιμοποιηθούν για την παρακολούθηση των επικοινωνιών, είτε μεταξύ των κόμβων «Mappers», είτε μεταξύ των «Mappers» με τον «Master» κόμβο, με

αντικειμενικό σκοπό την εκτόξευση, είτε «Replay» επιθέσεων, είτε «Man-in-the-Middle» επιθέσεων, είτε «Denial of Service» επιθέσεων σε αυτό το κατανεμημένο υπολογιστικό σύστημα. Επιπλέον μπορεί να χρησιμοποιηθούν, είτε για την υποκλοπή των δεδομένων, είτε για την τροποποίησή τους με απώτερο στόχο την αλλοίωση του τελικού αποτελέσματος.

3) Πλαστοί Κόμβοι Δεδομένων (Rogue Data Nodes).

Οι πλαστοί κόμβοι μπορεί να εισαχθούν στο σύστημα από τον επιτιθέμενο με σκοπό, είτε να υποκλέπτουν αντίγραφα των δεδομένων του συστήματος τα οποία αποστέλλονται σε αυτούς για επεξεργασία, είτε να διοχετεύουν κακόβουλο κώδικα στους λοιπούς «Mappers» και τον «Master» κόμβους, είτε να παράσχουν ψευδή στοιχεία για την αλλοίωση του τελικού αποτελέσματος. Η ικανότητα να δημιουργήσει ο επιτιθέμενος αντίγραφα κάποιου νόμιμου κόμβου και να τους εισάγει στο σύστημα, χωρίς να γίνει εύκολα αντιληπτός, είναι μια ευθεία απειλή που υφίστανται τα κατανεμημένα υπολογιστικά συστήματα, κυρίως αυτά που υποστηρίζονται από την «Νεφοϋπολογιστική» (Cloud) και τα εικονικά περιβάλλοντα (Virtual Environments).

Σύμφωνα λοιπόν με το ανωτέρω μοντέλο απειλών, υπάρχουν δυο κατευθύνσεις για την αντιμετώπισή τους. Αφενός θα πρέπει να εξασφαλιστεί η αξιοπιστία των κόμβων του συστήματος. Και αφετέρου θα πρέπει να προστατευτούν τα δεδομένα του συστήματος και η «Ιδιωτικότητά» τους, ακόμα και αν υπάρχουν στο σύστημα αυτό αναξιόπιστοι ή κακόβουλοι κόμβοι. Στο [51] αναφέρονται οι δέκα βέλτιστες πρακτικές που μπορούν να υλοποιηθούν για την αντιμετώπιση των ανωτέρω απειλών.

Αναλυτικότερα, η προστασία της αξιοπιστίας των κόμβων του συστήματος μπορεί να επιτευχθεί σε δυο βήματα [18]. Αρχικά, όταν ένας κόμβος του συστήματος στέλνει αίτημα στον «Master» κόμβο, θα πρέπει να εκτελεστεί αρχικά η διαδικασία της «Αμοιβαίας Αυθεντικοποίησης» (Mutual Authentication) [236] με την χρήση των αντίστοιχων μηχανισμών, όπως για παράδειγμα του μηχανισμού αυθεντικοποίησης «Kerberos». Δηλαδή θα πρέπει να αυθεντικοποιήσει ο «Master» κόμβος τον κόμβο που του έστειλε το αίτημα και αντίστροφα. Έτσι οι εργασίες του «Mapper» θα ανατίθενται μόνο σε αυθεντικοποιημένους κόμβους, οι οποίοι θα έχουν συγκεκριμένες ιδιότητες και θα ικανοποιούν τις προκαθορισμένες απαιτήσεις ασφαλείας. Παράλληλα θα πρέπει να εξεταστεί και η χρήση του «Trusted Platform Module - TPM» [237] για την ασφαλή αποθήκευση και προστασία των μυστικών

κλειδιών του κάθε κόμβου, που θα χρησιμοποιούνται κατά την διαδικασία της «Αμοιβαίας Αυθεντικοποίησης».

Σε συνέχεια της αρχικής αυθεντικοποίησης του κάθε κόμβου θα λαμβάνει χώρα περιοδικός έλεγχος των κόμβων του συστήματος από τον «Master» κόμβο, ως προς το κατά πόσο οι ρυθμίσεις ασφαλείας τους συμμορφώνονται με τις προκαθορισμένες πολιτικές ασφαλείας. Έτσι, αφενός θα είναι σε θέση να επισκευάσει τους κόμβους που δεν λειτουργούν σωστά ή έχουν πρόβλημα και αφετέρου θα μπορέσει να ανακαλύψει τους πλαστούς κόμβους. Μάλιστα, στην περίπτωση αυτή μπορεί να χρησιμοποιηθούν οι «Μηχανισμοί Ελέγχου Ακεραιότητας» (Integrity Check Mechanisms), που στηρίζονται στους αλγορίθμους της κρυπτογραφίας, για την ανίχνευση πλαστών ή και τροποποιημένων κόμβων του συστήματος.

Επιπρόσθετα η χρήση του μηχανισμού ελέγχου πρόσβασης «Mandatory Access Control – MAC» [238] μπορεί να διασφαλίσει ότι η πρόσβαση στα δεδομένα γίνεται μόνο από εξουσιοδοτημένους κόμβους του συστήματος «Big Data», οι οποίοι συμμορφώνονται πλήρως με την προκαθορισμένη πολιτική ασφαλείας για τα δεδομένα αυτά. Ωστόσο, ενώ αυτός ο μηχανισμός εγγυάται την ακεραιότητα των δεδομένων εισόδου στους «Mappers», δεν μπορεί να διασφαλίσει και να προλάβει την διαρροή δεδομένων στην έξοδο των αποτελεσμάτων.

Από την άλλη πλευρά, προκειμένου να προληφθεί η διαρροή πληροφοριών από τα αποτελέσματα της εξόδου των «Mappers» και συνεπώς να προστατευτεί η «Ιδιωτικότητα» των «Υποκειμένων» τους, απαιτείται η χρήση μηχανισμών «Αποχαρακτηρισμού» των δεδομένων (De-Identification Mechanisms) ή «Ανωνυμοποίησής» τους. Επομένως θα πρέπει να χρησιμοποιηθούν κατάλληλες τεχνικές που θα διασφαλίζουν την προστασία της «Ιδιωτικότητας» (Privacy Preserving Techniques) στα περιβάλλοντα αυτά, όπως για παράδειγμα η τεχνική «Ανωνυμοποίησης» «Differential Privacy». Η χρήση και η ασφάλεια αυτών των τεχνικών εξετάζεται επισταμένα στο 3^{ο} Κεφάλαιο της παρούσας μεταπτυχιακής διπλωματικής εργασίας.

Εντούτοις, αν και οι παραπάνω προτεινόμενες λύσεις αντιμετωπίζουν σε ικανοποιητικό βαθμό τα προβλήματα της αξιοπιστίας των κόμβων του συστήματος και την προστασία της «Ιδιωτικότητας» στα δεδομένα που διαχειρίζονται, εγείρονται

επιπλέον ζητήματα. Το πρώτο ζήτημα σχετίζεται με την επιβάρυνση στην απόδοση του συστήματος εξαιτίας της χρήσης του μηχανισμού «MAC». Το δεύτερο αφορά στο κατά πόσο η χρήση των τεχνικών αποχαρακτηρισμού των δεδομένων, όπως για παράδειγμα η χρήση της τεχνικής «Differential Privacy», εγγυάται την πλήρη προστασία της «Ιδιωτικότητας». Παράλληλα, τίθεται το ζήτημα μέχρι πιο βαθμό θα πρέπει να γίνει ο αποχαρακτηρισμός των δεδομένων αυτός, χωρίς να επηρεαστούν σημαντικά τα αποτέλεσμα που επιστρέφουν οι κόμβοι «Mappers» και να αλλοιωθεί το τελικό αποτέλεσμα. Όλα αυτά χρήζουν περαιτέρω διερεύνησης, προκειμένου εν τέλει να αντιμετωπιστούν αποτελεσματικά τα παραπάνω προβλήματα, χωρίς όμως να υπάρχει επιβάρυνση στην όλη λειτουργία αυτών των συστημάτων.

2.2.2 ΑΣΦΑΛΕΙΑ ΜΗ ΣΧΕΣΙΑΚΩΝ ΑΠΟΘΗΚΩΝ ΔΕΔΟΜΕΝΩΝ (NON RELATIONAL DATA STORES)

Οι υποδομές ασφαλείας για τις «Μη-Σχεσιακές Αποθήκες Δεδομένων» (Non-Relational Data Stores), τις οποίες υποστηρίζουν οι «NoSQL» βάσεις δεδομένων (NoSQL Databases), είναι ακόμα υπό εξέλιξη και βελτίωση [19]. Επιπρόσθετα, οι ίδιες οι «NoSQL» βάσεις δεδομένων αντιμετωπίζουν προβλήματα σε σχέση με την ασφάλειά τους. Έτσι για παράδειγμα, ανθεκτικές λύσεις για την αντιμετώπιση των «NoSQL Injections» δεν έχουν ωριμάσει ακόμα. Επιπλέον, το κάθε είδος των «NoSQL» βάσεων δεδομένων σχεδιάστηκε για να εξυπηρετήσει διαφορετική απαίτηση, ανάλογα με το εκάστοτε πρόβλημα που έπρεπε να επιλυθεί με τη βοήθεια των «Data Analytics». Ως εκ τούτου, το βάρος είχε δοθεί πρωτίστως στην λειτουργία τους παρά στην ασφάλεια κατά την σχεδίασή τους. Οπότε, καθώς η ασφάλεια των βάσεων αυτών δεν είχε ληφθεί αρχικά υπόψη, οι προγραμματιστές που τις χρησιμοποιούσαν συνήθως την ενσωμάτωναν εκ των υστέρων σε κάποια ενδιάμεση φάση της ανάπτυξής τους. Συνεπώς, οι «NoSQL» βάσεις δεδομένων δεν παρέχουν καθόλου εγγενή ασφάλεια κατά την λειτουργία τους, αλλά τουναντίον απαιτείται η υλοποίηση επιπρόσθετων λύσεων για την προστασία τους. Παράλληλα, το νέο περιβάλλον των «Big Data» εισάγει επιπλέον ζητήματα ασφαλείας, όσον αφορά αυτές τις λύσεις που υιοθετούνταν για την προστασία των «NoSQL» βάσεων δεδομένων, τα οποία χρήζουν περαιτέρω διερεύνησης.

Προκειμένου να τεθεί το πρόβλημα στην πραγματική του διάσταση θεωρούμε τις παρακάτω δυο περιπτώσεις [52]. Στην πρώτη περίπτωση, οι εταιρείες που

διαχειρίζονται μεγάλα «Μη Δομημένα» (Unstructured) σύνολα δεδομένων προφανώς θα επωφεληθούν με την μετάβαση από το παραδοσιακό «Σχεσιακό Μοντέλο» βάσεων δεδομένων (Traditional Relational Database - RDB) στο «Μη Σχεσιακό Μοντέλο» (NoSQL Database), καθώς αυτό θα τους δώσει την δυνατότητα να αποθηκεύσουν και να διαχειριστούν τεράστιες ποσότητες δεδομένων, είτε «Στατικών» (Static Data), είτε σε «Ροή/Κίνηση» (Streaming Data) προκειμένου να επιτύχουν τους επιχειρηματικούς τους στόχους. Ωστόσο οι «Απειλές» (Threats) που έχουν διαγνωστεί σε ευρέως χρησιμοποιούμενες «Μη Σχεσιακές» βάσεις δεδομένων, κατόπιν λεπτομερούς ανάλυσης των «Απειλών» σε αυτές (Threat Analysis) με τη βοήθεια των τεχνικών της «Ανάλυσης Απειλών» (Threat Modeling Techniques) [239], αποδεικνύουν ότι οι «Μη Σχεσιακές» βάσεις δεδομένων παρέχουν μια πολύ επιφανειακή ασφάλεια συγκρινόμενες με τις παραδοσιακές «Σχεσιακές» βάσεις δεδομένων. Γενικά διαπιστώνεται ότι η φιλοσοφία ως προς την ασφάλεια αυτών των βάσεων δεδομένων πηγάζει από την εφαρμογή εξωτερικών μηχανισμών ασφαλείας. Έτσι οι εταιρείες που τις χρησιμοποιούν, για να ελαττώσουν την πιθανότητα να συμβούν περιστατικά ασφαλείας, θα πρέπει παράλληλα να αναθεωρήσουν τις πολιτικές ασφαλείας για την προστασία τους και να θωρακίσουν από πλευράς ασφάλειας τις ίδιες τις «Μη Σχεσιακές» βάσεις δεδομένων σε τέτοιο βαθμό, που να εξισώνεται η ασφάλεια τους με αυτή των «Σχεσιακών» βάσεων δεδομένων, χωρίς όμως ταυτόχρονα αυτό να αποβαίνει σε βάρος των λειτουργικών χαρακτηριστικών τους. Από την άλλη πλευρά, η ικανότητα των «Μη Σχεσιακών» βάσεων δεδομένων να εκτελούν ανάλυση σε δομημένα και μη δεδομένα σε αποθήκευση είναι ασυγκρίτως μεγαλύτερη σε σχέση με αυτή των «Σχεσιακών» βάσεων δεδομένων. Συνεπώς, είναι σημαντικό τα κενά ασφαλείας που προκύπτουν από την χρήση των «Μη Σχεσιακών» βάσεων δεδομένων να καλύπτονται, χωρίς όμως αυτό να αποβαίνει σε βάρος της εξαιρετικής τους ικανότητας στον τομέα της «Αναλυτικής».

Η δεύτερη περίπτωση αφορά τους παραδοσιακούς παρόχους υπηρεσιών «Αναλυτικής» (Analytics As A Service - AaaS) μέσω «Νεφοϋπολογιστικής» (Cloud Infrastructure) [1]. Αυτό επιτυγχάνεται με τον συνδυασμό από την μια πλευρά μιας σουίτας εργαλείων, η οποία είναι ικανή να αναλύσει αποτελεσματικά, τόσο στατικά δεδομένα, όσο και δεδομένα σε κίνηση, και από την άλλη πλευρά των «Μη Σχεσιακών» βάσεων δεδομένων, οι οποίες χρησιμοποιούνται για την ενδιάμεση αποθήκευση των δεδομένων αυτών και των αποτελεσμάτων της ανάλυσης. Έτσι, στην περίπτωση αυτή, διάφοροι χρήστες της υπηρεσίας αλληλεπιδρούν με την

σουίτα εργαλείων, τροφοδοτώντας την είτε με στατικά, είτε με κινούμενα δεδομένα τα οποία θα τύχουν επεξεργασίας από τα αντίστοιχα εργαλεία «Αναλυτικής». Οπότε είναι αναγκαίο να αποθηκεύονται τα εκάστοτε σύνολα δεδομένων, που παρέχονται στην υπηρεσία από τους διάφορους χρήστες της, στις «Μη Σχεσιακές» βάσεις δεδομένων για την ενδιάμεση επεξεργασία τους. Αυτά τηρούνται σε αυτές μέχρι να επιστραφούν τα αποτελέσματα της «Αναλυτικής» επί αυτών στους ενδιαφερόμενους χρήστες. Αυτό όμως ενέχει κινδύνους, καθώς τα υπάρχοντα μέτρα ασφαλείας των «Μη Σχεσιακών» βάσεων δεδομένων καθιστούν σχεδόν αδύνατο τον διαχωρισμό και την προστασία των προσωπικών δεδομένων που ανήκουν στους διάφορους χρήστες της υπηρεσίας αυτής. Έτσι προκύπτει το πρόβλημα, πως θα διασφαλιστεί η προστασία της «Ιδιωτικότητας» των χρηστών από την χρήση των «Μη Σχεσιακών» βάσεων δεδομένων στις υπηρεσίες αυτές.

Από τα παραπάνω διαπιστώνεται ότι τα αξιοσημείωτα πλεονεκτήματα των «Μη Σχεσιακών» μοντέλων δεδομένων, τα οποία είναι η ευελιξία, η κλιμάκωση και η αποδοτικότητα, παράλληλα συνοδεύονται από σοβαρούς κινδύνους ασφαλείας [20]. Αυτό οφείλεται κατά κύριο λόγο στο γεγονός ότι τα «Μη Σχεσιακά» (NoSQL) μοντέλα δεδομένων αρχικά είχαν σχεδιαστεί με την προοπτική να διαχειρίζονται αποτελεσματικά μεγάλα σύνολα δεδομένων και ως εκ τούτου είχε δοθεί περιορισμένη έμφαση στην ασφάλεια τους [21]. Από την άλλη πλευρά όμως, η χρήση των βάσεων αυτών είναι μονόδρομος, ακριβώς λόγω των πλεονεκτημάτων τους. Έτσι, προκειμένου να αντιμετωπιστούν τα προβλήματα αυτά, ήταν επιτακτικό να σχεδιαστούν και να υλοποιηθούν «ad-hoc» λύσεις κάθε φορά.

Γενικά, το μοντέλο απειλών για τα «Μη Σχεσιακά» (NoSQL) μοντέλα δεδομένων [1] έχει τα παρακάτω έξι σενάρια:

1) Ακεραιότητα Συναλλαγών (Transactional Integrity).

Ένα από τα σημαντικότερα μειονεκτήματα των «Μη Σχεσιακών» (NoSQL) μοντέλων δεδομένων είναι η χαλαρή προσέγγιση που έχουν στην διασφάλιση της ακεραιότητας των συναλλαγών. Από την άλλη πλευρά όμως, η επιβολή πολύπλοκων περιορισμών ακεραιότητας στην λειτουργία τους θα προξενούσε την απώλεια των βασικών χαρακτηριστικών τους, ήτοι την αποδοτικότητα, την κλιμάκωση και την ευελιξία.

2) Χαλαροί Μηχανισμοί Αυθεντικοποίησης.

Τα «Μη Σχεσιακά» (NoSQL) μοντέλα δεδομένων χρησιμοποιούν ασθενείς μηχανισμούς αυθεντικοποίησης και χαλαρούς μηχανισμούς αποθήκευσης κωδικών

πρόσβασης «Passwords». Αυτό το γεγονός τα καθιστά ευάλωτα σε επιθέσεις «Replay» και «Password Brute Force». Συγκεκριμένα, σε αυτά χρησιμοποιείται η αυθεντικοποίηση «HTTP Basic» ή «Digest Based», η οποία είναι ευάλωτη σε «Replay» και «Man in the Middle» επιθέσεις. Επιπλέον, σε αυτά συνήθως χρησιμοποιείται για τις επικοινωνίες το πρωτόκολλο «REST» [240] που βασίζεται στο πρωτόκολλο «HTTP», γεγονός που το καθιστά ευάλωτο σε διάφορες επιθέσεις, όπως «Cross-Site Scripting», «Cross-Site Request Forgery», «Injection Attacks» κλπ. Από τα παραπάνω, εξάγεται το συμπέρασμα ότι σε πιθανό περιστατικό ασφαλείας είναι δυνατό ο επιτιθέμενος να αποκτήσει πρόσβαση, τόσο στο αρχείο διαχείρισης και παραμετροποίησης της βάσης δεδομένων, όσο και περαιτέρω στο σύστημα αρχείων αυτής. Επιπρόσθετα, άλλο ένα πρόβλημα είναι ότι δεν παρέχεται η δυνατότητα να ενσωματωθούν στα μοντέλα αυτά εργαλεία αυθεντικοποίησης από «Τρίτα» μέρη για την αντιμετώπιση του προβλήματος αυτού. Από την άλλη πλευρά, αν και σε κάποιες από τις υπάρχουσες «Μη Σχεσιακές» βάσεις δεδομένων παρέχεται η δυνατότητα αυτή τοπικά σε επίπεδο κόμβων, εντούτοις δεν υποστηρίζεται ευρύτερα και μεταξύ κόμβων που ανήκουν σε διαφορετικό τομέα «Cluster».

3) Ανεπαρκείς Μηχανισμοί Εξουσιοδότησης Πρόσβασης (Authorization).

Οι τεχνικές που χρησιμοποιούνται για την εξουσιοδότηση της πρόσβασης στα δεδομένα των «NoSQL» βάσεων δεδομένων διαφέρουν μεταξύ των διαφόρων υλοποιήσεων τους. Έτσι, στις περισσότερες από τις δημοφιλείς λύσεις προτιμάται η εφαρμογή της εξουσιοδότησης πρόσβασης στα υψηλότερα επίπεδα του «OSI», δηλαδή σε επίπεδο εφαρμογής, παρά στα χαμηλότερα, δηλαδή σε επίπεδο βάσης δεδομένων. Επιπλέον, τα συστήματα αυτά δεν υποστηρίζουν τους μηχανισμός ελέγχου πρόσβασης «Role - Based Access Control – RBAC», γιατί ο ορισμός ρόλων και ομάδων για τους χρήστες στα συστήματα αυτά καθίσταται εξαιρετικά επίπονη διαδικασία.

4) Ευαισθησία σε «Injection» Επιθέσεις.

Το γεγονός ότι τα «Μη Σχεσιακά» (NoSQL) μοντέλα δεδομένων συνδυάζουν χαλαρά, από πλευράς ασφάλειας, πρωτόκολλα και μηχανισμούς, τα καθιστά ευάλωτα σε πληθώρα επιθέσεων, όπως «JSON Injection», «Array Injection», «View Injection», «REST Injection», «GQL Injection», «Schema Injection» κλπ, που αποσκοπούν στην παράνομη απόκτηση πρόσβασης στο σύστημα αρχείων τους και στην περαιτέρω εκτέλεση κακόβουλων ενεργειών. Για παράδειγμα, μπορεί ένας επιτιθέμενος μέσω της επίθεσης «Schema Injection» να εισάγει στην βάση δεδομένων χιλιάδες στήλες με δεδομένα της επιλογής του, προκειμένου να επιτύχει ένα εύρος συνεπειών που

κυμαίνεται από την αλλοίωση των δεδομένων της, στοχεύοντας στην ακεραιότητά της (Integrity), έως να την καταστήσει μη διαθέσιμη (DoS Attack), στοχεύοντας στην διαθεσιμότητά της (Availability).

5) Απώλεια της Συνέπειάς της Βάσης Δεδομένων.

Η αδυναμία να εφαρμοστούν ταυτόχρονα σε μια κατανεμημένη βάση δεδομένων όλα τα χαρακτηριστικά του θεωρήματος «CAP» [241], ήτοι «Συνέπεια» (Consistency), «Διαθεσιμότητα» (Availability) και «Ευελιξία Κατανομής» (Partition Tolerance), υπονομεύει την αξιοπιστία των επιστρεφόμενων αποτελεσμάτων της βάσης αυτής. Καθώς λοιπόν οι κατανεμημένες βάσεις δεδομένων εστιάζουν στην «Διαθεσιμότητα» και την «Ευελιξία Κατανομής» εις βάρος της «Συνέπειας», σύμφωνα με το ανωτέρω θεώρημα, δεν υπάρχει εγγύηση στους χρήστες τους ότι τα αποτελέσματα που θα πάρουν σε κάποια δεδομένη στιγμή, ότι θα είναι συνεπή. Αυτό μπορεί να οφείλεται στο γεγονός ότι κάποιος κόμβος μπορεί να μην έχει συγχρονιστεί πλήρως με τον κόμβο που τηρεί το ενημερωμένο αντίγραφο μιας εγγραφής. Για παράδειγμα, αυτό μπορεί να προέρχεται από το γεγονός ότι οι αλγόριθμοι που χρησιμοποιούνται για την εγγυημένη αναπαραγωγή των δεδομένων στους κόμβους του συστήματος, βελτιώνοντας έτσι την συνέπεια της βάσης, μπορεί να αποτύχουν εξαιτίας της πιθανής αστοχίας κάποιου κόμβου, με συνέπεια την ανισορροπία των εγγραφών μεταξύ των κόμβων του συστήματος αυτού.

6) Απειλές «Εκ των Έσω» (Insider Attacks).

Οι «Μη Σχεσιακές» βάσεις δεδομένων καθίστανται ευάλωτες σε επιθέσεις «Εκ των Έσω» εξαιτίας των χαλαρών μηχανισμών ασφαλείας που έχουν υλοποιηθεί σε αυτές. Μάλιστα, οι επιθέσεις αυτές μπορεί να μην γίνουν αντιληπτές εξαιτίας της ανυπαρξίας ή της χαλαρής τήρησης ιστορικού ενεργειών (Logging) και της έλλειψης των αντίστοιχων μεθόδων για την ανάλυσή του (Log Analysis Methods). Συνεπώς, καθώς σε αυτές αποθηκεύονται κρίσιμα δεδομένα των χρηστών της, είναι δύσκολο να διασφαλιστεί ότι τελικά ο κάτοχος τους έχει και τον αποκλειστικό έλεγχό τους, ακριβώς λόγω της χαλαρής ασφάλειας που έχει υλοποιηθεί σε αυτές.

Για την αντιμετώπιση των ανωτέρω σοβαρών προβλημάτων διάφορες βέλτιστες πρακτικές αναφέρονται στα [1] και [51], οι οποίες παρουσιάζονται στη συνέχεια. Αναλυτικότερα, όσον αφορά την ακεραιότητα των συναλλαγών (Transactional Integrity), τεχνικές όπως η «Architectural Trade-off Analysis Method (ATAM)», οι οποίες εξειδικεύονται στον προσδιορισμό της βέλτιστης λύσης στον συσχετισμό των απαιτήσεων κατά τον σχεδιασμό των συστημάτων, θα αποβούν χρήσιμες στην

περίπτωση αυτή, καθώς θα συμβάλουν σημαντικά στην εκτίμηση του βέλτιστου επιπέδου ασφάλειας που θα πρέπει να υιοθετηθεί, χωρίς όμως ταυτόχρονα να επηρεαστεί σημαντικά η απόδοση των «Μη Σχεσιακών» (NoSQL) μοντέλων δεδομένων.

Από την άλλη πλευρά θα πρέπει να διασφαλιστεί η ακεραιότητα των δεδομένων (Data Integrity) που αποθηκεύονται στις «Μη Σχεσιακές» βάσεις δεδομένων. Έτσι, λαμβάνοντας υπ' όψιν τους ήδη χαλαρούς μηχανισμούς αυθεντικοποίησης και εξουσιοδότησης πρόσβασης που χρησιμοποιούνται, καθίσταται επιτακτικό να κρυπτογραφούνται τα δεδομένα που βρίσκονται αποθηκευμένα σε αυτές, παρά την επιβάρυνση που ενδεχομένως θα επιφέρει η ενέργεια αυτή στην απόδοση τους. Βέβαια, μία εναλλακτική λύση, η οποία θα καταστήσει την διαδικασία της κρυπτογράφησης πιο γρήγορη και η οποία θα εξομαλύνει τις αρνητικές επιπτώσεις της στην απόδοση του συστήματος, είναι η κρυπτογράφηση σε επίπεδο υλικού «Hardware Appliance Based Encryption». Ωστόσο και αυτή η λύση έχει αδυναμίες. Αφενός δημιουργεί δέσμευση με τον πάροχο αυτού του εξεζητημένου υλικού. Αφετέρου σε πολλές περιπτώσεις χρησιμοποιούνται ασθενή κλειδιά κρυπτογράφησης, γεγονός που την καθιστά ευάλωτη σε επιθέσεις για την εύρεση των κλειδιών αυτών. Παράλληλα θα πρέπει να διασφαλιστεί και η εμπιστευτικότητα των δεδομένων (Data Confidentiality) που είναι σε κίνηση. Η χρήση του πρωτοκόλλου «SSL/TLS» καθίσταται επιτακτική κατά την εγκαθίδρυση, τόσο των επικοινωνιών μεταξύ χρήστη και εξυπηρετητή, όσο και μεταξύ των κόμβων που συμμετέχουν στο σύστημα.

Επιπρόσθετα, θα πρέπει η εκάστοτε «Μη Σχεσιακή» βάση δεδομένων να υλοποιηθεί με τέτοιο τρόπο, που να υποστηρίζει την ενσωμάτωση μηχανισμών αυθεντικοποίησης και εξουσιοδότησης πρόσβασης από «Τρίτα» μέρη, έτσι ώστε να παρέχεται ασφάλεια σε όλα τα επίπεδα. Εναλλακτικά, θα πρέπει οι μηχανισμοί αυτοί να ενσωματωθούν, είτε σε κάποια εφαρμογή, είτε σε ένα ενδιάμεσο επίπεδο ασφαλείας, τα οποία θα παρεμβάλλονται μεταξύ των χρηστών και της βάσης δεδομένων. Μάλιστα, στους μηχανισμούς αυτούς δεν θα πρέπει ποτέ να αποθηκεύονται ή/και να μεταδίδονται οι κωδικοί πρόσβασης σε καθαρό κείμενο (Plaintext), αλλά αντιθέτως θα πρέπει να κρυπτογραφούνται με την χρήση ασφαλών κρυπτογραφικών ή «Hash» αλγορίθμων. Παράλληλα, η εφαρμογή των τεχνικών «Fuzzing» [244], με τις οποίες δίδονται ως είσοδοι στο σύστημα τυχαίες, ή μη-

έγκυρες, ή απροσδόκητες τιμές, θα αποτελέσει μια ιδανική μέθοδο για την αποκάλυψη των πιθανών αδυναμιών των μηχανισμών αυτών της «Μη Σχεσιακής» βάσης δεδομένων, στην οποία χρησιμοποιείται το πρωτόκολλο «HTTP» για την εγκαθίδρυση των συνδέσεων με τους χρήστες της [1].

Επίσης, για την προστασία της συνέπειας της βάσης, θα πρέπει κατ' αρχάς να διασφαλιστούν οι επικοινωνίες μεταξύ των κόμβων (Nodes) των διαφόρων τμημάτων (Clusters) του συστήματος αυτού. Αυτό θα επιτευχθεί αν ο εκάστοτε κόμβος, πριν την εγκαθίδρυση μιας ασφαλούς επικοινωνίας με κάποιο άλλο κόμβο του συστήματος, έχει προηγουμένως επιβεβαιώσει ότι ο κόμβος αυτός τηρεί το επιθυμητό επίπεδο εμπιστοσύνης. Οπότε στην συνέχεια, με την χρήση έξυπνων μηχανισμών, όπως για παράδειγμα με την χρήση των αλγορίθμων «Hash», θα μπορεί να εξασφαλιστεί ότι τα δεδομένα αναπαράγονται και τηρούνται ομαλά και σωστά στους κόμβους του συστήματος, ακόμα και σε περίπτωση αστοχίας κάποιων κόμβων του. Επιπλέον, προτείνεται τα «Μη Σχεσιακά» μοντέλα δεδομένων να λειτουργούν σε ασφαλή περιβάλλοντα «Trusted Environments» [242], τα οποία θα διασφαλίζουν ότι μόνο αξιόπιστα προγράμματα μπορούν να συνδεθούν σε αυτά [51]. Τέλος, κατάλληλες τεχνικές σήμανσης των δεδομένων (Data Tagging), οι οποίες με τη βοήθεια έξυπνων αλγορίθμων θα επιβάλλουν την χρονοσήμανση τους (Time Stamp) καθώς αυτά θα αντλούνται από τις διάφορες πηγές τους, θα αντιμετωπίσουν αποτελεσματικά την μη-εξουσιοδοτημένη τροποποίηση τους από κάποιο επιτιθέμενο. Επιπλέον οι τεχνικές αυτές θα πιστοποιούν και την αυθεντικότητα των αποθηκευμένων δεδομένων της βάσης [51].

Επιπλέον, θα πρέπει παράλληλα με τους μηχανισμούς ελέγχου πρόσβασης να χρησιμοποιηθούν και οι κατάλληλοι μηχανισμοί καταγραφής κίνησης (Logging Mechanisms) και ταχείας ανάλυσης της κίνησης (On-the-fly Log Analysis) [243], έτσι ώστε να ανιχνευτούν οι κακόβουλες ενέργειες, τόσο εκ των έσω, όσο και απ' έξω. Οπότε, μέσα από την ανάλυση των καταγεγραμμένων ενεργειών που πραγματοποιήθηκαν στην βάση, θα είναι δυνατό να συσχετιστούν οι ενέργειες αυτές και να αποκαλυφθούν πιθανές επιθέσεις στο σύστημα.

Είναι προφανές λοιπόν από τα ανωτέρω, ότι η ασφάλεια των «Μη Σχεσιακών» βάσεων δεδομένων και η προστασία των δεδομένων τους δεν μπορεί να στηριχτεί αποκλειστικά στην αδύναμη και εύκολα παραβιάσιμη ασφάλεια που προσφέρουν

εγγενώς οι ίδιες [22]. Τουναντίον, η ασφάλειά τους θα πρέπει να ενισχυθεί, ή με την δημιουργία ενός ενδιάμεσου επιπέδου ασφάλειας μεταξύ των χρηστών και της βάσης, ή με το να συμπεριληφθούν αυτές οι βάσεις μέσα σε ευρύτερα περιβάλλοντα και υλοποιήσεις (Frameworks). Οποιοδήποτε από αυτά υιοθετηθεί, θα επωμιστεί αποκλειστικά με το συνολικό βάρος της ασφάλειας της εκάστοτε «Μη Σχεσιακής» βάσης δεδομένων, καθώς σε αυτό θα είναι ενσωματωμένοι όλοι οι κατάλληλοι και απαραίτητοι μηχανισμοί ασφαλείας, προκειμένου να προστατευτεί, τόσο η ίδια η βάση δεδομένων, όσο και τα δεδομένα αυτής. Οπότε, η απαίτηση για την προστασία των δεδομένων θα επιβάλλεται ανάλογα από το ενδιάμεσο αυτό επίπεδο ασφαλείας ή υλοποίηση, το οποίο θα διασφαλίζει, αφενός ότι δεν υπάρχει άμεση πρόσβαση στα δεδομένα της βάσης και αφετέρου ότι αυτά είναι διαθέσιμα μόνο σε όποιον ικανοποιεί τις απαιτήσεις ασφαλείας του [1]. Παράδειγμα τέτοιας υλοποίησης είναι το κατανεμημένο σύστημα «Hadoop» που διανέμεται από την «Cloudera», στο οποίο χρησιμοποιείται ο ισχυρός μηχανισμός αυθεντικοποίησης «Kerberos» [245], ο οποίος μπορεί να:

✓ Αποτρέψει την δυνατότητα σε κάποιο επιτιθέμενο να προσποιηθεί ότι είναι κάποιος νόμιμος χρήστης του συστήματος.

✓ Επιβάλλει στον κάθε χρήστη να αυθεντικοποιείται οποτεδήποτε εκτελεί μεμακρυσμένη σύνδεση στο σύστημα (Remote Procedure Call - RPC).

✓ Διασφαλίζει τον κατάλληλο βαθμό απομόνωσης, καθώς όλες οι ενέργειες εκτελούνται υπό τον λογαριασμό αποκλειστικά του χρήστη που τις αιτήθηκε.

✓ Αποτρέπει τον εκάστοτε χρήστη από το να προβεί σε ανάλυση των ομάδων που αφορούν τους κόμβους του «Hadoop» συστήματος, δηλαδή τους «Hadoopmaster», «Cluster» και «Job Tracker» κόμβους του.

Μάλιστα, το πλεονέκτημα αυτών των λύσεων είναι ότι εξασφαλίζεται η υψηλή απόδοση των «Μη Σχεσιακών» βάσεων δεδομένων, καθώς και η δυνατότητα επέκτασής τους ανάλογα με την ζήτηση, ενώ παράλληλα επιτυγχάνεται ο βέλτιστος βαθμός ασφάλειας για αυτές. Παράλληλα, αυτές οι λύσεις θα είναι επιφορτισμένες και με την κρυπτογράφηση των δεδομένων που είναι αποθηκευμένα στις «Μη Σχεσιακές» βάσεις δεδομένων. Έτσι για παράδειγμα, το εργαλείο «Hadoop» χρησιμοποιεί κρυπτογράφηση σε επίπεδο αρχείων, για να παρέχει απαρέγκλιτη προστασία ανεξάρτητη του λειτουργικού συστήματος και του τρόπου αποθήκευσης των δεδομένων. Οπότε, με αυτές τις προσεγγίσεις επιτυγχάνεται περιορισμένη πρόσβαση στα αποθηκευμένα δεδομένα, ενώ ταυτόχρονα διατηρείται το επίπεδο της

βάσης δεδομένων στην προκαθορισμένη λειτουργία της, διατηρώντας έτσι την αναλυτική της ικανότητα. Παράλληλα, εξασφαλίζεται ότι οι ιδιοκτήτες των δεδομένων έχουν καλύτερο έλεγχο πάνω σε αυτά και ότι θα προλαμβάνονται ή θα ανιχνεύονται έγκαιρα οι επιθέσεις εκ των έσω.

Ωστόσο, οι παραπάνω λύσεις χρήζουν περαιτέρω έρευνας, καθώς υπάρχουν επιπλέον ζητήματα που παραμένουν άλυτα στο περιβάλλον των «Big Data». Έτσι, παρά την πληθώρα των εργαλείων κρυπτογράφησης, αυτά δεν είναι αποδοτικά στην διαχείριση των δεδομένων που βρίσκονται σε κίνηση (Streaming Data) και δεν μπορούν να εξασφαλίσουν την ταχεία επεξεργασία τους στη μνήμη (In-Memory Processing). Επιπλέον, αν και η κρυπτογράφηση των δεδομένων είναι η καλύτερη λύση για την αντιμετώπιση των ζητημάτων ασφαλείας, εντούτοις φαίνεται να επιφέρει σημαντικό κόστος στην απόδοση των συστημάτων αυτών.

2.3 ΕΞΕΤΑΣΗ ΤΗΣ ΑΣΦΑΛΕΙΑΣ ΕΠΙ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ (BIG DATA MANAGEMENT SECURITY)

Σε δεύτερη φάση εξετάζεται στην ενότητα αυτή η ασφάλεια της διαδικασίας για την επεξεργασία των δεδομένων από τα συστήματα «Big Data». Ειδικότερα μελετώνται τα ακόλουθα ζητήματα. Το πρώτο αφορά τον έλεγχο της προέλευσης των δεδομένων που θα εισέλθουν στο σύστημα για να τύχουν επεξεργασίας. Το δεύτερο σχετίζεται με την ασφαλή αποθήκευση των δεδομένων αυτών στις αποθήκες δεδομένων του συστήματος και παράλληλα στην τήρηση ιστορικού ενεργειών. Ενώ το τρίτο εστιάζει στην εκτέλεση αποτελεσματικών επιθεωρήσεων επί του συνόλου της διαδικασίας επεξεργασίας των δεδομένων αυτών. Αναλυτικότερα τα ζητήματα αυτά αναπτύσσονται στις επόμενες υποενότητες.

2.3.1 ΕΛΕΓΧΟΣ ΠΡΟΕΛΕΥΣΗΣ ΔΕΔΟΜΕΝΩΝ (DATA PROVENANCE)

Υπάρχουν διάφορες προσεγγίσεις για την έννοια της «Προέλευσης» των δεδομένων, οι οποίες την συνδέουν, άλλοτε με τον ιδιοκτήτη των δεδομένων αυτών, άλλοτε με το ποιος τα επιμελείται και άλλοτε με το που και πότε αυτά δημιουργήθηκαν. Ειδικότερα όμως για το περιβάλλον των «Big Data», στο οποίο το σύνολο των δεδομένων που προέρχεται από τις διάφορες πηγές μετατρέπεται σε χρήσιμη και πολύτιμη γνώση μέσα από την διαδικασία της «Αναλυτικής», οι

πληροφορίες που σχετίζονται με την προέλευση των δεδομένων αυτών, ήτοι τα «Μεταδεδομένα» (Metadata) [246], αποκτούν ιδιαίτερο ενδιαφέρον και σημασία, καθώς μπορούν να χρησιμοποιηθούν ταυτόχρονα για διάφορους σκοπούς, όπως να βεβαιώσουν την πηγή προέλευσής των δεδομένων αυτών και την αυθεντικότητά τους, να συμβάλουν στην αξιολόγηση των διαφόρων ισχυρισμών και παραδοχών που πιθανό να γίνουν κατά την ανάλυση των δεδομένων και να δικαιολογήσουν τα απροσδόκητα αποτελέσματα που μπορεί να προκύψουν από αυτή [3]. Παράλληλα, αυτά αποτελούν αναπόσπαστο μέρος της διαδικασίας διεξαγωγής επιθεωρήσεων στο σύστημα και επιπλέον βοηθούν, τόσο τους «Επιθεωρητές», όσο και τα ίδια τα «Υποκείμενα» των δεδομένων, να παρακολουθήσουν αδιάλειπτα την διαδρομή της επεξεργασίας των προσωπικών τους δεδομένων και να αποδώσουν τυχόν ευθύνες σε πιθανή κακή χρήση τους [23].

Ωστόσο, η ένταξη και η διαβίβαση των δεδομένων προέλευσης στα συστήματα «Big Data» μπορεί να συμβάλλει στην ανεπιθύμητη αποκάλυψη προσωπικών δεδομένων, ακόμα και μετά από πολλούς μετασχηματισμούς [3]. Έτσι για παράδειγμα, όταν ένα άτομο παρακολουθεί την πορεία των δεδομένων του με την χρήση των μεταδεδομένων, μπορεί τυχαία να αποκτήσει πρόσβαση και στα δεδομένα των άλλων χρηστών, αν τα δεδομένα τους αποτελούν μέρος της ίδιας διαδικασίας «Αναλυτικής». Άλλο παράδειγμα αποτελεί η πιθανή αποκάλυψη της ταυτότητας ενός ατόμου μέσα από τον συνδυασμό των μεταδεδομένων που ακολουθούν τα ιατρικά αρχεία, όπως είναι ο ιατρικός φορέας και η ημερομηνία δημιουργία τους, με κάποιες επιπλέον πληροφορίες, όπως τα αρχεία καταγραφής της θέσης του κινητού τηλεφώνου. Επίσης, η χρήση τους στο περιβάλλον των «Big Data» θα οδηγήσει στην παραγωγή ενός μεγάλου «Γραφήματος Μεταδεδομένων Προέλευσης» (Provenance Metadata Graph) με μεγάλη πολυπλοκότητα. Μάλιστα η ανάλυσή του για την ανίχνευση των αλληλεξαρτήσεων μεταξύ τους, για λόγους ασφαλείας, καθίσταται υπολογιστικά απαιτητική [1]. Παρά ταύτα, το βασικό πλεονέκτημα των «Μεταδεδομένων Προέλευσης» είναι ότι αυτά δεν ενοποιούνται με άλλα και επομένως οι πληροφορίες που παρέχουν είναι ακριβείς καθ' όλη την διάρκεια της διαδικασίας της «Αναλυτικής» [3].

Έτσι στην σύγχρονη ψηφιακή εποχή, αρκετές εφαρμογές απαιτούν τα ψηφιακά αρχεία που επεξεργάζονται να συνοδεύονται από πληροφορίες που σχετίζονται με την δημιουργία τους (Metadata) για λόγους ασφαλείας. Αν για παράδειγμα

θεωρήσουμε την περίπτωση που διερευνάται η οικονομική εκμετάλλευση εμπιστευτικών πληροφοριών σε κάποια χρηματοπιστωτική εταιρεία, ή την περίπτωση που διερευνάται κατά πόσο είναι ακριβή τα δεδομένα που παρέχει μια πηγή δεδομένων στην διεξαγωγή ερευνών, αυτό θα επιτευχθεί ύστερα από την ανάλυση των «Μεταδεδομένων Προέλευσης» (Provenance Metadata) που συνοδεύουν τα υπό διερεύνηση δεδομένα στην εκάστοτε περίπτωση. Ωστόσο, η ανάλυση αυτή είναι χρονοβόρα διαδικασία και απαιτεί την χρήση γρήγορων αλγορίθμων, προκειμένου να είναι αποτελεσματική και γρήγορη [52]. Επίσης, αυτά χρησιμοποιούνται κατά τις επιθεωρήσεις, αλλά και για να διαπιστωθεί η συμμόρφωση με την εκάστοτε νομοθεσία, τα συμβόλαια ή και τα πρότυπα.

Επομένως, για να είναι ασφαλής η χρησιμοποίηση των «Μεταδεδομένων Προέλευσης» στο περιβάλλον των «Big Data», απαιτείται τα αρχεία αυτά να είναι αξιόπιστα και ολοκληρωμένα, να προστατεύουν παράλληλα την ιδιωτική ζωή και να είναι ελεγχόμενη η πρόσβαση σε αυτά. Ταυτόχρονα, λόγω των χαρακτηριστικών των «Big Data», τα ζητήματα που σχετίζονται με την διαθεσιμότητα και την επεκτασιμότητα τους θα πρέπει να αντιμετωπίζονται με ιδιαίτερη προσοχή. Έτσι οι απειλές, που αντιμετωπίζει η χρήση των «Μεταδεδομένων Προέλευσης» στο περιβάλλον των «Big Data» [1], συνοψίζονται στις ακόλουθες κατηγορίες:

1) Δυσλειτουργία Τμημάτων της Υποδομής «Big Data».

Τα «Big Data» συστήματα αποτελούνται από πάρα πολλά επιμέρους τμήματα, με αποτέλεσμα η σύνθεση των «Μεταδεδομένων Προέλευσης» από αυτά να συνεπάγεται την δημιουργία μεγάλων και πολύπλοκων «Γραφημάτων Προέλευσης». Παράλληλα όμως, σε αυτό το πολύπλοκο σύστημα δεν μπορεί να αποκλειστεί η περίπτωση κατά την οποία κάποια τμήματα της υποδομής αυτής σποραδικά δυσλειτουργούν. Στην περίπτωση αυτή, η άμεση συνέπεια είναι τα «Μεταδεδομένα Προέλευσης» που παράγονται από τα τμήματα αυτά να μην είναι διαθέσιμα καθ' όλο το διάστημα της δυσλειτουργίας. Άμεσο επακόλουθο αυτού του γεγονότος είναι η δημιουργία ασυνέπειας στο «Γράφημα Μεταδεδομένων Προέλευσης», ενώ παράλληλα μειώνεται η διαθεσιμότητα και η αξιοπιστία των μεταδεδομένων που προέρχονται από τα δυσλειτουργούντα τμήματα της υποδομής αυτής.

2) Εξωτερικές Επιθέσεις στην Υποδομή «Big Data».

Το γεγονός ότι τα «Μεταδεδομένα Προέλευσης» είναι ζωτικής σημασίας για τα συστήματα «Big Data», αυτομάτως τα καθιστά πρωταρχικό στόχο για τους επιτιθέμενους στα συστήματα αυτά. Έτσι ένας εξωτερικός επιτιθέμενος έχει τις

δυνατότητες, είτε να τα πλαστογραφήσει, είτε να τα τροποποιήσει, είτε να τα αναπαράγει, είτε να καθυστερήσει αδικαιολόγητα την μετάδοσή τους προκειμένου να καταστρέψει την αξιοπιστία και την ακεραιότητά τους. Επιπλέον μπορεί να τα υποκλέψει και να τα αναλύσει, για να παραβιάσει την «Ιδιωτικότητα» των «Υποκειμένων» των δεδομένων στα οποία αυτά αναφέρονται.

3) Εσωτερικές Επιθέσεις στην Υποδομή «Big Data».

Αυτή η κατηγορία είναι παρεμφερής με αυτή της εξωτερικής επίθεσης. Ωστόσο, συγκρινόμενη με την εξωτερική επίθεση, η εσωτερική επίθεση έχει πιο καταστροφικές συνέπειες για την υποδομή των «Big Data». Έτσι σε αυτή την κατηγορία ένας εσωτερικός επιτιθέμενος είναι δυνατό να τροποποιήσει ή/και να διαγράψει, τόσο τα αποθηκευμένα «Μεταδεδομένα Προέλευσης», όσο και το ιστορικό των επιθεωρήσεων, αποσκοπώντας στην ολική καταστροφή του συστήματος τήρησης αρχείων με δεδομένα προέλευσης των συστημάτων «Big Data».

Συνεπώς, προκειμένου να αντιμετωπιστούν οι ανωτέρω απειλές, θα πρέπει να διασφαλιστεί η αξιοπιστία και η ευχρηστία των «Μεταδεδομένων Προέλευσης» στα συστήματα «Big Data». Αυτό θα επιτευχθεί, αφενός με την ασφάλιση της διαδικασίας της συλλογής τους και αφετέρου με την χρήση ενός λεπτομερούς μηχανισμού ελέγχου πρόσβασης σε αυτά.

Αναλυτικότερα, για να διασφαλιστεί η διαδικασία της συλλογής των «Μεταδεδομένων Προέλευσης», θα πρέπει οι πηγές του συστήματος που τα παράγουν να έχουν πρώτα αυθεντικοποιηθεί στο σύστημα. Αυτό μπορεί να επιτευχθεί με τον συνδυασμό μιας ελαφριάς τεχνικής αυθεντικοποίησης και ενός λεπτομερούς μηχανισμού ελέγχου πρόσβασης [24]. Στο πλαίσιο αυτό έχει προταθεί ο μηχανισμός ελέγχου πρόσβασης «Provenance Based Access Control - PBAC», ο οποίος άμεσα διατηρεί και αξιοποιεί τις απαραίτητες πληροφορίες που προέρχονται από τα μεταδεδομένα, για την δυναμική κατάτμηση των δικαιωμάτων πρόσβασης στους χρήστες του συστήματος [25], [26]. Επιπλέον θα πρέπει να διεξάγονται περιοδικοί έλεγχοι προκειμένου να διαπιστώνεται η κατάσταση των διαφόρων οντοτήτων του συστήματος «Big Data» (Status Updates), έτσι ώστε να διασφαλιστεί ότι αυτές, είτε δεν έχουν καταληφθεί από κάποιο επιτιθέμενο, είτε δεν έχουν κάποια δυσλειτουργία.

Παράλληλα, για να διασφαλιστεί η ακεραιότητα (Integrity) και η ακρίβεια (Accuracy) των «Μεταδεδομένων Προέλευσης» που αποστέλλονται από τις διάφορες οντότητες του συστήματος, θα πρέπει αυτά να συνοδεύονται από έλεγχο ακεραιότητας (Integrity Check), έτσι ώστε να επιβεβαιώνεται ότι δεν έχουν τροποποιηθεί ή πλαστογραφηθεί. Επίσης, θα πρέπει να επαληθεύεται η συνέπεια (Consistency) μεταξύ των δεδομένων αυτών και της πηγής προέλευσής τους, καθώς σε περίπτωση ασυνέπειας θα εξάγονται λανθασμένα αποτελέσματα. Αυτό μπορεί να υλοποιηθεί με την τήρηση «Ιστορικού Μεταδεδομένων» ανά οντότητα του συστήματος με την βοήθεια των «Hash Functions» και των «Hash Tables» [247]. Μάλιστα, μια επιπλέον καλή πρακτική είναι η χρήση τεχνικών που θα μετρούν τον βαθμό εμπιστοσύνης και φήμης των διαφορετικών οντοτήτων του συστήματος «Big Data». Με αυτές τις τεχνικές, θα βαθμολογούνται όλες οι οντότητες του συστήματος ανάλογα με την συμπεριφορά τους και την αξιοπιστία τους. Όσες από αυτές βρίσκονται κάτω από ένα προκαθορισμένο όριο θα αποκλείονται από το σύστημα.

Επιπρόσθετα, καθώς μερικές φορές υπάρχει περίπτωση στα «Μεταδεδομένα Προέλευσης» να συμπεριλαμβάνονται και ευαίσθητες πληροφορίες των πηγών τους, απαιτείται η κρυπτογράφησή τους, πριν την μετάδοσή τους, για την προστασία της εμπιστευτικότητας (Confidentiality) και της «Ιδιωτικότητας» (Privacy) της πηγής τους [27]. Επίσης, θα πρέπει να υλοποιηθούν ασφαλή κανάλια επικοινωνιών μεταξύ των διαφόρων τμημάτων της υποδομής, προκειμένου να επιτευχθεί ασφάλεια απ' άκρη σε άκρη (End-To-End). Τέλος, κρίνεται απαραίτητο η διαδικασία συλλογής των «Μεταδεδομένων Προέλευσης» στα συστήματα «Big Data» να είναι ευέλικτη (Flexibility) και προσαρμόσιμη (Scalability), έτσι ώστε να μπορεί να ενσωματώνει κάθε φορά τις διαρκώς μεταβαλλόμενες σε όγκο, ποικιλία και ταχύτητα πληροφορίες. Η διασφάλιση των ιδιοτήτων αυτών θα προσδώσει ασφάλεια στην διαδικασία συλλογής των «Μεταδεδομένων Προέλευσης», καθώς αυτή θα είναι δυναμικά προσαρμόσιμη, είτε αν αυτά αυξηθούν λόγω της επέκτασης του συστήματος, είτε αν αυτά ελαττωθούν εξαιτίας της μη διαθεσιμότητας κάποιων τμημάτων του [51].

Όσον αφορά την αντιμετώπιση των απειλών που σχετίζονται με τις επιθέσεις εναντίον της υποδομής, είτε εξωτερικά, είτε εκ των έσω, απαιτείται η ύπαρξη ενός λεπτομερούς μηχανισμού ελέγχου πρόσβασης. Για τα συστήματα «Big Data», στα αρχεία προέλευσης συμπεριλαμβάνονται όχι μόνο τα «Μεταδεδομένα Προέλευσης» από τις διάφορες οντότητες του συστήματος, αλλά επιπλέον και τα «Μεταδεδομένα

Προέλευσης» του ίδιου του συστήματος «Big Data». Αυτό συνεπάγεται ότι το πλήθος των «Μεταδεδομένων» σε αυτά τα συστήματα αυξάνεται εκθετικά. Έτσι, προκειμένου αυτά τα μεγάλα σε όγκο, πολυπλοκότητα και μερικές φορές ευαισθησία «Μεταδεδομένα Προέλευσης» να προστατευθούν, είναι απαραίτητο να υπάρχει έλεγχος της πρόσβασης σε αυτά. Σε αντίθετη περίπτωση, δεν θα ήταν εφικτό να αντιμετωπιστούν οι επιθέσεις αυτές. Οπότε ένας λεπτομερής μηχανισμός ελέγχου πρόσβασης, ο οποίος θα εκχωρεί διαφορετικά δικαιώματα στους διαφορετικούς ρόλους των οντοτήτων του συστήματος που αιτούνται πρόσβαση στα αρχεία αυτά, κρίνεται αναγκαίος. Επίσης, θα πρέπει αυτός ο λεπτομερής μηχανισμός ελέγχου πρόσβασης να είναι δυναμικός, επεκτάσιμος και να υποστηρίζει ευέλικτους μηχανισμούς ανάκλησης (Revocation Mechanisms). Τέτοιες πιθανές λύσεις μπορεί να είναι οι μηχανισμοί ελέγχου πρόσβασης «Role Based Access Control - RBAC», «Attribute Based Access Control - ABAC» και «Ciphertext Policy Attribute Based Encryption – CP-ABE» [51]. Επιπρόσθετα, ένας μηχανισμός λεπτομερούς ελέγχου πρόσβασης, όπως ο «Revocable Attribute Based Encryption Access Control» [28], θα πρέπει να ενσωματωθεί στο υπάρχον σύστημα αποθήκευσης των αρχείων προέλευσης [29] για να επιτευχθεί η ασφάλεια, τόσο κατά την αποθήκευση των αρχείων αυτών, όσο και κατά την διαδικασία πρόσβασης σε αυτά.

Τέλος, καθώς τα συστήματα «Big Data» μερικές φορές λειτουργούν παράλληλα με τις υποδομές «Cloud», θα πρέπει στα ανωτέρω ζητήματα, ήτοι της ασφαλούς συγκέντρωσης των «Μεταδεδομένων Προέλευσης» και της υλοποίησης κατάλληλου μηχανισμού για τον έλεγχο της πρόσβασης σε αυτά, να λαμβάνονται υπόψη, εκτός από τις απαιτήσεις για τα συστήματα «Big Data» και οι αντίστοιχες για τις υποδομές «Cloud». Έτσι, θα πρέπει να υλοποιηθεί μια ασφαλής διαδικασία συγκέντρωσης «Μεταδεδομένων Προέλευσης», η οποία θα χρησιμοποιεί μια γρήγορη και μη απαιτητική τεχνική αυθεντικοποίησης, που θα καλύπτει ταυτόχρονα τις απαιτήσεις και των «Big Data» και του «Cloud» [172].

2.3.2 ΑΣΦΑΛΗΣ ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ ΤΗΡΗΣΗ ΙΣΤΟΡΙΚΟΥ ΕΝΕΡΓΕΙΩΝ (STORAGE AND TRANSACTION LOGS)

Γενικά, στα παραδοσιακά πληροφοριακά συστήματα χρησιμοποιούνται διάφορα αποθηκευτικά μέσα, στα οποία αποθηκεύονται τα αρχεία δεδομένων, καθώς και το ιστορικό τήρησης ενεργειών των συστημάτων αυτών. Επίσης σε αυτά, στο πλείστο

των περιπτώσεων, υλοποιούνται διαφορετικά επίπεδα και πολιτικές ασφαλείας, ανάλογα με το είδος και την σημασία των δεδομένων που αποθηκεύονται στο καθένα από αυτά. Έτσι, στην περίπτωση που απαιτείται να μετακινηθούν αρχεία μεταξύ αυτών των μέσων αποθήκευσης, στα οποία μπορεί να έχει υλοποιηθεί διαφορετικό επίπεδο ασφαλείας, τότε αυτή πραγματοποιείται χειροκίνητα από τον ίδιο τον διαχειριστή του συστήματος, καθώς έτσι έχει πλήρη επίγνωση για το τι, που και πότε το μετακίνησε.

Ωστόσο, στο ιδιαίτερο περιβάλλον των «Big Data», όπου το μέγεθος των δεδομένων αυξάνει εκθετικά και τα ιδιαίτερα χαρακτηριστικά τους δημιουργούν νέες απαιτήσεις για ευελιξία, κλιμάκωση και διαθεσιμότητα των συστημάτων αυτών, επιβάλλεται η διαδικασία του χειρισμού και της αποθήκευσης των δεδομένων σε μέσα με διαφορετικά επίπεδα ασφαλείας να είναι αυτοματοποιημένη (Auto-Tiering) [248], καθώς η εκτέλεσή της χειροκίνητα από τον διαχειριστή του συστήματος καθίσταται αδύνατη. Αυτές όμως οι αυτοματοποιημένες λύσεις δεν τηρούν αρχείο καταγραφής για το τι και που είναι αποθηκευμένο, γεγονός που θέτει επιπλέον ζητήματα, τόσο για την ασφαλή αποθήκευσή τους, όσο και για τον βαθμό ελέγχου επ' αυτών. Παράλληλα τίθενται και ζητήματα για την προστασία της «Ιδιωτικότητας», καθώς ο αυτόματος διαμοιρασμός και ανταλλαγή των δεδομένων αυτών στα διάφορα επιμέρους πληροφοριακά συστήματα του περιβάλλοντος των «Big Data», στα οποία ενδεχομένως να έχουν υλοποιηθεί διαφορετικά επίπεδα και πολιτικές ασφαλείας, μπορεί να οδηγήσει στην υποβάθμιση του επιπέδου προστασίας των προσωπικών δεδομένων, ακόμα και αν αυτό ήταν υψηλό κατά την αρχική συλλογή τους για επεξεργασία [3]. Επομένως είναι επιτακτικό να βρεθούν νέοι μηχανισμοί ή να βελτιωθούν οι υπάρχοντες, έτσι ώστε, αφενός να βελτιστοποιείται η αυτοματοποιημένη διαδικασία για την αποθήκευση των δεδομένων, λαμβάνοντας υπόψη κάθε φορά τον βαθμό προστασίας των προσωπικών δεδομένων και την εκάστοτε πολιτική ασφαλείας, και αφετέρου να αντιμετωπίζεται αποτελεσματικά η μη εξουσιοδοτημένη πρόσβαση σε αυτά, διασφαλίζοντας παράλληλα την συνεχή διαθεσιμότητά τους. Επιπλέον, η τήρηση του ιστορικού όλων των ενεργειών, που έλαβαν χώρα κατά την εκτέλεση της διαδικασίας αυτής, θα συμβάλλει αποτελεσματικά, τόσο στην ανίχνευση των περιστατικών ασφαλείας, όσο και στην απόδοση τυχών ευθυνών [1].

Έτσι για παράδειγμα, αν ένας επιχειρηματίας θέλει να διαχειριστεί συνολικά τα δεδομένα από τα διαφορετικά τμήματα της επιχείρησής του, τότε η χρησιμοποίηση του αυτόματου συστήματος ιεράρχησης και αποθήκευσης των δεδομένων (Auto-Tier Storage) θα μειώσει το κόστος της ενέργειας αυτής. Επίσης, είναι προφανές ότι κάποια από αυτά ανακτώνται σπάνια, ενώ κάποια άλλα χρησιμοποιούνται πολύ συχνά από κάποια τμήματα της επιχείρησης. Οπότε είναι πολύ πιθανό για λόγους κόστους και κέρδους, τα σπανίως χρησιμοποιούμενα δεδομένα να αποθηκευτούν σε μέσο με χαμηλότερο και φθηνότερο επίπεδο ασφαλείας. Ωστόσο, σε αυτή την κατηγορία μπορεί να συμπεριλαμβάνονται και δεδομένα με κρίσιμες πληροφορίες της επιχείρησης τα οποία δεν ανακτώνται συχνά. Αυτό σημαίνει ότι η επιχείρηση θα πρέπει να μελετήσει προσεκτικά τις στρατηγικές για την αυτόματη ιεράρχηση και αποθήκευση των δεδομένων της, πριν την υλοποίηση της ενέργειας αυτής [52]. Επιπρόσθετα και τα «Μεταδεδομένα» (Metadata) αποτελούν κρίσιμες πληροφορίες που χρήζουν προστασίας, καθώς χρησιμοποιούνται σε διάφορες σημαντικές διαδικασίες του συστήματος, όπως αυτή της ερμηνείας των εξαγόμενων αποτελεσμάτων της «Αναλυτικής» και στην εκτέλεση επιθεωρήσεων. Όμοια, τα αρχεία τήρησης του ιστορικού κίνησης (Log Files) θα πρέπει να προστατευτούν από επιθέσεις αλλοίωσής τους (Log Poisoning), για να μην προκληθεί ασυνέπεια στα δεδομένα τους και περαιτέρω αδυναμία στην διερεύνηση των περιστατικών ασφαλείας και την απόδοση ευθυνών.

Οπότε, αν και τα «Κατανεμημένα σε Δίκτυο Συστήματα Αυτόματης Αποθήκευσης» των δεδομένων (Network Based Distributed Auto-Tier Storage Systems) σε μέσα αποθήκευσης με διαφορετικά επίπεδα ασφαλείας αποτελούν μια πολλά υποσχόμενη λύση στο περιβάλλον των «Big Data», καθώς παρέχουν διαφανείς υπηρεσίες με δυνατότητα κλιμάκωσης και ελαστικότητας, παράλληλα εισάγουν και νέα ζητήματα ασφαλείας που σχετίζονται με τις έννοιες της φυσικής κατοχής των δεδομένων, της μη αξιόπιστης υπηρεσίας αποθήκευσης και της ύπαρξης ασυνεπών πολιτικών ασφαλείας μεταξύ των διαφόρων μέσων αποθήκευσης των δεδομένων. Συγκεκριμένα, το μοντέλο απειλών για το σύστημα αυτό [1] περιλαμβάνει τα παρακάτω θέματα:

1) Εμπιστευτικότητα και Ακεραιότητα (Confidentiality and Integrity).

Παράλληλα με αυτούς που προσπαθούν να υποκλέψουν, ή να καταστρέψουν τα δεδομένα των χρηστών ενός πληροφοριακού συστήματος, και οι υπηρεσίες αποθήκευσης των δεδομένων μπορούν να θεωρηθούν ως μη έμπιστες τρίτες

οντότητες. Αυτό οφείλεται, όσον αφορά την εμπιστευτικότητα, στο γεγονός ότι για την αποθήκευση των δεδομένων στα διάφορα μέσα αποθήκευσης με διαφορετικά επίπεδα ασφαλείας, απαιτείται να παρέχονται στους διαχειριστές της υπηρεσίας στοιχεία που τους δίνουν την δυνατότητα να συσχετίσουν τις ενέργειες των χρηστών με κάποια σύνολα δεδομένων. Ακόμα και στην ειδική περίπτωση που χρησιμοποιούνται κρυπτογραφημένα δεδομένα, υπάρχει πιθανότητα να αποκαλυφθούν κάποιες ιδιότητες των ανωτέρω συσχετίσεων από τις αναγκαίες πληροφορίες που πρέπει να παρασχεθούν για την εκτέλεση της υπηρεσίας αποθήκευσης. Όσον αφορά την ακεραιότητα, αυτή σχετίζεται με το κατά πόσο εκτελέστηκε επιτυχώς η υπηρεσία αυτή, έτσι ώστε τα δεδομένα του συστήματος να είναι ενημερωμένα και ακριβή.

2) Προέλευση (Provenance).

Εξαιτίας του εξαιρετικά μεγάλου όγκου δεδομένων στο περιβάλλον των «Big Data», είναι πρακτικά αδύνατο να μεταφορτωθεί ολόκληρο το σύνολο των δεδομένων για να πιστοποιηθεί η διαθεσιμότητα και η ακεραιότητά του. Για αυτό, χρησιμοποιούνται εναλλακτικά πιο αποδοτικές τεχνικές, οι οποίες έχουν σχεδιαστεί να παρέχουν πιστοποίηση των ανωτέρω στοιχείων, με την χρήση προσεγγίσεων που στηρίζονται στις πιθανότητες και που συνεπάγονται μικρές επιβαρύνσεις στην υπολογιστική ισχύ και τις επικοινωνίες. Ωστόσο, επειδή αυτές στηρίζονται στην τυχαία δειγματοληψία επί του συνόλου των δεδομένων, τα ζητήματα που εγείρονται είναι σε ποιο βαθμό τα αποτελέσματα των τεχνικών αυτών ανταποκρίνονται στην πραγματικότητα και κατά πόσο αυτά προέρχονται από το αυθεντικό σύνολο των δεδομένων του συστήματος.

3) Διαθεσιμότητα (Availability).

Επίσης, το γεγονός, ότι οι μηχανισμοί αυτοί αποθηκεύουν τα δεδομένα σε μέσα αποθήκευσης με διαφορετικά επίπεδα ασφαλείας, δημιουργεί ζητήματα για την εγγύηση της διαθεσιμότητάς τους. Αυτό οφείλεται, αφενός στο ότι τα μέσα αποθήκευσης με τα χαμηλότερα επίπεδα ασφαλείας είναι ευάλωτα σε επιθέσεις, οι οποίες σκοπεύουν να τα καταστήσουν μη διαθέσιμα (Denial of Service – DoS). Αφετέρου, η ύπαρξη διαφορετικών χρόνων ανάκαμψης από καταστροφή για το κάθε μέσο αποθήκευσης, λόγω της διαφοράς που υπάρχει στα επίπεδα ασφαλείας μεταξύ τους, τα καθιστά μη διαθέσιμα στο σύστημα για διαφορετικούς χρόνους το καθένα σε περίπτωση που θα συμβεί κάποιο περιστατικό ασφαλείας σε αυτά, με άμεση συνέπεια να υπάρχουν σημαντικές διαφορές μεταξύ των αποδόσεων τους.

4) Συνέπεια (Consistency).

Ένα άλλο πολύ σημαντικό ζήτημα είναι η διατήρηση της συνέπειας μεταξύ των δεδομένων που αποθηκεύονται σε περισσότερα του ενός αντιγράφων στις διαφορετικές τοποθεσίες του συστήματος «Big Data». Για αυτό το λόγο, η τήρηση προτεραιότητας κατά την εγγραφή των δεδομένων στο σύστημα (Write Serializability) και η εφαρμογή μηχανισμών για την πολλαπλή ανάγνωση και εγγραφή (Multi Writer Multi Reader - MWMR) σε αυτά θα πρέπει να εξεταστεί επισταμένα.

5) Επιθέσεις Συνομωσίας (Collusion Attacks).

Συνήθως, όταν ο ιδιοκτήτης των δεδομένων αποθηκεύει τα δεδομένα του κρυπτογραφημένα σε ένα σύστημα αυτόματης αποθήκευσης, παράλληλα θα πρέπει να διανείμει τα κλειδιά κρυπτογράφησης στους υπόλοιπους χρήστες και να καθορίσει τα δικαιώματα πρόσβασής τους. Έτσι, ο κάθε χρήστης είναι εξουσιοδοτημένος να έχει πρόσβαση σε κάποιο συγκεκριμένο ποσοστό του συνόλου των δεδομένων αυτών, ενώ ο πάροχος της υπηρεσίας δεν είναι σε θέση να αποκρυπτογραφήσει τα δεδομένα αυτά, καθώς δεν γνωρίζει το κλειδί της κρυπτογράφησης. Ωστόσο, υπάρχει σημαντική πιθανότητα ο πάροχος της υπηρεσίας να συνωμοτήσει με κάποιους χρήστες, προκειμένου να μάθει το κλειδί κρυπτογράφησης και να αποκτήσει πρόσβαση σε ένα υποσύνολο των δεδομένων αυτών, για το οποίο δεν είχε αρχικά δικαίωμα πρόσβασης.

6) Επιθέσεις Υποβάθμισης της Ποιότητας των Αποθηκευμένων Δεδομένων (Roll Back Attacks).

Ο πάροχος μιας υπηρεσίας η οποία αριθμεί πολλούς χρήστες, δύναται να εκτοξεύσει επιθέσεις υποβάθμισης της ποιότητας των αποθηκευμένων δεδομένων τους σε αυτή (Roll Back Attack). Αυτή η επίθεση συνίσταται στην εξαπάτηση κάποιου «Υποκειμένου» δεδομένων, επιστρέφοντάς του σε ενδεχόμενη αίτησή του, όχι την ενημερωμένη έκδοση του συνόλου των δεδομένων του, την οποία προηγουμένως το ίδιο το «Υποκείμενο» είχε μεταφορτώσει και ενημερώσει, αλλά την προγενέστερη έκδοση και συνεπώς μη ενημερωμένη. Έτσι η ύπαρξη αποδείξεων, οι οποίες θα διαβεβαιώνουν τους χρήστες ότι τα δεδομένα τους είναι ορθά ενημερωμένα και παράλληλα θα τους προσδίδουν την δυνατότητα να διαπιστώσουν πιθανή ασυνέπεια, είναι επιτακτική. Αυτό εναλλακτικά αναφέρεται και ως «User's Freshness».

7) Διαμάχη (Disputes).

Πλέον των ανωτέρω, η έλλειψη τήρησης ιστορικού ενεργειών (Log) μπορεί να οδηγήσει σε διαμάχες, είτε μεταξύ των χρηστών, είτε μεταξύ των χρηστών και του παρόχου της υπηρεσίας αποθήκευσης. Έτσι σε περίπτωση απώλειας ή αλλοίωσης

των δεδομένων, τα αρχεία καταγραφής κίνησης (Transmission Logs/Records) θα αποβούν εξαιρετικά χρήσιμα για την διερεύνηση και εύρεση των υπευθύνων. Οπότε, για παράδειγμα, αν ένας κακόβουλος χρήστης, ο οποίος έχει αρχικά αποθηκεύσει κάποια δεδομένα, στη συνέχεια δηλώσει ψευδώς την απώλεια των δεδομένων αυτών και ζητά αποζημίωση για αυτή του την απώλεια, τότε ένα καλό σύστημα καταγραφής των ενεργειών που λαμβάνουν χώρα στο σύστημα θα μπορεί να προλάβει και να αντιμετωπίσει επιτυχώς τις περιπτώσεις αυτές εξαπάτησης.

Συνεπώς, η αντιμετώπιση των ανωτέρω απειλών καθίσταται κρίσιμη για την ασφάλεια στα συστήματα «Big Data». Γεγονός είναι ότι τις τελευταίες δεκαετίες έχουν συντελεστεί ραγδαίες εξελίξεις στους τομείς της διασφάλισης της ποιότητας των πληροφοριών και της ασφάλειας των διαδικτυακών πληροφοριακών υποδομών. Μάλιστα στο [51] αναφέρονται κάποιες εκλεπτυσμένες τεχνικές, για την αντιμετώπιση των παραπάνω ζητημάτων ασφαλείας.

Αναλυτικότερα, όσον αφορά τα ζητήματα της εμπιστευτικότητας και της ακεραιότητας, αυτά μπορούν να αντιμετωπιστούν με την χρήση ισχυρών κρυπτογραφικών τεχνικών και την χρήση της ψηφιακής σύνοψης (Message Digest). Ενώ η ανταλλαγή μηνυμάτων με την χρήση ψηφιακών υπογραφών (Signed Message Digests) μπορεί να χρησιμοποιηθεί για να αντιμετωπιστούν οι περιπτώσεις διαμάχης [30]. Παράλληλα, η διαβεβαίωση των χρηστών για την ενημερότητα των αρχείων τους (User's Freshness) μπορεί να συντελεστεί μέσω περιοδικών ελέγχων του συστήματος (Periodic Audits) [31]. Επίσης το πρόβλημα της τήρησης σειράς προτεραιότητας κατά την εγγραφή των αρχείων (Write Serializability) μπορεί να αντιμετωπιστεί με την χρήση των τεχνικών «Chain Hash» [32] ή «Persistent Authenticated Dictionary - PAD» [33]. Αλλά και η τεχνική «Secure Untrusted Data Repository - SUNDR» μπορεί να συμβάλλει στην επίλυση του προβλήματος αυτού, ενώ ταυτόχρονα μπορεί να χρησιμοποιηθεί για την ανίχνευση επιθέσεων που στοχεύουν στην δημιουργία ασυνέπειας στο σύστημα. Από την άλλη πλευρά, για την επίλυση του προβλήματος «Single-Write Multi-Read - SWMR» έχουν προταθεί δυο γραμμικά και παράλληλα «Lock Free» πρωτόκολλα [34]. Επιπρόσθετα, οι τεχνικές «Broadcast Encryption» [35] και «Key Rotation» [36] μπορούν να βελτιώσουν την δυνατότητα για κλιμάκωση του συστήματος. Επιπλέον, διάφορες λύσεις έχουν προταθεί από τους ερευνητές, για να αντιμετωπιστούν τα ζητήματα που σχετίζονται με την προέλευση των δεδομένων [37]. Συγκεκριμένα, η διαβεβαίωση για την

διαθεσιμότητα των αποθηκευμένων δεδομένων μπορεί να εκτελεστεί αποδοτικά και με καλή πιθανότητα με την χρήση των τεχνικών της «Απόδειξης της Δυνατότητας Ανάκτησης» (Proof of Retrievability - POR) και της «Απόδειξης Κατοχής» τους (Provable Data Possession - PDP) [38], [39].

Όσον αφορά τις επιθέσεις συνωμοσίας (Collusion Attacks), διακρίνουμε τις εξής περιπτώσεις. Αν οι χρήστες δεν ανταλλάσουν μεταξύ τους τα ιδιωτικά τους κλειδιά, τότε ένα σύστημα κρυπτογράφησης, που θα βασίζεται στην πολιτική ασφαλείας (Policy Based Encryption System - PBES), θα μπορέσει να εγγυηθεί ένα περιβάλλον απαλλαγμένο από τέτοιες επιθέσεις [40]. Αν οι χρήστες αποφασίσουν να ανταλλάξουν μεταξύ τους τα ιδιωτικά τους κλειδιά, χωρίς όμως να στέλνουν ο ένας στον άλλο και το αποκρυπτογραφημένο περιεχόμενο, τότε ένα σύστημα αποκρυπτογράφησης με διαμεσολάβηση (Mediated Decryption System) μπορεί να αποτρέψει αυτή την περίπτωση. Τέλος, αν οι χρήστες ανταλλάσουν μεταξύ τους, εκτός από τα ιδιωτικά τους κλειδιά και το αποκρυπτογραφημένο περιεχόμενο, τότε ένα σύστημα διαχείρισης των δεδομένων, σύμφωνα με τα ψηφιακά δικαιώματα του κάθε χρήστη (Digital Rights Management) [249], θα αντιμετωπίσει αποτελεσματικά αυτή την απειλή. Επιπλέον, προκειμένου να αντιμετωπιστούν τα ζητήματα διαμάχης που θα προκύψουν από τις ανωτέρω περιπτώσεις, έχουν προταθεί δυο πρωτόκολλα που θα εστιάζουν στην «Μη Αποποίηση» των ευθυνών των χρηστών (Non Repudiation Protocols) [41], [42].

Τέλος, οι «Αυτοματοποιημένοι Μηχανισμοί Επιβολής των Πολιτικών Ασφαλείας» στα δεδομένα του συστήματος (Automated Security Policy Enforcement Mechanisms) μπορεί κάλλιστα να συμβάλλουν την αντιμετώπιση των ανωτέρω απειλών, καθώς διεξάγεται έρευνα σύμφωνα με την οποία θα κρυπτογραφούνται τα δεδομένα του συστήματος με βάση την εκάστοτε πολιτική ασφαλείας του [47]. Συγκεκριμένα, τα δεδομένα κρυπτογραφούνται και αποκρυπτογραφούνται σε δυο φάσεις κάθε φορά. Η εξωτερική διαδικασία εκτελείται μόνο από την έμπιστη συσκευή ασφαλούς αποθήκευσης των δεδομένων (Tamper-Resistant Hardware Storage), που έχει στην κατοχή του ο νόμιμος χρήστης του συστήματος. Ενώ η εσωτερική διαδικασία εκτελείται μόνο από συγκεκριμένο λογισμικό και αφού προηγουμένως έχει ελεγχθεί ότι ικανοποιούνται οι απαιτήσεις ασφαλείας που επιβάλλει η συγκεκριμένη πολιτική ασφάλειας. Ωστόσο, τα βασικά προβλήματα που προκύπτουν είναι, αφενός ότι όλα τα συμβαλλόμενα μέρη θα πρέπει να θεωρούνται και να είναι έμπιστα,

γεγονός ανέφικτο στο περιβάλλον των «Big Data», και αφετέρου η παραμετροποίηση αυτών των μηχανισμών με γνώμονα για παράδειγμα την πολιτική ασφαλείας, την επιθυμία του «Υποκειμένου» των δεδομένων, τον τόπο αποθήκευσής τους κλπ κάθε φορά, καθίσταται μη αποδοτική σε αυτό το σύνθετο περιβάλλον. Επίσης, προκειμένου να είναι εφικτή η χρησιμοποίηση των εργαλείων αυτών, απαιτείται οι πολιτικές ασφαλείας να είναι αναγνώσιμες, τόσο από τις συσκευές, όσο και από τους ίδιους τους χρήστες, το οποίο προϋποθέτει την ύπαρξη αντίστοιχης γλώσσας.

Συνοψίζοντας παρατηρείται ότι, αν και γενικά οι ανωτέρω λύσεις εξασφαλίζουν την ικανοποίηση των γενικών απαιτήσεων ασφαλείας στο περιβάλλον των «Big Data», ήτοι την εμπιστευτικότητα, την ακεραιότητα και την διαθεσιμότητα των δεδομένων, υπάρχουν περαιτέρω θέματα που χρήζουν ιδιαίτερης προσοχής στο σύνθετο αυτό περιβάλλον. Συγκεκριμένα τα θέματα αυτά είναι:

- **Τα Δυναμικά Σύνολα Δεδομένων και η Διαθεσιμότητά τους.**

Τα σύνολα δεδομένων σε ένα αυτοματοποιημένο σύστημα αποθήκευσης είναι δυναμικά και μεταβάλλονται διαρκώς, καθώς ενέργειες όπως η τροποποίηση, η αντιγραφή, η διαγραφή και η εισαγωγή δεδομένων σε ένα σύστημα «Big Data», λαμβάνουν χώρα σε υψηλούς ρυθμούς. Εξαιτίας αυτού του γεγονότος, θα πρέπει να διασφαλίζεται κάθε φορά η διαθεσιμότητά τους. Μια εξελιγμένη δυναμική έκδοση της τεχνικής «Provable Data Possession - PDP» επιτυγχάνει υψηλή απόδοση προς την κατεύθυνση αυτή, καθώς στηρίζεται στην συμμετρική κρυπτογράφηση. Ωστόσο ο βασικός περιορισμός της είναι ότι μπορεί να υποστηρίξει περιορισμένο αριθμό επερωτήσεων και συνεπώς δεν μπορεί να υποστηρίξει πλήρως το ανωτέρω μεταβαλλόμενο περιβάλλον [43]. Παρά ταύτα, θεωρείται ότι βελτιώνει σημαντικά την πιθανότητα εντοπισμού των δεδομένων σε βάρος όμως της υπολογιστικής ισχύος του εξυπηρετητή. Από την άλλη μεριά έχει προταθεί μια βελτιωμένη έκδοση της τεχνικής «Proof of Retrievability - POR», η οποία αφενός υποστηρίζει την δημόσια επαλήθευση της διαθεσιμότητας των δεδομένων και αφετέρου ανταποκρίνεται αποτελεσματικά στο δυναμικό περιβάλλον των συστημάτων αυτών, καθιστώντας την μια καλή λύση για τα αυτοματοποιημένα συστήματα αποθήκευσης που υλοποιούνται σε περιβάλλον δικτύου (Network Based Auto Tier Storage Systems) [44].

- **Η Προστασία της «Ιδιωτικότητας».**

Γενικά είναι συνήθης πρακτική να ανατίθενται οι διαδικασίες για την επαλήθευση της διαθεσιμότητας των δεδομένων σε έμπιστους εξωτερικούς ελεγκτές (Third Party Auditors) και να δημοσιεύεται το πρωτόκολλο επαλήθευσης για να μπορεί να ελεγχτεί

η αξιοπιστία του δημόσια. Έτσι, εκ πρώτης όψεως φαίνεται οξύμωρο πως είναι δυνατό να προστατεύεται η «Ιδιωτικότητα» των χρηστών, ενώ παράλληλα καθίσταται δημόσια η δυνατότητα πιστοποίησης της διαθεσιμότητας των δεδομένων τους. Για αυτό το λόγο, ο [45] προτείνει ένα δημόσιο σχήμα, που θα μπορεί να ελέγχει και να πιστοποιεί την διαθεσιμότητα των δεδομένων, προστατεύοντας παράλληλα την «Ιδιωτικότητα» των χρηστών. Σε αυτό το σχήμα χρησιμοποιείται η τεχνική της «Γραμμικής Ομοιομορφικής Αυθεντικοποίησης» (Homomorphic Linear Authenticator) συνδυασμένη με την τεχνική της «Τυχαίας Απόκρυψης» των δεδομένων (Random Masking).

- **Ο Ασφαλής Χειρισμός των Κρυπτογραφημένων Δεδομένων.**

Η σύγχρονη τάση απαιτεί την εκτέλεση διαφόρων πράξεων πάνω στα κρυπτογραφημένα σύνολα δεδομένων, χωρίς προηγουμένως να απαιτείται η αποκρυπτογράφησή τους. Μάλιστα επιτάσσει οι υπολογισμοί να εκτελούνται από τρίτα μέρη που έχουν καλύτερη υπολογιστική ισχύ για λογαριασμό των χρηστών, χωρίς ταυτόχρονα να παραβιάζεται η «Ιδιωτικότητά» τους. Η λύση που υπόσχεται την επιτυχή εκτέλεση της διαδικασίας αυτής είναι η τεχνική της «Πλήρους Ομοιομορφικής Κρυπτογράφησης» (Fully Homomorphic Encryption), η οποία υποστηρίζει τις πιο πολλές πολύπλοκες πράξεις επί των κρυπτογραφημάτων [46].

Συμπερασματικά, διαπιστώνεται ότι, αν και φαίνεται να υπάρχει λύση για έκαστη απειλή στα «Αυτοματοποιημένα Συστήματα Αποθήκευσης» (Auto-Tier Storage Systems) σε μέσα με διαφορετικά επίπεδα ασφαλείας, εντούτοις δεν υπάρχει κάποια συστηματική και ολιστική προσέγγιση προκειμένου αυτά να αντιμετωπιστούν μαζικά και να προταθεί μια ενοποιημένη λύση. Επίσης, η ύπαρξη μη ομοιόμορφων πολιτικών ασφαλείας μεταξύ των διαφόρων μέσων αποθήκευσης έχει ως αποτέλεσμα αυτά να έχουν διαφορετικά επίπεδα ασφαλείας, γεγονός που δημιουργεί επιπλέον προκλήσεις. Οπότε είναι εμφανές ότι απαιτείται περαιτέρω έρευνα, προκειμένου να μετριαστούν τα διλήμματα που σχετίζονται με την ασφάλεια του συστήματος σε συνάρτηση με την πολυπλοκότητα, το κόστος και την αποδοτική λειτουργία του.

2.3.3 ΑΠΟΤΕΛΕΣΜΑΤΙΚΕΣ ΕΠΙΘΕΩΡΗΣΕΙΣ (AUDITS)

Αν και ο πρωταρχικός στόχος του ελέγχου της ασφάλειας της υποδομής σε πραγματικό χρόνο (Real Time Monitoring) είναι η άμεση ενημέρωση κατά τον χρόνο

που μια επίθεση πραγματοποιείται, στην πραγματικότητα αυτό είναι δύσκολο να επιτευχθεί, είτε λόγω του ότι οι επιθέσεις είναι «Zero-Day» [250] και επομένως δεν είναι γνωστές, είτε διότι εσφαλμένα αυτές παραβλέφθηκαν ή/και αγνοήθηκαν, ενώ ορθά οι ενδείξεις μαρτυρούσαν ότι μια επίθεση λαμβάνει χώρα (True Positives). Συνεπώς, προκειμένου να εντοπιστεί αν το πληροφοριακό μας σύστημα έχει δεχτεί επίθεση, είναι απαραίτητο να διεξάγονται λεπτομερείς και αποτελεσματικές επιθεωρήσεις. Αυτές είναι σημαντικές για δυο λόγους. Αφενός γιατί συμβάλουν στην εύρεση στοιχείων που σχετίζονται με απαντήσεις σε ερωτήματα όπως το τι έγινε, από πότε ξεκίνησε και τι πήγε λάθος. Αφετέρου αυτές επιβάλλονται ως απαίτηση για συμμόρφωση με τις νομικές υποχρεώσεις, τα διάφορα πρότυπα, αλλά και τις διαδικασίες των εγκληματολογικών ερευνών. Οι επιθεωρήσεις δεν είναι κάτι καινούριο. Ωστόσο το εύρος τους, δηλαδή σε τι έκταση θα γίνουν, και η λεπτομέρειά τους, δηλαδή σε τι βάθος θα φτάσουν, διαφοροποιούνται κάθε φορά ανάλογα με τον επιδιωκόμενο σκοπό και την δομή και την φύση του εκάστοτε πληροφοριακού συστήματος που είναι προς επιθεώρηση.

Έτσι για παράδειγμα οι απαιτήσεις συμμόρφωσης είτε σε νόμους, είτε σε συμβόλαια, είτε σε πρότυπα, απαιτούν από τους οικονομικούς οργανισμούς να παρέχουν λεπτομερές ιστορικό επιθεωρήσεων. Από την άλλη πλευρά, το προσωπικό των οικονομικών οργανισμών έχει πρόσβαση σε μεγάλα σύνολα δεδομένων, στα οποία ενδεχομένως να συμπεριλαμβάνονται, είτε πληροφορίες που ταυτοποιούν τα δεδομένα με τα υποκείμενά τους, όπως είναι για παράδειγμα ο «ΑΜΚΑ» ή το «ΑΦΜ», είτε τα προσωπικά τους δεδομένα και πιθανόν τα ευαίσθητα προσωπικά τους δεδομένα. Με δεδομένο λοιπόν ότι, σε περίπτωση διαρροής προσωπικών δεδομένων, το κόστος της απώλειας των αρχείων δεδομένων που περιέχουν ευαίσθητα προσωπικά δεδομένα εκτιμάται στο ποσό των 200€ ανά εγγραφή, ενώ παράλληλα θα υπάρξουν και νομικές επιπτώσεις ανάλογα με την κείμενη νομοθεσία της εκάστοτε γεωγραφικής περιοχής, καθίσταται αναγκαία η εκτέλεση λεπτομερούς επιθεώρησης, προκειμένου να εξεταστούν οι συνθήκες υπό τις οποίες συντελέστηκε η εν λόγω διαρροή και να εντοπιστούν οι υπεύθυνοι για να αποδοθούν ευθύνες [52].

Οι βασικές προϋποθέσεις όμως για την επιτυχία των επιθεωρήσεων [1] είναι οι παρακάτω:

1) Πληρότητα των Απαιτούμενων για την Επιθεώρηση Πληροφοριών.

Θα πρέπει ο επιθεωρητής να έχει στην διάθεσή του όλες τις απαιτούμενες πληροφορίες για την διεξαγωγή της επιθεώρησης. Σε αυτές συμπεριλαμβάνονται όλα τα αρχεία καταγραφής κίνησης και ενεργειών (Log Files) των διαφόρων συσκευών και εφαρμογών του συστήματος.

2) Έγκαιρη Πρόσβαση στις Πληροφορίες αυτές.

Αυτή η απαίτηση είναι ιδιαίτερα σημαντική, αφενός για να προληφθεί η ενδεχόμενη αλλοίωση ή απώλεια των αρχείων αυτών και αφετέρου γιατί ο χρόνος είναι πολύτιμος, τόσο σε περιπτώσεις που η επιθεώρηση συνοδεύεται και με εγκληματολογική έρευνα για τον εντοπισμό πειστηρίων, όσο και για την διαδικασία της ανάκαμψης από καταστροφή του πληροφοριακού συστήματος.

3) Ακεραιότητα των Πληροφοριών αυτών.

Σύμφωνα με αυτή την απαίτηση θα πρέπει οι πληροφορίες αυτές να μην υποστούν καμία τροποποίηση ή να απολεσθούν.

4) Εξουσιοδοτημένη Πρόσβαση στις Πληροφορίες αυτές.

Η απαίτηση αυτή διασφαλίζει ότι μόνο οι εξουσιοδοτημένοι επιθεωρητές θα έχουν πρόσβαση στις πληροφορίες αυτές, ενώ από το προσωπικό μόνο όσοι κρίνεται απαραίτητο αποκλειστικά για την εκτέλεση συγκεκριμένης εργασίας.

Επομένως, οι απειλές που στοχεύουν στην υπονόμευση των ανωτέρω προϋποθέσεων, όπως είναι για παράδειγμα η μη εξουσιοδοτημένη πρόσβαση στις απαιτούμενες πληροφορίες της επιθεώρησης, η εσκεμμένη ή μη τροποποίησή τους ή και αφαίρεσή τους, η εσκεμμένη απόκρυψή τους και η αδικαιολόγητη καθυστέρηση στην διάθεσή τους στους επιθεωρητές, θα θέσουν σε κίνδυνο, τόσο τις ίδιες τις πληροφορίες που απαιτούνται για την διεξαγωγή της επιθεώρησης, όσο και την ίδια την διαδικασία της.

Συνεπώς, η επιτυχία της επιθεώρησης εξαρτάται σε μεγάλο βαθμό από τις λύσεις που θα υλοποιηθούν στο σύστημα «Big Data» προκειμένου να διασφαλιστούν οι ανωτέρω προϋποθέσεις. Αυτές αναφέρονται στο [51] και παρουσιάζονται αναλυτικότερα στις επόμενες παραγράφους.

Κατ’ αρχάς, καθώς ο απώτερος σκοπός της επιθεώρησης είναι, αφενός να ανιχνευθεί αν έλαβε χώρα κάποια επίθεση στο πληροφοριακό σύστημα και αφετέρου, αν πράγματι το πληροφοριακό σύστημα έχει δεχτεί επίθεση, να δημιουργηθεί μια σφαιρική εικόνα για αυτή και να δοθούν απαντήσεις σε κρίσιμα ερωτήματα που θα

προκύψουν, απαραίτητη προϋπόθεση για την επιτυχή επίτευξή της είναι η χρήση όλων των πληροφοριών που προέρχονται από τα διάφορα τμήματα που συνθέτουν συνολικά το πληροφοριακό αυτό σύστημα. Βέβαια, η συγκέντρωση όλων αυτών των πληροφοριών εξαρτάται σε μεγάλο βαθμό από το κατά πόσο είναι δυνατό το εκάστοτε μεμονωμένο δομικό τμήμα του συστήματος να παράσχει αυτές τις απαραίτητες πληροφορίες. Επομένως, η βασική απαίτηση για την επίτευξη του σκοπού της επιθεώρησης είναι να υποστηρίζεται η δυνατότητα διεξαγωγής επιθεωρήσεων σε όλη την έκταση της υποδομής «Big Data». Αυτό επιτυγχάνεται αν όλα τα επιμέρους δομικά τμήματα της υποδομής αυτής είναι σε θέση να συγκεντρώσουν και να παράσχουν όλες εκείνες τις απαραίτητες πληροφορίες που χρειάζονται για την υποστήριξη της επιθεώρησης. Έτσι, θα πρέπει κατά την φάση του σχεδιασμού του πληροφοριακού συστήματος να καθορίζονται λεπτομερώς, τόσο η πλήρης διαδικασία που θα ακολουθηθεί για την εκτέλεση της επιθεώρησης, όσο και το είδος και η προέλευση όλων των αναγκαίων πληροφοριών που θα πρέπει να συγκεντρωθούν για την πλήρη υποστήριξή της. Οπότε, αυτό θα έχει ως επακόλουθο κατά την φάση της ανάπτυξης να υλοποιηθούν όλοι εκείνοι οι αντίστοιχοι μηχανισμοί για την υποστήριξη της διαδικασίας της επιθεώρησης. Παραδείγματα τέτοιων μηχανισμών αποτελούν η ύπαρξη αρχείων καταγραφής ενεργειών και κίνησης σε δρομολογητές (Syslog On Routers), ή σε εφαρμογές (Application Logging), ή σε βάσεις δεδομένων (Database Logging), ή ακόμα και σε επίπεδο λειτουργικού συστήματος (Operating System Logging).

Στη συνέχεια, τα δεδομένα που συλλέχθηκαν θα πρέπει να αναλυθούν, προκειμένου να απαντηθούν τα κρίσιμα ερωτήματα που προέκυψαν και να εξαχθούν πολύτιμα συμπεράσματα. Η ανάλυση μπορεί να πραγματοποιηθεί, είτε από τα εργαλεία που χρησιμοποιούνται στις ψηφιακές εγκληματολογικές έρευνες (Digital Forensics Investigation Tools), είτε από το εργαλείο «Security Information and Event Management - SIEM» [251]. Βέβαια, θα πρέπει να ληφθούν υπόψη οι εγγενείς περιορισμοί των εργαλείων αυτών στην επεξεργασία του όγκου και της ταχύτητας των δεδομένων που θα προέλθουν από τα συστήματα «Big Data». Παραδόξως, υπάρχει περίπτωση αυτά τα δεδομένα από μόνα τους να έχουν τα χαρακτηριστικά των «Big Data» και ως τέτοια, ενδεχομένως, να απαιτείται να τύχουν επεξεργασίας από μια τέτοια υποδομή. Ωστόσο, καθώς η βέλτιστη πρακτική είναι να διαχωρίζεται η λειτουργία της υποδομής «Big Data» από αυτή της χρήσης της ίδιας της υποδομής για την υποστήριξη και της διαδικασίας της επιθεώρησής της, προτείνεται, εφόσον

είναι εφικτό, τα εργαλεία που πρόκειται να χρησιμοποιηθούν για την επιθεώρηση να υλοποιηθούν και να χρησιμοποιηθούν, όχι από την ίδια την υποδομή «Big Data» που επιθεωρείται, αλλά από άλλη εξωτερική και ανεξάρτητη [53], [54], [55], [56].

Όσον αφορά τη διασφάλιση της ακεραιότητας των πληροφοριών που απαιτούνται για την επιτυχή διεξαγωγή της επιθεώρησης, αυτό μπορεί να επιτευχθεί με την χρήση των μηχανισμών ελέγχου της ακεραιότητας (Integrity Check Mechanisms), όπως είναι αυτός της ψηφιακής σύνοψης (Message Digest). Θα πρέπει να επισημανθεί ότι η ακεραιότητα των δεδομένων αυτών θα πρέπει να διασφαλίζεται σε όλες τις φάσεις της διαδικασίας της επιθεώρησης, ήτοι της συλλογής, της αποθήκευσης, της επεξεργασίας και της χρήσης τους. Ενώ, όσον αφορά την προστασία της εμπιστευτικότητας τους, αυτό μπορεί να επιτευχθεί με την κρυπτογράφησή τους, τόσο όταν βρίσκονται αποθηκευμένα, όσο και όταν αυτά διανέμονται στους εκάστοτε επιθεωρητές. Επιπρόσθετα, θα πρέπει να ελέγχεται και να περιορίζεται η πρόσβαση σε αυτά, έτσι ώστε να αποτραπεί η μη εξουσιοδοτημένη πρόσβαση που μπορεί να έχει ως συνέπεια το ενδεχόμενο αυτά να υποστούν αλλοίωση ή/και να διαγραφούν. Οπότε η χρήση ασφαλών μηχανισμών ελέγχου πρόσβασης (Access Control Mechanisms) μαζί με τους μηχανισμούς εξουσιοδότησης πρόσβασης (Authorization Mechanisms) κρίνεται επιτακτική. Μάλιστα, θα πρέπει η πρόσβαση σε αυτά να είναι άκρως περιορισμένη και να επιτρέπεται αποκλειστικά μόνο σε αρμόδια άτομα, ενώ παράλληλα θα πρέπει να τηρείται λεπτομερές αρχείο καταγραφής όλων των ενεργειών που λαμβάνουν χώρα σε αυτά τα δεδομένα.

Τέλος, άλλη μια λύση, που διερευνάται και είναι υπό ανάπτυξη, τείνει στην δημιουργία ενός ενδιάμεσου επιπέδου «Audit Layer/Orchestrator» που θα παρεμβάλλεται μεταξύ της υποδομής «Big Data» που επιθεωρείται και του επιθεωρητή [1], [51]. Τα πλεονεκτήματα αυτής της λύσης θα είναι, αφενός η υποβοήθηση και η διευκόλυνση του έργου των επιθεωρητών και αφετέρου η προστασία της εμπιστευτικότητας και της «Ιδιωτικότητας» των δεδομένων του συστήματος. Το πρώτο επιτυγχάνεται με την αυτοματοποιημένη συλλογή όλων των απαραίτητων για την επιθεώρηση πληροφοριών από όλα τα εμπλεκόμενα τμήματα της υποδομής και την διάθεσή τους στους εκάστοτε επιθεωρητές, κατόπιν συγκεκριμένου αιτήματός τους. Με αυτό τον τρόπο, επιτυγχάνεται κέρδος χρόνου και κόπου, καθώς οι πληροφορίες αυτές θα ήταν εξαιρετικά δύσκολο και επίπονο να

αναζητηθούν, να εντοπιστούν και να ανασυρθούν στο αχανές και πολύπλοκο περιβάλλον των συστημάτων «Big Data». Το δεύτερο πλεονέκτημα προκύπτει από το γεγονός ότι αυτό το ενδιάμεσο επίπεδο θα λειτουργεί ως ένα επίπεδο αφαίρεσης, τόσο των τεχνικών πληροφοριών του συστήματος, όσο και των δεδομένων των αρχείων καταγραφής, από τους επιθεωρητές. Συγκεκριμένα, αυτό το επίπεδο θα διεκπεραιώνει τα αιτήματα των επιθεωρητών, όπως για παράδειγμα ποιός είχε πρόσβαση στα δεδομένα του αντικειμένου «Χ» την ημερομηνία «Υ», με το να συλλέγει τις αναγκαίες πληροφορίες από τα δεδομένα καταγραφής ενεργειών των τμημάτων της υποδομής που εμπλέκονται κάθε φορά, ανάλογα με το αίτημα, και να τις επιστρέφει σε αυτούς.

2.4 ΕΞΕΤΑΣΗ ΕΙΔΙΚΩΝ ΖΗΤΗΜΑΤΩΝ ΑΣΦΑΛΕΙΑΣ ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Τέλος, σε τρίτη φάση εξετάζονται στην ενότητα αυτή κάποια ειδικά ζητήματα ασφάλειας που προκύπτουν από τα ιδιαίτερα χαρακτηριστικά του περιβάλλοντος των «Big Data». Το πρώτο σχετίζεται άμεσα με τον έλεγχο της ακεραιότητας και της αξιοπιστίας των ποικίλων δεδομένων που προέρχονται από τις διάφορες πηγές του συστήματος «Big Data» και τα οποία πρόκειται να τύχουν επεξεργασίας από τα συστήματα αυτά. Μάλιστα, ένα σημαντικό πρόβλημα, που τίθεται σε αυτή την υποενότητα, αφορά το πως είναι δυνατό αυτά να φιλτραριστούν, πριν εισέλθουν στο σύστημα για περαιτέρω επεξεργασία. Το δεύτερο ζήτημα αφορά στο αν και πως είναι δυνατό να χρησιμοποιηθεί αυτός ο μεγάλος και ποικιλόμορφος όγκος δεδομένων, εκτός από τον πρωταρχικό στόχο που οδήγησε στην χρήση αυτών των συστημάτων, ήτοι την επεξεργασία αυτού του ποικιλόμορφου όγκου δεδομένων για την εξαγωγή χρήσιμης πληροφορίας, ως είσοδος στο ίδιο ή και σε άλλο σύστημα «Big Data» για την εκτέλεση της σε πραγματικό χρόνο ανάλυσης του συστήματος για περιστατικά ασφαλείας και παράλληλα την παροχή ενδείξεων για την ασφαλή και ομαλή λειτουργία του.

2.4.1 ΕΛΕΓΧΟΣ ΕΙΣΕΡΧΟΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΦΙΛΤΡΑΡΙΣΜΑ (END-POINT VALIDATION AND FILTERING)

Τα συστήματα «Big Data» συλλέγουν δεδομένα από πλήθος πηγών, όπως τις διάφορες τερματικές συσκευές (End-Point Devices), τις ποικίλες διαδικτυακές

εφαρμογές και άλλα πληροφοριακά συστήματα. Παρόμοια, το σύστημα ασφαλείας «Security Information and Event Management - SIEM» συγκεντρώνει δεδομένα καταγραφής κίνησης από χιλιάδες συσκευές και εφαρμογές του πληροφοριακού συστήματος που επιτηρεί. Στις περιπτώσεις αυτές, η κύρια ανησυχία που προκύπτει είναι η επικύρωση των εισερχόμενων πληροφοριών ως προς την αξιοπιστία τους. Ερωτήματα, όπως μπορώ να εμπιστευτώ τα δεδομένα αυτά, πως μπορώ να επιβεβαιώσω ότι η πηγή που παρέχει αυτές τις πληροφορίες δεν έχει κακόβουλο σκοπό και πως μπορώ να φιλτράρω αυτά τα κακόβουλα δεδομένα κατά την συλλογή τους, εγείρονται. Συνεπώς, η ανάγκη να εξεταστεί η αξιοπιστία των εισερχόμενων πληροφοριών στο σύστημα «Big Data» και ταυτόχρονα αυτές να φιλτραριστούν αποτελεί ένα εξαιρετικά σημαντικό ζήτημα, το οποίο αποκτά ιδιαίτερη σημασία ειδικά σε αυτό το περιβάλλον, στο οποίο αντλούνται δεδομένα από πολλές μη έμπιστες πηγές.

Μάλιστα, το παραπάνω πρόβλημα είναι κοινό σε όλες τις περιπτώσεις, είτε συλλέγονται δεδομένα από αισθητήρες, είτε συλλέγονται σχόλια από κάποια εφαρμογή κινητού τηλεφώνου, είτε εφαρμόζεται στην επιχείρηση η πολιτική «Φέρε τη Δική σου Συσκευή» (Bring Your Own Device – BYOD) [52]. Ένας επιτιθέμενος με κίνητρο έχει την ικανότητα, είτε να δημιουργήσει ψεύτικα δεδομένα για να αλλοιώσει το τελικό αποτέλεσμα, είτε να εισάγει κακόβουλο κώδικα μέσα στο σύστημα για να παραβιάσει την ασφάλειά του και την «Ιδιωτικότητα» των χρηστών του. Μάλιστα, αυτό το πρόβλημα επιδεινώνεται σημαντικά στο περιβάλλον των «Big Data», στο οποίο το πλήθος των εισερχόμενων δεδομένων στο σύστημα αυξάνει εκθετικά σε όγκο. Οπότε, προκειμένου να αντιμετωπιστεί αποτελεσματικά το πρόβλημα αυτό, θα πρέπει να δημιουργηθούν νέοι αλγόριθμοι που θα μπορούν να αξιολογήσουν αποδοτικά τον τεράστιο όγκο των εισερχόμενων δεδομένων στο σύστημα.

Επομένως, το μοντέλο απειλών για τις περιπτώσεις αυτές περιλαμβάνει τα παρακάτω σενάρια [1]:

1) Ένας επιτιθέμενος μπορεί να παραβιάσει, είτε την συσκευή, είτε την εφαρμογή που συλλέγει και αποστέλλει δεδομένα, προκειμένου να εισάγει κακόβουλο κώδικα ή παραποιημένες πληροφορίες στο κεντρικό σύστημα αποθήκευσης των δεδομένων του συστήματος «Big Data».

2) Εναλλακτικά, ένας επιτιθέμενος δύναται να εκτοξεύσει επιθέσεις «Κλωνοποίησης Ταυτοτήτων» (ID Cloning Attacks) κατά την φάση της συλλογής των

δεδομένων από τις διάφορες πηγές. Αυτό μπορεί να επιτευχθεί με το να δημιουργήσει πολλαπλές ψεύτικες ταυτότητες και μέσω αυτών να εισάγει κακόβουλο κώδικα ή παραποιημένες πληροφορίες στο σύστημα. Η συχνότητα αυτών των επιθέσεων αυξάνει σημαντικά στην περίπτωση που εφαρμόζεται η πολιτική «Φέρε τη Δική σου Συσκευή» (BYOD), καθώς ο επιτιθέμενος μπορεί να συνδέσει την συσκευή του στο σύστημα, να πλαστογραφήσει την ταυτότητα μιας έμπιστης οντότητας και μέσω αυτής της πλαστής ταυτότητας να εισάγει το κακόβουλο περιεχόμενο στο σύστημα.

3) Ένα πιο πολύπλοκο σενάριο επίθεσης είναι αυτό κατά το οποίο ο επιτιθέμενος δύναται να αλλοιώσει τα δεδομένα που συλλέγουν διάφορες πηγές, όπως οι διάφοροι αισθητήρες ή οι συσκευές «GPS», χωρίς να απαιτείται προηγουμένως να τις έχει παραβιάσει όπως στο πρώτο σενάριο. Έτσι για παράδειγμα, αντί ο επιτιθέμενος να παραβιάσει ένα αισθητήρα καιρού για να αλλοιώσει τις μετρήσεις του, μπορεί κάλλιστα να μεταβάλλει τεχνητά τις τιμές της θερμοκρασίας που μετρώνται για μια τοποθεσία, τοποθετώντας δίπλα στον αισθητήρα μια πηγή θερμότητας, όπως για παράδειγμα μια θερμάστρα.

4) Επίσης, ένας επιτιθέμενος έχει την δυνατότητα να εκτοξεύσει επιθέσεις «Man-in-the-Middle» και «Replay» κατά την φάση της μετάδοσης των δεδομένων από μια έμπιστη πηγή προς το κεντρικό σύστημα συλλογής και αποθήκευσης των πληροφοριών του συστήματος «Big Data».

Σύμφωνα λοιπόν με το ανωτέρω μοντέλο, οι λύσεις για την αντιμετώπιση του προβλήματος αυτού μπορούν να ενταχθούν σε δυο κατηγορίες. Στην πρώτη κατηγορία εντάσσονται αυτές που θα αποτρέπουν ένα επιτιθέμενο από το να δημιουργήσει και να στείλει κακόβουλο περιεχόμενο στο σύστημα «Big Data». Ενώ στην δεύτερη κατηγορία εντάσσονται αυτές που θα εντοπίζουν και θα φιλτράρουν το κακόβουλο περιεχόμενο, που έχει ήδη αποθηκευτεί στο σύστημα «Big Data», αν ο επιτιθέμενος πέτυχε να το εισάγει σε αυτό. Οι λύσεις συνεπώς, που αναφέρονται στο [51], παρουσιάζονται συνοπτικά στις επόμενες παραγράφους.

Κατ’ αρχάς, για την αποτροπή του επιτιθέμενου από το να στείλει στο σύστημα «Big Data» κακόβουλο περιεχόμενο, απαιτείται να χρησιμοποιείται σε αυτό ασφαλές λογισμικό και να θωρακιστεί από πλευράς ασφάλειας το υλικό. Η χρήση ασφαλούς λογισμικού έχει μελετηθεί και υλοποιηθεί σε μεγάλο βαθμό στις εφαρμογές και το λογισμικό των σταθερών υπολογιστών, αν και το να υπάρξει λογισμικό πλήρως

απαλλαγμένο από αδυναμίες είναι αδύνατο, ενώ το αντίστοιχο των κινητών συσκευών είναι ακόμα υπό διερεύνηση και ως εκ τούτου αυτές οι συσκευές και το λογισμικό τους καθίσταται πρωταρχικός στόχος των επιτιθέμενων σήμερα. Συνεπώς, για να προστατευθεί περαιτέρω το λογισμικό και οι εφαρμογές των διαφόρων συσκευών του συστήματος «Big Data», κρίνεται επιτακτικό να χρησιμοποιηθούν σε αυτές μηχανισμοί ελέγχου πρόσβασης (Access Control Mechanisms), ενημερωμένα προγράμματα «Anti-Virus» και «Anti-Malware», μηχανισμοί «Intrusion Detection / Prevention Systems – IPS/IDS», καθώς και να λειτουργούν απαρέγκλιτα μηχανισμοί τήρησης αρχείων με δεδομένα κίνησης και ενεργειών (Logging and Monitoring Files). Παράλληλα, είναι σημαντικό να υλοποιηθούν στις διάφορες εφαρμογές των πηγών κατάλληλοι μηχανισμοί που θα ελέγχουν τα δεδομένα που δίδονται ως είσοδος σε αυτές, προκειμένου, κατόπιν εξέτασης και αξιολόγησης των ιδιοτήτων, της μορφής και των αναμενόμενων τιμών τους ανάλογα με την εκάστοτε περίπτωση, αυτά να απορρίπτονται αν δεν ικανοποιούν τις απαιτήσεις των προκαθορισμένων φίλτρων του αντίστοιχου μηχανισμού. Έτσι θα πραγματοποιείται ένα αρχικό φιλτράρισμα των δεδομένων, πριν αυτά εισέλθουν στο σύστημα. Ωστόσο, αν και αυτό από μόνο του δεν είναι ικανό να αποτρέψει πλήρως τις επιθέσεις αυτές, μπορεί εντούτοις να περιορίσει σε ικανοποιητικό βαθμό το εύρος τους. Όσον αφορά την θωράκιση του υλικού, η χρήση των «Trusted Platform Modules - TPMs» στις συσκευές του συστήματος μπορεί να εγγυηθεί την ακεραιότητα και την εμπιστευτικότητα των δεδομένων που αποστέλλονται στο σύστημα από αυτές, όπως για παράδειγμα στην περίπτωση των αισθητήρων [48]. Βέβαια, ακόμα και με την χρήση του «TPM» στις συσκευές, δεν διασφαλίζεται πλήρως ότι ο επιτιθέμενος δεν μπορεί να αλλοιώσει τα δεδομένα που στέλνουν οι συσκευές αυτές, όπως για παράδειγμα στην περίπτωση που περιγράφεται στο τρίτο σενάριο του μοντέλου απειλών.

Επιπρόσθετα, για την αντιμετώπιση και αποτροπή των επιθέσεων που διεξάγονται με την χρήση πλαστών ταυτοτήτων (ID Spoofing Attacks), προτείνεται η υιοθέτηση σχημάτων στα οποία χρησιμοποιούνται έμπιστα πιστοποιητικά (Trusted Certificates) και έμπιστες ασφαλείς συσκευές (Trusted Devices). Το πιο διαδεδομένο σχήμα προς αυτή την κατεύθυνση είναι η «Υποδομή Δημοσίου Κλειδιού» (Public Key Infrastructure - PKI), στο οποίο σε κάθε φυσική οντότητα αντιστοιχεί ένα ασφαλές ψηφιακό πιστοποιητικό (Digital Certificate) [173], [174]. Σύμφωνα με αυτό, δημιουργείται ένα ζεύγος ιδιωτικού (Private Key) και δημοσίου κλειδιού (Public Key), τα οποία διαχειρίζεται μια «Έμπιστη Τρίτη Οντότητα» (Trusted Third Party) η

«Certification Authority - CA» και τα οποία χρησιμοποιούνται κάθε φορά για την επιβεβαίωση της αυθεντικότητας της ταυτότητας του αποστολέα των δεδομένων στο σύστημα. Αν και θεωρείται αξιόπιστη λύση λόγω της ασφάλειας που παρέχει, το βασικό μειονέκτημά της είναι ότι καθίσταται μη αποδοτική και δυσχερής όσον αφορά την διαχείριση των πιστοποιητικών όλων των οντοτήτων που ενδεχομένως θα υπάρχουν στο περιβάλλον των «Big Data».

Εναλλακτικά, έχουν προταθεί διάφορες προσεγγίσεις που εστιάζουν πιο πολύ στην ανίχνευση των επιθέσεων αυτών, παρά στην αποτροπή τους. Σύμφωνα με αυτές, ελέγχονται οι δυνατότητες των πηγών του συστήματος (Resource Testing) [57], όπως για παράδειγμα η υπολογιστική ισχύς, οι δυνατότητες αποθήκευσης και το εύρος ζώνης του δικτύου τους. Αρχικά αντιστοιχίζονται σε κάθε πηγή του συστήματος οι υπολογιστικές δυνατότητες της συσκευής που την υποστηρίζει και οι οποίες είναι συγκεκριμένες. Οπότε, με δεδομένες τις υπολογιστικές δυνατότητες της κάθε πηγής, είναι εφικτό κάθε φορά να επαληθεύεται κατά πόσο η πηγή, που θέλει να στείλει δεδομένα στο σύστημα, είναι πραγματική και όχι πλαστή. Αυτό επιτυγχάνεται με την σύγκριση των δυνατοτήτων, που ισχυρίζεται ότι έχει η εν λόγω πηγή, με αυτές που αντιστοιχούν στην πραγματική πηγή του συστήματος [58]. Επίσης, άλλος τρόπος για τον εντοπισμό των πλαστών οντοτήτων ενός συστήματος είναι μέσω της μέτρησης των συνολικών δυνατοτήτων των υπό εξέταση οντοτήτων. Οπότε στη συνέχεια ελέγχεται αν αυτές διαθέτουν λιγότερους πόρους από αυτούς που θα αναμενόταν να έχουν οι αντίστοιχες ανεξάρτητες γνήσιες οντότητες, για να διαπιστωθεί η γνησιότητά τους. Ωστόσο, αυτές οι προσεγγίσεις παρέχουν ελάχιστη ασφάλεια στο σύστημα και δεν αναμένεται να συμβάλλουν σημαντικά στην αποτροπή των επιτιθέμενων από την εκτόξευση τέτοιων επιθέσεων.

Όσον αφορά την ανίχνευση και το φιλτράρισμα του κακόβουλου περιεχομένου που έχει ήδη εισαχθεί στο σύστημα «Big Data», αυτό κατά πάσα πιθανότητα θα εμφανιστεί ως ακραία τιμή. Οπότε στην περίπτωση αυτή μπορούν να χρησιμοποιηθούν διάφορες τεχνικές για την ανίχνευσή του, όπως η «Statistical Similarity Detection» για τον έλεγχο της ομοιότητας των τιμών και η «Outlier Detection» για την εύρεση των ακραίων τιμών [52]. Ωστόσο, οι τεχνικές για το πλήρες φιλτράρισμα των δεδομένων που έχουν ήδη εισαχθεί στο σύστημα «Big Data» και για την εν συνεχεία απόρριψη του κακόβουλου περιεχομένου, πριν την φάση της

επεξεργασίας του για την εξαγωγή πολύτιμης γνώσης, αποτελούν ανοιχτό ερευνητικό πεδίο.

Επομένως, από τα παραπάνω διαπιστώνεται ότι δεν υπάρχει κάποια συγκεκριμένη προσέγγιση που να είναι αποτελεσματική στην επικύρωση της αξιοπιστίας των εισερχόμενων πληροφοριών στο σύστημα «Big Data» και που ταυτόχρονα να μπορεί να τις φιλτράρει. Οπότε αυτό που προτείνεται στο [1] είναι να ακολουθούνται στην πράξη τα παρακάτω βήματα κατά τον σχεδιασμό του συστήματος «Big Data», ανάλογα με την εκάστοτε περίπτωση. Αρχικά θα πρέπει οι σχεδιαστές του συστήματος να δώσουν υπέρτατη προσοχή στην ανάπτυξη ασφαλών εφαρμογών και δομών για την συλλογή των δεδομένων από τις διάφορες πηγές. Παράλληλα θα πρέπει να αξιολογηθεί σοβαρά και με προσοχή η πολιτική «Φέρε τη Δική σου Συσκευή» (BYOD), προτού εφαρμοστεί. Στη συνέχεια, θα πρέπει να εντοπιστούν τα σημεία εκείνα του συστήματος στα οποία υπάρχει πιθανότητα να εκδηλωθούν «Επιθέσεις Πλαστών Ταυτοτήτων» (ID Spoofing Attacks) και παράλληλα να υιοθετηθούν λύσεις, που θα συνδυάζουν την αποδοτικότητα με το κόστος, για τον μετριασμό των επιπτώσεων των επιθέσεων αυτών. Τέλος, λαμβάνοντας υπόψη ότι ένας αποφασισμένος επιτιθέμενος τελικά θα καταφέρει να στείλει κακόβουλο περιεχόμενο στο σύστημα, θα πρέπει να σχεδιαστούν και να αναπτυχθούν αποδοτικοί αλγόριθμοι για την ανίχνευση και το φιλτράρισμα του περιεχομένου αυτού.

2.4.2 ΕΛΕΓΧΟΣ ΤΗΣ ΑΣΦΑΛΕΙΑΣ ΤΗΣ ΥΠΟΔΟΜΗΣ ΣΕ ΠΡΑΓΜΑΤΙΚΟ ΧΡΟΝΟ (REAL TIME SECURITY / COMPLIANCE MONITORING)

Όπως έχει αναφερθεί και σε προηγούμενη ενότητα, η σε «Πραγματικό Χρόνο Παρακολούθηση» της ασφάλειας ενός πληροφοριακού συστήματος είναι μια μεγάλη πρόκληση (Real Time Security/Compliance Monitoring). Ο βασικότερος λόγος είναι η ποσότητα των ειδοποιήσεων που παράγεται από τις διάφορες συσκευές ασφαλείας του συστήματος. Οι ειδοποιήσεις αυτές, στο πλείστο των περιπτώσεων, είτε συσχετιστούν, είτε όχι, συνήθως θεωρούνται εσφαλμένα ως «False Positives» εξαιτίας της περιορισμένης ικανότητας του ανθρώπου στο να τις αναλύσει και να τις αξιολογήσει αποτελεσματικά, με συνέπεια αυτές να αγνοούνται. Μάλιστα, το πρόβλημα αυτό αναμένεται να επιδεινωθεί στο περιβάλλον των «Big Data», αν λάβουμε υπόψη, αφενός τα ιδιαίτερα χαρακτηριστικά του, ήτοι τον όγκο, την ταχύτητα

και την ποικιλομορφία των δεδομένων που αυτό επεξεργάζεται, και αφετέρου την πολυπλοκότητα της αρχιτεκτονικής του, ήτοι το πλήθος των συσκευών από τις οποίες αυτό αποτελείται.

Εντούτοις, όσο και αν φαίνεται περίεργο, η λύση στο ανωτέρω πρόβλημα προέρχεται μέσα από την ίδια την τεχνολογία των «Big Data». Η δυνατότητά της για ταχεία επεξεργασία και αποτελεσματική ανάλυση του τεράστιου όγκου των ποικιλόμορφων τύπων δεδομένων αποτελεί σημαντικό πλεονέκτημα, που θα προσδώσει σημαντική ώθηση στην αποτελεσματική αντιμετώπιση του προβλήματος αυτού. Συνεπώς, διαπιστώνεται ότι πλέον η συσχέτιση των εννοιών των «Big Data» και της «Ασφάλειας» είναι αμφίδρομη. Έτσι, αυτές δεν συσχετίζονται μόνο προς την κατεύθυνση της «Ασφάλειας» των «Big Data» συστημάτων, αλλά αντίστροφα και προς την κατεύθυνση κατά την οποία χρησιμοποιούνται τα «Big Data» για να βελτιώσουν την «Ασφάλεια», είτε άλλων πληροφοριακών συστημάτων, είτε και της ίδιας της υποδομής τους. Οπότε πλέον, οι τεχνολογίες των «Big Data» και «Big Data Analytics» μετατρέπονται εκτός από αντικείμενα της «Ασφάλειας» και σε χρήσιμα εργαλεία της [53], [54], [55], [56].

Ωστόσο, το γεγονός αυτό εγείρει δυο εξαιρετικά σημαντικά ζητήματα.

- Το πρώτο σχετίζεται με το **αν** θα πρέπει να χρησιμοποιηθεί η ίδια η υποδομή «Big Data» για την σε «Πραγματικό Χρόνο Παρακολούθηση» της ασφάλειας της ίδιας της υποδομής της (Διαδικασία Αυτοελέγχου) παράλληλα με την εκτέλεση της αποστολής για την οποία αυτή αρχικά σχεδιάστηκε.
- Το δεύτερο σχετίζεται με το **πως** είναι εφικτό αυτό να πραγματοποιηθεί ασφαλώς. Δηλαδή με ποιο τρόπο είναι δυνατό να επιτευχθούν ταυτόχρονα και αποτελεσματικά, αφενός η ασφάλεια της «Διαδικασίας Αυτοελέγχου» και αφετέρου η ασφάλεια του συστήματος «Big Data» μέσω της «Διαδικασίας Αυτοελέγχου».

Όσον αφορά την απάντηση του πρώτου ερωτήματος, με δεδομένο ότι η σχεδίαση και η ανάπτυξη ενός συστήματος «Big Data» πραγματοποιείται κάθε φορά υπό το πρίσμα, τόσο των ιδιαίτερων συνθηκών του εκάστοτε προβλήματος που τίθεται, όσο και του επιδιωκόμενου σκοπού μέσα από την λύση του, γίνεται εύκολα αντιληπτό ότι η βέλτιστη λύση είναι αυτή του διαχωρισμού των ανωτέρω ενεργειών. Σε αυτό συντείνει και το γεγονός ότι, ανάλογα με το εκάστοτε πρόβλημα, κάθε φορά συλλέγονται διαφορετικά δεδομένα εισόδου, υλοποιούνται διαφορετικοί αλγόριθμοι

για την επεξεργασία και ανάλυσή τους και χρησιμοποιούνται διαφορετικά μοντέλα για την ερμηνεία τους. Επομένως, η βέλτιστη πρακτική είναι η χρήση, εάν είναι δυνατό, διαφορετικού συστήματος «Big Data» για την αποδοτική και αποτελεσματική σε «Πραγματικό Χρόνο Παρακολούθηση» της συνολικής ασφάλειας του αρχικού συστήματος «Big Data» που ήδη χρησιμοποιείται. Ωστόσο, αν αυτό δεν είναι εφικτό να πραγματοποιηθεί για διάφορους λόγους, τότε η χρήση της ίδιας της υποδομής «Big Data» για την αυτασφάλισή της είναι μονόδρομος. Οπότε, κατόπιν των παραπάνω, το δεύτερο ερώτημα απλοποιείται περαιτέρω και ανάγεται σε τελική ανάλυση και για τις δυο περιπτώσεις στην εγγενή ασφάλεια της υποδομής «Big Data» που θα πραγματοποιήσει την σε πραγματικό χρόνο παρακολούθηση της ασφάλειας, είτε του εαυτού της, είτε των άλλων πληροφοριακών συστημάτων μεταξύ των οποίων και άλλων συστημάτων «Big Data».

Καθώς λοιπόν η σε πραγματικό χρόνο παρακολούθηση της ασφάλειας ενός πληροφοριακού συστήματος για την έγκαιρη ανίχνευση περιστατικών ασφαλείας είναι κρίσιμη, τα οφέλη που θα προκύψουν από την χρήση των συστημάτων «Big Data» προς αυτή την κατεύθυνση θα είναι μεγάλα. Έτσι πλέον θα είναι εφικτό να απαντώνται σε πραγματικό χρόνο ερωτήσεις όπως ποιός ζητά πρόσβαση, σε ποιά δεδομένα, από ποιά πηγή και πότε, ή έχουμε δεχτεί επίθεση, ή υπάρχει παραβίαση της πολιτικής ασφαλείας εξαιτίας της τάδε ενέργειας κλπ. Παράλληλα, θα μειωθεί σημαντικά ο αριθμός των «False Positives» σε ένα σύστημα και θα αυξηθεί η ποιότητα των «True Positives», που μεταφράζεται σε ανίχνευση περισσότερων πραγματικών περιστατικών ασφαλείας για το σύστημα αυτό. Έτσι, για παράδειγμα οι οργανισμοί παροχής υπηρεσιών υγείας και ασφάλισης θα επωφεληθούν διπλά από την χρήση των τεχνολογιών αυτών. Αφενός θα είναι σε θέση να εντοπίζουν με καλύτερη ακρίβεια τις περιπτώσεις εξαπάτησης, εξοικονομώντας τεράστια ποσά. Αφετέρου θα προστατεύσουν αποτελεσματικά, τόσο τις πληροφορίες που τηρούν, όσο και την διαδικασία της επεξεργασίας αυτών από τα πληροφοριακά τους συστήματα. Αυτό θα επιτευχθεί χάρη στην δυνατότητα να ανιχνεύονται σε πραγματικό χρόνο οι μη φυσιολογικές συμπεριφορές, είτε αυτές προκαλούνται εκούσια, είτε ακούσια και ακολούθως να λαμβάνονται έγκαιρα οι δέουσες ενέργειες για την αποτροπή τους και για την αποφυγή της πρόκλησης περαιτέρω ζημιάς [52].

Συνεπώς, κατόπιν των παραπάνω, διαπιστώνεται ότι όσον αφορά την αποτελεσματική αντιμετώπιση του προβλήματος της σε «Πραγματικό Χρόνο

Παρακολούθησης» της ασφάλειας ενός πληροφοριακού συστήματος (Real Time Security / Compliance Monitoring), απαραίτητη προϋπόθεση είναι η εγγενής ασφάλεια του συστήματος «Big Data» που θα αναλάβει τον ρόλο αυτό. Ένα τέτοιο σύστημα όμως συνθέτεται από πολλά επιμέρους τμήματα. Οπότε η ασφάλειά του αναλύεται, αφενός στην ασφάλεια του εκάστοτε δομικού τμήματός του και αφετέρου στην ασφάλεια της διασύνδεσης των τμημάτων αυτών μεταξύ τους. Με γνώμονα αυτό, οι απειλές ενός τέτοιου συστήματος περιλαμβάνουν τα παρακάτω σενάρια [1]:

1) Ένας κακόβουλος διαχειριστής του συστήματος αυτού μπορεί να αποκτήσει πρόσβαση στις εφαρμογές και στους κόμβους του με απώτερο στόχο να το καταστήσει αναποτελεσματικό (Insider Attacks).

2) Ένας επιτιθέμενος μπορεί, εκμεταλλευόμενος τις διάφορες αδυναμίες των εφαρμογών του, να αποκτήσει παράνομη πρόσβαση σε αυτό και περαιτέρω να επηρεάσει σημαντικά την απόδοσή του.

3) Ένας επιτιθέμενος μπορεί να υποκλέψει τις επικοινωνίες του με επιθέσεις «Man-in-the-Middle».

4) Ένας επιτιθέμενος θα προσπαθήσει να το αποφύγει με «Evasion Attacks» επιθέσεις [49], έτσι ώστε να μην εντοπιστεί.

5) Ένας επιτιθέμενος θα προσπαθήσει να υποβαθμίσει την ποιότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση των αλγορίθμων του συστήματος αυτού, οι οποίοι αποσκοπούν στην ανίχνευση των διαφόρων απειλών και των περιστατικών ασφαλείας, με «Data Poisoning Attacks» επιθέσεις [50]. Απώτερος σκοπός αυτών των επιθέσεων είναι, είτε να οδηγήσουν τους αλγορίθμους προς την λάθος κατεύθυνση κατά την λήψη κάποιας απόφασης, είτε να προκαλέσουν την δυσλειτουργία τους, είτε να αναστείλουν την λειτουργία τους.

Επομένως, για την καλύτερη αντιμετώπιση των παραπάνω απειλών, σύμφωνα με το [51], προτείνονται οι παρακάτω λύσεις. Κατ' αρχάς, η χρήση της τεχνολογίας των «Big Data Analytics» μπορεί να συμβάλλει αποτελεσματικά στην ανίχνευση των ανώμαλων συμπεριφορών και των ύποπτων ενεργειών στο σύστημα αυτό. Η χρήση όμως αυτής της τεχνολογίας προϋποθέτει, αφενός την ύπαρξη των αρχείων καταγραφής κίνησης και ενεργειών του συστήματος (Log Files) και αφετέρου την λειτουργία του κατάλληλου εργαλείου, όπως για παράδειγμα του εργαλείου «Security Information and Event Management – SIEM», για την ανάλυση αυτών των αρχείων. Επίσης, προτείνεται να δημιουργηθούν κατάλληλοι αλγόριθμοι οι οποίοι μέσω της διαδικασίας του «Machine Learning» θα μπορούν να ανιχνεύουν αυτόματα, τόσο τις

ανώμαλες συμπεριφορές, όσο και τα πιθανά περιστατικά ασφαλείας. Μάλιστα, οι δυνατότητες των αλγορίθμων αυτών επεκτείνονται περαιτέρω και στην ικανότητα αυτοί να προσαρμόζουν την λειτουργία τους με βάση την «εμπειρία» που αποκτούν.

Όσον αφορά την προστασία των επιμέρους τμημάτων του συστήματος, αυτό μπορεί να επιτευχθεί μέσω της υλοποίησης των αντίστοιχων μέτρων ασφαλείας μπροστά από αυτά (Front-End Systems). Σε αυτά τα μέτρα συμπεριλαμβάνονται η χρήση του «Firewall», των «Application Level Firewalls», του «Antivirus», των «Anti-Malwares» κλπ, τα οποία μπορούν να ανιχνεύσουν και ταυτόχρονα να σταματήσουν κακόβουλες συμπεριφορές, είτε μέσω «Signatures», είτε μέσω του ελέγχου της συμπεριφοράς τους σε απομονωμένο περιβάλλον «Behavior Profiles». Μάλιστα, προτείνεται αυτά τα μέτρα να κλιμακώνονται σε διάφορα επίπεδα και να χρησιμοποιούνται σε συνδυασμό, ανάλογα με τα τμήματα που εμπλέκονται κάθε φορά. Παράλληλα, θα πρέπει να εξετάζονται ξεχωριστά και τα θέματα ασφαλείας της υποδομής «Cloud» που ενδεχομένως θα υποστηρίξει το σύστημα αυτό.

Επιπλέον, όσον αφορά την προστασία των συνδέσεων μεταξύ των διαφόρων τμημάτων του συστήματος και την εξασφάλιση ότι μόνο έγκυρες συνδέσεις θα πραγματοποιούνται μεταξύ των τμημάτων αυτών, η χρήση λύσεων ασφαλείας όπως το «TLS/SSL», το «IPsec», το «Secure Shell - SSH» κλπ, κρίνεται επιτακτική. Μάλιστα, αν απαιτείται, συστήνεται αυτές να υλοποιούνται ακόμα και μεταξύ των επιμέρους τμημάτων του κάθε βασικού δομικού στοιχείου του συστήματος αυτού. Επιπρόσθετα, η χρήση του ασφαλούς μηχανισμού ελέγχου πρόσβασης «Kerberos» μπορεί να συμβάλλει σημαντικά προς την κατεύθυνση αυτή, καθώς μέσω αυτού θα διασφαλίζεται η αυθεντικότητα των οντοτήτων, πριν την εγκατάσταση της σύνδεσης μεταξύ τους.

Από την άλλη πλευρά, η ασφάλεια των διαφόρων εφαρμογών του συστήματος αυξάνει όσο μειώνονται οι αδυναμίες τους. Έτσι για να μειωθούν οι ευπάθειες των δικτυακών εφαρμογών, προτείνεται να εφαρμόζονται οι βέλτιστες πρακτικές του οργανισμού «OWASP» [192]. Ενώ για τις υπόλοιπες εφαρμογές προτείνεται ο περιοδικός έλεγχος τους μέσω των διαδικασιών του «Vulnerability Assessment» και του «Application Level Penetration Test», έτσι ώστε να εξετάζεται τακτικά το επίπεδο ασφάλειας των εφαρμογών αυτών και σε περίπτωση που βρεθούν ευπάθειες να ενημερώνονται καταλλήλως.

Τέλος, όσοι ασχολούνται με τον σχεδιασμό και την ανάπτυξη των αλγορίθμων για «Data Mining» και «Data Analytics» θα πρέπει να λάβουν σοβαρά υπόψη τα προβλήματα που υπεισέρχονται σε αυτούς τους αλγόριθμους λόγω στατιστικής, προκειμένου να μετριάσουν τα αποτελέσματα των επιθέσεων «Evasion Attack» και «Poisoning Attack». Παράλληλα, για την αποτροπή των «Evasion Attacks», καθώς αυτές εξελίσσονται συνεχώς, η καλύτερη λύση στην παρούσα φάση είναι η δημιουργία ενός πολλαπλών επιπέδων σχεδίου άμυνας με την χρήση διαφορετικών στη λειτουργία αλγορίθμων για την ανίχνευσή τους. Ενώ για την αντιμετώπιση των «Data Poisoning Attacks» προτείνονται, αφενός η χρήση μηχανισμών που θα ελέγχουν τα δεδομένα που δίδονται ως είσοδο για την εκπαίδευση αυτών των αλγορίθμων και αφετέρου η χρήση μηχανισμών ελέγχου πρόσβασης, έτσι ώστε μόνο εξουσιοδοτημένες έμπιστες πηγές να παρέχουν τα δεδομένα αυτά στο σύστημα. Παράλληλα, σημαντική θεωρείται και η λεπτομερής καταγραφή, τόσο του ιστορικού κίνησης, όσο και των ενεργειών που λαμβάνουν χώρα, στα αντίστοιχα αρχεία του συστήματος.

2.5 ΣΥΝΟΨΗ ΚΕΦΑΛΑΙΟΥ

Τα «Big Data» χρησιμοποιούνται εδώ και δεκαετίες από τις κυβερνήσεις και τις μεγάλες επιχειρήσεις. Μάλιστα το γεγονός, ότι οι υποδομές τους ήταν συνήθως ιδιόκτητες και ημι-απομονωμένες από άλλα δίκτυα, καθιστούσε πιο εύκολη την διαχείριση των ζητημάτων ασφαλείας τους. Ωστόσο, οι ραγδαίες τεχνολογικές εξελίξεις που έχουν συντελεστεί στην σημερινή εποχή, όπως η διασύνδεση των διαφόρων δικτύων μεταξύ τους, η επίλυση των προβλημάτων που σχετίζονταν με τα μέσα αποθήκευσης και την υπολογιστική ισχύ, τα νέα εργαλεία και τεχνικές που έχουν εφευρεθεί για την αποτελεσματική επεξεργασία και ανάλυση αυτού του ποικιλόμορφου όγκου δεδομένων που παράγεται από πλήθος συσκευών, το «Cloud», καθώς και η δυνατότητα πρόσβασης σε αυτά από όλους τους οργανισμούς όλων των μεγεθών, διαμόρφωσαν νέα δεδομένα και συνθήκες, υπό τις οποίες οι παραδοσιακοί μηχανισμοί ασφαλείας πλέον καθίστανται αναποτελεσματικοί και ανεπαρκείς. Στο κεφάλαιο αυτό λοιπόν εξετάστηκαν τα ζητήματα που σχετίζονται με την ασφάλεια στο περιβάλλον των «Big Data», υπό το πρίσμα των νέων αυτών συνθηκών. Μάλιστα, για να είναι σφαιρική και πολύπλευρη η μελέτη της ασφάλειας

των «Big Data», αυτή επικεντρώθηκε σε επιμέρους τομείς, ήτοι την ασφάλεια της υποδομής τους, την ασφάλεια της διαδικασίας επεξεργασίας των δεδομένων και την εξέταση κάποιων ειδικών ζητημάτων που προκύπτουν στο νέο αυτό περιβάλλον. Αρχικά, εξετάστηκε η ασφάλεια της υποδομής που υποστηρίζει το περιβάλλον των «Big Data». Συγκεκριμένα μελετήθηκαν τα ζητήματα που σχετίζονται με την εκτέλεση ασφαλών υπολογισμών σε κατανεμημένα συστήματα, καθώς και με την ασφάλεια των «Μη - Σχεσιακών» αποθηκών δεδομένων που χρησιμοποιούνται ευρέως στα συστήματα αυτά. Έπειτα εξετάστηκε η ασφάλεια της διαδικασίας επεξεργασίας των δεδομένων. Έτσι διερευνήθηκαν τα θέματα που σχετίζονται με τον έλεγχο της προέλευσης των δεδομένων από τις διάφορες πηγές του συστήματος, την ασφάλεια κατά την αποθήκευση των δεδομένων αυτών και την τήρηση ιστορικού ενεργειών, καθώς και την διεξαγωγή αποτελεσματικών επιθεωρήσεων επί της διαδικασίας επεξεργασίας των δεδομένων. Τέλος, εξετάστηκαν τα ζητήματα που σχετίζονται, τόσο με την ακεραιότητα των δεδομένων εισόδου, όσο και με την επιτήρηση του συστήματος «Big Data» σε πραγματικό χρόνο. Από την έρευνα που διεξήχθη επ' αυτών, διαπιστώθηκε ότι για κάθε επιμέρους τομέα προκύπτουν διάφορα ζητήματα ασφάλειας, τα οποία και παρατέθηκαν. Μάλιστα, για το κάθε ένα ζήτημα παρουσιάστηκαν οι διάφορες λύσεις που έχουν προταθεί από την επιστημονική κοινότητα για την αντιμετώπισή τους. Εντούτοις, ορισμένες από τις προτεινόμενες λύσεις δημιουργούν πρόσθετα ζητήματα που σχετίζονται, τόσο με την αποτελεσματικότητά τους, όσο και με την επίδραση που έχουν στην απόδοση του συστήματος, και επομένως απαιτείται να τροποποιηθούν και να προσαρμοστούν στα νέα δεδομένα. Ενώ κάποιες άλλες, είτε βρίσκονται ακόμα σε ερευνητικό στάδιο, είτε αποτελούν μόνο προτάσεις. Επιπρόσθετα, υπάρχουν ορισμένα ζητήματα που παραμένουν ανοιχτά και άλυτα. Συνεπώς, όλα αυτά τα θέματα θα πρέπει να εξεταστούν ενδελεχώς και επισταμένα, προκειμένου να ανευρεθούν οι κατάλληλες λύσεις που θα καταστίσουν τα «Big Data» ασφαλή.

3. ΚΕΦΑΛΑΙΟ 3ο: ΠΡΟΣΤΑΣΙΑ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ (PRIVACY) ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

3.1 ΠΡΟΛΟΓΟΣ

Παράλληλα με την εξέταση της ασφάλειας στα συστήματα «Big Data» συνεξετάζονται και τα θέματα που σχετίζονται με την προστασία της «Ιδιωτικής Ζωής» και των «Προσωπικών Δεδομένων» στα συστήματα αυτά. Η εκτεταμένη συλλογή και περαιτέρω επεξεργασία από τα συστήματα «Big Data» αυτού του μεγάλου ποικιλόμορφου όγκου δεδομένων, στον οποίο συνήθως περιέχονται και προσωπικά δεδομένα, έχει προκαλέσει σοβαρές ανησυχίες και προβληματισμούς, αναδεικνύοντας το ζήτημα της προστασίας της «Ιδιωτικότητας» (Privacy) σε μείζον θέμα. Ειδικότερα, οι ανησυχίες αυτές εστιάζουν σε ζητήματα που σχετίζονται με την ευρείας κλίμακας ηλεκτρονική επιτήρηση, την αποκάλυψη προσωπικών ή/και ευαίσθητων προσωπικών δεδομένων σε τρίτους και την κατηγοριοποίηση και αντιμετώπιση των ανθρώπων με βάση το ηλεκτρονικό τους προφίλ (Profiling), που μπορεί να οδηγήσει περαιτέρω σε διακρίσεις και ρατσισμό [59].

Προκειμένου λοιπόν να απολαμβάνουμε τα πλεονεκτήματα που προκύπτουν από την χρήση των συστημάτων «Big Data» και παράλληλα να έχουμε τον μέγιστο δυνατό βαθμό προστασίας της «Ιδιωτικότητας», καθίσταται επιτακτικό, αφενός να περιοριστούν τα προβλήματα που προκύπτουν από την χρήση των συστημάτων αυτών και αφετέρου να ενσωματωθούν σε αυτά οι κατάλληλοι μηχανισμοί και δικλίδες ασφαλείας, που θα εγγυώνται την προστασία της ιδιωτικής ζωής του ατόμου. Παράλληλα, η έννοια της «Ιδιωτικότητας εκ Σχεδιασμού» (Privacy By Design) αποκτά ιδιαίτερη σημασία στο νέο αυτό περιβάλλον, καθώς αποτελεί την εγγύηση για την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων των ατόμων. Σύμφωνα με αυτή, θα πρέπει κατά την φάση του σχεδιασμού να προσδιοριστούν οι απαιτήσεις για την προστασία της «Ιδιωτικότητας» και εν συνεχεία κατά την φάση της ανάπτυξης που ακολουθεί να υλοποιηθούν όλοι εκείνοι οι απαραίτητοι τεχνικοί και οργανωτικοί μηχανισμοί που θα συμβάλουν αποτελεσματικά στην επίτευξή των απαιτήσεων αυτών. Συνεπώς, ο αντικειμενικός σκοπός της παρούσας ενότητας είναι η λεπτομερής εξέταση των ζητημάτων που σχετίζονται με την προστασία της «Ιδιωτικότητας» στο περιβάλλον των «Big Data» και κατ' επέκταση η υιοθέτηση και η

εφαρμογή κάθε φορά των κατάλληλων μηχανισμών και τεχνικών που θα την διασφαλίζουν στον μέγιστο δυνατό βαθμό.

3.2 ΟΡΙΣΜΟΙ «ΙΔΙΩΤΙΚΟΤΗΤΑΣ» (PRIVACY) ΚΑΙ «ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΕΚ ΣΧΕΔΙΑΣΜΟΥ» (PRIVACY BY DESIGN)

Κρίνεται σκόπιμο, πριν την εξέταση των ζητημάτων που άπτονται της προστασίας της «Ιδιωτικότητας» στο περιβάλλον των «Big Data», να παρατεθούν κατ' αρχάς οι ορισμοί της «Ιδιωτικότητας» (Privacy) και της «Ιδιωτικότητας εκ Σχεδιασμού» (Privacy By Design). Στη συνέχεια κρίνεται απαραίτητο να παρουσιαστεί το ισχύον νομικό πλαίσιο για την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων, τόσο στην Ευρωπαϊκή Ένωση, όσο και στην Ελλάδα. Ο ρόλος της κείμενης νομοθεσίας είναι ιδιαίτερα σημαντικός, γιατί αποτελεί τον βασικό πυλώνα για την ουσιαστική προστασία της «Ιδιωτικότητας» στο σύγχρονο περιβάλλον, καθώς λειτουργεί ως μοχλός άσκησης πίεσης για την υιοθέτηση και εφαρμογή των κατάλληλων μέτρων προς την κατεύθυνση αυτή. Τέλος, με γνώμονα τους ανωτέρω ορισμούς και το ισχύον νομικό πλαίσιο, εξετάζονται ενδελεχώς τα ζητήματα της προστασίας της «Ιδιωτικότητας» στα συστήματα «Big Data».

3.2.1 ΟΡΙΣΜΟΣ «ΙΔΙΩΤΙΚΟΤΗΤΑΣ» (PRIVACY)

Είναι γεγονός ότι για την έννοια της «Ιδιωτικότητας» (Privacy) έχουν δοθεί διάφοροι ορισμοί, οι οποίοι κάθε φορά εστιάζουν σε διαφορετική πτυχή της, καθώς η ερμηνεία της έννοιας αυτής είναι άμεσα συνυφασμένη με την εκάστοτε κουλτούρα, αντίληψη και γενική αποδοχή της από το ευρύτερο κοινωνικό σύνολο. Μάλιστα, η ερμηνεία της πολλές φορές διαφοροποιείται σημαντικά, ακόμα και μεταξύ των ατόμων της ίδιας κοινωνίας, εξαιτίας της διαφορετικής αντίληψης και των εμπειριών που μπορεί να έχει ο καθένας εξ' αυτών. Επίσης η έννοια αυτή είναι πολυδιάστατη, καθώς καλύπτει ένα ευρύ φάσμα, με άμεση συνέπεια η ερμηνεία της να διαφοροποιείται ανάλογα με την εκάστοτε περίπτωση.

Έτσι η έννοια της «Ιδιωτικότητας», ανάλογα με το που αναφέρεται, μπορεί ενδεικτικά να διαχωριστεί στις κάτωθι εκφάνσεις [60], [175]:

- ***Ιδιωτικότητα Πληροφοριών (Informational Privacy)***: η οποία αφορά στον έλεγχο του αν και πώς τα προσωπικά δεδομένα ενός προσώπου μπορούν να

συγκεντρωθούν, να αποθηκευτούν, να υποστούν επεξεργασία ή να διαδοθούν επιλεκτικά σε τρίτα άτομα.

- ***Εδαφική Ιδιωτικότητα (Μη Προσπελασιμότητα του Χώρου) (Territorial Privacy)***: η οποία αφορά στην προστασία της στενής φυσικής περιοχής που περιβάλλει ένα πρόσωπο, δηλαδή τα οικιακά και άλλα περιβάλλοντα όπως ο εργασιακός ή ο δημόσιος χώρος του ατόμου.
- ***Σωματική Ιδιωτικότητα (Bodily Privacy)***: η οποία αφορά στην προστασία ενός προσώπου από αδικαιολόγητη παρέμβαση όπως ο σωματικός έλεγχος, η υποχρεωτική υποβολή σε εξέταση/επέμβαση, η δοκιμή φαρμάκων και γενικότερα από οποιαδήποτε ενέργεια στοχεύει εναντίον της σωματικής υπόστασης του ατόμου, χωρίς την πρότερη άδειά του.
- ***Ιδιωτικότητα Επικοινωνίας (Communication Privacy)***: η οποία αφορά στην προστασία του θεμελιώδους δικαιώματος του ατόμου στην ελεύθερη επικοινωνία απαλλαγμένη από μη εξουσιοδοτημένη παρακολούθηση.

Αντίστοιχα, το [61] κατηγοριοποιεί όλους τους σύγχρονους ορισμούς και προσεγγίσεις για την έννοια της «Ιδιωτικότητας» στα παρακάτω θέματα:

- ***Στο Δικαίωμα να Παραμένεις Μόνος κατά Βούληση.***
- ***Στο Δικαίωμα να Περιορίζεις την Πρόσβαση των άλλων στα Προσωπικά σου Δεδομένα.***
- ***Στο Δικαίωμα του Απορρήτου και του να Αποκρύπτεις από τους άλλους τα Προσωπικά σου Δεδομένα.***
- ***Στο Δικαίωμα να Ελέγχεις το πώς χρησιμοποιούν οι άλλοι τα Προσωπικά σου Δεδομένα.***
- ***Στο Δικαίωμα να Ελέγχεις την Ιδιωτική σου Ζωή.***
- ***Στο Δικαίωμα του Σεβασμού της Προσωπικότητας και της Αυτονομίας.***
- ***Στο Δικαίωμα του Αυτοπροσδιορισμού, της Αυτοδιάθεσης και της Δυνατότητας για Προσωπική Εξέλιξη.***
- ***Στο Δικαίωμα της Προστασίας των Ενεργειών και της Ερωτικής Ζωής.***

Το σύγχρονο ηλεκτρονικό περιβάλλον όμως εισάγει νέες δυνατότητες και διαφοροποιεί σημαντικά την αντίληψη περί «Ιδιωτικής Ζωής», ακριβώς λόγω των ιδιαίτερων χαρακτηριστικών του. Επομένως η έννοια αυτή θα πρέπει να επανεξεταστεί και να επαναπροσδιοριστεί υπό το πρίσμα των νέων αυτών δεδομένων. Ωστόσο, αν και οι παραπάνω προσεγγίσεις αναφέρονται γενικά στην

έννοια της «Ιδιωτικότητας», εντούτοις κάποια στοιχεία τους μπορούν κάλλιστα να υιοθετηθούν και να χρησιμοποιηθούν για την ερμηνεία της έννοιας της «Ιδιωτικότητας» στο νέο αυτό περιβάλλον. Παράλληλα, προς αυτή την κατεύθυνση συμβάλλουν και οι παρακάτω δύο διαφοροποιημένες προσεγγίσεις της έννοιας της «Ιδιωτικότητας».

❖ **Προστασία της Ιδιωτικής Ζωής**:

Σύμφωνα με τους [62], ως **«Ιδιωτικότητα»** (Privacy) ορίζεται το δικαίωμα του ατόμου να παραμένει μόνο του, όταν το επιθυμεί. Υπό αυτή την άποψη, η προστασία της ιδιωτικής ζωής γίνεται κατανοητή ως ο συνδυασμός της «μοναξιάς» και της «μη-εισβολής». Επομένως αποκτά δύο διαστάσεις. Η μια αναφέρεται στο απόρρητο της σκέψης, της ιδιοκτησίας και των ενεργειών του ατόμου και η άλλη σχετίζεται με τα δεδομένα άλλων προσώπων και πως αυτά τον επηρεάζουν. Στην καθημερινή ζωή, αυτό το είδος της ιδιωτικής ζωής είναι σεβαστό και προστατεύεται μέσα από κοινωνικούς κανόνες. Καθώς παρόμοια διάσταση αποκτά η «Ιδιωτικότητα» και στον ηλεκτρονικό κόσμο, είναι σαφές, ότι πρέπει να οριστούν και σε αυτό τον κόσμο παρόμοιοι κανόνες. Επομένως, η προσέγγιση αυτή αφορά κυρίως την κοινωνική της διάσταση.

❖ **Προστασία των Προσωπικών Δεδομένων**:

Σύμφωνα με τον [63], την οποία πολλοί επιστήμονες της πληροφορικής την αναφέρουν και ως «Information Privacy», ορίζεται ως **«Ιδιωτικότητα»** το δικαίωμα του ατόμου να καθορίζει ποιες προσωπικές του πληροφορίες είναι διαθέσιμες, σε ποια άτομα και σε ποιο βαθμό. Έτσι, ο ορισμός αυτός εστιάζει στον απόλυτο έλεγχο που έχει το άτομο επί των προσωπικών του δεδομένων, των συνομιλιών του και των ενεργειών του. Συνεπώς, ο ίδιος ο κάτοχος των δεδομένων επιλέγει τι και σε ποιόν θα αποκαλύψει και είναι ο αποκλειστικός υπεύθυνος για τις πράξεις του. Επομένως, η προσέγγιση αυτή επικεντρώνεται περισσότερο στα προσωπικά δεδομένα ως ιδιοκτησία και ως εκ τούτου αποκτά τεχνικό προσανατολισμό.

Αξίζει να παρατηρηθεί ότι οι προσεγγίσεις αυτές αλληλοσυμπληρώνονται. Συνεπώς, αν λάβουμε υπόψη τα όσα αναφέραμε στην παρούσα ενότητα, είναι εφικτό να προσεγγίσουμε, σε μεγάλο και ικανοποιητικό βαθμό, την έννοια της **«Ιδιωτικότητας στο Ηλεκτρονικό Περιβάλλον»**. Έτσι αυτή μπορεί να ταυτιστεί, αφενός με την προστασία της ιδιωτικής ζωής στο ηλεκτρονικό περιβάλλον, ήτοι την

προστασία των διαφόρων ηλεκτρονικών ταυτοτήτων του ατόμου, των ηλεκτρονικών ενεργειών του, της επικοινωνίας του, των πνευματικών του δικαιωμάτων και των ηλεκτρονικών συσκευών, μέσω των οποίων εισέρχεται στο περιβάλλον αυτό, και αφετέρου με την προστασία των προσωπικών του δεδομένων, ευαίσθητων και μη [183].

Επιπρόσθετα, οι **απαιτήσεις** για την προστασία της «Ιδιωτικότητας» στα διάφορα Πληροφοριακά Συστήματα, όπως καθορίστηκαν και χρησιμοποιούνται ευρέως από την επιστημονική κοινότητα και οι οποίες προφανώς ισχύουν και για τα συστήματα «Big Data», είναι οι παρακάτω:

✓ ***Εμπιστευτικότητα (Confidentiality)***: η οποία αφορά στην προστασία από την αποκάλυψη των δεδομένων σε μη εξουσιοδοτημένες οντότητες.

✓ ***Ακεραιότητα (Integrity)***: η οποία αφορά στην προστασία των δεδομένων από μη εξουσιοδοτημένη εισαγωγή, τροποποίηση ή διαγραφή.

✓ ***Διαθεσιμότητα (Availability)***: η οποία αφορά στην προστασία από τη μη-διαθεσιμότητα των δεδομένων στον κάτοχό τους, όταν το επιθυμεί.

✓ ***Αυθεντικότητα (Authenticity)***: η οποία αφορά στη διασφάλιση της ταυτότητας της κάθε εμπλεκόμενης οντότητας σε οποιαδήποτε ενέργεια.

✓ ***Μη Αποποίηση (Non Repudiation)***: η οποία αφορά στην προστασία από πιθανή άρνηση μια οντότητας για την μη πραγματοποίηση κάποιας συγκεκριμένης δραστηριότητας, στην οποία όμως συμμετείχε.

Εκτός όμως από τις παραπάνω γενικές απαιτήσεις, που πρέπει να ικανοποιεί ένα Πληροφοριακό Σύστημα, καθορίστηκαν και επιπλέον τεχνικές απαιτήσεις για την διασφάλιση της «Ιδιωτικότητας» [64], [65], [66], [67], [68]. Αυτές είναι οι εξής:

✓ ***Αυθεντικοποίηση (Authentication)***: η διαδικασία μέσω της οποίας επιβεβαιώνεται η ταυτότητα μιας οντότητας. Αποτελεί κυρίως απαίτηση ασφάλειας παρά ιδιωτικότητας ενός Πληροφοριακού Συστήματος, ωστόσο έχει σημαντική συνεισφορά και στην ικανοποίηση των απαιτήσεων ιδιωτικότητας.

✓ ***Εξουσιοδότηση (Authorization)***: η διαδικασία μέσω της οποίας μία οντότητα αποκτά δικαίωμα πρόσβασης σε μια μεμονωμένη υπηρεσία, ή σε συγκεκριμένες υπηρεσίες ενός πληροφοριακού συστήματος.

✓ ***Αναγνώριση (Identification)***: η διαδικασία μέσω της οποίας ελέγχεται αν η υπηρεσία ή τα δεδομένα που ζητούνται απαιτούν αυθεντικοποίηση και στη συνέχεια εξουσιοδότηση, ή όχι.

✓ ***Προστασία Δεδομένων (Data Protection)***: η διαδικασία μέσω της οποίας διασφαλίζονται, σύμφωνα με την «Ευρωπαϊκή Οδηγία 1995/46/ΕΚ», οι κάτωθι αρχές:

- Αρχή της Νομιμότητας και της Δικαιοσύνης.
- Αρχή του καθορισμού του Σκοπού της συλλογής των δεδομένων και της επεξεργασίας αυτών για το Σκοπό που συλλέχθηκαν.
- Αρχή της Αναγκαιότητας της συλλογής και επεξεργασίας των δεδομένων.
- Αρχή της παροχής Πληροφόρησης, Ενημέρωσης και Πρόσβασης στους κατόχους των δεδομένων.
- Αρχή της Ασφάλειας και της Ακεραιότητας.
- Αρχή της Εποπτείας και της Επικύρωσης.

✓ ***Ανωνυμία (Anonymity)***: η διαδικασία μέσω της οποίας διασφαλίζεται ότι μία οντότητα μπορεί να χρησιμοποιήσει μια υπηρεσία ή να επικοινωνήσει με μια άλλη οντότητα, χωρίς να αποκαλύψει την ταυτότητά της.

✓ ***Ψευδωνυμία (Pseudonymity)***: η διαδικασία μέσω της οποίας προστατεύεται η αναγνώριση (Identification) μιας οντότητας από μη εξουσιοδοτημένες τρίτες οντότητες.

✓ **Μη-*Συνδεσιμότητα (Unlinkability)***: η διαδικασία μέσω της οποίας προστατεύεται η ιδιωτικότητα μιας οντότητας από πιθανούς επιτιθέμενους, απαγορεύοντας στους δεύτερους να συνδέσουν τμήματα σχετικών πληροφοριών μεταξύ τους, οδηγώντας έτσι στην αποκάλυψη της ταυτότητάς της.

✓ **Μη-*Παρατηρησιμότητα (Observability)***: η διαδικασία μέσω της οποία προστατεύεται η ιδιωτικότητα μιας οντότητας από πιθανούς επιτιθέμενους, απαγορεύοντας στους δεύτερους να παρατηρήσουν ή να εντοπίσουν ίχνη της πρώτης.

3.2.2 ΟΡΙΣΜΟΣ «ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΕΚ ΣΧΕΔΙΑΣΜΟΥ» (PRIVACY BY DESIGN)

Η έννοια της «Ιδιωτικότητας εκ Σχεδιασμού» (Privacy By Design), η οποία επινοήθηκε και παρουσιάστηκε ευρέως το 1990 από την Ann Cavoukian [69], αφορά την άμεση ενσωμάτωση των κατάλληλων μέτρων προστασίας και των «Τεχνολογιών Ενίσχυσης της Ιδιωτικότητας» (Privacy Enhancing Technologies - PETs) στο

Πληροφοριακό Σύστημα από την φάση κιόλας του σχεδιασμού του. Μάλιστα, στην περιγραφή της για το πώς είναι εφικτό να επιτευχθεί «Privacy by Design», έθεσε επτά κατευθυντήριες αρχές. Αυτές είναι η αρχή της Πρόληψης (Proactive), της Προεπιλογής (By Default), της Ενσωμάτωσης (Embedded), του Θετικού Αποτελέσματος (Positive Sum), της Προστασίας Ολόκληρου του Κύκλου Ζωής (Lifecycle Protection), της Διαφάνειας (Transparency) και του Σεβασμού προς τους Χρήστες (Respect for User) [70]. Επίσης, σύμφωνα με τον [71], το «Privacy by Design» είναι μια μηχανική και στρατηγική προσέγγιση διαχείρισης, η οποία δεσμεύεται με επιλεκτικό και βιώσιμο τρόπο να ελαχιστοποιήσει τους κινδύνους που στοχεύουν στην παραβίαση της «Ιδιωτικότητας» στα πληροφοριακά συστήματα, μέσω της χρήσης προληπτικών τεχνικών και μηχανισμών διαχειριστικού ελέγχου.

Ωστόσο, με την πάροδο του χρόνου η έννοια αυτή απέκτησε πολυδιάστατη σημασία. Έτσι στα διάφορα νομικά έγγραφα περιγράφεται με πολύ γενικούς όρους ως μια γενική αρχή για την προστασία της «Ιδιωτικότητας» στα Πληροφοριακά Συστήματα. Ενώ αντιθέτως η επιστημονική κοινότητα στον τομέα της πληροφορικής και οι μηχανικοί των ηλεκτρονικών υπολογιστών εξισώνουν την έννοια αυτή με την χρήση συγκεκριμένων «Τεχνολογιών Ενίσχυσης της Ιδιωτικότητας» (Privacy Enhancing Technologies - PETs). Πέρα από τις ανωτέρω προσεγγίσεις όμως, η έννοια του «Privacy by Design» δεν είναι ούτε μια συλλογή από γενικές αρχές, ούτε μπορεί να περιοριστεί μόνο στις «Τεχνολογίες Ενίσχυσης της Ιδιωτικότητας». Στην πραγματικότητα πρόκειται για μια διαδικασία που περιλαμβάνει διάφορα τεχνολογικά και οργανωτικά μέτρα, τα οποία θα ενσωματώσουν και θα εφαρμόσουν τις αρχές για την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων ήδη από την φάση του σχεδιασμού του πληροφοριακού συστήματος.

Συνεπώς, η έννοια του «Privacy by Design» είναι μια προσέγγιση, η οποία λαμβάνει υπόψη την προστασία της «Ιδιωτικότητας» καθ' όλη την διαδικασία του σχεδιασμού και της ανάπτυξης ενός πληροφοριακού συστήματος, αλλά και σε κάθε φάση της ξεχωριστά. Έτσι, σύμφωνα με την προσέγγιση αυτή, η προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων είναι ενσωματωμένη στο πληροφοριακό σύστημα καθ' όλη την διαδικασία του αρχικού σχεδιασμού, της ανάπτυξης, της λειτουργίας και της τελικής απόρριψής του [176].

3.3 ΡΥΘΜΙΣΤΙΚΟ ΚΑΙ ΚΑΝΟΝΙΣΤΙΚΟ ΠΛΑΙΣΙΟ ΓΙΑ ΤΗΝ ΠΡΟΣΤΑΣΙΑ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ (PRIVACY LEGISLATION)

Εν συνεχεία παρουσιάζεται και εξετάζεται ενδελεχώς το υφιστάμενο ρυθμιστικό και κανονιστικό πλαίσιο για την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων, τόσο στην Ευρωπαϊκή Ένωση, όσο και στην Ελλάδα. Αυτό γίνεται, προκειμένου να διαπιστωθεί κατά πόσο η κείμενη νομοθεσία επιβάλει την υιοθέτηση των κατάλληλων μέτρων για την προστασία της «Ιδιωτικότητας» στα διάφορα πληροφοριακά συστήματα, ελέγχει την υλοποίηση των μέτρων αυτών ανά περίπτωση, επιλαμβάνεται των τυχόν διενέξεων που θα προκύψουν μεταξύ των ενδιαφερόμενων μερών και τέλος προστατεύει, από νομικής πλευράς, το αναφαίρετο δικαίωμα του ατόμου για την προστασία της ιδιωτικής του ζωής και των προσωπικών του δεδομένων.

3.3.1 ΝΟΜΙΚΟ ΠΛΑΙΣΙΟ ΣΤΗΝ ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ

Συγκεκριμένα, το ισχύον νομικό πλαίσιο της Ευρωπαϊκής Ένωσης για την αντιμετώπιση των ζητημάτων προστασίας της «Ιδιωτικότητας» στα διάφορα Πληροφοριακά Συστήματα και ως εκ τούτου και στα συστήματα «Big Data» συνθέτεται από την «Οδηγία 95/46/ΕΚ» (Directive 95/46/EC) [72], [177] (Περί Προστασίας των Φυσικών Προσώπων Έναντι της Επεξεργασίας Δεδομένων Προσωπικού Χαρακτήρα και της Ελεύθερης Κυκλοφορίας των Δεδομένων Αυτών), η οποία αντικαταστάθηκε και επεκτάθηκε από τον «Κανονισμό (ΕΕ) 2016/679» (Regulation (EU) 2016/679) [73], [178], και από την «Οδηγία 2002/58/ΕΚ» (Directive 2002/58/EC) [74], [179] (Περί της Επεξεργασίας των Δεδομένων Προσωπικού Χαρακτήρα και της Προστασίας της Ιδιωτικής Ζωής στον Τομέα των Ηλεκτρονικών Επικοινωνιών), η οποία τροποποιήθηκε από την «Οδηγία 2009/136/ΕΚ» (Directive 2009/136/EC) [75], [180].

Οι προϋποθέσεις και το πεδίο εφαρμογής του ανωτέρου νομικού πλαισίου ορίζονται στο Άρθρο 4 και Άρθρο 3 αντίστοιχα της «Οδηγίας 95/46/ΕΚ» και στο Άρθρο 3 και Άρθρο 2 αντίστοιχα του «Κανονισμού (ΕΕ) 2016/679». Η νομοθεσία αυτή λοιπόν εφαρμόζεται σε περιπτώσεις «Επεξεργασίας Δεδομένων Προσωπικού Χαρακτήρα», όπως αυτή ορίζεται στο Άρθρο 2 της «Οδηγίας 95/46/ΕΚ» και στο Άρθρο 4 του «Κανονισμού (ΕΕ) 2016/679», και συγκεκριμένα αφορά την

επεξεργασία οποιασδήποτε πληροφορίας προσδιορισμένης ή προσδιορίσιμης ενός φυσικού προσώπου [76], [181]. Όσον αφορά τα συστήματα «Big Data», η προσοχή εστιάζεται περισσότερο στην έμμεση αναγνώριση. Ειδικότερα στο [77], [182] επισημαίνονται οι ακόλουθες περιπτώσεις:

- Η δυνατότητα απομόνωσης ορισμένων ή και όλων των αρχείων που προσδιορίζουν ένα άτομο σε ένα σύνολο δεδομένων (Singling out of a Record).
- Η διασύνδεση τουλάχιστον δυο αρχείων που προσδιορίζουν το ίδιο άτομο, είτε στην ίδια βάση δεδομένων, είτε σε διαφορετικές (Linkability of two Records).
- Η δυνατότητα να συναχθεί, με πολύ καλή πιθανότητα, η τιμή ενός χαρακτηριστικού κάποιου συγκεκριμένου ατόμου μέσα από το πλήθος των τιμών ενός συνόλου χαρακτηριστικών (Inference of Attributes).

Επομένως, γίνεται εύκολα αντιληπτό ότι τα άτομα και τα προσωπικά τους δεδομένα μπορεί να αναγνωριστούν μέσα από ένα πλήθος ενεργειών, που κυμαίνονται από το «Reverse Engineering» των κακών ψευδωνύμων, μέχρι την αναγνώριση των μοναδικών μοντέλων κινητικότητας (Mobility Patterns) ή τον συνδυασμό χαρακτηριστικών, προτιμήσεων, συνηθειών, ακόμη και του τρόπου γραφής.

Επίσης η «Οδηγία 95/46/ΕΚ» και συνεπώς και ο «Κανονισμός (ΕΕ) 2016/679» αναφέρονται στις περιπτώσεις εκείνες, κατά τις οποίες η επεξεργασία των δεδομένων στο περιβάλλον των «Big Data» μπορεί να μην υπόκειται στο ευρωπαϊκό νομικό πλαίσιο. Απαραίτητη προϋπόθεση για αυτό είναι τα δεδομένα να έχουν απογυμνωθεί από επαρκή στοιχεία, έτσι ώστε το «Υποκείμενο των Δεδομένων» (Data Subject) να μην μπορεί πλέον να προσδιοριστεί. Αυτό επιτυγχάνεται μέσω της τεχνικής της «Ανωνυμοποίησης» (Anonymization), κατά την οποία τα προσωπικά δεδομένα υποβάλλονται σε τέτοια επεξεργασία, χρησιμοποιώντας όλα τα εύλογα δυνατά μέσα, έτσι ώστε να μην καθίσταται πλέον εφικτό να ταυτοποιηθεί κάποιο φυσικό πρόσωπο. Ένα σημαντικό στοιχείο, που χρήζει επισήμανσης, είναι ότι η οδηγία δεν επιβάλλει τον τρόπο που θα πραγματοποιηθεί η διαδικασία της «Ανωνυμοποίησης». Απλά εστιάζει στο αποτέλεσμα της διαδικασίας αυτής. Συνεπώς, τα δεδομένα θα πρέπει να καταστούν σε τέτοια μορφή που να μην επιτρέπουν την άμεση ή/και την έμμεση ταυτοποίηση ενός φυσικού προσώπου, χρησιμοποιώντας όλες τις δυνατές και εφικτές λύσεις που υπάρχουν για τον σκοπό αυτό.

Ζωτική σημασία στην εφαρμογή της ευρωπαϊκής νομοθεσίας, για την προστασία των προσωπικών δεδομένων και στο περιβάλλον των «Big Data», αποκτά η έννοια

του «Υπεύθυνου Επεξεργασίας» (Data Controller). Σύμφωνα με το Άρθρο 2(δ) της «Οδηγίας 95/46/ΕΚ», ο «Υπεύθυνος Επεξεργασίας» διαδραματίζει ένα πολύ σημαντικό ρόλο, καθώς καθορίζει τους στόχους και τα μέσα της επεξεργασίας των προσωπικών δεδομένων και ως εκ τούτου έχει και συγκεκριμένες υποχρεώσεις. Μάλιστα στο Άρθρο 24 του «Κανονισμού (ΕΕ) 2016/679» καθορίζονται οι ευθύνες και οι υποχρεώσεις του «Υπεύθυνου Επεξεργασίας». Ενώ στο Άρθρο 26 αυτές επεκτείνονται και στην περίπτωση των από «Κοινού Υπεύθυνων Επεξεργασίας». Έτσι διαχωρίζονται και καθορίζονται επακριβώς οι ευθύνες του κάθε «Υπεύθυνου Επεξεργασίας», που συμμετέχει στην δημιουργία και λειτουργία ενός πολύπλοκου συστήματος «Big Data». Επιπρόσθετα, καθορίζονται στο Τμήμα 1 του Κεφαλαίου IV τα καθήκοντα του «Εκτελούντα την Επεξεργασία» (Processor) και στο Τμήμα 4 τα καθήκοντα του «Υπεύθυνου Προστασίας των Δεδομένων» (Data Protection Officer). Επιπλέον, στο Τμήμα 2 ορίζονται τα τεχνικά και τα οργανωτικά μέτρα για την ασφάλεια της διαδικασίας επεξεργασίας των προσωπικών δεδομένων και ταυτόχρονα προβλέπονται οι διαδικασίες σε περίπτωση παραβίασης. Στο Τμήμα 3 εισάγεται η διαδικασία της «Εκτίμηση Αντίκτυπου σχετικά με την Προστασία των Δεδομένων» (Data Protection Impact Assessment - DPIA), κατά την οποία εκτιμώνται όλες οι απειλές που σχετίζονται με την «Ιδιωτικότητα» και λαμβάνονται όλα τα αναγκαία μέτρα για τον περιορισμό τους, στοχεύοντας με αυτό τον τρόπο στην επίτευξη του «Privacy By Design». Τέλος, στο Τμήμα 5 ορίζεται η έννοια της «Πιστοποίησης» (Certification), η οποία θα αποδεικνύει την συμμόρφωση των πράξεων επεξεργασίας των «Υπεύθυνων Επεξεργασίας» και των «Εκτελούντων την Επεξεργασία» με τον «Κανονισμό (ΕΕ) 2016/679».

Επιπρόσθετα, στο Άρθρο 6 της «Οδηγίας 95/46/ΕΚ» όπως και στο Άρθρο 5 του «Κανονισμού (ΕΕ) 2016/679» ορίζονται οι αρχές για την επεξεργασία των προσωπικών δεδομένων, οι οποίες ισχύουν και για τα «Big Data». Κατ' αρχάς ορίζεται ότι η συλλογή και η επεξεργασία των δεδομένων θα πρέπει να γίνεται με τρόπο νόμιμο, θεμιτό και διαφανή. Η αρχή της «Θεμιτής» συλλογής δεδομένων συνεπάγεται ότι τα προσωπικά δεδομένα δεν πρέπει ποτέ να υποβάλλονται σε επεξεργασία, αν προηγουμένως το «Υποκείμενο» αυτών δεν έχει ενημερωθεί. Η αρχή του «Σκοπού» προϋποθέτει ότι η συλλογή των δεδομένων μπορεί να πραγματοποιηθεί, μόνο όταν ο σκοπός συλλογής είναι καθορισμένος, σαφής και νόμιμος. Μάλιστα, αυτός ο σκοπός θα πρέπει να έχει καθοριστεί πριν από την συλλογή και την επεξεργασία των δεδομένων αυτών και οποιοσδήποτε άλλος

σκοπός, που δεν συμβιβάζεται με τον αρχικό, θα είναι παράνομος σύμφωνα με την ευρωπαϊκή νομοθεσία. Η αρχή της «Αναλογικότητας» καθορίζει ότι τα δεδομένα που θα πρέπει να συλλεχθούν για την επίτευξη του συγκεκριμένου σκοπού, ο οποίος καθορίστηκε εξαρχής από τον «Υπεύθυνο Επεξεργασίας», θα πρέπει να είναι αυστηρά τα απολύτως κατάλληλα, συναφή και ανάλογα με τον σκοπό αυτό. Επίσης, προβλέπει ότι ο χρόνος τήρησης των επεξεργασμένων δεδομένων είναι συγκεκριμένος και μάλιστα δεν μπορεί να υπερβαίνει τον χρόνο που απαιτείται για την επίτευξη του σκοπού αυτού. Έτσι προβλέπει την διαγραφή τους με την παρέλευση του χρόνου αυτού. Τέλος, καθορίζει ότι ο «Υπεύθυνος Επεξεργασίας» είναι υπεύθυνος να τα τηρεί επικαιροποιημένα και να λαμβάνει όλα τα ενδεδειγμένα μέτρα ασφαλείας για την προστασία τους από μη εξουσιοδοτημένη ή παράνομη επεξεργασία.

Στο Άρθρο 7 της «Οδηγίας 95/46/ΕΚ» και στο Άρθρο 6 του «Κανονισμού (ΕΕ) 2016/679» προβλέπεται ότι, για να είναι νόμιμη η επεξεργασία των προσωπικών δεδομένων, θα πρέπει να πληρείται κάποια από τις απαιτήσεις που αναφέρονται σε αυτά. Στην πραγματικότητα, όσον αφορά τα συστήματα «Big Data», αυτές οι απαιτήσεις μπορεί να εστιαστούν στις ακόλουθες τρείς, ήτοι στην συγκατάθεση του «Υποκειμένου των Δεδομένων», στην εκπλήρωση των υποχρεώσεων που απορρέουν από τα συμβόλαια και στην προάσπιση του έννομου συμφέροντος. Για να είναι έγκυρη η συγκατάθεση, θα πρέπει να συντρέχουν οι παρακάτω προϋποθέσεις:

- Να έχει δοθεί ελεύθερα από το «Υποκείμενο των Δεδομένων», δηλαδή θα πρέπει να δίδεται η ελεύθερη επιλογή, είτε να αποδεχθεί, είτε να αρνηθεί, να τύχουν επεξεργασίας τα προσωπικά του δεδομένα.
- Να έχει προηγηθεί η πλήρης ενημέρωση του «Υποκειμένου», για να έχει ολοκληρωμένη γνώση πριν δώσει την συγκατάθεσή του.
- Να είναι ακριβής, δηλαδή αυτή η βούληση θα πρέπει να αφορά αποκλειστικά τον αρχικό σκοπό.
- Να είναι σαφής.

Μάλιστα, στα Άρθρα 10 και 11 της «Οδηγίας 95/46/ΕΚ» και στα Άρθρα 13 και 14 του «Κανονισμού (ΕΕ) 2016/679» καθορίζονται οι απαιτήσεις διαφάνειας που βαραίνουν τους «Υπεύθυνους Επεξεργασίας», προκειμένου να είναι έγκυρη η συγκατάθεση του «Υποκειμένου» των δεδομένων. Για να είναι θεμιτή η εκπλήρωση των υποχρεώσεων που απορρέουν από τα συμβόλαια, θα πρέπει να υπάρχει άμεση και αντικειμενική

σχέση μεταξύ των αποτελεσμάτων της επεξεργασίας και της ικανοποίησης των επιθυμιών του «Υποκειμένου των Δεδομένων. Η προάσπιση του έννομου συμφέροντος των «Υπεύθυνων Επεξεργασίας», ή «Τρίτων Μερών», ή «Μερών» στα οποία έχουν κοινοποιηθεί αυτά τα δεδομένα, προϋποθέτει ότι αυτά δεν έρχονται σε σύγκρουση με το συμφέρον και τις θεμελιώδεις ελευθερίες και δικαιώματα του «Υποκειμένου των Δεδομένων» αυτών.

Συμπληρωματικό στην ανωτέρω παράγραφο, όσον αφορά την επεξεργασία των δεδομένων στα συστήματα «Big Data», είναι και το Άρθρο 5(3) της «Οδηγίας 2002/58/ΕΚ», το οποίο εφαρμόζεται στις περιπτώσεις που οι «Υπεύθυνοι Επεξεργασίας» των συστημάτων αυτών επιθυμούν να αποκτήσουν πρόσβαση στις ήδη αποθηκευμένες πληροφορίες στις συσκευές των χρηστών τους. Σύμφωνα με την διάταξη αυτή, απαραίτητη προϋπόθεση, για να είναι νόμιμη η ενέργεια αυτή από τον «Υπεύθυνο Επεξεργασίας», είναι να έχει δοθεί προηγουμένως η συγκατάθεση του «Υποκειμένου» αυτών των δεδομένων, εκτός αν αυτές οι ενέργειες είναι απολύτως αναγκαίες για την παροχή μιας υπηρεσίας, που έχει ρητά ζητηθεί από τον ίδιο τον συνδρομητή. Μάλιστα, η απαίτηση αυτή αποκτά ιδιαίτερη σημασία για τα περιβάλλοντα «Big Data», στα οποία πολλοί ενδιαφερόμενοι μπορεί να εμπλέκονται στην παροχή μιας υπηρεσίας. Έτσι, για να αποκτήσουν πρόσβαση οι διάφοροι ενδιαφερόμενοι στα δεδομένα αυτά, που αποθηκεύονται στις συσκευές των χρηστών, προκειμένου να υποστηρίξουν την υπηρεσία τους, απαιτείται να υπάρχει προηγούμενη συγκατάθεση του «Υποκειμένου» των δεδομένων εξίσου για τον κάθε ένα από αυτούς τους ενδιαφερόμενους.

Στο Άρθρο 17 της «Οδηγίας 95/46/ΕΚ» και στο Άρθρο 32 του «Κανονισμού (ΕΕ) 2016/679» καθορίζεται ότι όλοι οι εμπλεκόμενοι στην διαδικασία επεξεργασίας των προσωπικών δεδομένων, που αξιολογούνται ως «Υπεύθυνοι Επεξεργασίας», είναι πλήρως υπεύθυνοι για την ασφάλεια της διαδικασίας αυτής και υποχρεούνται να εφαρμόσουν όλα τα κατάλληλα τεχνικά και οργανωτικά μέτρα για να το επιτύχουν, λαμβάνοντας κάθε φορά υπόψη τις τεχνολογικές εξελίξεις, αλλά και τους κινδύνους που απορρέουν από την επεξεργασία και την φύση των δεδομένων αυτών. Επίσης οι ίδιες υποχρεώσεις και ευθύνες ισχύουν και στην περίπτωση που η επεξεργασία διενεργείται από «Τρίτα Μέρη» για λογαριασμό του «Υπεύθυνου Επεξεργασίας». Αυτό έχει άμεση εφαρμογή και στα συστήματα «Big Data», στα οποία

πραγματοποιούνται πολλαπλές μορφές επεξεργασίας από διάφορους εμπλεκόμενους.

Τέλος, στην «Οδηγία 95/46/ΕΚ» θεμελιώνονται τα δικαιώματα των «Υποκειμένων» των δεδομένων, που ισχύουν και στα περιβάλλοντα «Big Data». Αυτά είναι το «Δικαίωμα Πρόσβασης» στο Άρθρο 12 και το «Δικαίωμα Εναντίωσης» στο Άρθρο 14. Ειδικότερα, το δικαίωμα πρόσβασης δίνει την δυνατότητα στο «Υποκείμενο» των δεδομένων να ενημερώνεται από τον «Υπεύθυνο Επεξεργασίας» για όλες τις λεπτομέρειες της διαδικασίας επεξεργασίας, γεγονός που ενισχύει την διαφάνεια στα περιβάλλοντα «Big Data». Ενώ το δικαίωμα εναντίωσης δίνει την δυνατότητα το «Υποκείμενο» των δεδομένων να ανακαλέσει την προηγούμενη συναίνεσή του και να αντιταχθεί στην επεξεργασία των δεδομένων του. Πέρα από τα ανωτέρω δικαιώματα, ο «Κανονισμός (ΕΕ) 2016/679» εισάγει στο Κεφάλαιο 3 και επιπλέον δικαιώματα και δίνει περισσότερη έμφαση στην διαφάνεια. Έτσι, εισάγει το «Δικαίωμα Διόρθωσης» στο Άρθρο 16, προκειμένου το «Υποκείμενο» να απαιτήσει την διόρθωση των δεδομένων του, το «Δικαίωμα Διαγραφής» στο Άρθρο 17, με το οποίο το «Υποκείμενο» απαιτεί την διαγραφή των δεδομένων του, τόσο από τον ίδιο τον «Υπεύθυνο Επεξεργασίας», όσο και από τους άλλους στους οποίους μπορεί αυτά να έχουν κοινοποιηθεί, το «Δικαίωμα του Περιορισμού της Επεξεργασίας» στο Άρθρο 18, με το οποίο το «Υποκείμενο» μπορεί να διακόψει προσωρινά την επεξεργασία των δεδομένων του, και το «Δικαίωμα της Φορητότητας» στο Άρθρο 20, με το οποίο το «Υποκείμενο» των δεδομένων μπορεί να διαβιβάσει τα προσωπικά του δεδομένα από κάποιο «Υπεύθυνο Επεξεργασίας» σε κάποιον άλλο όποτε θέλει, χωρίς ο αρχικός «Υπεύθυνος Επεξεργασίας» να μπορεί να φέρει αντίρρηση.

Από τα παραπάνω συμπεραίνουμε ότι ο «Κανονισμός (ΕΕ) 2016/679», που τέθηκε σε ισχύ πρόσφατα, βελτίωσε και επέκτεινε το νομικό πλαίσιο για την επεξεργασία των προσωπικών δεδομένων, αλλά και την προστασία του «Υποκειμένου» τους, λαμβάνοντας υπόψη το σύγχρονο τεχνολογικό περιβάλλον που έχει διαμορφωθεί από σύγχρονες τεχνολογίες, όπως τα «Big Data». Επίσης, αξίζει να επισημανθεί ότι ο ίδιος ο Κανονισμός προβλέπει την τακτική επισκόπηση και αναθεώρηση του, προκειμένου να ενσωματώνονται σε αυτόν, τόσο οι εξελίξεις στην τεχνολογία των πληροφοριών, όσο και η πρόοδος που έχει συντελεστεί στην κοινωνία της πληροφορίας.

3.3.2 ΝΟΜΙΚΟ ΠΛΑΙΣΙΟ ΣΤΗΝ ΕΛΛΑΔΑ

Αντίστοιχα στην Ελλάδα, το ισχύον νομικό πλαίσιο για την προστασία της «Ιδιωτικότητας» στα διάφορα Πληροφοριακά Συστήματα συνθέτεται από τον «Νόμο 2472/1997» [78] (Περί Προστασίας του Ατόμου από την Επεξεργασία Δεδομένων Προσωπικού Χαρακτήρα), σε εφαρμογή της κοινοτικής «Οδηγίας 95/46/ΕΚ», και από τον «Νόμο 3471/2006» [79] (Περί Προστασίας των Δεδομένων Προσωπικού Χαρακτήρα και της Ιδιωτικής Ζωής στον Τομέα των Ηλεκτρονικών Επικοινωνιών), σε εφαρμογή της κοινοτικής «Οδηγίας 2002/58/ΕΚ».

Αναλυτικότερα, στο Άρθρο 3 καθορίζεται το πεδίο εφαρμογής του «Νόμου 2472/1997». Έτσι, αυτός εφαρμόζεται κατά την «Επεξεργασία των Δεδομένων Προσωπικού Χαρακτήρα», όπως αυτή ορίζεται στο Άρθρο 2δ. Μάλιστα, αυτή επεκτείνεται, σύμφωνα με τα Άρθρα 2α και 2γ, στην επεξεργασία οποιασδήποτε πληροφορίας μπορεί να χρησιμοποιηθεί για την άμεση ή έμμεση εξακρίβωση της ταυτότητας ενός φυσικού προσώπου. Επιπλέον, στο Άρθρο 2στ ορίζεται ως «Διασύνδεση» αρχείων, η συσχέτιση των δεδομένων ενός αρχείου με δεδομένα από αρχεία που τηρούνται από διαφορετικούς «Υπεύθυνους Επεξεργασίας», ή ακόμα και με αρχεία που τηρούνται από τον ίδιο τον «Υπεύθυνο Επεξεργασίας», για την επίτευξη σκοπού διαφορετικού από αυτούς για τους οποίους είχαν αρχικά δημιουργηθεί τα αρχεία αυτά. Συνεπώς, συμπεριλαμβάνονται στον «Νόμο 2472/1997» οι περιπτώσεις που αναφέρθηκαν παραπάνω στην ενότητα «3.3.1», δηλαδή αυτές του έμμεσου προσδιορισμού ενός φυσικού προσώπου, και οι οποίες σχετίζονται άμεσα με τον τρόπο επεξεργασίας των δεδομένων στα συστήματα «Big Data».

Επίσης, ο «Νόμος 2472/1997» ορίζει στο Άρθρο 2ζ την έννοια του «Υπεύθυνου Επεξεργασίας», ο οποίος είναι ο έναντι του νόμου υπεύθυνος για τον καθορισμό του σκοπού και του τρόπου επεξεργασίας των προσωπικών δεδομένων, και στο Άρθρο 2η την έννοια του «Εκτελούντα την Επεξεργασία». Μάλιστα στα Άρθρα 6, 7 και 7Α καθορίζονται λεπτομερώς οι ευθύνες και οι υποχρεώσεις του «Υπεύθυνου Επεξεργασίας», ανάλογα με την φύση των δεδομένων, που πρόκειται να υποβληθούν σε επεξεργασία, και την εκάστοτε περίπτωση, που προβλέπεται στα άρθρα αυτά. Επιπρόσθετα, στο Άρθρο 8 προβλέπονται επακριβώς οι ευθύνες και οι

υποχρεώσεις των από «Κοινού Υπεύθυνων Επεξεργασίας» ή του «Υπεύθυνου Επεξεργασίας» για την ειδική περίπτωση της «Διασύνδεσης» δυο ή περισσότερων αρχείων, το οποίο τυγχάνει εφαρμογής και στα πολύπλοκα συστήματα των «Big Data».

Επιπλέον, στο Άρθρο 4 του «Νόμου 2472/1997» ορίζονται οι αρχές για την επεξεργασία των προσωπικών δεδομένων, οι οποίες ισχύουν και για τα «Big Data» και οι οποίες ταυτίζονται με αυτές που αναγράφονται στο Άρθρο 6 της «Οδηγίας 95/46/ΕΚ». Κατ' αρχάς προβλέπεται ότι η συλλογή και η επεξεργασία των δεδομένων θα πρέπει να γίνεται με τρόπο νόμιμο και θεμιτό. Εν συνεχεία ορίζονται και θεμελιώνονται οι αρχές του «Σκοπού», της «Αναλογικότητας», του «Περιορισμένου Χρόνου Τήρησης» και της «Ορθότητας» των δεδομένων. Μάλιστα, στην παράγραφο 2 του Άρθρου 4 προβλέπεται σαφώς ότι η τήρηση των αρχών αυτών, κατά την επεξεργασία των προσωπικών δεδομένων, βαρύνει τον «Υπεύθυνο Επεξεργασίας».

Επιπρόσθετα, στο Άρθρο 5 του «Νόμου 2472/1997» προβλέπεται ότι, για να είναι νόμιμη η επεξεργασία των προσωπικών δεδομένων, απαραίτητη προϋπόθεση είναι η συγκατάθεση του «Υποκειμένου των Δεδομένων», πλην των περιπτώσεων της παραγράφου 2 του Άρθρου 5 για τις οποίες αυτό δεν ισχύει. Μάλιστα, στο Άρθρο 2ια καθορίζεται ότι, για να είναι έγκυρη η συγκατάθεση, θα πρέπει να ισχύουν τα παρακάτω στο σύνολό τους:

- Να έχει δοθεί ελεύθερα από το «Υποκείμενο των Δεδομένων», δηλαδή θα πρέπει να δίδεται η ελεύθερη επιλογή, είτε να αποδεχθεί, είτε να αρνηθεί, να τύχουν επεξεργασίας τα προσωπικά του δεδομένα.
- Να έχει προηγηθεί η πλήρης ενημέρωση του «Υποκειμένου», για να έχει ολοκληρωμένη γνώση πριν δώσει την συγκατάθεσή του. Μάλιστα καθορίζονται και τα απαιτούμενα εκείνα στοιχεία για τα οποία θα πρέπει να έχει ενημερωθεί το «Υποκείμενο» των δεδομένων από τον «Υπεύθυνο Επεξεργασίας», πριν δώσει την συγκατάθεσή του.
- Να είναι ακριβής, δηλαδή αυτή η βούληση θα πρέπει να αναφέρεται αποκλειστικά στον αρχικό σκοπό.
- Να είναι ρητή και σαφής.

Συμπληρωματικό στην προηγούμενη παράγραφο είναι το Άρθρο 4(5) του «Νόμου 3471/2006» κατ' αντιστοιχία του Άρθρου 5(3) της «Οδηγίας 2002/58/ΕΚ»,

στο οποίο καθορίζεται ότι η αποθήκευση πληροφοριών ή η απόκτηση πρόσβασης στις ήδη αποθηκευμένες πληροφορίες στον τερματικό εξοπλισμό συνδρομητή ή χρήστη επιτρέπεται μόνο, αν ο χρήστης έχει δώσει προηγουμένως την ρητή συγκατάθεσή του και εφόσον έχει προηγηθεί εκτενής και σαφής ενημέρωσή του από τον «Υπεύθυνο Επεξεργασίας». Αυτό αποκτά ιδιαίτερη σημασία στο περιβάλλον των «Big Data», καθώς ρυθμίζει την περίπτωση που κάποιος «Πάροχος» θα θελήσει να χρησιμοποιήσει τα τερματικά των χρηστών της υπηρεσίας του ως «Πηγές Δεδομένων» για να αντλήσει, για περαιτέρω επεξεργασία, τα δεδομένα που αποθηκεύονται σε αυτά. Εκτός από τα προσωπικά δεδομένα, που μπορεί να αποθηκεύονται στα τερματικά, συμπεριλαμβάνονται και τα δεδομένα θέσης και κίνησης, τα οποία πολλές φορές μπορεί να τύχουν επεξεργασίας από τα συστήματα «Big Data» και να οδηγήσουν με πολύ καλή πιθανότητα στην αποκάλυψη της ταυτότητας κάποιου φυσικού προσώπου.

Στην παράγραφο 3 του Άρθρου 10 του «Νόμου 2472/1997» καθορίζεται ότι ο «Υπεύθυνος Επεξεργασίας» οφείλει να λαμβάνει όλα τα κατάλληλα τεχνικά και οργανωτικά μέτρα για την ασφάλεια των δεδομένων και την προστασία τους κατά τις διάφορες φάσεις επεξεργασίας τους. Μάλιστα, καθορίζεται ότι θα πρέπει να εξασφαλίζεται επίπεδο ασφάλειας ανάλογο των κινδύνων, που συνεπάγεται η επεξεργασία και η φύση των δεδομένων που τυγχάνουν αντικείμενο επεξεργασίας από το συγκεκριμένο πληροφοριακό σύστημα. Επίσης, οι ίδιες υποχρεώσεις και ευθύνες ισχύουν και για την περίπτωση, κατά την οποία η επεξεργασία διενεργείται από «Τρίτα Μέρη» για λογαριασμό του «Υπεύθυνου Επεξεργασίας». Αυτό έχει άμεση εφαρμογή και στο περιβάλλον των «Big Data», στο οποίο πραγματοποιούνται πολλαπλές μορφές επεξεργασίας από διάφορους εμπλεκόμενους.

Τέλος, στο Κεφάλαιο Γ του «Νόμου 2472/1997» θεμελιώνονται τα δικαιώματα του «Υποκειμένου» των δεδομένων, τα οποία είναι:

- Το «Δικαίωμα Ενημέρωσης», στο Άρθρο 11, στο οποίο προβλέπεται η ακριβής διαδικασία κατά περίπτωση, που οφείλει να ακολουθήσει ο «Υπεύθυνος Επεξεργασίας» για την ενημέρωση του «Υποκειμένου» των δεδομένων κατά την επεξεργασία τους.
- Το «Δικαίωμα Πρόσβασης», στο Άρθρο 12, στο οποίο παρέχεται η δυνατότητα στο «Υποκείμενο» των δεδομένων να ενημερώνεται από τον «Υπεύθυνο

Επεξεργασίας» για όλες τις λεπτομέρειες της διαδικασίας επεξεργασίας, γεγονός που ενισχύει την διαφάνεια.

- Το «Δικαίωμα Αντίρρησης», στο Άρθρο 13, το οποίο δίνει την δυνατότητα στο «Υποκείμενο» των δεδομένων, οποτεδήποτε, να υποβάλλει αίτημα αντίρρησης στον «Υπεύθυνο Επεξεργασίας» και να αξιώνει είτε την διόρθωση, είτε την προσωρινή μη χρησιμοποίηση, είτε την δέσμευση, είτε την μη διαβίβαση, είτε την διαγραφή των προσωπικών του δεδομένων.

Ωστόσο, η έκδοση του κοινοτικού «Κανονισμού (ΕΕ) 2016/679», ο οποίος έχει τεθεί ήδη σε ισχύ και θα αρχίσει να εφαρμόζεται από το 2018, επιβάλλει την σύμπνοια της κείμενης νομοθεσίας των κρατών μελών με αυτόν, καθώς καταργεί την «Οδηγία 95/46/ΕΚ» και καθίσταται δεσμευτικός για όλα τα κράτη μέλη. Αυτό συνεπάγεται ότι ο «Νόμος 2472/1997» θα πρέπει να εναρμονιστεί με τον κοινοτικό αυτό κανονισμό, ενσωματώνοντας τόσο τις τροποποιήσεις που έχουν υιοθετηθεί λόγω των σύγχρονων τεχνολογικών εξελίξεων, όσο και τις νέες διατάξεις που έχουν εισαχθεί για την αντιμετώπιση των νέων συνθηκών που αυτές έχουν επιβάλλει. Παρά ταύτα όμως, διαπιστώνεται ότι το ισχύον νομικό πλαίσιο στην Ελλάδα, όπως και στην Ευρωπαϊκή Ένωση, καλύπτει το ιδιαίτερο περιβάλλον των «Big Data» όσον αφορά την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων των πολιτών.

3.4 ΕΞΕΤΑΣΗ ΤΗΣ ΠΡΟΣΤΑΣΙΑΣ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ (PRIVACY) ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Τα προβλήματα που σχετίζονται με την προστασία της «Ιδιωτικότητας» στα πληροφοριακά συστήματα και ως εκ τούτου και στα συστήματα «Big Data» είναι παρεμφερή με αυτά που συναντώνται στο περιβάλλον του «Cloud». Έτσι για παράδειγμα, προβλήματα όπως που αποθηκεύονται τα δεδομένα, ποιος τα επεξεργάζεται και για ποιο σκοπό, πως είναι δυνατό να εκτελεστεί επιθεώρηση σε αυτό το πολύπλοκο πληροφοριακό σύστημα κλπ, είναι κοινά. Επιπρόσθετα, λόγω της ιδιαιτερότητας των τεχνολογιών «Big Data», ήτοι της συμμετοχής και της αλληλεπίδρασης σε αυτά πάρα πολλών διαφορετικών ενδιαφερομένων μερών, καθίσταται εξαιρετικά δύσκολο για τις «Ρυθμιστικές Αρχές» (π.χ. Α.Π.Δ.Π.Χ.), τους «Υπεύθυνους Επεξεργασίας Προσωπικών Δεδομένων», αλλά και τα ίδια τα «Υποκείμενα των Δεδομένων», να ακολουθήσουν τις ροές των δεδομένων τους και να επιβάλλουν ή να ελέγξουν τα αντίστοιχα μέτρα για την προστασία των δεδομένων

αυτών. Έτσι, ενώ οι «Αρχές» για την προστασία της «Ιδιωτικότητας» παραμένουν ίδιες και για τον τομέα των «Big Data», τα προβλήματα που προκύπτουν από την προσπάθεια διασφάλισής τους στο νέο αυτό περιβάλλον οξύνονται αναλογικά με την κλίμακα και τα χαρακτηριστικά της νέας αυτής τεχνολογίας. Επομένως, είναι σημαντικό να εξεταστούν όλες εκείνες οι τεχνικές λύσεις που θα διασφαλίζουν την προστασία της «Ιδιωτικότητας» στα περιβάλλοντα αυτά.

Κατόπιν των παραπάνω, στην παρούσα ενότητα εξετάζεται αρχικά η διαδικασία της «Εκτίμησης του Αντίκτυπου σχετικά με την Προστασία των Δεδομένων» (DPIA), που προβλέπεται στον «Ευρωπαϊκό Κανονισμό (ΕΕ) 2016/679» και η οποία αποσκοπεί στην επίτευξη του «Privacy by Design». Εν συνεχεία εξετάζονται αναλυτικά οι τεχνολογικές λύσεις της «Ανωνυμοποίησης», της «Κρυπτογραφίας», της υλοποίησης μηχανισμών «Ελέγχου Πρόσβασης», της υιοθέτησης μεθόδων «Ενημέρωσης και Διαφάνειας» και της υλοποίησης μηχανισμών «Συναίνεσης, Δήλωσης Κατοχής και Ελέγχου». Απώτερος στόχος είναι να διερευνηθεί κατά πόσο αυτές οι σύγχρονες τεχνολογίες μπορούν να χρησιμοποιηθούν και να προστατεύσουν την «Ιδιωτικότητα» στα συστήματα «Big Data» και να εξεταστεί αν απαιτείται περαιτέρω έρευνα για την προσαρμογή τους στο ιδιαίτερο αυτό περιβάλλον.

3.4.1 ΜΕΘΟΔΟΣ ΓΙΑ ΤΗΝ ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΕΠΙΠΤΩΣΕΩΝ ΕΠΙ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ (PRIVACY IMPACT ASSESSMENT)

Καθώς λοιπόν τα προβλήματα, που σχετίζονται με την προστασία της «Ιδιωτικότητας» στο περιβάλλον των «Big Data», μεγεθύνονται εξαιτίας των ιδιαίτερων χαρακτηριστικών της τεχνολογίας αυτής, αλλά και της πολυπλοκότητας των συστημάτων της, κρίνεται πιο αναγκαίο από ποτέ τα συστήματα αυτά να αναπτύσσονται έχοντας ως βασική αρχή την επίτευξη του «Privacy by Design». Παράλληλα, ως πληροφοριακά συστήματα που είναι, υπόκεινται στο ρυθμιστικό και κανονιστικό πλαίσιο της Ευρωπαϊκής Ένωσης και της Ελλάδας. Μάλιστα, στο Τμήμα 3 του Κεφαλαίου IV του «Ευρωπαϊκού Κανονισμού (ΕΕ) 2016/679» [73] εισάγεται η έννοια της «Εκτίμησης του Αντίκτυπου σχετικά με την Προστασία των Δεδομένων» (Data Protection Impact Assessment - DPIA), κατά την οποία εκτιμώνται όλες οι απειλές που σχετίζονται με την προστασία των δεδομένων προσωπικού χαρακτήρα και λαμβάνονται όλα τα αναγκαία μέτρα για τον περιορισμό τους, πετυχαίνοντας έτσι

«Privacy By Design». Συνεπώς, για τους ανωτέρω λόγους, η χρήση μεθόδων για την «Εκτίμηση του Αντίκτυπου σχετικά με την Προστασία των Δεδομένων» κατά την σχεδίαση και ανάπτυξη των συστημάτων «Big Data» είναι επιβεβλημένη, προκειμένου να υπάρχει και συμμόρφωση με την κείμενη νομοθεσία, αλλά και ουσιαστική προστασία της «Ιδιωτικότητας» κατά την λειτουργία των συστημάτων αυτών.

Συγκεκριμένα, οι μεθοδολογίες αυτές, ακολουθώντας την αρχή του «Privacy by Design», εστιάζουν στην διεξαγωγή συστηματικής ανίχνευσης και αξιολόγησης των κινδύνων που απειλούν την «Ιδιωτικότητα» του ατόμου κατά την φάση της ανάπτυξης του πληροφοριακού συστήματος, με σκοπό την πλήρη διερεύνηση των επιπτώσεών τους, τόσο στις διάφορες διεργασίες του πληροφοριακού συστήματος, όσο και στις πρακτικές που χρησιμοποιούνται για την διαχείριση των προσωπικών δεδομένων των χρηστών. Απώτερος στόχος τους είναι η αναγνώριση των τεχνικών και οργανωτικών απειλών, έτσι ώστε στη συνέχεια να επιλεγούν, να υιοθετηθούν και να υλοποιηθούν όλα τα αναγκαία και κατάλληλα μέτρα και τεχνικές πρόληψης για τον μετριασμό τους. Τυπικά, αυτές οι αξιολογήσεις ολοκληρώνονται νωρίς κατά την ανάπτυξη ενός πληροφοριακού συστήματος. Ωστόσο, παρά το γεγονός ότι πλέον αυτές οι μεθοδολογίες καθίστανται αναγκαίες, δεν υπάρχει κάποια τυποποιημένη και ευρέως αποδεκτή μεθοδολογία που να είναι εφαρμόσιμη και αποτελεσματική προς την κατεύθυνση αυτή [80], [81].

Εντούτοις, μια σημαντική προσέγγιση πραγματοποιήθηκε στα [80], [81], στα οποία παρουσιάζεται και ταυτόχρονα αξιολογείται μια ολοκληρωμένη μεθοδολογία για την «Εκτίμηση του Αντίκτυπου ως προς την Προστασία της Ιδιωτικής Ζωής» (Privacy Impact Assessment - PIA) ως μετεξέλιξη των ήδη υπαρχουσών. Ο βασικός σκοπός της είναι να χρησιμοποιείται προληπτικά κατά την φάση του σχεδιασμού ή της αναβάθμισης ενός πληροφοριακού συστήματος, προκειμένου να επιτευχθεί το «Privacy By Design». Μάλιστα, η μεθοδολογία αυτή ικανοποιεί τις απαιτήσεις που καθορίστηκαν στο [82] και οι οποίες είναι:

✓ Η διεξαγωγή συστηματικής αναγνώρισης όλων των ζητημάτων και των κινδύνων που σχετίζονται με την προστασία της «Ιδιωτικότητας» στα διάφορα πληροφοριακά συστήματα ή προγράμματα, πριν όμως αυτά τεθούν σε λειτουργία ή τροποποιηθούν.

✓ Η αξιολόγηση των επιπτώσεων σε όρια ευρύτερα από αυτά της συμμόρφωσης με την κείμενη νομοθεσία.

✓ Η διεξαγωγή της με γνώμονα την διαδικασία (Process-Oriented), παρά το εξαγόμενο αποτέλεσμα (Output-Oriented).

✓ Η διεξαγωγή της μεθοδικά και συστηματικά.

Ωστόσο, το ζήτημα που εγείρεται είναι κατά πόσο η μεθοδολογία που προτείνεται στον «Ευρωπαϊκό Κανονισμό (ΕΕ) 2016/679», η οποία αφορά την «Εκτίμηση του Αντίκτυπου σχετικά με την Προστασία των Προσωπικών Δεδομένων» (DPIA), συμπεριλαμβάνει και μπορεί να ανταποκριθεί και για την περίπτωση της «Εκτίμησης του Αντίκτυπου σχετικά με την Προστασία της Ιδιωτικής Ζωής» (PIA) και αντιστρόφως. Σύμφωνα με τον ορισμό της «Ιδιωτικότητας», ο οποίος συντίθεται από την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων, προκύπτει ότι η έννοια της προστασίας της ιδιωτικής ζωής (Privacy) αποτελεί υπερσύνολο της έννοιας της προστασίας των προσωπικών δεδομένων (Data Protection). Έτσι, προκειμένου να διερευνηθούν περαιτέρω οι ομοιότητες και οι διαφορές μεταξύ των κανόνων για την προστασία των προσωπικών δεδομένων και των κανόνων για την προστασία της ιδιωτικής ζωής, στο [81] έγινε σύγκριση μεταξύ της λίστας του [83], η οποία αποτελεί την πιο πλήρη λίστα των απειλών κατά της ιδιωτικής ζωής, και του «Ευρωπαϊκού Κανονισμού (ΕΕ) 2016/679». Μάλιστα, στο Παράρτημα Α του [81] διερευνήθηκε κατά πόσο η αυστηρή εφαρμογή του «Ευρωπαϊκού Κανονισμού (ΕΕ) 2016/679» συνεπάγεται και την επαρκή αντιμετώπιση όλων των γνωστών απειλών της ιδιωτικής ζωής που αναφέρει ο [83]. Οπότε, εάν ο «Ευρωπαϊκός Κανονισμός» είναι επαρκής, τότε η προσπάθεια για την ηλεκτρονική προστασία της ιδιωτικής ζωής θα επιτευχθεί ταυτόχρονα με την προστασία των προσωπικών δεδομένων. Η αξιολόγηση που έγινε στα [80], [81] κατέληξε στα εξής συμπεράσματα:

➢ Ότι αν εφαρμοστεί επακριβώς ο «Ευρωπαϊκός Κανονισμός», ιδίως με την υιοθέτηση και εφαρμογή προτύπων υψηλής ποιότητας για τον χειρισμό και την ασφάλεια των προσωπικών δεδομένων, τότε όλες οι απειλές που έχουν αναγνωριστεί για την προστασία της ιδιωτικής ζωής [83] μπορούν να αντιμετωπιστούν αποτελεσματικά.

➢ Ότι ο «Ευρωπαϊκός Κανονισμός» παρέχει στους ανθρώπους πρόσθετα δικαιώματα για την αυτοδιάθεση των πληροφοριών τους, τα οποία εκτείνονται πέρα από τον τομέα της προστασίας της ιδιωτικής τους ζωής.

Από τα ανωτέρω συμπεράσματα εξάγεται το ζητούμενο, ότι δηλαδή όντως η εφαρμογή του «Ευρωπαϊκού Κανονισμού» για την προστασία των προσωπικών δεδομένων υποστηρίζει παράλληλα και τον μετριασμό των πιθανών απειλών κατά της ιδιωτικής ζωής. Συνεπώς, αυτό αποδεικνύει ότι η μεθοδολογία «PIA» είναι πράγματι μια μεθοδολογία για την «Εκτίμηση του Αντίκτυπου σχετικά με την Ιδιωτική Ζωή» και όχι απλά μια μεθοδολογία που εστιάζει στην «Εκτίμηση του Αντίκτυπου για τα Προσωπικά Δεδομένα», την οποία επιβάλλει η ανάγκη συμμόρφωσης προς την νομοθεσία, και επιπλέον ότι όντως την υπερκαλύπτει. Επομένως, μέσω της μεθοδολογίας «PIA» επιτυγχάνεται και «Privacy By Design» και παράλληλα υπάρχει συμμόρφωση με την κείμενη νομοθεσία, η οποία προβλέπει την εκτέλεση «DPIA».

Αναλυτικότερα, η μεθοδολογία «PIA» τίθεται σε εφαρμογή από τη στιγμή που ένα πληροφοριακό σύστημα θα αρχίσει, είτε να σχεδιάζεται, είτε να αναβαθμίζεται. Πρόκειται για μια επαναλαμβανόμενη διαδικασία, η οποία διεξάγεται παράλληλα με την εξέλιξη του πληροφοριακού συστήματος. Έτσι, κάθε φορά που το πληροφοριακό σύστημα αλλάζει, τότε η μεθοδολογία «PIA» ενεργοποιείται και εκτελείται. Παράλληλα, ενημερώνεται και ελέγχεται για την εγκυρότητά της η υπάρχουσα τεκμηρίωση.

Τα στάδια της μεθοδολογίας αυτής [80], [81] (Εικόνα 11) είναι τα εξής:

1. ΒΗΜΑ 1^{ο}: Χαρακτηρισμός του συστήματος.

Απαραίτητη προϋπόθεση για την επιτυχία της μεθοδολογίας αυτής είναι η ολοκληρωμένη και λεπτομερής περιγραφή και καταγραφή του συστήματος υπό το πρίσμα της προστασίας της «Ιδιωτικότητας». Συνεπώς, για να είναι πλήρης, αυτή εκτελείται υπό τις οπτικές γωνίες της δομής του συστήματος (πχ. Εφαρμογές, Υλικό, Λογισμικό, Διεπαφές κλπ.), της λειτουργίας του (πχ. Διαδικασίες, Ρόλοι κλπ.), της ροής των δεδομένων του (πχ. Διαγράμματα Ροής, Είδος Δεδομένων, Εκτελεστές κλπ.) και του φυσικού περιβάλλοντός του (πχ. Τοποθεσία, Μέτρα Φυσικής Ασφάλειας κλπ.). Το αποτέλεσμα αυτού του βήματος είναι μια λεπτομερής τεκμηρίωση, η οποία δίνει έμφαση στις ροές και στη φύση των δεδομένων που τυγχάνουν επεξεργασίας από το σύστημα.

2. ΒΗΜΑ 2^{ο}: Καθορισμός των «Στόχων Ιδιωτικότητας» (Privacy Targets).

Στην συνέχεια, με γνώμονα τα αποτελέσματα του προηγούμενου βήματος, καθορίζονται αναλυτικά όλα εκείνα τα στοιχεία του συστήματος, τα οποία σχετίζονται τόσο με τα ίδια τα δεδομένα, όσο και με τις διαδικασίες για την επεξεργασίας τους,

που κρίνονται κρίσιμα για την προστασία της «Ιδιωτικότητας» και επομένως θα πρέπει να προστατευθούν. Το αποτέλεσμα αυτού του βήματος είναι να εξαχθεί μια λεπτομερής λίστα με όλους τους «Στόχους Ιδιωτικότητας» (Privacy Targets).

3. ΒΗΜΑ 3ο: Αξιολόγηση του βαθμού προστασίας του κάθε Στόχου.

Έπειτα αυτοί οι στόχοι αξιολογούνται στα πλαίσια της λειτουργίας και του περιβάλλοντος στο οποίο εντάσσεται το εκάστοτε σύστημα, με γνώμονα τις επιπτώσεις που θα έχει η μη επίτευξή τους, αφενός στους τομείς των οικονομικών επιπτώσεων και της φήμης για τον «Υπεύθυνο Επεξεργασίας» του συστήματος και αφετέρου στους τομείς των οικονομικών επιπτώσεων, της διασφάλισης των δικαιωμάτων και της φήμης για τα ίδια τα «Υποκείμενα» των δεδομένων αυτών. Η αξιολόγηση πραγματοποιείται με ποιοτικά κριτήρια, τα οποία εκτιμούν αν η ζημιά που θα προκληθεί σε αυτούς για κάθε ένα τομέα θα είναι περιορισμένη, σημαντική ή καταστροφική και ανάλογα με το υψηλότερο αποτέλεσμα, που θα έχουμε ανά περίπτωση, καταγράφεται ο αντίστοιχος βαθμός προστασίας του, ήτοι χαμηλός, μεσαίος ή υψηλός. Αυτή η διαδικασία έχει ως αποτέλεσμα την δημιουργία ενός πίνακα, στον οποίο κατατάσσονται οι στόχοι και καθορίζεται η προτεραιότητα επίτευξής τους με βάση τον βαθμό προστασίας τους.

4. ΒΗΜΑ 4ο: Προσδιορισμός των απειλών για κάθε Στόχο.

Στην συνέχεια αναλύονται οι στόχοι για να εξεταστεί, γιατί είναι ευάλωτοι, και για κάθε έναν καταγράφονται συστηματικά οι απειλές του. Για να είναι πλήρης αυτή η διαδικασία θα πρέπει, είτε να μην υπάρχει απειλή που να μην αντιστοιχίζεται σε κάποιο στόχο, είτε να μην υπάρχει στόχος χωρίς απειλές. Παράλληλα, καθώς η πιθανότητα εκδήλωσης μιας απειλής εξαρτάται από πάρα πολλούς παράγοντες, αυτοί αξιολογούνται ενδελεχώς για να καθοριστεί η πιθανότητα εκδήλωσης της εκάστοτε απειλής. Αυτή καθορίζεται για τον τομέα της προστασίας της «Ιδιωτικότητας» ως προς το αν είναι πιθανό να υπάρχει ή όχι μια απειλή, χωρίς να εξαχθεί κάποιο ποσοτικό κριτήριο. Έτσι, η ύπαρξή της συνεπάγεται την ανάγκη αντιμετώπισή της. Όλα τα παραπάνω, ήτοι οι αιτίες της τρωτότητας, οι απειλές και η πιθανότητα εκδήλωσής τους, συνδυάζονται και αξιολογούνται με βάση την αξία που έχει η επίτευξη του εκάστοτε στόχου. Το αποτέλεσμα όλης αυτής της διαδικασίας είναι η παραγωγή μιας λίστας, στην οποία κατηγοριοποιούνται όλες οι απειλές με γνώμονα τα ανωτέρω στοιχεία.

5. ΒΗΜΑ 5ο: Καθορισμός των κατάλληλων μέτρων προστασίας.

Αυτό το βήμα είναι κρίσιμο, καθώς σε αυτό καθορίζονται όλα εκείνα τα αναγκαία μέτρα που θα συμβάλουν στην ελαχιστοποίηση, τον περιορισμό ή την εξάλειψη των

απειλών του προηγούμενου βήματος. Τα μέτρα μπορεί να είναι τεχνικά ή μη τεχνικά. Τα τεχνικά ενσωματώνονται άμεσα στο σύστημα, ενώ τα μη τεχνικά αφορούν οργανωτικές και διαχειριστικές διαδικασίες. Επίσης, αυτά διακρίνονται σε προληπτικά και κατασταλτικά. Καθώς όμως ο στόχος είναι η επίτευξη του «Privacy by Design», μεγαλύτερη βαρύτητα θα πρέπει να δίνεται στην υλοποίηση προληπτικών μέτρων. Τα μέτρα που θα υιοθετηθούν κατηγοριοποιούνται σε ικανοποιητικά, ισχυρά και πολύ ισχυρά, ανάλογα με τον βαθμό προστασίας του εκάστοτε στόχου (Βήμα 3) και την κατηγοριοποίηση των απειλών (Βήμα 4). Το αποτέλεσμα αυτού του βήματος είναι η παραγωγή μιας λίστας με μέτρα.

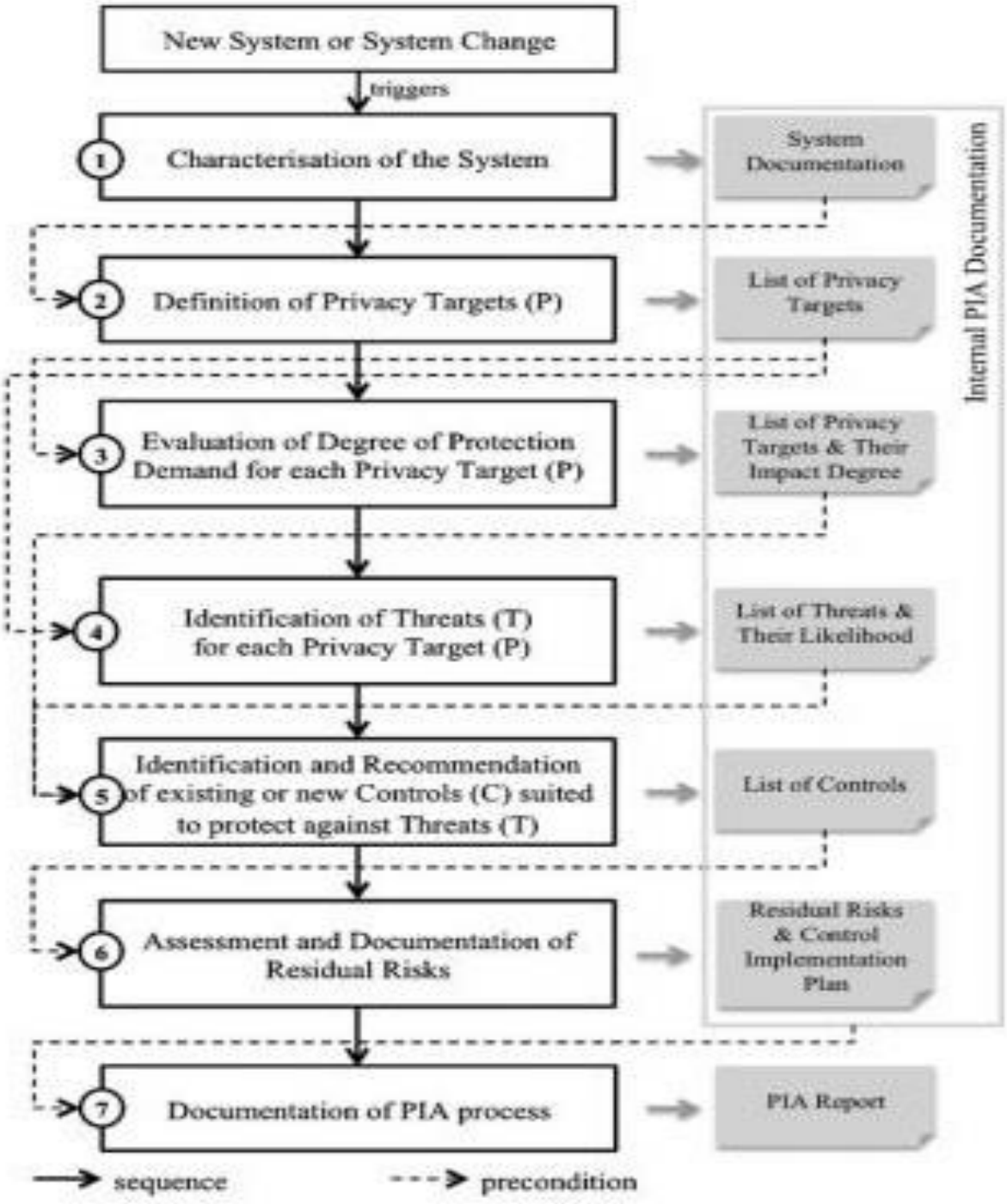
6. ΒΗΜΑ 6ο: Αξιολόγηση και τεκμηρίωση «Υπολειπόμενου Κινδύνου».

Τα μέτρα, που προέκυψαν στο προηγούμενο βήμα, αξιολογούνται από τα διάφορα ενδιαφερόμενα μέρη ως προς το αν είναι εφικτά και αποτελεσματικά και μπορεί ανάλογα, είτε να γίνουν αποδεκτά, είτε να προταθούν εναλλακτικά. Μετά τον καθορισμό των μέτρων προστασίας της «Ιδιωτικότητας» συντάσσεται το «Σχέδιο Υλοποίησής» τους (Control Implementation Plan), στο οποίο καταγράφονται πλήρως, τόσο ο «Υπολειπόμενος Κίνδυνος» (Residual Risk) και η αποδοχή του από τον «Υπεύθυνο Επεξεργασίας», γεγονός που τον καθιστά υπόλογο σε περίπτωση παραβίασης, όσο και το πλήρες χρονοδιάγραμμα για την υλοποίησή τους. Ο «Υπολειπόμενος Κίνδυνος» συνθέτεται από τον κίνδυνο που προέρχεται, τόσο από τις απειλές που παραμένουν άλυτες, όσο και από τις απειλές που δεν εξαλείφονται πλήρως από τα μέτρα προστασίας που έχουν προταθεί.

7. ΒΗΜΑ 7ο: Τεκμηρίωση της διαδικασίας «Privacy Impact Assessment».

Τέλος, στο βήμα αυτό συντάσσονται οι αναφορές, οι οποίες περιέχουν στοιχεία ανάλογα με το κοινό στο οποίο απευθύνονται. Έτσι απαιτείται, αφενός η έκδοση μιας πλήρους αναφοράς που θα περιλαμβάνει όλα τα στοιχεία, μεταξύ των οποίων και εμπιστευτικών, όπως η λεπτομερής περιγραφή του συστήματος, το είδος των δεδομένων και οι ροές τους, οι «Στόχοι Ιδιωτικότητας» και οι απειλές τους, τα μέτρα προστασίας τους και το αντίστοιχο σχέδιο υλοποίησης, ο «Υπολειπόμενος Κίνδυνος» καθώς και η συμμόρφωση με την κείμενη νομοθεσία, η οποία θα απευθύνεται, τόσο στον ίδιο τον «Υπεύθυνο Επεξεργασίας», όσο και στις «Ρυθμιστικές Αρχές» που θα διεξάγουν επιθεωρήσεις και ελέγχους συμμόρφωσης. Και αφετέρου, μια έκδοση με γενικά στοιχεία και λιγότερες λεπτομέρειες, που θα περιλαμβάνει στοιχεία όπως η γενική περιγραφή του συστήματος και ο σκοπός του, τα αποτελέσματα της μεθοδολογίας «PIA», οι απειλές και τα μέτρα που υιοθετήθηκαν, οι υπεύθυνοι σε περίπτωση παραβίασης, ο χρόνος εκτέλεσης και η διάρκειά της, η οποία θα

απευθύνεται, τόσο στα ίδια τα «Υποκείμενα των Δεδομένων», όσο και στα Μέσα Μαζικής Ενημέρωσης.



ΕΙΚΟΝΑ 11: PROCESS DEFINITION OF PIA METHODOLOGY

Συνοψίζοντας, αν και η «Privacy Impact Assessment» αποτελεί μια πολλά υποσχόμενη ολοκληρωμένη μεθοδολογία στην κατεύθυνση της προστασίας της «Ιδιωτικότητας», που ταυτόχρονα θα εγγυάται, τόσο την συμμόρφωση με την κείμενη ευρωπαϊκή νομοθεσία, όσο και την επίτευξη του «Privacy By Design», εντούτοις

υπάρχουν κάποιες επιπλέον ενέργειες που θα πρέπει να γίνουν, προκειμένου αυτή να καταστεί ευρέως αποδεκτή. Αρχικά, θα πρέπει να αξιολογηθεί στην πράξη ως προς την αποτελεσματικότητά της, τόσο από την ευρύτερη επιστημονική κοινότητα, όσο και από τον επιχειρηματικό κόσμο. Έπειτα, αφού αξιολογηθεί επιτυχώς, ακολουθούν οι απαραίτητες ενέργειες προκειμένου αυτή να καταστεί πρότυπο. Στις ενέργειες αυτές, ενδεικτικά, περιλαμβάνονται ο καθορισμός σαφών και ευρέως αποδεκτών «Στόχων Ιδιωτικότητας», η καταγραφή λεπτομερών οδηγιών για τη χρήση της, η βελτίωση και συμπλήρωση των διαδικασιών που πραγματοποιούνται σε κάθε βήμα, η καταγραφή όλων των δυνατών μέτρων προστασίας και η αναλυτική περιγραφή τους, και ο σαφής καθορισμός των αναγκαίων εγγράφων που θα πρέπει να την συνοδεύουν, κατ' αντιστοιχία του προτύπου «ISO/IEC 27001». Άλλωστε, η προτυποποίηση αποτελεί βασικό εργαλείο για την υποστήριξη της διαδικασίας της «Πιστοποίησης» (Certification), που προβλέπεται στο Άρθρο 42 του Τμήματος 5 του Κεφαλαίου IV του «Ευρωπαϊκού Κανονισμού». Στόχος της «Πιστοποίησης» είναι, αφενός η διερεύνηση και αφετέρου η διαβεβαίωση της συμμόρφωσης του πληροφοριακού συστήματος και ως εκ τούτου του «Υπεύθυνου Επεξεργασίας» του με τα καθοριζόμενα στον «Ευρωπαϊκό Κανονισμό (ΕΕ) 2016/679».

3.4.2 ΑΝΩΝΥΜΟΠΟΙΗΣΗ (ANONYMIZATION)

Στα συστήματα «Big Data» η προσοχή εστιάζεται, είτε στην δυνατότητα απομόνωσης ορισμένων ή και όλων των αρχείων που προσδιορίζουν ένα άτομο σε ένα σύνολο δεδομένων (Singling out of a Record), είτε στην διασύνδεση τουλάχιστον δυο αρχείων που προσδιορίζουν το ίδιο άτομο στην ίδια ή σε διαφορετικές βάσεις δεδομένων (Linkability of two Records), είτε στην δυνατότητα να συναχθεί, με πολύ καλή πιθανότητα, η τιμή ενός χαρακτηριστικού κάποιου συγκεκριμένου ατόμου μέσα από το πλήθος των τιμών ενός συνόλου χαρακτηριστικών (Inference of Attributes) [77]. Άρα θα πρέπει να εξεταστεί, πως είναι δυνατό να αποτραπούν οι περιπτώσεις του να ταυτοποιηθεί εκ νέου το «Υποκείμενο» των δεδομένων (Re-Identification), να αποκαλυφθεί η τιμή κάποιου χαρακτηριστικού του (Attribute Disclosure) και να συσχετιστούν μεταξύ τους διαφορετικά αρχεία, που περιέχουν δεδομένα του (Linking Records). Η απάντηση στο ερώτημα αυτό είναι μέσω της τεχνικής της «Ανωνυμοποίησης». Συνεπώς, η «Ανωνυμοποίηση» (Anonymization) είναι η διαδικασία εκείνη, κατά την οποία τα σύνολα δεδομένων τροποποιούνται με τέτοιο τρόπο, έτσι ώστε, αφενός το «Υποκείμενο» των δεδομένων να μην είναι εφικτό να

ταυτοποιηθεί εκ νέου (Re-Identification) και αφετέρου να μην υπάρχει η δυνατότητα να διαρρεύσει οποιαδήποτε πληροφορία του (Attribute Disclosure) από την διασύνδεση αυτών μεταξύ τους [110], [183].

Επομένως, όσον αφορά την «Ανωνυμοποίηση», η «Ιδιωτικότητα» ενός ατόμου απειλείται, είτε αν αποκαλυφθεί η ταυτότητά του, είτε αν αποκαλυφθεί κάποιο από τα προσωπικά του δεδομένα. Οπότε οι επιθέσεις εναντίον της «Ανωνυμοποίησης» εντάσσονται σε μια από τις παρακάτω κατηγορίες [3]:

1) Αποκάλυψη της Ταυτότητας (Identity Disclosure).

Σε αυτή την κατηγορία ανήκουν οι επιθέσεις εκείνες, κατά τις οποίες ο επιτιθέμενος έχει την δυνατότητα να αντιστοιχήσει τα αρχεία, που βρίσκονται σε ένα σύνολο δεδομένων που έχει δημοσιοποιηθεί, με το αντίστοιχο «Υποκείμενό» τους. Αυτό είναι γνωστό και ως «Αποκάλυψη της Οντότητας» (Entity Disclosure). Επίσης, εδώ εντάσσονται και οι επιθέσεις που στοχεύουν στην «Επαναταυτοποίηση» του ατόμου (Re-Identification).

2) Αποκάλυψη του Χαρακτηριστικού ενός Ατόμου (Attribute Disclosure).

Σε αυτή την κατηγορία ανήκουν οι επιθέσεις εκείνες, κατά τις οποίες ο επιτιθέμενος έχει την δυνατότητα να βελτιώσει τις γνώσεις του σχετικά με την τιμή κάποιου χαρακτηριστικού ενός ατόμου. Στην ίδια κατηγορία εντάσσονται και οι περιπτώσεις εκείνες, κατά τις οποίες ο επιτιθέμενος μπορεί να εντοπίσει αν τα προσωπικά δεδομένα κάποιου ατόμου συμπεριλαμβάνονται μέσα στο συγκεκριμένο σύνολο δεδομένων [84].

Στο σημείο αυτό, είναι αναγκαίο να επισημανθεί ότι οι κατηγορίες αυτές είναι διαφορετικές και ανεξάρτητες μεταξύ τους. Δηλαδή η αποκάλυψη της ταυτότητας ενός ατόμου, δεν συνεπάγεται ότι έχει προηγηθεί απαραίτητα και αποκάλυψη κάποιου χαρακτηριστικού του. Οπότε, είναι δυνατό να προκύψει αποκάλυψη της ταυτότητας ενός ατόμου, κατά την διαδικασία διασύνδεσης διαφορετικών συνόλων δεδομένων, μόνο με την χρήση των ήδη γνωστών πληροφοριών. Κατ' αντιστοιχία, από την άλλη πλευρά, είναι δυνατό να αποκαλυφθεί κάποιο χαρακτηριστικό ενός ατόμου, χωρίς προηγουμένως να απαιτείται να έχει αποκαλυφθεί η ταυτότητά του, όπως στην περίπτωση κατά την οποία ένα σύνολο ατόμων μοιράζεται την ίδια τιμή για κάποιο κοινό χαρακτηριστικό τους.

Επιπλέον πρέπει να σημειωθεί ότι η μελέτη του ρίσκου για τις δυο αυτές περιπτώσεις είναι δύσκολη στο περιβάλλον των «Big Data». Έτσι στην πρώτη περίπτωση, αν και θεωρείται ότι ο κίνδυνος να συμβεί είναι υπερτιμημένος και επομένως δεν θα πρέπει να αποτελεί τροχοπέδη στην δημοσιοποίηση ενός συνόλου δεδομένων [86], [87], εντούτοις μια απλή παραβίαση του συστήματος είναι ικανή να φέρει όλες τις πληροφορίες κοντά και να οδηγήσει στην εκ νέου αναγνώριση ενός ή και περισσότερων ατόμων. Ενώ στην δεύτερη περίπτωση το βασικό ζήτημα που τίθεται, ιδιαίτερα στους τεχνικούς της στατιστικής και της αναλυτικής, αφορά την επαύξηση του ρίσκου για την αποκάλυψη κάποιου χαρακτηριστικού ενός συνόλου ατόμων, ή ακόμα και ενός ατόμου, μετά την δημοσιοποίηση ενός συνόλου δεδομένων, στο οποίο περιέχονται επαρκείς πληροφορίες, οι οποίες μπορεί να οδηγήσουν στην αποκάλυψη, με καλή ακρίβεια, της τιμής κάποιου χαρακτηριστικού των/του ατόμων/ου αυτών/ου [85]. Επομένως, σε κάθε περίπτωση, είναι αναγκαίο να διεξάγεται λεπτομερής ανάλυση για τον κίνδυνο αποκάλυψης, πριν από την δημοσιοποίηση ενός «Ανωνυμοποιημένου» συνόλου δεδομένων.

Έτσι, με βάση την ανωτέρω κατηγοριοποίηση των επιθέσεων, τα μοντέλα «Ανωνυμοποίησης» μπορούν να ενταχθούν σε δυο βασικές ομάδες [3]:

❖ Στην πρώτη ομάδα εντάσσονται η «k-anonymity» [88], η οποία εστιάζει στην μη αποκάλυψη της ταυτότητας ενός ατόμου, και οι επεκτάσεις της, όπως η «p-sensitive k-anonymity» [89], η «l-diversity» [90], η «t-closeness» [91], η «(n,t)-closeness» [92] κλπ, οι οποίες την συμπληρώνουν και αποσκοπούν στην μη αποκάλυψη κάποιου χαρακτηριστικού του. Επιπρόσθετα, με αυτά τα μοντέλα αντιμετωπίζονται και οι επιθέσεις «Επαναταυτοποίησης» του ατόμου.

❖ Στην δεύτερη ομάδα εντάσσονται η «ε-differential privacy» [93] με τις παραλλαγές της, όπως η «crowd-blending privacy» [94], ή ο «BlowFish» [95], οι οποίες αποσκοπούν στον περιορισμό της συμμετοχής του ατόμου στην εξαγωγή του αποτελέσματος της ανάλυσης. Αρχικά η «ε-differential privacy» χρησιμοποιούταν μόνο για την «Ανωνυμοποίηση» των αποτελεσμάτων που επιστρέφονταν κατά την διεξαγωγή διαδραστικών επερωτήσεων σε βάσεις δεδομένων. Για αυτό αποκτά ιδιαίτερο ενδιαφέρον στο περιβάλλον των «Big Data», καθώς στοχεύει στην αποτροπή της αποκάλυψης της τιμής του χαρακτηριστικού ενός ατόμου στον χρήστη εκείνο, που εκτελεί την διαδικασία της ανάλυσης επί του συνόλου δεδομένων του συστήματος.

Συγκρίνοντας τα δυο αυτά μοντέλα για το περιβάλλον των «Big Data», υπάρχει η αίσθηση ότι αυτά είναι ανταγωνιστικά μεταξύ τους. Μάλιστα, το μοντέλο «k-anonymity» έχει κατακριθεί λόγω των αδυναμιών που έχει, με αποτέλεσμα η «ε-differential privacy» να θεωρείται ως η καλύτερη λύση για το πρόβλημα της «Ανωνυμοποίησης» [97]. Εντούτοις, υπάρχει μια ειδοποιός διαφορά μεταξύ των δυο [98]. Το μοντέλο «k-anonymity» στοχεύει στο να καταστήσει τα σύνολα δεδομένων ανώνυμα, πριν από την διάθεσή τους για περαιτέρω ανάλυση. Ενώ το μοντέλο «ε-differential privacy» στοχεύει στο να προστατέψει την «Ιδιωτικότητα» του ατόμου στις απαντήσεις που θα επιστραφούν, μετά από τις επερωτήσεις που θα γίνουν (Query) στα πλαίσια ενός προκαθορισμένου τύπου ανάλυσης. Επιπλέον, αν και θεωρείται ότι η «ε-differential privacy» είναι ανώτερη από την «k-anonymity», εντούτοις αυτή δεν είναι εφικτό να εφαρμοστεί σε όλες τις περιπτώσεις εξαιτίας του ότι περιορίζει σημαντικά την χρηστικότητα του συνόλου δεδομένων [106]. Αυτό σημαίνει ότι ακόμα το μοντέλο «k-anonymity» διαδραματίζει σημαντικό ρόλο, ειδικά στην «Ανωνυμοποίηση» των συνόλων δεδομένων που πρόκειται να διατεθούν για περαιτέρω επεξεργασία, η οποία δεν είναι εφικτό να πραγματοποιηθεί μέσω της διαδικασίας των επερωτήσεων. Εν κατακλείδι, διαπιστώνεται ότι και τα δυο είναι χρήσιμα στο σύγχρονο περιβάλλον, παρόλο τα μειονεκτήματά τους, καθώς τυγχάνουν εφαρμογής καλύπτοντας διαφορετικές περιπτώσεις το κάθε ένα. Αυτό συνεπάγεται ότι απαιτείται περαιτέρω έρευνα, προκειμένου να διατηρηθεί μια καλή ισορροπία μεταξύ της χρηστικότητας (Utility) των συνόλων δεδομένων και της προστασίας της «Ιδιωτικότητας» (Privacy) από την χρήση των μοντέλων αυτών στις διάφορες περιπτώσεις στο περιβάλλον των «Big Data».

Επίσης, όσον αφορά τις μεθόδους «Ανωνυμοποίησης», αυτές είναι [3]:

❖ Η Μέθοδος της Συγκάλυψης των Δεδομένων (Data Masking).

Με την μέθοδο αυτή παράγεται μια νέα τροποποιημένη έκδοση «Χ΄» του αρχικού συνόλου «Χ». Αυτή μπορεί να προκύψει με δύο τρόπους. Είτε με την διατάραξη των δεδομένων (Perturbative) μέσω των τεχνικών της προσθήκης θορύβου (Noise Addition), ή της αντικατάστασης με τις μέσες τιμές ισόποσων ομάδων τιμών (Micro Aggregation), ή της εναλλαγής τους (Data Swapping), ή της στοχαστικής αντικατάστασης των τιμών με τη βοήθεια «Μαρκοβιανών» αλυσίδων (Post Randomization). Είτε χωρίς την διατάραξη των δεδομένων (Non Perturbative) μέσω των τεχνικών της διάθεσης δείγματος του συνόλου (Sampling), ή της γενικοποίησής

του (Generalization), ή της απαλοιφής των άκρων (Top / Bottom Coding), ή της τοπικής συμπίεσης των τιμών (Local Suppression).

❖ Η Μέθοδος της Σύνθεσης των Δεδομένων (Data Synthesis).

Με αυτή την μέθοδο παράγεται ένα νέο τεχνητό σύνολο «Χ΄», το οποίο όμως διατηρεί κάποιες προεπιλεγμένες ιδιότητες, στατιστικά και συσχετίσεις του αρχικού συνόλου «Χ». Αυτό επιτυγχάνεται με τις ακόλουθες μεθόδους, είτε της παραγωγής πλήρως συνθετικών συνόλων δεδομένων (Fully Synthetic Data Generation), είτε της παραγωγής μερικώς συνθετικών συνόλων δεδομένων (Partially Synthetic Data Generation), είτε της υβριδικής μεθόδου (Hybrid Data), κατά την οποία συνθέτονται σε ένα νέο σύνολο το αρχικό και το τεχνητό σύνολο. Οι δυο τελευταίες μέθοδοι είναι οι μοναδικές της κατηγορίας αυτής, που είναι εφαρμόσιμες στην περίπτωση της «Ανωνυμοποίησης» ενός συνόλου δεδομένων στην πηγή του, πριν αυτό χρησιμοποιηθεί στη συνέχεια από το σύστημα «Big Data».

Ωστόσο και εδώ υπάρχουν κάποια ζητήματα που σχετίζονται με τις μεθόδους αυτές και που χρήζουν περαιτέρω έρευνας [111]. Αυτά αφορούν:

✓ Την Συγκρισιμότητα.

Δηλαδή ποιά μέθοδος «Ανωνυμοποίησης» παρέχει κάθε φορά τον καλύτερο βαθμό χρηστικότητας στο σύνολο δεδομένων που πρόκειται να διασυνδεθεί και ταυτόχρονα ικανοποιητικό επίπεδο στην προστασία από τον κίνδυνο αποκάλυψης κάποιου χαρακτηριστικού, ή της ταυτότητας του ατόμου. Αυτό το ζήτημα πηγάζει από το γεγονός ότι η κάθε μέθοδος στηρίζεται σε διαφορετική αρχή για την υλοποίησή της.

✓ Την Επαληθευσιμότητα από τα ίδια τα «Υποκείμενα» των Δεδομένων.

Δηλαδή κατά πόσο συμμετέχει και ενημερώνεται το ίδιο το «Υποκείμενο» των δεδομένων για τις διαδικασίες «Ανωνυμοποίησης», που εφαρμόζει ο εκάστοτε «Υπεύθυνος Επεξεργασίας» στα δεδομένα του.

✓ Τις Δυνατότητες του Επιτιθέμενου.

Δηλαδή πόσο αποτελεσματική είναι η εκάστοτε μέθοδος «Ανωνυμοποίησης», που εφαρμόζεται σε ένα σύνολο δεδομένων, συγκριτικά με τις δυνατότητες του εκάστοτε επιτιθέμενου.

✓ Την Παροχή Διαφάνειας στα «Υποκείμενα» των Δεδομένων.

Δηλαδή πόσες και ποιές λεπτομέρειες θα πρέπει να γνωστοποιηθούν στο «Υποκείμενο» των δεδομένων, όσον αφορά την μέθοδο «Ανωνυμοποίησης» που χρησιμοποιήθηκε.

Γενικά, είναι αποδεκτό ότι επίτευξη τέλειας «Ανωνυμοποίησης», χωρίς παράλληλα να απαξιωθεί η χρηστικότητα του συνόλου των δεδομένων, είναι εξαιρετικά δύσκολο να επιτευχθεί. Μάλιστα, αυτό αναμένεται να επιδεινωθεί στο περιβάλλον των «Big Data» εξαιτίας των ιδιαίτερων χαρακτηριστικών του, ήτοι του όγκου, της ποικιλομορφίας και της ταχύτητας των δεδομένων του. Έτσι από την μια πλευρά, η χαμηλού επιπέδου «Ανωνυμοποίηση» δεν είναι επαρκής για να αποτρέψει την αποκάλυψη της ταυτότητας ή κάποιου χαρακτηριστικού του ατόμου [107], [108]. Ενώ από την άλλη πλευρά, η υψηλού επιπέδου «Ανωνυμοποίηση», αν και αποτρέπει την διασύνδεση διαφορετικών συνόλων δεδομένων, τα οποία αφορούν το ίδιο άτομο ή την ίδια ομάδα ατόμων και τα οποία προέρχονται από διαφορετικές πηγές, εντούτοις καθιστά τα δεδομένα αυτά μη χρηστικά, ματαιώνοντας έτσι τα οφέλη που απορρέουν από την τεχνολογία των «Big Data». Αυτό συνεπάγεται ότι θα πρέπει να συμβιβαστούν και να μετριαστούν οι δυο αυτές αντικρουόμενες καταστάσεις.

Συνεπώς, τα ζητήματα που προκύπτουν από την εφαρμογή της τεχνικής της «Ανωνυμοποίησης» στο περιβάλλον των «Big Data» [3] είναι τα εξής:

❖ Η Ελεγχόμενη Συνδεσιμότητα (Controlled Linkability).

Η «Ανωνυμοποίηση» χρησιμοποιείται, για να προληφθεί η αποκάλυψη της ταυτότητας ή κάποιου χαρακτηριστικού ενός ατόμου από την διασύνδεση διαφορετικών συνόλων δεδομένων μεταξύ τους. Αυτό σημαίνει ότι η τεχνική αυτή επηρεάζει σε μεγάλο βαθμό την χρηστικότητα (Utility) ενός συνόλου δεδομένων, καθώς περιορίζει σημαντικά την δυνατότητα διασύνδεσής του με άλλα σύνολα. Από την άλλη πλευρά όμως, ένα σύνολο δεδομένων, για να είναι χρήσιμο στην εξαγωγή πολύτιμης πληροφορίας στο περιβάλλον των «Big Data», θα πρέπει να έχει την δυνατότητα να διασυνδέεται με τα υπόλοιπα. Έτσι το ζήτημα που προκύπτει είναι πώς μπορεί να διατηρείται το επιθυμητό επίπεδο συνδεσιμότητας ενός συνόλου δεδομένων, ενώ ταυτόχρονα αυτό έχει «Ανωνυμοποιηθεί» για να προστατεύσει την εκ νέου ταυτοποίηση ενός ατόμου ή την αποκάλυψη κάποιου χαρακτηριστικού του.

❖ Η Δυνατότητα Σύνθεσης (Composability).

Η «Συνθεσιμότητα» αφορά στο κατά πόσο ένα σύνολο δεδομένων, που προκύπτει από την διασύνδεση διαφορετικών συνόλων δεδομένων, τα οποία όμως έχουν «Ανωνυμοποιηθεί» πριν διασυνδεθούν μεταξύ τους με την χρήση κάποιου μοντέλου «Ανωνυμοποίησης», μπορεί να θεωρηθεί ότι είναι και αυτό «Ανωνυμοποιημένο». Μάλιστα εστιάζει και στο αν συνεχίζει το προκύπτον διασυνδεδεμένο σύνολο να

διατηρεί τον ίδιο βαθμό «Ανωνυμοποίησης» με αυτόν που είχαν τα αρχικά σύνολα δεδομένων. Το γεγονός αυτό αποκτά ιδιαίτερη σημασία στο περιβάλλον των «Big Data», στο οποίο διασυνδέονται ποικίλα σύνολα δεδομένων από διάφορες ανεξάρτητες πηγές.

❖ Η «Ανωνυμοποίηση» των Δυναμικών και Κινούμενων Δεδομένων (Anonymization of Dynamic / Streaming Data).

Η οποία αφορά, αφενός στο κατά πόσο είναι εφικτό να εφαρμοστούν οι διάφορες τεχνικές της «Ανωνυμοποίησης» πάνω σε δυναμικά και κινούμενα δεδομένα και αφετέρου ποιά θα είναι η χρηστικότητα των δεδομένων αυτών μετά την εφαρμογή της εκάστοτε τεχνικής. Επιπλέον, εξετάζεται αν και πώς είναι δυνατό να περιοριστεί ο κίνδυνος της αποκάλυψης της ταυτότητας ή κάποιου χαρακτηριστικού ενός ατόμου στην περίπτωση αυτή.

❖ Η Δυνατότητα Εκτέλεσης Υπολογισμών σε Μεγάλα Σύνολα Δεδομένων (Computability for Large Data Volumes).

Η οποία αφορά στο κατά πόσο είναι υπολογιστικά αποδοτικό και εφικτό να εφαρμοστούν οι διάφορες τεχνικές της «Ανωνυμοποίησης» πάνω στον τεράστιο όγκο των δεδομένων που διαχειρίζονται τα συστήματα «Big Data».

❖ Η Αποκεντρωμένη ή Κεντρική «Ανωνυμοποίηση» (Decentralized or Centralized Anonymization).

Η οποία αφορά το ποίος θα καταστήσει «Ανώνυμο» το σύνολο δεδομένων. Το ίδιο το «Υποκείμενο» των δεδομένων, πριν τα διαθέσει στο σύστημα «Big Data» για περαιτέρω επεξεργασία, ή κεντρικά ο «Υπεύθυνος Επεξεργασίας» του συστήματος αυτού, αφού προηγουμένως τα έχει συγκεντρώσει.

Αναλυτικότερα, δεδομένου ότι η «Ανωνυμοποίηση» επηρεάζει σε μεγάλο βαθμό την χρηστικότητα (Utility) ενός συνόλου δεδομένων, το πρώτο ζήτημα σχετίζεται με την διατήρηση της στο επιθυμητό επίπεδο κάθε φορά. Ιδανικά, θα έπρεπε τα αποτελέσματα που εξάγουν οι «Υπεύθυνοι Επεξεργασίας» μέσα από την επεξεργασία των «Ανωνυμοποιημένων» συνόλων δεδομένων να είναι τα ίδια σε ακρίβεια συγκρινόμενα με αυτά που θα προέκυπταν από την επεξεργασία των αντίστοιχων μη «Ανωνυμοποιημένων» συνόλων. Οπότε, το ερώτημα που γεννάται κατ' αρχάς είναι αν και πως μπορεί να μετρηθεί η χρηστικότητα ενός «Ανωνυμοποιημένου» συνόλου δεδομένων, έτσι ώστε να μην υπάρξει απώλεια, ή αυτή να είναι η ελάχιστη δυνατή, στην ακρίβεια των εξαγόμενων αποτελεσμάτων κατά την ανάλυσή του από το σύστημα «Big Data». Για την επίλυση του

προβλήματος αυτού έχουν προταθεί διάφορες λύσεις, οι οποίες όμως ποικίλλουν ανάλογα με τον τύπο και την χρήση των δεδομένων [112], [113], [114], [115]. Παράλληλα, τα ιδιαίτερα χαρακτηριστικά των «Big Data» δημιουργούν προβλήματα, τα οποία σχετίζονται με τον αντικειμενικό υπολογισμό της χρηστικότητας ενός συνόλου δεδομένων σε εφαρμογή των λύσεων αυτών, καθώς και με την μέτρηση της απώλειας των πληροφοριών από την εφαρμογή της «Ανωνυμοποίησης», γεγονός που το καθιστά ανοιχτό ερευνητικό πεδίο.

Ειδικότερα για το περιβάλλον των «Big Data», η χρηστικότητα ενός συνόλου δεδομένων σχετίζεται με το κατά πόσο αυτό μπορεί να διασυνδεθεί με άλλα σύνολα. Γενικά, η έννοια της «Συνδεσιμότητας» (Linkability) στο περιβάλλον αυτό αποκτά εξαιρετική σημασία, καθώς εγγυάται την δυνατότητα διασύνδεσης διαφορετικών συνόλων δεδομένων, τα οποία προέρχονται από διαφορετικές και μεταξύ τους ανεξάρτητες πηγές, προκειμένου αυτά να τύχουν επεξεργασίας και να εξαχθεί πολύτιμη πληροφορία. Παράλληλα όμως, η «Συνδεσιμότητα» αποτελεί απειλή για την «Ιδιωτικότητα» του ατόμου. Αυτό σημαίνει ότι θα πρέπει τα σύνολα αυτά να έχουν «Ανωνυμοποιηθεί» από την ίδια την πηγή, προτού διατεθούν για επεξεργασία. Αν όμως αυτό γίνει από την κάθε πηγή ξεχωριστά, τότε υπάρχει σοβαρός κίνδυνος να περιοριστεί σημαντικά η συνδεσιμότητα των συνόλων αυτών και να μην είναι δυνατή η περαιτέρω επεξεργασία τους. Οπότε, προκύπτει το ερώτημα αν και πως μπορεί να μετράται κάθε φορά ο βαθμός «Συνδεσιμότητας» ενός συνόλου δεδομένων, που παράγεται από την εφαρμογή του εκάστοτε μοντέλου και μεθόδου «Ανωνυμοποίησης». Διάφοροι τρόποι έχουν εξεταστεί [116], ένας εξ' αυτών η μέτρηση κάθε φορά του αριθμού των διασυνδέσεων που μπορούν να επιτευχθούν με το «Ανωνυμοποιημένο» αυτό σύνολο. Είναι προφανές ότι ο αριθμός αυτός θα είναι μικρότερος από αυτόν του αντίστοιχου μη «Ανωνυμοποιημένου» συνόλου. Ωστόσο και αυτός ο τομέας αποτελεί ανοιχτό ερευνητικό πεδίο, ειδικά για το περιβάλλον των «Big Data».

Συνεπώς, εφόσον βρεθεί τρόπος να μετρηθούν η χρηστικότητα και η συνδεσιμότητα ενός «Ανωνυμοποιημένου» συνόλου δεδομένων, η επίλυση του ζητήματος, που αφορά το πως είναι δυνατό να επιτυγχάνεται κάθε φορά το επιθυμητό επίπεδο αυτών, έτσι ώστε να συντελείται η ελάχιστη δυνατή απώλεια στην ακρίβεια των εξαγόμενων αποτελεσμάτων από την διασύνδεση των διαφόρων «Ανωνυμοποιημένων» συνόλων δεδομένων μεταξύ τους, και ταυτόχρονα να

προλαμβάνεται η αποκάλυψη της ταυτότητας ή κάποιου χαρακτηριστικού ενός ατόμου, μπορεί να πραγματοποιηθεί με μια από τις παρακάτω προσεγγίσεις:

❖ Η πρώτη εστιάζει στην επίτευξη ικανοποιητικού βαθμού χρηστικότητας του συνόλου δεδομένων (Utility First Anonymization).

Σύμφωνα με αυτή, εφαρμόζεται στο αρχικό σύνολο δεδομένων το μοντέλο «Ανωνυμοποίησης» με τέτοιες παραμέτρους, έτσι ώστε να επιτυγχάνεται ο επιθυμητός βαθμός χρηστικότητας του συνόλου που θα προκύψει. Στην συνέχεια υπολογίζεται το ρίσκο για την αποκάλυψη κάποιου δεδομένου ή την επαναταυτοποίηση του ατόμου. Αν αυτό το ρίσκο είναι υψηλό, τότε επαναλαμβάνεται η ίδια διαδικασία αλλά με αυστηρότερα κριτήρια για το μοντέλο της «Ανωνυμοποίησης», γεγονός που συνεπάγεται την αντίστοιχη υποβάθμιση του αρχικού επιθυμητού βαθμού χρηστικότητας.

❖ Η δεύτερη εστιάζει στην επίτευξη του ανώτατου επιθυμητού ορίου ως προς την προστασία της «Ιδιωτικότητας» (Privacy First Anonymization).

Σύμφωνα με αυτή, εφαρμόζεται στο αρχικό σύνολο δεδομένων το μοντέλο «Ανωνυμοποίησης» με τέτοιες παραμέτρους, οι οποίες θα εγγυώνται την επίτευξη του ανώτατου επιθυμητού ορίου για την προστασία από το ρίσκο της επαναταυτοποίησης ή της αποκάλυψης κάποιου δεδομένου του ατόμου. Στη συνέχεια υπολογίζεται ο βαθμός χρηστικότητας του συνόλου αυτού. Αν αυτός είναι πολύ μικρός, τότε, είτε χρησιμοποιείται άλλο μοντέλο, είτε χαλαρώνονται οι παράμετροι που χρησιμοποιήθηκαν, προκειμένου το σύνολο αυτό να αποκτήσει υψηλότερο βαθμό χρηστικότητας, χωρίς όμως να μεταβληθεί το επιθυμητό επίπεδο προστασίας της «Ιδιωτικότητας» που τέθηκε εξ' αρχής.

Όσον αφορά το δεύτερο ζήτημα, ο [96] καθόρησε και αξιολόγησε τις απαιτήσεις που θα πρέπει να ικανοποιεί ένα μοντέλο «Ανωνυμοποίησης» για να είναι χρήσιμο στο περιβάλλον των «Big Data». Συγκεκριμένα, αυτές είναι:

- ✓ Η «Συνθεσιμότητα» (Composability).
- ✓ Η «Συνδεσιμότητα» (Linkability).
- ✓ Η «Δυνατότητα Εκτέλεσης Υπολογισμών» (Computability).

Επιπλέον αξιολόγησε τα μοντέλα «Ανωνυμοποίησης» «k-anonymity» και «ε-differential privacy» με γνώμονα τις απαιτήσεις αυτές.

Έτσι, από την μια πλευρά, το μοντέλο «Ανωνυμοποίησης» «k-anonymity» ικανοποιεί την απαίτηση για «Συνδεσιμότητα» σε επίπεδο συνόλου μεταξύ «k»

εγγραφών, αλλά όχι και την «Συνθεσιμότητα», δηλαδή ο συνδυασμός δυο «k-ανώνυμων» συνόλων δεν εγγυάται ότι και το «k΄» σύνολο που προκύπτει ότι θα είναι «k- ανώνυμο» για «k΄>1». Έτσι, για παράδειγμα, αν έχουμε δυο «k- ανώνυμα» σύνολα δεδομένων με τους ασθενείς από δυο διαφορετικά νοσοκομεία, στα οποία συμπεριλαμβάνονται ο ταχυδρομικός κώδικας, το εύρος ηλικιών και το είδος της ασθένειας, είναι εφικτό να ταυτοποιηθεί κάποιο άτομο, αν υπάρχει η προηγούμενη πληροφόρηση ότι το συγκεκριμένο άτομο επισκέφτηκε και τα δυο νοσοκομεία και επιπλέον είναι γνωστά η ηλικία και ο τόπος διαμονής του. Γενικά η «k-anonymity» δεν μπορεί να εγγυηθεί την προστασία της «Ιδιωτικότητας», αν στο σύνολο δεδομένων δεν υπάρχει ποικιλία στις τιμές των ευαίσθητων δεδομένων και κάποιες επιπλέον πληροφορίες είναι διαθέσιμες στον επιτιθέμενο [109].

Ενώ, από την άλλη πλευρά, το μοντέλο «Ανωνυμοποίησης» «ε-differential privacy» είναι ισχυρά «Συνθέσιμο». Δηλαδή από τον συνδυασμό ενός «ε1-differential private» συνόλου με ένα «ε2-differential private» σύνολο θα προκύψει ένα νέο σύνολο «ε1+ε2-differential private», το οποίο θα ικανοποιεί την «ε-differential private» ανωνυμία του, αλλά όμως με λιγότερο αυστηρή παράμετρο. Όσον αφορά την «Συνδεσιμότητα» των «ε-differential private» συνόλων, αυτή εξαρτάται κάθε φορά από την μέθοδο «Ανωνυμοποίησης» που χρησιμοποιείται. Έτσι αυτά δεν την ικανοποιούν, αν για παράδειγμα έχουν δημιουργηθεί με την μέθοδο της «Προσθήκης Θορύβου» (Noise Addition), ενώ αντίθετα την ικανοποιούν, αν έχουν δημιουργηθεί από την μέθοδο της «Μερικής Συνθετικής Παραγωγής Δεδομένων» (Partially Synthetic Data Generation). Τέλος, όσον αφορά την «Δυνατότητα Εκτέλεσης Υπολογισμών» (Computability), αυτή εξαρτάται και για τα δυο μοντέλα από την μέθοδο «Ανωνυμοποίησης» που θα χρησιμοποιηθεί κάθε φορά.

Το τρίτο ζήτημα σχετίζεται με την «Ανωνυμοποίηση» των δυναμικών και των κινούμενων δεδομένων του περιβάλλοντος των «Big Data». Με τον όρο «Δυναμικά» δεδομένα περιγράφονται τα δεδομένα εκείνα, τα οποία ενδέχεται να μεταβάλλονται συχνά με την πάροδο του χρόνου και τα οποία παράλληλα μπορεί να διατίθενται και να επαναχρησιμοποιούνται ταυτόχρονα από πολλά σημεία [184]. Έτσι για παράδειγμα, ως τέτοια θεωρούνται τα δεδομένα που αποθηκεύονται στις διάφορες αποθήκες δεδομένων του συστήματος «Big Data», καθώς αυτά μεταβάλλονται συχνά λόγω της συχνής επεξεργασίας τους από την αλληλεπίδραση των διαφόρων χρηστών με το σύστημα. Το βασικό πρόβλημα, που προκύπτει στην περίπτωση

αυτή, είναι η πιθανότητα να υπάρξει αποκάλυψη δεδομένων, αν δεν ληφθεί υπόψη ποιές πληροφορίες έχουν ήδη καταστεί διαθέσιμες στους χρήστες από την δημοσιοποίηση των προγενέστερων εκδόσεων αυτών των συνόλων δεδομένων [117]. Για αυτό θα πρέπει να υπάρχει πλήρης επίγνωση του τι έχει ήδη δημοσιοποιηθεί στις προηγούμενες εκδόσεις, πριν καταστεί διαθέσιμη η τρέχουσα ενημερωμένη έκδοση. Διάφοροι αλγόριθμοι για την επίλυση του προβλήματος αυτού εξετάζονται στα [118], [119], ωστόσο απαιτείται περαιτέρω έρευνα προς την κατεύθυνση αυτή, ακριβώς λόγω των ιδιαιτεροτήτων του περιβάλλοντος των «Big Data».

Από την άλλη πλευρά, ως «Κινούμενα» δεδομένα περιγράφονται τα δεδομένα εκείνα, τα οποία εισέρχονται συνεχώς στο σύστημα «Big Data» από τις διάφορες πηγές του και τα οποία τυγχάνουν άμεσης επεξεργασίας με το που εισέλθουν σε αυτό [185]. Έτσι για παράδειγμα, ως τέτοια θεωρούνται τα δεδομένα που παράγονται συνεχώς από τους διάφορους αισθητήρες και τα οποία θα πρέπει να τύχουν ταχείας επεξεργασίας, προκειμένου να υπάρξει άμεση ενημέρωση για το οποιοδήποτε συμβάν. Τα ζητήματα που εγείρονται από αυτή την ιδιαιτερότητα είναι ποικίλα. Το πρώτο αφορά στην αντικειμενική δυσκολία που υπάρχει στην αξιολόγηση των αλγορίθμων για την προστασία της «Ιδιωτικότητας», λόγω του ότι δεν είναι πλήρες το σύνολο των δεδομένων που εισέρχεται στο σύστημα, καθώς αυτό καταφθάνει τμηματικά και πιθανόν όχι στην σωστή σειρά. Για τον ίδιο λόγο, το δεύτερο ζήτημα σχετίζεται με την αδυναμία στην αποτελεσματική εφαρμογή των υπαρχουσών τεχνικών της «Ανωνυμοποίησης» σε αυτά τα δεδομένα. Το τρίτο ζήτημα σχετίζεται με την δυσκολία στο να επιλεγούν οι κατάλληλες παράμετροι για τα διάφορα μοντέλα και μεθόδους «Ανωνυμοποίησης» που θα χρησιμοποιηθούν σε αυτά, λόγω του δυναμικού τους χαρακτήρα, αλλά και της ύπαρξης πολλών απρόβλεπτων παραγόντων που μπορεί να υπεισέλθουν στην διαδικασία συλλογής τους. Συνεπώς, όλα αυτά τα ζητήματα χρήζουν διερεύνησης από την επιστημονική κοινότητα, προκειμένου να βρεθούν νέοι μηχανισμοί και τεχνικές για την αποτελεσματική «Ανωνυμοποίηση» των δεδομένων αυτών.

Στο τέταρτο ζήτημα εξετάζεται η «Δυνατότητα Εκτέλεσης Υπολογισμών σε Μεγάλα Σύνολα Δεδομένων», δηλαδή κατά πόσο είναι υπολογιστικά αποδοτικό και εφικτό να εφαρμοστούν τα διάφορα μοντέλα και μέθοδοι «Ανωνυμοποίησης» στον τεράστιο όγκο του συνόλου των δεδομένων, που τυγχάνουν επεξεργασίας στο

περιβάλλον των «Big Data» [120]. Αρχικά, τα μοντέλα και οι μέθοδοι αυτοί σχεδιάστηκαν και αναπτύχθηκαν λαμβάνοντας υπόψη διαφορετικές αρχές και τεχνικές, ανάλογα με τον τύπο και την φύση των δεδομένων στα οποία έπρεπε να εφαρμοστούν κάθε φορά. Όμως στο περιβάλλον των «Big Data» τα σύνολα δεδομένων είναι ποικιλόμορφα, δηλαδή μέσα στο ίδιο σύνολο δεδομένων εμπεριέχονται διαφορετικοί τύποι δεδομένων. Αυτό σημαίνει ότι τα μοντέλα και οι μέθοδοι αυτοί ενδεχομένως να μην είναι αποτελεσματικοί, αν εφαρμοστούν σε αυτά τα ποικιλόμορφα σύνολα. Επιπλέον, το περιβάλλον των «Big Data» είναι πολυσύνθετο και πολύπλοκο, καθώς αποτελείται από πολλά επιμέρους τμήματα, τα οποία μπορεί να είναι και ανεξάρτητα μεταξύ τους. Έτσι άλλο ένα πρόβλημα που γεννάται είναι πως θα μπορούσαν αυτά τα μοντέλα και μέθοδοι να εφαρμοστούν αποτελεσματικά στο σύνολο των δεδομένων, τα οποία μπορεί να βρίσκονται σε διαφορετικές τοποθεσίες, και τι υπολογιστικό κόστος συνεπάγεται αυτό. Συνεπώς, απαιτείται περαιτέρω έρευνα προκειμένου να αντιμετωπιστούν τα προβλήματα αυτά και να αναπτυχθούν νέες τεχνικές «Ανωνυμοποίησης», οι οποίες θα τυγχάνουν εφαρμογής στο ιδιαίτερο αυτό περιβάλλον.

Τέλος, το πέμπτο ζήτημα αφορά το ποιός θα εκτελέσει την διαδικασία της «Ανωνυμοποίησης» πάνω στα δεδομένα των συστημάτων «Big Data». Τα σενάρια που υπάρχουν για αυτό το ζήτημα είναι τα εξής:

- Κεντρική Ανωνυμοποίηση (Centralized Anonymization)

Στο σενάριο αυτό η διαδικασία της «Ανωνυμοποίησης» εκτελείται κεντρικά από τον ίδιο τον «Υπεύθυνο Επεξεργασίας», ο οποίος έχει πλήρη πρόσβαση σε ολόκληρο το αρχικό σύνολο δεδομένων. Ένα πλεονέκτημα του σεναρίου αυτού είναι ότι τα «Υποκείμενα» των δεδομένων δεν απαιτείται να «Ανωνυμοποιήσουν» τα δεδομένα τους, πριν τα διαθέσουν στο σύστημα «Big Data», καθώς αυτό θα εκτελεστεί από τον «Υπεύθυνο Επεξεργασίας» του συστήματος, ο οποίος θα έχει καλύτερη υπολογιστική ισχύ και πιθανόν μεγαλύτερη εμπειρία σε αυτό. Ένα άλλο πλεονέκτημα είναι ότι ο «Υπεύθυνος Επεξεργασίας» έχει πλήρη εικόνα του συνόλου των δεδομένων που πρόκειται να «Ανωνυμοποιηθεί» και ως εκ τούτου είναι σε θέση να καθορίσει τις κατάλληλες εκείνες παραμέτρους που χρειάζονται, προκειμένου να επιτευχθεί το καλύτερο δυνατό αποτέλεσμα, όσον αφορά την χρηστικότητά του συνόλου, αλλά και την μείωση του ρίσκου αποκάλυψης της ταυτότητας ή κάποιου χαρακτηριστικού των «Υποκειμένων» των δεδομένων αυτών. Ωστόσο, σε αυτό το σενάριο προκύπτουν κάποια σοβαρά προβλήματα. Το πρώτο σχετίζεται με το ότι θα

πρέπει ο «Υπεύθυνος Επεξεργασίας» να τυγχάνει της πλήρους εμπιστοσύνης όλων των «Υποκειμένων» των δεδομένων αυτών. Το δεύτερο αναφέρεται στο ότι η διεξαγωγή της «Ανωνυμοποίησης» μόνο από τον «Υπεύθυνο Επεξεργασίας» μπορεί να αποβεί εξαιρετικά απαιτητική υπολογιστικά, ειδικά στο περιβάλλον των «Big Data». Το τρίτο αφορά το γεγονός ότι είναι σχεδόν αδύνατο να υπάρχει κεντρική διαχείριση στο περιβάλλον αυτό, καθώς υπάρχουν πάρα πολλοί εμπλεκόμενοι.

❖ Τοπική Ανωνυμοποίηση (Local Anonymization)

Στο σενάριο αυτό η «Ανωνυμοποίηση» των συνόλων δεδομένων υλοποιείται από τα ίδια τα «Υποκείμενα» των δεδομένων, πριν τα διαθέσουν στο σύστημα «Big Data». Έτσι αντιμετωπίζονται αποτελεσματικά τα προβλήματα του προηγούμενου σεναρίου. Ωστόσο, ενώ επιτυγχάνεται υψηλότερος βαθμός στην προστασία των δεδομένων των «Υποκειμένων», καθώς η διαδικασία αυτή εκτελείται από τα ίδια τα «Υποκείμενά» τους, εντούτοις υφίσταται ο κίνδυνος να μειωθεί σημαντικά ο βαθμός χρηστικότητας του συνόλου αυτού, γεγονός που θα οδηγήσει σε απώλεια πληροφοριών και ως εκ τούτου στην μείωση της απόδοσης της διαδικασίας ανάλυσής τους και εξαγωγής πολύτιμων συμπερασμάτων.

❖ Ανωνυμοποίηση Συνεργασίας (Co-utile Collaborative Anonymization)

Ένα εναλλακτικό και πιο αποδοτικό σενάριο, συγκριτικά με τα προηγούμενα, είναι αυτό κατά το οποίο η «Ανωνυμοποίηση» του συνόλου των δεδομένων πραγματοποιείται με συνεργατικό και κατανεμημένο τρόπο. Προκειμένου να καταστεί αυτό δυνατό, θα πρέπει να ικανοποιούνται οι εξής προϋποθέσεις:

(1) Να μην υπάρξει καμία επιπλέον απώλεια πληροφορίας στο σύνολο που θα παραχθεί με τον τρόπο αυτό, συγκριτικά με το «Ανωνυμοποιημένο» σύνολο δεδομένων που θα προέκυπτε, αν η «Ανωνυμοποίηση» του αρχικού συνόλου γινόταν κεντρικά από τον ίδιο τον «Υπεύθυνο Επεξεργασίας» του συστήματος, με δεδομένο ότι διατηρείται σταθερό το επίπεδο προστασίας της «Ιδιωτικότητας» και για τα δυο αυτά σύνολα.

(2) Να μην αποκτήσει κανείς, ούτε τα «Υποκείμενα» των δεδομένων αυτών, ούτε ο εκάστοτε «Υπεύθυνος Επεξεργασίας», περισσότερη γνώση για τα προσωπικά δεδομένα οποιουδήποτε άλλου «Υποκειμένου Δεδομένων», από την γνώση που θα αποκτούσε με την δημοσιοποίηση του τελικού «Ανωνυμοποιημένου» συνόλου δεδομένων.

Συνοψίζοντας, είναι σημαντικό να επισημανθεί, ότι η «Ανωνυμοποίηση» δεν θα πρέπει να θεωρείται ότι από μόνη της αποτελεί την υπέρτατη λύση για την

προστασία της «Ιδιωτικότητας» στο περιβάλλον των «Big Data». Αν και αποτελεί μια σημαντική λύση προς την κατεύθυνση αυτή, εντούτοις υπάρχουν περιπτώσεις στις οποίες αυτή μπορεί να μην είναι αποτελεσματική και ενδεχομένως να μην είναι δυνατό να εφαρμοστεί. Επομένως, θα πρέπει να χρησιμοποιείται σε συνδυασμό και με άλλες τεχνολογίες, προκειμένου να επιτευχθούν τα βέλτιστα αποτελέσματα σχετικά με την προστασία της «Ιδιωτικότητας» στο περιβάλλον αυτό.

3.4.3 ΚΡΥΠΤΟΓΡΑΦΙΑ (CRYPTOGRAPHY)

Η κρυπτογραφία είναι η σημαντικότερη τεχνική για την προστασία των προσωπικών δεδομένων στο ιδιαίτερο και πολύπλοκο περιβάλλον των «Big Data». Το ιδιαίτερο χαρακτηριστικό της τεχνικής αυτής είναι ότι τα δεδομένα μετατρέπονται σε ακατανόητη μορφή, με την χρήση κάποιου κρυπτογραφικού αλγορίθμου, έτσι ώστε από την μια πλευρά να αποκρύπτονται από τις μη εξουσιοδοτημένες οντότητες και από την άλλη να είναι δυνατό αυτά να αποκαλυφθούν μόνο στις αντίστοιχες νόμιμες και εξουσιοδοτημένες, οι οποίες κατέχουν το μυστικό της αποκρυπτογράφησης τους [186]. Το χαρακτηριστικό αυτό την καθιστά αναπόσπαστο εσωτερικό δομικό στοιχείο στα περισσότερα μέτρα ασφαλείας, αλλά και μηχανισμούς προστασίας της «Ιδιωτικότητας». Μάλιστα, κρίνεται επιτακτική η χρήση της στο περιβάλλον των «Big Data» για την προστασία, τόσο του ίδιου του συστήματος, όσο και των δεδομένων που αυτό διαχειρίζεται. Ειδικότερα, πέρα από την πρωταρχική λειτουργία της, ήτοι της απόκρυψης των δεδομένων από τις μη εξουσιοδοτημένες οντότητες, είτε αυτά βρίσκονται αποθηκευμένα, είτε μεταδίδονται, πλέον αποκτά και επιπλέον ρόλους. Αυτοί είναι η ασφαλής αναζήτηση σε κρυπτογραφημένα δεδομένα, το ασφαλές φιλτράρισμά τους, η διασφάλιση της ακεραιότητάς τους, ο έλεγχος της πρόσβασης σε αυτά και η εκτέλεση ασφαλών υπολογισμών με αυτά. Ιδιαίτερη βαρύτητα δίδεται στο να εκτελούνται οι προαναφερθέντες ρόλοι, χωρίς να απαιτείται προηγουμένως αυτά να έχουν αποκρυπτογραφηθεί. Η επίτευξη των ρόλων αυτών άμεσα συνεπάγεται την ασφάλεια της διαδικασίας της «Αναλυτικής» (Big Data Analytics), καθώς η επιτυχής υλοποίησή τους θα επιτρέψει την ασφαλή επεξεργασία των συνόλων δεδομένων, ενόσω αυτά είναι σε κρυπτογραφημένη μορφή. Επομένως, η τεχνική της κρυπτογραφίας αποτελεί προάγγελο για την ασφαλή χρησιμοποίηση των δεδομένων στο σύγχρονο περιβάλλον των «Big Data».

Προκειμένου όμως να είναι ασφαλής η εκτέλεση των ανωτέρω ενεργειών, θα πρέπει να καθοριστεί σαφώς το μοντέλο απειλών ανά ενέργεια. Γενικά το μοντέλο απειλών στην κρυπτογραφία ορίζεται μαθηματικά από την αλληλεπίδραση του συστήματος, που υλοποιεί το εκάστοτε κρυπτογραφικό πρωτόκολλο, με τον επιτιθέμενο. Έτσι στο μοντέλο αυτό, ο επιτιθέμενος αλληλεπιδρά με το σύστημα και προσπαθεί να υπολογίσει με καλή πιθανότητα και σε πολυωνυμικό χρόνο τις παραμέτρους του κρυπτογραφικού αλγορίθμου. Συνεπώς, ένα σύστημα θεωρείται ότι είναι ασφαλές, όταν ο επιτιθέμενος έχει πολύ μικρή πιθανότητα να υπολογίσει κάποιες από τις παραμέτρους του αλγορίθμου κρυπτογράφησης σε πολυωνυμικό χρόνο.

Μάλιστα, ένα υποψήφιο κρυπτογραφικό πρωτόκολλο αποδεικνύεται ότι είναι ισχυρό απέναντι σε κάποια απειλή [1], είτε με την μέθοδο των «Επιχειρημάτων Μείωσης» (Reduction Arguments), είτε με την μέθοδο των «Επιχειρημάτων Προσομοίωσης» (Simulation Arguments). Με την πρώτη μέθοδο αποδεικνύεται ότι, αν ο επιτιθέμενος μπορεί να υπολογίσει ορισμένες από τις παραμέτρους του κρυπτογραφικού αλγορίθμου, τότε μπορεί να σπάσει και ένα αριθμό θεωρητικών παραμέτρων, οι οποίες ευρέως θεωρούνταν ισχυρές. Ή εναλλακτικά, αν μπορεί να σπάσει την ασφάλεια των απλούστερων κρυπτογραφικών δομικών στοιχείων, που χρησιμοποιήθηκαν για την κατασκευή του πρωτοκόλλου αυτού, τότε μπορεί να σπάσει και το ίδιο το πρωτόκολλο. Με την δεύτερη μέθοδο αποδεικνύεται ότι, αν ένας επιτιθέμενος έχει την ικανότητα να σπάσει τον υποψήφιο κρυπτογραφικό αλγόριθμο, τότε θα μπορεί να σπάσει και την ιδανική λειτουργία του στο πείραμα προσομοίωσης, το οποίο αποτελεί και τις επίσημες προδιαγραφές ασφάλειας του αλγορίθμου [99].

Επομένως, το μοντέλο απειλών για τους επιπλέον ρόλους της τεχνικής της κρυπτογραφίας στο περιβάλλον των «Big Data» έχει ως εξής [1]:

1) Ένας κρυπτογραφικός μηχανισμός ελέγχου πρόσβασης θεωρείται ασφαλής, όταν ένας επιτιθέμενος δεν μπορεί να αποφασίσει με καλή πιθανότητα σε πολυωνυμικό χρόνο, από ποιό κείμενο εισόδου (Plaintext) έχει προκύψει το αντίστοιχο κρυπτοκείμενο (Ciphertext), ακόμα και αν του δοθεί η δυνατότητα να επιλέξει μεταξύ δύο, του σωστού κειμένου εισόδου και του λάθους. Η παραπάνω συνθήκη θα πρέπει να ικανοποιείται, ακόμα και στην περίπτωση που κάποια μέρη, τα οποία εξαιρούνται από τον μηχανισμό ελέγχου πρόσβασης, αποφασίσουν να συνεργαστούν και μεταξύ τους, αλλά και με τον επιτιθέμενο.

2) Τα κρυπτογραφικά πρωτόκολλα αναζήτησης και φιλτραρίσματος των κρυπτογραφημένων δεδομένων θεωρούνται ασφαλή, όταν ο επιτιθέμενος δεν μπορεί να εξάγει καμία απολύτως πληροφορία για τα κρυπτογραφημένα δεδομένα, παρά μόνο να διαπιστώσει αν το εν λόγω αίτημα ικανοποιήθηκε. Περαιτέρω έρευνα πέτυχε να αποκρύψει και αυτή την πληροφορία, οπότε πλέον ο επιτιθέμενος δεν μαθαίνει απολύτως τίποτα, ούτε για τα κρυπτογραφημένα δεδομένα, ούτε για τα κριτήρια αναζήτησης.

3) Τα κρυπτογραφικά πρωτόκολλα που εκτελούν υπολογισμούς πάνω σε κρυπτογραφημένα δεδομένα θεωρούνται ασφαλή, όταν ένας επιτιθέμενος δεν μπορεί να αποφασίσει με καλή πιθανότητα και σε πολυωνυμικό χρόνο από ποιό κείμενο εισόδου (Plaintext) έχει προκύψει το αντίστοιχο κρυπτοκείμενο (Ciphertext), ακόμα και αν του δοθεί η δυνατότητα να επιλέξει μεταξύ δύο, του σωστού κειμένου εισόδου και του λάθους. Αξίζει να σημειωθεί ότι αυτή η συνθήκη είναι μια πολύ αυστηρή απαίτηση, καθώς ο επιτιθέμενος έχει την δυνατότητα να εκτελέσει μια σειρά κρυπτογραφήσεων των αυθεντικών δεδομένων, χρησιμοποιώντας στον αλγόριθμο κρυπτογράφησης διάφορες αυθαίρετες παραμέτρους. Στο σενάριο αυτό, το πιο ισχυρό μοντέλο απειλών της απλής κρυπτογράφησης των αποθηκευμένων δεδομένων, που είναι η «Chosen Ciphertext Attack», δεν έχει νόημα, οπότε περαιτέρω έρευνα απαιτείται για τον ορισμό ενός νέου αντίστοιχου μοντέλου [121].

4) Για τα κρυπτογραφικά πρωτόκολλα που διασφαλίζουν την ακεραιότητα των δεδομένων, που προέρχονται από κάποια ήδη ταυτοποιημένη πηγή, υπάρχει μια ποικιλία από μοντέλα απειλών. Η βασική απαίτηση σε όλα αυτά είναι να μην έχει την ικανότητα ο επιτιθέμενος να πλαστογραφήσει δεδομένα. Δηλαδή να μην μπορεί να στείλει δεδομένα προσποιούμενος ότι αυτά έχουν σταλεί από την ήδη ταυτοποιημένη πηγή, ενώ στην πραγματικότητα αυτά δεν προέρχονται από αυτή. Επίσης, μια άλλη απαίτηση είναι αυτή της υιοθέτησης κάποιου βαθμού ανωνυμίας. Αυτό επιτυγχάνεται, όταν η πηγή των δεδομένων μπορεί να προσδιοριστεί μόνο ως μέλος μιας ομάδας. Ενδέχεται όμως σε κάποιες περιπτώσεις, όπως για λόγους λογοδοσίας (Accountability), να απαιτείται μια «Έμπιστη Τρίτη Οντότητα» (Trusted Third Party) να έχει την δυνατότητα να συνδέσει απευθείας τα δεδομένα με την ακριβή πηγή προέλευσής τους.

Κατόπιν των παραπάνω, στη συνέχεια παρατίθενται κάποιες από τις υπάρχουσες τεχνολογικές, αλλά και ερευνητικές λύσεις, για την αποτελεσματική υποστήριξη των επιπλέον ρόλων της κρυπτογραφίας στο περιβάλλον των «Big

Data», οι οποίες ικανοποιούν τις απαιτήσεις ασφαλείας των αντίστοιχων μοντέλων απειλών τους [3], [51]. Σκοπός τους είναι η ασφαλής επεξεργασία των συνόλων δεδομένων στο περιβάλλον αυτό και μάλιστα η εκτέλεση της επεξεργασίας επί των κρυπτογραφημάτων αυτών, προστατεύοντας έτσι την «Ιδιωτικότητα» των «Υποκειμένων» τους.

Αναλυτικότερα, όσον αφορά την διασφάλιση της εμπιστευτικότητας των δεδομένων που είναι αποθηκευμένα ή μεταδίδονται, έχουν προταθεί και χρησιμοποιούνται διάφορες λύσεις, οι οποίες εξετάστηκαν σε προηγούμενες ενότητες της παρούσα μεταπτυχιακής διπλωματικής εργασίας. Επιπλέον, η «Υβριδική» (Hybrid) κρυπτογράφηση χρησιμοποιείται ολοένα και περισσότερο [187], καθώς συνδυάζει τα πλεονεκτήματα της συμμετρικής κρυπτογράφησης, ήτοι την αποδοτικότητα σχετικά με την υπολογιστική ισχύ και την ταχύτητα εκτέλεσης, και της ασύμμετρης, ήτοι την ασφαλή διαχείριση των κλειδιών κρυπτογράφησης και της δυνατότητας κλιμάκωσής της σε μεγάλο αριθμό οντοτήτων, εξαλείφοντας ταυτόχρονα τα προβλήματά τους. Ωστόσο, εκτός από την παραδοσιακή κρυπτογράφηση, που βασίζεται στην λογική του «κρυπτογράφησέ τα όλα ή τίποτα», η υποστήριξη των νέων ρόλων της στο περιβάλλον των «Big Data» επιβάλλει την υιοθέτηση μιας νέας λογικής. Αυτή θα εστιάζει στην «Κρυπτογράφηση σε Περισσότερα και με Μεγαλύτερη Λεπτομέρεια Επίπεδα» των δεδομένων (Fine Grained Encryption), έτσι ώστε η εκάστοτε κατηγορία χρηστών να έχει διαφορετικά δικαιώματα πρόσβασης, διαμοιρασμού και επεξεργασίας σε διαφορετικό υποσύνολο του συνόλου δεδομένων κάθε φορά, ανάλογα με την εκάστοτε πολιτική ασφαλείας, προστατεύοντας ταυτόχρονα την «Ιδιωτικότητα» των «Υποκειμένων» τους. Οι τεχνικές που μπορούν να ανταπεξέλθουν στην λογική αυτή είναι:

- Η «Attribute Based Encryption - ABE»

Αυτή συνδυάζει τον έλεγχο της πρόσβασης στα δεδομένα και την «Υποδομή Δημοσίου Κλειδιού» (Public Key Cryptography - PKI). Σε αυτή, το κλειδί της κρυπτογράφησης (Key), αλλά και το ίδιο το κρυπτοκείμενο (Ciphertext), εξαρτώνται αποκλειστικά από προκαθορισμένα χαρακτηριστικά. Έτσι, αν τα χαρακτηριστικά που θα χρησιμοποιηθούν για την αποκρυπτογράφηση δεν ταιριάζουν απόλυτα με τα προκαθορισμένα, τότε αυτή αποτυγχάνει, ακόμα και αν συνεργαστούν οι χρήστες μεταξύ τους (Collusion Attack) [28], [101].

- Η «Identity Based Encryption - IBE»

Αποτελεί μια απλούστερη παραλλαγή της προηγούμενης. Σε αυτή το κλειδί της κρυπτογράφησης (Key), αλλά και το ίδιο το κρυπτοκείμενο (Ciphertext), εξαρτώνται αποκλειστικά από κάποια ταυτότητα των χρηστών, όπως το «e-mail». Όμοια με πριν, η αποκρυπτογράφηση πραγματοποιείται μόνο από την συγκεκριμένη ταυτότητα και όχι από άλλες, ακόμα και αν συνεργαστούν διάφοροι χρήστες μεταξύ τους (Collusion Attack) [100], [122].

❖ Η «Functional Encryption - FE»

Αποτελεί μετεξέλιξη της «ABE». Σε αυτή, ο εκάστοτε χρήστης δημιουργεί το δικό του κλειδί, χρησιμοποιώντας τα συγκεκριμένα χαρακτηριστικά του. Έτσι το κλειδί αυτό του δίνει την δυνατότητα να εκτελέσει συγκεκριμένες λειτουργίες πάνω στα κρυπτογραφημένα δεδομένα. Αυτό αποκτά ιδιαίτερη εφαρμογή σε περιπτώσεις, όπου ένα κρυπτογραφημένο κείμενο είναι ευρέως ορατό, αλλά ο κάθε χρήστης μπορεί να αποκρυπτογραφήσει μόνο ένα ορισμένο τμήμα του και να εκτελέσει σε αυτό συγκεκριμένες πράξεις επεξεργασίας, δηλαδή αυτές που του επιτρέπει το ιδιωτικό του κλειδί. Έτσι για παράδειγμα, αν κάποιος χρήστης θέλει να μετρήσει τον αριθμό των ασθενών για μια συγκεκριμένη ασθένεια σε ένα κρυπτογραφημένο σύνολο ιατρικών δεδομένων, τότε θα είναι δυνατό να το εκτελέσει με την τεχνική αυτή, χωρίς προηγουμένως να χρειάζεται να το αποκρυπτογραφήσει [123].

Εντούτοις, αφενός οι παραπάνω τεχνικές είναι ακόμα υπό ανάπτυξη και διερεύνηση και αφετέρου οι αλγόριθμοι που χρησιμοποιούνται για την υλοποίησή τους είναι υπολογιστικά μη αποδοτικοί.

Εκτός από τον προηγούμενο ρόλο της κρυπτογραφίας, εξίσου σημαντικός είναι και αυτός της ασφαλούς αναζήτησης και φιλτραρίσματος. Μάλιστα, προκειμένου να προστατευτεί η «Ιδιωτικότητα» στο περιβάλλον των «Big Data», ιδιαίτερη βαρύτητα δίδεται στο να εκτελείται αποτελεσματικά πάνω σε κρυπτογραφημένα δεδομένα, δηλαδή να μην χρειάζεται αυτά προηγουμένως να αποκρυπτογραφηθούν. Αυτό θα την καταστήσει ένα πανίσχυρο εργαλείο στο νέο αυτό περιβάλλον. Προς αυτή την κατεύθυνση έχουν προταθεί διάφορες τεχνικές, οι οποίες όμως βρίσκονται ακόμα σε ερευνητικό στάδιο. Συνεπώς, αν και η καθεμία τους επιτυγχάνει διαφορετικά επίπεδα όσον αφορά την προστασία της «ιδιωτικότητας», την απόδοση και την δυνατότητα εκτέλεσης επερωτήσεων [124], αυτές συνοπτικά είναι οι παρακάτω:

❖ Η «Property Preserving Encryption - PPE»

Η τεχνική αυτή βασίζεται στην ιδέα του να κρυπτογραφούνται τα δεδομένα με τέτοιο τρόπο, έτσι ώστε να διατηρούνται κάποιες ιδιότητες των αρχικών δεδομένων και στο

κρυπτογράφημα. Έτσι για παράδειγμα, αν «A>B» τότε θα είναι και «Enc(A)>Enc(B)», ή αν «A=B» τότε θα είναι και «Enc(A)=Enc(B)». Η τεχνική αυτή, αν και είναι αποδοτική και επιτρέπει την εκτέλεση πιο πολύπλοκων επερωτήσεων στα κρυπτογραφημένα δεδομένα, εντούτοις δεν εγγυάται πλήρως την προστασία της «Ιδιωτικότητας», καθώς είναι ευάλωτη στις επιθέσεις «Data Inference Attacks» [125].

❖ Η «Structured Encryption - SE»

Η τεχνική αυτή βασίζεται στην εκτέλεση επερωτήσεων, για την αναζήτηση κάθε φορά κάποιας λέξης/κλειδιού (Keyword) στο κρυπτογράφημα. Αν και είναι αποδοτική και εγγυάται την προστασία της «Ιδιωτικότητας», εντούτοις υστερεί στην εκφραστικότητα των επερωτήσεων. Μάλιστα, υπάρχουν δυο εκδοχές της ανάλογα με τον τύπο της κρυπτογράφησης. Η πρώτη είναι αν εφαρμόζεται συμμετρική κρυπτογράφηση, οπότε προκύπτει η «Symmetric Searchable Encryption - (SSE)» [126]. Σε αυτή κρυπτογραφούνται τα δεδομένα με κάποιο συμμετρικό αλγόριθμο κρυπτογράφησης, έτσι ώστε να επιτρέπεται μεταγενέστερα η αναζήτηση λέξεων/κλειδιών επί του κρυπτογραφήματος. Αυτό μπορεί να επιτευχθεί για παράδειγμα με την χρήση ενός ευρετηρίου, το οποίο κρυπτογραφείται με τον ίδιο αλγόριθμο. Οπότε, σε κάθε αναζήτηση ελέγχεται αν υπάρχει το αποτέλεσμα της κρυπτογράφησης της λέξης/κλειδιού στο ευρετήριο αυτό, προκειμένου να επιστραφούν τα αποτελέσματα της επερώτησης. Η μέθοδος αυτή είναι κυρίως χρήσιμη για αυτόν που κρυπτογράφησε τα δεδομένα και είναι αποδοτική στο περιβάλλον των «Big Data». Η δεύτερη είναι αν εφαρμόζεται ασύμμετρη κρυπτογράφηση, οπότε προκύπτει η «Public Key Searchable Encryption - (PEKS)» [102]. Σε αυτή τα δεδομένα κρυπτογραφούνται με κάποιο ασύμμετρο αλγόριθμο κρυπτογράφησης, έτσι ώστε να επιτρέπεται η μεταγενέστερη αναζήτηση λέξεων/κλειδιών επί του κρυπτογραφήματος. Μάλιστα, υποστηρίζει επερωτήσεις σύγκρισης (Comparison Queries), επερωτήσεις υποσυνόλων (Subset Queries), καθώς και οποιονδήποτε συνδυασμό αυτών. Επίσης μπορεί να χρησιμοποιηθεί από οποιονδήποτε χρήστη του συστήματος. Εντούτοις, αν και επιτρέπει καλύτερη εκφραστικότητα των επερωτήσεων συγκρινόμενη με την «SSE», υστερεί σε αποδοτικότητα και ασφάλεια σε σχέση με αυτή.

❖ Η «Relational Encryption - RE»

Η τεχνική αυτή επιτρέπει την εκτέλεση αναζητήσεων για διάφορες λέξεις/κλειδιά πάνω σε κρυπτογραφημένα δεδομένα, τα οποία όμως μπορεί να έχουν κρυπτογραφηθεί με διαφορετικά κλειδιά. Έτσι, σε αυτή το κάθε «Υποκείμενο» των δεδομένων μπορεί να χρησιμοποιήσει διαφορετικό κλειδί για την κρυπτογράφηση

των δεδομένων του. Παράλληλα, η τρίτη οντότητα, η οποία θα επωμιστεί το βάρος να εκτελέσει την αναζήτηση για λογαριασμό κάποιου χρήστη, δεν θα μπορεί να τα αποκρυπτογραφήσει. Συνεπώς, σε αυτή την τεχνική προστατεύεται η εμπιστευτικότητα των δεδομένων [127].

Επίσης, σημαντικό είναι να εκτελούνται και ασφαλείς υπολογισμοί επί των κρυπτογραφημένων δεδομένων στο περιβάλλον των «Big Data», χωρίς να απαιτείται αυτά προηγουμένως να έχουν αποκρυπτογραφηθεί, έτσι ώστε να προστατεύεται η «Ιδιωτικότητα» των «Υποκειμένων» των δεδομένων σε αυτό. Αν και σε αυτό τον τομέα έχουν προταθεί κάποιες τεχνικές λύσεις, εντούτοις αυτές βρίσκονται ακόμα σε ερευνητικό στάδιο. Επιπλέον, η απόδοση και η αποτελεσματικότητά τους είναι πολύ χαμηλές, ειδικά για το περιβάλλον των «Big Data». Ωστόσο, αυτές κρίνονται πολύ σημαντικές και ελπιδοφόρες. Συγκεκριμένα, οι τεχνικές αυτές είναι:

❖ Η «Homomorphic Encryption – HE»

Η «Ομοιομορφική Κρυπτογράφηση» (Homomorphic Encryption) βασίζεται στο ότι θα πρέπει το αποτέλεσμα μιας πράξης πάνω στο κρυπτογράφημα να είναι ακριβώς το ίδιο με αυτό που θα παραγόταν, αν η πράξη αυτή γινόταν στο αρχικό κείμενο, χωρίς όμως να έχει προηγουμένως διαμοιραστεί το μυστικό κλειδί, προκειμένου αυτά να αποκρυπτογραφηθούν για να εκτελεστεί η πράξη. Υπάρχουν δυο τύποι της τεχνικής αυτής. Η πρώτη είναι η «Πλήρης Ομοιομορφική Κρυπτογράφηση» (Fully Homomorphic Encryption -FHE) [103], η οποία επιτρέπει σε κάποιον να εκτελέσει οποιονδήποτε αριθμό πράξεων πάνω στα κρυπτογραφημένα δεδομένα, χωρίς να χρειάζεται προηγουμένως να τα αποκρυπτογραφήσει. Εντούτοις, η τεχνική αυτή δεν είναι υπολογιστικά αποδοτική στο περιβάλλον των «Big Data». Η δεύτερη είναι η «Μερική Ομοιομορφική Κρυπτογράφηση» (Somewhat or Partial Homomorphic Encryption – SHE or PHE) [104], η οποία είναι παρεμφερής με την προηγούμενη, καθώς ακολουθεί την ίδια λογική, ωστόσο υποστηρίζει μικρότερο αριθμό πράξεων σε σχέση με αυτή. Αν και αυτό βελτιώνει αισθητά την απόδοσή της συγκρινόμενη με την προηγούμενη, εντούτοις και αυτή ακόμα δεν μπορεί να χρησιμοποιηθεί στο περιβάλλον αυτό.

❖ Η «Oblivious RAM - ORAM»

Η τεχνική αυτή αποκρύπτει τις ταυτότητες των χρηστών που προσπελαύνουν τα δεδομένα κάποιου άλλου χρήστη, ο οποίος τα έχει αποθηκεύσει σε μια ή περισσότερες τοποθεσίες [128]. Με τον τρόπο αυτό, αφενός προστατεύονται οι ταυτότητες των χρηστών και αφετέρου αποτρέπεται η αποκάλυψη κάποιου

δεδομένου, ή η εξαγωγή συμπερασμάτων από την ανάλυση του μοτίβου της συχνότητας προσπέλασης των δεδομένων αυτών. Δηλαδή η τεχνική αυτή είναι ανθεκτική στις «Data Inference Attacks». Παράλληλα, επιτρέπει την εκτέλεση ασφαλών υπολογισμών πάνω σε κρυπτογραφημένα δεδομένα, καθιστώντας την ιδιαίτερα σημαντική για το περιβάλλον των «Big Data». Ωστόσο, και αυτή έχει ακόμα χαμηλή απόδοση για το περιβάλλον αυτό.

❖ Η «Secure Multi Party Computation - SMPC»

Η τεχνική αυτή επιτρέπει σε διαφορετικούς χρήστες να υπολογίσουν το αποτέλεσμα μιας πράξης, η οποία θα πάρει ως είσοδο τα δεδομένα του κάθε χρήστη, χωρίς όμως αυτά να αποκαλυφθούν στους υπόλοιπους χρήστες που συμμετέχουν στον υπολογισμό αυτό. Έτσι η τεχνική αυτή επιτρέπει για παράδειγμα να υπολογίσουν τρείς χρήστες «X», «Y», «Z» το ποιός παίρνει τον μεγαλύτερο μισθό (max(x,y,z)), χωρίς όμως να αποκαλύψουν ο ένας στον άλλο τον μισθό του. Αυτό το σενάριο γενικεύεται για περισσότερους χρήστες, με πολλαπλά δεδομένα εισόδου και εξόδου [137]. Υπάρχουν διάφοροι τρόποι για την επίτευξη του, όπως το «Zero Knowledge Proof», το «Oblivious Transfer», το «Yao's Millionaire Protocol» κλπ. Ωστόσο, η τεχνική αυτή παρουσιάζει αδυναμία ως προς την προστασία της «Ιδιωτικότητας» στο περιβάλλον των «Big Data», καθώς είναι ευάλωτη σε επιθέσεις συνωμοσίας (Collusion Attacks). Επίσης, αντιμετωπίζει και άλλα προβλήματα, τα οποία την καθιστούν μη χρησιμοποιήσιμη στο περιβάλλον αυτό.

Συνεπώς, απαιτείται περαιτέρω έρευνα, αφενός για την εύρεση νέων και απλούστερων τεχνικών και αφετέρου για την βελτίωση των ήδη υπαρχουσών, έτσι ώστε αυτές να υποστηρίζουν την εκτέλεση ασφαλών υπολογισμών πάνω στα κρυπτογραφημένα δεδομένα στο περιβάλλον των «Big Data».

Τέλος, η προστασία της ακεραιότητας των δεδομένων προϋποθέτει να έχει προηγουμένως ταυτοποιηθεί η πηγή τους. Αυτό όμως αντιβαίνει στην προστασία της «Ιδιωτικότητας». Εντούτοις, είναι δυνατό να βρεθεί μια μέση λύση, η οποία θα εγγυάται τον έλεγχο της ταυτότητας μιας οντότητας και παράλληλα θα προστατεύει εν μέρει την ανωνυμία της. Συνεπώς, όσον αφορά την προστασία της ακεραιότητας των δεδομένων και παράλληλα την προστασία της «Ιδιωτικότητας» των «Υποκειμένων» τους στο περιβάλλον των «Big Data», έχουν προταθεί οι παρακάτω τεχνικές:

❖ Των «Group Signatures»

Η τεχνική αυτή επιτρέπει σε κάποια οντότητα, που παράλληλα συμμετέχει και σε κάποια ομάδα, να υπογράψει τα δεδομένα της με τέτοιο τρόπο, έτσι ώστε να μπορεί

να αναγνωρίζεται από τις υπόλοιπες οντότητες μόνο μέσω της ομάδας στην οποία ανήκει. Ο μοναδική οντότητα που θα μπορεί άμεσα να εντοπίσει την ταυτότητα της οντότητας αυτής θα είναι μια «Έμπιστη Τρίτη Οντότητα» (Trusted Third Party). Επιπλέον, αυτή είναι υπεύθυνη για την δημιουργία, αλλαγή και διαγραφή των ομάδων, καθώς και για τον υπολογισμό και των διαμοιρασμό των ομαδικών κλειδιών στους χρήστες της εκάστοτε ομάδας [105]. Ωστόσο, οι αλγόριθμοι για την υλοποίησή της είναι υπολογιστικά μη αποδοτικοί για το περιβάλλον των «Big Data».

❖ Των «Ring Signatures»

Η τεχνική αυτή είναι μια παραλλαγή της προηγούμενης, στην οποία όμως υπάρχουν μόνο χρήστες και όχι «Έμπιστες Τρίτες Οντότητες». Αυτό σημαίνει ότι σε αυτή δεν υπάρχουν προκαθορισμένες ομάδες χρηστών και διαδικασίες για τον ορισμό, την αλλαγή ή την διαγραφή κάποιας ομάδας. Επιπλέον δεν υπάρχει τρόπος να διανεμηθούν κάποια ειδικά κλειδιά, όπως δεν υπάρχει και τρόπος να ανακληθεί η ανωνυμία του πραγματικού υπογράφοντα. Συνεπώς, το μόνο που διασφαλίζει η τεχνική αυτή είναι η ανωνυμία του κάθε μέλος της, στα πλαίσια του δημοσίου κλειδιού που το ίδιο θα φτιάξει. Συγκεκριμένα, για να παραχθεί αυτό το δημόσιο κλειδί, θα πρέπει ο πραγματικός υπογράφων να επιλέξει ένα αυθαίρετο σύνολο από πιθανούς υπογράφοντες, μέσα στους οποίους συμπεριλαμβάνει και τον εαυτό του. Στη συνέχεια, θα πρέπει να υπολογίσει το τελικό δημόσιο κλειδί του, χρησιμοποιώντας μόνο το ιδιωτικό του κλειδί και τα δημόσια κλειδιά των υπολοίπων οντοτήτων που επέλεξε αυθαίρετα [129]. Ωστόσο, και αυτή η τεχνική αποτελεί ανοιχτό ερευνητικό πεδίο για το περιβάλλον των «Big Data».

3.4.4 ΕΛΕΓΧΟΣ ΠΡΟΣΒΑΣΗΣ (ACCESS CONTROL)

Παράλληλα με την κρυπτογραφία και η χρήση των «Μηχανισμών Ελέγχου Πρόσβασης» (Access Control Mechanisms) είναι εξίσου σημαντική για την προστασία των προσωπικών δεδομένων. Οι μηχανισμοί αυτοί έχουν σχεδιαστεί για να ελέγχουν την πρόσβαση στα δεδομένα και ως εκ τούτου να προστατεύουν το απόρρητό τους. Έτσι, αφενός απαγορεύουν την πρόσβαση στις οντότητες, οι οποίες δεν έχουν τέτοιο δικαίωμα, και αφετέρου την επιτρέπουν σε αυτές, οι οποίες έχουν την εξουσιοδότηση να τα προσπελάσουν. Μάλιστα, στους μηχανισμούς αυτούς υλοποιούνται σειριακά οι διαδικασίες της ταυτοποίησης και της πιστοποίησης της ταυτότητας μιας οντότητας (Identification and Authentication), του ελέγχου της εξουσιοδότησής της για πρόσβαση στα δεδομένα (Authorization), της έγκρισης της

πρόσβασης σε αυτά, με την προϋπόθεση ότι είναι εξουσιοδοτημένη (Access Approval), και της καταγραφής των ενεργειών της για λόγους λογοδοσίας (Accountability) [188]. Επομένως, οι μηχανισμοί αυτοί χρησιμοποιούνται ανά περίπτωση, προκειμένου να μετριαστεί από την μια πλευρά η απαίτηση να διαμοιραστούν τα δεδομένα με άλλες οντότητες, οι οποίες είναι εξουσιοδοτημένες να τα επεξεργαστούν, και από την άλλη πλευρά αυτά να προστατευτούν από τις μη εξουσιοδοτημένες οντότητες, οι οποίες θα θελήσουν να τα αποκτήσουν για να τα χρησιμοποιήσουν για διάφορους ενδεχομένως κακόβουλους σκοπούς.

Ειδικότερα για το περιβάλλον των «Big Data», η επίτευξη του βέλτιστου μετριασμού των παραπάνω απαιτήσεων είναι ένα σοβαρό πρόβλημα. Εκ σχεδιασμού, η λειτουργία των «Big Data Analytics» προϋποθέτει τον διαμοιρασμό και την επεξεργασία των συνόλων δεδομένων που προέρχονται από διαφορετικές και ανεξάρτητες μεταξύ τους πηγές. Επίσης, είναι προφανές ότι ο αριθμός των οντοτήτων του συστήματος αυτού είναι εξαιρετικά μεγάλος. Επιπλέον, στον τεράστιο αυτό όγκο δεδομένων συμπεριλαμβάνονται πλήθος προσωπικών δεδομένων από τις διαφορετικές αυτές οντότητες, τα οποία θα πρέπει να προστατεύονται σε τέτοιο βαθμό, όσο επιβάλλει η εφαρμογή των απαιτήσεων, τόσο των αντίστοιχων «Υποκειμένων» τους, όσο και αυτών που πηγάζουν από την νομοθεσία, τα συμβόλαια και τις πολιτικές ασφαλείας. Έτσι, η χρήση ενός τυπικού «Μηχανισμού Ελέγχου Πρόσβασης» μπορεί να αποτρέψει τον διαμοιρασμό των αρχείων, τα οποία υπό άλλες συνθήκες θα ήταν απαραίτητα για την διαδικασία της «Αναλυτικής» και θα διαμοιράζονταν, καθώς δεν θα υπήρχε πρόβλημα για την προστασία του απορρήτου τους. Συνεπώς, στο περιβάλλον των «Big Data» προκύπτει η ανάγκη να υλοποιηθεί ένας «Μηχανισμός Ελέγχου Πρόσβασης με Μεγάλο Βαθμό Λεπτομέρειας» (Granular Access Control Mechanism), ο οποίος θα παρέχει την δυνατότητα να διαμοιράζονται τα δεδομένα του συστήματος αυτού στις διάφορες οντότητές του με μεγαλύτερη ακρίβεια, χωρίς παράλληλα να παραβιάζεται το απόρρητό τους.

Η βασική ιδέα για την επαύξηση του βαθμού λεπτομέρειας των «Μηχανισμών Ελέγχου Πρόσβασης» έγκειται στην αξιολόγηση κάθε φορά περισσότερων στοιχείων, αφενός για την έγκριση της πρόσβασης μιας οντότητας στα δεδομένα και αφετέρου για τον ακριβή καθορισμό σε ποια από αυτά θα έχει πρόσβαση και με τι δικαιώματα για το καθένα. Για να επιτευχθεί αυτό λοιπόν, είναι αναγκαίο να καταγράφονται λεπτομερώς όλες εκείνες οι πληροφορίες που σχετίζονται, τόσο με τις απαιτήσεις

απορρήτου της κάθε πληροφορίας ξεχωριστά, όσο και με τα ιδιαίτερα χαρακτηριστικά της κάθε οντότητας του συστήματος, έτσι ώστε αυτά να αξιολογούνται υπό το πρίσμα της εκάστοτε περίπτωσης, αλλά και του επιδιωκόμενου σκοπού της, προκειμένου να ληφθεί η κρίσιμη απόφαση για το αν θα δοθεί δικαίωμα πρόσβασης στα δεδομένα, σε ποιά συγκεκριμένα από αυτά και με τι δικαιώματα για το καθένα.

Έτσι για παράδειγμα, προκειμένου να γίνει πιο σαφής η λειτουργία του μηχανισμού αυτού, ας εξετάσουμε την λειτουργία του σε διάφορες περιπτώσεις, στις οποίες κάποιος ιατρός ενός συγκεκριμένου τμήματος μιας υγειονομικής υπηρεσίας θέλει να αποκτήσει πρόσβαση στα ιατρικά αρχεία ενός ασθενή. Ως γνωστόν τα ιατρικά αρχεία περιέχουν ευαίσθητα προσωπικά δεδομένα και ως εκ τούτου οι απαιτήσεις απορρήτου τους καθορίζονται, τόσο από τις απαιτήσεις του ίδιου του «Υποκειμένου» τους, όσο και από την κείμενη νομοθεσία, την ισχύουσα πολιτική ασφαλείας και το συμβόλαιο, που έχει υπογραφεί μεταξύ της υγειονομικής υπηρεσίας και του «Υποκειμένου» αυτού. Στην περίπτωση λοιπόν, που ο ιατρός αυτός θέλει να αποκτήσει πρόσβαση στα αρχεία αυτά με σκοπό να εξετάσει την υγεία του ασθενούς, ο μηχανισμός αυτός, αφού αξιολογήσει όλες τις απαιτήσεις απορρήτου των δεδομένων αυτών, τα συγκεκριμένα χαρακτηριστικά του ιατρού, όπως για παράδειγμα την ειδικότητά του και το αν εργάζεται στο συγκεκριμένο τμήμα της υπηρεσίας, καθώς και τον σκοπό του, που στην συγκεκριμένη περίπτωση είναι η εξέταση, θα του δώσει πρόσβαση μόνο στα απαραίτητα στοιχεία που του χρειάζονται για τον σκοπό αυτό και όχι σε όλα και μάλιστα με διαφορετικά δικαιώματα για το καθένα από αυτά. Τέτοια στοιχεία μπορεί να είναι το ιστορικό, η πάθηση, η φαρμακευτική αγωγή και το ΑΜΚΑ του ασθενούς. Αν στη συνέχεια ένας άλλος ιατρός της ίδιας ειδικότητας και του ίδιου τμήματος θέλει να αποκτήσει πρόσβαση στα αρχεία αυτά, με σκοπό να διεξάγει έρευνα για κάποια ασθένεια σχετική με την ειδικότητά του, τότε ο ίδιος ο μηχανισμός, αφού εξετάσει εκ νέου τις απαιτήσεις απορρήτου των δεδομένων αυτών, τα χαρακτηριστικά του ιατρού αυτού, και τον σκοπό του, δεν θα του επιτρέψει την πρόσβαση σε όλες τις πληροφορίες της προηγούμενης περίπτωσης, παρά μόνο σε αυτές που του χρειάζονται για την συγκεκριμένη έρευνα, ήτοι την ανάγνωση μόνο της ασθένειας, προκειμένου να χρησιμοποιηθεί για την εξαγωγή του στατιστικών της έρευνας. Τέλος, αν οποιοσδήποτε άλλος ιατρός άλλης ειδικότητας αλλά του ίδιου τμήματος θέλει να αποκτήσει πρόσβαση στα αρχεία αυτά, ενδεχομένως ο μηχανισμός να του απαγορεύσει την πρόσβαση αυτή.

Επομένως, η ένταξη των «Μηχανισμών Ελέγχου Πρόσβασης με Μεγάλο Βαθμό Λεπτομέρειας» στο ιδιαίτερο περιβάλλον των «Big Data» αυξάνει την δυνατότητα διαμοιρασμού των δεδομένων, ενώ παράλληλα προστατεύει στον καλύτερο δυνατό βαθμό το απόρρητό τους. Εντούτοις όμως, τα ιδιαίτερα χαρακτηριστικά του περιβάλλοντος αυτού, ήτοι ο όγκος των δεδομένων, το πλήθος των οντοτήτων, οι διαφορετικές απαιτήσεις απορρήτου για τα δεδομένα και η πολυπλοκότητα των συστημάτων αυτών, εισάγουν νέες προκλήσεις όσον αφορά τον σχεδιασμό, την υλοποίηση και την χρήση των μηχανισμών αυτών. Μάλιστα, είναι προφανές ότι όσο αυξάνει ο βαθμός λεπτομέρειας των μηχανισμών αυτών, τόσο αυξάνει και η πολυπλοκότητα των προκλήσεων, ακριβώς λόγω των χαρακτηριστικών του περιβάλλοντος των «Big Data». Συγκεκριμένα, οι προκλήσεις αυτές είναι οι εξής [1]:

1) Η Διαχείριση των Απαιτήσεων Απορρήτου της Κάθε Πληροφορίας του Συστήματος.

Η πρόκληση αυτή σχετίζεται με τις ενέργειες, αφενός της παρακολούθησης-ενημέρωσης και αφετέρου της διάθεσης-τήρησης των απαιτήσεων απορρήτου του κάθε στοιχείου πληροφορίας. Αναλυτικότερα, το πρώτο πρόβλημα αφορά την παρακολούθηση και την ενημέρωση των απαιτήσεων απορρήτου, που συνοδεύουν το κάθε στοιχείο δεδομένων. Μάλιστα, ένα επιπλέον πρόβλημα που τίθεται είναι αυτό του καθορισμού των απαιτήσεων απορρήτου για το στοιχείο εκείνο, το οποίο θα προκύψει από την επεξεργασία δυο ή περισσότερων άλλων στοιχείων. Το δεύτερο πρόβλημα σχετίζεται με την διάθεση κάθε φορά των απαιτήσεων απορρήτου στα ενδιαφερόμενα μέρη, προκειμένου να ληφθεί απόφαση για το αν και πως τα δεδομένα στα οποία αναφέρονται θα διαμοιραστούν στις διάφορες οντότητες του συστήματος, που θα θέλουν να αποκτήσουν πρόσβαση σε αυτά. Αυτό συνδέεται άμεσα με το πρόβλημα της τήρησης των ετικετών μαζί με τα δεδομένα στα οποία αντιστοιχούν, τόσο κατά τις διάφορες φάσεις της ανάλυσης, όσο και κατά τους ποικίλους μετασχηματισμούς τους.

2) Η Τήρηση, η Ενημέρωση και η Διάθεση των Χαρακτηριστικών των Διαφόρων Οντοτήτων του Συστήματος.

Η πρόκληση αυτή σχετίζεται με την ασφαλή αποθήκευση, ενημέρωση και κοινοποίηση των χαρακτηριστικών των διαφόρων οντοτήτων του συστήματος, καθώς και των ρόλων και των εξουσιοδοτήσεών τους. Επιπρόσθετα, υπάρχει περίπτωση κάποια οντότητα να έχει διαφορετικές εξουσιοδοτήσεις στο σύστημα «Big Data», λόγω των διαφορετικών ρόλων που μπορεί να έχει στα επιμέρους ανεξάρτητα

τμήματά του. Αυτό δημιουργεί σημαντικό πρόβλημα για την προστασία του απορρήτου των δεδομένων, αλλά και για την σωστή λειτουργία του μηχανισμού ελέγχου πρόσβασης. Συνεπώς, ο καθορισμός με σαφήνεια του εύρους των εξουσιοδοτήσεων για μια οντότητα, το οποίο θα ισχύει σε ολόκληρο το σύστημα «Big Data» και το οποίο θα σέβεται τους διαφορετικούς ρόλους της οντότητας αυτής, είναι ένα επιπλέον πρόβλημα.

3) Ο Σωστός Σχεδιασμός και Υλοποίηση του Μηχανισμού Ελέγχου Πρόσβασης σε Επίπεδο Εφαρμογής.

Η τρίτη πρόκληση σχετίζεται με την σωστό σχεδιασμό και υλοποίηση της λειτουργίας του μηχανισμού ελέγχου πρόσβασης. Ο μηχανισμός ελέγχου πρόσβασης είναι ένα λογικό φίλτρο, το οποίο, αφού λάβει υπόψη τις απαιτήσεις απορρήτου που σχετίζονται με τα δεδομένα και τα χαρακτηριστικά των οντοτήτων του συστήματος, θα λάβει απόφαση για το αν θα επιτραπεί ή όχι η πρόσβαση, σε ποια δεδομένα και με τι δικαιώματα για το καθένα από αυτά. Επίσης, λόγω του ότι ο μηχανισμός αυτός συνήθως υλοποιείται σε επίπεδο εφαρμογών και του ότι το πλήθος των εφαρμογών στο περιβάλλον των «Big Data» είναι μεγάλο, δημιουργείται μια επιπλέον πρόκληση, η οποία οφείλεται κυρίως στον ανθρώπινο παράγοντα και που αφορά την πιθανότητα να μην υλοποιηθούν σωστά οι μηχανισμοί αυτοί σε όλες τις εφαρμογές του συστήματος.

Συνεπώς, για να επιτύχουν τους στόχους τους οι μηχανισμοί αυτοί και να είναι αποτελεσματικοί, απαιτείται η επίλυση των ανωτέρω προκλήσεων κατά τις φάσεις του σχεδιασμού και της ανάπτυξή τους. Μάλιστα, απαραίτητη προϋπόθεση για την υιοθέτηση των διαφόρων προτεινόμενων λύσεων [51] είναι να έχει καθοριστεί σαφώς ο βαθμός λεπτομέρειας, ο οποίος σχετίζεται, αφενός με τις απαιτήσεις απορρήτου της κάθε πληροφορίας ξεχωριστά και αφετέρου με τα ιδιαίτερα χαρακτηριστικά, τους ρόλους και τις εξουσιοδοτήσεις της κάθε οντότητας του συστήματος. Σκοπός κάθε φορά είναι να καθορίζεται τέτοιος βαθμός λεπτομέρειας, όσος χρειάζεται προκειμένου να αυξηθεί η ακρίβεια στον διαμοιρασμό των δεδομένων στις διάφορες οντότητες του συστήματος, χωρίς παράλληλα να παραβιάζεται το απόρρητό τους.

Συγκεκριμένα, οι απαιτήσεις απορρήτου των δεδομένων αποθηκεύονται σε ετικέτες για το κάθε στοιχείο δεδομένων και το ακολουθούν στις διάφορες φάσεις της επεξεργασίας του. Είναι σημαντικό να επισημανθεί, ότι οι απαιτήσεις απορρήτου εξαρτώνται κάθε φορά από το ποιό είναι το ελάχιστο στοιχείο δεδομένων. Έτσι, η

απαίτηση απορρήτου σε επίπεδο γραμμής, όπου η γραμμή αναπαριστά τις τιμές των χαρακτηριστικών μιας συγκεκριμένης οντότητας, συχνά συνδέεται με τις απαιτήσεις της οντότητας αυτής. Ενώ η απαίτηση απορρήτου σε επίπεδο στήλης, όπου η στήλη αναπαριστά ένα συγκεκριμένο σύνολο τιμών κάποιου χαρακτηριστικού όλων των οντοτήτων, συνήθως συνδέεται με την ευαισθησία του χαρακτηριστικού αυτού, όπως για παράδειγμα αν αυτό ανήκει στην κατηγορία των ευαίσθητων προσωπικών δεδομένων. Από την άλλη πλευρά, ο συνδυασμός των δυο παραπάνω περιπτώσεων, αν και προσδίδει μεγαλύτερο βαθμό λεπτομέρειας, εντούτοις είναι πιθανό να καταρρεύσει κατά την διαδικασία της ανάλυσης. Επιπλέον, ο μετασχηματισμός των πινάκων ή των συνόλων δεδομένων σε όψεις, για την απάντηση σε κάποια επερώτηση, συνήθως δεν εμπίπτει στις ανωτέρω κατηγορίες και επομένως απαιτείται η εύρεση άλλης κατάλληλης λύσης για τον ορισμό των απαιτήσεων απορρήτου στην περίπτωση αυτή. Τέλος, οι απαιτήσεις απορρήτου σε επίπεδο κελιού, όπου το κελί αναπαριστά την ελάχιστη ποσότητα πληροφορίας, μπορεί να συμπεριλάβουν τις απαιτήσεις, τόσο των «Υποκειμένων» τους, όσο και αυτών που επιβάλλουν η νομοθεσία, τα συμβόλαια και οι πολιτικές ασφαλείας, υποστηρίζοντας ταυτόχρονα ένα ευρύ σύνολο μετασχηματισμών επί των δεδομένων αυτών, χωρίς να υπάρχει κίνδυνος κατάρρευσής τους κατά τις διάφορες φάσεις της ανάλυσης [1].

Έτσι, όσον αφορά το πρόβλημα της διάθεσης και της διατήρησης των ετικετών με τα αντίστοιχα δεδομένα τους, προκειμένου να χρησιμοποιηθούν άμεσα οποτεδήποτε χρειαστεί να ληφθεί απόφαση για την έγκριση της πρόσβασης στα δεδομένα αυτά, είναι αναγκαίο αυτά να παραμένουν συνενωμένα μεταξύ τους σε όλο τον κύκλο ζωής των δεδομένων αυτών. Μάλιστα, στην περίπτωση που θα προκύψει ένα νέο στοιχείο δεδομένων από την επεξεργασία δυο ή περισσότερων άλλων στοιχείων, τα οποία μπορεί να προέρχονται από διαφορετικές πηγές, είναι επιτακτικό, αφενός να καθοριστεί με σαφήνεια ποιες θα είναι οι νέες απαιτήσεις απορρήτου για αυτό και αφετέρου θα πρέπει αυτές να αποθηκευτούν σε αντίστοιχη ετικέτα, η οποία άμεσα θα συνενωθεί με το στοιχείο αυτό. Ένα ζήτημα, που εγείρεται στην περίπτωση αυτή, είναι πως θα προκύψουν οι νέες απαιτήσεις και τι σχέση θα έχουν με αυτές των συστατικών του στοιχείων. Συγκεκριμένα, αυτό το στοιχείο θα έχει ως απαιτήσεις απορρήτου τις ισχυρότερες από τις απαιτήσεις απορρήτου των συστατικών του στοιχείων, τις λιγότερο ισχυρές από αυτές, ή θα του καθοριστούν νέες, οι οποίες δεν θα συμπίπτουν με αυτές των συστατικών του στοιχείων. Προφανώς, το τι θα

εφαρμοστεί εξαρτάται από την εκάστοτε περίπτωση, ωστόσο ο σκοπός, δηλαδή η διασφάλιση του απορρήτου, τόσο του ίδιου του στοιχείου, όσο και των συστατικών του, παραμένει ίδιος. Ένα άλλο ζήτημα είναι αν θα πρέπει να συνενωθούν και να συνοδεύουν το νέο αυτό στοιχείο όλες συνολικά οι ετικέτες των επιμέρους συστατικών του για λόγους παρακολούθησης της προέλευσής του. Γενικά, στο πλείστο των περιπτώσεων, αυτό είναι χρήσιμο και αναγκαίο, καθώς η τήρησή τους συνεπάγεται εν μέρει την καταγραφή και την παρακολούθηση της προέλευσής του, ακόμα και μετά από πολλούς μετασχηματισμούς. Το πρόβλημα όμως, που προκύπτει από την συνένωση αυτή, είναι ποια ετικέτα από αυτές που συνοδεύουν το στοιχείο αυτό θα ισχύει κάθε φορά, έτσι ώστε να αξιολογηθούν οι αντίστοιχες απαιτήσεις απορρήτου από τον εκάστοτε μηχανισμό ελέγχου πρόσβασης, προκειμένου να ληφθεί η απόφαση για την εξουσιοδότηση της πρόσβασης στο δεδομένο αυτό. Για να επιλυθεί το πρόβλημα αυτό, θα πρέπει να χρησιμοποιηθεί ένας κατάλληλος μηχανισμός, ο οποίος θα αποσαφηνίζει ποια είναι η βασική απαίτηση απορρήτου για το στοιχείο αυτό με την επισήμανση της αντίστοιχης ετικέτας.

Από την άλλη πλευρά, η παρακολούθηση των απαιτήσεων απορρήτου και η ενημέρωσή τους είναι εξίσου σημαντική για την αποτελεσματικότητα του μηχανισμού ελέγχου πρόσβασης. Οι απαιτήσεις απορρήτου μπορεί να μεταβάλλονται με την πάροδο του χρόνου, οπότε θα πρέπει ο μηχανισμός να έχει την δυνατότητα να παρακολουθεί και να προσαρμόζεται γρήγορα στις μεταβολές αυτές για να λειτουργεί σωστά, χωρίς όμως αυτό να συνεπάγεται καθυστερήσεις στην λειτουργία του. Συνεπώς, θα πρέπει να σχεδιαστεί ένας μηχανισμός, στον οποίο θα αποθηκεύονται στις ετικέτες οι απαιτήσεις εκείνες, οι οποίες μεταβάλλονται δύσκολα με την πάροδο του χρόνου, ενώ αντίστοιχα θα ελέγχονται κατά την φάση του ελέγχου της εξουσιοδότησης πρόσβασης οι απαιτήσεις αυτές που μεταβάλλονται συχνά. Επίσης είναι εξίσου σημαντικό να ελέγχεται τακτικά η εγκυρότητα και η ισχύς των απαιτήσεων απορρήτου, που είναι αποθηκευμένες στις ετικέτες, αν και υποτίθεται ότι αυτές δεν μεταβάλλονται συχνά, έτσι ώστε να μην γίνονται λανθασμένες υποθέσεις κατά την φάση του ελέγχου της εξουσιοδότησης πρόσβασης.

Όσον αφορά τα χαρακτηριστικά, τους ρόλους και τις εξουσιοδοτήσεις των διαφόρων οντοτήτων του συστήματος, θα πρέπει κατ' αρχάς αυτά να ελέγχονται τακτικά και να μην θεωρούνται αμετάβλητα. Οπότε, αν διαπιστωθεί τροποποίηση, θα

πρέπει έγκαιρα αυτά να αναθεωρηθούν, έτσι ώστε να επανακαθοριστούν, τόσο οι ρόλοι και οι εξουσιοδοτήσεις των οντοτήτων, όσο και τα κλειδιά και τα χαρακτηριστικά τους. Επίσης, σημαντική είναι η ασφαλής αποθήκευση των στοιχείων αυτών σε κάποια βάση δεδομένων, η οποία θα προστατεύεται από ένα ισχυρό μηχανισμό ελέγχου πρόσβασης και η οποία θα είναι κρυπτογραφημένη. Παράλληλα, ιδιαίτερη προσοχή θα πρέπει να δοθεί και στην ασφάλεια των επικοινωνιών της βάσης αυτής, κατά την αποστολή των στοιχείων αυτών για την λειτουργία των μηχανισμών ελέγχου πρόσβασης.

Επιπλέον, η χρησιμοποίηση των μηχανισμών «Single Sign On - SSO» θα μειώσει σημαντικά την διαχειριστική επιβάρυνση της υποστήριξης αυτού του μεγάλου πλήθους των οντοτήτων στα συστήματα αυτά. Σκοπός τους είναι να διευρυνθεί η δυνατότητα διαμοιρασμού των δεδομένων, χωρίς όμως να αυξηθεί σημαντικά το διαχειριστικό κόστος. Έτσι μετά την επιτυχή αυθεντικοποίηση μιας οντότητας, τα χαρακτηριστικά, οι ρόλοι και οι εξουσιοδοτήσεις της θα αντληθούν από μια ή περισσότερες αξιόπιστες πηγές μέσω συστημάτων όπως το «LDAP», το «Active Directory», το «OAuth», το «OpenID» και πολλά άλλα [191].

Για να αντιμετωπιστεί η τρίτη πρόκληση, θα πρέπει να δοθεί ιδιαίτερη προσοχή, τόσο κατά την σχεδίαση, όσο και κατά την ανάπτυξη του μηχανισμού ελέγχου πρόσβασης. Για την φάση της σχεδίασης, το πρώτο ζήτημα σχετίζεται με την απαίτηση να μην παραληφθεί κανένα στοιχείο, που θα του είναι απαραίτητο για την αποτελεσματική λειτουργία του και την επίτευξη του επιθυμητού επιπέδου λεπτομέρειας. Για αυτό, θα πρέπει κατ' αρχάς να διαχωριστούν μεταξύ τους τα στοιχεία που σχετίζονται με τα δεδομένα, τις οντότητες και το εκάστοτε περιβάλλον της εφαρμογής και στην συνέχεια αυτά να αποθηκευτούν ανά κατηγορία, για να είναι πιο εύκολη η διαχείριση και ο έλεγχός τους. Έτσι, αυτά που σχετίζονται με τα δεδομένα θα πρέπει να αποθηκευτούν στο σύστημα των ετικετών, αυτά που σχετίζονται με τις οντότητες θα πρέπει να τηρούνται σε μια ασφαλή βάση δεδομένων και αυτά που σχετίζονται με την εφαρμογή θα πρέπει να παρέχονται τοπικά από την ίδια την εφαρμογή. Με τον τρόπο αυτό πετυχαίνουμε δυο πλεονεκτήματα. Αφενός η παρακολούθηση, ο έλεγχος, η διαχείριση και η διάθεσή τους θα γίνεται με αποδοτικότερο τρόπο, καθώς οι ενέργειες αυτές θα υποστηρίζονται από τα κατάλληλα εργαλεία της κάθε κατηγορίας. Αφετέρου, η κατάτμησή τους σε κατηγορίες καθιστά πιο εύκολη την αντιμετώπιση των επιμέρους προβλημάτων που μπορεί να

προκύψουν. Συνεπώς, ο τρόπος αυτός εγγυάται ότι κάθε φορά θα διατίθενται στον μηχανισμό ελέγχου πρόσβασης τα απαραίτητα εκείνα στοιχεία από τα αντίστοιχα εργαλεία της κάθε κατηγορίας για την αποτελεσματική του λειτουργία.

Το δεύτερο πρόβλημα αφορά την μοντελοποίηση της πιθανότητας να τροποποιηθεί κάποιο από τα παραπάνω στοιχεία με την πάροδο του χρόνου. Για παράδειγμα, υπάρχουν στοιχεία σχετικά αμετάβλητα, όπως οι απαιτήσεις απορρήτου, και στον αντίποδα στοιχεία ιδιαίτερα ευμετάβλητα, όπως αυτά που σχετίζονται με τον ρόλο και τις ιδιότητες κάποιας οντότητας. Έτσι, για να αντιμετωπιστεί αποτελεσματικά το πρόβλημα της απόδοσης, τόσο κατά την εγγραφή, όσο και κατά την ανάγνωση των στοιχείων αυτών, η προτεινόμενη λύση είναι να «Κανονικοποιηθούν» (Normalization) [189] οι ευμετάβλητες απαιτήσεις, για να μειωθεί το κόστος της ενημέρωσής τους κατά την εγγραφή τους, και να «Αποκανονικοποιηθούν» (De-normalization) [190] αντίστοιχα τα αμετάβλητα στοιχεία, για να βελτιωθεί η απόδοση κατά την ανάγνωσή τους. Οπότε, οι απαιτήσεις απορρήτου αποκανονικοποιούνται και αποθηκεύονται στις ετικέτες των δεδομένων, ενώ τα στοιχεία των χρηστών κανονικοποιούνται και αποθηκεύονται στην σχεσιακή βάση δεδομένων. Κατά την λειτουργία του μηχανισμού ελέγχου πρόσβασης, όλα τα στοιχεία αυτά θα συνενωθούν δυναμικά μεταξύ τους, προκειμένου να ληφθεί σε ικανοποιητικό χρόνο η απόφαση για την εξουσιοδότηση πρόσβασης στα δεδομένα.

Όσον αφορά την φάση της ανάπτυξης, το πρόβλημα που τίθεται είναι η πιθανότητα να μην υλοποιηθούν σωστά οι μηχανισμοί αυτοί σε κάποιες από τις εφαρμογές του συστήματος των «Big Data». Για να αντιμετωπιστεί το πρόβλημα αυτό, θα πρέπει να καθοριστούν επακριβώς κατά την φάση του σχεδιασμού τα βήματα για την υλοποίησή τους, έτσι ώστε αυτά να ακολουθούνται επακριβώς από τις διάφορες ομάδες ανάπτυξης λογισμικού. Επίσης υπάρχει σοβαρό ενδεχόμενο να παραβιαστεί το απόρρητο των δεδομένων εξαιτίας της μη σωστής υλοποίησης των μηχανισμών αυτών. Οι μηχανισμοί αυτοί επιτρέπουν την πρόσβαση κάθε φορά σε συγκεκριμένα δεδομένα, ανάλογα με τον εκάστοτε σκοπό, τα χαρακτηριστικά της οντότητας και τις απαιτήσεις απορρήτου των δεδομένων αυτών. Κανονικά, αν αλλάξει οποιοδήποτε από αυτά τα στοιχεία, θα πρέπει να αξιολογηθεί εκ νέου το αίτημα της πρόσβασης στα δεδομένα. Ωστόσο, αν ο μηχανισμός δεν έχει υλοποιηθεί σωστά, ώστε να λειτουργεί με αυτόν τον τρόπο, υπάρχει κίνδυνος να παραβιαστεί το απόρρητο των δεδομένων. Συνεπώς θα πρέπει να διασφαλίζεται ότι, κάθε φορά που

ολοκληρώνεται μια συνεδρία, θα διαγράφονται τα στοιχεία και τα δεδομένα που χρησιμοποιήθηκαν σε αυτή, πριν ξεκινήσει η καινούρια.

Επιπρόσθετα, είναι αναγκαίο να αναπτυχθεί ένα πρωτόκολλο για την διαχείριση των πολιτικών ασφαλείας και των απαιτήσεων απορρήτου, έτσι ώστε να υποστηρίξει την λειτουργία του μηχανισμού ελέγχου πρόσβασης ως προς τον τομέα αυτό. Ως γνωστόν, οι πολιτικές ασφαλείας και οι απαιτήσεις απορρήτου συνήθως είναι πολύπλοκες, γεγονός που καθιστά δύσκολη την εφαρμογή τους με ακρίβεια 100%. Παράλληλα, είναι σημαντικό αυτές να επανεξετάζονται τακτικά και να αναθεωρούνται, όταν κρίνεται απαραίτητο. Συνεπώς, το πρωτόκολλο, που θα χρησιμοποιηθεί για την υποστήριξη του μηχανισμού ελέγχου πρόσβασης, θα πρέπει να υποστηρίζει τις δυο αυτές διαφορετικές λειτουργίες. Οπότε η καλύτερη λύση είναι να διαχωριστεί σε δυο επιμέρους πρωτόκολλα. Το πρώτο πρωτόκολλο θα είναι αυτό που θα χρησιμοποιηθεί για την κωδικοποίηση των πολιτικών ασφαλείας και των απαιτήσεων απορρήτου, καθιστώντας εφικτή την επεξεργασία, τον σχολιασμό και την οπτικοποίησή τους. Για την επίτευξη του σκοπού αυτού, μπορεί να χρησιμοποιηθεί το ανοιχτό πρότυπο «eXtensible Access Control Markup Language - XACML» [130]. Το δεύτερο πρωτόκολλο θα είναι αυτό που θα χρησιμοποιηθεί για τον έλεγχο της αποτελεσματικότητας της πολιτικής ασφαλείας και των απαιτήσεων απορρήτου. Αυτό θα καταγράφει όλες τις αποφάσεις σχετικά με την παροχή πρόσβασης στα δεδομένα. Η ανάλυση όλων αυτών θα αποκαλύψει χρήσιμες πληροφορίες, όπως για παράδειγμα τι δεδομένα προσπελαύνουν οι χρήστες, ή ποια δεδομένα προσπαθούν να προσπελάσουν, ή πως προσπαθούν παρακάμψουν τον μηχανισμό αυτό. Τα συμπεράσματα της ανάλυση αυτής θα καθορίσουν, αν χρήζει αναθεώρησης η εκάστοτε πολιτική ασφαλείας, ή αν θα πρέπει να επανακαθοριστούν οι απαιτήσεις απορρήτου των δεδομένων.

Συνοψίζοντας, διαπιστώνεται ότι οι παραδοσιακές λύσεις για τον έλεγχο της πρόσβασης στα δεδομένα στο περιβάλλον των «Big Data», όπως για παράδειγμα αυτές που βασίζονται σε ρόλους «Role-Based Access Control - RBAC», ή σε λίστες «List Based Access Control - LBAC», καθίστανται ανεπαρκείς, όσον αφορά την επαύξηση της δυνατότητας διαμοιρασμού των δεδομένων του συστήματος στις διάφορες οντότητές του με μεγαλύτερη ακρίβεια, χωρίς παράλληλα να παραβιάζεται το απόρρητό τους, καθώς δεν έχουν σχεδιαστεί για να παρέχουν μεγάλο βαθμό λεπτομέρειας. Επομένως απαιτείται μια νέα πρόταση, η οποία θα υποστηρίζει τον

λεπτομερή έλεγχο πρόσβασης στα δεδομένα με την αξιολόγηση, αφενός των απαιτήσεων απορρήτου των δεδομένων και αφετέρου των χαρακτηριστικών των οντοτήτων που αιτούνται την πρόσβαση σε αυτά. Μια τέτοια λύση είναι ο μηχανισμός «Attribute Based Access Control - ABAC» [131]. Ωστόσο, τα προβλήματα της υψηλής διαχειριστικής επιβάρυνσης, ως απόρροια του πλήθους των οντοτήτων, και της πιθανής απειλής για την εξαγωγή του προφίλ των χρηστών, μέσα από την χρήση των χαρακτηριστικών τους για την απόκτηση πρόσβασης στα δεδομένα, τον καθιστούν μη αποδοτικό και χρησιμοποιήσιμο ακόμα για το περιβάλλον των «Big Data» και συνεπώς περαιτέρω έρευνα απαιτείται για την βελτίωσή του.

3.4.5 ΕΝΗΜΕΡΩΣΗ «ΥΠΟΚΕΙΜΕΝΩΝ» ΚΑΙ ΔΙΑΦΑΝΕΙΑ

Η αρχή της «Διαφάνειας», που προβλέπεται στον «Κανονισμό (ΕΕ) 2016/679» [73] και στον «Νόμο 2472/1997» [78], αποσκοπεί στην προστασία των προσωπικών δεδομένων κατά την συλλογή και επεξεργασία τους. Μάλιστα, σύμφωνα με την αρχή αυτή, τα προσωπικά δεδομένα δεν θα πρέπει ποτέ να υποβάλλονται σε οποιαδήποτε μορφή επεξεργασίας, αν προηγουμένως δεν έχει ενημερωθεί το «Υποκείμενο» τους και δεν έχει δώσει την «Συγκατάθεσή» του. Αυτό ισχύει, τόσο κατά την αρχική συλλογή τους για την επίτευξη κάποιου προκαθορισμένου σκοπού, όσο και μεταγενέστερα για την επίτευξη σκοπού διαφορετικού του αρχικού, για τον οποίο αυτά συλλέχθηκαν. Επομένως, οι έννοιες της «Ενημέρωσης» και της «Συγκατάθεσης» των «Υποκειμένων» αποκτούν ιδιαίτερη βαρύτητα για να είναι νόμιμη η επεξεργασία των προσωπικών δεδομένων. Μάλιστα, βασική προϋπόθεση για την νομιμότητα της επεξεργασίας είναι κατ' αρχάς να έχει προηγηθεί η πλήρης ενημέρωση του «Υποκειμένου» των δεδομένων από τον «Υπεύθυνο Επεξεργασίας» και στην συνέχεια να δοθεί η ελεύθερη, ρητή και σαφής «Συγκατάθεσή» του για αυτή. Σκοπός των ανωτέρω ενεργειών είναι να έχουν ενημερωθεί πλήρως τα «Υποκείμενα» των δεδομένων για το πως θα τύχουν επεξεργασίας τα προσωπικά τους δεδομένα, ποιος θα είναι ο «Υπεύθυνος Επεξεργασίας» και για ποιο «Σκοπό» αυτά συλλέγονται, έτσι ώστε να προβούν στην συνέχεια σε συνειδητές επιλογές.

Ειδικότερα, οι έννοιες της «Ενημέρωσης» και της «Συγκατάθεσης» των «Υποκειμένων» των δεδομένων, για την προώθηση της «Διαφάνειας» στο περιβάλλον των «Big Data», αποκτούν ιδιαίτερη σημασία. Ως γνωστόν, εκ σχεδιασμού η διαδικασία της «Αναλυτικής των Μαζικών Δεδομένων» (Big Data

Analytics) προϋποθέτει την διασύνδεση διαφορετικών μεταξύ τους συνόλων δεδομένων, τα οποία συνήθως προέρχονται από διαφορετικές και ανεξάρτητες μεταξύ τους πηγές, προκειμένου αυτά να τύχουν επεξεργασίας και να εξαχθούν πολύτιμα συμπεράσματα. Άρα στην διαδικασία αυτή, τα διάφορα σύνολα δεδομένων, που χρησιμοποιούνται κατά την διεξαγωγή της, τυγχάνουν επεξεργασίας για την επίτευξη διαφορετικού σκοπού από αυτούς που επέβαλλαν την αρχική συλλογή τους από τις αντίστοιχες πηγές τους. Οπότε για να είναι νόμιμη η διαδικασία αυτή, θα πρέπει να έχουν προηγηθεί η «Ενημέρωση» και η «Συγκατάθεση» των «Υποκειμένων» των δεδομένων αυτών. Συνεπώς, η ύπαρξη αντίστοιχων μηχανισμών, τόσο για την «Ενημέρωση», όσο και για την «Συγκατάθεση» των «Υποκειμένων» στο περιβάλλον αυτό, κρίνεται πιο αναγκαία από ποτέ. Η χρήση των μηχανισμών αυτών, αφενός θα καταστήσει νόμιμη την διαδικασία της «Αναλυτικής» και αφετέρου θα συμβάλλει στην προώθηση της «Διαφάνειας» στα πολύπλοκα συστήματα των «Big Data» και ως εκ τούτου στην προστασία του απορρήτου των δεδομένων αυτών. Οπότε, μέσω των μηχανισμών αυτών επεκτείνεται η έννοια της «Διαφάνειας» πέρα από την αρχική συλλογή των δεδομένων, συμπεριλαμβάνοντας και την μετέπειτα επεξεργασία τους στο περιβάλλον των «Big Data». Επομένως, ο σχεδιασμός και η ανάπτυξη των κατάλληλων μηχανισμών «Ενημέρωσης» και «Συγκατάθεσης» στο περιβάλλον αυτό, για τους ανωτέρω λόγους, κρίνεται επιτακτικός. Στην παρούσα ενότητα εξετάζονται διεξοδικά οι μηχανισμοί «Ενημέρωσης», ενώ στην επόμενη ενότητα μελετώνται οι μηχανισμοί «Συγκατάθεσης».

Όσον αφορά την «Ενημέρωση» των «Υποκειμένων» των δεδομένων, θα πρέπει κατ' αρχάς να βελτιωθεί η ίδια η διαδικασία που ακολουθείται για αυτό τον σκοπό και εν συνεχεία να αναπτυχθούν μηχανισμοί που θα την υποστηρίζουν στο περιβάλλον των «Big Data». Η συνήθης διαδικασία είναι να ενημερώνονται τα «Υποκειμένων» των δεδομένων μέσα από δυσκολονόητα και εκτενή κείμενα. Η διαδικασία αυτή όμως αποτυγχάνει να εκπληρώσει τον σκοπό της, καθώς είναι εξαιρετικά χρονοβόρα και αναποτελεσματική. Για να βελτιωθεί λοιπόν η αποτελεσματικότητά της, στο [132] προτείνεται μια νέα προσέγγιση. Σύμφωνα με αυτή, θα πρέπει να παρέχονται όλες οι αναγκαίες πληροφορίες, στις οποίες συμπεριλαμβάνονται ο εκάστοτε σκοπός και οι λεπτομέρειες της επεξεργασίας, στα αντίστοιχα «Υποκείμενα» των δεδομένων σε γλώσσα κατανοητή, υπό μορφή λίστας, ειδοποιήσεων ή και εικόνων, κατά τα διάφορα στάδια της επεξεργασίας των δεδομένων τους και μάλιστα με αυξανόμενο

επίπεδο λεπτομέρειας για το καθένα από αυτά. Σκοπός είναι να καταστούν ευανάγνωστες και ευκολονόητες όλες οι πληροφορίες που σχετίζονται με την εκάστοτε επεξεργασία, προκειμένου τα «Υποκείμενα» να προβούν στην συνέχεια σε συνειδητή επιλογή για το αν επιθυμούν να τύχουν τα δεδομένα τους επεξεργασίας ή όχι.

Έτσι οι μηχανισμοί, που έχουν προταθεί προς την κατεύθυνση αυτή, χρησιμοποιούν είτε λίστες, είτε ειδοποιήσεις, είτε εικόνες και εικονίδια. Ένας τέτοιος μηχανισμός είναι το «Disconnect Privacy Icons» [193]. Σε αυτόν χρησιμοποιούνται συγκεκριμένα εικονίδια, προκειμένου να ενημερωθούν οι χρήστες για όλες τις αναγκαίες λεπτομέρειες, που καθορίζονται στην κείμενη νομοθεσία και που αφορούν την συλλογή και επεξεργασία των προσωπικών δεδομένων. Επίσης, στις πληροφορίες αυτές μπορεί να συμπεριλαμβάνονται και επιπλέον πληροφορίες, όπως ο χρόνος διατήρησής τους, η υποστήριξη της δυνατότητας να μην καταγράφεται η τοποθεσία καθώς και η ενημέρωση για το αν αυτή καταγράφεται, τα μέτρα ασφάλειας της δικτυακής αυτής υπηρεσίας κλπ. Σε αυτό τον μηχανισμό τα εικονίδια δημιουργούνται κατόπιν ανάλυσης της πολιτικής ασφαλείας της δικτυακής αυτής υπηρεσίας. Ένας άλλος παρεμφερής μηχανισμός είναι και ο «Mozilla Privacy Icons» [194]. Εντούτοις, το πρόβλημα που προκύπτει είναι ότι δεν υποχρεούνται όλοι οι «Υπεύθυνοι Επεξεργασίας» να χρησιμοποιήσουν στις δικτυακές υπηρεσίες τους αυτούς τους μηχανισμούς. Έτσι, σε περίπτωση που οι μηχανισμοί εμφανίσουν αρνητικά εικονίδια για κάποια δικτυακή υπηρεσία, γεγονός που υποδηλώνει επισφαλή πολιτική ασφαλείας ως προς την επεξεργασία των προσωπικών δεδομένων, τότε ο «Υπεύθυνος» αυτός ενδεχομένως να μην τον χρησιμοποιήσει. Η λύση σε αυτό το πρόβλημα είναι να ενσωματωθούν οι μηχανισμοί αυτοί στα διάφορα προγράμματα «Περιήγησης» (Browsers) και να χρησιμοποιούνται αυτόματα από αυτά. Μάλιστα, εκτός από την εμφάνιση των εικονιδίων κατόπιν ανάλυσης της πολιτικής ασφαλείας της δικτυακής υπηρεσίας, θα πρέπει να εμφανίζονται τα αρνητικά εικονίδια και στην περίπτωση που αυτή δεν παρέχει την δυνατότητα να αναλυθεί η πολιτική ασφαλείας της. Επιπλέον, για την υποστήριξη της λειτουργίας αυτής μπορεί να χρησιμοποιηθούν οι τεχνολογίες, είτε της «EXtensible Markup Language-XML» [195], είτε το «Platform for Privacy Preferences-P3P» [196], είτε το «P3P Preference Exchange Language (APPEL)» [197].

Συνοψίζοντας, διαπιστώνεται ότι με την τροποποίηση της διαδικασίας της ενημέρωσης, αυτή βελτιστοποιείται και γίνεται πιο αποτελεσματική, προάγοντας έτσι την «Διαφάνεια» στα διάφορα πληροφοριακά συστήματα. Ενώ αντίστοιχα με την χρήση των μηχανισμών για την ενημέρωση των «Υποκειμένων», αυξάνεται η ενημερότητα των χρηστών ως προς την επεξεργασία των προσωπικών τους δεδομένων. Ωστόσο, θα πρέπει να εξεταστούν περαιτέρω οι λύσεις αυτές από πλευράς αποτελεσματικότητας και δυνατότητας να χρησιμοποιηθούν στο πολύπλοκο περιβάλλον των «Big Data», ακριβώς λόγω των ιδιαίτερων χαρακτηριστικών του. Έτσι το πλήθος των «Οντοτήτων», οι διαφορετικοί και μεταξύ τους ανεξάρτητοι «Υπεύθυνοι Επεξεργασίας», οι διάφορες πολιτικές ασφαλείας, οι ποικίλες και διαδοχικές μορφές επεξεργασίας που λαμβάνουν χώρα στα συστήματα αυτά, το γεγονός ότι τα ίδια δεδομένα μπορεί να συμμετέχουν σε περισσότερες από μία τέτοιες επεξεργασίες και η ανάγκη διαχείρισης του πλήθους των ενημερώσεων, το οποίο μπορεί να προκαλεί επιβαρύνσεις και καθυστερήσεις, συνθέτουν τα ζητήματα που θα πρέπει να εξεταστούν ενδελεχώς.

3.4.6 ΜΗΧΑΝΙΣΜΟΙ ΣΥΝΑΙΝΕΣΗΣ - ΔΗΛΩΣΗΣ ΚΑΤΟΧΗΣ - ΕΛΕΓΧΟΥ

Εκτός από την «Ενημέρωση» των «Υποκειμένων» των δεδομένων, για να είναι νόμιμη η επεξεργασία των προσωπικών τους δεδομένων, απαιτείται και η «Συγκατάθεσή» τους. Παράλληλα, στην κείμενη νομοθεσία προβλέπονται και τα δικαιώματα των «Υποκειμένων» των δεδομένων, όπως για παράδειγμα το δικαίωμα «Πρόσβασης», «Αντίρρησης», Φορητότητας κλπ. Σκοπός όλων αυτών είναι να αποκτήσει το «Υποκείμενο» των δεδομένων τον πλήρη έλεγχο επί των προσωπικών του δεδομένων. Συνεπώς, η ανάπτυξη των αντίστοιχων μηχανισμών, τόσο για την «Συγκατάθεση», όσο και για τον πλήρη «Έλεγχο», κρίνεται επιτακτική. Μάλιστα στο περιβάλλον των «Big Data» αυτό είναι πιο αναγκαίο από ποτέ, καθώς οι ύπαρξή τους θα επαυξήσει την δυνατότητα το ίδιο το «Υποκείμενο» των δεδομένων να προστατεύσει την «Ιδιωτικότητά» του.

Έτσι, η συνεχής τροποποίηση του αρχικού σκοπού, για τον οποίο συλλέχθηκε αρχικά το σύνολο δεδομένων, και η διαρκής επαναχρησιμοποίηση, είτε των ήδη επεξεργασμένων συνόλων, είτε και των αρχικών, για την επίτευξη κάθε φορά διαφορετικού σκοπού στο περιβάλλον αυτό, καθιστά τους παραδοσιακούς μηχανισμούς συγκατάθεσης παρωχημένους και ανεπαρκείς. Επομένως θα πρέπει να

αναπτυχθούν χρήσιμοι και πρακτικοί μηχανισμοί «Συγκατάθεσης», οι οποίοι θα είναι αποτελεσματικοί στο να ικανοποιούν την νομική αυτή απαίτηση στο νέο αυτό περιβάλλον. Πράγματι, τέτοιοι μηχανισμοί έχουν αναπτυχθεί τα τελευταία χρόνια και ήδη χρησιμοποιούνται στα πληροφοριακά συστήματα, καθιστώντας την δυνατότητα της συγκατάθεσης εφικτή. Τέτοιο παράδειγμα είναι η περίπτωση για την συγκατάθεση στην πολιτική ασφαλείας της «Google» [198]. Επίσης παρόμοιοι μηχανισμοί, φιλικοί προς τον χρήστη, που στηρίζονται κυρίως στις τεχνολογικές λύσεις των «Banners» [199], έχουν προταθεί από τις «Αρχές Προστασίας των Δεδομένων» (Data Protection Authorities).

Ωστόσο, η υιοθέτηση και η χρήση των υπαρχόντων μηχανισμών «Συγκατάθεσης» στο περιβάλλον των «Big Data» απαιτεί περισσότερη έρευνα, καθώς εγείρονται ποικίλα ζητήματα. Το πρώτο αφορά την αυτοματοποίηση των μηχανισμών αυτών για την συλλογή, αλλά και την ανάκληση της συγκατάθεσης του «Υποκειμένου» των δεδομένων, καθώς στο περιβάλλον αυτό το πλήθος τους αναμένεται να είναι εξαιρετικά μεγάλο και να δημιουργεί σοβαρά προβλήματα. Ένα άλλο ζήτημα σχετίζεται με την περίπτωση εκείνη, στην οποία κάποιος «Υπεύθυνος Επεξεργασίας» διαμοιράζει τα προσωπικά δεδομένα σε κάποιο «Τρίτο Μέρος» για περαιτέρω επεξεργασία. Το ερώτημα που τίθεται είναι αν και πώς θα λαμβάνεται κάθε φορά η συγκατάθεση του «Υποκειμένου» των δεδομένων αυτών. Επίσης, εξέτασης χρήζει το πως θα λαμβάνεται η συγκατάθεση του «Υποκειμένου» στην περίπτωση που κάποιος αισθητήρας ή έξυπνη συσκευή συλλέγει προσωπικά του δεδομένα. Στην περίπτωση αυτή, ενδεχομένως να απαιτείται να καθοριστούν συγκεκριμένες ενέργειες ως πράξεις συγκατάθεσης εκ μέρους των «Υποκειμένων».

Μια άλλη εξαιρετικά ενδιαφέρουσα λύση για την υποστήριξη της «Συγκατάθεσης» στον ηλεκτρονικό κόσμο είναι αυτή των «Προτιμήσεων Απορρήτου» (Privacy Preferences). Σύμφωνα με αυτή, το «Υποκείμενο» των δεδομένων και ο «Υπεύθυνος Επεξεργασίας» δηλώνουν εξ' αρχής μέσω των απαιτήσεων απορρήτου και των πολιτικών ασφαλείας αντίστοιχα, πως επιθυμούν να τύχουν επεξεργασίας τα δεδομένα. Έτσι τα «Υποκείμενα» των δεδομένων επισυνάπτουν σε αυτά, με την χρήση του μηχανισμού «Sticky Policies» [133], ετικέτες με τις προτιμήσεις απορρήτου που επιθυμούν, δηλώνοντας άμεσα σε κάθε «Υπεύθυνο Επεξεργασίας», που θα θέλει να τα επεξεργαστεί, υπό ποιες προϋποθέσεις θα συναινέσουν αυτά να τύχουν επεξεργασίας. Στις ετικέτες αυτές μπορούν να δηλώσουν για παράδειγμα

τους αποδεκτούς σκοπούς επεξεργασίας, τους παραλήπτες στους οποίους αυτά επιτρέπονται να διαμοιραστούν, τον μέγιστο χρόνο τήρησης τους κλπ. Από την άλλη πλευρά, οι «Υπεύθυνοι Επεξεργασίας» δηλώνουν στις πολιτικές ασφαλείας τους τον σκοπό και τις συνθήκες διεξαγωγής της επεξεργασίας. Οπότε, προκειμένου να εξάγεται απόφαση κάθε φορά για το αν τα δεδομένα αυτά θα τύχουν επεξεργασίας ή όχι, απαιτείται η σύγκριση των απαιτήσεων απορρήτου του «Υποκειμένου» τους με τις πολιτικές ασφαλείας των εκάστοτε «Υπευθύνων Επεξεργασίας». Το αποτέλεσμα της σύγκρισης θα καθορίζει κάθε φορά αν το «Υποκείμενο» των δεδομένων συναινεί ή όχι να τύχουν επεξεργασίας τα δεδομένα του και μάλιστα τα επιμέρους στοιχεία πάνω στα οποία έγινε η σύγκριση θα καθορίζουν και της ακριβείς λεπτομέρειες για την διεξαγωγή της επεξεργασίας αυτής. Εκ πρώτης όψεως, διαφαίνεται η λύση αυτή να επιλύει σε μεγάλο βαθμό τα προβλήματα της αυτοματοποίησης και του διαμοιρασμού των δεδομένων σε «Τρίτα Μέρη» στο περιβάλλον των «Big Data» συγκριτικά με την προηγούμενη.

Μάλιστα, τεχνολογίες όπως η «EXtensible Markup Language-XML» [195], η «Platform for Privacy Preferences-P3P» [196] και η «P3P Preference Exchange Language (APPEL)» [197], μπορούν να υποστηρίξουν την λύση των «Προτιμήσεων Απορρήτου» (Privacy Preferences). Επιπρόσθετα, έχουν προταθεί τεχνικές για την βελτίωση και επέκταση της λύσης αυτής [134], [135], [136], οι οποίες υποστηρίζουν την διεξαγωγή αυτοματοποιημένης διαδικασίας διαπραγμάτευσης μεταξύ των «Υποκειμένων» των δεδομένων και των «Υπεύθυνων Επεξεργασίας». Στις τεχνικές αυτές τα «Υποκείμενα» των δεδομένων ορίζουν μια σειρά από προτιμήσεις απορρήτου, οι οποίες συνθέτουν τις παραμέτρους της διαπραγμάτευσης. Οπότε, κατά την διαπραγμάτευση λαμβάνονται υπόψη αυτές οι παράμετροι, με αποτέλεσμα να επιτευχθεί μια κοινά αποδεκτή λύση, η οποία θα ικανοποιεί τις απαιτήσεις και των δυο μερών και η οποία θα καθορίζει τις λεπτομέρειες επεξεργασίας των προσωπικών δεδομένων του «Υποκειμένου» στην συγκεκριμένη περίπτωση.

Ωστόσο, αν και οι ανωτέρω λύσεις αυξάνουν τον βαθμό ελέγχου του «Υποκειμένου» επί των δεδομένων του, το πρόβλημα που προκύπτει είναι πως αυτός θα διασφαλίζεται κάθε φορά. Οι λύσεις αυτές από μόνες τους δεν μπορούν να τον εγγυηθούν, καθώς θα πρέπει και τα υπόλοιπα μέρη, που συμμετέχουν στην διαδικασία της επεξεργασίας των δεδομένων, να είναι έμπιστα και να σέβονται τις απαιτήσεις του «Υποκειμένου» τους. Για την επίλυση του προβλήματος αυτού, μια

λύση, που βρίσκεται σε ερευνητικό στάδιο, είναι να προκύψει το κλειδί της κρυπτογράφησης των δεδομένων από τις αντίστοιχες απαιτήσεις απορρήτου τους. Έτσι, μόνο όσοι «Υπεύθυνοι Επεξεργασίας» τις ικανοποιούν θα μπορούν να τα αποκρυπτογραφήσουν και να τα χρησιμοποιήσουν. Χρήσιμες προς αυτή την κατεύθυνση είναι διάφορες τεχνικές της κρυπτογραφίας, όπως η «Υβριδική Κρυπτογράφηση» (Hybrid Cryptography), η «Κρυπτογράφηση με Βάση την Ταυτότητα μιας Οντότητας» (Identity Based Encryption) και η «Κρυπτογράφηση με Βάση τα Χαρακτηριστικά» (Attribute Based Encryption). Βέβαια σε όλες αυτές τις λύσεις θα πρέπει να υπάρχουν μία ή και περισσότερες «Έμπιστες Τρίτες Οντότητες» (Trusted Third Parties) για τον διαμοιρασμό και την διαχείριση των κλειδιών.

Εναλλακτικά, μπορεί να χρησιμοποιηθούν τεχνολογικές λύσεις οι οποίες θα προσδίδουν πλήρη έλεγχο του «Υποκειμένου» επί των προσωπικών του δεδομένων. Συγκεκριμένα αυτές οι λύσεις, είτε ονομάζονται «Personal Data Vaults», είτε «Personal Data Lockers», είτε «Personal Data Stores», επιτρέπουν στο εκάστοτε «Υποκείμενο» δεδομένων που τις χρησιμοποιεί να συλλέγει, να αποθηκεύει, να ενημερώνει, να διορθώνει, να αναλύει, ή/και να μοιράζει τα προσωπικά του δεδομένα κατά βούληση κάθε φορά [137]. Έτσι, οι λύσεις αυτές του προσδίδουν την δυνατότητα να ασκεί πλήρη έλεγχο, τόσο κατά την διάθεσή τους στους διάφορους «Υπεύθυνους Επεξεργασίας», όσο και μεταγενέστερα κατά την χρήση τους για τους σκοπούς της επεξεργασίας. Επίσης, την λειτουργία τους θα πρέπει να υποστηρίξουν, αφενός ένα κατάλληλο πρωτόκολλο για την ανταλλαγή των δεδομένων (Semantic Data Interchange Protocol) [202] και αφετέρου τα αντίστοιχα έμπιστα πλαίσια (Trust Frameworks) [138], τα οποία θα επιτρέπουν την ενσωμάτωση των προτιμήσεων απορρήτου του «Υποκειμένου» των δεδομένων, έτσι ώστε αυτά να ικανοποιούνται από τον εκάστοτε «Υπεύθυνο Επεξεργασίας».

Προς αυτή την κατεύθυνση υπάρχουν σε ερευνητικό στάδιο λύσεις όπως το «Personal Data Vault» [139], το «Databox» [140], το «Virtual Individual Server» [141], το «OpenPDS» [142] και το «Enigma» [143]. Αντίστοιχες εμπορικές λύσεις είναι το «Higgins» [200], το «Mydex» [201], το «CyberAll» [144], το «Persona» [145] και το «Confab» [146]. Όμως όλες αυτές οι λύσεις διαφοροποιούνται μεταξύ τους ως προς το είδος των δεδομένων που διαχειρίζονται, τον τρόπο που έχουν υλοποιηθεί, το ποσοστό ελέγχου που δίνουν στο «Υποκείμενο» των δεδομένων και την συσκευή στην οποία εκτελούνται. Οπότε προκύπτουν τα εξής ερωτήματα, αφενός ποιες από

αυτές θα είναι αποτελεσματικές σε σχέση με τα ιδιαίτερα χαρακτηριστικά των συστημάτων «Big Data», ήτοι την ποικιλομορφία των δεδομένων, το πλήθος των οντοτήτων σε αυτό και την πολυπλοκότητά του, και αφετέρου πως θα πρέπει αυτές να χρησιμοποιηθούν, προκειμένου το εκάστοτε «Υποκείμενο» των δεδομένων να έχει εγγυημένα τον πλήρη έλεγχο επί των προσωπικών του δεδομένων. Επιπλέον, ένα άλλο ζήτημα είναι αυτό της διαχείρισης των πολυάριθμων αιτημάτων πρόσβασης στα δεδομένα από τα «Υποκείμενά» τους. Λόγω των μηχανισμών αυτών, αντί οι «Υπεύθυνοι Επεξεργασίας» να βασίζονται μόνο στα υπάρχοντα σύνολα δεδομένων που έχουν, μπορούν κάθε φορά να αιτούνται πρόσβαση στα δεδομένα αυτά άμεσα από τα ίδια τα «Υποκείμενά» τους. Επίσης, η κεντρική αποθήκευση και διαχείριση των δεδομένων αυτών από τα ίδια τα «Υποκείμενα» τους, ενέχει κινδύνους για την ασφάλειά τους (Single Point of Failure).

Συνεπώς, οι ανωτέρω μηχανισμοί προσδίδουν θετική προοπτική στην διασφάλιση της δυνατότητας του «Υποκειμένου» να ασκεί πλήρη έλεγχο στα δεδομένα του και ταυτόχρονα να επιλέγει τις απαιτήσεις απορρήτου τους στο περιβάλλον των «Big Data». Εντούτοις, αν και υπάρχουν κάποιες λύσεις προς την κατεύθυνση αυτή, οι οποίες έχουν υλοποιηθεί και χρησιμοποιούνται σε περιορισμένο όμως βαθμό, το πεδίο αυτό βρίσκεται ακόμα σε πολύ πρώιμο στάδιο. Ως εκ τούτου, προκειμένου να γίνουν πραγματικότητα αυτές οι δυνατότητες, απαιτείται περισσότερη καινοτομία και δημιουργικότητα στον τομέα αυτό.

3.5 ΣΥΝΟΨΗ ΚΕΦΑΛΑΙΟΥ

Ήδη τα «Big Data» χρησιμοποιούνται ευρέως σε ποικίλους τομείς της ζωής του σύγχρονου ανθρώπου, από το εμπόριο και τις «online» συναλλαγές, μέχρι την επιστήμη, την υγεία, την έρευνα και την ασφάλεια. Ωστόσο, η εκτεταμένη συλλογή δεδομένων από μια πληθώρα πηγών, στα οποία συνήθως περιλαμβάνονται και προσωπικά δεδομένα, έχει προκαλέσει σοβαρές ανησυχίες για την προστασία της «Ιδιωτικότητας». Μάλιστα το γεγονός, ότι αυτά τυγχάνουν επεξεργασίας από τις διάφορες τεχνικές των «Big Data Analytics» για να ληφθούν τεκμηριωμένες αποφάσεις επί διαφόρων ζητημάτων, αναδεικνύει το ζήτημα της προστασίας της «Ιδιωτικότητας» σε μείζον θέμα. Στο παρόν κεφάλαιο λοιπόν εξετάστηκαν λεπτομερώς τα θέματα, που σχετίζονται με την προστασία της στο ιδιαίτερο περιβάλλον των «Big Data». Αρχικά, παρατέθηκαν οι ορισμοί της «Ιδιωτικότητας» και

της «Ιδιωτικότητας εκ Σχεδιασμού» και παρουσιάστηκε συνοπτικά το ρυθμιστικό και κανονιστικό πλαίσιο που διέπει την προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων στην Ευρωπαϊκή Ένωση και στην Ελλάδα. Στην συνέχεια μελετήθηκαν επισταμένα τα επιμέρους ζητήματα, που σχετίζονται με την προστασία της «Ιδιωτικότητας» στο περιβάλλον αυτό. Συγκεκριμένα, τα ζητήματα που εξετάστηκαν είναι η χρήση της μεθόδου για την «Αξιολόγηση των Επιπτώσεων επί της Ιδιωτικότητας» κατά τον σχεδιασμό των συστημάτων αυτών, η εφαρμογή των διαφόρων τεχνικών «Ανωνυμοποίησης» στα ποικίλα σύνολα δεδομένων που αυτό διαχειρίζεται, η χρήση της «Κρυπτογραφίας», η υλοποίηση μηχανισμών «Ελέγχου Πρόσβασης», η υιοθέτηση των μηχανισμών «Ενημέρωσης και Διαφάνειας», καθώς και η υλοποίηση μηχανισμών «Συναίνεσης», «Δήλωσης Κατοχής» και «Ελέγχου». Από την έρευνα που διεξήχθη, διαπιστώθηκε ότι στο κάθε ένα από αυτά τα ζητήματα προκύπτουν ορισμένα επιπλέον θέματα, τα οποία και παρατέθηκαν. Μάλιστα, για το κάθε θέμα παρουσιάστηκαν οι διάφορες λύσεις, που έχουν προταθεί από την επιστημονική κοινότητα για την αντιμετώπισή τους. Εντούτοις, ορισμένες από τις προτεινόμενες λύσεις δημιουργούν πρόσθετα ζητήματα που σχετίζονται, τόσο με την αποτελεσματικότητά τους, όσο και με την επίδραση που έχουν στην απόδοση του συστήματος, με συνέπεια να απαιτείται να τροποποιηθούν και να προσαρμοστούν στα νέα δεδομένα. Ενώ κάποιες άλλες, είτε βρίσκονται ακόμα σε ερευνητικό στάδιο, είτε αποτελούν μόνο προτάσεις. Επιπρόσθετα, υπάρχουν ορισμένα θέματα που παραμένουν ανοιχτά και άλυτα. Συνεπώς, όλα αυτά τα θέματα θα πρέπει να διερευνηθούν σε βάθος, έτσι ώστε να επιλυθούν και να καταστούν βιώσιμα, συμβάλλοντας έτσι στην αποτελεσματική προστασία της «Ιδιωτικότητας» στο περιβάλλον των «Big Data».

4. ΚΕΦΑΛΑΙΟ 4ο: ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ ΓΙΑ ΤΗΝ ΠΡΟΣΤΑΣΙΑ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΩΝ ΜΑΖΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ

4.1 ΠΡΟΛΟΓΟΣ – ΠΕΡΙΓΡΑΦΗ ΜΕΛΕΤΗΣ ΠΕΡΙΠΤΩΣΗΣ

Είναι γεγονός ότι τα τελευταία χρόνια η μηχανοργάνωση των υπουργείων και των διαφόρων φορέων και οργανισμών της Ελληνικής Δημόσιας Διοίκησης συντελείται με γοργούς ρυθμούς. Επιπρόσθετα, στα πλαίσια της αναβάθμισης των υπηρεσιών που παρέχονται στον Έλληνα πολίτη, έχει προωθηθεί σε μεγάλο βαθμό η διασύνδεση όλων αυτών των διαφορετικών πληροφοριακών συστημάτων μεταξύ τους, ενίοτε με την συγκέντρωση κάποιων από τα σύνολα δεδομένων τους σε μία ενιαία αποθήκη δεδομένων και άλλοτε με την διάθεσή τους στο κοινό, αλλά και μεταξύ των διαφόρων φορέων του Δημοσίου Τομέα, μέσω διαδικτυακών υπηρεσιών. Τα οφέλη αυτής της διασύνδεσης είναι πολλαπλά και ενδεικτικά κάποια από αυτά είναι η ταχεία εξυπηρέτηση των πολιτών στα «Κέντρα Εξυπηρέτησης Πολιτών-ΚΕΠ», η καταπολέμηση της γραφειοκρατίας, η παροχή «Διαφάνειας» κατά την άσκηση της δημόσιας διοίκησης σε εφαρμογή της «Ανοιχτής Διακυβέρνησης» (Open Government) και η επιτάχυνση του έργου των διαφόρων φορέων και υπουργείων, μέσα από την αποδοτικότερη ανταλλαγή και επεξεργασία των διασυνδεδεμένων αυτών συνόλων δεδομένων. Συνεπώς, η διασύνδεση όλων των ανεξάρτητων πληροφοριακών συστημάτων της δημόσιας διοίκησης μεταξύ τους, για την περαιτέρω επεξεργασία των διασυνδεδεμένων αυτών συνόλων δεδομένων από τα διάφορα όργανά της, στο οποίο βρίσκεται αποθηκευμένο πλήθος ποικιλόμορφων δεδομένων και αρχείων, την καθιστά περιβάλλον «Big Data». Μάλιστα, είναι σημαντικό να τονιστεί ότι αυτό το περιβάλλον δεν είναι απομονωμένο, τουναντίον είναι επεκτάσιμο, καθώς σε αυτό μπορούν να προστεθούν και άλλες μη κυβερνητικές πηγές δεδομένων, όπως για παράδειγμα τα ευρετήρια τηλεφωνικού καταλόγου κλπ.

Ωστόσο, αν και η διασύνδεση αυτή συμβάλλει σε μεγάλο βαθμό στην καλύτερη άσκηση της δημόσιας διοίκησης, εντούτοις η εφαρμογή της «Ανοιχτής Διακυβέρνησης» (Open Government) στο περιβάλλον αυτό ενέχει κινδύνους για την προστασία της «Ιδιωτικότητας» των Ελλήνων πολιτών. Σκοπός λοιπόν του παρόντος κεφαλαίου είναι να εξεταστεί κατά πόσο είναι εφικτό, στα πλαίσια της «Διαφάνειας»,

να διαρρεύσουν τα προσωπικά δεδομένα τους μέσα από την συγκέντρωση και επεξεργασία των διαφόρων πληροφοριών, που αποθηκεύονται ή/και παρέχονται, τόσο από τα διάφορα πληροφοριακά συστήματα της Δημόσιας Διοίκησης, όσο και από άλλες ευρέως διαθέσιμες μη κυβερνητικές πηγές δεδομένων. Μάλιστα, αν και το ισχύον νομοθετικό πλαίσιο καθορίζει σαφώς τα ζητήματα που σχετίζονται με την προστασία των προσωπικών δεδομένων των Ελλήνων πολιτών, πολλές φορές η αμέλεια στην αποτελεσματική εφαρμογή των κανόνων, ή η περιστασιακή αδιαφορία για αυτούς, ή/και η ελλιπής ενημέρωση, τόσο των ίδιων των υπαλλήλων της δημόσιας διοίκησης, όσο και των ίδιων των πολιτών, μπορεί να θέσουν σε κίνδυνο το απόρρητο των δεδομένων αυτών.

Συγκεκριμένα, στο παρόν κεφάλαιο διερευνάται κατά πόσο είναι εφικτό να προκληθεί αποκάλυψη «Προσωπικών Δεδομένων», «Μοναδικών Αριθμών Ταυτοποίησης» (Unique Identification Numbers-UIN) και «Αναγνωριστικών Στοιχείων Ταυτότητας» (Personally Identifiable Information-PII) σε μη εξουσιοδοτημένους πολίτες μέσα από τον συνδυασμό των δεδομένων, που παρέχονται σε αυτούς, τόσο από μια σειρά κυβερνητικών πηγών δεδομένων στα πλαίσια της «Ανοιχτής Διακυβέρνησης», όσο και από άλλες ευρέως διαθέσιμες μη κυβερνητικές πηγές. Στις πηγές του Ελληνικού κράτους, οι οποίες προσφέρονται ως διαδικτυακές υπηρεσίες προς τους Έλληνες πολίτες, συμπεριλαμβάνονται το «Υπουργείο Οικονομικών» (ΑΦΜ), το «Υπουργείο Προστασίας του Πολίτη» (ΑΔΤ), το «Υπουργείο Εργασίας και Κοινωνικής Ασφάλισης» (ΑΜΚΑ), το «Υπουργείο Εσωτερικών» (Εκλογές, Δημοτολόγιο) και το «Υπουργείο Διοικητικής Ανασυγκρότησης» (Διαύγεια). Ενώ αντίστοιχα η ευρέως διαθέσιμη μη κυβερνητική πηγή, που χρησιμοποιείται για τις ανάγκες τις παρούσας εργασίας, είναι η «Υπηρεσία Τηλεφωνικού Καταλόγου του ΟΤΕ» (Τηλέφωνο). Στη συνέχεια του κεφαλαίου, αφού περιγραφούν οι ανωτέρω πηγές δεδομένων, εξετάζονται, αφενός αν είναι δυνατό να προκύψει διαρροή προσωπικών δεδομένων από τις ίδιες τις πηγές και αφετέρου αν μπορεί ο συνδυασμός των πληροφοριών, που περιέχονται σε αυτές τις διαφορετικές πηγές, να οδηγήσει σε επιπλέον αποκάλυψη προσωπικών δεδομένων. Τέλος, παρουσιάζονται τα αποτελέσματα και οι διαπιστώσεις της έρευνας αυτής, καθώς επίσης προτείνονται και κάποιες λύσεις για την αντιμετώπισή τους.

4.2 ΠΗΓΕΣ ΔΕΔΟΜΕΝΩΝ ΜΕΛΕΤΗΣ ΠΕΡΙΠΤΩΣΗΣ

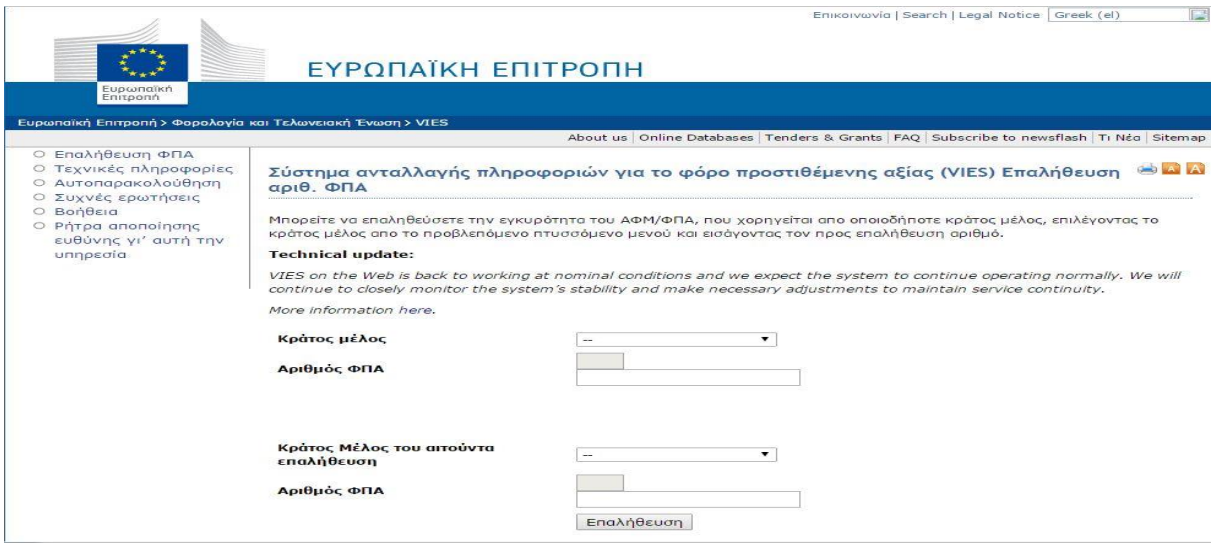
Στην ενότητα αυτή, παρουσιάζονται συνοπτικά οι πηγές δεδομένων που χρησιμοποιήθηκαν για τις ανάγκες της παρούσας έρευνας. Επιπλέον περιγράφεται ο σκοπός για τον οποίο δημιουργήθηκαν, καθώς και ο τύπος των δεδομένων που περιέχουν. Ωστόσο, πριν την περιγραφή τους, αρχικά παρατίθενται οι έννοιες των «Προσωπικών Δεδομένων», των «Ευαίσθητων Προσωπικών Δεδομένων», του «Μοναδικού Αριθμού Ταυτοποίησης» (Unique Identification Numbers-UIN/UID) και του «Αναγνωριστικού Στοιχείου Ταυτότητας» (Personally Identifiable Information-PII). Στην συνέχεια περιγράφονται με την σειρά, πρώτα οι κυβερνητικές πηγές δεδομένων και έπειτα οι μη κυβερνητική.

Έτσι κατ' αρχάς, σύμφωνα με το Άρθρο 2α του «Νόμου 2472/1997» [78], ως «Δεδομένα Προσωπικού Χαρακτήρα» ορίζονται κάθε πληροφορία που αναφέρεται στο «Υποκείμενο» των δεδομένων. Μάλιστα, στο Άρθρο 2γ καθορίζεται η αμφίδρομη ένα προς ένα (1-1) αντιστοίχιση του «Υποκειμένου» με τα δεδομένα του, η οποία μπορεί να γίνει, είτε αν η ταυτότητά του είναι γνωστή, είτε αν αυτή μπορεί να εξακριβωθεί άμεσα ή έμμεσα βάσει του αριθμού ταυτότητας ή και περισσότερων συγκεκριμένων στοιχείων, όπως ο «ΑΜΚΑ» και το «ΑΦΜ», ή από στοιχεία που χαρακτηρίζουν την υπόστασή του από άποψη φυσική, βιολογική, ψυχική, οικονομική, πολιτιστική, πολιτική ή κοινωνική. Περαιτέρω, στο Άρθρο 2β προσδιορίζεται ακριβώς η ειδική κατηγορία των «Ευαίσθητων Δεδομένων». Αυτά είναι τα δεδομένα που αφορούν την φυλετική ή εθνική προέλευση, τα πολιτικά φρονήματα, τις θρησκευτικές ή φιλοσοφικές πεποιθήσεις, την συμμετοχή σε συνδικαλιστική οργάνωση, την υγεία, την κοινωνική πρόνοια, την ερωτική ζωή, τα σχετικά με ποινικές διώξεις ή καταδίκες, καθώς και την συμμετοχή με συναφείς με τα ανωτέρω ενώσεις προσώπων. Τέλος, ο «Μοναδικός Αριθμός Ταυτοποίησης» (Unique Identification Number–UIN/UID) [218] και το «Αναγνωριστικό Στοιχείο Ταυτότητας» (Personally Identifiable Information-PII) [217] είναι ο αριθμός ή η πληροφορία αντίστοιχα, που συνδέονται αμφιμονοσήμαντα με κάποια οντότητα και ως εκ τούτου μπορούν κάλλιστα να χρησιμοποιηθούν για την άμεση αποκάλυψη της ταυτότητάς της. Σε αυτά συμπεριλαμβάνονται ο «Αριθμός Φορολογικού Μητρώου», ο «Αριθμός Μητρώου Κοινωνικής Ασφάλισης», ο «Αριθμός Δελτίου Ταυτότητας», ο «Ειδικός Εκλογικός Αριθμός», το «e-mail», το δακτυλικό αποτύπωμα, ο «Αριθμός Τηλεφώνου» κλπ.

Αναλυτικότερα, όσον αφορά τις κυβερνητικές πηγές δεδομένων της παρούσας έρευνας, αυτές είναι οι παρακάτω [147]:

❖ **Υπουργείο Οικονομικών (Αριθμός Φορολογικού Μητρώου-ΑΦΜ)**

Ο «Αριθμός Φορολογικού Μητρώου» (ΑΦΜ) είναι ένας μοναδικός αριθμός, ο οποίος εκδίδεται από το Υπουργείο Οικονομικών για κάθε πολίτη ή νομικό πρόσωπο και σχετίζεται άμεσα με την οποιαδήποτε οικονομική του δραστηριότητα στο κράτος. Μάλιστα, το γεγονός, ότι αυτός συνδέεται με κάποια φυσική ή νομική οντότητα μονοσήμαντα, τον καθιστά βασική προϋπόθεση για την διεκπεραίωση της οποιαδήποτε συναλλαγής του με τους διάφορους οργανισμούς και επιχειρήσεις, όπως τις τράπεζες και άλλους ασφαλιστικούς φορείς, καθώς χρησιμοποιείται ως επιπλέον μέθοδος ταυτοποίησής τους πέρα από τον «Αριθμό Δελτίου Ταυτότητας». Έτσι, προκειμένου να ελέγχεται κάθε φορά η εγκυρότητα του «ΑΦΜ», ο αλγόριθμος παραγωγής του έχει κοινοποιηθεί στους διάφορους κρατικούς φορείς για να τον χρησιμοποιήσουν για το σκοπό αυτό. Ο «ΑΦΜ» είναι ένας μη σειριακός αριθμός 9 ψηφίων, ο οποίος παράγεται από ένα αλγόριθμο. Ο «ΑΦΜ» χωρίζεται σε δυο κύριες κατηγορίες. Στην πρώτη κατηγορία ανήκουν τα φυσικά πρόσωπα, όπως οι μισθωτοί και οι συνταξιούχοι, ενώ στην δεύτερη κατηγορία συμπεριλαμβάνονται οι ελεύθεροι επαγγελματίες, οι επιχειρήσεις, τα διάφορα νομικά πρόσωπα, οι οργανισμοί και τα ιδρύματα. Μάλιστα, τόσο η «Γενική Γραμματεία Πληροφοριακών Συστημάτων-(ΓΓΠΣ)» [203], η οποία είναι υπεύθυνη για την μηχανοργάνωση του κρατικού φορέα, όσο και η Ευρωπαϊκή Ένωση, έχουν εισάγει μια νέα διαδικτυακή υπηρεσία [206], η οποία δίνει την δυνατότητα να εξετάζεται κάθε φορά η εγκυρότητα των ΑΦΜ της δεύτερης κατηγορίας, ενώ αυτό δεν υποστηρίζεται αντίστοιχα για την πρώτη. Ειδικότερα, η «ΓΓΠΣ» έχει συνδέσμους, τόσο στην καρτέλα που αφορά τους πολίτες [204], όσο και στην καρτέλα που αφορά τις επιχειρήσεις [205], οι οποίοι παραπέμπουν για την εκτέλεση της υπηρεσίας αυτής στην αντίστοιχη διαδικτυακή υπηρεσία της Ευρωπαϊκής Ένωσης [206] (Εικόνα 12). Επίσης, ο έλεγχος της εγκυρότητας των «ΑΦΜ» της δεύτερης κατηγορίας επιστρέφει τα εξής στοιχεία, το ονοματεπώνυμο, την επωνυμία και την διεύθυνση, που αντιστοιχούν στο συγκεκριμένο «ΑΦΜ».



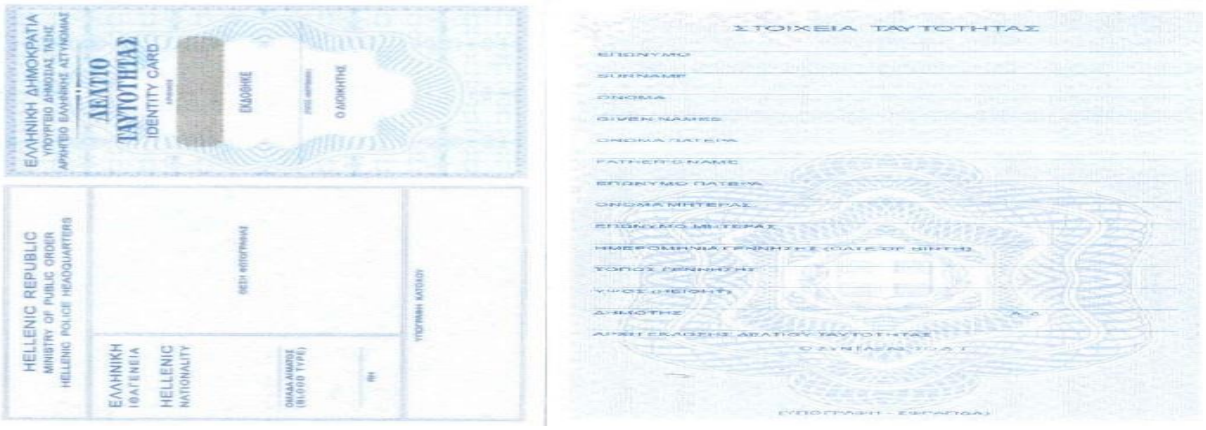
ΕΙΚΟΝΑ 12: ΕΥΡΩΠΑΙΚΗ ΕΠΙΤΡΟΠΗ – ΔΙΑΔΙΚΤΥΑΚΗ ΥΠΗΡΕΣΙΑ ΕΛΕΓΧΟΥ ΑΦΜ

Επιπρόσθετα θα πρέπει να σημειωθεί ότι, στα πλαίσια της προώθησης της ηλεκτρονικής διακυβέρνησης και της εξασφάλισης της διαλειτουργικότητας των διαφόρων πληροφοριακών συστημάτων του Ελληνικού κράτους μεταξύ τους, παρέχεται η δυνατότητα ελέγχου της εγκυρότητας του «ΑΦΜ» και του «Αριθμού Δελτίου Ταυτότητας-ΑΔΤ», τόσο των «Φυσικών» προσώπων, όσο και των «Μη Φυσικών» προσώπων, αποκλειστικά για τους «Φορείς του Δημοσίου». Έτσι, έχει αναπτυχθεί μια διαδικτυακή υπηρεσία για τον σκοπό αυτό, η οποία όμως είναι διαθέσιμη μόνο κατόπιν αιτήσεως του αρμόδιου φορέα από την «ΓΓΠΣ». Επίσης, η υπηρεσία αυτή, αν εγκριθεί, θα επιστρέψει τα βασικά στοιχεία του φορολογούμενου, τα στοιχεία επικοινωνίας του και κάποια πρόσθετα στοιχεία αν πρόκειται για «Μη Φυσικό» πρόσωπο [207].

❖ Υπουργείο Προστασίας του Πολίτη (Αριθμός Δελτίου Ταυτότητας)

Η ταυτότητα είναι επίσημο έγγραφο του Ελληνικού κράτους και εκδίδεται από την Ελληνική Αστυνομία, η οποία υπάγεται στο Υπουργείο Προστασίας του Πολίτη [208]. Στην πίσω όψη (Εικόνα 13) αναγράφονται το ονοματεπώνυμο του κατόχου της, τα ονοματεπώνυμα του πατέρα και της μητέρας του, η ημερομηνία και ο τόπος γέννησής του, το ύψος του, καθώς και ο Δήμος στον οποίο είναι εγγεγραμμένος με τον αριθμό δημοτολογίου του. Στην μπροστά όψη της (Εικόνα 13) περιλαμβάνονται η φωτογραφία του κατόχου, η υπογραφή του κατόχου, η ομάδα αίματος και ο «Αριθμός Δελτίου Ταυτότητας». Μάλιστα, πριν το 2005, στις ταυτότητες αναγράφονταν και επιπλέον στοιχεία, ήτοι το επάγγελμα, το θρήσκευμα, η διεύθυνση κατοικίας, το ονοματεπώνυμο συζύγου και το χρώμα των ματιών και των μαλλιών, ενώ στην

μπροστά όψη υπήρχε και το δακτυλικό αποτύπωμα του δείκτη του δεξιού χεριού. Ωστόσο, αυτά αφαιρέθηκαν για λόγους προστασίας των προσωπικών δεδομένων, ή γιατί δεν ήταν απαραίτητα για την ταυτοποίηση του προσώπου. Η ταυτότητα χρησιμοποιείται κυρίως για την ταυτοποίηση των πολιτών στις διάφορες δραστηριότητές τους. Για αυτό τον σκοπό, σε αυτή αντιστοιχεί ένας μοναδικός αριθμός, ο «Αριθμός Δελτίου Ταυτότητας». Ο αριθμός αυτός, κατά βάση, αποτελείται από δυο κεφαλαία ελληνικά γράμματα και στη συνέχεια από 6 αριθμούς. Ωστόσο, σε αντίθεση με το «ΑΜΚΑ» και το «ΑΦΜ», ο αριθμός αυτός αφορά αποκλειστικά την ταυτότητα και όχι το πρόσωπο αυτό καθ' αυτό. Έτσι, για παράδειγμα σε περίπτωση απώλειας ή κλοπής της, η επανέκδοση της ταυτότητας συνεπάγεται και την αλλαγή του μοναδικού αριθμού αυτού και ταυτόχρονα την ακύρωση του προηγούμενου. Αυτό σημαίνει ότι δεν υπάρχει μονοσήμαντη αντιστοίχιση του εκάστοτε πολίτη με τον «Αριθμό Δελτίου Ταυτότητας», οπότε είναι δυνατό σε ένα πολίτη να αντιστοιχούν περισσότεροι από έναν τέτοιοι αριθμοί, από τους οποίους όμως σε ισχύ είναι μόνο ο τελευταίος εκδοθείς κάθε φορά. Για αυτό τον λόγο, σε πολλές περιπτώσεις, τόσο στον δημόσιο, όσο και στον ιδιωτικό τομέα, ζητείται και καταχωρείται παράλληλα με τον «Αριθμό Δελτίου Ταυτότητας» και ο «ΑΦΜ» ή/και ο «ΑΜΚΑ». Αυτό σημαίνει ότι είναι δυνατό να βρεθεί στα διάφορα σύνολα δεδομένων, που τηρούνται στις διάφορες δημόσιες υπηρεσίες, ο συνδυασμός του εκάστοτε ισχύοντα «Αριθμού Δελτίου Ταυτότητας» με τα αντίστοιχα «ΑΦΜ» ή/και «ΑΜΚΑ».



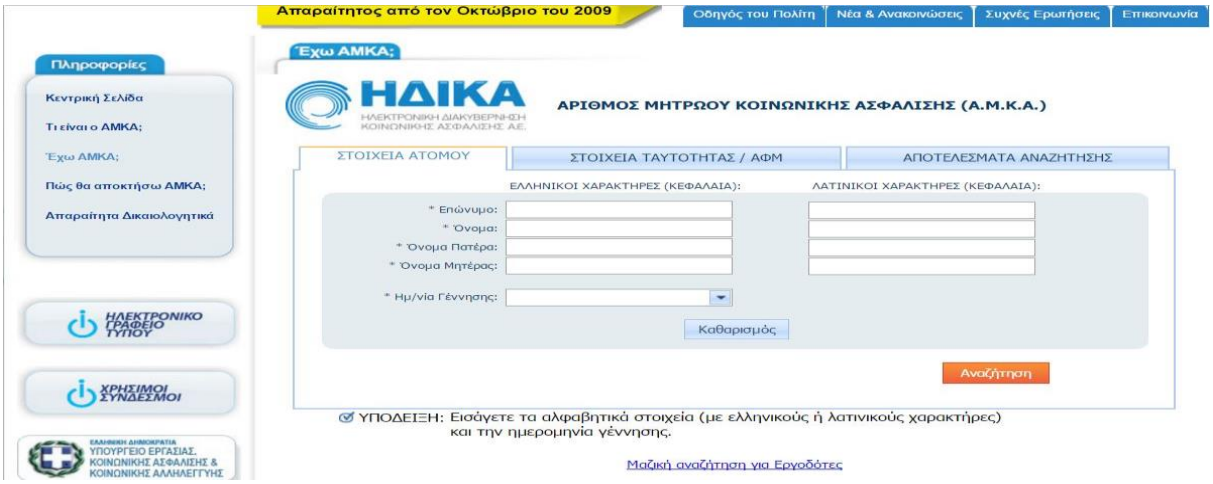
ΕΙΚΟΝΑ 13: ΕΜΠΡΟΣΘΙΑ ΚΑΙ ΟΠΙΣΘΙΑ ΟΨΗ ΔΕΛΤΙΟΥ ΤΑΥΤΟΤΗΤΑΣ

❖ **Υπουργείο Εργασίας και Κοινωνικής Ασφάλισης (Αριθμός Μητρώου Κοινωνικής Ασφάλισης- ΑΜΚΑ)**

Ο «Αριθμός Μητρώου Κοινωνικής Ασφάλισης» (ΑΜΚΑ) είναι ένας μοναδικός αριθμός που αντιστοιχεί στον κάθε Έλληνα πολίτη και ο οποίος εκδίδεται από το

Υπουργείο Εργασίας και Κοινωνικής Ασφάλισης. Αυτός χρησιμοποιείται ευρέως από τους ασφαλιστικούς φορείς, τις διάφορες υγειονομικές υπηρεσίες και στον εργασιακό τομέα. Μάλιστα είναι υποχρεωτικό ο κάθε Έλληνας πολίτης να αποκτήσει έναν τέτοιο αριθμό. Ο «ΑΜΚΑ» αποτελείται από 11 αριθμούς και έχει την μορφή «ΗΗΜΜΕΕΧΧΧΥΖ». Οι πρώτοι 6 αριθμοί αντιστοιχούν στην κωδικοποίηση της ημερομηνίας γέννησης του ατόμου (ΗΗ/ΜΜ/ΕΕ). Ενώ οι επόμενοι 4 αριθμοί (ΧΧΧΥ) ακολουθούν μια σειριακά αυξανόμενη ακολουθία, στην οποία τα 3 πρώτα ψηφία (ΧΧΧ) προκύπτουν από τον αύξοντα αριθμό των γεννήσεων στο συγκεκριμένο έτος και το 4^{ο} ψηφίο (Υ) αντιστοιχεί στο φύλλο του ατόμου. Το (Υ) είναι περιττός αριθμός για τους άνδρες και άρτιος για τις γυναίκες. Το τελευταίο ψηφίο (Ζ) αποτελεί χαρακτήρα ελέγχου για το ΑΜΚΑ [148].

Επίσης, δίδεται η δυνατότητα στους Έλληνες πολίτες να βρουν τον «ΑΜΚΑ» τους στην διαδικτυακή υπηρεσία [209] (Εικόνα 14). Σε αυτή, αρχικά παρέχουν το ονοματεπώνυμό τους, το πατρώνυμο, το μητρώνυμο και την πλήρη ημερομηνία γέννησής τους. Αν κάποιο πεδίο από αυτά δεν συμπληρωθεί σωστά ή καθόλου, τότε αυτή βγάζει μήνυμα λάθους ότι τα στοιχεία είναι ελλιπή ή λανθασμένα. Ενώ αν αυτά δεν αντιστοιχούν σε κάποιο πολίτη, τότε επιστρέφεται το αντίστοιχο μήνυμα λάθους, δηλαδή ότι δεν βρέθηκε άτομο καταχωρημένο στο εθνικό μητρώο «ΑΜΚΑ». Αν αυτά είναι σωστά, τότε στο επόμενο βήμα η υπηρεσία αυτή ζητάει να εισαχθούν ο «Αριθμός Δελτίου Ταυτότητας» και το «ΑΦΜ», προκειμένου να ολοκληρωθεί. Τέλος, αν όλα τα στοιχεία είναι σωστά, τότε στο τρίτο βήμα επιστρέφεται ο «ΑΜΚΑ» του πολίτη.



ΕΙΚΟΝΑ 14: ΗΔΙΚΑ – ΔΙΑΔΙΚΤΥΑΚΗ ΥΠΗΡΕΣΙΑ ΕΥΡΕΣΗΣ ΑΜΚΑ

❖ Υπουργείο Εσωτερικών (Ειδικός Εκλογικός Αριθμός-ΕΕΑ)

Το Υπουργείο Εσωτερικών έχει αναπτύξει μια διαδικτυακή υπηρεσία [210] (Εικόνα 15), η οποία βοηθά τους Έλληνες ψηφοφόρους να βρουν το εκλογικό κέντρο, στο οποίο πρέπει να ασκήσουν το εκλογικό τους δικαίωμα την ημέρα των εκλογών. Μάλιστα, η υπηρεσία αυτή είναι συνεχώς διαθέσιμη στους πολίτες. Επιπλέον, σε κάθε Έλληνα πολίτη αντιστοιχεί ένας μοναδικός «Ειδικός Εκλογικός Αριθμός». Έτσι, η εύρεση του εκλογικού κέντρου με την υπηρεσία αυτή μπορεί να γίνει με δυο τρόπους. Στον πρώτο τρόπο θα πρέπει οι πολίτες να εισάγουν τον «ΕΕΑ» και το επώνυμό τους ολογράφως. Ενώ στον δεύτερο τρόπο θα πρέπει να εισάγουν ολογράφως το επώνυμό τους, τα δυο πρώτα γράμματα από το όνομα, το πατρώνυμο και το μητρώνυμό τους και το έτος της ημερομηνίας γέννησής τους. Μάλιστα, αξίζει να σημειωθεί ότι το μητρώνυμο είναι προαιρετικό και για τις δυο αυτές περιπτώσεις. Οπότε, αν τα στοιχεία εισαχθούν σωστά με έναν από τους δυο παραπάνω τρόπους, τότε επιστρέφονται ο «ΕΕΑ», τα πλήρη στοιχεία του πολίτη, ήτοι ολογράφως το ονοματεπώνυμο, το πατρώνυμο, το μητρώνυμο και ο αριθμός δημοτολογίου του, και αναλυτικά ο τόπος άσκησης του εκλογικού του δικαιώματος.

ΕΙΚΟΝΑ 15: ΥΠΕΣ – ΔΙΑΔΙΚΤΥΑΚΗ ΥΠΗΡΕΣΙΑ ΣΤΟΙΧΕΙΩΝ ΕΚΛΟΓΙΚΟΥ ΣΩΜΑΤΟΣ ΕΛΛΗΝΩΝ ΕΚΛΟΓΕΩΝ

❖ Υπουργείο Εσωτερικών (Εθνικό Δημοτολόγιο)

Παράλληλα, το Υπουργείο Εσωτερικών έχει αναπτύξει άλλη μια διαδικτυακή υπηρεσία, με την οποία ο εκάστοτε πολίτης μπορεί να βρει σε πιο δήμο είναι εγγεγραμμένος [211] (Εικόνα 16). Για την είσοδο στην υπηρεσία αυτή απαιτείται η εισαγωγή του επωνύμου ολογράφως, τα δυο πρώτα γράμματα από το όνομα, το πατρώνυμο και το μητρώνυμο και το έτος της ημερομηνίας γέννησης. Όμοια, και σε

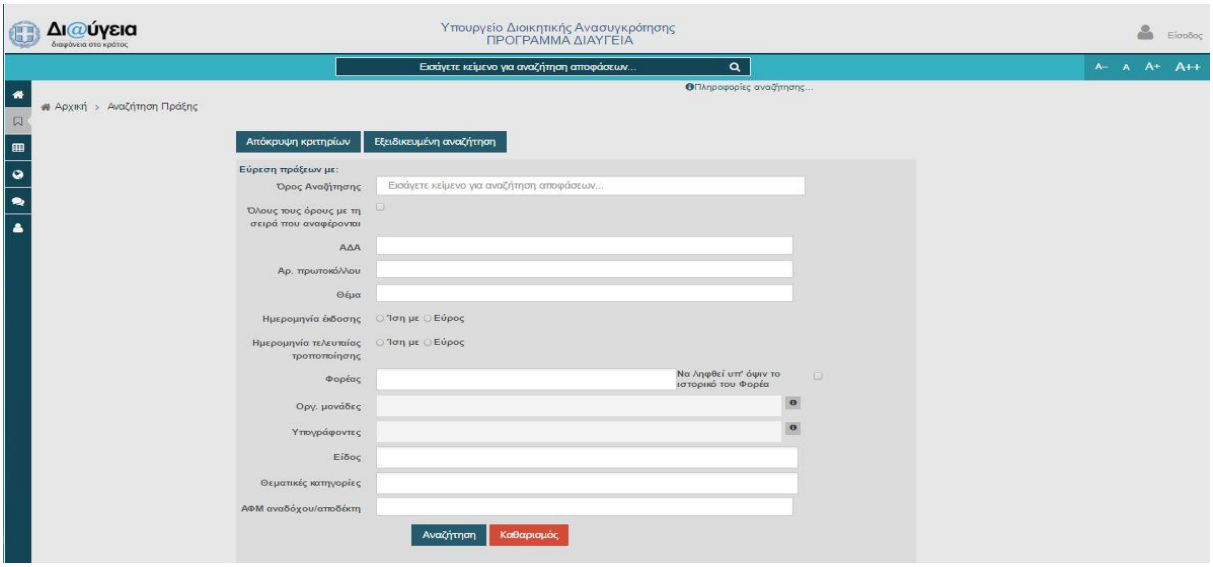
αυτή την υπηρεσία το μητρώνυμο είναι προαιρετικό. Αν τα στοιχεία είναι σωστά, τότε επιστρέφονται ολογράφως το ονοματεπώνυμο του πολίτη, το πατρώνυμο, το μητρώνυμο, η πλήρης ημερομηνία γέννησης, ο δήμος και ο νομός στον οποίο είναι εγγεγραμμένος, καθώς και ο «Αριθμός Δημοτολογίου» του.

ΕΙΚΟΝΑ 16: ΥΠΕΣ – ΔΙΑΔΙΚΤΥΑΚΗ ΥΠΗΡΕΣΙΑ ΔΗΜΟΤΟΛΟΓΙΚΩΝ ΣΤΟΙΧΕΙΩΝ ΕΚΛΟΓΙΚΟΥ ΣΩΜΑΤΟΣ

❖ Υπουργείο Διοικητικής Ανασυγκρότησης (Διαύγεια)

Το πρόγραμμα «Διαύγεια» [212] (Εικόνα 17) αποτελεί μια διαδικτυακή υπηρεσία του Υπουργείου Διοικητικής Ανασυγκρότησης από τον Οκτώβριο του 2010, σε εφαρμογή του «Νόμου 3861/2010». Στην υπηρεσία αυτή αναρτώνται υποχρεωτικά, στα πλαίσια της διαφάνειας και της λογοδοσίας κατά την άσκηση της δημόσιας διοίκησης και εξουσίας, όλες οι αποφάσεις και πράξεις όλων των Οργανισμών και των Φορέων Διοίκησης του Ελληνικού Κράτους. Μάλιστα, σε αυτή αναρτούν στοιχεία 4538 φορείς του στενού και του ευρύτερου Δημοσίου τομέα και το σύνολο αυτών ανέρχεται μέχρι τώρα στον αριθμό των 23.300.000 πράξεων, σύμφωνα με τα στατιστικά της στοιχεία. Ειδικότερα, στην υπηρεσία αυτή δημοσιεύουν τις πράξεις τους διάφοροι φορείς, όπως Υπουργεία, Οργανισμοί Τοπικής Αυτοδιοίκησης Α΄ και Β΄ βαθμού, Πανεπιστήμια, Ανεξάρτητες Αρχές, Αποκεντρωμένες Διοικήσεις, Ανώνυμες Εταιρείες, Ιδρύματα, Μ.Κ.Ο., Νοσοκομεία, ΔΕΚΟ κλπ. Επίσης, λαμβάνεται ειδική μέριμνα για τις πράξεις που αφορούν στην εθνική άμυνα και τα ευαίσθητα προσωπικά δεδομένα, σύμφωνα με την σελίδα των πληροφοριών για την υπηρεσία αυτή [213]. Από το 2014, η εν λόγω υπηρεσία αναβαθμίστηκε στο «Διαύγεια ΙΙ» και παρέχει βελτιώσεις συγκριτικά με την προηγούμενη έκδοση σε θέματα που σχετίζονται με την αναζήτηση και τα κριτήρια αναζήτησης, την ευχρηστία της

υπηρεσίας, την πληρότητα και την ποιότητα των παρεχόμενων πληροφοριών, την διασφάλιση της εγκυρότητας και της ακεραιότητας των εγγράφων που παρέχει στους πολίτες και την διασύνδεσή της με τα υπόλοιπα πληροφοριακά συστήματα του δημοσίου τομέα. Σε αυτή την υπηρεσία έχουν δικαίωμα πρόσβασης ανεξαιρέτως όλοι οι πολίτες του κράτους, προκειμένου να λάβουν γνώση για τα πεπραγμένα της δημόσιας διοίκησης, αλλά και να ασκήσουν έλεγχο σε αυτή. Παράλληλα, γίνεται εύκολα αντιληπτό, ότι σε αυτή βρίσκεται αποθηκευμένο πλήθος πληροφοριών διαφόρων τύπων. Συνεπώς, διαπιστώνεται ότι το πρόγραμμα «Διαύγεια» έχει τα χαρακτηριστικά ενός περιβάλλοντος «Big Data» από μόνο του, καθώς διαχειρίζεται και παρέχει στους πολίτες τις πράξεις και τις αποφάσεις από τα διάφορα ανεξάρτητα πληροφοριακά συστήματα των φορέων και υπηρεσιών του ευρύτερου δημοσίου τομέα, κατόπιν διασύνδεσής τους σε μία ενιαία υπηρεσία. Μάλιστα, η υπηρεσία αυτή υποστηρίζει τρείς τρόπους αναζήτησης, οι οποίοι επικεντρώνονται στους τίτλους των εγγράφων ή/και στα μεταδεδομένα τους. Ο πρώτος τρόπος είναι η «Ελεύθερη Αναζήτηση», στην οποία εισάγεται μία ή και περισσότερες λέξεις αναζήτησης στο κεντρικό πεδίο αναζήτησης. Ο δεύτερος τρόπος είναι η «Σύνθετη Αναζήτηση», στην οποία ο χρήστης πρέπει να συμπληρώσει τα συγκεκριμένα πεδία αυτής για μια πιο εστιασμένη αναζήτηση. Και στις δυο περιπτώσεις ο χρήστης μπορεί στα επιστρεφόμενα αποτελέσματα να εξειδικεύσει περαιτέρω την αναζήτηση του με την εφαρμογή μιας σειράς επιπλέον φίλτρων που εμφανίζονται. Ο τρίτος τρόπος είναι μέσω μιας διεπαφής, η οποία δίνει την δυνατότητα αναζητήσεων προγραμματιστικά μέσω εφαρμογών που έχουν αναπτυχθεί για τον σκοπό αυτό.

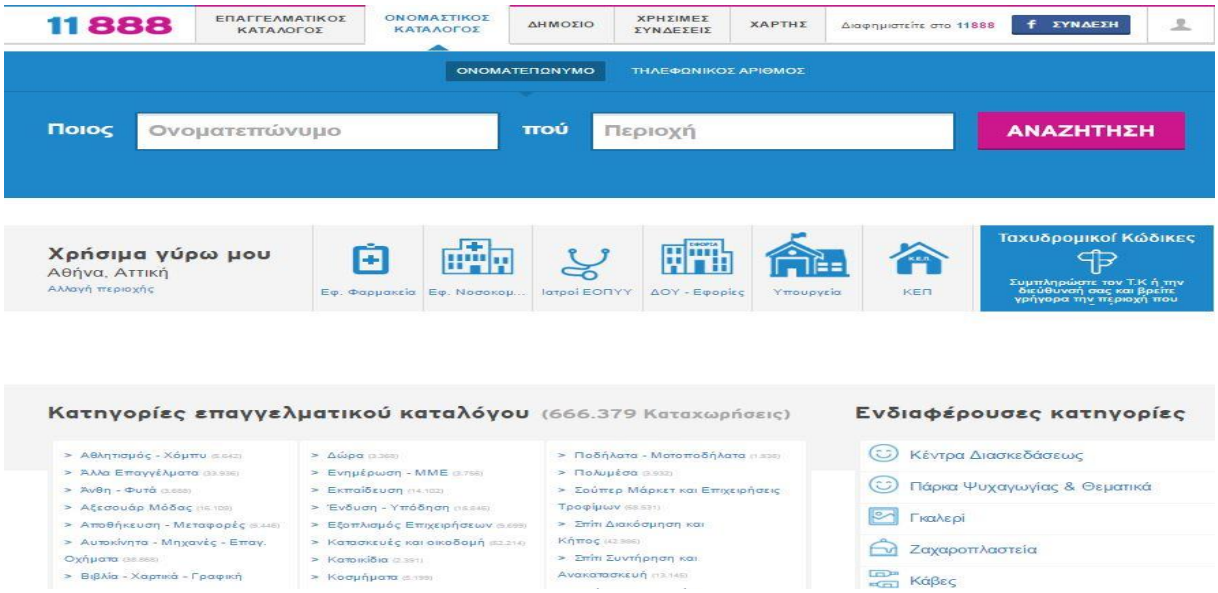


ΕΙΚΟΝΑ 17: ΥΔΑ – ΠΡΟΓΡΑΜΜΑ ΔΙ@ΥΓΕΙΑ

Ενώ αντίστοιχα, η μη κυβερνητική πηγή δεδομένων είναι η εξής:

❖ **Υπηρεσία Τηλεφωνικού Καταλόγου ΟΤΕ (Τηλέφωνο)**

Οι υπηρεσίες τηλεφωνικών καταλόγων χρησιμοποιούνται για την εύρεση κινητών και σταθερών τηλεφώνων. Μάλιστα σε κάποιες περιπτώσεις, αυτές παρέχονται και μέσω διαδικτύου, όπως για παράδειγμα το «Ευρετήριο Τηλεφωνικού Καταλόγου» του ΟΤΕ [214], του κύριου τηλεπικοινωνιακού φορέα στην Ελλάδα (Εικόνα 18). Συγκεκριμένα, η υπηρεσία αυτή παρέχει την δυνατότητα στον οποιονδήποτε να αναζητήσει, είτε επαγγελματίες στην καρτέλα «Επαγγελματικός Κατάλογος», είτε και ιδιώτες στην καρτέλα «Ονομαστικός Κατάλογος». Ειδικότερα, για την αναζήτηση ιδιωτών στην καρτέλα «Ονομαστικός Κατάλογος» παρέχονται δυο τρόποι. Στον πρώτο τρόπο η αναζήτηση γίνεται με βάση το «Ονοματεπώνυμο» και ενδεχομένως την «Περιοχή», όπως φαίνεται στην καρτέλα «Ονοματεπώνυμο», ενώ στον δεύτερο τρόπο υποστηρίζεται η αντίστροφη αναζήτηση, δηλαδή η εύρεση σε ποιόν ανήκει κάποιος αριθμός τηλεφώνου, όπως φαίνεται στην καρτέλα «Τηλεφωνικός Αριθμός». Ωστόσο, και στις δυο περιπτώσεις αυτό που επιστρέφεται είναι η αντιστοίχιση του τηλεφωνικού αριθμού με το πλήρες ονοματεπώνυμο του νόμιμου κατόχου του. Επίσης, στο αποτέλεσμα αυτό αναγράφονται και επιπλέον πληροφορίες, όπως το αρχικό γράμμα του πατρώνυμου, η πλήρης διεύθυνση διαμονής του και η τοποθεσία στον χάρτη.



ΕΙΚΟΝΑ 18: ΟΤΕ – ΕΥΡΕΤΗΡΙΟ ΤΗΛΕΦΩΝΙΚΟΥ ΚΑΤΑΛΟΓΟΥ

4.3 ΕΠΕΞΕΡΓΑΣΙΑ ΠΗΓΩΝ ΚΑΙ ΕΞΑΓΩΓΗ ΔΕΔΟΜΕΝΩΝ

Στην συνέχεια, εξετάζεται αν και κατά πόσο είναι δυνατό να εξαχθούν προσωπικά δεδομένα, «Αναγνωριστικά Στοιχεία Ταυτότητας» (Personally Identifiable Information-PII) και «Μοναδικοί Αριθμοί Ταυτοποίησης» (Unique Identification Number–UIN/UID), είτε απευθείας από τις ανωτέρω πηγές δεδομένων, είτε κατόπιν του συνδυασμού των δεδομένων των πηγών αυτών μεταξύ τους.

Συγκεκριμένα, ο σκοπός της παρούσας ενότητας είναι να εξακριβωθεί η ακρίβεια των κάτωθι ισχυρισμών:

✓ Αν είναι εφικτό να επιβεβαιωθούν οι πληροφορίες που έχουν συλλεχθεί για κάποιο πολίτη από κάποια άλλη πηγή.

✓ Αν μπορεί κάποιος να εξάγει οποιαδήποτε πληροφορία, που σχετίζεται με πραγματικές ταυτότητες πολιτών, μέσα από αλλεπάλληλες δοκιμές διαφόρων τιμών στα αντίστοιχα πεδία της κάθε διαδικτυακής υπηρεσίας των ανωτέρω πηγών (Brute Force), οι οποίες χρησιμοποιούνται ως πειράματα.

✓ Αν εξάγεται επιπλέον πληροφορία από ότι θα έπρεπε στα αποτελέσματα που επιστρέφονται από αυτές τις διαδικτυακές υπηρεσίες.

✓ Αν είναι δυνατό μέσα από τον συνδυασμό των πληροφοριών των διαφόρων πηγών μεταξύ τους, να εξαχθούν και άλλες επιπλέον πληροφορίες.

Η πρώτη πηγή δεδομένων που εξετάζεται είναι η διαδικτυακή υπηρεσία για τον έλεγχο της εγκυρότητας του «ΑΦΜ» [206]. Αυτή αναπτύχθηκε προκειμένου να ελέγχεται κάθε φορά η εγκυρότητα των «ΑΦΜ» της δεύτερης κατηγορίας, ενώ αντίστοιχα αυτό δεν υποστηρίζεται για τα «ΑΦΜ» της πρώτης. Έτσι, αν αυτό ανήκει στην δεύτερη κατηγορία, τότε επιστρέφονται το ονοματεπώνυμο, η επωνυμία και η διεύθυνση που αντιστοιχούν στο συγκεκριμένο «ΑΦΜ», ενώ αν ανήκει στην πρώτη, τότε επιστρέφεται μήνυμα ότι το εν λόγω «ΑΦΜ» είναι μη έγκυρος αριθμός, ακόμα και αν αυτός ο «ΑΦΜ» είναι όντως υπαρκτός. Από την διαφοροποίηση αυτή όμως μπορεί να εξαχθεί έμμεσο συμπέρασμα για την εργασιακή κατάσταση κάποιου πολίτη. Οπότε σε περίπτωση εισαγωγής ενός «ΑΦΜ», αν επιστραφεί αποτέλεσμα, αμέσως εξάγεται το συμπέρασμα ότι ανήκει στην δεύτερη κατηγορία και μάλιστα στο αποτέλεσμα θα συμπεριλαμβάνονται το ονοματεπώνυμο, η επωνυμία και η διεύθυνση που αντιστοιχούν σε αυτό. Δηλαδή, στην περίπτωση που δεν ήταν

γνωστά τα ανωτέρω στοιχεία, αυτά διατίθενται από την ίδια την υπηρεσία, γεγονός που σημαίνει ότι εξάγεται περισσότερη πληροφορία από την ήδη γνωστή κάποιες φορές. Αυτό αποκτά ιδιαίτερη σημασία για τους ελεύθερους επαγγελματίες και τους αυτοαπασχολούμενους, όπως για παράδειγμα τους υδραυλικούς, τους αγρότες κλπ, στον οποίων το «ΑΦΜ» συνήθως αντιστοιχεί η διεύθυνση διαμονής τους και το πραγματικό τους ονοματεπώνυμο. Αν δεν επιστραφεί αποτέλεσμα τότε υπάρχουν δυο ενδεχόμενα, είτε να μην υπάρχει όντως ο συγκεκριμένος «ΑΦΜ», είτε να αντιστοιχεί σε φυσικό πρόσωπο της πρώτης κατηγορίας. Αν όμως είναι επιβεβαιωμένο ότι ο εν λόγω «ΑΦΜ» όντως υπάρχει, το συμπέρασμα που εξάγεται εμμέσως είναι ότι ο πολίτης, στον οποίο αυτός αντιστοιχεί, ανήκει στην πρώτη κατηγορία, ήτοι μπορεί να είναι για παράδειγμα ή μισθωτός, ή συνταξιούχος. Μάλιστα, στην περίπτωση που διατίθενται επιπλέον στοιχεία, όπως το ονοματεπώνυμο, το πατρώνυμο, το μητρώνυμο, η πλήρης ημερομηνία γέννησης και ο «ΑΔΤ», τότε μπορεί να εξακριβωθεί η εγκυρότητά του με την χρήση της υπηρεσίας εύρεσης του «ΑΜΚΑ».

Η δεύτερη πηγή δεδομένων σχετίζεται με τις ταυτότητες που εκδίδονται από την Ελληνική Αστυνομία στους πολίτες του Ελληνικού κράτους. Όπως προαναφέρθηκε, σε αυτές συμπεριλαμβάνονται το ονοματεπώνυμο του κατόχου της, τα ονοματεπώνυμα του πατέρα και της μητέρας του, η ημερομηνία και ο τόπος γέννησής του, το ύψος του, ο Δήμος στον οποίο είναι εγγεγραμμένος με τον αριθμό δημοτολογίου του, η φωτογραφία του, η υπογραφή του, η ομάδα αίματος του και ο «Αριθμός Δελτίου Ταυτότητας». Προφανώς, η ασφάλεια αυτής της πηγής δεδομένων ισοδυναμεί με την παρεχόμενη ασφάλεια, που απορρέει από την φυσική κατοχή της από τον νόμιμο ιδιοκτήτη της. Ωστόσο, αν και δεν υπάρχει κάποια διαδικτυακή υπηρεσία που να σχετίζεται με τις ταυτότητες, το γεγονός, ότι αυτή χρησιμοποιείται ευρέως στις περισσότερες συναλλαγές του δημόσιου βίου και μάλιστα στο πλήθος των περιπτώσεων εκδίδονται αντίγραφα αυτής, εγκυμονεί κινδύνους. Πέρα από τον προφανή κίνδυνο να δημιουργηθεί κάποια πλαστή ταυτότητα με τα έγκυρα αυτά στοιχεία, οι πληροφορίες που αναγράφονται σε αυτή μπορεί να οδηγήσουν στην αποκάλυψη και άλλων πληροφοριών, που σχετίζονται με τον νόμιμο κάτοχό της. Τέτοιες πληροφορίες που μπορούν άμεσα να εξαχθούν είναι ο «ΕΕΑ», ο αριθμός τηλεφώνου με την διεύθυνση, ο αριθμός δημοτολογίου με τον Δήμο εγγραφής και ο «ΑΜΚΑ» με την χρήση των αντίστοιχων διαδικτυακών υπηρεσιών. Μάλιστα στην περίπτωση του «ΑΜΚΑ», επειδή ζητείται και το «ΑΦΜ», αυτό μπορεί να βρεθεί

σχετικά εύκολα, απλά με την εκτέλεση αλλεπάλληλων δοκιμών για όλους τους πιθανούς συνδυασμούς των 9ψήφιων αριθμών (Brute Force). Επιπλέον, τα στοιχεία αυτά μπορεί να χρησιμοποιηθούν και στα προγράμματα «Διαύγεια» και «Υπερδιαύγεια», προκειμένου να εξαχθούν επιπλέον πληροφορίες για τον ιδιοκτήτη της συγκεκριμένης ταυτότητας, αν αυτός συμπεριλαμβάνεται σε κάποιο από τα ανηρτημένα έγγραφα της Δημόσιας Διοίκησης.

Η τρίτη πηγή δεδομένων που μελετάται είναι η διαδικτυακή υπηρεσία για την εύρεση του «ΑΜΚΑ». Όπως προαναφέρθηκε, τα στοιχεία που απαιτούνται, προκειμένου να επιστραφεί ο «ΑΜΚΑ», είναι το ονοματεπώνυμο, το πατρώνυμο, το μητρώνυμο, η πλήρης ημερομηνία γέννησης, ο «ΑΦΜ» και ο «ΑΔΤ». Κατά την χρήση όμως της υπηρεσίας αυτής διαπιστώνονται κάποιες παρατυπίες, όσον αφορά την λειτουργία της, οι οποίες είναι οι εξής:

- Για να είναι έγκυρη η μετάβαση από το πρώτο στο δεύτερο βήμα, θα πρέπει στα πεδία «Επώνυμο», «Όνομα», «Όνομα Πατέρα» και «Όνομα Μητέρας» να υπάρχουν τουλάχιστον 2 χαρακτήρες, ενώ αντίστοιχα το πεδίο «Ημερομηνία Γέννησης» θα πρέπει να είναι συμπληρωμένο με την πλήρη ημερομηνία γέννησης. Ωστόσο, προκειμένου να είναι εφικτός ο έλεγχος και να επιστραφούν αποτελέσματα, θα πρέπει κατ' ελάχιστο να είναι συμπληρωμένα ολογράφως το πεδίο «Επώνυμο» και να έχει δοθεί η πλήρης ημερομηνία γέννησης. Στην περίπτωση αυτή, αν τα στοιχεία αντιστοιχούν σε κάποιο πολίτη, ο οποίος έχει εγγραφεί στο Εθνικό Μητρώο «ΑΜΚΑ», τότε η υπηρεσία αυτή μεταβαίνει στο δεύτερο βήμα, στο οποίο ζητάει την εισαγωγή των έγκυρων τιμών του «ΑΔΤ» και του «ΑΦΜ» του, προκειμένου να εκτυπώσει στο τρίτο βήμα τον «ΑΜΚΑ» του. Αντιθέτως, αν αυτά δεν αντιστοιχούν σε κάποιο εγγεγραμμένο πολίτη στο Εθνικό Μητρώο «ΑΜΚΑ», τότε επιστρέφεται το αντίστοιχο μήνυμα. Αυτό το γεγονός όμως μπορεί να μετατρέψει την υπηρεσία αυτή σε πείραμα. Αφενός μπορεί να χρησιμοποιηθεί για να εξακριβωθεί η εγκυρότητα των στοιχείων κάποιου πολίτη, ακόμα και αν αυτά δεν είναι πλήρη. Έτσι για παράδειγμα, είναι δυνατό να ελεγχθεί η εγκυρότητα των στοιχείων κάποιου πολίτη, με την προϋπόθεση ότι είναι γνωστά το επώνυμο και η ημερομηνία γέννησής του, ακόμα και αν δεν είναι γνωστές οι πλήρεις τιμές του ονόματος, του πατρώνυμου και του μητρώνυμού του, παρά μόνο το αρχικό ή τα 2 πρώτα τους γράμματα. Μάλιστα, στην περίπτωση που είναι γνωστά μόνο τα πρώτα γράμματα, είναι εύκολο να βρεθούν και τα δεύτερα γράμματα απλά με την εκτέλεση αλλεπάλληλων δοκιμών των 24 γραμμάτων του ελληνικού αλφάβητου. Αυτό το γεγονός οδηγεί στην αποκάλυψη

επιπλέον πληροφορίας, η οποία δεν ήταν προηγουμένως γνωστή, καθώς πλέον με τα δυο γράμματα γνωστά είναι πιο εύκολο να βρεθούν τα ακριβή ονόματα. Αφετέρου μπορεί να χρησιμοποιηθεί για να πιστοποιηθεί ότι όντως ο συνδυασμός κάποιων συγκεκριμένων τυχαίων τιμών των αντίστοιχων πεδίων της υπηρεσίας αυτής αντιστοιχεί σε πραγματικό φυσικό πρόσωπο. Μάλιστα, στην περίπτωση αυτή, εκτός από το ότι διαπιστώνεται η ύπαρξη κάποιου πραγματικού προσώπου, ταυτόχρονα γίνονται γνωστά και τα προσωπικά του δεδομένα, τα οποία προφανώς δεν ήταν γνωστά πριν. Ωστόσο, η εκτέλεση αλλεπάλληλων δοκιμών (Brute Force), προκειμένου να βρεθεί ποιοι συνδυασμοί τιμών των πεδίων της υπηρεσίας αυτής αντιστοιχούν σε πραγματικό φυσικό πρόσωπο, είναι χρονοβόρα και πολύπλοκη. Εντούτοις, στις περισσότερες περιπτώσεις συνήθως είναι γνωστά κάποια στοιχεία, όπως για παράδειγμα το ονοματεπώνυμο, οπότε η πολυπλοκότητα μειώνεται σημαντικά. Μάλιστα, μέσα από την διαδικασία αυτή αποκαλύπτονται και επιπλέον πληροφορίες για το συγκεκριμένο άτομο, οι οποίες δεν ήταν προηγουμένως γνωστές.

- Επιπλέον, στο δεύτερο βήμα της υπηρεσίας αυτής παρατηρήθηκε ότι, αν ένας πολίτης ανήκει σε κάποια ειδική κατηγορία, είναι για παράδειγμα στρατιωτικός, τότε αντί για το πεδίο «Αριθμός Ταυτότητας» αναγράφεται «Αριθμός Στρατιωτικής Ταυτότητας». Αυτό έχει δυο σοβαρές συνέπειες. Αφενός αποκαλύπτει το επάγγελμα του εν λόγω φυσικού προσώπου, οδηγώντας έτσι στην απόκτηση επιπλέον γνώσης, η οποία δεν ήταν προηγουμένως γνωστή. Αφετέρου παρέχεται επιπλέον πληροφορία στον εκάστοτε επιτιθέμενο, ως προς το τι τύπου δοκιμές θα πρέπει να κάνει κάθε φορά, χρησιμοποιώντας την υπηρεσία αυτή ως πείραμα, προκειμένου να βρει κάποια έγκυρη τιμή για το πεδίο αυτό, με την προϋπόθεση ότι όλα τα υπόλοιπα στοιχεία είναι γνωστά. Προφανώς, η εφαρμογή του αλγορίθμου εύρεσης της τιμής του «ΑΔΤ», ο οποίος έχει την μορφή (ΑΑ111111), δεν θα είχε απόδοση για την εύρεση της έγκυρης τιμής του «Αριθμού Στρατιωτικής Ταυτότητας», ο οποίος ακολουθεί διαφορετική τεχνική από αυτή. Επιπρόσθετα, δεν υφίσταται λόγος να γίνεται τέτοια διευκρίνηση, καθώς ο εκάστοτε πολίτης γνωρίζει ποια τιμή θα πρέπει να βάλλει κάθε φορά στο αντίστοιχο πεδίο.
- Επίσης, με την προϋπόθεση ότι είναι γνωστά όλα τα λοιπά στοιχεία, είναι δυνατό να βρεθεί ο έγκυρος «ΑΦΜ» με την εκτέλεση αλλεπάλληλων δοκιμών, χρησιμοποιώντας την υπηρεσία αυτή ως πείραμα.
- Τέλος, για να ολοκληρωθεί η μετάβαση από το δεύτερο στο τρίτο βήμα και με την προϋπόθεση ότι όλα τα στοιχεία του πρώτου βήματος είναι συμπληρωμένα σωστά, απαιτείται να δοθούν οι σωστές και έγκυρες τιμές των «ΑΔΤ» και «ΑΦΜ»,

που αντιστοιχούν στο συγκεκριμένο πολίτη. Ωστόσο παρατηρήθηκε ότι, αν σε ένα από τα δυο αυτά πεδία δοθεί η έγκυρη και σωστή τιμή, ενώ στο άλλο κάποια έγκυρη μεν, αλλά όχι η σωστή τιμή, τότε η διαδικασία ολοκληρώνεται επιτυχώς και επιστρέφεται ο αντίστοιχος «ΑΜΚΑ», παρόλο που η μια από τις δυο αυτές τιμές δεν αντιστοιχεί στον εν λόγω πολίτη και δεν είναι σωστή. Έτσι για παράδειγμα, με δεδομένο ότι τα πεδία του πρώτου βήματος είναι συμπληρωμένα σωστά, αν δοθεί το σωστό «ΑΦΜ» και ένα έγκυρο αλλά όχι το σωστό «ΑΔΤ», τότε η υπηρεσία θα επιστρέψει τον «ΑΜΚΑ» του πολίτη αυτού. Το ίδιο ισχύει και αντιστρόφως, δηλαδή αν είναι σωστό το «ΑΔΤ», ενώ το «ΑΦΜ» είναι μεν έγκυρο αλλά όχι το σωστό.

Οι επόμενες πηγές που εξετάζονται είναι οι διαδικτυακές υπηρεσίες του Υπουργείου Εσωτερικών, ήτοι η υπηρεσία «Δημοτολογικά Στοιχεία Εκλογικού Σώματος» και η υπηρεσία «Στοιχεία Εκλογικού Σώματος Ελλήνων Εκλογέων». Αξίζει να σημειωθεί ότι και στις δύο υπηρεσίες, προκειμένου να επιστραφούν αποτελέσματα στον εκάστοτε χρήστη τους, απαιτείται να είναι συμπληρωμένα ολογράφως το πεδίο «Επώνυμο», τα 2 πρώτα γράμματα τουλάχιστον στα πεδία «Όνομα», «Όνομα Πατέρα» και «Όνομα Μητέρας» και το έτος γέννησης. Μάλιστα, και για τις δυο υπηρεσίες το πεδίο «Όνομα Μητέρας» είναι προαιρετικό και συνεπώς δεν απαιτείται η συμπλήρωσή του. Επιπλέον, η υπηρεσία «Στοιχεία Εκλογικού Σώματος Ελλήνων Εκλογέων» υποστηρίζει εναλλακτικά την εισαγωγή του «ΕΑΑ» και του επωνύμου ολογράφως στα αντίστοιχα πεδία, για να επιστρέψει αποτελέσματα στον εκάστοτε χρήστη. Επίσης, και οι δυο υπηρεσίες, στην περίπτωση που οι τιμές που θα δοθούν στα αντίστοιχα πεδία δεν αντιστοιχούν σε κάποιο υπαρκτό φυσικό πρόσωπο που να είναι καταχωρημένο σε αυτές, επιστρέφουν το μήνυμα ότι δεν βρέθηκαν εγγραφές. Αυτές όμως οι λειτουργίες εγκυμονούν κινδύνους, καθώς μπορεί κάλλιστα οι υπηρεσίες αυτές να μετατραπούν σε πειράματα, τόσο για την εξακρίβωση της εγκυρότητας των στοιχείων κάποιου πολίτη, όσο και για την εύρεση πολιτών μαζί με τις τιμές των αντίστοιχων χαρακτηριστικών τους, τα οποία δεν ήταν προηγουμένως γνωστά, κατόπιν εκτέλεσης αλλεπάλληλων δοκιμών με διάφορες τιμές στα αντίστοιχα πεδία των υπηρεσιών αυτών. Μάλιστα, τα χαρακτηριστικά αυτά μπορούν να χρησιμοποιηθούν περαιτέρω και σε άλλες διαδικτυακές υπηρεσίες, προκειμένου να εξαχθεί επιπλέον πληροφορία. Αναλυτικότερα, οι υπηρεσίες αυτές μπορεί να χρησιμοποιηθούν ως εξής:

- Προφανώς και στις δυο υπηρεσίες είναι εύκολο να εξακριβωθεί η ακρίβεια της εγκυρότητας των στοιχείων κάποιου πολίτη, απλά με την εισαγωγή των αντίστοιχων

τιμών στα πεδία της υπηρεσίας και τον έλεγχο του αποτελέσματός της. Έτσι, αν δεν επιστραφεί αποτέλεσμα, τότε ο πολίτης με τα χαρακτηριστικά αυτά δεν υφίσταται. Ενώ αν επιστραφεί, τότε όχι μόνο επιβεβαιώνεται η ύπαρξή του, αλλά επιπλέον και οι δυο υπηρεσίες αποκαλύπτουν πληροφορίες, οι οποίες δεν ήταν προηγουμένως γνωστές. Συγκεκριμένα η υπηρεσία «Δημοτολογικά Στοιχεία Εκλογικού Σώματος» αποκαλύπτει το πλήρες όνομα, πατρώνυμο, μητρώνυμο και ημερομηνία γέννησης, καθώς και τον Δήμο που είναι εγγεγραμμένος με τον αριθμό δημοτολογίου του. Ενώ αντίστοιχα, η υπηρεσία «Στοιχεία Εκλογικού Σώματος Ελλήνων Εκλογέων» αποκαλύπτει τα ίδια στοιχεία με την προηγούμενη υπηρεσία, με την διαφορά ότι σε αυτά συμπεριλαμβάνεται ο «ΕΑΑ» και όχι η πλήρης ημερομηνία γέννησής του.

- Επίσης, και στις δυο υπηρεσίες είναι δυνατό να εκτελεστούν αλλεπάλληλες δοκιμές, προκειμένου να αποκαλυφθούν υπαρκτά φυσικά πρόσωπα με τα αντίστοιχα χαρακτηριστικά τους. Οι δοκιμές αυτές απαιτούν κατ' ελάχιστο να είναι γραμμένο κάθε φορά ολογράφως το επώνυμο, να υπάρχουν τα δυο πρώτα γράμματα στα πεδία «Όνομα» και «Πατρώνυμο» και να έχει δοθεί μόνο το έτος γέννησης. Έτσι, στην περίπτωση που θα επιστραφεί αποτέλεσμα, τότε επιβεβαιώνεται η ύπαρξη του φυσικού προσώπου και μάλιστα αποκαλύπτονται τα πλήρη χαρακτηριστικά του. Τα χαρακτηριστικά αυτά ποικίλουν ανάλογα με την υπηρεσία που χρησιμοποιήθηκε για την εκτέλεση των δοκιμών. Προφανώς, η εκτέλεση της ενέργειας αυτής είναι χρονοβόρα και πολύπλοκη. Εντούτοις, στις περισσότερες περιπτώσεις κάποια από τα στοιχεία αυτά είναι γνωστά, οπότε η πολυπλοκότητα μειώνεται σημαντικά, ενώ ταυτόχρονα αυξάνει η απόδοσή της.
- Επιπρόσθετα, όσον αφορά την υπηρεσία «Στοιχεία Εκλογικού Σώματος Ελλήνων Εκλογέων», η αποκάλυψη υπαρκτών φυσικών προσώπων με τα αντίστοιχα χαρακτηριστικά τους μπορεί να πραγματοποιηθεί εναλλακτικά μόνο με την χρήση των πεδίων «Επώνυμο» και «ΕΕΑ». Έτσι, με αλλεπάλληλες δοκιμές τυχαίων τιμών στα πεδία αυτά, ανάλογα με το αποτέλεσμα, επιβεβαιώνεται η ύπαρξη ή όχι κάποιου φυσικού προσώπου. Μάλιστα, αποκαλύπτονται και επιπλέον πληροφορίες για αυτόν, οι οποίες δεν ήταν προηγουμένως γνωστές.

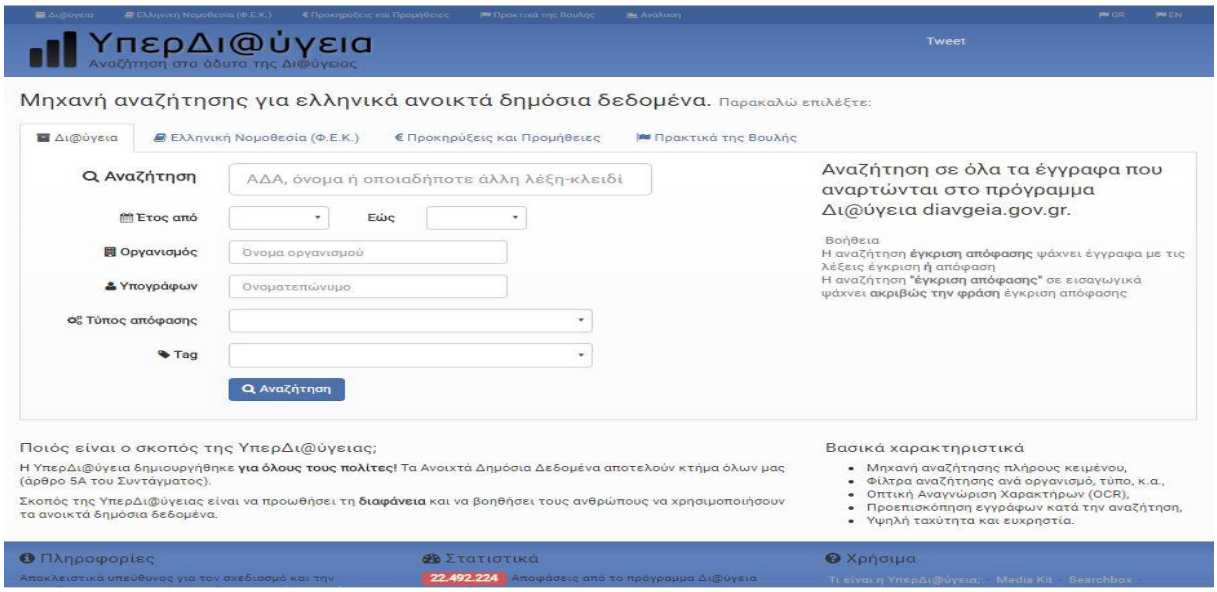
Έπειτα μελετάται η διαδικτυακή υπηρεσία «Διαύγεια» του Υπουργείου Διοικητικής Ανασυγκρότησης, η οποία δίδει την δυνατότητα αναζήτησης κάποιου εγγράφου ανάμεσα σε όλες τις αποφάσεις και πράξεις του ευρύτερου Δημοσίου τομέα. Η υπηρεσία αυτή διαθέτει τρείς τρόπους αναζήτησης, οι οποίοι εστιάζουν στους τίτλους και τα μεταδεδομένα των εγγράφων αυτών. Ωστόσο, οι μηχανισμοί αυτοί δεν

υποστηρίζουν την αναζήτηση και μέσα στο περιεχόμενο των ανηρτημένων αυτών εγγράφων. Έτσι, προκειμένου να εκτελεστεί εις βάθος έρευνα και να διερευνηθεί αν πράγματι στα ανηρτημένα δημόσια έγγραφα του προγράμματος «Διαύγεια» συμπεριλαμβάνονται προσωπικά δεδομένα και τι είδους, χρησιμοποιήθηκε συμπληρωματικά με τους μηχανισμούς αυτούς και το εργαλείο «Υπερδιαύγεια» [219] (Εικόνα 19). Το εργαλείο αυτό, σύμφωνα με τον κατασκευαστή του, υποστηρίζει γρήγορους και εύχρηστους μηχανισμούς αναζήτησης, οι οποίοι εστιάζουν, εκτός από τους τίτλους και τα μεταδεδομένα των εγγράφων αυτών, και στο περιεχόμενό τους, παρόλο που τα αρχεία αυτά αναρτώνται σε μορφή «PDF». Αυτό επιτυγχάνεται χάρη στην χρήση των μηχανισμών «Οπτικής Αναγνώρισης Χαρακτήρων» (OCR), οι οποίοι επιτρέπουν την αναζήτηση μιας συγκεκριμένης συμβολοσειράς (String) μέσα στο περιεχόμενο του εγγράφου με την σάρωσή του. Αυτό συνεπάγεται την αύξηση του εύρους των πληροφοριών που είναι δυνατό να αναζητηθούν και να αντληθούν από ένα έγγραφο. Μάλιστα, οι δυνατότητες αναζήτησης του εργαλείου αυτού επεκτείνονται, εκτός από τα έγγραφα του προγράμματος «Διαύγεια», και στα έγγραφα της Ελληνικής Νομοθεσίας, των Προκηρύξεων και Προμηθειών που δημοσιεύονται στο «Κεντρικό Ηλεκτρονικό Μητρώο Δημόσιων Συμβάσεων» (ΚΗΜΔΗΣ) και των Πρακτικών της Βουλής. Συνεπώς, προκειμένου να εξεταστεί αν τα έγγραφα που αναρτώνται στο πρόγραμμα «Διαύγεια» περιέχουν προσωπικά δεδομένα, ή/και ευαίσθητα προσωπικά δεδομένα, εκτελέστηκε μια αλληλουχία αναζητήσεων με την δοκιμή κάθε φορά στα εργαλεία αυτά διαφορετικής συμβολοσειράς (String). Οι συμβολοσειρές που χρησιμοποιήθηκαν είναι «ΑΜΚΑ», «ΑΦΜ», «ΑΔΤ», «ΕΕΑ», «ΑΜΕΑ», «Αναπηρία», «Μονογονεϊκή», «Τρίτεκνη», «Πολύτεκνη», «Παλιννοστούντες», «Βορειοηπειρώτες», «ΡΟΜΑ», «Απεξαρτημένα», «Εξαρτησιογόνες Ουσίες» κλπ. Τα αποτελέσματα των αναζητήσεων αυτών, στα οποία παρατίθενται και κάποια ενδεικτικά αντιπροσωπευτικά έγγραφα, είναι τα παρακάτω:

- Στην πλειοψηφία των εγγράφων συναντώνται στοιχεία, όπως το ονοματεπώνυμο, το πατρώνυμο, το μητρώνυμο και η ημερομηνία γέννησης, είτε πλήρης, είτε μόνο το έτος [220], [221].
- Σε άλλα έγγραφα, πλέον των παραπάνω, αναγράφονται στοιχεία όπως το «ΑΦΜ», το «ΑΜΚΑ» και το «ΑΔΤ» [222], [223], [226].
- Επιπλέον βρέθηκαν έγγραφα που περιέχουν ευαίσθητα προσωπικά δεδομένα, τα οποία κυρίως σχετίζονται με την υγεία, την κοινωνική πρόνοια και την φυλετική ή

εθνική προέλευση, καθώς και λοιπές πληροφορίες όπως ο αριθμός τέκνων, η διεύθυνση κατοικίας κλπ [224], [225], [227], [228], [229].

Προφανώς, όλα αυτά τα στοιχεία μπορεί κάλλιστα να συνδυαστούν με τις λοιπές διαδικτυακές υπηρεσίες που περιγράφηκαν στο παρόν κεφάλαιο προκειμένου, αφενός να εξαχθούν και περαιτέρω στοιχεία για τους εκάστοτε πολίτες και αφετέρου να διασταυρωθεί η ακρίβειά τους. Μάλιστα, το πιο πιθανό σενάριο είναι τα στοιχεία, που θα ανασυρθούν από το πρόγραμμα «Διαύγεια», να αποτελέσουν το εφαλτήριο για την διεξαγωγή περαιτέρω έρευνας. Ωστόσο, θα πρέπει να επισημανθεί ότι κατά την διάρκεια της έρευνας βρέθηκαν και ορισμένα «Ανωνυμοποιημένα» έγγραφα [230], αν και σε κάποια από αυτά υπήρχαν λάθη, ή αυτή είχε υλοποιηθεί μερικώς [231], [232], όπως και «Ψευδωνυμοποιημένα» έγγραφα [233], γεγονός που δείχνει την προσπάθεια ορισμένων φορέων του Δημοσίου στο να προστατεύσουν τα προσωπικά δεδομένα των πολιτών, στα πλαίσια της «Ανοιχτής Διακυβέρνησης» στον Ελληνικό Δημόσιο Τομέα.



ΕΙΚΟΝΑ 19: ΥΠΕΡΔΙ@ΥΓΕΙΑ

Τέλος εξετάζεται το «Ευρετήριο Τηλεφωνικού Καταλόγου» του ΟΤΕ, στο οποίο υποστηρίζονται δυο τρόποι αναζήτησης. Στον πρώτο τρόπο η αναζήτηση γίνεται με βάση το «Ονοματεπώνυμο» και ενδεχομένως την «Περιοχή», ενώ στον δεύτερο τρόπο υποστηρίζεται η αντίστροφη αναζήτηση, δηλαδή η εύρεση σε ποιόν ανήκει κάποιος αριθμός τηλεφώνου. Ωστόσο, αυτό που παρατηρείται είναι ότι και στις δυο περιπτώσεις εξάγεται περισσότερη πληροφορία από την ήδη γνωστή. Έτσι, ενώ παρέχονται κατ' ελάχιστον, είτε το ονοματεπώνυμο, είτε ο αριθμός τηλεφώνου, στο

τελικό αποτέλεσμα επιστρέφονται η αντιστοίχιση του τηλεφωνικού αριθμού με το πλήρες ονοματεπώνυμο του νόμιμου κατόχου του, το αρχικό γράμμα του πατρώνυμου, η πλήρης διεύθυνση διαμονής του και η τοποθεσία στον χάρτη. Επίσης, αξίζει να σημειωθεί ότι για τον πρώτο τρόπο είναι δυνατό να εκτελεστεί αναζήτηση, είτε εισάγοντας μόνο το όνομα ή το επώνυμο, είτε δηλώνοντας μόνο την τοποθεσία, όπως για παράδειγμα την πόλη, ή την οδό και την πόλη, και ένα γράμμα της αλφαβήτου στο πεδίο όνομα. Πάντα το αποτέλεσμα αυτών των αναζητήσεων είναι η επιστροφή ενός συνόλου εγγραφών, που ικανοποιούν τα κριτήρια αναζήτησης. Προφανώς αυτά μπορούν να χρησιμοποιηθούν, είτε για να επιβεβαιωθούν τα στοιχεία κάποιου πολίτη, είτε για να χρησιμοποιηθούν ως είσοδος στις υπόλοιπες διαδικτυακές υπηρεσίες για την εξαγωγή περισσότερων στοιχείων, που είναι και το πιο σύνηθες.

4.4 ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΚΑΙ ΠΡΟΤΕΙΝΟΜΕΝΕΣ ΛΥΣΕΙΣ

Συνεπώς, κατόπιν των παραπάνω, διαπιστώνεται ότι εν τέλει είναι δυνατό να εξαχθούν τα προσωπικά δεδομένα, τα «Αναγνωριστικά Στοιχεία Ταυτότητας» (Personally Identifiable Information-PII) και οι «Μοναδικοί Αριθμοί Ταυτοποίησης» (Unique Identification Number–UIN/UID) για κάποιο Έλληνα πολίτη, είτε απευθείας από τις ίδιες τις πηγές δεδομένων, είτε κατόπιν του συνδυασμού των δεδομένων των πηγών αυτών μεταξύ τους. Έτσι, η εφαρμογή της «Ανοιχτής Διακυβέρνησης» (Open Government) στην Ελληνική Δημόσια Διοίκηση πράγματι ενέχει κινδύνους για την προστασία της «Ιδιωτικότητας» των Ελλήνων πολιτών.

Συγκεκριμένα, οι ισχυρισμοί της προηγούμενης ενότητας επιβεβαιώθηκαν και είναι δυνατό, μέσω της χρήσης των διαδικτυακών υπηρεσιών των πηγών δεδομένων του παρόντος κεφαλαίου, να επιτευχθούν τα κάτωθι:

✓ Να επιβεβαιωθούν οι πληροφορίες που έχουν συλλεχθεί για κάποιο Έλληνα πολίτη από κάποια άλλη πηγή.

✓ Να εκτελεστούν αλλεπάλληλες δοκιμές διαφόρων τιμών στα αντίστοιχα πεδία της κάθε διαδικτυακής υπηρεσίας των ανωτέρω πηγών (Brute Force), οι οποίες χρησιμοποιούνται ως πειράματα, προκειμένου να βρεθούν οι πραγματικές ταυτότητες των πολιτών, καθώς και τα αντίστοιχα προσωπικά τους δεδομένα.

✓ Να αντληθεί επιπλέον πληροφορία από την ήδη γνωστή, μέσα από τα αποτελέσματα που επιστρέφουν αυτές οι διαδικτυακές υπηρεσίες.

✓ Να εξαχθούν επιπλέον πληροφορίες μέσα από τον συνδυασμό των αποτελεσμάτων των πηγών αυτών μεταξύ τους.

Επιπλέον, παρατηρήθηκε ότι η πρόσβαση στις διαδικτυακές υπηρεσίες «Εύρεση του ΑΜΚΑ», «Δημοτολογικά Στοιχεία Εκλογικού Σώματος» και «Στοιχεία Εκλογικού Σώματος Ελλήνων Εκλογέων» βασίζεται σε ασθενείς μηχανισμούς ελέγχου πρόσβασης, στους οποίους απαιτείται η εισαγωγή των προσωπικών στοιχείων του εκάστοτε πολίτη. Ενώ οι αντίστοιχοι μηχανισμοί δεν υφίστανται στις υπόλοιπες διαδικτυακές υπηρεσίες που εξετάστηκαν. Μάλιστα, η ασφάλεια των μηχανισμών αυτών βασίζεται στην παραδοχή ότι μόνο ο κάτοχος των εν λόγω πληροφοριών μπορεί να τις γνωρίζει. Ωστόσο, η παραδοχή αυτή είναι εξαιρετικά ασθενής, καθώς στο πλείστο των περιπτώσεων οι πληροφορίες αυτές μπορεί, είτε να αναζητηθούν στις υπόλοιπες διαδικτυακές υπηρεσίες, είτε να βρεθούν ύστερα από πολλαπλές δοκιμές (Brute Force). Επομένως, από τα παραπάνω εξάγεται το συμπέρασμα ότι ένας κακόβουλος χρήστης, ακόμα και αν δεν έχει επαρκείς πόρους στην διάθεσή του, εξακολουθεί να θεωρείται σοβαρή απειλή για την προστασία των προσωπικών δεδομένων των πολιτών, στα πλαίσια της «Ανοιχτής Διακυβέρνησης» στην Ελληνική Δημόσια Διοίκηση.

Επομένως, το γεγονός, ότι τα προσωπικά δεδομένα των Ελλήνων πολιτών βρίσκονται σε κίνδυνο, επιβάλλει την υιοθέτηση μέτρων για την προστασία τους. Γενικά, θεωρείται δύσκολο να συμβιβαστούν οι έννοιες της «Διαφάνειας» μέσω της «Ανοιχτής Διακυβέρνησης» και της «Προστασίας των Προσωπικών Δεδομένων». Εντούτοις, προκειμένου να περιοριστεί η έκταση της διαρροής των προσωπικών δεδομένων στο περιβάλλον αυτό, απαιτείται η εφαρμογή κατάλληλων μέτρων, που θα συμβάλλουν αποτελεσματικά στην επίτευξη του σκοπού αυτού. Κάποιες προτεινόμενες λύσεις προς αυτή την κατεύθυνση είναι οι παρακάτω:

➢ Να περιοριστεί η δυνατότητα εκτέλεσης αλλεπάλληλων δοκιμών στις διαδικτυακές υπηρεσίες (Brute Force), με τον καθορισμό ενός μέγιστου αριθμού αιτημάτων ανά ημέρα που θα μπορεί να υποβάλει μια συγκεκριμένη «IP» διεύθυνση στην εκάστοτε διαδικτυακή υπηρεσία. Αυτό θα περιορίσει την εκτέλεση επερωτήσεων στην διαδικτυακή αυτή υπηρεσία, με αποτέλεσμα να καταστεί πιο δύσκολη από άποψη χρόνου η εκτέλεση των αλλεπάλληλων δοκιμών. Μάλιστα, σε περίπτωση που κάποια συγκεκριμένη «IP» διεύθυνση υπερβαίνει συχνά αυτό το όριο, θα μπορούσε να σημανθεί, είτε ως κακόβουλη και να μπλοκαριστεί, είτε να απενεργοποιηθεί για

κάποιο χρονικό διάστημα η δυνατότητα αποστολής αιτημάτων από αυτή. Συμπληρωματικά, μπορούν να χρησιμοποιηθούν και οι τεχνικές «CAPTCHA» [234], έτσι ώστε να αποφευχθεί να πραγματοποιούνται οι δοκιμές αυτές μέσω προγραμμάτων.

- Να υποστούν τα δημόσια έγγραφα, στα οποία αναγράφονται προσωπικά δεδομένα, «Ανωνυμοποίηση» [183] ή «Ψευδωνυμοποίηση» [235] ανάλογα με την εκάστοτε περίπτωση, πριν αυτά αναρτηθούν στην διαδικτυακή υπηρεσία «Διαύγεια». Παράδειγμα «Ανωνυμοποιημένου» εγγράφου είναι το [230], και αντίστοιχο παράδειγμα «Ψευδωνυμοποιημένου» εγγράφου είναι το [233]. Ωστόσο, μόνο η εφαρμογή των τεχνικών αυτών δεν εγγυάται την πλήρη προστασία των προσωπικών δεδομένων, καθώς θα πρέπει να μετριάζεται κάθε φορά το αποτέλεσμά τους σε συνάρτηση με την απαίτηση για «Διαφάνεια».
- Να επανασχεδιαστούν οι διαδικτυακές υπηρεσίες που μελετήθηκαν στο παρόν κεφάλαιο, έτσι ώστε να διορθωθούν τα προβλήματα που ανιχνεύτηκαν στην λειτουργία τους, να υιοθετηθούν ισχυρότεροι μηχανισμοί ελέγχου πρόσβασης και να επανακαθοριστούν οι πληροφορίες που αυτές θα επιστρέφουν κάθε φορά, προκειμένου να ελαχιστοποιηθεί το πρόβλημα της διαρροής προσωπικών δεδομένων από την υφιστάμενη λειτουργία τους.
- Να επιτρέπεται η χρήση όλων των διαδικτυακών υπηρεσιών, που παρέχονται από τους διάφορους φορείς του ευρύτερου Δημοσίου Τομέα, μόνο αν έχει προηγουμένως αυθεντικοποιηθεί ο εκάστοτε πολίτης με την παροχή των έγκυρων διαπιστευτηρίων του. Αυτό θα βοηθήσει στην απόδοση ευθυνών σε περίπτωση κακής χρήσης ή κατάχρησης των υπηρεσιών αυτών.
- Να διεξάγονται σεμινάρια τακτικά πάνω σε θέματα που άπτονται, τόσο των διαδικασιών που θα πρέπει να ακολουθούνται για να προστατεύονται τα προσωπικά δεδομένα κατά την ανάρτηση των δημοσίων εγγράφων στο «Διαύγεια», όσο και για τις κυρώσεις που θα επιβάλλονται σε περίπτωση πλημμελούς εκτέλεσης των καθηκόντων αυτών, έτσι ώστε όλοι οι δημόσιοι λειτουργοί, που εμπλέκονται στην διαδικασία αυτή, να είναι ενημερωμένοι. Παράλληλα, και οι τεχνικοί υπάλληλοι, οι οποίοι είναι υπεύθυνοι για την ανάπτυξη και την λειτουργία αυτών των διαδικτυακών υπηρεσιών, θα πρέπει να παρακολουθούν σεμινάρια τακτικά, που να σχετίζονται με την ασφάλεια και την προστασία της «Ιδιωτικότητας», έτσι ώστε να είναι ενημερωμένοι με τις τρέχουσες εξελίξεις στον χώρο αυτό.

4.5 ΣΥΝΟΨΗ ΚΕΦΑΛΑΙΟΥ

Η εφαρμογή της «Ανοιχτής Διακυβέρνησης» (Open Government) στον Ελληνικό Δημόσιο Τομέα είναι γεγονός. Ωστόσο, η πιθανότητα ύπαρξης ευπαθειών στα διάφορα πληροφοριακά συστήματα, που την υποστηρίζουν, εγείρει ζητήματα που σχετίζονται με την προστασία των προσωπικών δεδομένων των πολιτών στο περιβάλλον αυτό. Έτσι στο παρόν κεφάλαιο εξετάστηκε, κατά πόσο είναι εφικτό να προκληθεί αποκάλυψη «Προσωπικών Δεδομένων», «Μοναδικών Αριθμών Ταυτοποίησης» (Unique Identification Numbers-UIN) και «Αναγνωριστικών Στοιχείων Ταυτότητας» (Personally Identifiable Information-PII) σε μη εξουσιοδοτημένους πολίτες μέσα από την χρήση των διαφόρων διαδικτυακών υπηρεσιών της Δημόσιας Διοίκησης, οι οποίες αναπτύχθηκαν στα πλαίσια της «Ανοιχτής Διακυβέρνησης». Μάλιστα, οι πηγές δεδομένων που διερευνήθηκαν είναι το «Υπουργείο Οικονομικών» (ΑΦΜ), το «Υπουργείο Προστασίας του Πολίτη» (ΑΔΤ), το «Υπουργείο Εργασίας και Κοινωνικής Ασφάλισης» (ΑΜΚΑ), το «Υπουργείο Εσωτερικών» (Εκλογές, Δημοτολόγιο), το «Υπουργείο Διοικητικής Ανασυγκρότησης» (Διαύγεια) και η «Υπηρεσία Τηλεφωνικού Καταλόγου του ΟΤΕ» (Τηλέφωνο). Από την έρευνα που διεξήχθη, αποκαλύφθηκε ότι εν τέλει είναι δυνατό να διαρρεύσουν τα προσωπικά δεδομένα των πολιτών. Ενδεχομένως, αυτό μπορεί να οφείλεται, αφενός στην έλλειψη καθοδήγησης, συντονισμού και συνεργασίας κατά τις διάφορες φάσεις της ανάπτυξης και διασύνδεσης των διαφόρων αυτών πληροφοριακών συστημάτων και διαδικτυακών υπηρεσιών του δημοσίου τομέα μεταξύ τους, και αφετέρου στην μη τήρηση των κανόνων για την προστασία των προσωπικών δεδομένων των Ελλήνων πολιτών. Απόρροια αυτού είναι από την μια πλευρά να περιλαμβάνονται ποικίλα προσωπικά δεδομένα στα διάφορα δημόσια έγγραφα που αναρτώνται στο διαδίκτυο, και από την άλλη πλευρά οι διάφορες διαδικτυακές υπηρεσίες να έχουν αδυναμίες. Συνεπώς, θα πρέπει να ληφθούν μέτρα για την επίλυση του προβλήματος, ενώ ταυτόχρονα θα ικανοποιείται η απαίτηση για «Διαφάνεια». Διάφορες λύσεις προς αυτή την κατεύθυνση είναι η χρήση μηχανισμών για την απαγόρευση εκτέλεσης πολλαπλών δοκιμών (Brute Force), η εφαρμογή των τεχνικών της «Ανωνυμοποίησης» και «Ψευδωνυμοποίησης» στα δεδομένα, ο επανασχεδιασμός των διαδικτυακών υπηρεσιών, έτσι ώστε να διορθωθούν τα προβλήματα, η χρήση μηχανισμών ελέγχου πρόσβασης μέσω των οποίων θα είναι επιτρεπτή η πρόσβαση στις υπηρεσίες αυτές και τέλος η ενημέρωση και η εκπαίδευση των υπαλλήλων.

Ωστόσο, καθώς αυτό είναι ένα πολύπλοκο πρόβλημα με πολλούς εμπλεκόμενους, η επίλυσή του χρήζει περαιτέρω έρευνας και συζήτησης.

5. ΚΕΦΑΛΑΙΟ 5ο: ΕΠΙΛΟΓΟΣ

5.1 ΣΥΝΟΨΗ - ΣΥΜΠΕΡΑΣΜΑΤΑ

Κατά την διάρκεια των τελευταίων ετών, η έλευση των «Big Data» έχει επιφέρει ραγδαίες εξελίξεις σε ποικίλους τομείς της σύγχρονης ζωής, όπως το εμπόριο, την επιστήμη, την υγεία, την έρευνα, τον επιχειρηματικό τομέα και την εσωτερική ασφάλεια. Μάλιστα, καινοτόμες τεχνολογίες έχουν εισαγάγει νέους τρόπους για την αποτελεσματική διαχείριση του τεράστιου όγκου δεδομένων που παράγεται, σχεδόν σε πραγματικό χρόνο, από μια πληθώρα πηγών όπως τους διάφορους αισθητήρες του «Internet of Things», τα μέσα κοινωνικής δικτύωσης, τις κινητές ή/και σταθερές συσκευές, τους δορυφόρους, τις συσκευές εντοπισμού θέσης κλπ. Σκοπός όλων αυτών των τεχνολογιών είναι να εξαχθεί πολύτιμη πληροφορία μέσα από αυτό το «πληροφοριακό χάος» και να εντοπιστούν νέοι συσχετισμοί ή να συλληφθούν νέες και απροσδόκητες χρήσεις των δεδομένων αυτών. Ωστόσο, αν και τα αποτελέσματα από την εφαρμογή των «Big Data» και «Big Data Analytics» στους διάφορους τομείς της σύγχρονης κοινωνικής ζωής είναι θετικά, εντούτοις δεν θα πρέπει να παραβλεφθούν, αφενός τα σοβαρά ζητήματα που σχετίζονται με την ασφάλεια των συστημάτων αυτών και αφετέρου οι ανησυχίες που αναδύονται για την προστασία, τόσο της ιδιωτικής ζωής, όσο και των προσωπικών δεδομένων των ατόμων.

Με γνώμονα τα παραπάνω, στην παρούσα μεταπτυχιακή διπλωματική εργασία διερευνήθηκαν, τόσο τα ζητήματα που άπτονται της ασφάλειας στα συστήματα «Big Data», όσο και τα θέματα που σχετίζονται με την προστασία της «Ιδιωτικότητας» στο περιβάλλον αυτό. Αρχικά παρατέθηκαν κάποιοι ορισμοί της έννοιας των «Big Data» και παρουσιάστηκαν, τόσο η διαδικασία που ακολουθείται για την επεξεργασία των δεδομένων στο περιβάλλον αυτό, όσο και η αρχιτεκτονική τους. Έπειτα εξετάστηκαν τα ζητήματα ασφαλείας, τόσο σε επίπεδο υποδομής, όσο και σε επίπεδο διαδικασίας, ενώ επιπλέον μελετήθηκαν και κάποια ειδικά ζητήματα που εγείρονται στο ιδιαίτερο αυτό περιβάλλον. Στην συνέχεια, αφού ορίστηκε η έννοια της «Ιδιωτικότητας» και παρατέθηκε η ευρωπαϊκή και ελληνική νομοθεσία που την προστατεύει, διερευνήθηκαν τα θέματα που σχετίζονται με την προστασία της στο ιδιαίτερο περιβάλλον των «Big Data». Μάλιστα, στο επόμενο κεφάλαιο εξετάστηκε κατά πόσο

είναι εφικτό να διαρρεύσουν τα προσωπικά δεδομένα των Ελλήνων πολιτών από την εφαρμογή της «Ανοιχτής Διακυβέρνησης» στην Ελληνική Δημόσια Διοίκηση.

Από την έρευνα που διεξήχθη, διαπιστώθηκε ότι, όσον αφορά τα ζητήματα ασφάλειας, αλλά και τα θέματα προστασίας της «Ιδιωτικότητας», δημιουργούνται πολλά επιμέρους προβλήματα, τα οποία στο σύνολό τους χρήζουν περαιτέρω διερεύνησης και επίλυσης. Συγκεκριμένα, ορισμένες από τις προτεινόμενες λύσεις της επιστημονικής κοινότητας για την αντιμετώπιση των προβλημάτων αυτών, είτε είναι αναποτελεσματικές στο περιβάλλον αυτό και χρήζουν τροποποίησης, είτε εισάγουν νέα προβλήματα, σχετικά με την μείωση της απόδοσης του συστήματος και την αύξηση της διαχειριστικής επιβάρυνσης σε αυτό, και επομένως πρέπει να διερευνηθούν περαιτέρω. Ενώ κάποιες άλλες εξ' αυτών, είτε βρίσκονται ακόμα σε ερευνητικό στάδιο, είτε αποτελούν μόνο προτάσεις. Επιπλέον, κάποια άλλα προβλήματα παραμένουν ακόμα άλυτα και ανοιχτά. Από την άλλη πλευρά, όσον αφορά την μελέτη της πιθανότητας να διαρρεύσουν προσωπικά δεδομένα μέσα από τις διάφορες διαδικτυακές υπηρεσίες της Ελληνικής Δημόσιας Διοίκησης, στα πλαίσια της «Ανοικτής Διακυβέρνησης», εξήχθη το συμπέρασμα ότι αυτό είναι εφικτό. Συνεπώς, απαιτείται πολύ προσπάθεια ακόμα για να καταστούν τα συστήματα «Big Data» ασφαλή και ταυτόχρονα να προστατεύουν και την «Ιδιωτικότητα».

5.2 ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ - ΚΑΤΕΥΘΥΝΣΕΙΣ

Κατόπιν των ανωτέρω, ένα άμεσο συμπέρασμα, που εξάγεται, είναι ότι η υλοποίηση ενός ασφαλούς συστήματος «Big Data», που ταυτόχρονα θα προστατεύει την «Ιδιωτικότητα» και μάλιστα αυτό θα έχει υλοποιηθεί εκ σχεδιασμού και όχι εκ των υστέρων, αποτελεί ένα εξαιρετικά δύσκολο πόνημα και για την επίτευξή του απαιτείται πολύ έρευνα προς την κατεύθυνση αυτή. Για να επιτευχθεί αυτό λοιπόν, επιβάλλεται όλοι οι εμπλεκόμενοι φορείς να αναλάβουν πρωτοβουλία και μέσα από την έρευνα και την συνεργασία μεταξύ τους να εκτελέσουν όλα εκείνα τα απαραίτητα βήματα, προκειμένου να ενσωματωθούν στον σχεδιασμό και την ανάπτυξη των συστημάτων αυτών οι έννοιες της «Ασφάλειας» και του «Privacy By Design».

Συγκεκριμένα, στα ζητήματα, που χρήζουν περαιτέρω διερεύνησης από την επιστημονική κοινότητα, συμπεριλαμβάνονται όλα τα επιμέρους προβλήματα που σχετίζονται, τόσο με την ασφάλεια του συστήματος «Big Data», όσο και με την

προστασία της «Ιδιωτικότητας» στο περιβάλλον αυτό. Μάλιστα, θα πρέπει έκαστη εκ των λύσεων, που έχουν προταθεί για την αντιμετώπιση των επιμέρους αυτών προβλημάτων, να μελετηθεί σε βάθος. Συνεπώς, στις μελλοντικές προεκτάσεις συγκαταλέγονται, αφενός η μελέτη και ανάπτυξη όλων εκείνων των επιμέρους λύσεων που είτε χρήζουν τροποποίησης, είτε βρίσκονται σε ερευνητικό στάδιο, είτε αποτελούν απλά προτάσεις, έτσι ώστε αυτές να καταστούν πραγματικότητα, και αφετέρου να μελετηθούν ενδελεχώς όλα εκείνα τα ζητήματα που παραμένουν άλυτα και ανοιχτά ερευνητικά πεδία.

Τέλος, όσον αφορά το πρόβλημα της διαρροής των προσωπικών δεδομένων από την εφαρμογή της «Ανοιχτής Διακυβέρνησης» στην Ελληνική Δημόσια Διοίκηση, θα πρέπει οι αρμόδιοι Φορείς να εξετάσουν σε βάθος το πρόβλημα που προκύπτει και να μελετήσουν τις προτεινόμενες λύσεις, ή και άλλες, προκειμένου αυτό να αντιμετωπιστεί αποτελεσματικά.

Συνοψίζοντας, τα «Big Data» ήρθαν για να μείνουν. Οπότε η επίλυση όλων των ζητημάτων, που παρουσιάστηκαν στην παρούσα μεταπτυχιακή διπλωματική εργασία, είναι μονόδρομος, καθώς δεν χωράνε παραλείψεις και εκπτώσεις ούτε στα ζητήματα της ασφάλειας των συστημάτων αυτών, ούτε και στην προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων. Συνεπώς, το να καταστούν οι έννοιες της «Ασφάλειας» (Security) και της «Προστασίας της Ιδιωτικότητας» (Privacy Preserving) βασικές αρχές κατά την σχεδίαση και ανάπτυξη των συστημάτων «Big Data», θα πρέπει να είναι ο υπέρτατος στόχος όλων των εμπλεκομένων.

6. ΒΙΒΛΙΟΓΡΑΦΙΑ - ΑΝΑΦΟΡΕΣ

ΔΗΜΟΣΙΕΥΣΕΙΣ

1. Cloud Security Alliance. Expanded Top Ten Big Data Security and Privacy Challenges. Big Data Working Group, April 2013.

2. ENISA. Big Data Threat Landscape and Good Practice Guide. January 2016.

3. ENISA. Privacy By Design in Big Data. An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics. December 2015.

4. Moura, Serrão. Security and Privacy Issues of Big Data.

5. Jaseena, David. Issues, Challenges and Solutions: Big Data Mining. Department of Computer Applications, M.E.S College, Marampally, Aluva, Cochin, India. 2014.

6. International Journal of Advanced Research in Computer Science and Software Engineering. Big Data: Concept, Challenges and Management Tools. Ranjana Bahri. February 2015.

7. Christos Doulkeridis. Big Data: More than just size!. Lectures in Piraeus University, Digital Systems, 2015-2016.

8. Laney, Douglas. 3D Data Management: Controlling Data Volume, Velocity and Variety. Gartner. Issued: 6 February 2001.

9. Manyika. Big Data: The Next Frontier for Innovation, Competition and Productivity. McKinsey Global Institute, pp. 1-137, 2011.

10. ENISA. Big Data Security: Good Practices and Recommendations on the Security of Big Data Systems. December 2015.

11. EMC/IDC. The Digital Universe. Study, 2014.

12. Nasser, Tariq. Big Data Challenges. Journal of Computer Engineering & Information Technology. September 2015.

13. Snijders, Matzat, Reips. Big Data: Big Gaps of Knowledges in the Field of Internet. International Journal of Internet Science. 2012.

14. Lohr. The Origins of “Big Data”. An Etymological Detective Story. New York Times, 1 February 2013.

15. Hurwitz, Nugent, Dr. Halper, Kaufman. Big Data for Dummies. Wiley Brand, 2013.

16. SANS Institute. Enabling Big Data by Removing Security and Compliance Barriers. April 2015.

17. US Leading Researchers Community. Big Data White Paper. Challenges and Opportunities with Big Data. November 2011 to February 2012.

18. Roy, Setty, Kilzer, Shmatikov, Witchel. Airavat: Security and Privacy for MapReduce. USENIX Conference on Networked Systems Design and Implementation, 2010.

19. Okman, Gal-Oz, Gonen, Gudes, Abramov. Security Issues in NoSQL Databases. TrustCom IEEE Conference, 2011.

20. Penchikala. Virtual Panel: Security Considerations in Accessing NoSQL Databases. November 2011.

21. Sullivan. NoSQL, But Even Less Security. 2011.

22. Chickowski. Does NoSQL Mean No Security? January 2012.

23. Kot. Tracking Personal Data Use: Provenance and Trust. Seventh Biennial Conference on Innovative Data Systems Research, 2015.

24. Shea. Anything but Meta: Planning for Data Provenance. 2013.

25. Park, Nguyen, Sandhu. A Provenance-based Access Control Model. Tenth Annual International Conference on Privacy, Security and Trust, Paris, 2012.

26. Park, Nguyen, Sandhu. A Provenance-based Access Control Model for Dynamic Separation of Duties. Eleventh Annual International Conference on Privacy, Security and Trust, Tarragona, 2013.

27. Lu, Lin, Liang, Shen. Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing. Fifth ACM Symposium on Information, Computer and Communications Security, New York, USA, 2010.

28. Goyal, Pandey, Sahai, Waters. Attribute-based Encryption for Fine-Grained Access Control of Encrypted Data. ACM Conference on Computer and Communications Security, 2006.

29. Muniswamy-Reddy, Holland, Braun, Seltzer. Provenance-aware Storage Systems. USENIX Annual Technical Conference, General Track, 2006.

30. Feng, Chen, Liu. Bridging the Missing Link of Cloud Data Storage Security in AWS. Seventh IEEE Consumer Communications and Networking Conference, Las Vegas, Nevada, USA, January 2010.

31. Popa, Lorch, Molnar, Wang, Zhuang. Enabling Security in Cloud Storage SLA's with CloudProof. Microsoft TechReport, May, 2010.

32. Onieva, Lopez, Zhou. Secure Multi-Party Non-Repudiation Protocols and Applications. Advances in Information Security Series, Springer, 2009.

33. Anagnostopoulos, Goodrich, Tamassia. Persistent Authenticated Dictionaries and their Applications. Fourth International Conference on Information Security, October, 2001.

34. Majuntke, Dobre, Serafini, Suri. Abortable Fork Linearizable Storage. OPODIS, 2009.

35. Boneh, Gentry, Waters. Collusion Resistant Broadcast Encryption with Short Ciphertexts and Private Keys. Lecture Notes in Computer Science, 2005.

36. Kallahalla, Riedel, Swaminathan, Wang, Fu. Plutus: Scalable Secure File Sharing on Untrusted Storage. USENIX Conference on File and Storage Technologies, 2003.

37. Muniswamy-Reddy, Macko, Seltzer. Provenance for the Cloud. Eighth USENIX Conference on File and Storage Technologies, February 2010.

38. Agreiter, Hafner, Breu. A Fair Non-Repudiation Service in a Web Service Peer-to-Peer Environment. Computer Standards & Interfaces, August 2008.

39. Juels, Pors. Proofs of Retrievability for Large Files. ACM CCS, 2007.

40. Bagga, Molva. Collusion-Free Policy-Based Encryption. ISC, Springer, 2006.

41. Feng, Chen, Summerville. A Fair Multi-Party Non-Repudiation Scheme for Storage Clouds. International CTS, Philadelphia, USA, May 2011.

42. Feng, Chen, Ku, Liu. Analysis of Integrity Vulnerabilities and a Non-Repudiation Protocol for Cloud Data Storage Platforms. International Workshop on Security in Cloud Computing, San Diego, California, USA, 2010.

43. Ateniese, Di Pietro, Mancini, Tsudik. Scalable and Efficient Provable Data Possession. SecureComm, New York, USA, 2008.

44. Wang, Wang, Ren, Lou, Li. Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing. Parallel and Distributed Systems, IEEE Transactions, May 2011.

45. Wang, Wang, Ren, Lou. Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing. INFOCOM, March 2010.

46. Gentry. Computing Arbitrary Functions of Encrypted Data. Communications of the ACM, 2010.

47. Lambert. Measures of Disclosure Risk and Harm. Journal of Official Statistics, 1993.

48. Gilbert, Jung, Lee, Qin, Sharkey, Sheth, Cox. YouProve: Authenticity and Fidelity in Mobile Sensing. ACM SenSys, Seattle, November 2011.

49. Ptacek, Newsham. Insertion, Evasion and Denial of Service: Eluding Network Intrusion Detection. Tech Report, 1998.

50. Barreno, Nelson, Sears, Joseph, Tygar. Can Machine learning be Secure? Proc. Of the 2006 ACM Symposium on Information, Computer and Communications Security, March 2006.

51. Cloud Security Alliance. Big Data. Security and Privacy Handbook: 100 Best Practices in Big Data Security and Privacy. Presented by Big Data working Group Guidance, 2016.

52. Cloud Security Alliance. Top Ten Big Data Security and Privacy Challenges. Big Data Working Group, November 2012.

53. HP, Slashdotmedia, Sourceforge. 10 Ways to Build a Better Big Data Security Strategy. IT Manager's Journal, January 2014.

54. Cloud Security Alliance. Big Data Analytics for Security Intelligence. Big Data Working Group, September 2013.

55. Piper, CISSP. Big Data Security for Dummies, Blue Coat Second Edition. Wiley Brand, 2015.

56. Goldberg. Big Data for Security: How Can I Put Big Data to Work for Me?. Lecture in RSA Conference, Europe, 2013.

57. Doucer. The Sybil Attack. International Workshop on Peer-to-Peer Systems. Springer Berlin Heidelberg, 2002.

58. Balachandran, Sanyal. A Review of Techniques to Mitigate Sybil Attacks. ArXiv Preprint ArXiv: 1207.2617, 2012.

59. Felix T. Wu. Big Data Threats.

60. Rosenberg. The Social Impact of Computers. New York: Academic Press, 1992.

61. Solove. Understanding Privacy. Cambridge, Mass.: Harvard University Press, 2008.

62. Warren, Brandeis. The Right to Privacy. In: Harvard Law Review, 1890.

63. Westin. Privacy and Freedom. Athenaeum, New York, 1967.

64. Pfitzmann, Hansen. A Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity and Identity Management. August 2010.

65. Fischer-Hübner. IT-Security and Privacy-Design and Use of Privacy-Enhancing Security Mechanisms. Springer Scientific Publishers, Lecture Notes of Computer Science, May 2001.

66. Cannon. Privacy: What Developers and IT Professionals Should Know. Addison-Wesley Professional, 2004.

67. Kalloniatis. Privacy Enhancing Technologies. The PRIS Method. Lectures Aegean University, Lectures Piraeus University, 2016.

68. Kalloniatis, Kavakli, Gritzalis. Addressing Privacy Requirements in System Design: the PriS Method. Springer, 2008.

69. Cavoukian. Privacy by Design. Take the Challenge. Information and Privacy Commissioner of Ontario, Canada, 2009.

70. Cavoukian. Privacy by Design: the 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada, 2009.

71. Spiekermann. The Challenges of Privacy by Design. Communications of the ACM, 2012.

72. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

73. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance).

74. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

75. Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on

cooperation between national authorities responsible for the enforcement of consumer protection laws (Text with EEA relevance).

76. Article 29 Data Protection Working Party, "Opinion 4/2007 on the concept of personal data," 2007.

77. Article 29 Data Protection Working Party, "Opinion 05/2014 on Anonymization Techniques," 2014.

78. Νόμος 2472/1997 (Περί Προστασίας του Ατόμου από την Επεξεργασία Δεδομένων Προσωπικού Χαρακτήρα με Ενσωματωμένες τις Τροποποιήσεις), Νοέμβριος 2011.

79. Νόμος 3471/2006 (Περί Προστασίας των Δεδομένων Προσωπικού Χαρακτήρα και της Ιδιωτικής Ζωής στον Τομέα των Ηλεκτρονικών Επικοινωνιών και Τροποποίηση του Νόμου 2472/1997), Μάιος 2012.

80. Oetzel, Spiekermann. Privacy-by-Design Through Systematic Privacy Impact Assessment. A Design Science Approach. European Conference on Information Systems, 2012.

81. Oetzel, Spiekermann. A Systematic Methodology for Privacy Impact Assessments: a Design Science Approach. European Journal of Information Systems, 2013.

82. Bennett, Bayley. Privacy Impact Assessments: International Study of their Application and Effects. Loughborough University, UK, 2007.

83. Solove. A Taxonomy of Privacy. University of Pennsylvania Law Review, 2006.

84. Dalenius. Towards a Methodology for Statistical Disclosure Control. Statistik Tidskrift, 1977.

85. Vaidya, Clifton, Zhu. Privacy Preserving Data Mining. 2006.

86. Bambauer. Tragedy of the Data Commons. Harvard Journal of Law and Technology, 2011.

87. Polonetsky, Wolf, Brennan. Comments of the Future of Privacy Forum. Washington, 2014.

88. Samarati, Sweeney. Protecting Privacy when Disclosing Information: k-Anonymity and its Enforcement Through Generalization and Suppression. 1998.

89. Truta, Vinay. Privacy Protection: p-Sensitive k-Anonymity Property. In 22nd International Conference on Data Engineering Workshops, Atlanta, 2006.

90. Machanavajjhala, Kifer, Abowd, Gehrke, Vilhuber. L-Diversity: Privacy Beyond k-Anonymity. ACM Transactions on Knowledge Discovery from Data, 2007.

91. Ninghui, Tiancheng, Venkatasubramanian. T-Clossness: Privacy Beyond k-Anonymity and l-Divesity. In IEEE 23rd International Conference on Data Engineering, Istanbul, 2007.

92. Ninghui, Tiancheng, Venkatasubramanian. Clossness: A New Privacy Measurefor Data Publishing. IEEE Transactions on Knowledge and Data Engineering, 2010.

93. Dwork. Differential Privacy. In 33rd International Colloquium, Venice, 2006.

94. Gehrke, Hay, Lui, Pass. Crowd-Blending Privacy. In 32nd Annual Cryptology Conference, Santa Barbara, 2012.

95. Machanavajjhala, Kifer. Designing Statistical Privacy for your Data. Communications of the ACM, 2015.

96. Soria-Comas, Domingo-Ferrer. Big Data Privacy: Challenges to Privacy Principles and Models. Data Science and Engineering, 2015.

97. National Science Foundation. Theoretical Computer Science Provides Answers to Data Privacy Problem. 2015.

98. Clifton, Tassa. On Syntactic Anonymity and Differential Privacy. Transactions on Data Privacy, 2013.

99. Canetti. Universally Composable Security: A New Paradigm for Cryptographic Protocols. Foundations of Computer Science, 2001.

100. Boneh, Franklin. Identity-Based Encryption from the Weil Pairing. SIAM Journal on Computing, 2003.

101. Sahai, Waters. Fuzzy Identity-Based Encryption. Advances in Cryptology – EUROCRYPT 2005, Springer, May 2005.

102. Boneh, Waters. Conjunctive, Subset and Range Queries on Encrypted Data. Theory of Cryptography, 2007.

103. Gentry. Fully Homomorphic Encryption Using Ideal Lattices. Proceedings of the 41st Annual ACM Symposium on Symposium on Theory of Computing, ACM Press, 2009.

104. Boneh, Goh, Nissim. Evaluating 2-DNF Formulas on Ciphertexts. Theory of Cryptography, Springer, February 2005.

105. Boneh, Boyen, Shacham. Short Group Signatures. Advances in Cryptology-CRYPTO 2004, Springer Berlin/Heidelberg, 2004.

106. Narayanan, Shmatikov. Myths and Fallacies of Personally Identifiable Information. Communications of the ACM, 2010.

107. Barbaro, Zeller. A Face is Exposed for AOL Searcher No. 4417749. The New York Times, 2006.

108. Hansell. AOL Removes Search Data on Vast Group of Web Users. The New York Times, 2006.

109. Castelluccia, Privatics. (Big) Data Anonymization. 2014.

110. United States Department of Justice. Privacy Technology Focus Group Report. Final Report and Recommendations, September 2006.

111. Hundepool, Domingo-Ferrer, Franconi, Giessing, Schulte-Nordholt, Spicer, Wolf. Statistical Disclosure Control. Wiley, 2012.

112. Cook, Holder. Mining Data Graph. Wiley-Interscience, 2007.

113. Domingo-Ferrer, Torra. Disclosure Control Methods and Information Loss for Microdata in Confidentiality, Disclosure and Data Access. North-Holland, 2001.

114. Woo, Reiter, Oganian, Karr. Global Measures of Data Utility Masked for Disclosure Limitation. The Journal of Privacy and Confidentiality, 2009.

115. Dasseni, Verykios, Elmagarmid, Bertino. Hiding Association Rules by Using Confidence and Support. In 4th International Workshop on Information Hiding, 2001.

116. Li, Liu, Tian, Shen, Mao. A Storage Solution for Massive IoT Data Based on NoSQL. In IEEE International Conference on Green Computing and Communications, Besancon, 2012.

117. Stokes, Torra. N-Confusion: a Generalization of k-Anonymity. In 2012 Joint EDBT/ICDT Workshops, 2012.

118. Nergiz, Clifton, Nergiz. Multirelational k-Anonymity. IEEE Transactions on Knowledge and Data Engineering, 2009.

119. Truta, Campan. K-Anonymization Incremental Maintenance and Optimization Techniques. In 2007 ACM Symposium on Applied Computing, 2007.

120. Muntes-Mulero, Nin. Privacy and Anonymization for Very Large Datasets. In 18th ACM Conference on Information and Knowledge Management, 2009.

121. Loftus, May, Smart, Vercauteren. On CCA-Secure Fully Homomorphic Encryption. Cryptology ePrint Archive, Report, 2010.

122. Shamir. Identity-Based Cryptosystems and Signature Schemes. Advances in Cryptology – CRYPTO84, August 1985.

123. Boneh, Sahai, Waters. Functional Encryption: Definitions and Challenges. In 8th Theory of Cryptography Conference, 2011.

124. Kamara. How to Search on Encrypted Data. 2015.

125. Naveed, Kamara, Wright. Inference Attacks on Property Preserving Encrypted Datasets. 2015.

126. Cash, Jarecki, Jutla, Krawczyk, Rosu, Steiner. Highly-Scalable Searchable Symmetric Encryption with Support for Boolean Queries. Springer, 2013.

127. Mandal, Roy. Relational Hash: Probabilistic Hash for Verifying Relations, Secure Against Forgery and More. Springer, Berlin Heidelberg, 2015.

128. Stefanov, Shi, Song. Towards Practical Oblivious RAM. 2011.

129. Rivest, Shamir, Tauman. How to Leak a Secret. Springer, December 2001.

130. OASIS. OASIS eXtensible Access Control Markup Language (XACML).

131. Chibba, Cavoukian, Williamson, Gerfuson. The Importance of ABAC: Attribute-Based Access Control to Big Data: Privacy and Context. 2015.

132. Article 29 Data Protection Working Party. Opinion 02/2013 on Apps on Smart Devices. 2013.

133. Karjoth, Schunter, Waidner. Platform for Enterprise Privacy Practices: Privacy Enabled Management of Customer Data. In 2nd International Conference on Privacy Enhancing Technologies, 2002.

134. Lopes, Novais, Coelho. Towards an Interdisciplinary Framework for Automated Negotiation. In 9th International Conference on e-Commerce and Web Technologies, 2008.

135. Jennings et al. Automated Negotiation: Prospects, Methods and Challenges. Group Decision and Negotiation, 2001.

136. Tubaro, Casilli, Sarabi. Three Approaches to Privacy: As Penetration, Regulation and Negotiation. In Against the Hypothesis of the End of Privacy, Springer, 2014.

137. Larsen, Brochot, Lewis, Eisma, Brunini. Study on Personal Data Stores Conducted at the Cambridge University Judge Business School, 2015.

138. Cavoukian, Jonas. Privacy by Design in the Age of Big Data. 2012.

139. Mun et al. Personal Data Vaults: a Locus of Control for Personal Data Streams. In 6th International Conference on Emerging Networking Experiments and Technologies, Philadelphia, 2010.

140. Haddadi et al. Personal Data: Thinking Inside the Box. 2015.

141. Caceres, Cox, Lim, Varshavsky, Shakimov. Virtual Individual Servers as Privacy Preserving Proxies for Mobile Devices. In 1st ACM Workshop on Networking, Systems and Applications for Mobile Handhelds, 2009.

142. Montjoye, Shmueli, Wang, Pentland. OpenPDS: Protecting the Privacy of Metadata Through SafeAnswers. PLoS ON, 2014.

143. Zyskind, Nathan, Pentland. Enigma: Decentralized Computation Platform with Guaranteed Privacy. 2015.

144. Bell. A Personal Digital Store. Communications of the ACM, 2001.

145. Baden, Bender, Spring, Bhattacharjee, Starin. Persona: an Online Social Network with User Defined Privacy. In ACM SIGCOMM 2009 Conference on Data Communication, 2009.

146. Hong, Landay. An Architecture for Privacy Sensitive Ubiquitous Computing. In 2nd International Conference on Mobile Systems, Applications and Services, 2004.

147. Tzermias, Prevelakis, Ioannidis. Privacy Risks from Public Data Sources. 2014.

148. Gessiou, Labrinidis, Ioannidis. A Greek (Privacy) Tragedy: The Introduction of Social Security Numbers in Greece. In Proceedings of the 8th Annual ACM Workshop on Privacy in the Electronic Society. ACM, 2009.

149. Oracle and Intel. Enterprise Security for Big Data Environments: A White Paper from Oracle and Intel. A Multi-Layered Architecture for Defense-in-Depth Protection. July 2016.

150. MIT Technology Review Custon, Oracle. Securing the Big Data Life Cycle. 2015.

151. Kaisler, Armour, Espinosa, Money. Big Data: Issues and Challenges Moving Forward. 46th Hawaii International Conference on System Sciences, 2013.

152. Hongjun, Wenning, Dengchao, Yuxing. Survey of Research on Information Security in Big Data. XXXIV CSBC, 2014.

153. Securosis. Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments. October 2012.

154. Gaddam. Securing Your Big Data Environment. US 2015.

155. Bell, Rotman, VanDenBerg. Navigating Big Data's Privacy and Security Challenges. KPMG, 2014.

156. Jitendrakumar. Security Issues in Big Data: In Context with Hadoop. International Journal of Innovative and Emerging Research in Engineering. 2015.

157. HITPC Privacy and Security Workgroup. Health Big Data Recommendations. August 2015.

WEBSITES

158. https://www.domo.com/blog/data-never-sleeps-2-0/

159. https://www.domo.com/blog/data-never-sleeps-4-0/

160. http://23.66.85.199/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf

161. https://www.youtube.com/yt/press/statistics.html (YouTube Statistics 14 Feb 2015)

162. http://www.statisticbrain.com/youtube-statistics/ (YouTube Statistics 1 Sep 2016)

163. http://www.statisticbrain.com/facebook-statistics/ (Facebook Statistics 1 Aug 2016)

164. http://www.statisticbrain.com/twitter-statistics/ (Twitter Statistics 1 Sep 2016)

165. http://www.statisticbrain.com/instagram-company-statistics/ (Instagram Statistics 1 Sep 2016)

166. http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#28ca1e5d55da

167. http://www.gartner.com/it-glossary/big-data

168. https://en.wikipedia.org/wiki/Big_data

169. https://en.wikipedia.org/wiki/Analytics

170. https://en.wikipedia.org/wiki/Data_analysis

171. https://www.techopedia.com/definition/28659/big-data-analytics

172. http://www.pasoa.org/ (Provenance Aware Service Oriented Architecture - PASOA)

173. https://tools.ietf.org/html/rfc5280

174. https://en.wikipedia.org/wiki/X.509

175. https://en.wikipedia.org/wiki/Privacy

176. https://en.wikipedia.org/wiki/Privacy_by_design

177. http://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:31995L0046&qid=1487758077031&from=EN

178. http://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:32016R0679&qid=1487758569952&from=EN

179. http://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:32002L0058&qid=1487759098354&from=EN

180. http://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:32009L0136&qid=1487759257254&from=EN

181.http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

182. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

183. https://en.wikipedia.org/wiki/Data_anonymization

184. https://en.wikipedia.org/wiki/Dynamic_data

185. https://en.wikipedia.org/wiki/Streaming_media

186. https://en.wikipedia.org/wiki/Cryptography

187. https://en.wikipedia.org/wiki/Hybrid_cryptosystem

188. https://en.wikipedia.org/wiki/Access_control

189. https://en.wikipedia.org/wiki/Database_normalization

190. https://en.wikipedia.org/wiki/Denormalization

191. https://en.wikipedia.org/wiki/Single_sign-on

192. https://en.wikipedia.org/wiki/OWASP

193. https://disconnect.me/icons

194. https://wiki.mozilla.org/Privacy_Icons

195. https://www.w3.org/XML/

196. https://www.w3.org/P3P/

197. https://www.w3.org/TR/P3P-preferences/

198. https://www.google.com/about/company/user-consent-policy.html

199. http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3167654

200. http://www.eclipse.org/higgins/

201. https://mydex.org/

202. https://lists.oasis-open.org/archives/xdi/201011/pdf00000.pdf

203. http://www.gsis.gr/gsis/info/gsis_site/

204. http://www.gsis.gr/gsis/info/gsis_site/Services/polites.html

205. http://www.gsis.gr/gsis/info/gsis_site/Services/epixeiriseis.html

206. http://ec.europa.eu/taxation_customs/vies/vatRequest.html

207. http://www.gsis.gr/gsis/info/gsis_site/PublicIssue/wnsp/wsg2g/mitrwo.html

208. https://en.wikipedia.org/wiki/Greek_identity_card

209. http://www.amka.gr/AMKAGR/

210. http://www.ypes.gr/services/eea/eeagr/eea.htm

211. http://www.ypes.gr/services/eea/dimos/eea.htm

212. https://diavgeia.gov.gr/

213. https://diavgeia.gov.gr/info

214. https://www.11888.gr/home?_sm=wp

215. http://www.emc.com/leadership/digital-universe/iview/big-data-2020.htm

216. http://www.emc.com/leadership/digital-universe/2012iview/big-data-2020.htm (IDC 2012, Big Data in 2020)

217. https://en.wikipedia.org/wiki/Personally_identifiable_information

218. https://de.wikipedia.org/wiki/Unique_Identification_Number

219. https://yperdiavgeia.gr/

220. https://diavgeia.gov.gr/doc/61%CE%93%CE%A94653%CE%A04-4%CE%92%CE%A3?inline=true

221.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/19120375

222.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/20352639

223. https://diavgeia.gov.gr/doc/%CE%A9%CE%A1%CE%A9%CE%A6%CE%9B-%CE%9D%CE%946?inline=true

224.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/20291995

225.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/21672132

226.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/18854982

227. https://diavgeia.gov.gr/doc/%CE%92%CE%9526469%CE%97%CE%9C5-851?inline=true

228.https://diavgeia.gov.gr/doc/%CE%92%CE%99%CE%9E%CE%9E%CE%A9%CE%A84-%CE%9B%CE%9C%CE%A6?inline=true

229.https://diavgeia.gov.gr/doc/%CE%A90%CE%988%CE%9F%CE%95%CE%99%CE%94-3%CE%99%CE%9D?inline=true

230.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/20124243

231.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/20135157

232.https://yperdiavgeia.gr/pdfjs/web/viewer.html?file=/decisions/downloadPdf/20274669

233. http://www.oaed.gr/nea-arthro/-/asset_publisher/ebcGfvDrjPsQ/content/prosorinos-pinakas-katataxes-anergon-gia-24-251-theseis-plerous-apascholeses-tou-programmatos-koinophelous-charaktera-se-olous-tous-demous-tes-choras?_101_INSTANCE_ebcGfvDrjPsQ_redirect=http%3A%2F%2Fwww.oaed.gr%2Fnea-arthro%3Fp_p_id%3D101_INSTANCE_ebcGfvDrjPsQ%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-3%26p_p_col_count%3D1%26p_r_p_564233524_categoryId%3D15508%26p_r_p_564233524_resetCur%3Dtrue&redirect=http%3A%2F%2Fwww.oaed.gr%2Fnea-arthro%3Fp_p_id%3D101_INSTANCE_ebcGfvDrjPsQ%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-3%26p_p_col_count%3D1%26p_r_p_564233524_categoryId%3D15508%26p_r_p_564233524_resetCur%3Dtrue

234. https://en.wikipedia.org/wiki/CAPTCHA

235. https://en.wikipedia.org/wiki/Pseudonymity

236. https://en.wikipedia.org/wiki/Mutual_authentication

237. https://en.wikipedia.org/wiki/Trusted_Platform_Module

238. https://en.wikipedia.org/wiki/Mandatory_access_control

239. https://en.wikipedia.org/wiki/Threat_model

240. https://en.wikipedia.org/wiki/Representational_state_transfer

241. https://en.wikipedia.org/wiki/CAP_theorem

242. https://en.wikipedia.org/wiki/Trusted_execution_environment

243. https://en.wikipedia.org/wiki/Log_analysis

244. https://en.wikipedia.org/wiki/Fuzzing

245. https://en.wikipedia.org/wiki/Kerberos_(protocol)

246. https://en.wikipedia.org/wiki/Metadata

247. https://en.wikipedia.org/wiki/Hash_table

248. https://en.wikipedia.org/wiki/Automated_tiered_storage

249. https://en.wikipedia.org/wiki/Digital_rights_management

250. https://en.wikipedia.org/wiki/Zero-day_(computing)

251. https://en.wikipedia.org/wiki/Security_information_and_event_management

