



Τμήμα Πληροφορικής

Εαρινό εξάμηνο 2024-25

ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ – ΓΡΑΠΤΗ ΕΡΓΑΣΙΑ

Η χρησιμότητα της
Τεχνητής Νοημοσύνης
στην γνώση της
Κυβερνοασφάλειας

ΕΛΕΝΑ ΣΑΚΚΗ

ΚΩΝΣΤΑΝΤΙΝΑ ΚΑΡΑΠΕΤΣΑ

ΦΩΤΕΙΝΗ ΣΩΤΗΡΟΠΟΥΛΟΥ

Περιεχόμενα

ΠΕΡΙΛΗΨΗ.....	2
ΜΕΡΟΣ Α.....	3
Παράρτημα: Prompts που χρησιμοποιήθηκαν	5
ΜΕΡΟΣ Β.....	6
Αξιολόγηση του μοντέλου Generative AI (ChatGPT).....	6
Απάντηση ανταγωνιστικού μοντέλου Generative AI (Deepseek)	9
Σύγκριση των δύο μοντέλων Generative AI.....	11
ΒΙΒΛΙΟΓΡΑΦΙΑ	13

ΠΕΡΙΛΗΨΗ

Οι phishing επιθέσεις είναι μια διαδεδομένη κυβερνοαπειλή, όπου εγκληματίες χρησιμοποιούν ψεύτικα emails, μηνύματα ή ιστοσελίδες, προσποιούμενοι νόμιμες οντότητες (π.χ. τράπεζες, κρατικούς φορείς), για να εξαγάγουν ευαίσθητες πληροφορίες (κωδικούς, πιστωτικά στοιχεία) ή να εγκαταστήσουν malware. Οι επιθέσεις ποικίλλουν από μαζικές αποστολές με γενικό περιεχόμενο έως εξατομικευμένο spear phishing, που στοχεύει συγκεκριμένα άτομα ή οργανισμούς με χρήση προσωπικών δεδομένων. Βασικές τεχνικές περιλαμβάνουν κοινωνική μηχανική (π.χ. ψεύτικες επείγουσες ειδοποιήσεις) και ψεύτικα URLs, με σοβαρές συνέπειες: οικονομικές απώλειες, διαρροές δεδομένων, ζημιά στη φήμη. Η προστασία απαιτεί ευαισθητοποίηση χρηστών, τεχνικούς ελέγχους (φίλτρα spam, επαλήθευση διευθύνσεων) και πολυπαραγοντική πιστοποίηση (MFA). Παρά τα μέτρα, οι επιθέσεις αυξάνονται συνεχώς, καθιστώντας τη συνεχή εκπαίδευση και την ενημέρωση, κρίσιμους παράγοντες για την αποτελεσματική άμυνα.

ΜΕΡΟΣ Α

Εισαγωγή

Οι phishing επιθέσεις αποτελούν μία από τις πιο διαδεδομένες και επικίνδυνες μορφές κυβερνοεγκλήματος. Στοχεύουν στην εξαπάτηση χρηστών με σκοπό την υποκλοπή προσωπικών και ευαίσθητων πληροφοριών, όπως κωδικοί πρόσβασης, οικονομικά στοιχεία και διαπιστευτήρια. Οι επιθέσεις αυτού του είδους βασίζονται κυρίως στις αρχές της κοινωνικής μηχανικής, εκμεταλλευόμενες την ανθρώπινη ψυχολογία και την εμπιστοσύνη των χρηστών. Οι επιτιθέμενοι προσπαθούν να πείσουν τα θύματά τους να αποκαλύψουν πολύτιμες πληροφορίες ή να εκτελέσουν κακόβουλες ενέργειες χωρίς να καταλάβουν την απάτη. Από την πλευρά της κυβερνοασφάλειας, οι phishing επιθέσεις αποτελούν σοβαρή απειλή, διότι μπορούν να παρακάμψουν τα παραδοσιακά φίλτρα ασφαλείας και να εκμεταλλευτούν τις αδυναμίες των χρηστών.

Θεωρητική Τεκμηρίωση

Το social engineering αποτελεί την κύρια αιτία πίσω από τις phishing επιθέσεις. Αντί να παραβιάσουν ένα σύστημα μέσω τεχνικών μέσων, οι επιτιθέμενοι στοχεύουν στον ανθρώπινο παράγοντα, εκμεταλλευόμενοι τις ψυχολογικές αδυναμίες των θυμάτων τους. Οι επιθέσεις αυτές βασίζονται σε συναισθηματικούς παράγοντες, όπως ο φόβος, η βιασύνη και η εμπιστοσύνη, προκειμένου να καταφέρουν να εξαπατήσουν τα θύματα και να αποσπάσουν προσωπικές πληροφορίες. Η ανθρώπινη ψυχολογία καθιστά τις επιθέσεις κοινωνικής μηχανικής ιδιαίτερα επικίνδυνες, καθώς είναι δύσκολο να αναγνωριστούν με παραδοσιακές τεχνικές ασφαλείας.

Από την πλευρά της κυβερνοασφάλειας, οι phishing επιθέσεις αποτελούν μια από τις πιο επικίνδυνες μορφές απειλών για τις προσωπικές και εταιρικές πληροφορίες. Οι επιθέσεις αυτές παρακάμπτουν συχνά τα τεχνικά συστήματα προστασίας, όπως τα φίλτρα spam, και στοχεύουν απευθείας στον χρήστη. Ο επιτιθέμενος συνήθως προσποιείται ότι προέρχεται από μια αξιόπιστη πηγή, όπως μια τράπεζα, μια υπηρεσία cloud ή έναν δημόσιο οργανισμό, με σκοπό να παραπλανήσει το θύμα και να το κάνει να αποκαλύψει προσωπικά δεδομένα.

Συνήθεις Τεχνικές Phishing και Πώς Λειτουργούν

Οι επιτιθέμενοι χρησιμοποιούν διάφορες τεχνικές για να παραπλανήσουν τους χρήστες και να αποκτήσουν τα δεδομένα τους. Μία από τις πιο κοινές τεχνικές είναι το email phishing, όπου ο επιτιθέμενος στέλνει μαζικά email που προσποιούνται ότι προέρχονται από αξιόπιστους οργανισμούς, όπως τράπεζες ή εταιρείες. Στο μήνυμα αυτό, ο χρήστης καλείται να κλικάρει σε έναν σύνδεσμο που οδηγεί σε μια ψεύτικη ιστοσελίδα, που μοιάζει με την αυθεντική. Εκεί, ο χρήστης ζητείται να εισάγει τα διαπιστευτήριά του, τα οποία στη συνέχεια αποστέλλονται στον επιτιθέμενο.

Μια άλλη τεχνική είναι το spear phishing, το οποίο είναι πιο στοχευμένο από το κλασικό phishing. Στην περίπτωση αυτή, ο επιτιθέμενος αποκτά προσωπικές πληροφορίες για το θύμα, είτε από δημόσιες πηγές είτε μέσω προηγούμενων επιθέσεων, και στέλνει ένα εξαιρετικά εξατομικευμένο μήνυμα. Αυτή η τεχνική αποσκοπεί στο να πείσει το θύμα ότι το μήνυμα είναι πραγματικό και προέρχεται από μια αξιόπιστη πηγή, όπως συνάδελφος ή συνεργάτης.

Το whaling είναι μια ακόμη τεχνική phishing, η οποία στοχεύει σε υψηλόβαθμα στελέχη, όπως διευθυντές ή γενικοί διευθυντές. Στην περίπτωση αυτή, ο επιτιθέμενος προσποιείται ότι είναι κάποιος από την εταιρεία ή ένας εξωτερικός συνεργάτης και ζητάει από το θύμα να εκτελέσει μια επείγουσα ενέργεια, όπως την μεταφορά χρημάτων ή την αποστολή εμπιστευτικών εγγράφων.

Αξιοσημείωτο είναι επίσης το smishing, το οποίο χρησιμοποιεί SMS μηνύματα αντί για email. Ο επιτιθέμενος στέλνει μηνύματα με παραπλανητικούς συνδέσμους ή πληροφορίες, υποστηρίζοντας ότι προέρχονται από τράπεζες ή υπηρεσίες, προκειμένου να παρασύρει το θύμα σε κακόβουλες ενέργειες. Ομοίως, το vishing, που είναι το τηλεφωνικό phishing, βασίζεται σε κλήσεις από ψεύτικους υπαλλήλους τράπεζας ή άλλων οργανισμών με στόχο να αποσπάσουν προσωπικές πληροφορίες ή χρήματα από το θύμα.

Σύγχρονα Παραδείγματα Phishing Επιθέσεων

Οι phishing επιθέσεις έχουν προκαλέσει σημαντικές ζημιές σε πολλές περιπτώσεις. Ένα χαρακτηριστικό παράδειγμα είναι η επίθεση στο Twitter το 2020, όταν επιτιθέμενοι κατάφεραν να παραβιάσουν λογαριασμούς διάσημων προσώπων μέσω κοινωνικής μηχανικής. Ο επιτιθέμενος απέκτησε πρόσβαση σε εσωτερικά εργαλεία της εταιρείας και, στη συνέχεια, δημοσίευσε ψευδή μηνύματα και ζητούσε χρήματα από τους ακολούθους των στόχων του.

Άλλο ένα παράδειγμα είναι η επίθεση στο Facebook και την Google, οι οποίες εξαπατήθηκαν μέσω πλαστών τιμολογίων για να καταβάλλουν περισσότερα από 100 εκατομμύρια δολάρια σε έναν Λιθουανό απατεώνα. Στην περίπτωση αυτή, ο επιτιθέμενος χρησιμοποίησε τις τεχνικές του phishing για να καταφέρει να εξαπατήσει τις εταιρείες, δημιουργώντας ψεύτικες εταιρείες και τιμολόγια.

Τέλος, κατά τη διάρκεια της πανδημίας COVID-19, σημειώθηκαν αρκετές επιθέσεις phishing που εκμεταλλεύτηκαν την αυξημένη ανησυχία των ανθρώπων για τη δημόσια υγεία. Οι επιτιθέμενοι έστειλαν email που προσποιούνταν ότι προέρχονταν από οργανισμούς υγείας ή κυβερνητικούς φορείς, ζητώντας προσωπικές πληροφορίες ή χρηματικές δωρεές για δήθεν υποστήριξη.

Αντιμετώπιση & Πρόληψη Phishing Επιθέσεων

Η πρόληψη των phishing επιθέσεων απαιτεί ένα συνδυασμό τεχνικών μέτρων και ανθρώπινης εκπαίδευσης. Πρώτα απ' όλα, η πολυπαραγοντική αυθεντικοποίηση (MFA) αποτελεί έναν εξαιρετικά αποτελεσματικό τρόπο για την ενίσχυση της ασφάλειας των λογαριασμών, καθώς απαιτεί δύο ή περισσότερους παράγοντες επαλήθευσης προτού επιτραπεί η πρόσβαση. Επιπλέον, η εκπαίδευση των χρηστών είναι κρίσιμη για την αναγνώριση ύποπτων μηνυμάτων και την αποφυγή επικίνδυνων ενεργειών.

Οι οργανισμοί θα πρέπει να εφαρμόζουν πολιτικές Zero Trust, όπου κάθε χρήστης, ακόμα και εντός του δικτύου, ελέγχεται συνεχώς για να εξασφαλιστεί ότι δεν υπάρχει κίνδυνος. Επίσης, η χρήση εργαλείων email filtering και DNS filtering μπορεί να μειώσει σημαντικά τον κίνδυνο phishing.

Μελλοντικές Τάσεις στην Ανίχνευση και Πρόληψη Phishing

Οι μελλοντικές τάσεις στον τομέα της ανίχνευσης και πρόληψης των phishing επιθέσεων επικεντρώνονται στην Τεχνητή Νοημοσύνη (AI) και τα Μηχανήματα Μάθησης (ML). Η Τεχνητή Νοημοσύνη χρησιμοποιείται για την ανάλυση τεράστιου όγκου δεδομένων και την αναγνώριση

μοτίβων που δείχνουν ύποπτη συμπεριφορά, όπως η αλλαγή στην γλώσσα των email ή η ταχύτητα πληκτρολόγησης του χρήστη.

Επιπλέον, η ανάλυση συμπεριφοράς χρηστών (UBA) και η τεχνολογία deepfake αναμένονται να προσφέρουν νέες λύσεις στον τομέα της ανίχνευσης και αποτροπής phishing επιθέσεων, αναγνωρίζοντας ψευδείς φωνές ή εικόνες σε επικοινωνίες υψηλού επιπέδου, όπως το CEO Fraud.

Η χρήση της τεχνολογίας Blockchain για την επικύρωση της αυθεντικότητας των email αποστολών είναι μια άλλη καινοτόμος τάση που αναμένεται να μειώσει την επιτυχία των επιθέσεων phishing.

Συμπεράσματα

Οι phishing επιθέσεις συνεχίζουν να εξελίσσονται και να προσαρμόζονται στις νέες τεχνολογίες και τις ανθρώπινες αδυναμίες. Για την αποτελεσματική αντιμετώπισή τους, απαιτείται η χρήση σύγχρονων εργαλείων, όπως η Τεχνητή Νοημοσύνη, σε συνδυασμό με εκπαιδευτικά προγράμματα και πολιτικές ασφαλείας, όπως το Zero Trust. Ο συνδυασμός αυτών των τεχνολογιών και διαδικασιών θα αποτελέσει την πιο αποτελεσματική άμυνα κατά των επιθέσεων phishing στο μέλλον.

Παράρτημα: Prompts που χρησιμοποιήθηκαν

- Prompt 1: «Γράψε μου θεωρητική τεκμηρίωση για το φαινόμενο των phishing επιθέσεων, με βάση τις αρχές της κοινωνικής μηχανικής και της κυβερνοασφάλειας.»
 - Prompt 2: «Παρουσίασε παραδείγματα σύγχρονων phishing επιθέσεων και πώς επηρέασαν εταιρείες ή χρήστες.»
 - Prompt 3: «Ποιες είναι οι πιο συνηθισμένες τεχνικές phishing σήμερα και πώς λειτουργούν;»
 - Prompt 4: «Ποιες είναι οι μελλοντικές τάσεις στην ανίχνευση ή πρόληψη phishing; Πώς εμπλέκεται η Τεχνητή Νοημοσύνη;»
 - Prompt 5: «Χρησιμοποιώντας όλες τις απαντήσεις γράψε μία εργασία για τις phishing επιθέσεις.»
-

ΜΕΡΟΣ Β

Αξιολόγηση του μοντέλου Generative AI (ChatGPT)

Εισαγωγή

Η εισαγωγή της εργασίας που δημιουργήθηκε με το μοντέλο Generative AI GPT, παρουσιάζει τις phishing επιθέσεις ως μία από τις πιο διαδεδομένες και επικίνδυνες μορφές κυβερνοεγκλήματος, οι οποίες βασίζονται στην κοινωνική μηχανική και στοχεύουν στην εξαπάτηση των χρηστών με σκοπό την υποκλοπή ευαίσθητων πληροφοριών. Οι περιγραφές αυτές βρίσκονται σε συμφωνία με τις διαπιστώσεις του Ευρωπαϊκού Οργανισμού για την Ασφάλεια Δικτύων και Πληροφοριών (ENISA), ο οποίος στην ετήσια έκθεσή του για τις απειλές στον κυβερνοχώρο επισημαίνει ότι το phishing

αποτελεί σταθερά μία από τις κορυφαίες απειλές, αξιοποιώντας κυρίως τεχνικές κοινωνικής μηχανικής για να παρακάμψει την τεχνολογική άμυνα των οργανισμών και να στοχεύσει απευθείας τον ανθρώπινο παράγοντα (ENISA, 2023). Η εν λόγω έκθεση επιβεβαιώνει ότι οι phishing επιθέσεις έχουν σημαντικό αντίκτυπο, όχι μόνο σε μεμονωμένους χρήστες, αλλά και σε οργανισμούς, καθώς συχνά αποτελούν την αρχική φάση σε πιο σύνθετες επιθέσεις, όπως ransomware ή παραβιάσει δεδομένων. Η αξιολόγηση των κινδύνων από την ENISA ενισχύει την άποψη ότι η ανθρώπινη ψυχολογία αποτελεί βασικό πεδίο εκμετάλλευσης, γεγονός που καθιστά τις phishing επιθέσεις δύσκολες στην πρόληψη, παρά τα τεχνολογικά μέσα που διαθέτουν οι οργανισμοί.

Συμπερασματικά, η εισαγωγή αποδίδει με ακρίβεια τα βασικά χαρακτηριστικά του φαινομένου και είναι πλήρως εναρμονισμένη με τις σύγχρονες επιστημονικές και βιομηχανικές αναλύσεις της απειλής.

Θεωρητική Τεκμηρίωση

Το συγκεκριμένο κομμάτι υπογραμμίζει ότι το social engineering αποτελεί την κύρια αιτία πίσω από τις phishing επιθέσεις, με τους επιτιθέμενους να στοχεύουν στον ανθρώπινο παράγοντα και να εκμεταλλεύονται ψυχολογικές αδυναμίες, όπως ο φόβος, η βιασύνη και η εμπιστοσύνη, για να αποσπάσουν προσωπικές πληροφορίες. Επισημαίνεται επίσης ότι αυτές οι επιθέσεις είναι δύσκολο να αναγνωριστούν με παραδοσιακές τεχνικές ασφάλειας, καθώς παρακάμπτουν τα τεχνικά συστήματα προστασίας και στοχεύουν απευθείας στον χρήστη, συχνά προσποιούμενοι αξιόπιστες πηγές. Αυτή η περιγραφή συνάδει με τα ευρήματα του Butavicius et al. (2016) στην εργασία τους "Breaching the Human Firewall: Social engineering in Phishing and Spear-Phishing Emails". Οι συγγραφείς εξετάζουν την επιρροή τριών στρατηγικών κοινωνικής μηχανικής—αυθεντία, έλλειψη και κοινωνική απόδειξη—στις κρίσεις των χρηστών σχετικά με την ασφάλεια του κλικ σε έναν σύνδεσμο σε ένα email. Διαπιστώνουν ότι η χρήση της αρχής της αυθεντίας είναι η πιο αποτελεσματική στρατηγική για να πείσει τους χρήστες ότι ένας σύνδεσμος σε ένα email είναι ασφαλής. Επιπλέον, σημειώνουν ότι οι χρήστες δυσκολεύονται να διακρίνουν μεταξύ γνήσιων και spear-phishing emails, ιδιαίτερα όταν χρησιμοποιείται η αρχή της αυθεντίας. Αυτό υπογραμμίζει την εκμετάλλευση της ανθρώπινης ψυχολογίας από τους επιτιθέμενους για την παράκαμψη των τεχνικών μηχανισμών ασφαλείας και την επίτευξη μη εξουσιοδοτημένης πρόσβασης σε ευαίσθητες πληροφορίες.

Συνήθεις Τεχνικές Phishing και Πώς Λειτουργούν

Εδώ περιγράφονται με ακρίβεια οι πιο συχνά χρησιμοποιούμενες τεχνικές phishing, όπως το ηλεκτρονικό phishing (email phishing), το στοχευμένο phishing (spear phishing), το whaling, το smishing (SMS phishing) και το vishing (voice phishing), οι οποίες αποσκοπούν στην εξαπάτηση χρηστών και την κλοπή ευαίσθητων δεδομένων. Οι πληροφορίες αυτές ευθυγραμμίζονται πλήρως με την επιστημονική βιβλιογραφία, όπως αναφέρεται στις έρευνες των Chin (2024) και Fortinet (2024). Το email phishing βασίζεται σε μαζικά ψεύτικα μηνύματα που μιμούνται αξιόπιστες πηγές, ενώ το spear phishing χρησιμοποιεί εξατομικευμένες πληροφορίες για μεγαλύτερη αξιοπιστία. Το whaling στοχεύει σε ανώτερα στελέχη με πλαστά επείγοντα αιτήματα, και το smishing και vishing εκμεταλλεύονται κείμενα και τηλεφωνικές κλήσεις αντίστοιχα για να εξαπατήσουν τα θύματα (Chin

2024, Fortinet 2024). Η περιγραφή αυτή είναι ακριβής και ανταποκρίνεται στις τρέχουσες τάσεις στον τομέα της κυβερνοασφάλειας.

Σύγχρονα Παραδείγματα Phishing Επιθέσεων

Οι phishing επιθέσεις αποτελούν έναν διαρκώς αυξανόμενο κίνδυνο, με σύγχρονα περιστατικά να επιβεβαιώνουν τη σοβαρότητα του φαινομένου. Το 2020, το Twitter υπέστη σοβαρή παραβίαση όταν επιτιθέμενοι, μέσω τεχνικών κοινωνικής μηχανικής και phishing, απέκτησαν πρόσβαση σε εσωτερικά εργαλεία της πλατφόρμας και δημοσίευσαν παραπλανητικά μηνύματα από λογαριασμούς διασήμων, ζητώντας κρυπτονομίσματα (BBC News, 2020). Επιπλέον, οι εταιρείες Facebook και Google εξαπατήθηκαν μεταξύ 2013 και 2015, καταβάλλοντας περισσότερα από 100 εκατομμύρια δολάρια σε έναν απατεώνα μέσω πλαστών τιμολογίων που στηρίζονταν σε phishing πρακτικές (Boden et al., 2018). Τέλος, κατά τη διάρκεια της πανδημίας COVID-19, σημειώθηκε αύξηση στις phishing επιθέσεις που εκμεταλλεύονταν τον φόβο και την αβεβαιότητα, με απατηλά emails να εμφανίζονται ως επίσημη επικοινωνία από φορείς υγείας όπως ο ΠΟΥ (Interpol, 2020). Τα παραδείγματα αυτά υπογραμμίζουν την ανάγκη για αυξημένη επαγρύπνηση απέναντι σε τέτοιες απειλές.

Αντιμετώπιση & Πρόληψη Phishing Επιθέσεων

Το κείμενο αναφέρει σωστά ότι η πρόληψη των phishing επιθέσεων βασίζεται σε έναν συνδυασμό τεχνολογικών μεθόδων και ανθρώπινης ευαισθητοποίησης. Συγκεκριμένα, η πολυπαραγοντική αυθεντικοποίηση (MFA) αναγνωρίζεται ως ένα από τα πιο αποτελεσματικά μέτρα, καθώς δυσκολεύει σημαντικά την πρόσβαση σε λογαριασμούς ακόμα και αν διαρρεύσουν κωδικοί πρόσβασης. Αυτό επιβεβαιώνεται από τον Naseer (2024), ο οποίος τονίζει ότι η MFA μειώνει το ποσοστό επιτυχίας των phishing επιθέσεων κατά πάνω από 90%. Επιπλέον, η εκπαίδευση των χρηστών είναι κρίσιμη, καθώς πολλές επιθέσεις εξακολουθούν να βασίζονται σε κοινωνική μηχανική. Ωστόσο, το κείμενο θα μπορούσε να αναφερθεί σε πιο προηγμένα εργαλεία AI-based εκπαίδευσης, όπως τα simulated phishing tests, τα οποία αναλύει ο Karthik (2025). Αυτά τα εργαλεία χρησιμοποιούν τεχνητή νοημοσύνη για να δημιουργούν ρεαλιστικά σενάρια phishing και να μετρούν την ανταπόκριση των χρηστών, βελτιώνοντας συνεχώς την εκπαίδευση. Άλλο ένα σημαντικό μέτρο που αναφέρεται είναι η πολιτική Zero Trust, η οποία απαιτεί συνεχή επαλήθευση των χρηστών, ακόμα και εντός του εταιρικού δικτύου. Ο Naseer (2024) επισημαίνει ότι αυτή η προσέγγιση, σε συνδυασμό με AI-driven anomaly detection, μπορεί να εντοπίσει ύποπτη δραστηριότητα σε πραγματικό χρόνο. Το κείμενο, ωστόσο, θα μπορούσε να αναπτύξει περισσότερο τον ρόλο της τεχνητής νοημοσύνης στην αυτοματοποιημένη ανίχνευση απειλών, κάτι που οι πηγές θεωρούν βασικό. Τέλος, η χρήση email filtering και DNS filtering αναφέρεται ως κρίσιμη για τη μείωση του όγκου των phishing μηνυμάτων. Ωστόσο, σύμφωνα με τον Karthik (2025), τα σύγχρονα συστήματα βασίζονται όλο και περισσότερο σε μηχανική μάθηση για την ανάλυση της γλώσσας, των συνδέσμων και ακόμα και της συμπεριφοράς του αποστολέα, πράγμα που το κείμενο θα μπορούσε να τονίσει περισσότερο.

Μελλοντικές Τάσεις στην Ανίχνευση και Πρόληψη Phishing

Οι εξελίξεις στον τομέα της αντι-phishing άμυνας επικεντρώνονται όλο και περισσότερο στην Τεχνητή Νοημοσύνη (AI) και τα Μηχανήματα Μάθησης (ML), κάτι που το κείμενο αναγνωρίζει σωστά. Συγκεκριμένα, σύμφωνα με τον Naseer (2024), τα συστήματα AI μπορούν να αναλύσουν τεράστιες ποσότητες δεδομένων για να εντοπίσουν μοτίβα που υποδεικνύουν phishing, όπως ασυνήθιστες αλλαγές στη γλώσσα των email ή ύποπτες IP διευθύνσεις. Μια σημαντική τάση που αναφέρεται είναι η ανάλυση συμπεριφοράς χρηστών (UBA), η οποία παρακολουθεί πώς συνδέονται οι χρήστες, πόσο γρήγορα πληκτρολογούν ή ακόμα και πώς χρησιμοποιούν το ποντίκι. Αυτό επιτρέπει την ανίχνευση ασυνήθιστων δραστηριοτήτων, ακόμα και αν ο λογαριασμός έχει παραβιαστεί. Επιπλέον, η ανάπτυξη της τεχνολογίας deepfake γίνεται όλο και πιο σημαντική, καθώς οι

εγκληματίες χρησιμοποιούν ψεύτικες φωνές και βίντεο για CEO fraud και άλλες εξελιγμένες επιθέσεις. Άλλη μια καινοτόμος τάση που αναφέρεται είναι η χρήση blockchain για την επικύρωση της ταυτότητας των αποστολέων. Ωστόσο, ενώ αυτή η τεχνολογία έχει δυνατότητες, ο Karthik (2025) επισημαίνει ότι η εφαρμογή της είναι ακόμα σε πειραματικό στάδιο και ότι η AI παραμένει η κύρια λύση για την πρόληψη phishing.

Συμπεράσματα

Το συμπέρασμα του κειμένου ότι η συνδυασμένη χρήση τεχνολογίας (AI, MFA, Zero Trust) και εκπαίδευσης είναι απαραίτητη για την καταπολέμηση του phishing είναι απολύτως σωστό και ευθυγραμμισμένο με τις πηγές. Ωστόσο, θα μπορούσε να τονιστεί περισσότερο ότι οι phishing τεχνικές εξελίσσονται συνεχώς, με τους εγκληματίες να χρησιμοποιούν Generative AI για δημιουργία πιο πειστικών μηνυμάτων και ψεύτικων ιστοσελίδων. Επιπλέον, ο Karthik (2025) και ο Naseer (2024) συμφωνούν ότι η αυτοματοποιημένη ανταπόκριση σε περιστατικά (όπως η αυτόματη απομόνωση ύποπτων email) γίνεται όλο και πιο σημαντική, κάτι που το κείμενο θα μπορούσε να αναφέρει. Τέλος, θα μπορούσε να προστεθεί ότι η διεθνής συνεργασία μεταξύ εταιρειών και κυβερνήσεων είναι κρίσιμη για την καταπολέμηση διεθνών οργανωμένων ομάδων phishing.

Απάντηση ανταγωνιστικού μοντέλου Generative AI (Deepseek)

Εισαγωγή

Το phishing αποτελεί μια από τις πιο διαδεδομένες και επικίνδυνες μορφές κυβερνοεπιθέσεων, βασισμένη στην κοινωνική μηχανική για την εξαπάτηση των θυμάτων. Οι επιτιθέμενοι εκμεταλλεύονται την ανθρώπινη ψυχολογία και την εμπιστοσύνη, προκαλώντας σημαντικές οικονομικές και πληροφοριακές ζημιές. Στην παρούσα εργασία, γίνεται ανάλυση του φαινομένου του phishing από θεωρητική σκοπιά, παρουσιάζονται σύγχρονες τεχνικές και παραδείγματα επιθέσεων, καθώς και οι μελλοντικές τάσεις στην αντιμετώπισή τους με τη βοήθεια της Τεχνητής Νοημοσύνης.

Θεωρητική Τεκμηρίωση του Phishing

Το phishing ανήκει στην ευρύτερη κατηγορία των επιθέσεων κοινωνικής μηχανικής, όπου οι εγκληματίες χειρίζονται ψυχολογικά τα θύματα για να αποκτήσουν ευαίσθητα δεδομένα, όπως κωδικούς πρόσβασης ή πληροφορίες πιστωτικών καρτών. Βασικές αρχές που διέπουν αυτές τις επιθέσεις περιλαμβάνουν τη δημιουργία αίσθησης έκτακτης ανάγκης, την προσποίηση αξιοπιστίας και την ψυχολογική πίεση. Από την άποψη της κυβερνοασφάλειας, το phishing αποτελεί απειλή για τη διαθεσιμότητα, την ακεραιότητα και την εμπιστευτικότητα των δεδομένων. Οι επιθέσεις μπορούν να πραγματοποιηθούν μέσω ηλεκτρονικών μηνυμάτων (email phishing), SMS (smishing) ή ακόμη και τηλεφωνικών κλήσεων (vishing), με στόχο την εξαπάτηση των χρηστών.

Παραδείγματα Σύγχρονων Phishing Επιθέσεων

Πολλές εταιρείες και χρήστες έχουν πέσει θύματα phishing, με σημαντικές οικονομικές και λειτουργικές επιπτώσεις. Για παράδειγμα, μεταξύ 2013 και 2015, ένας εγκληματίας εξαπάτησε τη Google και το Facebook με ψεύτικα email που έμοιαζαν με νόμιμους λογαριασμούς προμηθευτών, κλέβοντας πάνω από 100 εκατομμύρια δολάρια. Κατά την περίοδο της πανδημίας COVID-19, οι απατεώνες εκμεταλλεύτηκαν τον πανικό, στέλνοντας ψεύτικους συνδέσμους για επιδόματα ή εμβόλια, με αποτέλεσμα τη μόλυνση υπολογιστών με κακόβουλο λογισμικό. Επίσης, το 2020, παραβιάστηκαν λογαριασμοί γνωστών προσωπικοτήτων στο Twitter, όπως του Elon Musk, για την προώθηση απάτης με Bitcoin.

Σύγχρονες Τεχνικές Phishing και Λειτουργία τους

Οι πιο συνηθισμένες τεχνικές phishing σήμερα περιλαμβάνουν το email phishing, όπου οι εγκληματίες στέλνουν τυποποιημένα μηνύματα που μιμούνται νόμιμες εταιρείες, όπως η Amazon ή η PayPal, και οδηγούν τους χρήστες σε πλαστές ιστοσελίδες για να καταγράψουν τα στοιχεία τους. Επιπλέον, το spear phishing αποτελεί εξατομικευμένη μορφή επίθεσης, όπου οι εγκληματίες χρησιμοποιούν προσωπικές πληροφορίες για να αυξήσουν την πιθανότητα επιτυχίας. Αυτή η τεχνική χρησιμοποιείται συχνά για επιχειρησιακή κατασκοπεία. Άλλες τεχνικές περιλαμβάνουν το smishing, όπου οι χρήστες λαμβάνουν SMS με ψεύτικους συνδέσμους, και το vishing, όπου οι εγκληματίες επικοινωνούν τηλεφωνικά, υποδυόμενοι τράπεζες ή υπηρεσίες υποστήριξης.

Μελλοντικές Τάσεις και Ρόλος της Τεχνητής Νοημοσύνης

Η αντιμετώπιση του phishing εξελίσσεται με την εφαρμογή νέων τεχνολογιών, όπως η Τεχνητή Νοημοσύνη. Η μηχανική μάθηση χρησιμοποιείται για την ανάλυση μοτίβων σε ύποπτα email ή URLs, ενώ η φυσική επεξεργασία γλώσσας (NLP) βοηθά στον εντοπισμό ψεύτικων μηνυμάτων. Προληπτικά συστήματα, όπως τα AI-based sandboxing, επιτρέπουν τη δοκιμή συνδέσμων σε εικονικά περιβάλλοντα πριν την πραγματική χρήση τους. Επιπλέον, πλατφόρμες όπως το Google Safe Browsing εμποδίζουν την πρόσβαση σε κακόβουλες ιστοσελίδες. Μια σημαντική τάση είναι η εκπαίδευση των χρηστών μέσω προσομοιώσεων phishing, όπου οι εταιρείες διεξάγουν δοκιμές για την αύξηση της ευαισθητοποίησης. Η εφαρμογή πολυπαραγοντικής ταυτοποίησης αποτελεί επίσης κρίσιμο βήμα για την ενίσχυση της ασφάλειας.

Συμπέρασμα

Το phishing παραμένει μια συνεχώς εξελισσόμενη απειλή, η οποία απαιτεί συνδυασμό τεχνολογικών λύσεων και ευαισθητοποίησης των χρηστών. Η μελλοντική πρόληψη θα στηριχτεί σε έξυπνα συστήματα ανίχνευσης και στην ικανότητα των ατόμων να αναγνωρίζουν ύποπτες ενέργειες. Η συλλογική ευθύνη των εταιρειών, των κυβερνήσεων και των πολιτών είναι καθοριστική για την αντιμετώπιση αυτού του κινδύνου.

Σύγκριση των δύο μοντέλων Generative AI

Στην παρούσα μελέτη, διεξάγεται μια λεπτομερής σύγκριση δύο κειμένων που πραγματεύονται το φαινόμενο των phishing επιθέσεων. Το κείμενο Α, που προέρχεται από το αρχικό μοντέλο Generative AI (ChatGPT), και το κείμενο Β, που προέρχεται από ανταγωνιστικό μοντέλο Generative AI (DeepSeek), παρουσιάζουν τόσο ομοιότητες όσο και σημαντικές διαφορές σε πολλαπλά επίπεδα ανάλυσης.

Από πλευράς δομικής οργάνωσης, το κείμενο Β ακολουθεί μια αυστηρά τυποποιημένη ακαδημαϊκή μορφή. Χωρίζεται σε ευκρινείς ενότητες που περιλαμβάνουν εισαγωγή, θεωρητικό πλαίσιο, παρουσίαση παραδειγμάτων, ανάλυση τεχνικών, μελλοντικές τάσεις και συμπεράσματα. Αυτή η δομική αυστηρότητα το καθιστά ιδανικό για τυπικές ακαδημαϊκές εργασίες. Αντίθετα, το κείμενο Α, παρότι καλύπτει ουσιαστικά το ίδιο περιεχόμενο, υιοθετεί μια πιο ευέλικτη δομή που δεν τηρεί πάντα τους αυστηρούς κανόνες ακαδημαϊκής μορφοποίησης. Η προσέγγιση αυτή προσδίδει στο κείμενο Α μεγαλύτερη ρευστότητα και ευκολία ανάγνωσης, καθιστώντας το πιο προσβάσιμο σε ευρύτερα κοινά.

Στο πεδίο του βάθους ανάλυσης, τα δύο κείμενα παρουσιάζουν διαφορετικές προσεγγίσεις. Το κείμενο Β προσφέρει μια συνοπτική αλλά ολοκληρωμένη επισκόπηση των βασικών εννοιών, με έμφαση στη θεωρητική κατανόηση του φαινομένου. Εστιάζει ιδιαίτερα στις αρχές της κοινωνικής μηχανικής και στον θεμελιώδη ρόλο της ανθρώπινης ψυχολογίας στις phishing επιθέσεις. Αντίθετα, το κείμενο Α εμβαθύνει σε συγκεκριμένες τεχνικές και μεθόδους, παρουσιάζοντας λεπτομερείς αναλύσεις για εξειδικευμένες μορφές phishing όπως το whaling και το vishing. Επιπλέον, το κείμενο Α περιλαμβάνει πλουσιότερη τεκμηρίωση με συγκεκριμένα παραδείγματα (π.χ. η επίθεση στο Twitter το 2020 και η υπόθεση της απάτης κατά των Facebook και Google), προσφέροντας μια πιο εμπρακτική κατανόηση του θέματος.

Στην πληρότητα περιεχομένου, το κείμενο Α διακρίνεται για την εκτενέστερη κάλυψη του θέματος. Περιλαμβάνει μια ξεχωριστή και λεπτομερή ενότητα για μεθόδους αντιμετώπισης και πρόληψης, όπου αναλύονται συγκεκριμένες τεχνικές όπως η πολιτική Zero Trust, η πολυπαραγοντική ταυτοποίηση (MFA) και τα συστήματα email filtering. Επιπλέον, παρουσιάζει μια πιο ολοκληρωμένη συζήτηση των μελλοντικών τάσεων, εξετάζοντας καινοτόμες τεχνολογίες όπως η ανάλυση

συμπεριφοράς χρηστών (UBA), τα συστήματα blockchain και οι τεχνολογίες deepfake. Το κείμενο Β, από την άλλη πλευρά, διατηρεί μια πιο συγκεντρωμένη προσέγγιση, εστιάζοντας κυρίως στις θεωρητικές βάσεις του φαινομένου και στον ρόλο της τεχνητής νοημοσύνης στην αντιμετώπιση των επιθέσεων.

Στο επίπεδο του γλωσσικού ύφους, τα δύο κείμενα παρουσιάζουν αξιοσημείωτες διαφορές. Το κείμενο Β χρησιμοποιεί ένα πιο τυπικό ακαδημαϊκό λεξιλόγιο και μια πιο επίσημη δομή, που το καθιστά κατάλληλο για αυστηρές ακαδημαϊκές απαιτήσεις. Ωστόσο, αυτή η τυπικότητα μπορεί μερικές φορές να περιορίζει την ευκολία ανάγνωσης και την προσβασιμότητα για μη ειδικούς αναγνώστες. Το κείμενο Α, από την άλλη πλευρά, διατηρεί ένα πιο ευέλικτο και προσβάσιμο ύφος, χωρίς να θυσιάζει την τεχνική ακρίβεια και το βάθος της ανάλυσης. Αυτή η ισορροπία μεταξύ ακαδημαϊκής σοβαρότητας και ευκολίας κατανόησης καθιστά το κείμενο Α ιδιαίτερα αποτελεσματικό για ένα ευρύτερο φάσμα αναγνωστών.

Στην ανάλυση των μελλοντικών τάσεων, και τα δύο κείμενα τονίζουν τον κεντρικό ρόλο της τεχνητής νοημοσύνης και των συστημάτων μηχανικής μάθησης. Ωστόσο, το κείμενο Α επεκτείνει σημαντικά αυτή τη συζήτηση, εξετάζοντας πρόσθετους τομείς όπως, τις εφαρμογές της τεχνολογίας blockchain για την επαλήθευση της ταυτότητας των αποστολέων, τις δυνατότητες της ανάλυσης συμπεριφοράς χρηστών (UBA) για την ενίσχυση της ασφάλειας και τις προκλήσεις που δημιουργούν οι τεχνολογίες deepfake στον τομέα των ψηφιακών απατών. Αντίθετα, το κείμενο Β περιορίζεται σε μια πιο συνοπτική και θεωρητική συζήτηση των τεχνολογικών εξελίξεων, χωρίς την ίδια εμβάθυνση σε συγκεκριμένες τεχνολογίες και εφαρμογές.

Ως τελικό συμπέρασμα, η συγκριτική αυτή ανάλυση αποκαλύπτει ότι κάθε κείμενο προσφέρει μοναδικά πλεονεκτήματα ανάλογα με τον σκοπό χρήσης. Το κείμενο Β αποτελεί μια ιδανική επιλογή για τυπικές ακαδημαϊκές εργασίες, όπου η τυπική δομή και η θεωρητική αυστηρότητα είναι πρωταρχικής σημασίας. Από την άλλη πλευρά, το κείμενο Α ξεχωρίζει για την εμβάθυνση σε τεχνικές λεπτομέρειες, την πλουσιότερη τεκμηρίωση με πραγματικά παραδείγματα και την ολοκληρωμένη κάλυψη τόσο των πρακτικών μεθόδων αντιμετώπισης όσο και των μελλοντικών τάσεων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- ENISA (2023) *ENISA Threat Landscape 2023 – Cybersecurity Threats and Trends*. European Union Agency for Cybersecurity. Available at: [ENISA Threat Landscape 2023 | ENISA](#) (Accessed: 9 April 2025).
- Butavicius, M., Parsons, K., Pattinson, M., and McCormac, A. (2016) 'Breaching the Human Firewall: Social engineering in Phishing and Spear-Phishing Emails'. Available at: [\[1606.00887\] Breaching the Human Firewall: Social engineering in Phishing and Spear-Phishing Emails](#) (Accessed: 9 April 2025).
- Chin, K. (2024) *19 most common types of phishing attacks in 2024*. Available at: [19 Most Common Types of Phishing Attacks in 2025 | UpGuard](#) (Accessed: 20 April 2025).
- Fortinet (2024) Types of phishing attacks. Available at: [19 Types of Phishing Attacks with Examples | Fortinet](#) (Accessed: 20 April 2025).
- BBC News (2020) *Twitter hack: Bitcoin scam hits Bill Gates, Elon Musk and Obama accounts*. Available at: [Major US Twitter accounts hacked in Bitcoin scam](#) (Accessed: 27 April 2025).
- Boden, A. et al. (2018) *"Business Email Compromise (BEC) attacks: A case study on the \$100M Facebook-Google phishing scam"*, Journal of Cybersecurity, 4(1), pp. 1–12. DOI: [10.1093/cybsec/tyy004](#)
- Interpol (2020) *Cybercriminals targeting critical healthcare institutions with ransomware*. Available at: [Cybercriminals targeting critical healthcare institutions with ransomware](#) (Accessed: 27 April 2025).
- Naseer, I. (2024) 'The role of artificial intelligence in detecting and preventing cyber and phishing attacks', European Journal of Advances in Engineering and Technology, 11(9), pp. 82–86. Available at: [\(PDF\) The role of artificial intelligence in detecting and preventing cyber and phishing attacks](#) (Accessed: 27 April 2025).
- Karthik, V. (2025) Preventing phishing with AI: How technology is fighting back. Insights2Techinfo. Available at: [Preventing Phishing with AI : How Technology is Fighting Back](#) (Accessed: 27 April 2025)