ENISA Threat Landscape 2023 - Threats Against Data

(DIT172 | Project 2 | itp22104 / Anastasios Kotronis)

Τα τελευταία χρόνια με την διεύρυνση της χρήσης του διαδικτύου αλλά και του γενικότερου ψηφιακού μετασχηματισμού παρατηρείται μια τεράστια αύξηση του όγκου των δεδομένων στους διάφορους τομείς παραγωγής.

Η κατάσταση αυτή αποτυπώνεται γλαφυρά από τη δήλωση του Βρετανού μαθηματικού Clive Humby το 2006 σε συνέδριο της Association of National Advertisers: *data is the new oil* αλλά και τη δημοσίευση στο Economist το 2017 με τίτλο *The world's most valuable resource is no longer oil, but data*.

Ενδεικτική του μεγέθους των δεδομένων που δημιουργούνται αλλά και της υπερσυγκέντρωσής τους από εταιρίες κολοσσούς είναι η ανοφορά του Economist ότι

> - 5 συγκεκριμένες εταιρίες (Google, Amazon, Apple, Facebook, Microsoft) συγκέντρωσαν συνολικό καθαρό κέρδος πάνω από 25 δισεκατομμύρια δολλάρια το πρώτο τρίμηνο του 2017
> - Η Amazon συγκεντρώνει μισό από το συνολικό ποσό σε δολάρια που ξοδεύεται online στην Αμερική
> - Οι Google και Facebook καρπώθηκαν σχεδόν τη συνολική αύξηση εσόδων (revenue growth) της ψηφιακής διαφήμισης στην Αμερική το 2016

το γεγονός ότι τον Απρίλιο του 2022 το 63% του παγκόσμιου πληθυσμού (περίπου 5 δισεκατομμύρια) ήταν χρήστες του διαδυκτύου, καθώς επίσης και το ότι σύμφωνα με την πλατφόρμα συλλογής και οπτικοποίησης δεδομένων Statista, η συνολική ποσότητα δεδομένων που επρόκειτο να δημιουργηθούν/αντιγραφούν/καταναλωθούν, παγκοσμίως ήταν 97 και 181 zetabytes (zetabyte = 1 billion terabytes) αντίστοιχα για το 2022 και το 2025.

Παράλληλα με τα παραπάνω, η Μηχανική Μάθηση και η Τεχνητή Νοημοσύνη παίζουν εξέχοντα ρόλο στη νέα ψηφιακή εποχή εισάγοντας νέες επαναστατικές μεθόδους στις διαδικασίες διαχείρισης, ανάλυσης και πρόβλεψης δεδομένων με αντίκτυπο σε όλους τους τομείς της ανθρώπινης δραστηριότητας από την υγεία και την οικονομία μέχρι τις μεταφορές, τις κατασκευές και αλλού, μεταξύ άλλων

- αυτοματοποιώντας διαδικασίες,
- εντοπίζοντας μοτίβα,
- βελτιστοποιώντας τη λήψη αποφάσεων,
- αυξάνοντας την παραγωγικότητα,
- μειώνοντας κόστη

Ο βασικός ρόλος που κατέχουν τα ψηφιακά δεδομένα σε αυτό το τοπίο και η δύναμη που δίνουν στους κατόχους / διαχειριστές τους, νομοτελειακά τα καθιστούν στόχο διαφόρων ειδών επιθέσεων ορισμένοι από τους οποίους συνίστανται στον αποκλεισμό της πρόσβασης σε αυτά ή και της αλλοίωσής τους με σκοπό το οικονομικό ή άλλο όφελος. Σχετικά παραδείγματα είναι η χρήση λογισμικών/μεθόδων όπως:

> - *Ransom Denial of Service (RDoS)*, όπου το κακόβουλο λογισμικό είναι σχεδιασμένο να μπλοκάρει την πρόσβαση σε ένα σύστημα μέχρι να πληρωθεί κάποιο ποσό, ή
> - *Distributed Denial of Service (DDoS)*, όπου δημιουργείται τεχνητά έντονη κίνηση σε ένα σύστημα προκαλώντας τη δυσλειτουργία του

Σύμφωνα με το GDPR, η παραβίαση δεδομένων (data breach) ορίζεται ως

> Κάθε παραβίαση της ασφάλειας που οδηγεί στην ακούσια ή έκνομη καταστροφή, απώλεια, αλλοίωση ή μη εξουσιοδοτημένη έκθεση ή πρόσβαση σε προσωπικά δεδομένα που μεταδίδονται, αποθηκεύονται ή επεξεργάζονται κατά οποιονδήποτε τρόπο.

Πέρα από τον παραπάνω ορισμό, οι απειλές στις οποίες υπόκεινται τα δεδομένα χωρίζονται σε δυο κύριες κατηγορίες που αφορούν κατά βάση τον τρόπο με τον οποίο πραγματοποιούνται:

- **Data breach**: Η ηθελημένη επίθεση προς έναν οργανισμό/υπηρεσία/εταιρία με σκοπό την απόκτηση πρόσβασης σε δεδομένα και κλοπής τους
  Παραδείγματα:
  - Μάρτιος 2020: Η CAM4 πάροχος περιεχομένου streaming όπου ο server της παραβιάστηκε εκθέτοντας 10.88 δισεκατομμύρια καταχωρήσεις.
  - Οκτώβριος 2017: Η Yahoo δέχθηκε επίθεση από χάκερς εκθέτοντας 3 δισεκατομμύρια λογαριασμούς.
  - Μάρτιος 2018: Στοιχεία όπως ονοματεπώνυμα, τραπεζικοί λογαριασμοί κ.α. 1.1 δισεκατομμυρίων ανθρώπων εκτέθηκαν από βάση βιομετρικών δεδομένων που διαχειριζόταν κρατική εταιρία στην Ινδία.
  - Ιούλιος 2022: Πάνω από 23 terabytes δεδομένων παραβιάστηκαν από τη βάση δεδομένων πελάτών της Alibaba εκθέτοντας στοιχεία όπως ονοματεπώνυμα, κωδικοί χρηστών, τηλέφωνα, διευθύνσεις κ.ά.
- **Data leak**: Η απόκτηση πρόσβασης σε δεδομένα λόγω εντοπισμού υφιστάμενης αδυναμίας στην ασφάλειά στους, ή κατόπιν της έκθεσής τους λόγω λάθους.
  Παραδείγματα

- Microsoft Power Apps (Εργαλεία δημουργίας και διαχείρισης εφαρμογών) data leak in 2021: Όπου κατέστη επιτρεπτή η δημόσια πρόσβαση σε δεδομένα που περιλάμβαναν
  - Προσωπικές πληροφορίες για εντοπισμό επαφών λόγω COVID-19,
  - Ραντεβού εμβολιασμού COVID-19
  - ΑΦΜ πολιτών οι οποίοι είχαν κάνει αίτησεις εργασίας
  - Διάφοροι κωδικοί υπαλλήλων
  - Εκατομμύρια ονοματεπώνυμων και διευθύνσεων email

Σύμφωνα με την έρευνα.pdf) της Tenable (εταιρία κυβερνοασφάλειας) το 2022 για τις παραβιάσεις δεδομένων, οι καταχωρήσεις, τα αρχεία και τα δεδομένα που εκτέθηκαν για το 2021 κι 2022 αντίστοιχα είναι 40.000.000.000 και 2.296.941.687, 1.800.000.000 και 389.127.450, 260 terabytes και 257 terabytes με τη μεγαλύτερη ποσοστιαία έκθεση ανά γεωγραφική περιοχή να έχει καταγραφεί στην Ασία (68%), ανά τομέα στις υπηρεσίες υγείας (35.4%) (λόγω της νομικής υποχρέωσης των εμπλεκομένων φορέων να αναφέρουν ανάλογα περιστατικά για λόγους διαφάνειας) και ανά είδος από Ransomware (29%) αμέσως μετά τις απροσδιόριστου είδους (33%).

Η υφιστάμενη τάση της χρήσης δεδομένων την τελευταία 10ετία αποτυπώνεται στην καταγραφή Data Never Sleeps του ιστοτόπου domo.com, σύμφωνα με την οποία παρατηρείται κατακόρυφη αύξηση από έτος 2013 έως το 2022 στη μετάδοση δεδομένων ανά λεπτό της ημέρας σε εφαρμόγές όπως

> - *Το Google*: 2 εκατομμύρια vs 5.9 εκατομμύρια queries
> - *Το Youtube*: 48 vs 500 ώρες video upload
> - *Το Instagram*: 3.6 χιλιάδες vs 66 χιλιάδες φωτογραφίες κοινοποιήθηκαν
> - *Το Tweets*: 100 χιλιάδες vs 347 χιλιάδες tweets κοινοποιήθηκαν
> - *Το Facebook*: 684 χιλιάδες vs 1.7 εκατομμύρια tweets κοινοποιήθηκαν
> - *Emails*: 204 εκατομμύρια vs 231 εκατομμύρια emails στάλθηκαν

Ο όγκος αυτός τοποθετεί τα δεδομένα στο επίκεντρο παραβιάσεων οι οποίες σύμφωνα με τον ιστότοπο καταγραφής παραβιάσεων δεδομένων SurfShark, από το 2004 έως τον Ιούνιο του 2023 αφορούν περίπου 16 δισεκατομμύρια προσωπικούς λογαριασμούς εκ των οποίων περίπου τα 6 δισεκατομμύρια είναι μοναδικές ηλεκτρονικές διευθύνσεις.
Επιπλέον κατά μέσο όρο:

> - Μια μοναδική ηλεκτρονική διεύθυνση παραβιάζεται 3 φορές,
> - Για κάθε 100 άτομα 75 μοναδικές ηλεκτρονικές διευθύνσεις παραβιάζονται και
> - 207 λογαριασμοί παραβιάζονται ανά 100 άτομα καθώς επίσης, ανά γεωγραφική επικράτεια οι περισσότερες παραβιάσεις σε προσωπικούς λογαριασμούς παρατηρούνται σε
> - Αμερική (2.767.381.100)
> - Ρωσία (2.332.257.266)
> - Κίνα (1.032.311.473)

Προκειμένου να κατανοήσει κανείς τη φύση των παραβιάσεων δεδομένων, ενδιαφέρον παρουσιάζει το να εστιάσει στα μέσα με τα οποία πραγματοποιούνται, τα κίνητρα και τους στόχους τους.
Οργανισμοί όπως το μη κερδοσκοπικό ITRC (Identity Theft Resource Center), ή εταιρίες κυβερνοασφάλειας όπως η Verison δημοσιεύουν σχετικές αναφορές (ITRC, Verison) στις οποίες μπορεί κανείς να δει λεπτομερή στοιχεία.

Απο την αναφορά της ITRC με βάση έρευνα που πραγματοποιήθηκε στην Αμερική, μπορεί να δει κανείς ότι αναφορικά με τα μέσα παραβιάσεων, το 2022 πρώτα έρχεται το phishing/smishing/BEC (461 περιστατικά) και ακολουθεί το ransomware (276 περιστατικά), όπου συνοπτικά

> - *Phishing*: Οι επιτιθέμενοι εξαπατούν χρήστες ώστε να αποκαλύψουν ευαίσθητα προσωπικά δεδομένα μέσω emails ή ψεύτικων ιστότοπων
> - *Smishing*: Οι επιτιθέμενοι εξαπατούν χρήστες ώστε να αποκαλύψουν ευαίσθητα προσωπικά δεδομένα μέσω text messages ή sms
> - *BEC (Business Email Compromise)*: Οι επιτιθέμενοι υποδυόμενοι πελάτες ή υπαλλήλους εταιρειών αποκτούν πρόσβαση σε ευαίσθητα εταιρικά δεδομένα
> - *Ransomware*: Κακόβουλα λογισμικά/μέθοδοι τύπου RDoS, DDoS που αναφέρθηκαν παραπάνω

Σύμφωνα με την ίδια έρευνα, στις βιομηχανίες/τομείς που γίνονται στόχοι επιθέσεων, πρώτο βρίσκουμε τον τομέα υγείας (344 περιστατικά), δεύτερο τον τομέα οικονομικών υπηρεσιών (268 περιστατικά) και τρίτο τον τομέα κατασκευών (249 περιστατικά)

Ως προς την εξέλιξη των παραβιάσεων σε σχέση με το χρόνο παρατηρούμε πτώση το 2022 σε σχέση με την προηγούμενη χρονιά (1802 από 1862 περιστατικά) γεγονός που ενδεχομένως να οφείλεται στην έναρξη του πολέμου στην Ουκρανία το Φεβρουάριο του 2022, κάτι που δείχνει να συμβαδίζει και με την πορεία των περιστατικών μέσα στο 2022 όπου η αυξητική τάση του δεύτερου εξαμήνου ακολούθησε την πτωτική του πρώτου.

Σε σχέση με τα κίνητρα, παρατηρεί κανείς στην δεύτερη έρευνα ότι, μάλλον αναμενόμενα, πρώτο έρχεται το προσωπικό ή οικονομικό όφελος με 96% σε όλες και 71% στις μεγάλες επιχιρήσεις ενώ δεύτερη η διαμαρτυρία/προσωπική διαφωνία με 3% και 25% αντίστοιχα. Τα μοτίβα που αναδεικνύονται εδώ είναι ότι

> - Οι περισσότερες επιθέσεις πραγματοποιούνται από επαγγελματίες που μετατρέπουν την πληροφορία σε κέρδος
> - Υπάρχει σημαντική παρουσία του ακτιβισμού μέσω επιθέσεων σε μεγάλες επιχειρήσεις

Στα μέσα/εργαλεία με τα οποία πραγματοποιούνται κυβερνοεπιθέσεις εμπίπτουν και τα

> - *Identity Abuse*: Όπου ο επιτιθέμενος κάνει χρήση της ταυτότητας/των αναγνωριστικών κάποιου νόμιμου χρήστη προκειμένου να αποκτήσει πρόσβαση σε κάποιο σύστημα, όπως για παράδειγμα με το να αποσπάσει τα αναγνωριστικά ενός χρήστη με phising.
> - *Living-off-the-Land (LOTL) Tactic*: Όπου ο επιτιθέμενος αντί να εγκατήσει κακόβουλο λογισμικό, χρησιμοποιεί υπάρχοντα εργαλεία/εφαρμογές σε ένα σύστημα προκειμένου να πραγματοποιήσει κακόβουλες δραστηριότητες, όπως για παράδειγμα την εκτέλεση shell scripts από τα Shells του λειτουργικού συστήματος.

Επιπλέον, ως προς την έκθεση/χρήση προσωπικών δεδομένων γίνεται η ακόλουθη διάκριση:

> - *Misused identity*: Όταν υπάρχουν στοιχεία που καταδεικνύουν ότι τα προσωπικά δεδομένα κάποιου έχουν χρησιμοποιηθεί κακόβουλα.
> - *Compromised identity*: Όταν τα προσωπικά δεδομένα κάποιου έχουν εκτεθεί και ο κίνδυνος κακόβουλης χρήσης τους είναι υπαρκτός αλλά όχι αποδεδειγμένος

Σε σχέση με τα παραπάνω και με βάση την αναφορά Trends in Identity Report, επίσης της ITRC, μεταξύ άλλων ευρυμάτων παρατηρούμε, ότι

> - 40% των καταναλωτών αναφέρουν ότι προσωπικά τους δεδομένα έχουν κλαπεί ή χρησιμοποιηθεί κακόβουλα το 2022
> - Οι περισσότερες αναφορές το 2021 και το 2022 για compromised identity αφορούσαν την Google Voice (53% – 3.926 και 61% – 4.081 θύματα αντίστοιχα)
> - Στην κακόβουλη χρήση ταυτότητας μεσω υπάρχοντος λογαριασμού έναντι δημιουργίας νέου, πρώτα έρχονται τα κοινωνικά δίκτυα έναντι των πιστωτικών καρτών

Λαμβάνονται μέτρα για την αντιμετώπιση των επιθέσεων;

Για να ληφθούν μέτρα είναι απαραίτητη η δημιουργία μιας αντιπροσωπευτικής εικόνας του φαινομένου, και αυτό είναι εφικτό όταν αυτά τα φαινόμενα αναφέρονται και καταγράφονται.

Ποια είναι η γνώση μας σχεικά με αυτά;

Η ITRC δίνει πάλι μια εικόνα της κατάστασης η οποία δείχνει να μην είναι ελπιδοφόρα, καθώς το 2022 οι δημόσιες αναφορές περιστατικών στην Αμερική με πληροφορίες για το θύμα και την επίθεση κατέβηκαν στο 34%, νούμερο το οποία είναι το μικρότερο της τελευταίας 5ετίας, κάτι που δείχνει

> - Γενικότερα πόσο ευάλωτοι είναι οι πολίτες και οι επιχειρήσεις/οργανισμοί σε τέτοια φαινόμενα
> - Ειδικότερα πόσο δύσκολο είναι να εκτιμήσουν την κατάσταση και να λάβουν μέτρα αφού στα 2/3 περίπου των αναφορών λείπει η απαραίτητη πληροφορία για να το κάνουν

Πιθανοί λόγοι για το παραπάνω γεγονός σύμφωνα με την έρευνα μπορεί να είναι:

- Αποφάσεις διαφόρων δικαστηρίων στην Αμερική δίνουν την δυνατότητα στις εταιρίες να κοινοποιούν μια ελάχιστη ποσότητα πληροφορίας για σχετικά περιστατικά
- Εταιρίες συνειδητά αποφασίζουν να μην αποκαλύψουν αρκετές πληροφορίες για περιστατικά έκθεσης δεδομένων (Παράδειγμα οι εταιρίες Samsung, DoorDash, LastPass των οποίων δεδομένα εκτέθηκαν το 2022 και αποφάσισαν να αποκρύψουν πληροφορίες για το περιστατικό)
- Υπάρχει εγγενής δυσκολία στον εντοπισμό της αιτίας και της εκτίμησης των συνεπειών τέτοιων περιστατικών λόγω του αυξανόμενου όγκου των κυβερνοεπιθέσεων
  - (Διαμεσος ημερών μέχρι τον εντοπισμό ενός περιστατικού είναι 207 ημέρες σύμφωνα με την έρευνα IBM Cost of Data Breach Report 2022)

Μια τάση η οποία τοποθετεί την ανάγκη λήψης μέτρων ενάντια στις επιθέσεις σε ευρύτερο πλαίσιο είναι αυτή της μεταφοράς υποδομών και δεδομένων υπηρεσιών οργανισμών και εταιριών σε υπηρεσίες cloud. Παρόλο το ότι η χρήση υπηρεσιών cloud αυξάνεται, τα σχετικά μέτρα ασφάλειας στις υποδομές των υπηρεσιών αυτών φαίνεται να μην ακολουθούν την ίδια πορεία.
Σύμφωνα με την έρευνα Thales, 2023 Data Threat Report της Thalesgroup (υπηρεσία παροχής υπηρεσιών ασφάλειας δεδομένων), το 2021 μόνο 17% των ερωτώμενων απάντησε ότι πάνω 50% των ευαίσθητων δεδομένων τους σε υποδομές cloud είναι κρυπτογραφημένο, ενώ το 2022 μόνο 22% για πάνω από 60% που δείχνει ότι μεγάλος όγκος δεδομένων εξακολουθεί να παραμένει ευάλωτος σε επιθέσεις.

Πιθανές αιτίες γιαυτό δείχνουν να είναι

- η έλλειψη σχετικά καταρτισμένου προσωπικού
- η δυσκολία/πολυπλοκότητα της εφαρμογής μέτρων ασφάλειας σε υποδομές cloud

αν λάβει κανείς υπόψη ότι σύμφωνα με την ίδια έρευνα

> - το 51% των ερωτώμενων απάντησε ότι αντιμετωπίζει μεγαλύτερη δυσκολία στη διαχείριση ασφάλειας σε cloud υπηρεσίες σε σχέση με on-premise υποδομές
> - το 79% των ερωτώμενων απάντησε ότι χρησιμοποιεί υπηρεσίες cloud από πολλαπλούς παρόχους

αλλά και ότι με βάση την έρευνα *In 451 Research's VotE: Information Security, Organizational Dynamics 2021* υπάρχει σημαντική

έλλειψη ειδίκευσης σε θέματα ασφαλείας για πλατφόρμες cloud.

Την ανάγκη λήψης μέτρων προστασίας δεδομένων, την ευαλωτότητα τους και την έκταση των συνεπειών που μπορεί να έχει μια επίθεση δείχουν και τα περιστατικά έκθεσης δεδομένων από εταιρείς κολοσσούς παρόχους υπηρεσιών cloud:

> - Περίπτωση Microsoft 2022: όπου προβληματική παραμετροποίηση στο Azure εξέθεσε 2.4 terabytes δεδομένων που ανήκαν σε 150.000 εταιρίες από 123 χώρες με υπογεγραμμένα έγγραφα, emails και άλλες ευαίσθητες πληροφορίες
> - Περίπτωση Amazon 2022: όπου βάση δεδομένων μη προστατευμένη με κωδικό ήταν δημόσια προσβάσιμη και εξέθεσε 215 εκατομμύρια εγγραφές με πληροφορίες streaming της Prime σχετικές με τη συνδρομή, τη συσκευή μετάδοσης και άλλα.

Με την εδραίωση της Μηχανικής Μάθησης προστίθενται νέες κατηγορίες απειλών/επιθέσεων για τα δεδομένα, οι οποίες σχετίζονται με τη χρήση της στους διάφορους τομείς της ψηφιακής δραστηριότητας.
Μια τέτοια κατηγορία είναι το **Data manipulation**, το οποίο με τη σειρά του περιλαμβάνει υποκατηγορίες όπως οι

> - **Data poisoning**: Επιθέσεις κατά τις οποίες αλλοιώνονται τα δεδομένα εκπαίδευσης με σκοπό να επηρεάσουν την ακρίβεια των προβλέψεων των μοντέλων
> - **Adversarial attacks**: Επιθέσεις κατά τις οποίες αλλοιώνονται τα δεδομένα εισόδου κατά τη διαδικασία πρόβλεψης μοντέλων σε νέα δεδομένα.
> - **Information manipulation**: Συνίσταται στη δημιουργία σύγχυσης/ψευδών εντυπώσεων σε ομάδα ατόμων μέσω της διάδοσης διαστρεβλωμένης ή αλλοιωμένης πληροφορίας.

Επιπλέον, η εφαρμογή της σε εργαλεία τα οποία σχετίζονται με ολοένα και μεγαλύτερους όγκους δεδομένων όπως τα chatbots (ChatGPT, Google Bard κλπ), καταδυκνύει τους με αντίστοιχο ρυθμό αυξανόμενους κινδύνους που προκύπτουν και τον αντίκτυπο που μπορεί να έχει η έκθεση δεδομέων τόσο μεγάλης έκτασης, καθώς και τη δύναμη που μπορεί να δώσει σε όσους θα εκμεταλλευτούν τέτοια περιστατικά.

Προς στήριξη των ανωτέρω, ενδεικτικά μπορεί κανείς να αναφέρει ότι

- Σύμφωνα με τον Florian Tramèr, αναπληρωτή καθηγητή του τμήματος επιστήμης υπολογιστών του πανεπιστημίου της Ζυρίχης,
  - για να εκπαιδευτούν μοντέλα όπως αυτά που χρησιμοποιούνται στα chatbots, χρειάζονται ουσιαστικά ένα τεράστιο μέρος των δεδομένων ολόκληρου του διαδικτύου.
  - αν όταν τα μοντέλα αυτά χρησιμοποιούνται σε μηχανές αναζήτησης, κάποια παρέμβαση στη διαδικασία παραγωγής αποτελέσματός τους, τα οδηγήσει σε μεροληψία, το οικονομικό όφελος όποιου το εκμεταλλετεί θα είναι τεράστιο
- Η έκθεση δεδομένων αυτής τη έκτασης δεν είναι υποθετικό σενάριο, αλλά πραγματικό γεγονός, καθώς η OpenAI, δημιουργός εταιρία του ChatGPT, επιβεβαίωσε το 2023 ότι ένα bug σε κάποια βιβλιοθήκη της οδήγησε σε έκθεση δεδομένων κατά την οποία χρήστες μπορούσαν να δουν τμήματα του ιστορικού συνομιλίας με το bot άλλων χρηστών εφόσων ήταν και οι δυό ενεργοί την ίδια στιγμή, καθώς επίσης και πληροφορίες πληρωμής που περιλάμβαναν
- Ονοματεπώνυμα
- Emails
- Διευθύνσεις πληρωμής
- Τύπους πιστωτικής κάρτας όπως και τμήματα του αριθμού τους