ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

“ΑΣΦΑΛΕΙΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ”

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«ΑΣΦΑΛΕΙΑ ΚΑΙ ΠΡΟΣΤΑΣΙΑ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΤΑ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ»

ΤΕΤΡΙΜΕΛΗ ΒΑΣΙΛΙΚΗ

ΜΤΕ 2128

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ : ΣΤΕΦΑΝΟΣ ΓΚΡΙΤΖΑΛΗΣ

ΑΘΗΝΑ 2024

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ

Τα μεγάλα δεδομένα υπόσχονται να παράγουν αναλυτικές γνώσεις που θα διευρύνουν το σώμα της ερευνητικής και κοινωνικής επιστημονικής γνώσης, θα ενισχύσουν δραματικά την ανθρώπινη αυτογνωσία και κατανόηση και θα βελτιώσουν σημαντικά τη λήψη αποφάσεων στον δημόσιο και ιδιωτικό τομέα. Έχουν ήδη οδηγήσει στην ανάπτυξη εντελώς νέων κατηγοριών προϊόντων και υπηρεσιών, πολλές από τις οποίες έχουν γίνει αποδεκτές με ενθουσιασμό τόσο από ιδρύματα όσο και από ανθρώπους.

Τα μεγάλα δεδομένα αναφέρονται ρητά σε σύνολα δεδομένων που είναι τόσο τεράστια ή περίπλοκα που το συμβατικό λογισμικό επεξεργασίας δεδομένων δεν επαρκεί για να τα χειριστεί. Κάθε μέρα, μια εταιρεία κατακλύζεται από τον τεράστιο όγκο δεδομένων, οργανωμένων και μη. Λόγω των πρόσφατων τεχνικών εξελίξεων, ο όγκος των δεδομένων που παράγονται από το Διαδίκτυο, τους ιστότοπους κοινωνικής δικτύωσης, τα δίκτυα αισθητήρων, τις εφαρμογές υγειονομικής περίθαλψης και πολλές άλλες επιχειρήσεις αυξάνεται δραματικά κάθε μέρα.

Κατά τη διαδικασία συλλογής, αποθήκευσης και χρήσης δεδομένων, αυτό το γεγονός μπορεί εύκολα να οδηγήσει σε διαρροή προσωπικών πληροφοριών και στο γεγονός ότι τα δεδομένα είναι δύσκολο να διακριθούν. Ο τρόπος διασφάλισης της ασφάλειας των μεγάλων δεδομένων και της προστασίας του απορρήτου έχει γίνει ένα από τα καυτά ζητήματα στο τρέχον στάδιο της έρευνας.

Μερικά από τα πλεονεκτήματα της χρήσης των Μεγάλων Δεδομένων, περιλαμβάνουν, την απόδοση του συστήματος χωρίς την ανάγκη διαγραφής ακυρωμένων λογαριασμών ή παλαιών αρχείων καταγραφής μετά από ένα ορισμένο χρονικό διάστημα, ιδίως επειδή αυτά μπορεί να είναι χρήσιμα για τους σκοπούς των εγκληματολογικών ερευνών αργότερα, καθώς και τη δυνατότητα εκτέλεσης περίπλοκων και προηγμένων ερωτημάτων σε μεγάλα και αδόμητα σύνολα δεδομένων, δυνατότητα λήψης αποφάσεων σε πραγματικό χρόνο, συστήματα αυτόματης άμυνας και μείωσης κινδύνου με την πρόβλεψη επιθέσεων στο μέλλον, και τέλος ταχύτερη, καλύτερη και φθηνότερη ασφάλεια σε σύγκριση με τις παραδοσιακές μεθόδους.

ΚΕΦΑΛΑΙΟ 1^{Ο} - ΕΙΣΑΓΩΓΗ

1.1 Πρόλογος

Στην ακαδημαϊκή βιβλιογραφία, υπάρχουν αρκετοί ορισμοί και απόψεις αναφορικά με τη σχέση μεταξύ δεδομένων, πληροφοριών, γνώσης και νοημοσύνης. Η πιο κοινή ιδέα είναι ότι υπάρχει μια ιεραρχία, όπου τα δεδομένα είναι το χαμηλότερο δομικό στοιχείο προς τη γνώση. Σύμφωνα με αυτή την ιδέα, τα δεδομένα ενσωματώνουν μια συλλογή απλών παρατηρήσεων και γεγονότων, τα οποία πρέπει να έχουν πριν δημιουργήσει πληροφορίες, και μόνο μετά από αυτά τα δύο στάδια, μπορεί να επιτευχθεί η γνώση (Tuomi 1999-2000).

Έτσι, τα δεδομένα από μόνα τους θεωρούνται ότι έχουν τη μικρότερη αξία, αφού που αποτελούνται από απλούς αριθμούς, χαρακτήρες κειμένου και σήματος. Στην αλυσίδα αξίας πληροφοριών, λειτουργούν ως πρώτη ύλη για τα ανώτερα, πιο εξελιγμένα επίπεδα στοιχείων πληροφοριών. Μεταφέρονται εύκολα από λειτουργικά συστήματα σε διάφορες βάσεις δεδομένων χωρίς να χάσουν το περιεχόμενό του. Στη συνέχεια, συγκεντρώνονται σε μια *αποθήκη δεδομένων* και χρησιμοποιούνται περαιτέρω ως βάση για αναφορά ή ανάλυση (Kaario και Peltola 2008).

Οι πληροφορίες θεωρούνται ένα επίπεδο υψηλότερα στην ιεραρχία. Οι Davenport και Prusak (1998) ορίζουν τις πληροφορίες ως δεδομένα στα οποία αποδίδεται σημαντικό νόημα. Ανάλογα με τον δέκτη, η πληροφορία περιέχει πάντα κάποιο είδος μηνύματος, το οποίο μπορεί να ερμηνευτεί υποκειμενικά, και έτσι, ο Pirttimäki (2007) συμπεραίνει ότι είναι η πληροφορία πιο πολύτιμη για τον δέκτη από τα δεδομένα. Η μεταφορά πληροφοριών είναι επίσης πιο δύσκολη μέσω πληροφοριακών συστημάτων από τον ένα χρήστη στον άλλο και κατά συνέπεια παρουσιάζει μια πρόκληση για την αυτοματοποίηση των διαδικασιών πληροφοριών (Kaario και Peltola 2008).

Όταν αποδίδεται σημασία στην πληροφορία, αυτή μετατρέπεται σε γνώση. Η γνώση προσδιορίζεται ως ο πιο πολύτιμος τύπος πληροφοριών στην αλυσίδα αξίας, επειδή είναι πιο κοντά στη διαδικασία λήψης αποφάσεων και χρησιμοποιείται ως βάση τόσο επιχειρησιακών όσο και στρατηγικών ενεργειών (Pirttimäki 2007).

Σε αυτό το στάδιο, τα απλά γεγονότα στα δεδομένα, έχουν αποδοθεί ένα συγκεκριμένο πλαίσιο λειτουργίας. Όπως το έθεσε ο Tuomi (1999-2000), τα γεγονότα υπάρχουν πλέον στην ανθρώπινη νοητική δομή και έχουν τη δυνατότητα να γίνουν μέρος της οργανωτικής μνήμης που μπορεί να οδηγήσει σε συνειδητές προβλέψεις μελλοντικών συνεπειών ή άλλα συμπεράσματα της παρούσας κατάστασης, καθιστώντας τη συμπεριφορά λήψης αποφάσεων έξυπνος.

Σε αντίθεση με δύο κατώτερα δομικά στοιχεία στην ιεραρχία, η γνώση μπορεί να είναι δομημένη ή μη, και σιωπηρή ή ρητή ανάλογα με το αν μπορεί εύκολα να διαδοθεί και να αναπαρασταθεί με αριθμούς και κείμενο ή αν σχετίζεται με προσωπικές εμπειρίες και τεχνογνωσία (Pirttimäki 2007).

Όταν μετράται με το περιεχόμενο πληροφοριών, η γνώση είναι ο μεγαλύτερος πόρος πληροφοριών σε έναν οργανισμό, αλλά είναι επίσης ο πιο δύσκολος να αξιοποιηθεί και να μοιραστεί μεταξύ των χρηστών πληροφοριών λόγω της υποκειμενικής φύσης της (Kaario, Peltola 2008). Είναι επίσης πιο δύσκολο να προστατευθεί και να διατηρηθεί στον οργανισμό, επειδή συχνά εξαρτάται από τους υπαλλήλους και τα κίνητρά τους για να το χρησιμοποιήσουν και να το προωθήσουν.

Ο όρος λοιπόν των Μεγάλων Δεδομένων, ως διαδικασία και ως οργανωτική λειτουργία, συμμετέχει ενεργά στην αλυσίδα αξίας των πληροφοριών, επεξεργαζόμενος δεδομένα και πληροφορίες σε νοημοσύνη που μπορούν να αξιοποιηθούν σε όλα τα επίπεδα ενός οργανισμού, οδηγώντας σε καλύτερες, έγκαιρες επιχειρηματικές αποφάσεις (Kaario, Peltola 2008, Clark et al. 2007). Η νοημοσύνη είναι η γνώση και η πρόγνωση του εσωτερικού και εξωτερικού επιχειρηματικού περιβάλλοντος στο οποίο δραστηριοποιείται ο οργανισμός (Herring 1988).

Για να γίνει μια διαφορά μεταξύ γνώσης και νοημοσύνης, η ευφυΐα μπορεί να ειπωθεί ότι περιλαμβάνει πληροφορίες σχετικά με κρίσιμες τάσεις και πρότυπα καθώς και σχέσεις μεταξύ των ενεργειών του οργανισμού και των πελατών που υποδηλώνουν ορισμένες αιτίες και αναμενόμενες αλλαγές. Αυτές οι αλλαγές θα μπορούσαν να οδηγήσουν σε νέες ευκαιρίες και απειλές στις επιχειρήσεις, οι οποίες κατευθύνουν τη λήψη στρατηγικών αποφάσεων (Pirttimäki 2007).

Η συμβατική ιεραρχία πληροφοριών ωστόσο, έχει επικριθεί αρκετά στο παρελθόν. Ο Tuomi (1999-2000) δηλώνει ότι «*δεν μπορεί να υπάρχουν μεμονωμένα μέρη απλών γεγονότων εκτός εάν κάποιος τα έχει δημιουργήσει, χρησιμοποιώντας τις γνώσεις του. Τα δεδομένα μπορούν να προκύψουν μόνο εάν μια δομή νοήματος, ή σημασιολογία, πρώτα καθοριστεί και στη συνέχεια χρησιμοποιηθεί για την αναπαράσταση πληροφοριών*».

Αυτή η ιδέα προτείνει μια αντίστροφη ιεραρχία, η οποία είναι η σχετική έρευνα κατά το σχεδιασμό συστημάτων πληροφοριών και την κατασκευή μιας σημασιολογικά καθορισμένης βάσης δεδομένων. Ωστόσο, στην εφαρμογή των *Μεγάλων Δεδομένων* η συμβατική ιεραρχία της δημιουργίας αξίας πληροφοριών, μπορεί να θεωρηθεί κατάλληλη, καθώς ο στόχος είναι να γίνουν νέες παρατηρήσεις από προϋπάρχοντα δεδομένα εντός και εκτός του οργανισμού.

Τα Big Data ή διαφορετικά Μεγάλα Δεδομένα είναι μια αναδυόμενη περιοχή που εφαρμόζεται για τη διαχείριση συνόλων δεδομένων των οποίων το μέγεθος αναφέρεται πέρα από την ικανότητα των εργαλείων λογισμικού που χρησιμοποιούνται, συνήθως να συλλάβουν, να διαχειριστούν και να αναλύσουν έγκαιρα αυτόν τον συγκεκριμένο όγκο δεδομένων. Η ποσότητα των δεδομένων που αναλύεται στις μέρες μας, αναμένεται να διπλασιάζεται κάθε δύο χρόνια (IDC, 2012).

Όλα αυτά τα δεδομένα αναφέρονται πολύ συχνά ως αδόμητα και από ποικίλες πηγές, όπως τα μέσα κοινωνικής δικτύωσης, αισθητήρες, επιστημονικές εφαρμογές, αρχεία βίντεο και εικόνων, ευρετηρίαση αναζήτησης στο Διαδίκτυο, ιατρικά αρχεία, επιχειρηματικές συναλλαγές και αρχεία καταγραφής συστήματος. Τα μεγάλα δεδομένα κερδίζουν ολοένα και μεγαλύτερη προσοχή, καθώς ο αριθμός των συσκευών που συνδέονται με το λεγόμενο «Internet of Things» (IoT), εξακολουθεί να αυξάνεται σε απρόβλεπτα επίπεδα, παράγοντας μεγάλες ποσότητες δεδομένων που πρέπει να μετατραπούν σε πολύτιμες πληροφορίες.

Επιπλέον, είναι πολύ δημοφιλής στις μέρες μας, η αγορά πρόσθετης υπολογιστικής ισχύος και αποθήκευσης κατ' απαίτηση από δημόσιους παρόχους cloud για την εκτέλεση μιας εντατικής παράλληλης επεξεργασίας δεδομένων. Με αυτόν τον τρόπο, τα θέματα ασφάλειας και απορρήτου, μπορούν δυνητικά να ενισχυθούν από τον όγκο, την ποικιλία και την ευρεία ανάπτυξη της υποδομής του συστήματος για την υποστήριξη εφαρμογών Big Data σε επιχειρήσεις.

1.2 Ορισμός των Big Data (Μεγάλα Δεδομένα)

Τα μεγάλα δεδομένα υπόσχονται να παράγουν αναλυτικές γνώσεις που θα διευρύνουν το σώμα της ερευνητικής και κοινωνικής επιστημονικής γνώσης, θα ενισχύσουν δραματικά την ανθρώπινη αυτογνωσία και κατανόηση και θα βελτιώσουν σημαντικά τη λήψη αποφάσεων στον δημόσιο και ιδιωτικό τομέα. Έχουν ήδη οδηγήσει στην ανάπτυξη εντελώς νέων κατηγοριών προϊόντων και υπηρεσιών, πολλές από τις οποίες έχουν γίνει αποδεκτές με ενθουσιασμό τόσο από ιδρύματα όσο και από ανθρώπους.

Ωστόσο, όταν αυτά τα δεδομένα δεσμεύονται να συλλέγουν πληροφορίες για την ανθρώπινη δραστηριότητα, έχουν θεωρηθεί ότι θέτουν μια πρόκληση για τις βασικές αρχές της αυτονομίας, της δικαιοσύνης, της δίκαιης διαδικασίας, της ιδιοκτησίας, της αλληλεγγύης και, ίσως το πιο σημαντικό, της ιδιωτικής ζωής.

Δεδομένης αυτής της αντίθεσης, ορισμένοι έχουν ζητήσει πλήρεις απαγορεύσεις σε διαφορετικές τεχνικές μεγάλων δεδομένων, ενώ άλλοι βρήκαν

σοβαρούς λόγους να κάνουν επιτέλους την προσοχή και την προστασία της ιδιωτικής ζωής στην υπόθεση ότι τα μεγάλα δεδομένα θα αναπληρώσουν τα πιθανά μειονεκτήματά τους. Φυσικά, υπάρχουν ακόμη άλλοι που επιδιώκουν μια προσέγγιση αρχών για την ιδιωτικότητα που σέβεται τα σημαντικά ιδανικά που διατηρεί η ιδιωτικότητα, παρέχοντας παράλληλα την απαιτούμενη ευελιξία για να εκπληρωθούν αυτές οι υποσχέσεις (Lane, 2014).

Ωστόσο, καθώς τα Big Data επεκτείνονται με τη βοήθεια των δημόσιων υπηρεσιών του cloud, οι παραδοσιακές λύσεις ασφαλείας προσαρμοσμένες σε ιδιωτικές υποδομές υπολογιστών, που περιορίζονται σε μια ορθά καθορισμένη περίμετρο ασφαλείας, όπως *τείχη προστασίας* και *αποστρατιωτικοποιημένες ζώνες* δεν είναι πιο αποτελεσματικές. Χρησιμοποιώντας Big Data, απαιτούνται συναρτήσεις ασφαλείας για να λειτουργήσουν πάνω στην ετερογενή σύνθεση διαφορετικού υλικού, λειτουργικών συστημάτων και τομέων δικτύου.

Σε αυτό το υπολογιστικό περιβάλλον τύπου *παζλ*, η ικανότητα αφαίρεσης του Software-Defined Networking (SDN) φαίνεται πολύ σημαντικό χαρακτηριστικό που μπορεί να επιτρέψει την αποτελεσματική ανάπτυξη ασφαλών υπηρεσιών Big Data πάνω από την ετερογενή υποδομή. Η τεχνική SDN εισάγει την αφαίρεση αυτή επειδή διαχωρίζει το επίπεδο ελέγχου από την υποκείμενη υποδομή του συστήματος που εποπτεύεται και ελέγχεται αντίστοιχα.

Ο διαχωρισμός της λογικής ελέγχου ενός δικτύου από τους υποκείμενους φυσικούς δρομολογητές και διακόπτες που προωθούν την κυκλοφορία, επιτρέπει στους διαχειριστές συστήματος να δημιουργούν προγράμματα ελέγχου υψηλού επιπέδου που να καθορίζουν τη συμπεριφορά ενός ολόκληρου δικτύου, σε αντίθεση με τα συμβατικά δίκτυα, σύμφωνα με τα οποία οι διαχειριστές πρέπει να κωδικοποιήσουν τη λειτουργικότητα τους, όσον αφορά τη διαμόρφωση της συσκευής χαμηλού επιπέδου.

Το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (NIST) εγκαινίασε πρόσφατα ένα πλαίσιο με ένα σύνολο εθελοντικών κατευθυντήριων γραμμών για να βοηθήσει τους οργανισμούς να κάνουν τις επικοινωνίες και τις υπολογιστικές τους λειτουργίες ασφαλέστερες (NIST, 2014). Αυτό θα μπορούσε να επιτευχθεί μέσω μιας συστηματικής επαλήθευσης της υποδομής του κάθε συστήματος, όσον αφορά την εκτίμηση κινδύνου, την προστασία από απειλές και τις δυνατότητες απόκρισης και ανάκτησης από επιθέσεις.

Ακολουθώντας τις τελευταίες αρχές επαλήθευσης, ο Οργανισμός Defense Advanced Research Projects Agency (DARPA) δημιουργεί ένα πρόγραμμα που

ονομάζεται Mining and Understanding Software Enclaves (MUSE) για να βελτιώσει την ποιότητα του λογισμικού του αμερικανικού στρατού.

Αυτό το πρόγραμμα έχει σχεδιαστεί για να παράγει πιο ισχυρό λογισμικό που μπορεί να λειτουργήσει με μεγάλα σύνολα δεδομένων χωρίς να προκαλεί σφάλματα ή συντριβή κάτω από τον τεράστιο όγκο πληροφοριών (DARPA, 2014). Επιπλέον, η ασφάλεια και το απόρρητο γίνονται πολύ επείγουσες πτυχές Big Data που πρέπει να αντιμετωπιστούν (Agrawal, Das, & El Abbadi, 2011).

Για να καταδείξουν αυτό το γεγονός, τα κοινωνικά δίκτυα έδωσαν τη δυνατότητα στους ανθρώπους να μοιράζονται και να διανέμουν πολύτιμο ψηφιακό περιεχόμενο που προστατεύεται από πνευματικά δικαιώματα, με έναν πολύ εύκολο τρόπο. Κατά συνέπεια, οι συμπεριφορές παραβίασης πνευματικών δικαιωμάτων, όπως η παράνομη αντιγραφή, η κακόβουλη διανομή, η μη εξουσιοδοτημένη πρόσβαση και χρήση και η δωρεάν κοινή χρήση ψηφιακού περιεχομένου που προστατεύεται από πνευματικά δικαιώματα, θα γίνουν πολύ πιο συνηθισμένο φαινόμενο.

Για να μετριαστούν αυτά τα προβλήματα, τα Big Data θα πρέπει να έχουν σταθερές λύσεις για την υποστήριξη του απορρήτου του περιεχομένου τους και των πνευματικών δικαιωμάτων του συγγραφέα (Marques & Serrão, 2013a). Επίσης, οι χρήστες μοιράζονται όλο και περισσότερα προσωπικά δεδομένα και περιεχόμενο που δημιουργείται από τους χρήστες μέσω των φορητών συσκευών και των υπολογιστών τους σε κοινωνικά δίκτυα και υπηρεσίες cloud, χάνοντας τον έλεγχο δεδομένων και περιεχομένου με σοβαρό αντίκτυπο στο απόρρητό τους.

Τέλος, μια δυνητικά πολλά υποσχόμενη προσέγγιση είναι η δημιουργία πρόσθετης αβεβαιότητας για τους εισβολείς, αλλάζοντας δυναμικά τις ιδιότητες του συστήματος σε αυτό που ονομάζεται κινούμενος στόχος στον κυβερνοχώρο (MT) (Okhravi, et al., 2014). Παρουσιάζουν μια σύνοψη πολλών τύπων τεχνικών MT, εξετάζουν τα πλεονεκτήματα και τις αδυναμίες του καθενός και κάνουν συστάσεις για μελλοντική έρευνα σε αυτόν τον τομέα.

Τα μεγάλα δεδομένα αναφέρονται ρητά σε σύνολα δεδομένων που είναι τόσο τεράστια ή περίπλοκα που το συμβατικό λογισμικό επεξεργασίας δεδομένων δεν επαρκεί για να τα χειριστεί. Κάθε μέρα, μια εταιρεία κατακλύζεται από τον τεράστιο όγκο δεδομένων, οργανωμένων και μη. Λόγω των πρόσφατων τεχνικών εξελίξεων, ο όγκος των δεδομένων που παράγονται από το Διαδίκτυο, τους ιστότοπους κοινωνικής δικτύωσης, τα δίκτυα αισθητήρων, τις εφαρμογές υγειονομικής περίθαλψης και πολλές άλλες επιχειρήσεις αυξάνεται δραματικά κάθε μέρα. Με άλλα λόγια, τα μεγάλα δεδομένα αναφέρονται στον τεράστιο όγκο δεδομένων που

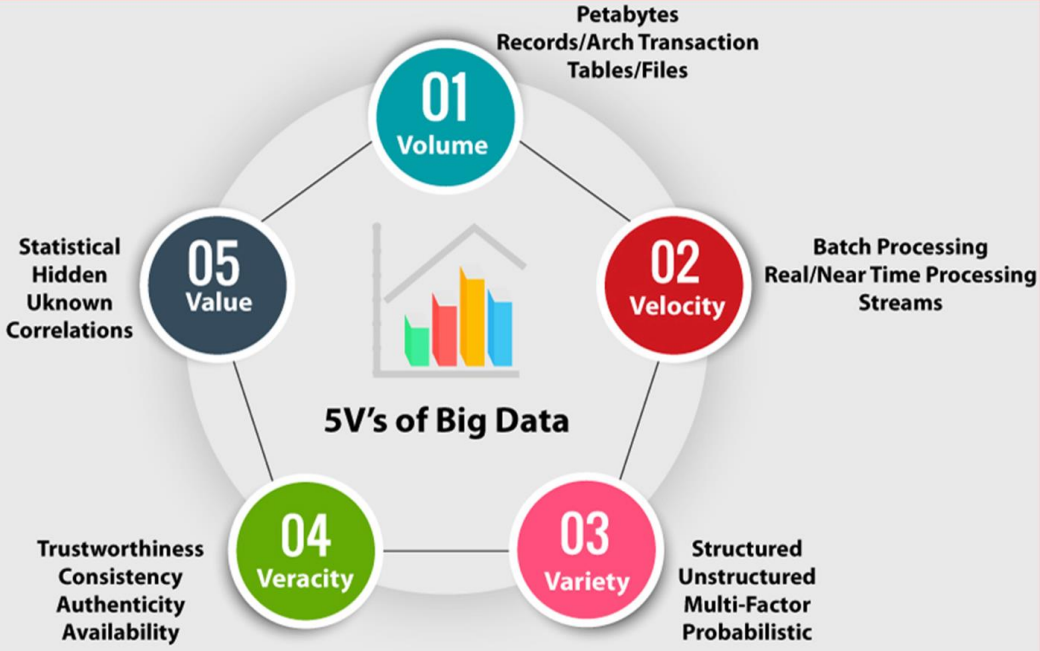
δημιουργούνται από πολλές πηγές σε πολλές διαφορετικές μορφές σε εξαιρετικά γρήγορες ταχύτητες (Jain, 2016).

Ο όρος "μεγάλα δεδομένα" αναφέρεται σε μια νέα γενιά τεχνολογιών και αρχιτεκτονικών που έχουν σχεδιαστεί για να διαχωρίζουν οικονομικά την αξία από τους πολύ μεγάλους όγκους δεδομένων. Αυτές οι τεχνολογίες επιτρέπουν τη λήψη, την ανακάλυψη και την ανάλυση υψηλής ταχύτητας. Μερικές από τις βασικές ιδιότητες των μεγάλων δεδομένων περιλαμβάνουν τον όγκο, την ταχύτητα και την ποικιλία.

Μεταγενέστερες μελέτες τόνισαν ότι ο ορισμός των 3Vs (όγκος, ταχύτητα και ποικιλία) δεν είναι αρκετά συγκεκριμένος για να εξηγήσει τα μεγάλα δεδομένα που αντιμετωπίζουμε τώρα. Έτσι, άλλοι όροι προστέθηκαν στον ορισμό, όπως η αλήθεια, η εγκυρότητα, η αξία, η μεταβλητότητα, ο τόπος, το λεξιλόγιο και η ασάφεια (Jain, 2016).

Η ποικιλομορφία των δεδομένων, που μπορεί να περιλαμβάνει κείμενο, ήχο, εικόνα και βίντεο, είναι ένα κοινό θέμα στα μεγάλα δεδομένα. Η ποικιλία χρησιμεύει ως μεταφορά για τα διάφορα χαρακτηριστικά των δεδομένων. Αρκετές στρατηγικές έχουν αναπτυχθεί τα τελευταία χρόνια για να εγγυηθεί το απόρρητο τεράστιων δεδομένων. Τα στάδια του κύκλου ζωής των μεγάλων δεδομένων, συμπεριλαμβανομένης της παραγωγής, αποθήκευσης και επεξεργασίας δεδομένων, μπορούν να χρησιμοποιούνται για την ταξινόμηση αυτών των μεθόδων. Οι περιορισμοί πρόσβασης και οι τεχνικές παραποίησης δεδομένων χρησιμοποιούνται κατά τη φάση δημιουργίας δεδομένων για τη διασφάλιση του απορρήτου (Jain, 2016).

Ο όρος *Μεγάλα Δεδομένα (Big Data)* χρησιμοποιείται συνήθως για μεγάλα και σύνθετα σύνολα δεδομένων που δεν μπορούν να υποβληθούν σε επεξεργασία/διαχείριση από τυπικό λογισμικό το οποίο χαρακτηρίζεται μέσω 5Vs συγκεκριμένα ως όγκος (μέγεθος δεδομένων), ταχύτητα (υψηλή ταχύτητα δεδομένων), ποικιλία (διαφορετικοί τύποι δεδομένων και πηγές), την ακρίβεια (συνέπεια και αξιοπιστία των δεδομένων) και την αξία (εκροές που λαμβάνονται από το σύνολο δεδομένων) (Agrawal, Das, & El Abbadi, 2011). Η εικόνα No.1 δείχνει τους διαφορετικούς χαρακτήρες των Big Data μέσω 5Vs.



Εικόνα No.1 – Τα 5 Στοιχεία των Μεγάλων Δεδομένων

1.Όγκος: Η ικανότητα επεξεργασίας μεγάλων ποσοτήτων δεδομένων είναι μια κρίσιμη πτυχή των Μεγάλων Δεδομένων, ειδικά επειδή ο όγκος είναι μια από τις μεγαλύτερες προκλήσεις των συμβατικών δομών πληροφορικής στις οποίες οι εταιρείες δεν είναι σε θέση να επεξεργαστούν μεγάλες ποσότητες αρχειοθετημένων αρχείων καταγραφής δεδομένων. Ένα παράδειγμα τέτοιων επιχειρήσεων είναι η Wal-Mart, η οποία συνήθιζε να αποθηκεύει 1.000 terabytes δεδομένων το 1999 σε αντίθεση με πάνω από 2,5 petabytes δεδομένων το έτος 2012.

2.Ταχύτητα: Αυτό δείχνει την υψηλή ταχύτητα με την οποία τα δεδομένα δημιουργούνται, επεξεργάζονται, αποθηκεύονται και αναλύονται από σχεσιακή βάση δεδομένων, επιπλέον της ταχύτητας με την οποία δημιουργούνται και διακινούνται νέα δεδομένα, όπως ο τρόπος με τον οποίο οι πληροφορίες στα μέσα κοινωνικής δικτύωσης γίνονται viral στην ουσία. δευτερόλεπτα ή τις εκατό ώρες περιεχομένου βίντεο που ανεβαίνουν καθημερινά στο YouTube.

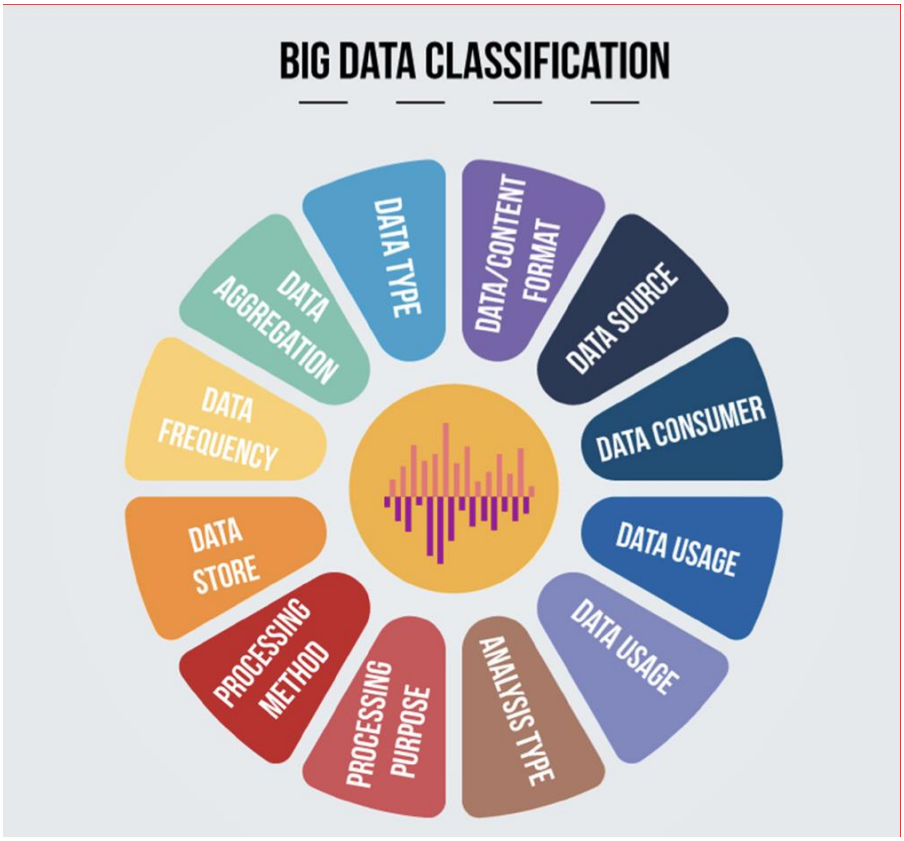
3.Ποικιλία: Η ποικιλία είναι μια άλλη ενδιαφέρουσα πτυχή των Μεγάλων Δεδομένων, που σημαίνει ότι αυτά τα δεδομένα μπορούν να έρθουν σε δομημένη, αδόμητη ή ημι-δομημένη μορφή, καθιστώντας την εξαιρετικά δύσκολη την τοποθέτηση σε μια σχεσιακή βάση δεδομένων, ειδικά επειδή στο 90% των περιπτώσεων, τα δεδομένα που δημιουργούνται είναι σε μη δομημένη μορφή,

γεγονός που καθιστά σημαντικό για τους αναλυτές δεδομένων να γνωρίζουν την κατηγορία στην οποία ανήκουν τα Μεγάλα Δεδομένα.

4.Ακρίβεια Δεδομένων: Όταν ασχολείται κανείς με *Μεγάλα Δεδομένα*, υπάρχει πάντα η πιθανότητα λήψης περίπλοκων δεδομένων. Η ποιότητα των δεδομένων και η ακρίβεια της ανάλυσης εξαρτάται σε μεγάλο βαθμό από την ακρίβεια της πηγής τους.

5.Αξία: Παρόλο που υπάρχουν μεγάλες δυνητικές αξίες στη χρήση των Μεγάλων Δεδομένων, εκτός εάν υπάρχει απόδοση επένδυσης (παράγεται αξία) για την εταιρεία. Θα ήταν πολύ δαπανηρό (και άχρηστο) η εφαρμογή συστημάτων υποδομής πληροφορικής για την αποθήκευση στα *Μεγάλα Δεδομένα*.

Μπορούμε να χρησιμοποιήσουμε διαφορετικές προσεγγίσεις για την απόκτηση, επεξεργασία, αποθήκευση και ανάλυση Μεγάλων Δεδομένων. Ωστόσο, είναι σημαντικό να έχετε κατά νου ότι υπάρχουν διαφορετικά χαρακτηριστικά των πηγών από τις οποίες λαμβάνονται Μεγάλα Δεδομένα, όπως τύπος δεδομένων, μέγεθος, ταχύτητα, συνέπεια/αξιοπιστία και συχνότητα. Επιπλέον, η επιλογή και η κατασκευή μιας λύσης μπορεί να είναι προκλητική λόγω παραγόντων όπως η διακυβέρνηση, η ασφάλεια και οι πολιτικές]. Τα Μεγάλα Δεδομένα μπορούν να κατηγοριοποιηθούν σύμφωνα με τις ακόλουθες ταξινομήσεις: τύπος δεδομένων, περιεχόμενο, πηγή, καταναλωτής, χρήση, τύπος ανάλυσης, σκοπός επεξεργασίας, μέθοδος επεξεργασίας, αποθήκευση και συχνότητα όπως φαίνεται στο Σχήμα No.2 παρακάτω.



Εικόνα No.2 - Ταξινόμηση των Μεγάλων Δεδομένων

1.3 Θεμιτή Χρήση των Big Data

Αποτελεί γεγονός πως μεγάλοι όγκοι δεδομένων παράγονται και συλλέγονται συνεχώς από επιχειρήσεις και κυβερνητικούς οργανισμούς. Η τρέχουσα αυξανόμενη έμφαση σε μεγάλους όγκους δεδομένων, θα δημιουργήσει πιθανώς ευκαιρίες και δρόμους για την κατανόηση του τρόπου επεξεργασίας τέτοιων δεδομένων σε πολλούς διαφορετικούς κλάδους. Ωστόσο, η υπόσχεση για μεγάλα δεδομένα έχει κόστος, αφού το απόρρητο του πελάτη βρίσκεται συνεχώς σε κίνδυνο.

Οι τρέχουσες τεχνικές ανάλυσης μεγάλων δεδομένων και *εξόρυξης*, έχουν περιορισμούς όσον αφορά τη διασφάλιση της συμμόρφωσης με τους όρους και τη νομοθεσία περί απορρήτου. Ανεξάρτητα από τροποποιήσεις στις εφαρμογές και τους κανόνες απορρήτου, οι προγραμματιστές θα πρέπει να μπορούν να επιβεβαιώνουν ότι οι εφαρμογές τους συμμορφώνονται με τις συμφωνίες απορρήτου και ότι τα ευαίσθητα δεδομένα διατηρούνται ιδιωτικά. Η ανάγκη για νέες συνεισφορές στους τομείς των επίσημων τεχνικών και των διαδικασιών δοκιμών θα πρέπει να εντοπιστεί για να αντιμετωπιστούν αυτές οι δυσκολίες (Jain, 2016).

Μαζί με την προστασία του ατομικού απορρήτου, τα μεγάλα δεδομένα επανάφεραν το «συλλογικό απόρρητο» στο προσκήνιο. Αυτή η ιδέα υποδηλώνει ότι οι κοινωνικές ομάδες πρέπει επίσης να έχουν το δικαίωμα στην ιδιωτική ζωή, όχι μόνο τα άτομα. Η ανάγκη για προστασία της ιδιωτικής ζωής έχει συχνά συγκριθεί με τη σημασία της δημόσιας ασφάλειας, όπως σημειώνουν οι ακαδημαϊκοί, αφού «*πρέπει να συμβιβαστούν δύο ηθικές υποχρεώσεις: η προώθηση των ανθρωπίνων δικαιωμάτων και η αύξηση της ανθρώπινης ευημερίας». Αλλά διαφωνεί με την ιδέα ότι το τελευταίο θα ήταν ένα πολιτικό ζήτημα που επηρεάζει το ευρύ κοινό και το πρώτο ένα ηθικό ζήτημα που επηρεάζει τα δικαιώματα των ανθρώπων*» (Richterich, 2018).

1.4 Big Data Management

Αποτελεί γεγονός πως την τελευταία δεκαετία, πολλοί οργανισμοί έχουν αναπτύξει Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα για να βοηθήσουν τη διαχείριση του οργανισμού τόσο επιχειρησιακά όσο και στρατηγικά. Το 2014, μια έρευνα της Gartner έδειξε ότι η ζήτηση για Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα στους οργανισμούς, συνεχίζει να αυξάνεται και οι παγκόσμιες δαπάνες για τα δεδομένα αυτά, έφτασαν τα 14 δισεκατομμύρια δολάρια το 2013 (Sommer, Sood, 2014).

Η παγκόσμια αγορά για Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα και αναλυτικών στοιχείων αναμένεται να αυξηθεί στα 20,8 δισεκατομμύρια δολάρια μέχρι το 2018, σύμφωνα με προβλέψεις της παγκόσμιας εταιρείας έρευνας αγοράς και συμβούλων (Markets, Markets (2013). Με τη διαθεσιμότητα μεγάλων δεδομένων, πιο εξελιγμένων και ισχυρών υπολογιστικών δυνατοτήτων και αναλυτικού λογισμικού, η διαθεσιμότητα επιχειρηματικής ευφυΐας για τη λήψη στρατηγικών αποφάσεων φαίνεται να γίνεται όλο και πιο σημαντική. Ωστόσο, το ερώτημα είναι αν τα ανώτερα στελέχη αξιοποιούν στο έπακρο τους διαθέσιμους πόρους ή άλλοι παράγοντες θεωρούνται πιο σημαντικοί για την ποιότητα της λήψης στρατηγικών αποφάσεων.

Τα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα αναπτύσσονται ως μια τεχνολογική λύση για την αποθήκευση, την ενσωμάτωση και την ανάλυση των πληροφοριών που απαιτούνται για την υποστήριξη της λήψης αποφάσεων του οργανισμού (Popovic et al, 2012). Η έννοια ενός συστήματος για τα Μεγάλα Δεδομένα ωστόσο, δεν είναι μια πρόσφατη εξέλιξη και ο Power (2007) εξηγεί ότι τα συστήματα αυτά εξελίχθηκαν από συστήματα υποστήριξης αποφάσεων που

εμφανίστηκαν κατά τη δεκαετία του 1960 για να βοηθήσουν τον προγραμματισμό και τη λήψη αποφάσεων.

Ωστόσο, πριν από την επίσημη αναγνώριση στα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα, η ιδέα συζητιόταν ήδη. Το 1958, ο Luhn περιέγραψε για πρώτη φορά τον όρο Μεγάλα Δεδομένα, χρησιμοποιώντας τον ορισμό της νοημοσύνης από το λεξικό Webster: «*Η ικανότητα να κατανοεί κανείς τις αλληλεπιδράσεις των παρουσιαζόμενων γεγονότων με τέτοιο τρόπο ώστε να καθοδηγεί τη δράση προς έναν επιθυμητό στόχο*» (Luhn, 1958).

Επιπλέον, ο Luhn παρείχε επίσης μια επισκόπηση των βασικών στοιχείων ενός συστήματος για τα Μεγάλα Δεδομένα, πολλά από τα οποία εξακολουθούν να είναι αναγνωρίσιμα σήμερα. Τέλος, προτείνεται ότι τα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα θα συλλέγουν αυτόματα δεδομένα από πολλές πηγές και θα επικοινωνεί πληροφορίες για να παρέχει ευφυΐα που να επιτρέπει την επίλυση προβλημάτων.

Σε συνέχεια της περιγραφής του συστήματος από τον Luhn, μεταξύ της δεκαετίας του 1960 και του 1980 αναπτύχθηκαν διάφορες μορφές Συστημάτων Υποστήριξης Αποφάσεων (DSS) για να βοηθήσουν στη λήψη αποφάσεων και τον προγραμματισμό. Από το DSS άρχισαν να αναπτύσσονται αποθήκες δεδομένων, Συστήματα Πληροφοριών Διαχείρισης, Συστήματα Σχεδιασμού Πόρων.

Κατά τη διάρκεια της δεκαετίας του 1990, ο Όμιλος Gartner εισήγαγε τον σύγχρονο όρο Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα που ορίζει την «επιχειρηματική ευφυΐα» ως γενικό όρο για να περιγράψει «*έννοιες και μεθόδους για τη βελτίωση της λήψης επιχειρηματικών αποφάσεων χρησιμοποιώντας συστήματα υποστήριξης που βασίζονται σε γεγονότα*» (Power 2007). Τα τελευταία χρόνια, με την αυξανόμενη πολυπλοκότητα στα περιβάλλοντα της αγοράς και τον αυξανόμενο ανταγωνισμό μεταξύ των προμηθευτών συστημάτων για τα Μεγάλα Δεδομένα, οι προγραμματιστές αύξησαν τις δυνατότητες των σύγχρονων Συστημάτων Διαχείρισης για τα Μεγάλα Δεδομένα να αποθηκεύουν, να συνθέτουν, να αναλύουν και να επικοινωνούν δεδομένα, πληροφορίες και γνώσεις για τη λήψη αποφάσεων (Power 2007, Ranjan 2008).

Το σύγχρονο σύστημα συχνά περιλαμβάνει μια αποθήκη δεδομένων για αποθήκευση. μετασχηματισμός εξαγωγής και ικανότητα φόρτωσης για τη μετατροπή ακατέργαστων δεδομένων από εταιρικά συστήματα (κατά την προετοιμασία για την αποθήκη δεδομένων χρησιμοποιώντας μια σειρά καθορισμένων ορισμών και δομών δεδομένων), την ικανότητα εξόρυξης δεδομένων, πίνακες εργαλείων για την

αποτελεσματική επικοινωνία των δεδομένων και ένα αναλυτικό εργαλείο για την πρόβλεψη και την ανάπτυξη άλλων γνώσεων (Aruldoss, Travis και Venkatesan 2014).

Συνοπτικά, τα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα, είτε λαμβάνοντας υπόψη το σύγχρονο εργαλείο είτε την ιστορική ιδέα, επικεντρώνονταν πάντα στην παροχή πληροφοριών στους λήπτες αποφάσεων για την ενημέρωση της λήψης αποφάσεων. Για τους σκοπούς αυτής της μελέτης, Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα ορίζονται ως ένα σύστημα πληροφορικής το οποίο:

- ✓ Αποθηκεύει δεδομένα
- ✓ Μετατρέπει αυτά τα δεδομένα σε επιχειρηματική ευφυΐα (συμπεριλαμβανομένης της ανάλυσης δεδομένων και ανάπτυξη ενόρασης)· και
- ✓ Παρέχει επιχειρηματική ευφυΐα στους υπεύθυνους λήψης αποφάσεων.

Αν και τα οφέλη από την ύπαρξη ενός τέτοιου συστήματος σε έναν οργανισμό μπορεί να φαίνονται αυτονόητα, είναι χρήσιμο να τα εξετάσουμε πιο επίσημα. Ειδικότερα, όσον αφορά τα οφέλη που μπορεί να σχετίζονται με αυτή τη μελέτη. Η ανάπτυξη στα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα παρουσιάζει μια σημαντική επένδυση για τους οργανισμούς και ως εκ τούτου η επίτευξη θετικής απόδοσης επένδυσης είναι σημαντική (Elbashir et al., 2008).

Τα οφέλη υλοποίησης στα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα έχουν καθοριστεί καλά και περιλαμβάνουν αυτά που σχετίζονται με τη βελτιωμένη λήψη αποφάσεων (Braeutigam, Gerlach and Miller 2006, Elbashir, Collier and Davern 2008, Popvic, Hackney, Coelho and Jaklic 2012, Yeoh, Koronios and Gao 2008). Ο Negash (2004) υποστηρίζει ότι τα συστήματα BI έχουν εισαχθεί για να βελτιώσουν την επικαιρότητα και την ποιότητα των εισροών της διαδικασίας λήψης αποφάσεων.

Οι Aruldoss, Travis και Venkatesan (2014) τόνισαν ότι ο θεμελιώδης σκοπός ενός συστήματος για τα Μεγάλα Δεδομένα πρέπει να είναι να παρέχει στους οργανισμούς την ικανότητα να παρακολουθούν την απόδοση και τις λειτουργίες της επιχείρησης και να βοηθούν τη διοίκηση στην ανάπτυξη επιχειρηματικής στρατηγικής και ενεργειών. Συνεχίζουν περιγράφοντας ένα από τα βασικά πλεονεκτήματα οποιουδήποτε συστήματος είναι η παροχή των σωστών πληροφοριών την κατάλληλη στιγμή για να καταστεί δυνατή η λήψη αποφάσεων να λαμβάνει αποφάσεις χρησιμοποιώντας την επιχειρηματική ευφυΐα.

Ως αποτέλεσμα, η εισαγωγή σε Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα σε οργανισμούς, συχνά καθοδηγείται από την ανάγκη να υπάρχουν διαθέσιμες βελτιωμένες πληροφορίες σχετικά με τις επιχειρήσεις για να βοηθήσουν στη λήψη αποφάσεων. Επιπλέον, οι Gangadharan και Swami (2004) εξηγούν ότι τα δεδομένα μετατρέπονται σε εταιρικό πόρο και η εστίαση μετατοπίζεται από την ποσότητα στην ποιότητα της γνώσης.

Κατά συνέπεια, τα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα έχουν σχεδιαστεί για να παρέχουν περισσότερα από ακατέργαστα δεδομένα. Πράγματι, προσθέτουν αξία μέσω της μετατροπής των ακατέργαστων δεδομένων σε πληροφορίες, οι οποίες μπορούν στη συνέχεια να χρησιμοποιηθούν πιο εύκολα για τη βελτίωση της λήψης στρατηγικών και επιχειρησιακών αποφάσεων. Ωστόσο, οι Popovic et al (2012) εξηγούν ότι η μέτρηση των οφελών που προσφέρει η BI Systems μπορεί να είναι δύσκολη καθώς οι αποδόσεις εμφανίζονται μακροπρόθεσμα και συχνά είναι έμμεσες.

Ο βαθμός στον οποίο πραγματοποιούνται αυτά τα οφέλη εξαρτάται σε μεγάλο βαθμό από τον τρόπο και την έκταση στην οποία χρησιμοποιείται ένα σύστημα. Μια μελέτη από τους Turpin και Marais (2004) βρήκε ότι οι υπεύθυνοι λήψης αποφάσεων που έλαβαν συνέντευξη, πολλοί από τους οποίους είχαν επίσημη εκπαίδευση στη χρήση της εξελιγμένης τεχνολογίας υποστήριξης αποφάσεων, συχνά χρησιμοποιούσαν αυτή την τεχνολογία σπάνια κατά τη λήψη αποφάσεων. Οι υπεύθυνοι λήψης αποφάσεων που ερωτήθηκαν ισχυρίστηκαν ότι η διαίσθηση και την ευαισθησία στο πολιτικό περιβάλλον, είχε προτεραιότητα έναντι της ορθολογικής διαδικασίας λήψης αποφάσεων.

Τέλος, οι Jasperson, Carter and Zmund (2005), Mabert and Soni (2001), Venkatesh, Morris, Davies and Davies (2003) και Ramamurthy, Sen και Sinha (2008) έχουν όλοι προτείνει ότι η υποχρησιμοποίηση νέων συστημάτων που εφαρμόστηκαν γενικά σημαίνει ότι οφέλη δεν προκύπτουν στον ενδιαφερόμενο οργανισμό.

Οι Popovic et al (2012) συμφώνησαν επίσης και υποστήριξαν ότι τα οφέλη ενός Συστήματος για τα Μεγάλα Δεδομένα, σε σχέση με τη βελτιωμένη λήψη αποφάσεων, μπορούν να συγκεντρωθούν μόνο εάν το Σύστημα για τα Μεγάλα Δεδομένα χρησιμοποιείται από τους υπεύθυνους λήψης αποφάσεων. Αυτό υποδηλώνει ότι παρόλο που οι οργανισμοί μπορεί να επενδύουν σημαντικούς πόρους στην ανάπτυξη Συστημάτων για τα Μεγάλα Δεδομένα, τα οφέλη μπορεί να μην αποκομιστούν εάν τα Συστήματα Διαχείρισης για τα Μεγάλα Δεδομένα στην

πραγματικότητα δεν υιοθετηθούν και δεν χρησιμοποιηθούν από τους υπεύθυνους λήψης αποφάσεων.

1.5 Κίνδυνοι των Big Data

Αναφερόμενοι σχετικά στους κινδύνους των Μεγάλων Δεδομένων, θα πρέπει να σημειωθεί πως οι παραδοσιακές προσεγγίσεις ασφάλειας και απορρήτου, δεν είναι σε θέση να αντιμετωπίσουν πλήρως τις αλλαγές που έχουν εισαγάγει τα δεδομένα αυτά στον ψηφιακό κόσμο, τα οποία κυμαίνονται από τον όγκο των δεδομένων που συλλέγονται, αποθηκεύονται έως τον χειρισμό τους στις επιχειρήσεις (Agrawal, Das, & El Abbadi, 2011). Τα μέτρα ασφαλείας, όπως σύνθετοι αλγόριθμοι κρυπτογράφησης, περιορισμοί ελέγχου πρόσβασης, τείχη προστασίας και συστήματα ανίχνευσης εισβολών για την ασφάλεια δικτύου μπορούν να σπάσουν, ακόμη και ανώνυμα δεδομένα θα μπορούσαν να επαναπροσδιοριστούν και να συσχετιστούν με έναν συγκεκριμένο χρήστη για κακόβουλη χρήση (IDC, 2012).

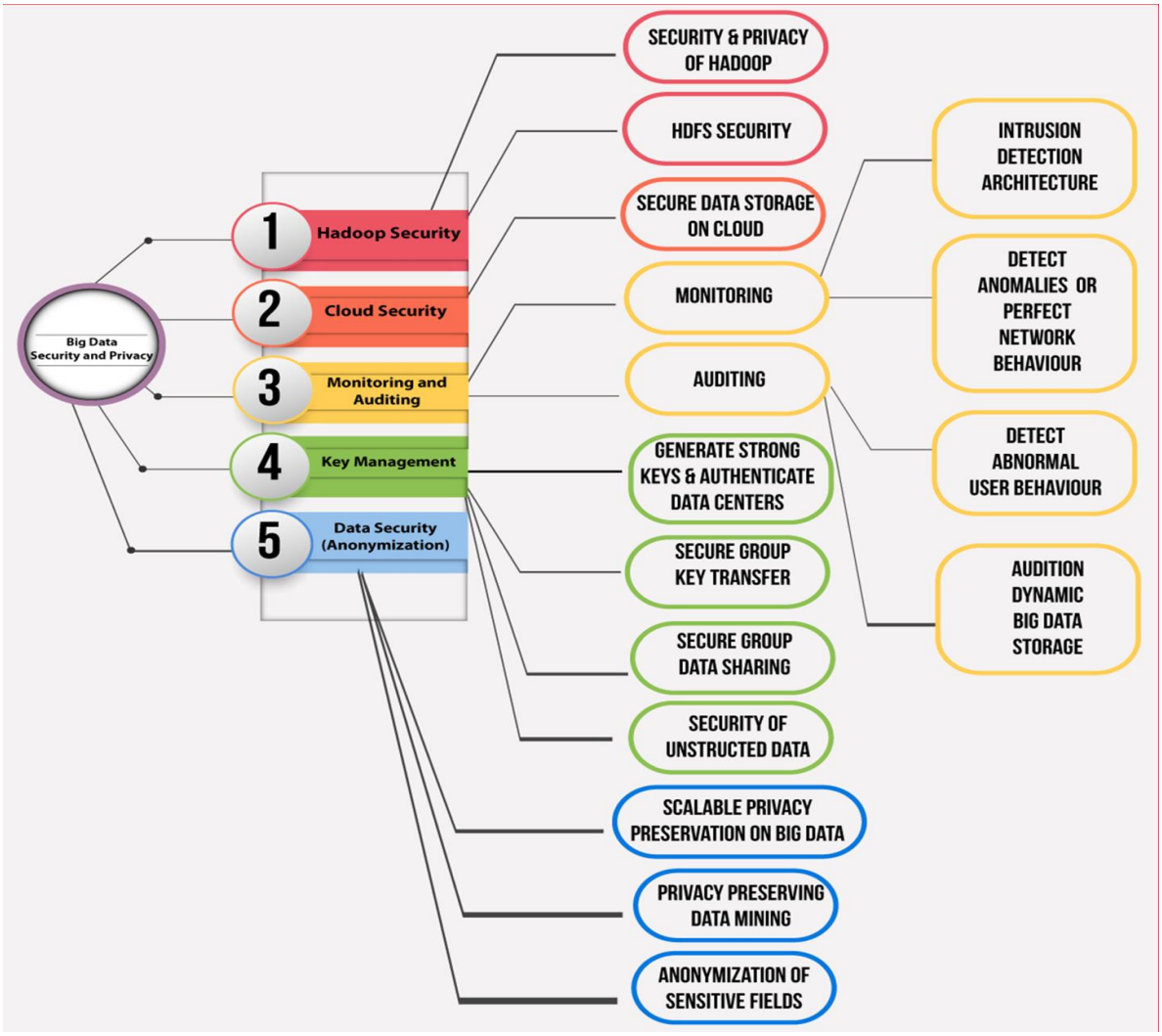
Υπάρχουν ωστόσο, αρκετοί νέοι κανονισμοί που προτείνονται ειδικά για την αντιμετώπιση προκλήσεων που έχουν εισαγάγει τα Big Data στο απόρρητο των ατόμων, προκλήσεις όπως, η εξαγωγή συμπερασμάτων και η συγκέντρωση που καθιστά δυνατή την εκ νέου αναγνώριση ατόμων ακόμη και μετά την αφαίρεση των αναγνωριστικών από ένα σύνολο δεδομένων. Ωστόσο, υπάρχουν περιπτώσεις στις οποίες οι προκαθορισμένοι κανονισμοί μπορεί να οδηγήσουν σε παραβίαση του απορρήτου, όπως η διατήρηση δεδομένων email για ορισμένο χρονικό διάστημα (σε περιπτώσεις έως και 5 χρόνια) που απλώς αφήνει ανοιχτή την πόρτα για πιθανές παραβιάσεις της ιδιωτικής ζωής (IDC, 2012).

Ωστόσο, εδώ αντιμετωπίζει κανείς ένα παλαιότερο δίλημμα, δηλαδή το τρίγωνο ασφαλείας που δηλώνει ότι καθώς κανείς εφαρμόζει δυσκολότερα μέτρα ασφαλείας, επηρεάζει αντίστοιχα αρνητικά τη λειτουργικότητα και την ευκολία χρήσης των συστημάτων, για παράδειγμα, εάν ένας συγκεκριμένος κανονισμός περιορίζει την πρόσβαση των εταιρειών στην ανάλυση και τον χειρισμό ακατέργαστων δεδομένων, οι εταιρείες δεν θα είναι σε θέση να βελτιώσουν τις δραστηριότητές τους.

Ως εκ τούτου, απαιτείται να προτείνει κανείς μια ισορροπημένη προσέγγιση ως προς τους κανονισμούς και τα αναλυτικά στοιχεία που διασφαλίζει το δικαίωμα των εταιρειών στα αναλυτικά στοιχεία καθώς και το απόρρητο των ατόμων. Με λίγα λόγια, ολόκληρο το οικοσύστημα των Μεγάλων Δεδομένων, από την υποδομή και τη διαχείριση έως τις πολιτικές εμπιστοσύνης, την ακεραιότητα και την ποιότητα των

δεδομένων πρέπει να επανεξεταστεί και να εξεταστεί περαιτέρω σε σχέση με θέματα ασφάλειας και ιδιωτικότητας (IDC, 2012).

Οι κίνδυνοι των Μεγάλων Δεδομένων λοιπόν, αναφέρονται σε ορισμένα ζητήματα ασφάλειας και απορρήτου. Ωστόσο, εξακολουθεί να υπάρχει ανάγκη για μια ολοκληρωμένη έρευνα για τον ενδελεχή εντοπισμό και αντιμετώπιση αυτών των ανησυχιών. Επίσης, για να βεβαιωθεί κανείς ότι τα μέτρα ασφαλείας ενσωματώνονται σε όλες τις τεχνολογίες που έχουν αναπτυχθεί για την διαχείριση των Μεγάλων Δεδομένων, όπως τεχνολογίες για υποδομές, διαδικασίες παρακολούθησης και ελέγχου, εφαρμογές και προέλευση δεδομένων. Στο πλαίσιο αυτό, εξετάζονται οι προκλήσεις Big Data (ασφάλεια και απόρρητο) από 5 διαφορετικές οπτικές γωνίες και συγκεκριμένα ως πλαίσιο (Hadoop), υποδομή (Cloud), παρακολούθηση και έλεγχος, διαχείριση κλειδιών και ασφάλεια δεδομένων (ανωνυμοποίηση), όπως φαίνεται στο σχήμα No.3.



Εικόνα No.3 – Περιοχές Ασφάλειας και Απορρήτου Μεγάλων Δεδομένων

1. Hadoop Security

Το πλαίσιο Hadoop είναι ένα κατανεμημένο πλαίσιο διαδικασίας ανοιχτού κώδικα που χρησιμοποιεί το μοντέλο MapReduce για την επεξεργασία μεγάλων συνόλων δεδομένων και χρησιμοποιείται ευρέως από μεγάλες εταιρείες όπως η Google, το LinkedIn, το Facebook και το Yahoo για επεξεργασία δεδομένων ( ). Ωστόσο, αυτό το πλαίσιο δεν αναπτύχθηκε αρχικά για λειτουργία σε μη αξιόπιστο περιβάλλον. Ως εκ τούτου, δεν ενσωματώθηκαν τα απαραίτητα μέτρα ασφαλείας. Η έλλειψη κατάλληλης προστασίας ασφαλείας σε πολλές τεχνολογίες που αναπτύχθηκαν για Big Data, όπως το Hadoop, το Twitter Storm, το Pig, το Hive, το MapReduce, το Mahout και το Cassandra, έχει μετατρέψει την υποδομή σε πρόκληση ασφάλειας για τη διαχείριση και ανάλυση Big Data (IDC, 2012). Ωστόσο, παρά τις ελλείψεις του στην ασφάλεια, το Hadoop έλαβε μεγάλο ενδιαφέρον και επιλέχθηκε ως μία από τις κύριες πλατφόρμες για Big Data, καθιστώντας υποχρεωτικό τον εντοπισμό των τρόπων με τους οποίους μπορούν να προστεθούν οι απαραίτητες προφυλάξεις ασφαλείας, ειδικά επειδή οι χάκερ συνήθως στοχεύουν δεδομένα που είναι αποθηκευμένα στο cloud. Στην περίπτωση αυτή, αναφέρονται οι δύο μεγάλες *αδυναμίες* λειτουργίας ασφαλείας του Hadoop (1-Acessing Data on Cloud, 2-HDFS Security) και συζητιούνται εν συντομία οι τεχνικές που μπορούν να χρησιμοποιηθούν κατά την ανάπτυξη ενός συστήματος Hadoop για την εγγύηση της ασφάλειας και του απορρήτου των δεδομένων:

1. Hadoop Security and Privacy: Ένας τρόπος με τον οποίο παρέχεται ασφαλής πρόσβαση των χρηστών στα δεδομένα που είναι αποθηκευμένα στο cloud, είναι μέσω του ελέγχου ταυτότητας χρήστη πριν από την παραχώρηση πρόσβασης σε έναν κόμβο ονόματος, σε αυτόν τον μηχανισμό τόσο ο χρήστης όσο και ο κόμβος ονόματος δημιουργούν μια συνάρτηση κατακερματισμού χρησιμοποιώντας αλγόριθμους όπως ο SHA-256, ο κόμβος ονόματος εκτελεί μια σύγκριση μεταξύ της τιμής κατακερματισμού που αποστέλλεται από τον χρήστη και αυτής που δημιουργείται και χορηγεί πρόσβαση εάν οι τιμές είναι σωστές. Αυτός ο μηχανισμός καταφέρνει να παρέχει πρόσβαση σε κόμβους δεδομένων. Ένας άλλος εύκολος και κοινώς χρησιμοποιούμενος τρόπος για να διασφαλιστεί η ασφάλεια των δεδομένων και να περιοριστεί η μη εξουσιοδοτημένη πρόσβαση, είναι η εκτέλεση

κρυπτογράφησης και αποκρυπτογράφησης χρησιμοποιώντας αλγόριθμους τυχαίας κρυπτογράφησης όπως AES, Triple DES, RSA, RC6, IDEA και Rijndael, όπως χρησιμοποιούνται από το MapReduce.

2. Ασφάλεια HDFS: Αυτό είναι το κατανεμημένο σύστημα αρχείων του Hadoop το οποίο έχει τρία κύρια στοιχεία, δηλαδή ως κόμβο ονόματος (κύριο κόμβο), κόμβο δεδομένων και δευτερεύοντα κόμβο ονόματος. Το HDFS δημιουργεί πολλά αντίγραφα κάθε μπλοκ δεδομένων προκειμένου να διασφαλίσει τη διαθεσιμότητα και τον γρήγορο χρόνο απόκρισης. Ωστόσο, το HDFS έχει ορισμένα προβλήματα σχετικά με τον έλεγχο ταυτότητας για τα οποία έχει προταθεί η χρήση του Kerberos (πρωτόκολλο ελέγχου ταυτότητας) για να επιτρέπεται στους κόμβους να αποδείξουν την ταυτότητά τους μεταξύ τους. Για να διασφαλιστεί η ασφάλεια των αναπαραγόμενων δεδομένων και να βεβαιωθεί κανείς ότι η πρόσβαση παρέχεται μόνο στους εξουσιοδοτημένους χρήστες, ο αλγόριθμος Bullseye χρησιμοποιείται για την παρακολούθηση δεδομένων.

2. Cloud Security

Το cloud computing χρησιμοποιείται ευρέως σε συνδυασμό με τα Μεγάλα Δεδομένα, λόγω των πολυάριθμων πλεονεκτημάτων που παρέχει, συγκεκριμένα ως διαθεσιμότητα υπηρεσιών κατ' απαίτηση/σε πραγματικό χρόνο, ευρεία πρόσβαση και κοινή χρήση πόρων. Ωστόσο, η χρήση του υπολογιστικού νέφους συνοδεύεται από έναν τεράστιο αριθμό προκλήσεων ασφαλείας, καθώς αυτή η τεχνολογία περιλαμβάνει πολλαπλούς τομείς και αρχές όπως δικτύωση, κοινή χρήση πόρων, βάσεις δεδομένων, εικονικοποίηση, λειτουργικά συστήματα κ.λπ., επομένως ζητήματα ασφάλειας αυτών των συστημάτων και τεχνολογιών ισχύουν για το cloud computing.

Ένα από τα κύρια ζητήματα με το cloud είναι η ασφάλεια των δεδομένων αποθήκευσης. Στο εξής, οι πάροχοι υπηρεσιών cloud έχουν προτείνει ασφαλείς τρόπους για την κοινή χρήση Big Data στην πλατφόρμα cloud (Agrawal, Das, & El Abbadi, 2011). Αυτοί οι πάροχοι διαβεβαιώνουν ότι οι πελάτες τους δεν αντιμετωπίζουν ζητήματα όπως απώλεια δεδομένων ή κλοπή, που προκαλούνται από πλαστοπροσωπία χρήστη (IDC, 2012). Εδώ έχουμε διαχωρίσει τις προκλήσεις για την ασφάλεια του cloud σε τρεις κατηγορίες, συγκεκριμένα ως επίπεδο δικτύου, επίπεδο ελέγχου ταυτότητας χρήστη και ζητήματα επιπέδου δεδομένων.

1. Επίπεδο δικτύου: Τα ζητήματα πρωτοκόλλου και ασφάλειας σε επίπεδο δικτύου, περιλαμβάνουν συνήθως τομείς όπως οι επικοινωνίες μεταξύ των κόμβων και οι κατανεμημένοι κόμβοι και δεδομένα. Ως εκ τούτου, συνιστάται η κρυπτογράφηση όλων των επικοινωνιών δικτύου από το Secure Sockets Layer (SSL) για την εγγύηση της ασφάλειας των πακέτων και τη διασφάλιση ότι μπορούν να προκύψουν χρήσιμες πληροφορίες ακόμη και αν ένας μη εξουσιοδοτημένος χρήστης αποκτήσει πρόσβαση στις επικοινωνίες δικτύου.

2. Επίπεδο ελέγχου ταυτότητας: Τα ζητήματα ασφαλείας επιπέδου ελέγχου ταυτότητας χρήστη, περιλαμβάνουν συνήθως τομείς όπως μεθόδους ελέγχου ταυτότητας, όπως καταγραφή, δικαιώματα διαχείρισης κόμβων, έλεγχος ταυτότητας εφαρμογών και τεχνικές που χρησιμοποιούνται για κρυπτογράφηση/αποκρυπτογράφηση. Για την αντιμετώπιση αυτών των προβλημάτων, είναι σημαντικό να καταγράφονται πάντα οι δραστηριότητες τροποποίησης δεδομένων που εκτελούνται από τους χρήστες και να ελέγχονται τακτικά για να διαπιστώνεται εάν τα δεδομένα έχουν παραποιηθεί. Επιπλέον, είναι σημαντικό να επικυρώσετε την αυθεντικότητα των κόμβων χρησιμοποιώντας τεχνολογίες όπως το Kerberos πριν από την ένταξη σε ένα σύμπλεγμα και, ως δευτερεύον μέτρο, να ορίσει κανείς ορισμένους κόμβους εντός συστάδων για να *παγιδεύσουν* τους χάκερ σε περίπτωση που κατάφεραν να *περάσουν* τον έλεγχο ταυτότητας.

3. Επίπεδο δεδομένων: Τα ζητήματα ασφάλειας σε επίπεδο δεδομένων συνήθως περιλαμβάνουν τομείς όπως η κατανεμημένη προστασία δεδομένων για τη διασφάλιση της διαθεσιμότητας και της ακεραιότητας των δεδομένων. Είναι σημαντικό να έχει κανείς πάντα τουλάχιστον τρεις διαφορετικούς Back-Up διακομιστές, έτοιμους να συνδεθούν στο διαδίκτυο, σε περίπτωση που ο κύριος διακομιστής καταστεί μη διαθέσιμος λόγω τεχνικών προβλημάτων, επιθέσεων ή φυσικών καταστροφών. Επιπλέον, τα διανεμημένα δεδομένα θα πρέπει πάντα να διατηρούνται σε κρυπτογραφημένη και συμπιεσμένη μορφή για να αποφευχθούν ζητήματα ασφαλείας.

Το σχήμα κρυπτογραφίας που χρησιμοποιείται εδώ, *εκμεταλλεύεται* την εικονική χαρτογράφηση για να χωρίσει τα δεδομένα σε διαφορετικά μέρη και να τα τοποθετήσει σε πολλαπλούς αποθηκευτικούς χώρους ώστε να είναι αδύνατο για τους χάκερ να αποκτήσουν πλήρη πρόσβαση σε αυτά.

2. Παρακολούθηση και Έλεγχος

Η παρακολούθηση και ο έλεγχος αποτελούν αναπόσπαστο μέρος της διαχείρισης ασφάλειας δικτύου. που βοηθά τους παρόχους υπηρεσιών να αποτρέπουν παραβιάσεις ασφάλειας απλώς ελέγχοντας την κυκλοφορία του δικτύου και χρησιμοποιώντας τις πληροφορίες που αποκτήθηκαν για την προσαρμογή ή την εφαρμογή ορισμένων μέτρων ασφαλείας. Ενώ η παρακολούθηση του δικτύου επικεντρώνεται κυρίως στη συλλογή και τη μελέτη γεγονότων για την πρόβλεψη/ανίχνευση εισβολών. Ο έλεγχος δικτύου θεωρείται ως μια συστηματική και μετρήσιμη πολιτική ασφάλειας, η οποία έχει μεγάλο αντίκτυπο στην ασφάλεια του δικτύου.

1. Παρακολούθηση Δικτύου: Υπάρχουν παράγοντες που πρέπει να αναλυθούν εάν πρόκειται να εφαρμοστεί μια *Αρχιτεκτονική Ανίχνευσης και Πρόληψης Εισβολής*, παρακολουθώντας με επιτυχία ολόκληρη την κίνηση του δικτύου, ορισμένες από τις περιοχές που πρέπει να συμπεριληφθούν στη διαδικασία παρακολούθησης περιλαμβάνουν την κυκλοφορία HTTP και DNS, εγγραφές της ροής IP και τα δεδομένα που συγκεντρώθηκαν από τα honeypots τοποθετήθηκαν ως παγίδα για τους εισβολείς. Το προτεινόμενο σύστημα ανίχνευσης εισβολής (IDS) αποθηκεύει και επεξεργάζεται δεδομένα χρησιμοποιώντας τρεις μετρήσεις κακόβουλης πιθανότητας για εύρεση

Ο εντοπισμός μη φυσιολογικής συμπεριφοράς χρήστη και *ύποπτης* συμπεριφοράς δεδομένων, είναι ένας άλλος σημαντικός τομέας εντός του δικτύου για τον οποίο έχει παρουσιαστεί ένα Σύστημα Αυτοδιασφάλισης. Αυτό το σύστημα διατηρεί μια βιβλιοθήκη λέξεων-κλειδιών που σχετίζονται με μη αξιόπιστη συμπεριφορά και δημιουργεί ένα χαμηλό κρίσιμο αρχείο καταγραφής πληροφοριών αναγνώρισης των χρηστών rmation που εκτελεί μια ύποπτη ενέργεια βάσει της οποίας παράγεται ένα δεύτερο αρχείο καταγραφής (υψηλό κρίσιμο ημερολόγιο) ελέγχοντας τη συχνότητα εμφάνισης των χαμηλών κρίσιμων αρχείων καταγραφής και αποφασίζοντας εάν έχει φτάσει στο μέγιστο όριο. Στο τελευταίο βήμα, το σύστημα αυτοδιασφάλισης αποκλείει την πρόσβαση ύποπτων χρηστών στο δίκτυο.

Λόγω της ετερογενούς φύσης των Μεγάλων Δεδομένων, είναι σημαντικό να αναπτυχθεί ένα σύστημα παρακολούθησης ικανό να ανιχνεύει ανωμαλίες μέσα από τη ροή δεδομένων, απλώς με τη συλλογή των αρχείων καταγραφής του δικτύου, την ταξινόμηση/φιλτράρισμα αυτών και, τέλος, την ανάλυση και την απόκτηση νοήματος/συσχέτισης αυτά τα δεδομένα βάσει των οποίων παράγονται τα απαραίτητα στατιστικά στοιχεία και σχηματίζονται σωστές προβλέψεις σε σχέση με τη συμπεριφορά και τα συμβάντα του δικτύου.

2. Έλεγχος δικτύου. Όλοι γνωρίζουν ότι τα Μεγάλα Δεδομένα (με τα ιδιαίτερα χαρακτηριστικά τους) έχουν επηρεάσει βαθιά τον τρόπο με τον οποίο γίνεται η ανάλυση δεδομένων γενικά. Ωστόσο, υπάρχουν ορισμένες προκλήσεις όσον αφορά τον έλεγχο δεδομένων όσον αφορά την ακεραιότητα και τη διαθεσιμότητα των μεγάλων δεδομένων. Ενώ η διαθεσιμότητα των Μεγάλων Δεδομένων, μπορεί εύκολα να επιτευχθεί με τη διατήρηση ενός αριθμού αντιγράφων για την εξασφάλιση εύκολης και γρήγορης πρόσβασης, μπορεί να προκαλέσει ορισμένα προβλήματα με την ακεραιότητα των δεδομένων, ειδικά επειδή τα έξοδα για την επαλήθευση ενημέρωσης των δυναμικών συνόλων δεδομένων είναι τεράστια και δεν υπάρχει κανένα σχέδιο ακεραιότητας για ταυτόχρονο έλεγχο και έλεγχος ταυτότητας των δεικτών μπλοκ.

4. Διαχείριση κλειδιών

Η βελτίωση της ασφάλειας και του απορρήτου των Μεγάλων Δεδομένων, συνοδεύεται από διαφορετικό αριθμό προκλήσεων, ειδικά επειδή η δυναμική δημιουργία κλειδιών για τα Μεγάλα Δεδομένα δεν είναι αποτελεσματική, χρησιμοποιώντας τις τρέχουσες κρυπτογραφικές τεχνικές. Στη σημερινή κοινωνία της κινητής τηλεφωνίας, οι χρήστες των κέντρων δεδομένων μπορούν να βρίσκονται οπουδήποτε, γεγονός που καθιστά απαραίτητο να υπάρχει ένα συγκεκριμένο σύστημα διαχείρισης κλειδιών για την ταυτόχρονη ασφάλεια των δεδομένων και του καναλιού που χρησιμοποιούνται για μετάδοση μεταξύ κόμβων. Για την επίλυση αυτού του προβλήματος, τα κέντρα δεδομένων χρειάζονται αποτελεσματική Κβαντική Κρυπτογραφία, χρησιμοποιώντας τον αλγόριθμο του Grover για κατάλληλες προσεγγίσεις ελέγχου ταυτότητας για τη βελτίωση της ασφάλειας και της ιδιωτικής ζωής με λιγότερη πολυπλοκότητα σε κινητά ή σταθερά κέντρα δεδομένων.

Παρόλο που η χρήση του Quantum μοντέλου μπορεί να αυξήσει την αποτελεσματικότητα του συστήματος (μειώνοντας τον αριθμό των βασικών λειτουργιών αναζήτησης) και την ασφάλεια (μείωση του αριθμού των επιθέσεων). Είναι σημαντικό να θυμάται κανείς ότι οι επικοινωνίες μεγάλων δεδομένων απαιτούν πρωτόκολλα ασφαλούς μεταφοράς κλειδιών ομάδας για να αντέχουν σε επιθέσεις. Ως εκ τούτου, προτείνεται η χρήση ενός διαδικτυακού κέντρου παραγωγής κλειδιών που βασίζεται στη συμφωνία κλειδιού Diffie-Hellman.

Τα Μεγάλα Δεδομένα μπορούν γενικά να χωριστούν σε δύο γενικούς τύπους, δηλαδή ως δομημένα και μη δομημένα, όπου είναι πολύ πιο δύσκολο να διασφαλιστεί η ασφάλεια για μη δομημένα δεδομένα. Εδώ θα προτείναμε τι μπορεί να χρησιμοποιηθεί για να εγγυηθεί την ασφάλεια των μη δομημένων δεδομένων. Σε αυτή την προσέγγιση, τα δεδομένα εξετάζονται, φιλτράρονται, ομαδοποιούνται και τελικά ταξινομούνται με βάση τον τύπο και το επίπεδο ευαισθησίας τους, στη συνέχεια δημιουργούνται συγκεκριμένοι κόμβοι δεδομένων στη βάση δεδομένων.

Για την παροχή ασφάλειας στους κόμβους δεδομένων, σχεδιάστηκε μια *σουίτα* ασφαλείας που ενσωματώνει διαφορετικά πρότυπα ασφαλείας και αλγόριθμους σύμφωνα με τον τύπο του κόμβου δεδομένων. Σε αυτό το στάδιο, ο καταλληλότερος αλγόριθμος εκχωρείται στον κόμβο δεδομένων με βάση τον τύπο δεδομένων/τις απαιτήσεις του (εμπιστευτικότητα, έλεγχος ταυτότητας και ακεραιότητα) και το επίπεδο ευαισθησίας (ευαίσθητο, εμπιστευτικό, δημόσιο) από τη σουίτα ασφαλείας.

5. Ασφάλεια Δεδομένων (Ανωνυμοποίηση)

Η συλλογή δεδομένων για ανάλυση, έχει προκαλέσει πάρα πολλά προβλήματα όσον αφορά το δικαίωμα του χρήστη στην ιδιωτική ζωή. Μία από τις πιο σημαντικές αρμοδιότητες των δεδομένων είναι να διασφαλίζουν την ασφάλεια και το απόρρητο των δεδομένων. παρόλο που αυτό μπορεί να αποδειχθεί ανέφικτο σε ορισμένες χρονικές στιγμές. Το Privacy Preserving Data Publishing (PPDP) εξετάζει τους τρόπους με τους οποίους μπορούν να δημοσιευτούν τα δεδομένα διασφαλίζοντας παράλληλα το δικαίωμα των χρηστών στην ιδιωτικότητα.

Παρόλο που γίνεται όλο και πιο δύσκολο για τους εκδότες δεδομένων να κρύψουν τις Προσωπικές Αναγνωριστικές Πληροφορίες (PII) λόγω της ταχύτητας με την οποία

κοινοποιούνται τα δεδομένα, υπάρχει απόλυτη ανάγκη να σχεδιαστούν πολιτικές στις οποίες οι εταιρείες θα λογοδοτούν για να διασφαλιστεί η ανωνυμοποίηση (αποπροσδιορίστηκε) και ασφαλής μεταφορά των προσωπικών δεδομένων των χρηστών.

Δυστυχώς και μετά κατά την εκτέλεση της διαδικασίας ανωνυμοποίησης, υπάρχουν τρόποι (χρήση ισχυρών αλγορίθμων και ανάλυση τεχνητής νοημοσύνης) με τους οποίους μπορούν να επαναπροσδιοριστούν οι χρήστες. Για να αποφευχθεί αυτό το ζήτημα, χρησιμοποιήθηκαν μετρήσεις που βασίζονται στην ανωνυμία για την απόκρυψη ευαίσθητων πεδίων. Εδώ τα προσωπικά αναγνωριστικά αφαιρούνται από τα αρχεία καταγραφής χρήσης για την προστασία του απορρήτου των χρηστών. Η ανωνυμοποίηση ευαίσθητων πεδίων επιτυγχάνεται με τη χρήση συμμετρικής κρυπτογράφησης κλειδιού AES, η οποία αποθηκεύεται σε HDFS για ανάλυση. Σε περιπτώσεις όπου απαιτείται εκ νέου αναγνώριση των δεδομένων, τα αποθηκευμένα αρχεία καταγραφής μετακινούνται και αποκρυπτογραφούνται χρησιμοποιώντας το κλειδί (IDC, 2012).
Η Εξόρυξη Δεδομένων Διατήρησης Απορρήτου (PPDM) είναι ένα άλλο θέμα που έχει κερδίσει έλξη λόγω της αυξημένης χρήσης των αναλυτικών στοιχείων και των ανησυχιών περί απορρήτου. Είναι σημαντικό να αποκτήσετε απόρρητο χωρίς να διακυβεύεται το περιεχόμενο δεδομένων ή η ακρίβεια εξόρυξης. Επομένως, εδώ συμβουλεύουμε τη χρήση ενός αλγορίθμου με το όνομα Adaptive Utility based Anonymization model (AUA) για την αντιμετώπιση του κινδύνου αποκάλυψης δεδομένων χωρίς να επηρεάζεται η ακρίβεια ταξινόμησης (Agrawal, Das, & El Abbadi, 2011).

Η επεκτασιμότητα και ο τεράστιος όγκος είναι ένας άλλος λόγος που οι συνήθεις μέθοδοι ανωνυμοποίησης είναι ανεπιτυχείς στην απόκρυψη ευαίσθητων πληροφοριών όταν πρόκειται για μεγάλα δεδομένα. Ως εκ τούτου, εδώ συμβουλεύουμε τη χρήση ενός υβριδικού μοντέλου ανωνυμοποίησης υποδέντρων από πάνω-κάτω και από κάτω-επάνω για να αυξηθεί η ικανότητα επεκτασιμότητας της μεθόδου (IDC, 2012).

Ωστόσο, με βάση μια διαφορετική μελέτη, υπάρχει μια διαφορετική επιλογή για την αύξηση της επεκτασιμότητας που ονομάζεται αλγόριθμος ομαδοποίησης δύο φάσεων. Αυτή η προσέγγιση περιλαμβάνει έναν αλγόριθμο ομαδοποίησης και έναν αλγόριθμο συγκεντρωτικής ομαδοποίησης με επίγνωση της εγγύτητας που έχει

σχεδιαστεί με το MapReduce για να αποκτήσει μεγαλύτερη επεκτασιμότητα. Αυτή η προσέγγιση βελτιώνει την επεκτασιμότητα και τη χρονική απόδοση της ανωνυμοποίησης τοπικής επανακωδικοποίησης, επιπλέον της ικανότητας άμυνας έναντι παραβιάσεων της ιδιωτικής ζωής εγγύτητας (IDC, 2012).

ΚΕΦΑΛΑΙΟ 2^{Ο} – BIG DATA SECURITY

2.1 Πρόλογος

Αποτελεί γεγονός στις μέρες μας πως ένας μεγάλος αριθμός επιχειρήσεων, χρησιμοποιούν τα Μεγάλα Δεδομένα για σκοπούς μάρκετινγκ και έρευνας. Ωστόσο, οι περισσότερες επιχειρήσεις από αυτές δεν διαθέτουν θεμελιώδη περιουσιακά στοιχεία όσον αφορά την ασφάλεια. που μπορεί να οδηγήσει σε σοβαρές αγωγές και ζημιά στη φήμη, εάν συμβεί παραβίαση ασφάλειας (IDC, 2012).

Ως εκ τούτου, είναι προφανές ότι οι οργανισμοί απαιτούν νέους μηχανισμούς και κανονισμούς για να εγγυηθούν την ασφάλεια των συστημάτων και των δεδομένων τους, ακόμη και ιδιαίτερα επειδή οι παραδοσιακές τεχνικές είναι αναποτελεσματικές όσον αφορά τις προκλήσεις για την ασφάλεια των μεγάλων δεδομένων και την προστασία της ιδιωτικής ζωής.

Λαμβάνοντας υπόψη όλα τα αναφερόμενα εδώ, είναι ακόμα σημαντικό να κατανοήσουμε ότι ο ανοιχτός κώδικας ή οι πιο πρόσφατες τεχνολογίες μπορεί να έχουν τα δικά τους μειονεκτήματα, όπως η δημιουργία μιας πίσω πόρτας ή προεπιλεγμένων διαπιστευτηρίων, γεγονός που καθιστά απαραίτητο να εξετάζεται προσεκτικά και να διασφαλίζεται ότι η διαθεσιμότητα, η ακεραιότητα και το απόρρητο των δεδομένων παραμένουν ανέπαφα πριν από τη χρήση οποιουδήποτε προϊόντος. Υπάρχουν διάφορες τεχνικές που χρησιμοποιούνται για το σκοπό αυτό όπως η κρυπτογράφηση, η καταγραφή και η ανίχνευση δεδομένων (Agrawal, Das, & El Abbadi, 2011).

Το φαινόμενο Big Data δεν αντιμετωπίζει μόνο προκλήσεις ασφάλειας αλλά και προβλήματα απορρήτου δεδομένων. Αυτές τις μέρες πολλές εταιρείες παλεύουν με προκλήσεις και υποχρεώσεις σχετικά με το απόρρητο. Ωστόσο, σε αντίθεση με την ασφάλεια, το απόρρητο θεωρείται περιουσιακό στοιχείο, γεγονός που το καθιστά σημείο πώλησης τόσο για τους πελάτες όσο και για τους ενδιαφερόμενους. Η ευρεία χρήση των τεχνολογιών των Μεγάλων Δεδομένων, είχε ως αποτέλεσμα την αποθήκευση και ανάλυση petabytes δεδομένων, καθιστώντας την ταξινόμηση των πληροφοριών ακόμη πιο κρίσιμη από πριν (IDC, 2012).

Το καλό είναι ότι τα Big Data analytics (χρησιμοποιώντας πιο εξελιγμένη ανάλυση προτύπων και ανάλυση πολλαπλών δεδομένων) μπορούν να βοηθήσουν

τους οργανισμούς με τον εντοπισμό σε πρώιμο στάδιο και την πρόληψη προηγμένων απειλών και κακόβουλων εισβολέων.

Με βάση τα τελευταία νέα, η Εθνική Υπηρεσία Ασφαλείας των Ηνωμένων Πολιτειών (NSA), συλλέγει με συνέπεια προσωπικά δεδομένα ατόμων από βάσεις δεδομένων μεγάλων εταιρειών που δραστηριοποιούνται είτε στο διαδίκτυο είτε στον τομέα των τηλεπικοινωνιών, παραβιάζοντας το απόρρητο των ανθρώπων στο όνομα της προστασίας των πολιτών των ΗΠΑ. Για την αντιμετώπιση τέτοιων πολύπλοκων προκλήσεων, υπάρχει απόλυτη ανάγκη οι νόμοι και οι κανονισμοί να επιβάλλουν ξεκάθαρα όρια όσον αφορά τη μη εξουσιοδοτημένη πρόσβαση, την κοινή χρήση δεδομένων και την κακή χρήση των χρηστών. προσωπικά δεδομένα (IDC, 2012).

Με βάση μια μελέτη που πραγματοποιήθηκε από την Cloud Security Alliance, οι προκλήσεις ασφάλειας και απορρήτου στα Big Data χωρίζονται σε τέσσερις κατηγορίες και συγκεκριμένα ως:

1) Ασφάλεια υποδομής,
2) Απόρρητο δεδομένων
3) Διαχείριση δεδομένων και
4) Ακεραιότητα και αντιδραστική ασφάλεια όπως εξηγείται παρακάτω:

1. Η ασφάλεια υποδομής περιλαμβάνει κατανεμημένο προγραμματισμό, κόμβους, δεδομένα, επικοινωνία μεταξύ των κόμβων και πρακτικές ασφάλειας για τις μη σχεσιακές αποθήκες δεδομένων.
2. Το απόρρητο δεδομένων περιλαμβάνει αναλύσεις δεδομένων διατήρησης απορρήτου, κρυπτογράφηση του κέντρου δεδομένων και έλεγχο πρόσβασης.
3. Η διαχείριση δεδομένων αναφέρεται στην ασφάλεια αποθήκευσης δεδομένων, στις συναλλαγές καταγραφής, στην προέλευση των δεδομένων και στον έλεγχο.
4. Η ακεραιότητα και η αντιδραστική ασφάλεια συνίστανται σε παρακολούθηση δεδομένων και ενεργειών σε πραγματικό χρόνο, φιλτράρισμα και επικύρωση.

Με βάση όλες τις πληροφορίες που αναφέρονται στο πλαίσιο αυτό, είναι απαραίτητο να εφαρμοστούν μηχανισμοί εξουσιοδότησης και επαλήθευσης ταυτότητας για χρήστες και εφαρμογές για τον έλεγχο της πρόσβασης σε ευαίσθητα δεδομένα, καθώς και τεχνικές κρυπτογράφησης και κάλυψης δεδομένων (ανωνυμοποίηση) σε μεταφορές δεδομένων και σύνολα δεδομένων (IDC, 2012).

2.2 Big Data Management Security

2.2.1 Έλεγχος Προέλευσης Δεδομένων (Big Data Infrastructure)

Στον σημερινό κόσμο που βασίζεται στα δεδομένα, οι οργανισμοί συλλέγουν τεράστιες ποσότητες πληροφοριών με πρωτοφανή ρυθμό. Αυτό το φαινόμενο οδήγησε στη γέννηση και την ταχεία εξέλιξη των «μεγάλων δεδομένων», ενός όρου που περιγράφει σύνολα δεδομένων τόσο μεγάλα και πολύπλοκα που δεν μπορούν να υποστούν επεξεργασία χρησιμοποιώντας παραδοσιακά εργαλεία διαχείρισης δεδομένων. Με αυτό το κύμα δεδομένων προκύπτει η ανάγκη για μια ισχυρή υποδομή για την αποτελεσματική αποθήκευση, επεξεργασία και ανάλυση τους (IDC, 2012).

Για να κατανοήσουμε την υποδομή μεγάλων δεδομένων, είναι σημαντικό πρώτα να κατανοήσουμε τι εννοούμε με τον όρο μεγάλα δεδομένα. Τα μεγάλα δεδομένα αναφέρονται σε τεράστιες ποσότητες δομημένων, ημιδομημένων και μη δομημένων δεδομένων, που παράγονται με ταχύτητα από πολλαπλές πηγές και με τεράστια ποικιλία. Αυτά τα δεδομένα είναι πολύ περίπλοκα και μεγάλης κλίμακας για να τα χειρίζεται το παραδοσιακό λογισμικό επεξεργασίας δεδομένων. Για μια εις βάθος ματιά στα μεγάλα δεδομένα και την πολυπλοκότητά τους, συμπεριλαμβανομένων παραδειγμάτων και βέλτιστων πρακτικών, μη διστάσετε να εξερευνήσετε τη λεπτομερή ανάρτηση ιστολογίου μας σχετικά με τα μεγάλα δεδομένα (IDC, 2012).

Η υποδομή μεγάλων δεδομένων είναι ένα θεμέλιο που έχει σχεδιαστεί για τη διαχείριση, αποθήκευση και ανάλυση αυτού του τεράστιου όγκου δεδομένων. Περιλαμβάνει εξαιρετικά επεκτάσιμους πόρους αποθήκευσης, προηγμένες δυνατότητες επεξεργασίας δεδομένων και εξελιγμένα εργαλεία ανάλυσης. Η υποδομή επιτρέπει στους οργανισμούς να συλλαμβάνουν, να επιμελούνται, να διαχειρίζονται και να επεξεργάζονται δεδομένα μέσα σε ένα ανεκτό χρονικό διάστημα που έχει παρέλθει για να αντλήσουν πολύτιμες γνώσεις για τη λήψη αποφάσεων (IDC, 2012).

Αυτή η υποδομή δεν είναι απλώς μια ενιαία τεχνολογία ή εργαλείο, αλλά ένα περίπλοκο και εξελισσόμενο οικοσύστημα που περιλαμβάνει στοιχεία υλικού όπως διακομιστές και συστήματα αποθήκευσης, λογισμικό για διαχείριση και ανάλυση δεδομένων, λύσεις δικτύωσης για τη διαχείριση της μεταφοράς δεδομένων και πόρους cloud που προσφέρουν επεκτασιμότητα και ευελιξία. Επιπλέον, η υποδομή μεγάλων δεδομένων πρέπει να λάβει υπόψη τα πρωτόκολλα ασφαλείας για την προστασία της ακεραιότητας των δεδομένων και τη συμμόρφωση με τους κανονισμούς απορρήτου.

Αξιοποιώντας τη σωστή υποδομή μεγάλων δεδομένων, οι επιχειρήσεις και οι οργανισμοί μπορούν να επεξεργάζονται μεγάλους όγκους δεδομένων γρήγορα, αποκτώντας τις γνώσεις που χρειάζονται για την προώθηση της καινοτομίας, της αποτελεσματικότητας και του ανταγωνιστικού πλεονεκτήματος.
Η κατανόηση των θεμελιωδών στοιχείων της υποδομής μεγάλων δεδομένων είναι απαραίτητη για την κατανόηση του τρόπου με τον οποίο υποστηρίζει το χειρισμό μαζικών και πολύπλοκων συνόλων δεδομένων. Θα περιγράψουμε τα βασικά στοιχεία και τις αρχές που στηρίζουν αποτελεσματικά συστήματα διαχείρισης μεγάλων δεδομένων.

2.2.2 Αρχιτεκτονική Μαζικών Δεδομένων

Αρχιτεκτονικά, ένα σύστημα λειτουργίας των Μεγάλων Δεδομένων, μπορεί να χωριστεί σε δύο μέρη, τις τεχνολογίες back-end αποθήκευσης δεδομένων και τεχνολογίες front-end πρόσβασης δεδομένων, συμπεριλαμβανομένων εργαλείων ανάλυσης δεδομένων, αναφοράς και διανομής (Popovic et al. 2010). Πρώτον, οι δραστηριότητες αποθήκευσης μεγάλων δεδομένων ενσωματώνουν και επικυρώνουν τα δεδομένα που εξάγονται, μετασχηματίζονται και φορτώνονται (δηλαδή μέσω της διαδικασίας ETL) από λειτουργικά συστήματα σύμφωνα με μια καθορισμένη λογική και συλλέγουν σχετικά δεδομένα σε ένα αποθετήριο (Hovi et al. 2009, Herschel, Jones 2005).

Οι Sen και Sinha (2005) ορίζουν ότι μια *αποθήκη* λειτουργίας των Μεγάλων Δεδομένων, είναι «*μια συλλογή δεδομένων προσανατολισμένη στο θέμα, ολοκληρωμένη, χρονικά μεταβλητή και μη ασταθής που υποστηρίζει τη λήψη διοικητικών αποφάσεων*». Επιτρέπει στα ανόμοια λειτουργικά συστήματα να επικοινωνούν μεταξύ τους (Isik et al. 2011) και παρέχει μια ομοιόμορφη και ολοκληρωμένη άποψη για τα οργανωτικά δεδομένα (Popovic et al. 2010).

Δεύτερον, οι τεχνολογίες front-end παρέχουν στους επιχειρησιακούς χρήστες πρόσβαση στην αποθήκη δεδομένων μέσω εργαλείων όπως πίνακες εργαλείων, πελάτες ερωτημάτων SQL και σχεσιακούς κύβους (Hovi et al. 2009).

Στο πλαίσιο αυτό, οι Hovi et al. (2009) περιορίζουν τον ορισμό της λειτουργίας των Μεγάλων Δεδομένων στην πρόσβαση και ανάλυση δεδομένων και συζητούν την αποθήκευση δεδομένων ως ξεχωριστή δραστηριότητα που περιλαμβάνει τη διαδικασία ETL και το σχεδιασμό και την υλοποίηση της αποθήκης δεδομένων. Σύμφωνα με αυτόν τον ορισμό, μια αποθήκη δεδομένων και ένα

διαφορετικό λογισμικό της σύστημα λειτουργίας των Μεγάλων Δεδομένων, είναι μόνο η τεχνολογία που εκτελεί τη διαδικασία, επομένως λαμβάνονται υπόψη τόσο οι οργανωτικές όσο και οι τεχνολογικές πτυχές του όρου. Αυτό είναι σύμφωνο με τον ορισμό του BI που χρησιμοποιείται γενικά στη Φινλανδία (Pirttimäki 2007).

Ο Greene (1966) όρισε επίσης το τελικό προϊόν της σύστημα λειτουργίας των Μεγάλων Δεδομένων ως «*επεξεργασμένες πληροφορίες που ενδιαφέρουν τη διοίκηση σχετικά με το παρόν και το μελλοντικό περιβάλλον στο οποίο λειτουργεί η επιχείρηση*». Αυτός ο ορισμός έχει δύο σημαντικές συνέπειες. Οι αναφερόμενες πληροφορίες υποβάλλονται σε επεξεργασία, κάτι που συνεπάγεται αναλυόμενες και στοχευμένες πληροφορίες από τον χρήστη, όχι μόνο οποιαδήποτε μαζικά δεδομένα. Η άλλη συνέπεια είναι ότι η διοίκηση, στην οποία αναφέρονται οι πληροφορίες, εμπλέκεται ουσιαστικά στην εφαρμογή της λειτουργίας των Μεγάλων Δεδομένων.

Η διαχείριση περιγράφει τις προδιαγραφές του προϊόντος των Μεγάλων Δεδομένων και αποφασίζει τι εμπίπτει στο πεδίο εφαρμογής. Η διοίκηση είτε καθορίζει είτε παραβιάζει τον σκοπό της λειτουργίας των Μεγάλων Δεδομένων, καθώς η εμπλοκή τους θα επιλύσει εάν οι συλλεγόμενες πληροφορίες είναι σχετικές πληροφορίες ή μόνο περιττές σειρές δεδομένων (Popovic et al. 2010).

Συνοπτικά λοιπόν, η εφαρμογή της λειτουργίας των Μεγάλων Δεδομένων, είναι μια φιλοσοφία διαχείρισης, η οποία συχνά αναφέρεται ως οι τεχνολογίες, οι διαδικασίες, οι πρακτικές και οι λειτουργίες που αναλύουν κρίσιμα επιχειρηματικά δεδομένα για να βοηθήσουν έναν οργανισμό να αυξήσει την επιχειρηματική του ευαισθητοποίηση, να αποκτήσει μια ολιστική άποψη των δυνατοτήτων και του επιχειρηματικού περιβάλλοντος και επομένως, λαμβάνει κανείς πιο επίκαιρες αποφάσεις (Chen et al. 2012, Kaario, Peltola 2008, Lönnqvist, Pirttimäki 2006). Περιλαμβάνει μεθόδους όπως η διαχείριση της στρατηγικής απόδοσης και η ανταγωνιστική ευφυΐα και παρέχει εργαλεία για οπτικοποίηση και χαρτογράφηση δεδομένων που βοηθούν στην παρουσίαση των αποτελεσμάτων σε μια πλούσια σε πληροφορίες, εφαρμόσιμη μορφή (Smith, Lindsay 2012).

Σύμφωνα με τον Thomas (2001), οι πρωταρχικοί στόχοι των λειτουργιών των Μεγάλων Δεδομένων είναι η αποφυγή εκπλήξεων, ο εντοπισμός απειλών και ευκαιριών, η κατανόηση των ευάλωτων οργανισμών, η μείωση του χρόνου αντίδρασης στις αλλαγές στο λειτουργικό περιβάλλον, η αντιμετώπιση του ανταγωνισμού και η προστασία του πνευματικού κεφαλαίου. Με άλλα λόγια, ο στόχος είναι να παράσχει στην ανώτερη διοίκηση μια κατάσταση υγείας 360° του οργανισμού.

Στο πλαίσιο της επιχειρηματικής αξίας των Μεγάλων Δεδομένων, οι Ramakrishnan et al. (2012) συζητούν τους τρεις γενικούς σκοπούς για τους οποίους εφαρμόζεται ο όρος της λειτουργίας των Μεγάλων Δεδομένων. Πρώτον, ένας οργανισμός θέλει να αποκτήσει διορατικότητα. Η ανταγωνιστική πίεση στην αγορά αυξάνει την αβεβαιότητα και οι συγγραφείς υποστηρίζουν ότι τα συστήματα της λειτουργίας των Μεγάλων Δεδομένων, γίνονται γρήγορα αναγκαιότητα για έναν οργανισμό με σκοπό να μπορεί να αντιμετωπίσει το όλο και πιο δυναμικό επιχειρηματικό περιβάλλον (Ramakrishnan et al. 2012).

Η εφαρμογή της λειτουργίας των Μεγάλων Δεδομένων έχει γίνει η βασική δραστηριότητα που βοηθά τους Διευθυντές Information Officers (CIOs) στην πρόβλεψη της συμπεριφοράς της αγοράς, έτσι ώστε ένας οργανισμός να μπορεί να προσαρμοστεί στις μεταβαλλόμενες επιχειρηματικές συνθήκες (Smith, Lindsay 2012). Ο όρος της λειτουργίας των Μεγάλων Δεδομένων, παρέχει στη διοίκηση καλύτερη κατανόηση σχετικά με τις υποκείμενες τάσεις και εξαρτήσεις που επηρεάζουν το περιβάλλον στο οποίο δραστηριοποιούνται.

Οι άλλοι δύο σκοποί της εφαρμογή της λειτουργίας των Μεγάλων Δεδομένων όπως ορίζονται από τους Ramakrishnan et al. (2012), προσφέρονται σχετίζονται με τη συνοχή των οργανωτικών πληροφοριών. Οι συγγραφείς αναφέρουν ότι η εφαρμογή της λειτουργίας των Μεγάλων Δεδομένων παρέχει σε έναν οργανισμό μια ενιαία εκδοχή αλήθειας και μπορεί επίσης να διευκολύνει τον οργανωτικό μετασχηματισμό. Τα δεδομένα των επιχειρήσεων αλλάζουν συνεχώς, ειδικά καθώς οι εταιρείες διέρχονται από συγχωνεύσεις και εξαγορές.

Οι οργανωτικές αλλαγές φέρνουν νέους καταναλωτές πληροφοριών με πιθανώς ολοκαίνουργιες ανάγκες πληροφόρησης. Η απόκτηση μιας ενιαίας εκδοχής της αλήθειας διευκολύνει την επικοινωνία μεταξύ αυτών των ατόμων όταν όλα έχουν πρόσβαση στις ίδιες πληροφορίες. Η σαφής επιχειρηματική λογική των αριθμών, των υπολογισμών και των όρων βελτιώνει επίσης την ποιότητα των δεδομένων και εξοικονομεί χρόνο για καλύτερη ανάλυση. (Ramakrishnan et al. 2012)

Η εφαρμογή της λειτουργίας των Μεγάλων Δεδομένων βοηθά επίσης τους οργανισμούς να αντιμετωπίσουν τις αυστηρότερες ρυθμιστικές απαιτήσεις (Hervonen 2010). Για παράδειγμα, το ρυθμιστικό πρότυπο της Βασιλείας III για την κεφαλαιακή επάρκεια των τραπεζών, τις προσομοιώσεις ακραίων καταστάσεων και τον κίνδυνο ρευστότητας της αγοράς στον τραπεζικό κλάδο και η οδηγία Solvency II στον ασφαλιστικό κλάδο απαιτούν αυξημένη διαφάνεια στις αναφορές και αυστηρά συντονισμένες διαδικασίες στη διαχείριση πληροφοριών (Hovi et al. 2009).

Οι Kaario και Peltola (2008) διαχωρίζουν τα οφέλη που προκύπτουν από την καλύτερη διαχείριση πληροφοριών σε τρεις κατηγορίες:

1) αυξημένη αποτελεσματικότητα (δηλαδή υψηλότερα επίπεδα αυτοματισμού, ταχύτερη πρόσβαση στις πληροφορίες και καλύτερη χρήση του κεφαλαίου πληροφοριών),
2) βελτιωμένη ποιότητα πληροφοριών και διαχείριση κινδύνου. δηλαδή μείωση σφαλμάτων, καλύτερη συμμόρφωση με τα απαιτούμενα πρότυπα και ασφάλεια συστήματος) και
3) υψηλότερα επίπεδα υπηρεσιών (δηλαδή προστιθέμενη αξία στις υπάρχουσες υπηρεσίες, αυξημένη διαθεσιμότητα και ταχύτερες διαδικασίες).

Τέλος, οι Cooper et al. (2000) προσθέτουν ότι η εφαρμογή της λειτουργίας των Μεγάλων Δεδομένων, επιτρέπει την αποτελεσματικότερη χρήση των πληροφοριών πελατών, τον εντοπισμό των πιο κερδοφόρων τμημάτων πελατών και την ανάπτυξη δομών και στρατηγικών τιμολόγησης για την επέκταση των σχέσεων με τους πελάτες (π.χ. cross-selling).

ΚΕΦΑΛΑΙΟ 3^{Ο} – ΠΡΟΣΤΑΣΙΑ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ (PRIVACY)

3.1 Πρόλογος

Η ιδιωτικότητα μπορεί να θεωρηθεί ο έλεγχος των προσωπικών πληροφοριών ή ακόμη και ως η ελευθερία από την κρίση των άλλων. Σύμφωνα με την Solove, ο όρος «privacy» (ιδιωτικότητα) είναι ένας όρος - ομπρέλα, που αναφέρεται σε μια ευρεία και ανόμοια ομάδα σχετικών στοιχείων και δεν μπορεί να κατανοηθεί ανεξάρτητα από την κοινωνία, καθώς η ιδιωτικότητα, στον πυρήνα της, είναι ένα κοινωνικό τεχνούργημα και χωρίς το πλαίσιο της κοινωνίας, δεν θα υπήρχε ανάγκη για ιδιωτικότητα (Richterich, 2018).

Οι ανησυχίες σχετικά με τη χρήση μεγάλων δεδομένων, πιθανότατα εγείρονται συχνότερα όταν πρόκειται για το απόρρητο και την ασφάλεια των επιχειρήσεων ή των ατόμων. Η ιδιωτικότητα αναφέρεται στην ικανότητα ενός ατόμου να ορίζει και να περιορίζει την πρόσβαση στα προσωπικά του στοιχεία. Αυτό μπορεί να έχει να κάνει με ενέργειες, όπως η κρυφή είσοδος σε ιδιωτικούς χώρους ή με δεδομένα που προέρχονται από τα ψηφιακά αποτυπώματα των ανθρώπων (Richterich, 2018).

Η ιδιωτικότητα των πληροφοριών αναφέρεται ως το δικαίωμα σε κάποιο βαθμό ελέγχου σχετικά με τη συλλογή και τη χρήση των προσωπικών δεδομένων. Το απόρρητο των πληροφοριών, είναι η ικανότητα ενός ατόμου ή μιας ομάδας να αποτρέπει την αποκάλυψη πληροφοριών για τον εαυτό του σε άτομα ή ομάδες εκτός αυτών στους οποίους αποκαλύπτονται οι πληροφορίες (Jain, 2016). Η αναγνώριση προσωπικών πληροφοριών κατά τη μετάδοση μέσω Διαδικτύου, είναι ένα σημαντικό πρόβλημα για το απόρρητο των χρηστών (Jain, 2016). Από την άλλη πλευρά, σύμφωνα με τους Belanger και Crossler, «*το απόρρητο των πληροφοριών είναι ένα σύνολο της συνολικής έννοιας της ιδιωτικής ζωής, η οποία έχει διερευνηθεί και συζητηθεί για αιώνες*» (Belanger & Crossler, 2011).

3.2 Ορισμός Ιδιωτικότητας (Privacy)

Σύμφωνα με τον Skinner, οι περισσότερες ερμηνείες του όρου «ιδιωτικότητα» αφορούν ένα ανθρώπινο δικαίωμα. Αυτές οι συνθήκες ενέπνευσαν τον Clarke (1999) να ονομάσει τέσσερις τύπους απορρήτου, το απόρρητο ενός ατόμου, το απόρρητο

της προσωπικής του συμπεριφοράς, η ιδιωτικότητα των προσωπικών του επικοινωνιών και το απόρρητο των προσωπικών του δεδομένων. Το απόρρητο της προσωπικής επικοινωνίας και το απόρρητο των δεδομένων μπορούν πλέον να συνδυαστούν στην έννοια του απορρήτου των πληροφοριών καθώς η πλειοψηφία των επικοινωνιών είναι πλέον ψηφιακές και αποθηκεύονται ως πληροφορίες (Belanger & Crossler, 2011).

Αν ρίξουμε μια πιο βαθιά ματιά, η ιδιωτικότητα των πληροφοριών ορίζεται με διάφορους τρόπους, αλλά υπάρχει μικρή διαφορά μεταξύ των βασικών στοιχείων των ορισμών, τα οποία συχνά περιλαμβάνουν κάποιο είδος ελέγχου στις πιθανές δευτερεύουσες χρήσεις των προσωπικών πληροφοριών ενός ατόμου. Η διαδικασία χρήσης δεδομένων για λόγους άλλους από αυτούς για τους οποίους αποκτήθηκαν αρχικά είναι γνωστή ως δευτερεύουσα χρήση. Οι τέσσερις πτυχές του απορρήτου των πληροφοριών που εντόπισε ο Smith (1996), περιλαμβάνουν τη συλλογή, μη εξουσιοδοτημένη δευτερεύουσα χρήση, ακατάλληλη πρόσβαση και σφάλματα.

Η συλλογή, η επεξεργασία, η διανομή και η εισβολή πληροφοριών περιλαμβάνονται σε μια διαφορετική ταξινόμηση. Οι διαστάσεις του χρόνου, της ύλης και του χώρου της προτεινόμενης ταξινόμησης του απορρήτου πληροφοριών από τον Skinner σε συνεργατικά περιβάλλοντα, αντικατοπτρίζουν τη δομική άποψη του απορρήτου στις πληροφορίες, που περιλαμβάνει ατομικό, ομαδικό και οργανωτικό απόρρητο. Η διάσταση του χώρου αντικατοπτρίζει επίσης τη δομική άποψη της ιδιωτικότητας στις πληροφορίες. Ο Clarke δήλωσε ξεκάθαρα ότι «το ενδιαφέρον που έχει ένα άτομο να ελέγχει, ή τουλάχιστον να επηρεάζει σημαντικά, την επεξεργασία δεδομένων για τον εαυτό του» είναι αυτό που εννοούσε με το απόρρητο των πληροφοριών (Belanger & Crossler, 2011).

Το δικαίωμα στην ιδιωτική ζωή θεωρείται αστικό στις δημοκρατίες. Πολλά εθνικά συντάγματα εγγυώνται ρητά ή έμμεσα το δικαίωμα στην ιδιωτική ζωή. Το δικαίωμα στην ιδιωτική ζωή θεωρείται συχνά ότι διευρύνεται για να συμπεριλάβει την προστασία των προσωπικών δεδομένων. Ωστόσο, ο Χάρτης των Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης αντιμετωπίζει και τα δύο ανεξάρτητα, με το άρθρο 8 να επικεντρώνεται στην προστασία των δεδομένων και το άρθρο 7 να αφορά το σεβασμό της ιδιωτικής και οικογενειακής ζωής του ατόμου (The European Union Agency for Fundamental Rights 2007) (Richterich, 2018).

Οι επικριτές των μεγάλων δεδομένων, έχουν υπογραμμίσει πώς οι άνθρωποι δεν κατανοούν και δεν ελέγχουν τις προσωπικές πληροφορίες που συλλέγονται κατά τη χρήση διαδικτυακών υπηρεσιών, οι οποίες αφορούν το απόρρητο των ανθρώπων. Ενώ οι υποστηρικτές των μεγάλων δεδομένων, ιδιαίτερα οι εταιρικοί πάροχοι

υπηρεσιών, ισχυρίζονται ότι οι πληροφορίες των χρηστών παραμένουν ανώνυμες, οι επικριτές αμφισβητούν εάν είναι ακόμη δυνατό να ανωνυμοποιηθούν δεδομένα με τόσο ευρύ φάσμα χαρακτηριστικών σε τόσο μεγάλη κλίμακα (Richterich, 2018).

Εάν εστιάσουμε στον ορισμό του απορρήτου από άλλους μελετητές, είναι το δικαίωμα σε κάποιο βαθμό ελέγχου σχετικά με τη συλλογή και τη χρήση των προσωπικών σας πληροφοριών. Το απόρρητο των πληροφοριών είναι η ικανότητα ενός ατόμου ή μιας ομάδας να εμποδίζει τις πληροφορίες για τον εαυτό τους να είναι γνωστές σε άλλα άτομα εκτός από αυτά στα οποία αποκαλύπτουν τις πληροφορίες. Ο Jain υποδεικνύει ότι «*ένα σοβαρό ζήτημα απορρήτου των χρηστών είναι η αναγνώριση προσωπικών πληροφοριών κατά τη μετάδοση μέσω Διαδικτύου*» (Jain, 2016).

3.2.1 Προστασία Ιδιωτικότητας Δεδομένων

Υπάρχει μια φυσική σχέση μεταξύ του δικαιώματος στην ιδιωτική ζωή και του δικαιώματος στην προστασία των δεδομένων. Ωστόσο, υπάρχει πολλή συζήτηση στον ακαδημαϊκό κόσμο για το πώς το δικαίωμα στην ιδιωτική ζωή και το δικαίωμα με την προστασία δεδομένων σχετίζονται (Kulhari, 2018). Ο Kulhari επισημαίνει ότι «το δικαίωμα στην ιδιωτική ζωή διαφέρει από το δικαίωμα στην προστασία δεδομένων, επειδή το πρώτο έχει τις ρίζες του στα ανθρώπινα δικαιώματα, ενώ το δεύτερο αντιμετωπίζεται ως οικονομικό ζήτημα είναι τουλάχιστον μια κόκκινη γραμμή.

Το θεμέλιο του δικαιώματος στην προστασία δεδομένων είναι η ιδέα ότι η επεξεργασία δεδομένων πραγματοποιείται χωρίς προγραμματισμό. Ως αποτέλεσμα, το GDPR περιέχει συγκεκριμένους κανονισμούς που αφορούν τα καθήκοντα του υπεύθυνου επεξεργασίας και του εκτελούντος την επεξεργασία δεδομένων. Προτείνεται ότι αυτοί οι περιορισμοί στο δικαίωμα προστασίας δεδομένων και σε ό,τι χαρακτηρίζεται ως νόμιμη επεξεργασία αντιπροσωπεύουν συμβιβασμό μεταξύ πολλών έννομων συμφερόντων (Kulhari, 2018).

Παρόλο που το δικαίωμα στην προστασία δεδομένων συνδέεται στενά με το δικαίωμα στην ιδιωτική ζωή, καθώς και με τα δικαιώματα στην ελευθερία της έκφρασης, σε αποτελεσματικό ένδικο μέσο και δίκαιη δίκη, έχει επίσης ορισμένα μοναδικά χαρακτηριστικά που υποστηρίζουν τον χαρακτηρισμό του ως αυτόνομο δικαίωμα. Όπως αναφέρεται στο άρθρο 8 του Χάρτη των Θεμελιωδών Δικαιωμάτων της ΕΕ του 2012, αυτές οι συνιστώσες περιλαμβάνουν: το δικαίωμα πρόσβασης και διόρθωσης των συλλεγόμενων δεδομένων. έλεγχος από ανεξάρτητη αρχή· και την απαίτηση για δίκαιη επεξεργασία των δεδομένων, για συγκεκριμένους σκοπούς, και

μόνο με τη συγκατάθεση του ενδιαφερόμενου προσώπου ή άλλη νομική αιτιολόγηση (McDermott, 2017).

Ας δούμε μερικές από τις βασικές αρχές που καθοδηγούν το δικαίωμα στην προστασία δεδομένων στην έννομη τάξη της ΕΕ (McDermott, 2017):

- Μυστικότητα: Όπως αναφέρθηκε προηγουμένως, είναι προφανές ότι το δικαίωμα στην προστασία των δεδομένων αποσκοπεί στη διασφάλιση της αξίας της ιδιωτικής ζωής, η οποία αποτελεί από μόνο του βασικό δικαίωμα.

- Αυτονομία: Η αυτονομία του ατόμου είναι μια άλλη κρίσιμη αξία που κατοχυρώνει το δικαίωμα στην προστασία δεδομένων, όπως αποδεικνύεται από τη συνεχή σημασία της ενημερωμένης συναίνεσης στην ευρωπαϊκή πολιτική προστασίας δεδομένων. Τα φυσικά πρόσωπα «θα πρέπει να έχουν τον έλεγχο των προσωπικών τους δεδομένων», σύμφωνα με την αιτιολογική σκέψη 7 του GDPR. Η ιδέα της αξιοπρέπειας συνδέεται αναμφισβήτητα με την έννοια της αυτονομίας και την έμφαση που δίνει στη συναίνεση.

- Διαφάνεια: Σήμερα, η διαφάνεια είναι μια βασική αρχή που κατοχυρώνεται στο Άρθ. 5(1)(α) του GDPR που ορίζει ότι τα προσωπικά δεδομένα πρέπει να «επεξεργάζονται νομίμως, δίκαια και με διαφανή τρόπο σε σχέση με το υποκείμενο των δεδομένων», υποδεικνύοντας έτσι τη στενή σύνδεση μεταξύ διαφάνειας, νομιμότητας και δικαιοσύνης. (Felzmann et al., 2019). Δεδομένων των ζητημάτων που αναφέρθηκαν προηγουμένως με συναίνεση και γνώση, αρκετοί συγγραφείς έχουν αντιμετωπίσει τις ανισότητες που υπάρχουν στον τομέα της προστασίας δεδομένων. Ως απάντηση σε αυτήν την πραγματικότητα, ο GDPR ορίζει τη «συγκατάθεση» ως «οποιαδήποτε ελεύθερα δοθείσα, συγκεκριμένη, ενημερωμένη και σαφή ένδειξη των επιθυμιών του υποκειμένου των δεδομένων με την οποία αυτό ή αυτή, με μια δήλωση ή με μια σαφή θετική ενέργεια, υποδηλώνει συμφωνία με την επεξεργασία προσωπικών δεδομένων που τον αφορούν».

- Μη διάκριση: Αναγνωρίζοντας ότι η συλλογή και η επεξεργασία δεδομένων πρέπει να γίνεται με τρόπο που να αποτρέπει τις διακρίσεις εις βάρος ανθρώπων "*με βάση τη φυλετική ή εθνική καταγωγή, τις πολιτικές πεποιθήσεις, τη θρησκεία ή τις πεποιθήσεις, τη συμμετοχή σε συνδικάτα, τη γενετική ή υγειονομική κατάσταση ή τον σεξουαλικό προσανατολισμό*" την

αρχή της διαφάνειας που χρησιμεύει ως θεμέλιο του GDPR (McDermott, 2017).

Το δικαίωμα στην προστασία δεδομένων αντιμετωπίζει ορισμένα ιδιαίτερα δύσκολα ζητήματα στο τρέχον περιβάλλον των διάχυτων τεχνικών επιτήρησης και της αυξανόμενης έμφασης στη συλλογή και χρήση Μεγάλων Δεδομένων (McDermott, 2017). Ωστόσο, αναφέρονται δύο συγκεκριμένες προκλήσεις. Η πρώτη πρόκληση είναι η ευρεία χρήση «εθελοντών» δεδομένων για το δικαίωμα προστασίας δεδομένων στη σύγχρονη εποχή, ειδικά με τον αυξανόμενο αριθμό φορητών τεχνολογίας και πλατφορμών μέσων κοινωνικής δικτύωσης, παρόλο που οι χρήστες αυτών των συστημάτων ενδέχεται να μην αντιλαμβάνονται ότι παρέχουν δεδομένα σε οι υπόλοιποι. Δεδομένα κοινωνικών δικτύων, συσκευές Internet of Things (IoT) και άλλες μέθοδοι θα μπορούσαν να οδηγήσουν στη συλλογή πληροφοριών σχετικά με το περιβάλλον του χρήστη εκτός από τον μεμονωμένο χρήστη. Δεδομένου ότι τα δεδομένα μπορεί να συλλέγονται για έναν σκοπό και στη συνέχεια να χρησιμοποιηθούν για έναν άλλο, ένα παράδειγμα που βασίζεται μόνο στη συγκατάθεση για την επεξεργασία δεδομένων δεν μπορεί να εγγυηθεί πλήρως ότι θα προστατεύονται (McDermott, 2017).Δεύτερον, η μαζική παρακολούθηση χρησιμοποιείται για την πρόληψη μελλοντικών εγκλημάτων όπως τρομοκρατικές επιθέσεις και κυβερνοεπιθέσεις.

Ως αποτέλεσμα, οι «παρατηρητές» είναι λιγότερο εμφανείς από ό,τι στο παρελθόν, επειδή ένα μεγάλο μέρος της παρακολούθησής τους διεξάγεται μέσω της «παρακολούθησης δεδομένων», η οποία περιλαμβάνει παρακολούθηση επικοινωνιών και διαδικτυακής δραστηριότητας και όχι πραγματικών κινήσεων. Σύμφωνα με τους ερευνητές, οποιαδήποτε «θεωρία των δικαιωμάτων» θα αντιμετωπίσει δυσκολίες σχετικά με τα συμφέροντα που προσδιορίζει ως δικαιώματα και τους όρους με τους οποίους τα προσδιορίζει σε διαφωνίες σχετικά με τη σωστή ισορροπία που πρέπει να επιτευχθεί μεταξύ κάποιου ατομικού συμφέροντος και ορισμένων αντισταθμιστικών κοινωνικών εκτιμήσεων» (McDermott, 2017).

3.3 GDPR Data Privacy – Νομικό Πλαίσιο της Προστασίας της Ιδιωτικότητας

Ο κύριος στόχος του Νέου Κανονισμού GDPR – General Data Protection Regulation, είναι να προστατεύσει τους πολίτες της Ε.Ε. από οργανισμούς που

χρησιμοποιούν παράνομα προσωπικά αναγνωρίσιμα στοιχεία (PII). Οι κυρώσεις για παραβιάσεις δεδομένων, έχουν επίσης αυξηθεί και οι επιχειρήσεις έχουν νέες απαιτήσεις για ειδοποιήσεις παραβίασης δεδομένων. Οι επιχειρήσεις που δεν συμμορφώνονται με το Νέο Κανονισμό GDPR, θα αντιμετωπίσουν κυρώσεις ύψους 20 εκατ. ευρώ ή 4% του συνολικού ετήσιου κύκλου εργασιών τους.

Οι νέοι κανόνες του GDPR, πρέπει επίσης να βοηθήσουν τους επιχειρησιακούς οργανισμούς να προετοιμάσουν τις σωστές πολιτικές και διαδικασίες για την αντιμετώπιση περιστατικών ασφάλειας στον κυβερνοχώρο. Επιπλέον, η εφαρμογή του Νέου Κανονισμού GDPR, θα αλλάξει τον τρόπο με τον οποίο οι οργανισμοί επεξεργάζονται και αποθηκεύουν προσωπικές αναγνωρίσιμες πληροφορίες. Τα δικαιώματα των πολιτών της Ε.Ε., πρόκειται να επεκταθούν και το GDPR ισχύει για όλους τους οργανισμούς που επεξεργάζονται τα προσωπικά αναγνωρίσιμα στοιχεία των κατοίκων της Ε.Ε. (Γενικός κανονισμός της ΕΕ για την προστασία των δεδομένων (GDPR), Duncan 2018).

Επίσης η εφαρμογή του Νέου Κανονισμού GDPR, τυποποιεί την προστασία των προσωπικών αναγνωρίσιμων πληροφοριών σε κάθε ευρωπαϊκή χώρα. Οι οργανισμοί πρέπει να εξετάσουν το τι ακριβώς τα προσωπικά αναγνωρίσιμα στοιχεία, επεξεργάζονται και πώς πρέπει να το προστατεύσουν. Με την εφαρμογή του Νέου Κανονισμού GDPR, υπάρχουν διαφορετικοί ρόλοι για κάθε οργανισμό που είναι υπεύθυνος επεξεργασίας δεδομένων των ατόμων και πελατών του. Ο κάθε ελεγκτής πρέπει να καθορίσει τον τρόπο και τον λόγο για τον οποίο επεξεργάζεται τα προσωπικά αναγνωρίσιμα στοιχεία και επεξεργάζεται τον επεξεργαστή.

Ακόμη και οι επιχειρησιακές οργανώσεις εκτός της Ευρωπαϊκής Ένωσης που δραστηριοποιούνται στην επικράτεια αυτής, πρέπει να εφαρμόζουν τις απαιτήσεις του κανονισμού. Μετά την ημερομηνία λήξης του χρονικού περιθωρίου για την εφαρμογή του Νέου Κανονισμού GDPR, κάθε οργανισμός πρέπει να χειρίζεται τα προσωπικά δεδομένα νόμιμα και με διαφάνεια. Επιπλέον, η επεξεργασία στα προσωπικά αναγνωρίσιμα στοιχεία, πρέπει να έχει πραγματικό σκοπό. Όταν οι προσωπικές αναγνωρίσιμες πληροφορίες δεν απαιτούνται πλέον, οι οργανισμοί θα πρέπει να τις καταργήσουν και ουσιαστικά να τις αποσύρουν από τα αρχεία τους (Curtis 2018).

Το άρθρο 5 του κανονισμού GDPR, περιγράφει τις αρχές στις οποίες πρέπει να δίνουν προσοχή οι επιχειρήσεις κατά την επεξεργασία των προσωπικών στοιχείων (PII).

Παρακάτω αναφέρεται ένας κατάλογος ορισμένων αρχών επεξεργασίας των προσωπικών στοιχείων PII ως εξής (Κανονισμός (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου 2016):

- ✓ Η επεξεργασία στα στοιχεία PII πρέπει να γίνεται με νόμιμη, δίκαιη και διαφανή μέθοδο.
- ✓ Τα στοιχεία PII πρέπει να συλλέγονται για συγκεκριμένους, σαφείς και νόμιμους σκοπούς και όχι για επεξεργασία που δεν είναι συμβατή με αυτούς τους σκοπούς.
- ✓ Τα στοιχεία PII πρέπει να είναι επαρκής, σχετικά και περιορισμένα για τους σκοπούς για τους οποίους είναι επεξεργασία.
- ✓ Τα στοιχεία PII πρέπει να είναι ακριβείς, ενημερωμένα και οι οργανισμοί θα πρέπει να διασφαλίζουν τα δεδομένα δεν είναι ανακριβή.
- ✓ Τα στοιχεία PII πρέπει να διατηρούνται σε μορφή που επιτρέπει την αναγνώριση των υποκειμένων των δεδομένων μόνο για ο χρόνος που είναι απαραίτητος.
- ✓ Τα στοιχεία PII πρέπει να έχουν υποβληθεί σε επεξεργασία με τρόπο που να εξασφαλίζεται και να προστατεύεται μη εξουσιοδοτημένη ή παράνομη μεταποίηση και όχι τυχαία χαμένη, καταστραφεί ή σκάρτος.

Ο κάθε ελεγκτής είναι υπεύθυνος για την απόδειξη της συμμόρφωσης των δεδομένων με τον κανονισμό. Η νομιμότητα, η δικαιοσύνη και η διαφάνεια μπορούν να εξηγηθούν με τον ακόλουθο τρόπο, όπου ένας οργανισμός πρέπει να ενημερώσει ένα άτομο για τις μεθόδους επεξεργασίας δεδομένων. Επιπλέον, πρέπει επίσης να ενημερώσουν το είδος των δεδομένων που υποβάλλονται σε επεξεργασία. Οι μέθοδοι επεξεργασίας πρέπει να ταιριάζουν με μια αναφορά ασφάλειας δεδομένων που προσφέρεται από έναν οργανισμό. Στο σχήμα No.3 παρατίθενται έξι αρχές προστασίας της ιδιωτικής ζωής του GDPR. (Οι έξι αρχές προστασίας προσωπικών δεδομένων του GDPR 2017).

Σχήμα No.1. Αρχές προστασίας προσωπικών δεδομένων του GDPR (Οι έξι αρχές προστασίας προσωπικών δεδομένων του GDPR 2017)

Οι περιορισμοί στόχων, σημαίνουν ότι οι προσωπικές αναγνωρίσιμες πληροφορίες (PII), μπορούν να υποβάλλονται σε επεξεργασία για συγκεκριμένους, σαφείς και νόμιμους σκοπούς. Το υποκείμενο των δεδομένων γνωρίζει τους προαναφερθέντες σκοπούς και οι PII δεν χρησιμοποιείται για περαιτέρω ενέργειες χωρίς τη συγκατάθεση του χρήστη. Συλλέγονται μόνο δεδομένα που είναι απαραίτητα και τίποτα περισσότερο (ελαχιστοποίηση δεδομένων).

Η ακρίβεια σημαίνει ότι οι PII θα πρέπει να ενημερώνονται και να είναι ακριβές. Οι περιορισμοί αποθήκευσης σημαίνουν ότι τα δεδομένα θα αποθηκεύονται μόνο για τον απαιτούμενο χρόνο και όχι πλέον. Όταν δεν υπάρχει πια σκοπός για την αποθήκευση των PII, τα δεδομένα πρέπει να διαγραφούν.

Η ακεραιότητα και η εμπιστευτικότητα σημαίνουν ότι τα PII, θα πρέπει να αντιμετωπίζεται κατά τρόπο που να διασφαλίζεται από την παράνομη επεξεργασία ή την τυχαία καταστροφή ή ζημία. (Οι έξι αρχές προστασίας προσωπικών δεδομένων του GDPR 2017). Το έκτο άρθρο του κανονισμού GDPR, ορίζει τη νομιμότητα της επεξεργασίας, τα παραδείγματα ακολουθούν (Κανονισμός (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου 2016):

- Το υποκείμενο των δεδομένων, έχει δώσει τη συγκατάθεσή του για την επεξεργασία των προσωπικών του προσωπικών στοιχείων για τουλάχιστον έναν ή περισσότερους σκοπούς

- Το υποκείμενο των δεδομένων είναι μέρος μιας σύμβασης που απαιτεί επεξεργασία προσωπικών στοιχείων.
- Τα ζωτικά συμφέροντα των υποκειμένων ή των ζωτικών συμφερόντων άλλου φυσικού προσώπου πρέπει να προστατεύονται.
- Η επεξεργασία είναι απαραίτητη από την άποψη του δημόσιου συμφέροντος ή της δημόσιας αρχής.
- Όταν πρέπει να προστατεύονται τα νόμιμα συμφέροντα του υπεύθυνου επεξεργασίας ή του τρίτου μέρους.

ΚΕΦΑΛΑΙΟ 4^{Ο} – ISO / IEC 20547 – 4:2020

4.1 Πρόλογος

Ο ISO (ο Διεθνής Οργανισμός Τυποποίησης) είναι μια παγκόσμια ομοσπονδία εθνικών φορέων τυποποίησης. Το έργο της προετοιμασίας των Διεθνών Προτύπων, πραγματοποιείται συνήθως μέσω των τεχνικών επιτροπών ISO. Κάθε όργανο μέλος που ενδιαφέρεται για ένα θέμα για το οποίο έχει συσταθεί τεχνική επιτροπή έχει το δικαίωμα να εκπροσωπείται σε αυτήν την επιτροπή. Στις εργασίες συμμετέχουν και διεθνείς οργανισμοί, κυβερνητικοί και μη, σε συνεργασία με το ISO (ISO / IEC 20547/4).

Ο ISO συνεργάζεται στενά με τη Διεθνή Ηλεκτροτεχνική Επιτροπή (IEC) για όλα τα θέματα ηλεκτροτεχνικής τυποποίησης. Οι διαδικασίες που χρησιμοποιούνται για την ανάπτυξη αυτού του εγγράφου και εκείνες που προορίζονται για την περαιτέρω συντήρησή του περιγράφονται στις Οδηγίες ISO/IEC ως εξής (ISO / IEC 20547/4).

Μέρος 1. Ειδικότερα, θα πρέπει να σημειωθούν τα διαφορετικά κριτήρια έγκρισης που απαιτούνται για τους διαφορετικούς τύπους εγγράφων ISO.

Μέρος 2. Τεχνολογία πληροφοριών - Αρχιτεκτονική αναφοράς μεγάλων δεδομένων - Μέρος 4: Ασφάλεια και απόρρητο.

Το NEN-ISO-IEC 20547-4 καθορίζει τις πτυχές ασφάλειας και απορρήτου που ισχύουν για την αρχιτεκτονική αναφοράς μεγάλων δεδομένων (BDRA), συμπεριλαμβανομένων των ρόλων, των δραστηριοτήτων και των λειτουργικών στοιχείων μεγάλων δεδομένων και παρέχει επίσης καθοδήγηση σχετικά με τις λειτουργίες ασφάλειας και απορρήτου για μεγάλα δεδομένα. Ως εκ τούτου, αναφέρονται στοιχεία πιστοποιήσεων, ως εξής

ISO/IEC 20546, Τεχνολογία πληροφοριών — Μεγάλα δεδομένα — Επισκόπηση και λεξιλόγιο.
ISO/IEC 20547-3, Τεχνολογία πληροφοριών — Αρχιτεκτονική αναφοράς μεγάλων δεδομένων.

ISO/IEC/IEEE 15288, Μηχανική συστημάτων και λογισμικού — Διαδικασίες κύκλου ζωής συστήματος.

4.2 Ανησυχίες Σχετικά με την Ασφάλεια των Big Data και την Προστασία της Ιδιωτικότητας

Αποτελεί γεγονός πως με τον πολλαπλασιασμό των συσκευών που συνδέονται στο Διαδίκτυο καθώς και μεταξύ τους όπως κι ο όγκος των δεδομένων που συλλέγονται, αποθηκεύονται και επεξεργάζονται, αυξάνεται καθημερινά, γεγονός που φέρνει επίσης νέες προκλήσεις όσον αφορά την ασφάλεια των πληροφοριών.

Στην πραγματικότητα, οι μηχανισμοί ασφαλείας που χρησιμοποιούνται σήμερα, όπως τα *τείχη προστασίας* και τα *DMZ* δεν μπορούν να χρησιμοποιηθούν στην υποδομή των Big Data, επειδή οι μηχανισμοί ασφαλείας θα πρέπει να εκτείνονται εκτός της περιμέτρου του δικτύου του οργανισμού για να πληρούν τις απαιτήσεις κινητικότητας χρήστη/δεδομένων (ISO / IEC 20547/4).

Λαμβάνοντας υπόψη αυτά τα νέα στοιχεία, το σχετικό ερώτημα είναι ποιες πολιτικές και τεχνολογίες ασφάλειας και απορρήτου, είναι πιο κατάλληλες για την εκπλήρωση των σημερινών κορυφαίων απαιτήσεων απορρήτου και ασφάλειας Big Data (Cloud Security Alliance, 2013). Αυτές οι προκλήσεις μπορούν να οργανωθούν σε τέσσερις πτυχές εφαρμογής των Big Data, όπως ασφάλεια υποδομής (π.χ. ασφαλείς κατανεμημένοι υπολογισμοί χρησιμοποιώντας MapReduce), απόρρητο δεδομένων (π.χ. εξόρυξη δεδομένων που διατηρεί το απόρρητο/συγκεκριμένη πρόσβαση), διαχείριση δεδομένων (π.χ. ασφαλής προέλευση και αποθήκευση δεδομένων) και, ακεραιότητα και αντιδραστική ασφάλεια (π.χ. παρακολούθηση σε πραγματικό χρόνο ανωμαλιών και επιθέσεων).

Λαμβάνοντας υπόψη τα μεγάλα δεδομένα, υπάρχει ένα σύνολο τομέων κινδύνου που πρέπει να ληφθούν υπόψη. Αυτά περιλαμβάνουν τον κύκλο ζωής των πληροφοριών (προέλευση, ιδιοκτησία και ταξινόμηση δεδομένων), τη διαδικασία δημιουργίας και συλλογής δεδομένων και την έλλειψη διαδικασιών ασφαλείας. Σε τελική ανάλυση, οι στόχοι ασφάλειας των μεγάλων δεδομένων δεν διαφέρουν από οποιουσδήποτε άλλους τύπους δεδομένων – για τη διατήρηση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητάς τους.

Όντας τα Big Data ένα τόσο σημαντικό και πολύπλοκο θέμα, είναι σχεδόν φυσικό να προκύψουν τεράστιες προκλήσεις ασφάλειας και ιδιωτικότητας (Michael &

Miller, 2013, Tankard, 2012). Τα Big Data έχουν συγκεκριμένα χαρακτηριστικά που επηρεάζουν την ασφάλεια των πληροφοριών, όπως ποικιλία, όγκος, ταχύτητα, τιμή, μεταβλητότητα και ακρίβεια (Agrawal, Das, & El Abbadi, 2011). Αυτές οι προκλήσεις έχουν άμεσο αντίκτυπο στον σχεδιασμό λύσεων ασφαλείας που απαιτούνται για την αντιμετώπιση όλων αυτών των χαρακτηριστικών και απαιτήσεων (Demchenko, et al., 2014).

Στο πλαίσιο αυτό, η Cloud Secure Alliance (CSA), ένας μη κερδοσκοπικός οργανισμός με αποστολή να προωθήσει τη χρήση βέλτιστων πρακτικών για την παροχή διασφάλισης ασφάλειας στο Cloud Computing, δημιούργησε μια ομάδα εργασίας Big Data που έχει επικεντρωθεί στις κύριες προκλήσεις για την εφαρμογή ασφαλών υπηρεσιών Big Data (Cloud Security Alliance, 2013). Η CSA έχει κατηγοριοποιήσει τις διαφορετικές προκλήσεις ασφάλειας και απορρήτου σε τέσσερις διαφορετικές πτυχές του οικοσυστήματος Big Data. Αυτές οι πτυχές είναι η ασφάλεια υποδομής, το απόρρητο δεδομένων, η διαχείριση δεδομένων και η ακεραιότητα και η αντιδραστική ασφάλεια. Κάθε μία από αυτές τις πτυχές αντιμετωπίζει τις ακόλουθες προκλήσεις ασφαλείας, σύμφωνα με την CSA:

<u>Ασφάλεια Υποδομών</u>

1. Ασφαλής Κατανεμημένη Επεξεργασία Δεδομένων
2. Ασφάλεια Βέλτιστες ενέργειες για μη σχεσιακές βάσεις δεδομένων

<u>Απόρρητο δεδομένων</u>

3. Ανάλυση Δεδομένων μέσω Εξόρυξης Δεδομένων Διατήρηση Προστασίας Δεδομένων
4. Κρυπτογραφικές λύσεις για την ασφάλεια δεδομένων
5. Κοκκώδης Έλεγχος Πρόσβασης

<u>Διαχείριση δεδομένων και ακεραιότητα</u>

6. Ασφαλής αποθήκευση δεδομένων και αρχεία καταγραφής συναλλαγών
7. Κοκκώδεις Έλεγχοι
8. Προέλευση δεδομένων

<u>Αντιδραστική ασφάλεια</u>

9. Φιλτράρισμα & επικύρωση από άκρο σε άκρο
10. Επίβλεψη του επιπέδου ασφαλείας σε πραγματικό χρόνο

Αυτές οι προκλήσεις ασφάλειας και απορρήτου καλύπτουν ολόκληρο το φάσμα του κύκλου ζωής Big Data όπως πηγές παραγωγής δεδομένων (συσκευές), τα ίδια τα δεδομένα, επεξεργασία δεδομένων, αποθήκευση δεδομένων, μεταφορά δεδομένων και χρήση δεδομένων σε διαφορετικές συσκευές. Μια ιδιαίτερη πτυχή της ασφάλειας και του απορρήτου των μεγάλων δεδομένων, πρέπει να σχετίζεται με την άνοδο του Διαδικτύου των Πραγμάτων (IoT).

Το IoT, που ορίζεται από την Oxford1 ως «μια προτεινόμενη ανάπτυξη του Διαδικτύου στην οποία τα καθημερινά αντικείμενα έχουν συνδεσιμότητα δικτύου, που τους επιτρέπει να στέλνουν και να λαμβάνουν δεδομένα», είναι ήδη πραγματικότητα – η Gartner εκτιμά ότι 26 δισεκατομμύρια συσκευές IoT θα εγκατασταθούν έως το 2020, δημιουργώντας πρόσθετα έσοδα 300 δισεκατομμυρίων δολαρίων (Rivera & van der Meulen, 2014).

Η τεράστια αύξηση του αριθμού των συνδεδεμένων συσκευών (αυτοκίνητα, συστήματα φωτισμού, ψυγεία, τηλέφωνα, γυαλιά, συστήματα ελέγχου κυκλοφορίας, συσκευές παρακολούθησης της υγείας, συστήματα SCADA, τηλεοράσεις, συστήματα οικιακής ασφάλειας, συστήματα οικιακού αυτοματισμού και πολλά άλλα) έχει οδηγήσει τους κατασκευαστές να προωθήσει στην αγορά, σε σύντομο χρονικό διάστημα, ένα μεγάλο σύνολο συσκευών, συστημάτων cloud και εφαρμογών για κινητές συσκευές για να εκμεταλλευτεί αυτή την ευκαιρία. Παρόλο που παρουσιάζει τεράστια οφέλη και ευκαιρίες για τους τελικούς χρήστες, είναι επίσης υπεύθυνο για τις προκλήσεις ασφάλειας.

Επίσης η επιχείρηση HP διεξήγαγε πρόσφατα μια μελέτη σχετικά με λύσεις IoT που είναι διαθέσιμες στην αγορά και κατέληξε στο συμπέρασμα ότι το 70% αυτών περιέχουν προβλήματα ασφάλειας. Αυτά τα προβλήματα ασφάλειας σχετίζονταν με ζητήματα απορρήτου, ανεπαρκή εξουσιοδότηση, έλλειψη μεταφοράς π κρυπτογράφηση, ανασφαλής διεπαφή ιστού και ανεπαρκής προστασία λογισμικού (HP, 2014).

Με βάση ορισμένα από αυτά τα ευρήματα, η HP έχει ξεκινήσει ένα έργο στο OWASP (Open Web Application Security Project) που τιτλοφορείται «*OWASP Internet of Things Top Ten*» (OWASP, 2014) του οποίου ο στόχος είναι να βοηθήσει τους προμηθευτές IoT να εντοπίσουν τα δέκα (10) κορυφαία προβλήματα των συσκευών ασφάλειας και πώς να τα αποφύγουν. Αυτό το έργο, παρόμοιο με το OWASP Top 10, εντόπισε τα ακόλουθα προβλήματα ασφάλειας:

Μη ασφαλής διεπαφή Ιστού, η οποία μπορεί να επιτρέψει σε έναν εισβολέα να εκμεταλλευτεί μια διεπαφή ιστού διαχείρισης (μέσω δέσμης ενεργειών μεταξύ τοποθεσιών, πλαστογράφησης αιτημάτων μεταξύ τοποθεσιών και έγχυσης SQL) και να αποκτήσει μη εξουσιοδοτημένη πρόσβαση για τον έλεγχο της συσκευής IoT.

Ανεπαρκής έλεγχος ταυτότητας/εξουσιοδότηση, η οποία μπορεί να επιτρέψει σε έναν εισβολέα να εκμεταλλευτεί μια κακή πολιτική κωδικών πρόσβασης, να σπάσει αδύναμους κωδικούς πρόσβασης και να αποκτήσει πρόσβαση σε προνομιακές λειτουργίες στη συσκευή IoT.

Υπηρεσίες ανασφαλούς δικτύου, οι οποίες μπορεί να οδηγήσουν σε έναν εισβολέα να εκμεταλλευτεί περιττές ή αδύναμες υπηρεσίες που εκτελούνται στη συσκευή ή να χρησιμοποιήσει αυτές τις υπηρεσίες ως σημείο μετάβασης για να επιτεθεί σε άλλες συσκευές στο δίκτυο IoT.

Έλλειψη κρυπτογράφησης μεταφοράς, η οποία επιτρέπει σε έναν εισβολέα να κρυφακούει δεδομένα κατά τη μεταφορά μεταξύ συσκευών IoT και συστημάτων υποστήριξης.

Ανησυχίες περί απορρήτου, προκύπτουν από το γεγονός ότι οι περισσότερες συσκευές και συστήματα υποστήριξης IoT συλλέγουν προσωπικά δεδομένα από χρήστες και αποτυγχάνουν να προστατεύσουν αυτά τα δεδομένα.

Insecure Cloud Interface, χωρίς κατάλληλους ελέγχους ασφαλείας ένας εισβολέας μπορεί να χρησιμοποιήσει πολλαπλά διανύσματα επίθεσης (ανεπαρκής έλεγχος ταυτότητας, έλλειψη κρυπτογράφησης μεταφοράς, απαρίθμηση λογαριασμού) για πρόσβαση σε δεδομένα ή στοιχεία ελέγχου μέσω του ιστότοπου cloud.

Μη ασφαλής διεπαφή κινητής τηλεφωνίας, χωρίς κατάλληλους ελέγχους ασφαλείας ένας εισβολέας μπορεί να χρησιμοποιήσει πολλαπλά διανύσματα επίθεσης (ανεπαρκής έλεγχος ταυτότητας, έλλειψη κρυπτογράφησης μεταφοράς, απαρίθμηση λογαριασμού) για πρόσβαση σε δεδομένα ή στοιχεία ελέγχου μέσω της διεπαφής για κινητά.

Ανεπαρκής παραμετροποίηση ασφαλείας λόγω της έλλειψης ή των κακών μηχανισμών διαμόρφωσης, ένας εισβολέας μπορεί να έχει πρόσβαση σε δεδομένα ή στοιχεία ελέγχου της συσκευής.

Μη ασφαλές λογισμικό, όπου οι εισβολείς μπορούν να επωφεληθούν από μη κρυπτογραφημένες και μη επαληθευμένες συνδέσεις για να παραβιάσουν ενημερώσεις συσκευών IoT και να εκτελέσουν κακόβουλη ενημέρωση που μπορεί να θέσει σε κίνδυνο τη συσκευή, ένα δίκτυο συσκευών και τα δεδομένα που διατηρούν.

Κακή φυσική ασφάλεια, όπου εάν η συσκευή IoT είναι φυσικά προσβάσιμη από έναν εισβολέα μπορεί να χρησιμοποιήσει θύρες USB, κάρτες SD ή άλλα μέσα αποθήκευσης για να αποκτήσει πρόσβαση στο λειτουργικό σύστημα της συσκευής και ενδεχομένως σε τυχόν δεδομένα που είναι αποθηκευμένα στη συσκευή.

Είναι σαφές λοιπόν βάσει των ανωτέρω, πως τα Big Data παρουσιάζουν ενδιαφέρουσες ευκαιρίες για χρήστες και επιχειρήσεις, ωστόσο αυτές οι ευκαιρίες αντιμετωπίζονται από τεράστιες προκλήσεις όσον αφορά το απόρρητο και την ασφάλεια (Cloud Security Alliance, 2013). Οι παραδοσιακοί μηχανισμοί ασφάλειας είναι ανεπαρκείς για να δώσουν μια ικανή απάντηση σε αυτές τις προκλήσεις.

4.3 Θέσπιση των Στόχων που Περιβάλουν τις Έννοιες “Security and Privacy”

Αναφερόμενοι στη θέσπιση των στόχων που περιβάλλουν τις έννοιες “Security and Privacy”, υπάρχει μια ενιαία μαγική λύση για την επίλυση των προκλήσεων ασφάλειας και απορρήτου των Big Data όπως και οι παραδοσιακές λύσεις ασφάλειας, οι οποίες είναι αφιερωμένες κυρίως στην προστασία μικρών ποσοτήτων στατικών δεδομένων και δεν επαρκούν στις νέες απαιτήσεις που επιβάλλουν οι υπηρεσίες Big Data (Cloud Security Alliance, 2013).

Υπάρχει λοιπόν η ανάγκη να κατανοήσουμε πώς μπορεί να προστατευτεί η συλλογή μεγάλων ποσοτήτων πολύπλοκων δομημένων και μη δομημένων στοιχείων. Η μη εξουσιοδοτημένη πρόσβαση σε αυτά τα δεδομένα για τη δημιουργία νέων σχέσεων, τον συνδυασμό διαφορετικών πηγών δεδομένων και τη διάθεση τους σε κακόβουλους χρήστες, αποτελεί σοβαρό κίνδυνο για τα Μεγάλα Δεδομένα.

Η βασική και πιο κοινή λύση για αυτό το γεγονός, περιλαμβάνει την κρυπτογράφηση των πάντων για την ασφάλεια των δεδομένων ανεξάρτητα από το πού βρίσκονται τα δεδομένα (κέντρο δεδομένων, υπολογιστής, φορητή συσκευή ή οποιαδήποτε άλλη). Καθώς τα Μεγάλα Δεδομένα αναπτύσσονται και η επεξεργασία τους γίνεται γρηγορότερη, η κρυπτογράφηση, η κάλυψη και το tokenization είναι κρίσιμα στοιχεία για την προστασία ευαίσθητων δεδομένων.

Λόγω των χαρακτηριστικών του, τα έργα Big Data πρέπει να έχουν ένα ολιστικό όραμα για την ασφάλεια (Tankard, 2012). Τα έργα Big Data πρέπει να λαμβάνουν υπόψη τον προσδιορισμό των διαφορετικών πηγών δεδομένων, την προέλευση και τους δημιουργούς των δεδομένων, καθώς και το ποιος επιτρέπεται να έχει πρόσβαση στα δεδομένα. Είναι επίσης απαραίτητο να διεξαχθεί μια σωστή ταξινόμηση για τον εντοπισμό κρίσιμων δεδομένων και την ευθυγράμμιση με την πολιτική ασφάλειας πληροφοριών του οργανισμού όσον αφορά την επιβολή πολιτικών ελέγχου πρόσβασης και διαχείρισης δεδομένων.

Ως σύσταση, διαφορετικοί μηχανισμοί ασφάλειας, θα πρέπει να είναι πιο κοντά στις πηγές δεδομένων και τα ίδια τα δεδομένα, προκειμένου να παρέχεται ασφάλεια ακριβώς στην προέλευση των δεδομένων, και οι μηχανισμοί ελέγχου και πρόληψης της αρχειοθέτησης, πρόληψης διαρροής δεδομένων και ελέγχου πρόσβασης θα πρέπει να συνεργάζονται (Kindervag, Balaouras, Hill, & Mak, 2012).

Οι νέες λύσεις ασφάλειας Big Data θα πρέπει να επεκτείνουν την ασφαλή περίμετρο από την επιχείρηση στο δημόσιο cloud (Juels & Oprea, 2013). Με αυτόν τον τρόπο, θα πρέπει επίσης να δημιουργηθεί ένας αξιόπιστος μηχανισμός προέλευσης δεδομένων σε όλους τους τομείς. Επιπλέον, παρόμοιοι μηχανισμοί με αυτούς που χρησιμοποιούνται σε (Luo, Lin, Zhang, & Zukerman, 2013) μπορεί να χρησιμοποιηθεί για τον μετριασμό των κατανεμημένων επιθέσεων άρνησης υπηρεσίας (DDoS) που εξαπολύονται εναντίον υποδομών Big Data. Επίσης, η ασφάλεια και το απόρρητο των μεγάλων δεδομένων είναι απαραίτητα για τη διασφάλιση της αξιοπιστίας των δεδομένων καθ' όλη τη διάρκεια του κύκλου ζωής των δεδομένων – από τη συλλογή δεδομένων έως τη χρήση.

Η δυνατότητα εξατομίκευσης ορισμένων υπηρεσιών Big Data και ο αντίκτυπός τους στο απόρρητο των χρηστών συζητούνται στο (Hasan, et al., 2013). Συζητούν αυτά τα θέματα στο πλαίσιο του EEXCESS, ενός συγκεκριμένου έργου που στοχεύει τόσο στην παροχή συστάσεων υψηλού επιπέδου όσο και στο σεβασμό του απορρήτου των χρηστών. Μια πρόσφατη εργασία περιγράφει προτεινόμενες επεκτάσεις απορρήτου στην UML για να βοηθήσουν τους μηχανικούς λογισμικού να

οπτικοποιήσουν γρήγορα τις απαιτήσεις απορρήτου και να τις σχεδιάσουν σε εφαρμογές Big Data (Jutla, Bodorik, & Ali, 2013).

Ενώ προσπαθεί κανείς να αξιοποιήσει στο έπακρο τα Μεγάλα Δεδομένα, όσον αφορά την ασφάλεια και το απόρρητο, καθίσταται υποχρεωτικό να πληρούνται οι μηχανισμοί που καλύπτουν νομικές απαιτήσεις σχετικά με τη διαχείριση δεδομένων. Πρέπει να χρησιμοποιείται ασφαλής τεχνολογία κρυπτογράφησης για την προστασία όλων των εμπιστευτικών δεδομένων (Προσωπικές πληροφορίες ταυτοποίησης (PII), Προστατευμένες Πληροφορίες Υγείας (PHI) και Πνευματική Ιδιοκτησία (IP) και προσεκτικές πολιτικές διαχείρισης πρόσβασης κρυπτογραφικού υλικού (κλειδιά), για να διασφαλιστεί το σωστό κλείδωμα και ξεκλείδωμα των δεδομένων – αυτό είναι ιδιαίτερα σημαντικό για τα αποθηκευμένα δεδομένα.

Για να είναι επιτυχείς αυτοί οι μηχανισμοί πρέπει να είναι διαφανείς στον τελικό χρήστη και να έχουν χαμηλό αντίκτυπο στην απόδοση και την επεκτασιμότητα των δεδομένων (λογισμικό και υλικό- πρέπει να ληφθούν υπόψη οι βασικοί μηχανισμοί κρυπτογράφησης) (Advantech, 2013). Όπως αναφέρθηκε προηγουμένως, η παραδοσιακή κρυπτογράφηση και η ανωνυμοποίηση δεδομένων δεν επαρκούν για την επίλυση προβλημάτων Big Data. Είναι επαρκείς για την προστασία των στατικών πληροφοριών, αλλά δεν είναι επαρκείς όταν εμπλέκεται ο υπολογισμός δεδομένων (MIT, 2014).

Ως εκ τούτου, πρέπει να χρησιμοποιηθούν άλλες τεχνικές, που επιτρέπουν συγκεκριμένο και στοχευμένο υπολογισμό δεδομένων, διατηρώντας τα δεδομένα μυστικά. Secure Function Evaluation (SFE) (Lindell & Pinkas, 2002), Fully Homomorphic Encryption (FHE) (Gentry, 2009) and Functional Encryption (FE) (Goldwasser et al., 2014) και διαμερισμός δεδομένων σε μη επικοινωνούντα κέντρα δεδομένων, μπορεί να βοηθήσει στην επίλυση των περιορισμών των παραδοσιακών τεχνικών ασφαλείας.

Η ομομορφική κρυπτογράφηση είναι μια μορφή κρυπτογράφησης που επιτρέπει σε συγκεκριμένους τύπους υπολογισμών (π.χ. αλγόριθμο κρυπτογράφησης δημόσιου κλειδιού RSA) να εκτελούνται σε κρυπτογραφημένο κείμενο και να δημιουργούν ένα κρυπτογραφημένο αποτέλεσμα το οποίο, όταν αποκρυπτογραφείται, ταιριάζει με το αποτέλεσμα των λειτουργιών που εκτελούνται στο απλό κείμενο (Gentry, 2010). Η πλήρως ομομορφική κρυπτογράφηση έχει πολλές εφαρμογές, όπως αναφέρεται στο (Van Dijk, et al., 2010).

Αυτό το γεγονός επιτρέπει κρυπτογραφημένα ερωτήματα σε βάσεις δεδομένων, τα οποία διατηρούν μυστικές ιδιωτικές πληροφορίες χρήστη όπου αυτά

τα δεδομένα αποθηκεύονται κανονικά (κάπου στο σύννεφο - στο όριο που ένας χρήστης μπορεί να αποθηκεύσει τα δεδομένα του σε οποιονδήποτε μη αξιόπιστο διακομιστή, αλλά σε κρυπτογραφημένη μορφή, χωρίς να ανησυχεί για τα δεδομένα μυστικότητα) (Ra Popa & Redfield, 2011).

Επιτρέπει επίσης ιδιωτικά ερωτήματα σε μια μηχανή αναζήτησης - ο χρήστης υποβάλλει ένα κρυπτογραφημένο ερώτημα και η μηχανή αναζήτησης υπολογίζει μια συνοπτική κρυπτογραφημένη απάντηση χωρίς ποτέ να κοιτάξει το ερώτημα καθαρά, το οποίο θα μπορούσε να περιέχει ιδιωτικές πληροφορίες χρήστη, όπως τον αριθμό της εθνικής υπηρεσίας υγειονομικής περίθαλψης.

Η ομομορφική κρυπτογράφηση επιτρέπει επίσης την αναζήτηση σε κρυπτογραφημένα δεδομένα - ένας χρήστης αποθηκεύει κρυπτογραφημένα αρχεία σε έναν απομακρυσμένο διακομιστή αρχείων και μπορεί αργότερα να ζητήσει από τον διακομιστή να ανακτήσει μόνο αρχεία που (όταν αποκρυπτογραφούνται) ικανοποιούν κάποιο περιορισμό, παρόλο που ο διακομιστής δεν μπορεί να αποκρυπτογραφήσει τα αρχεία από μόνος του. Γενικότερα, η πλήρως ομομορφική κρυπτογράφηση βελτιώνει την αποτελεσματικότητα του ασφαλούς πολυμερούς υπολογισμού.

Μια σημαντική πρόκληση ασφάλειας και απορρήτου για τα Big Data, σχετίζεται με την αποθήκευση και την επεξεργασία κρυπτογραφημένων δεδομένων. Η εκτέλεση ερωτημάτων σε μια κρυπτογραφημένη βάση δεδομένων είναι μια βασική απαίτηση ασφαλείας για ασφαλή Big Data, ωστόσο είναι μια πρόκληση. Αυτό εγείρει ερωτήματα όπως α) εάν η βάση δεδομένων είναι κρυπτογραφημένη με ένα μόνο ή πολλαπλά κλειδιά. β) χρειάζεται να αποκρυπτογραφηθεί η βάση δεδομένων πριν από την εκτέλεση του ερωτήματος; γ) τα ερωτήματα πρέπει επίσης να είναι κρυπτογραφημένα, δ) ποιος ως τα δικαιώματα για την αποκρυπτογράφηση της βάσης δεδομένων? και πολλά άλλα.

Πρόσφατα ένα σύστημα που αναπτύχθηκε στο MIT, δίνει απαντήσεις σε μερικά από αυτά τα ερωτήματα. Το CryptDB επιτρέπει στους ερευνητές να εκτελούν ερωτήματα βάσης δεδομένων πάνω από κρυπτογραφημένα δεδομένα (Ra Popa & Redfield, 2011). Οι αξιόπιστες εφαρμογές που σκοπεύουν να υποβάλουν ερωτήματα σε κρυπτογραφημένα δεδομένα θα περάσουν αυτά τα ερωτήματα σε έναν διακομιστή μεσολάβησης CryptDB (που βρίσκεται μεταξύ της εφαρμογής και της βάσης δεδομένων) που ξαναγράφει αυτά τα ερωτήματα με συγκεκριμένο τρόπο, ώστε να μπορούν να εκτελεστούν έναντι της κρυπτογραφημένης βάσης δεδομένων.

Η βάση δεδομένων επιστρέφει τα κρυπτογραφημένα αποτελέσματα πίσω στον διακομιστή μεσολάβησης, ο οποίος κρατά ένα κύριο κλειδί και θα αποκρυπτογραφήσει τα αποτελέσματα, στέλνοντας την τελική απάντηση πίσω στην

εφαρμογή. Το CryptDB υποστηρίζει πολυάριθμες μορφές σχημάτων κρυπτογράφησης που επιτρέπουν διαφορετικούς τύπους λειτουργιών στα δεδομένα (RA Popa & Redfield, 2012). Με βάση το CryptDB, η Google έχει αναπτύξει το Encryp ted Big Query Client που θα επιτρέπει κρυπτογραφημένα μεγάλα ερωτήματα έναντι της υπηρεσίας BigQuery που ενεργοποιεί εξαιρετικά ερωτήματα τύπου SQL έναντι πινάκων μόνο με προσάρτημα, χρησιμοποιώντας την επεξεργαστική ισχύ της υποδομής της Google (Google, 2014).

Εκτός από πιο συγκεκριμένες συστάσεις ασφάλειας, είναι επίσης σημαντικό να ληφθεί υπόψη η ασφάλεια της ίδιας της υποδομής πληροφορικής. Μία από τις κοινές πρακτικές ασφαλείας είναι η τοποθέτηση ελέγχων ασφαλείας στην άκρη των δικτύων, ωστόσο, εάν ένας εισβολέας παραβιάσει αυτήν την περίμετρο ασφαλείας, θα έχει πρόσβαση σε όλα τα δεδομένα μέσα σε αυτό. Επομένως, είναι απαραίτητη μια νέα προσέγγιση για να μετακινηθούν αυτά τα στοιχεία ελέγχου ασφαλείας κοντά στα δεδομένα (ή να προστεθούν επιπλέον).

Η παρακολούθηση, η ανάλυση και η μάθηση από τη χρήση και την πρόσβαση δεδομένων είναι επίσης μια σημαντική πτυχή για τη συνεχή βελτίωση της ασφάλειας της υποδομής αποθήκευσης δεδομένων και την αξιοποίηση των ήδη υπαρχουσών λύσεων ασφαλείας (Kindervag et al., 2012, Kindervag, Wang, Balaouras, & Coit, 2011).

Οι ερευνητικές προκλήσεις σε αυτό το οικοσύστημα Big Data κυμαίνονται από τη δημιουργία δεδομένων (και τις πηγές Big Data - συσκευές), την αποθήκευση και μεταφορά δεδομένων, τον μετασχηματισμό και την επεξεργασία δεδομένων και, τέλος, τη χρήση δεδομένων. Για την υποστήριξη αυτού του κύκλου ζωής, θα χρειαστεί μια αρχιτεκτονική υψηλής χωρητικότητας και υψηλής κατανομής, εκτεθειμένη σε ένα εχθρικό περιβάλλον που υπόκειται σε κάθε είδους επιθέσεις.

Η προσέγγιση SDN όπως προτείνεται σε αυτό το κεφάλαιο είναι μια πιθανή λύση για την αντιμετώπιση αυτών των απειλών, ωστόσο πρέπει να διεξαχθεί περαιτέρω έρευνα, ιδίως σχετικά με τις ανησυχίες για την αυτόματη προσαρμογή των πολιτικών ασφαλείας που βασίζονται στη μεταγωγή και τη συμπεριφορά (Chen, Jorgen, & Yuan, 2011, Dohi & Uemura, 2012).

Υπάρχουν επίσης σημαντικές ερευνητικές προκλήσεις σχετικά με τη διατήρηση της ασφάλειας και του απορρήτου των δεδομένων από άκρο σε άκρο. Διασφάλιση ότι τα δεδομένα δεν αποκαλύπτονται ποτέ με σαφήνεια, ιδίως σε μη εξουσιοδοτημένα μέρη, σε οποιοδήποτε σημείο του κύκλου ζωής των Μεγάλων Δεδομένων. Μεταβαίνοντας από τα δεδομένα στα προγράμματα, υπάρχουν τεχνικές

για την προστασία του απορρήτου κατά την περιήγηση, την αναζήτηση, τις κοινωνικές αλληλεπιδράσεις και τη γενική χρήση μέσω μεθόδων συσκότισης.

Ωστόσο, πρέπει να διεξαχθεί περισσότερη έρευνα σχετικά με την επεξεργασία κρυπτογραφημένων δεδομένων και την προστασία της ιδιωτικής ζωής στο πλαίσιο τόσο προγραμμάτων υπολογιστών όσο και συστημάτων που βασίζονται στο διαδίκτυο. Περισσότερες ερευνητικές προκλήσεις στην περιοχή Big Data περιλαμβάνουν την ανάπτυξη τεχνικών για την εκτέλεση διαφανών υπολογισμών σε κρυπτογραφημένα δεδομένα με πολλαπλά κλειδιά, από πολλές πηγές και πολλούς χρήστες.

Όσον αφορά την έρευνα, θα ήταν δύσκολο να μελετηθούν και να αναπτυχθούν τρόποι ανάθεσης περιορισμένων λειτουργιών σε κρυπτογραφημένα δεδομένα, έτσι ώστε τρίτα μέρη να μπορούν να τα αναλύσουν. Όλες οι πτυχές που σχετίζονται με τη διαχείριση κλειδιών, την ανάθεση εξουσιοδότησης, τη διαχείριση δικαιωμάτων, είναι θέματα που απαιτούν περαιτέρω έρευνα σε αυτόν τον τομέα.

Όταν εξετάζει κανείς το ενδεχόμενο ασφαλούς και ιδιωτικού συστήματος, η εμπιστοσύνη είναι το παν. Ειδικότερα, στην περίπτωση των Big Data, θα πρέπει να δημιουργηθεί ένα αξιόπιστο περιβάλλον για τα περισσότερα σενάρια (υγειονομική περίθαλψη, υποβοηθούμενη διαβίωση, συστήματα SCADA και πολλά άλλα). Είναι ιδιαίτερα δύσκολο από πλευράς ερευνητικών κατευθύνσεων πώς μπορεί να επιτευχθεί αυτό το περιβάλλον.

Η εμπιστοσύνη σε εφαρμογές που είναι σε θέση να διερευνούν και να επεξεργάζονται Big Data και να εξάγουν γνώση από αυτά και, να εμπιστεύονται συσκευές που συλλέγουν όλα τα δεδομένα από πολλαπλές πηγές, αποτελεί βασική απαίτηση ασφαλείας. Η κατανόηση του τρόπου με τον οποίο μπορεί να δημιουργηθεί εμπιστοσύνη μεταξύ των τελικών χρηστών, οι συσκευές (IoT) και οι εφαρμογές είναι ένα καυτό ερευνητικό θέμα για τα επόμενα χρόνια. Όσον αφορά τα Μεγάλα Δεδομένα, αυτές οι ερευνητικές προκλήσεις αντιπροσωπεύουν μόνο την κορυφή του παγόβουνου σχετικά με τα προβλήματα που πρέπει ακόμη να μελετηθούν και να επιλυθούν σχετικά με την ανάπτυξη ενός ασφαλούς οικοσυστήματος Big Data με επίγνωση της ιδιωτικής ζωής.

ΚΕΦΑΛΑΙΟ 5ο – ΕΠΙΛΟΓΟΣ

Σύμφωνα με όσα αναφέρθηκαν παραπάνω, θα λέγαμε πως στην εποχή των *μεγάλων δεδομένων* (big data), ο τρόπος ζωής, οι καθημερινές συνήθειες και οι τρόποι σκέψης των ανθρώπων, έχουν υποστεί ριζικές αλλαγές. Τα μεγάλα δεδομένα (big data), έχουν γίνει ένα σημαντικό θέμα για την έρευνα στη βιομηχανία και τον ακαδημαϊκό κόσμο.

Κατά τη διαδικασία συλλογής, αποθήκευσης και χρήσης δεδομένων, αυτό το γεγονός μπορεί εύκολα να οδηγήσει σε *διαρροή* προσωπικών πληροφοριών και στο γεγονός ότι τα δεδομένα είναι δύσκολο να διακριθούν. Ο τρόπος διασφάλισης της ασφάλειας των μεγάλων δεδομένων και της προστασίας του απορρήτου έχει γίνει ένα από τα καυτά ζητήματα στο τρέχον στάδιο της έρευνας.

Στην εποχή των μεγάλων δεδομένων (big data) λοιπόν, οι άνθρωποι είναι οι ουσιαστικοί ωφελούμενοι της τεχνολογίας του Διαδικτύου. Τα δεδομένα έχουν μεγάλη εμπορική αξία για τους παρόχους υπηρεσιών Διαδικτύου, αλλά η ανάλυση και η εφαρμογή τους, θα είναι πιο περίπλοκη και πιο δύσκολη στη διαχείριση και το προσωπικό απόρρητο θα απειληθεί.

Με την ταχεία ανάπτυξη του Διαδικτύου, οι άνθρωποι αφήνουν καθημερινά πολλά ίχνη δεδομένων στο Διαδίκτυο καθημερινά. Αυτό προσφέρει στους εγκληματίες την ευκαιρία να συλλέγουν πληροφορίες στο Διαδίκτυο και στη συνέχεια να διεξάγουν παράνομες δραστηριότητες όπως μεταπώληση, απάτη κ.λπ., όχι μόνο για ανθρώπους. Στην εποχή των μεγάλων δεδομένων, ο τρόπος αντιμετώπισης ζητημάτων ασφάλειας και απορρήτου στο πλαίσιο των μεγάλων δεδομένων είναι επιτακτική ανάγκη για τους ανθρώπους να έχουν μια καλή λύση.

Η προέλευση των μεγάλων δεδομένων (big data), προέρχεται από το Διαδίκτυο. Οι ερευνητές δημιουργούν διαφοροποιημένα μοντέλα με βάση τις ανάγκες της επιχείρησης και στη συνέχεια εξάγουν σημαντικά διανύσματα με βάση τα μοντέλα για να βρουν τρόπους αντιμετώπισης ανθρώπων ή πραγμάτων σε διαφορετικούς ρόλους. Αυτή είναι η πηγή και τα χαρακτηριστικά των μεγάλων δεδομένων.

Σύμφωνα με τις πηγές των μεγάλων δεδομένων, αυτά μπορούν να διαχωριστούν σε τρεις κατηγορίες, Πρώτον, όλα τα είδη δεδομένων που προέρχονται από άτομα, όπως τα άτομα στη διαδικασία χρήσης του Διαδικτύου, συμπεριλαμβανομένων βίντεο, εικόνων, κειμένου κ.λπ. δεύτερον, από τα σχετικά μηχανήματα. Τα δεδομένα που παράγονται από διάφορους τύπους υπολογιστών κατά τη διάρκεια των εργασιών είναι με τη μορφή πολυμέσων, βάσεων δεδομένων,

GPS, έξυπνων κατοικιών, εγγράφων κ.λπ. Το τρίτο είναι από αντικείμενα. Τα δεδομένα που συλλέγονται κατά τη λειτουργία διαφόρων τύπων ψηφιακών συσκευών, όπως τα ψηφιακά σήματα που λαμβάνονται από την κάμερα.

Αναφέρονται επίσης προκλήσεις ασφάλειας μεγάλων δεδομένων όπως οι κίνδυνοι της ιδιωτικότητας. Ενώ οι άνθρωποι απολαμβάνουν την άνεση που προσφέρουν τα μεγάλα δεδομένα, αντιμετωπίζουν επίσης πολλές δυσκολίες. Εάν τα μεγάλα δεδομένα δεν προστατεύονται καλά για τα δεδομένα χρήστη κατά τη διαδικασία χρήσης, θα απειλήσουν άμεσα το απόρρητο των χρηστών και την ασφάλεια των δεδομένων. Σύμφωνα με το διαφορετικό περιεχόμενο προστασίας, μπορεί να χωριστεί σε ανώνυμα αναγνωριστικά, ανώνυμη προστασία και προστασία απορρήτου.

Στην εποχή των μεγάλων δεδομένων, τα προβλήματα ασφάλειας δεδομένων των ανθρώπων, δεν είναι μόνο τα παραδοσιακά ζητήματα προσωπικού απορρήτου, αλλά βασίζονται περισσότερο στην ανάλυση και έρευνα των δεδομένων των ανθρώπων και στη στοχευμένη πρόβλεψη της κατάστασης και της συμπεριφοράς των ανθρώπων. Για παράδειγμα, οι έμποροι λιανικής μπορούν να συγκρίνουν. Οι γονείς γνωρίζουν καλύτερα τις συνήθειες δαπανών των παιδιών τους κ.λπ., και έτσι δημοσιεύουν σχετικές διαφημιστικές πληροφορίες.

Ένα άλλο παράδειγμα είναι το status περιεχομένου που δημοσιεύεται από χρήστες στο Διαδίκτυο και μπορεί να αναλύσει τις πολιτικές πληροφορίες αυτού του ατόμου, όπως την ομάδα και τις συνήθειες δαπανών. Προς το παρόν, πολλές εταιρείες πιστεύουν ότι μετά την ανώνυμη επεξεργασία των πληροφοριών, τα αναγνωριστικά θα κρυφτούν και στη συνέχεια οι πληροφορίες θα δημοσιοποιηθούν. Ωστόσο, η πραγματικότητα είναι ότι η προστασία της ιδιωτικής ζωής δεν μπορεί να επιτευχθεί αποτελεσματικά μόνο μέσω της ανώνυμης προστασίας.

Για παράδειγμα, μια εταιρεία μπορεί να χρησιμοποιήσει ορισμένες από τις εγγραφές του ιστορικού αναζήτησής της, με ανώνυμο τρόπο εντός 3 μηνών για χρήση από άτομα. Αν και οι πληροφορίες αναγνώρισης που περιέχονται σε αυτές τις αναζητήσεις, έχουν γίνει προσεκτικά, τα περιεχόμενα πολλών από τις εγγραφές που περιέχονται σε αυτές μπορούν να οριστούν με ακρίβεια.

Σημαντική είναι επίσης η περίπτωση όπου η αξιοπιστία των μεγάλων δεδομένων πρέπει να επιβεβαιωθεί. Για παράδειγμα, ορισμένοι ιστότοποι περιέχουν ψευδή σχόλια και οι χρήστες μπορούν εύκολα να αγοράσουν αυτά τα αγαθά και υπηρεσίες, αφού διακρίνουν αυτά τα ψεύτικα σχόλια. Σε συνδυασμό με την τρέχουσα δημοτικότητα της τεχνολογίας του Διαδικτύου, ο αντίκτυπος αυτών των ψευδών

πληροφοριών, είναι ανυπολόγιστος και η χρήση της τεχνολογίας ασφάλειας πληροφοριών για τον έλεγχο αυτών των δεδομένων είναι επίσης πολύ δύσκολη.

Επίσης τα μεγάλα δεδομένα μπορεί να αλλοιωθούν κατά τη διαδικασία διάδοσης. Επομένως, η διασφάλιση της αυθεντικότητας και της αξιοπιστίας των δεδομένων είναι εξαιρετικά σημαντική. Επίσης, αναφέρεται πως στην εποχή των μεγάλων δεδομένων, οι πληροφορίες διαχέονται με εξαιρετικά γρήγορους ρυθμούς.

Ταυτόχρονα με τη μετάδοση πληροφοριών, λόγω της αδύναμης εποπτείας των πληροφοριών δεδομένων, της έλλειψης τεχνικής υποστήριξης, του ατελούς συστήματος εποπτείας και της ευπάθειας απώλειας πληροφοριών, η χρήση των πληροφοριών δεδομένων δεν είναι μεγάλης αξίας και τα δεδομένα μειώνονται. Η ίδια η αξία θα επιφέρει πολλές αρνητικές και αρνητικές επιπτώσεις σε άτομα, επιχειρήσεις ακόμα και στην κοινωνία, με αποτέλεσμα μεγαλύτερες οικονομικές απώλειες.

Στο πλαίσιο αυτό επίσης, σημειώνεται πως η ασφάλεια των έξυπνων τερματικών έχει γίνει επίσης ένα σοβαρό πρόβλημα για τους χρήστες. Τα διαδικτυακά μέσα που δημιουργήθηκαν στην εποχή των μεγάλων δεδομένων, έχουν γίνει το πιο σημαντικό κανάλι διαπροσωπικής επικοινωνίας. Επίσης οι περισσότερες δυτικές χώρες έχουν δημιουργήσει ειδικές υπηρεσίες προστασίας της ιδιωτικής ζωής για την προστασία του απορρήτου και των πληροφοριών των πολιτών. Με τη δημιουργία μιας υπηρεσίας προστασίας της ιδιωτικής ζωής, όχι μόνο μπορεί να παρακολουθείται αποτελεσματικά η διαδικτυακή συμπεριφορά των ανθρώπων, αλλά και ο σκοπός της διάδοσης του νόμου.

Ως εκ τούτου, με τη συνεχή πρόοδο της εποχής των μεγάλων δεδομένων, ο αριθμός των πληροφοριών δεδομένων, έχει αυξηθεί σημαντικά. Οι πολίτες πρέπει να προσαρμοστούν στις αλλαγές της εποχής και να αυξήσουν σταδιακά τον αλφαβητισμό και την ευαισθητοποίησή τους στα δεδομένα. Η παιδεία στα δεδομένα αυτά, απευθύνεται κυρίως σε επιστημονικούς ερευνητές και δημόσιους υπαλλήλους. Απαιτεί ότι όταν έρχονται σε επαφή με τις πληροφορίες των πολιτών, μπορούν να διαχειρίζονται αποτελεσματικά τις πληροφορίες των πολιτών και να αναλαμβάνουν την πρωτοβουλία να αναλάβουν την ευθύνη της προστασίας της ιδιωτικής ζωής των πολιτών, ώστε να προστατεύεται αποτελεσματικά η ιδιωτική ζωή των πολιτών.

Η ευαισθητοποίηση δεδομένων απευθύνεται κυρίως στο ευρύ κοινό και απαιτεί από τους πολίτες να συνειδητοποιήσουν τη σημασία των μεγάλων δεδομένων. Η έλευση της εποχής των μεγάλων δεδομένων όχι μόνο παρείχε σημαντικές ευκαιρίες για κοινωνική πρόοδο, αλλά επέφερε επίσης πολλές απειλές για

την ασφάλεια των πληροφοριών στην κοινωνία, καθιστώντας την προστασία του απορρήτου των προσωπικών δεδομένων ανησυχητικό.

Για να πραγματοποιηθεί η ασφάλεια και η προστασία της ιδιωτικής ζωής των πληροφοριών μεγάλων δεδομένων, απαιτείται όχι μόνο ένας μεγάλος αριθμός επαγγελματικών τεχνολογιών ιδιωτικής ασφάλειας πληροφοριών, αλλά και η συνειδητοποίηση της προστασίας της ιδιωτικής ζωής των πολιτών στη κάθε χώρα πρέπει να ενισχυθεί, ώστε να μπορεί να εφαρμοστεί η ασφάλεια των πληροφοριών απορρήτου.

Ο κύριος σκοπός της ανάλυσης Big Data είναι να αποκτήσει χρήσιμες πληροφορίες από έναν μεγάλο όγκο ετερογενών δεδομένων [30]. Ωστόσο, η πρόσβαση σε μεγάλης κλίμακας, κατανεμημένα σύνολα δεδομένων παρουσιάζει ορισμένες ανησυχίες σχετικά με το απόρρητο και την ασφάλεια που έχουμε συζητήσει εν συντομία σε αυτό το έγγραφο. Ερευνήσαμε επίσης πώς τα Big Data έχουν διαφορετικές απαιτήσεις όσον αφορά την ασφάλεια και το απόρρητο σε διαφορετικούς τομείς όπως η συλλογή δεδομένων, η αποθήκευση, η ανάλυση και η μεταφορά.

Επιπλέον, έχουμε αναθεωρήσει συγκριτικά έναν αριθμό μελετών που έγιναν σχετικά με την ασφάλεια και το απόρρητο των μεγάλων δεδομένων, βάσει των οποίων συνήχθη το συμπέρασμα ότι είναι σημαντικό να παρακολουθείται συνεχώς η κυκλοφορία του δικτύου προκειμένου να εντοπίζονται γρήγορα ύποπτες συμπεριφορές. Με τον τύπο δεδομένων, οι χρήστες και οι συσκευές πρέπει να έχουν πρόσβαση για να μπορούν να χρησιμοποιούν πόρους, όλες οι επικοινωνίες θα πρέπει να πραγματοποιούνται μέσω ασφαλών καναλιών και τα προσωπικά δεδομένα θα πρέπει να καλύπτονται πριν από τη δημοσίευση του συνόλου δεδομένων.

Η ιδιωτικότητα και η ασφάλεια των μεγάλων δεδομένων είναι ένας από τους πιο σημαντικούς τομείς για περαιτέρω συζήτηση και έρευνα στο μέλλον. Είναι προφανές ότι πλέον υπάρχει ανάγκη για ανάπτυξη νέων ή αναβάθμιση σύγχρονων τεχνικών, τεχνολογιών και λύσεων σε σχέση με τις τρέχουσες ανάγκες. Ωστόσο, όπως αναφέρθηκε στην προηγούμενη ενότητα, πρέπει να έχουμε κατά νου ότι τα Big Data μπορούν να συγκριθούν με ένα γεμάτο όπλο, μπορεί να προκαλέσει βλάβη εάν δεν χρησιμοποιηθεί με ασφαλή τρόπο με σωστή ρύθμιση, αλλά μπορεί επίσης να παρέχει ασφάλεια και ασφάλεια εάν χρησιμοποιείται σωστά.

Η δραματική αύξηση του όγκου των αποθηκευμένων και των ροών και η ικανότητα ανάλυσής τους μπορούν να χρησιμοποιηθούν σε μεγάλο βαθμό σε τομείς ασφάλειας πληροφοριών, όπως ο εντοπισμός ή η πρόβλεψη ανωμαλιών, η εισβολή

και η απάτη, απλώς εξετάζοντας αρχεία καταγραφής/συμβάντων/και επισκεψιμότητας συστήματος, δικτύου και ιστότοπων. Για το σκοπό αυτό, θα πρέπει να συλλεχθεί μεγάλος όγκος και ποικιλία δεδομένων που σχετίζονται με το ιστορικό του δικτύου και να αναλυθούν για αναγνώριση προτύπων.

Μερικά από τα πλεονεκτήματα της χρήσης των Μεγάλων Δεδομένων, περιλαμβάνουν, την απόδοση του συστήματος χωρίς την ανάγκη διαγραφής ακυρωμένων λογαριασμών ή παλαιών αρχείων καταγραφής μετά από ένα ορισμένο χρονικό διάστημα, ιδίως επειδή αυτά μπορεί να είναι χρήσιμα για τους σκοπούς των εγκληματολογικών ερευνών αργότερα, καθώς και τη δυνατότητα εκτέλεσης περίπλοκων και προηγμένων ερωτημάτων σε μεγάλα και αδόμητα σύνολα δεδομένων, δυνατότητα λήψης αποφάσεων σε πραγματικό χρόνο, συστήματα αυτόματης άμυνας και μείωσης κινδύνου με την πρόβλεψη επιθέσεων στο μέλλον, και τέλος ταχύτερη, καλύτερη και φθηνότερη ασφάλεια σε σύγκριση με τις παραδοσιακές μεθόδους.

Τέλος, η ανάπτυξη κατάλληλων συστημάτων, τεχνολογιών και λύσεων για την αντιμετώπιση των προκλήσεων που σχετίζονται με τα μεγάλα δεδομένα, μπορεί να συμβάλει στην περαιτέρω άμβλυνση των σημείων συμφόρησης στους τομείς της ασφάλειας και της ιδιωτικής ζωής, όχι μόνο για το σήμερα, αλλά και για το μέλλον.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Agrawal, D., Das, S., & El Abbadi, A. (2011). Big data and cloud computing. In Proceedings of the 14th International Conference on Extending Database Technology - EDBT/ICDT ’11 (p. 530). New York, New York, USA: ACM Press. doi:10.1145/1951365.1951432

Carlos Serrão, Neves, D., Trevor Barker, & Massimo Balestri. (2003). OpenSDRM -- An Open and Secure Digital Rights Management Solution. In Proceedings of the IADIS International Conference e- Society.

Chen, P., Jorgen, B., & Yuan, Y. (2011). Software behavior based trusted attestation. In Proceedings - 3rd International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2011 (Vol. 3, pp. 298–301). doi:10.1109/ICMTMA.2011.645

Chen, X., & Shi, S. (2009). A literature review of privacy research on social network sites. In Multimedia Information Networking and Security, 2009. MINES’09. International Conference on (Vol. 1, pp. 93–97).

Cloud Security Alliance. (2013). Expanded Top Ten Security and Privacy Challenges. Retrieved from https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_an d_Privacy_Challenges.pdf

CTL. (2014). Computation tree logic. Retrieved from http://en.wikipedia.org/wiki/Computation_tree_logic

DARPA. (2014). MINING AND UNDERSTANDING SOFTWARE ENCLAVES (MUSE). Retrieved August from http://www.darpa.mil/Our_Work/I2O/Programs/Mining_and_Understanding_Software_Enclaves_(MUS E).aspx

De Cristofaro, E., Soriente, C., Tsudik, G., & Williams, A. (2012). Hummingbird: Privacy at the time of twitter. In Security and Privacy (SP), 2012 IEEE Symposium on (pp. 285–299).

Demchenko, Y., Ngo, C., Laat, C. de, Membrey, P., & Gordijenko, D. (2014). Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure. In W. Jonker & M. Petković (Eds.), Secure Data Management (pp. 76–94). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-06811-4_13

Dohi, T., & Uemura, T. (2012). An adaptive mode control algorithm of a scalable intrusion tolerant architecture. In Journal of Computer and System Sciences (Vol. 78, pp. 1751–1754).

Feamster, N. (2014). Software Defined Networking. Retrieved from https://www.coursera.org/course/sdn

Gentry, C. (2009). A fully homomorphic encryption scheme. Stanford University. Retrieved from http://cs.au.dk/~stm/local-cache/gentry-thesis.pdf

Gentry, C. (2010). Computing arbitrary functions of encrypted data. Communications of the ACM. doi:10.1145/1666420.1666444

Goldwasser, S., Gordon, S. D., Goyal, V., Jain, A., Katz, J., Liu, F.-H., ... Zhou, H.-S. (2014). Multi- input functional encryption. In Advances in Cryptology--EUROCRYPT 2014 (pp. 578–602). Springer.

Google. (2014). Encrypted Big Query Client. Retrieved from https://code.google.com/p/encrypted-bigquery-client/

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In Proceedings of the 2005 ACM workshop on Privacy in the electronic society (pp. 71–80).

Hand, R., Ton, M., & Keller, E. (2013). Active security. In Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks - HotNets-XII (pp. 1–7). New York, New York, USA: ACM Press. doi:10.1145/2535771.2535794

Hasan, O., Habegger, B., Brunie, L., Bennani, N., & Damiani, E. (2013). A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case. In 2013 IEEE International Congress on Big Data (pp. 25–30). IEEE. doi: 10.1109/BigData.Congress.2013.13

HP. (2014). Internet of Things Research Study (p. 4). Retrieved from http://fortifyprotect.com/HP_IoT_Research_Study.pdf

IBM big data platform - Bringing big data to the Enterprise. (2014, July). CT000.

IDC. (2012). Big Data in 2020. Retrieved from http://www.emc.com/leadership/digital-universe/2012iview/big-data-2020.htm

Juels, A., & Oprea, A. (2013). New approaches to security and availability for cloud data. Communications of the ACM, 56(2), 64. doi:10.1145/2408776.2408793

Jutla, D. N., Bodorik, P., & Ali, S. (2013). Engineering Privacy for Big Data Apps with the Unified Modeling Language. In 2013 IEEE International Congress on Big Data (pp. 38–45). IEEE. doi: 10.1109/BigData.Congress.2013.15

Kim, C., Jin, M.-H., Kim, J., & Shin, N. (2012). User perception of the quality, value, and utility of user- generated content. Journal of Electronic Commerce Research, 13(4), 305–319.

Kindervag, J., Balaouras, S., Hill, B., & Mak, K. (2012). Control and Protect Sensitive Information In the Era of Big Data.

Kindervag, J., Wang, C., Balaouras, S., & Coit, L. (2011). Applying Zero Trust to The Extending Enterprise.

Kreutz, D., Ramos, F. M. V., Verissimo, P., Rothenberg, C. E., Azodolmolky, S., & Uhlig, S. (2014). Software-Defined Networking: A Comprehensive Survey, 49. Networking and Internet Architecture. Retrieved from http://arxiv.org/abs/1406.0440

Lindell, Y., & Pinkas, B. (2002). Privacy Preserving Data Mining. Journal of Cryptology, 15(3), 177– 206. doi:10.1007/s00145-001-0019-2

Luo, H., Lin, Y., Zhang, H., & Zukerman, M. (2013). Preventing DDoS attacks by identifier/locator separation. IEEE Network, 27(6), 60–65. doi:10.1109/MNET.2013.6678928

Marques, J., & Serrão, C. (2013a). Improving Content Privacy on Social Networks Using Open Digital Rights Management Solutions. Procedia Technology, 9, 405–410. doi: 10.1016/j.protcy.2013.12.045

Marques, J., & Serrão, C. (2013b). Improving user content privacy on social networks using rights management systems. Annals of Telecommunications - Annales Des Télécommunications, 69(1-2), 37–45. doi:10.1007/s12243-013-0388-1

McKenzie, P. J., Burkell, J., Wong, L., Whippey, C., Trosow, S. E., & McNally, M. B. (2012, June 6). User-generated online content: overview, current state and context. First Monday. Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/3912/3266

Michael, K., & Miller, K. W. (2013). Big Data: New Opportunities and New Challenges [Guest editors' introduction]. Computer, 46(6), 22–24. doi:10.1109/MC.2013.196

MIT. (2014). Big Data Privacy Workshop, Advancing the state of the art in Technology and Practice - Workshop summary report. Retrieved from http://web.mit.edu/bigdata-priv/images/MITBigDataPrivacyWorkshop2014_final05142014.pdf

Monsanto, C., Reich, J., Foster, N., Rexford, J., & Walker, D. (2013). Composing software-defined networks. Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation, 1–14. Retrieved from http://dl.acm.org/citation.cfm?id=2482626.2482629\nhttp://www.frenetic-lang.org/pyretic/

NIST. (2014). Framework for Improving Critical Infrastructure Cybersecurity. Retrieved from http://www.nist.gov/cyberframework/upload/cybersecurity-framework-021214-final.pdf

NuSMV. (2014). An overview of NuSMV. Retrieved from http://nusmv.fbk.eu/NuSMV/

Okhravi, H., Hobson, T., Bigelow, D., & Streilein, W. (2014). Finding Focus in the Blur of Moving- Target Techniques. IEEE Security & Privacy, 12(2), 16–26. doi:10.1109/MSP.2013.137

OWASP. (2014). OWASP Internet of Things Top Ten Project. Retrieved from https://www.owasp.org/index.php/OWASP_Internet_of_Things_Top_Ten_Project

Popa, R., & Redfield, C. (2011). Cryptdb: protecting confidentiality with encrypted query processing. Proceedings of the ..., 85–100. doi:10.1145/2043556.2043566

Popa, R., & Redfield, C. (2012). CryptDB: Processing queries on an encrypted database. Communications of the ..., 55, 103. doi:10.1145/2330667.2330691

Pyretic. (2014). Pyretic Language. Retrieved from https://github.com/frenetic-lang/pyretic/wiki/Language-Basics

Python. (2014). Python Language. Retrieved from https://www.python.org/ Rivera, J., & van der Meulen, R. (2014). Gartner Says the Internet of Things Will Transform the Data Center. Retrieved from http://www.gartner.com/newsroom/id/2684915

Rodríguez, E., Rodríguez, V., Carreras, A., & Delgado, J. (2009). A Digital Rights Management approach to privacy in online social networks. In Workshop on Privacy and Protection in Web-based Social Networks (within ICAIL'09), Barcelona.

Serrão, C. (2008). IDRM - Interoperable Digital Rights Management: Interoperability Mechanisms for Open Rights Management Platforms. Universitat Politècnica de Catalunya. Retrieved from http://repositorio-iul.iscte.pt/handle/10071/1156

Serrão, C., Dias, J. M. S., & Kudumakis, P. (2005). From OPIMA to MPEG IPMP-X: A standard's history across R&D projects. Signal Processing: Image Communication, 20(9), 972–994.

Serrão, C., Dias, M., & Delgado, J. (2005). Using Web-Services to Manage and Control Access to Multimedia Content. ISWS05-The 2005 International Symposium on Web Services and Applications, Las Vegas, USA.

Serrão, C., Rodriguez, E., & Delgado, J. (2011). Approaching the rights management interoperability problem using intelligent brokerage mechanisms. Computer Communications, 34(2), 129–139.

Stephen A Thomas. (2000). SSL & TLS Essentials: Securing the Web (Pap/Cdr., p. 224). Wiley.

Tankard, C. (2012). Big data security. Network Security, 2012(7), 5–8. doi:10.1016/S1353- 4858(12)70063-6

Van Dijk, M., Gentry, C., Halevi, S., & Vaikuntanathan, V. (2010). Fully homomorphic encryption over the integers. In Advances in Cryptology– EUROCRYPT ’10 (pp. 24–43). doi:10.1007/978-3-642-38348- 9_20

ISO / IEC 20547/4 (2020). International Standard / International Technology.