Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακή Διατριβή
στα Πληροφοριακά και Επικοινωνιακά Συστήματα

Συστηματοποίηση της εξέλιξης των τεχνικών CAPTCHA από την άποψη της ασφάλειας.

Μαρία Σερεσιώτη

Επιβλέπων Καθηγητής
Ηλίας Αθανασόπουλος

Ιούνιος 2017

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Συστηματοποίηση της εξέλιξης των τεχνικών CAPTCHA από την άποψη της ασφάλειας.

Μαρία Σερεσιώτη

Επιβλέπων Καθηγητής
Ηλίας Αθανασόπουλος

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε
προς μερική εκπλήρωση των απαιτήσεων για απόκτηση

μεταπτυχιακού τίτλου σπουδών
στα Πληροφοριακά Συστήματα

από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών
του Ανοικτού Πανεπιστημίου Κύπρου

Ιούνιος 2017

Περίληψη

Τα τελευταία χρόνια πλήθος δεδομένων προέρχεται από το διαδίκτυο το οποίο συνέχεια προσελκύει όλο και περισσότερους χρήστες. Καθημερινές είναι οι κακόβουλες επιθέσεις που δέχονται οι χρήστες του διαδικτύου από επιτήδειους οι οποίοι έχουν ως στόχο να κερδίσουν χρήματα μέσω της συστηματικής κατάχρησης των διαφόρων δωρεάν υπηρεσιών που υπάρχουν στις ιστοσελίδες. Η ανάγκη για προστασία από τους κίνδυνους του διαδικτύου έχει αναγκάσει τους χρήστες να πρέπει να αποδεικνύουν την ανθρώπινη υπόστασή τους όταν τις χρησιμοποιούν. Πολλά κακόβουλα προγράμματα προσπαθούν να αποκτήσουν πρόσβαση στα mail ή σε άλλες υπηρεσίες που χρησιμοποιούν οι χρήστες με επανειλημμένες προσπάθειες, συνδυάζοντας γράμματα ή αριθμούς και σχηματίζοντας κωδικούς πρόσβασης, μέχρι να βρουν τον σωστό ή χρησιμοποιούν τα έγκυρα μέιλ προκειμένου να στέλνουν μαζικά διαφημιστικά μηνύματα (spam). Για την αντιμετώπιση του προβλήματος χρησιμοποιούνται ευρέως τα CAPTCHA, με σκοπό να εξακριβώσουν αν μία αίτηση σε μία υπηρεσία γίνεται από έναν χρήστη ή από ένα αυτοματοποιημένο πρόγραμμα. Τα CAPTCHA εμφανίζονται στο διαδίκτυο ως παραμορφωμένες εικόνες που παρουσιάζουν συνδυασμούς γραμμάτων και αριθμών σε διάφορα μεγέθη, χρώματα κλπ. Όταν ο χρήστης θέλει να πραγματοποιήσει κάποια εγγραφή, καλείται να αναγνωρίσει την εικόνα και το περιεχόμενο της και να συμπληρώσει σωστά τη φόρμα. Αν και το captcha θεωρείται μια καλή τεχνική δεν είναι απολυτά ασφαλές διότι κάποιοί κακόβουλοι χρήστες έχουν βρει τρόπο να το ξεπερνούν. Αντικείμενο της διατριβής αυτής θα είναι η συστηματοποίηση (systemization) των διαφόρων τεχνολογιών CAPTCHA, με έμφαση στην ασφάλεια. Πιο συγκεκριμένα, θα γίνει μελέτη και καταγραφή των γνωστών μέχρι σήμερα τεχνολογιών με συγκριτική αποτίμηση των πλεονεκτημάτων και των μειονεκτημάτων τους και, κατά πόσο η χρήση τους καθιστά ασφαλείς τις διαδικτυακές εφαρμογές. Ειδικότερα θα διερευνηθούν οι τεχνολογίες που χρησιμοποιούνται και ποιες είναι οι επιθέσεις που δέχονται ώστε, κάποιος να μπορεί να αποφανθεί αν είναι ασφαλείς ή όχι.

Λεξεις Κλειδια:

CAPTCHA, ασφάλεια, επιθέσεις, Spamming.

Summary

In recent years, a lot of data come from the internet, which is constantly attracting an increasing number of users. The malicious attacks that the users experience from other deft users are daily and are aimed at earning money through the systematic abuse of various free services on websites. The need to protect against the dangers of the internet has forced users to demonstrate their human nature when using them. Many malicious programs try to repeatedly access emails or other services that users use by combining letters or numbers and creating passwords until they find the correct one or use the valid emails to send unsolicited advertising electronic messages know as spam. To deal with the problem, CAPTCHA is widely used to determine if a request to a service is made by a user or an automated program. CAPTCHA are displayed on the web as distorted images that present combinations of letters and numbers in different sizes, colors, etc. When users want to make a registration, they are asked to recognize the image and its contents and fill in the form correctly. Although captcha is considered a good technique, it is not completely safe because some malicious users have found a way to overcome it. The subject of this dissertation will be the systemization of the various CAPTCHA technologies, with emphasis on safety. More specifically, there will be a study and recording of the technologies known today by comparing their advantages and disadvantages and whether their use makes web applications safe. In particular, this paper will explore the technologies used and the attacks they face so that one can decide whether they are safe or not.

Keywords:

CAPTCHA, Security, Attacks, Spamming.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Η. Αθανασόπουλο για την καθοδήγηση του καθ' όλη τη διάρκεια της εκπόνησης της εργασίας αυτής.

Τέλος θα ήθελα να εκφράσω την ευγνωμοσύνη μου στον σύζυγο μου και στους φίλους μου για την υπομονή και συμπαράσταση που έχουν δείξει σε όλη την διάρκεια των σπουδών μου.

Περιεχόμενα

Κεφάλαιο 1
Εισαγωγή

Τα CAPTCHA είναι ένα είδος αντίστροφου τεστ Turing με στόχο να επιβεβαιώσει ότι η απάντηση σε αυτό προέρχεται από έναν άνθρωπο και όχι από έναν υπολογιστή. Στην πιο διαδεδομένη τους μορφή είναι παραμορφωμένες εικόνες κειμένου και η συμπλήρωσή τους είναι απαραίτητη σε πάρα πολλές υπηρεσίες, από την υποβολή ενός σχολίου για κάποιο κείμενο και την δημιουργία ενός λογαριασμού ηλεκτρονικού ταχυδρομείου μέχρι την προσπάθεια ανάκτησης κωδικού σε υπηρεσίες ηλεκτρονικού εμπορίου, με σκοπό να σταματήσουν τις αυτοματοποιημένες δέσμες ενεργειών που χρησιμοποιούνται για την κακόβουλη χρήση αυτών των υπηρεσιών [38].

Η ραγδαία ανάπτυξη του διαδικτύου και οι πολλές υπηρεσίες που προσφέρουν πολλές ιστοσελίδες οδηγούν στην καθημερινή κατάχρησή τους. Για να αντιμετωπιστούν διάφορα προβλήματα δημιουργήθηκαν τα CAPTCHA με σκοπό να ξεχωρίσουν αν οι ενέργειες γίνονται από τον άνθρωπο ή από κάποια μηχανή. Η χρήση των CAPTCHA έχει γίνει πολύ βασική για κάθε ιστοσελίδα προκειμένου να θωρακιστεί απέναντι σε οποιαδήποτε επιχείρηση εισβολής

Η ενσωμάτωση των CAPTCHA στις υπηρεσίες των ιστοσελίδων αυτόματα σήμανε και την έναρξη ενός αγώνα δρόμου με σκοπό την ανάπτυξη μιας τεχνολογίας που θα βοηθά στην λύση τους. Αυτό είχε σαν αποτέλεσμα, πέρα από την εξέλιξη των βασικών οπτικών δοκιμασιών, την

ανάπτυξη νέων ειδών CAPTCHA που βασίζονται σε διαφορετικές τεχνικές. Όσο εξελίσσονται οι τεχνικές για τη δημιουργία ενός συστήματος CAPTCHA αυτόματα αυξάνεται και ο βαθμός δυσκολίας όσον αφορά τις γνώσεις και τον χρόνο που απαιτείται. Ο συνηθέστερος δρόμος είναι η χρήση μίας έτοιμης λύσης, αλλά πλέον ακόμη και αυτό χρειάζεται ιδιαίτερη προσπάθεια [42].

Η μεγάλη ποικιλία λύσεων σε συνδυασμό με την ανάγκη της χρήσης CAPTCHA στις περισσότερες ιστοσελίδες εμφανίζει μια καινούργια δυσκολία για τους δημιουργούς των ηλεκτρονικών υπηρεσιών. Θα πρέπει να επιλέξουν να ενσωματώσουν ένα CAPTCHA το οποίο θα είναι πολύ εύκολο για τον άνθρωπο και πολύ σκληρό για τους υπολογιστές. Στόχος τους είναι η μεγιστοποίηση τη δυσκολία των αυτοματοποιημένων προγραμμάτων να περάσουν τις δοκιμές χωρίς όμως να δυσκολεύουν τους επισκέπτες τους.

1.1 Σκοπός και Δομή Διατριβής

Στο πλαίσιο της διατριβής θα γίνει μελέτη και καταγραφή των γνωστών μέχρι σήμερα τεχνολογιών με συγκριτική αποτίμηση των πλεονεκτημάτων και των μειονεκτημάτων τους και, κατά πόσο η χρήση τους καθιστά ασφαλής τις web εφαρμογές. Ειδικότερά θα διερευνηθούν οι τεχνολογίες που χρησιμοποιούνται και ποιες είναι οι επιθέσεις που δέχονται ώστε, κάποιος να μπορεί να αποφανθεί αν είναι ασφαλής ή όχι.

Τέλος θα ολοκληρώσουμε με την ανάλυση σχετικά με τις τυποποιημένες λύσεις που πιθανόν να ενσωματωθούν στην εκάστοτε ιστοσελίδα, επισημαίνοντας τις λεπτομέρειες αλλά και τα προβλήματα υλοποίησης.

Η Δομή της Διατριβής όπως διαρθρώνεται ανά κεφάλαιο είναι η ακόλουθη:

Στο **κεφάλαιο 2** παρουσιάζονται οι βασικές επιθέσεις που πραγματοποιούνται σε φόρμες ιστοσελίδων γεγονός που καθιστά αναγκαία την χρήση των CAPTCHA

Στο **κεφάλαιο 3** αρχικά δίνεται ο ορισμός των CAPTCHA και παρουσιάζονται οι λόγοι χρήσης τους. Στην συνέχεια γίνεται αναφορά στις μεθόδους παραγωγής και στις δυσκολίες της χρήσης των CAPTCHA.

Στο **κεφάλαιο 4** παρουσιάζονται οι τεχνικές που υπάρχουν μέχρι σήμερα για τη δημιουργία των CAPTCHA, τα πλεονεκτήματα και τα μειονεκτήματα κάθε τεχνικής, καθώς και τυπικές επιθέσεις που μπορούν να δεχτούν.

Στο **κεφάλαιο 5** αναφέρονται διάφορα εργαλεία επιθέσεων σε CAPTCHA (λογισμικά, browser extensions). Έπειτα πραγματοποιούνται επιθέσεις σε CAPTCA ιστοσελίδων με την χρήση μερικών από τα εργαλεία που παρουσιάζονται.

Στο **κεφάλαιο 6** γίνεται σύνοψη των αποτελεσμάτων της ερευνάς.

Κεφάλαιο 2
Κακόβουλη Χρήση των Υπηρεσιών

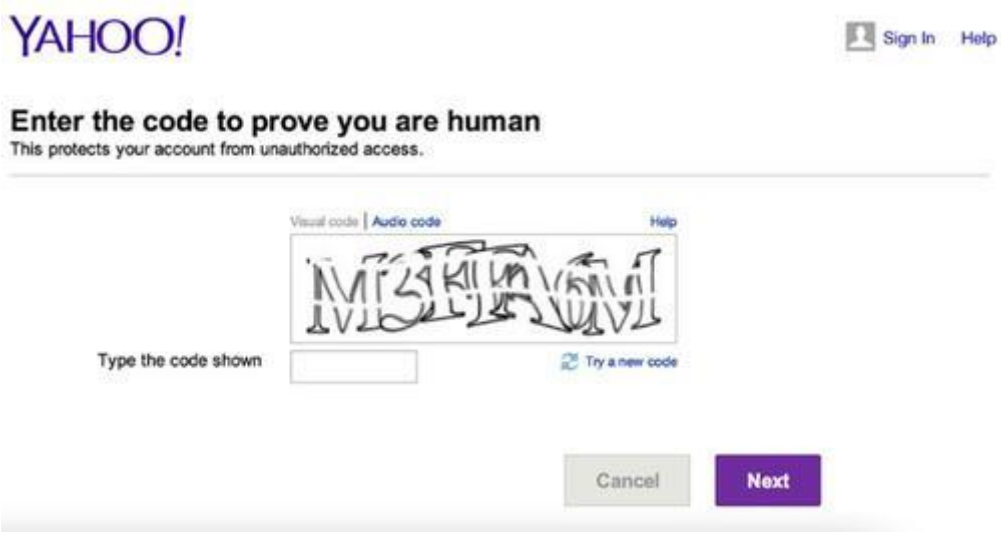
Η πρόσβαση στο Διαδίκτυο σήμερα δεν είναι ακίνδυνη, καθημερινά οι χρήστες του χρησιμοποιούν τις υπηρεσίες του για να βρουν πληροφορίες που χρειάζονται. Μερικοί ιστότοποι εμφανίζουν πληροφορίες οι οποίες φαινομενικά είναι ακριβείς ή αναφέρουν απόλυτα αξιόπιστους δημιουργούς ή πηγές. Κακόβουλοι χρήστες προσπαθούν να καταχραστούν τις υπηρεσίες που παρέχουν με στόχο είτε την αποκομιδή ιδίου οφέλους είτε της παραπλάνησης των αγνώστων χρηστών. Για την επίτευξη των σκοπών τους οι επιτιθέμενοι πολλές φορές κάνουν χρήση αυτοματοποιημένων εργαλείων για τη μαζική επίτευξη των στόχων τους

2.1 Κύριες Τοποθεσίες Επιθέσεων

Παρακάτω παρατίθενται οι κυρίες τοποθεσίες επιθέσεων σε ιστοσελίδες και όπως αυτές αναλύονται δείχνουν άμεσα την ανάγκη προστασίας των υπηρεσιών τους και την ανάπτυξη της τεχνολογίας CAPTCHA.

2.1.1 Δωρεάν Υπηρεσίες Ηλεκτρονικού Ταχυδρομείου

Οι εταιρείες όπως Yahoo, Microsoft που προσφέρουν υπηρεσίες δωρεάν ηλεκτρονικού ταχυδρομείου, υποφέρουν οι περισσότερες από αυτές από ένα συγκεκριμένο είδος επίθεσης: "bots" κατά την οποία κάθε λεπτό γίνεται εγγραφή χιλιάδων λογαριασμών. Αυτού του είδους η επίθεση μπορεί να περιοριστεί απαιτώντας από τους χρήστες να αποδείξουν ότι είναι άνθρωποι πριν να αποκτήσουν δωρεάν λογαριασμό ηλεκτρονικού ταχυδρομείου ή κάνουν χρήση κάποιας άλλης υπηρεσίας που προσφέρει η ιστοσελίδα τους. Οι εταιρίες για την αποτροπή των bots χρησιμοποιούν τις περισσότερες φορές ένα CAPTCHA που ζητά από το χρήστη να διαβάσει μια παραμορφωμένη λέξη όπως αυτή που φαίνεται παρακάτω στην παρακάτω εικόνα. Τα τρέχοντα προγράμματα υπολογιστών συνήθως δεν είναι τόσο καλά όσο οι άνθρωποι στην ανάγνωση ενός διαστρεβλωμένου κειμένου [17].



Εικόνα 2.1: Παράδειγμα Yahoo CAPTCHA

2.1.2 Διαδικτυακό Marketing

Οι spammers έχουν ως στόχο να διαφημίσουν τα προϊόντα που αντιπροσωπεύουν μαζικά σε emails αλλά και σε forums, socialnetworks καθώς και σε διάφορες άλλες online υπηρεσίες. Για αυτό το λόγο δημιουργούν διάφορα bots, τα οποία αναλαμβάνουν είτε να σαρώνουν το Διαδίκτυο για την συλλογή διευθύνσεων email, είτε να εκτελούν μαζική δημοσίευση διαφημιστικών σχολίων στις διάφορες υπηρεσίες, αφού πρώτα δημιουργήσουν αυτόματα τους ανάλογους λογαριασμούς που θα επιτρέψουν τις δημοσιεύσεις αυτές [17].

2.1.3 Μηχανές Αναζήτησης

Το Internet σήμερα αποτελεί έναν χώρο που αναπτύσσετε δυναμικά με αποτέλεσμα η πλειονότητα των χρηστών γνωρίζει ένα μικρό μέρος από διευθύνσεις ιστοσελίδων. Πιο ειδικά γνωρίζουν ιστοσελίδες που το περιεχόμενο τους έχει άμεση σχέση με τα ενδιαφέροντα τους. Σε κάθε άλλη περίπτωση είναι υποχρεωμένοι να στραφούν στις μηχανές αναζήτησης μέσω των διαφόρων φυλομετρητών(Browsers). Μέχρι το πρόσφατο παρελθόν όσα περισσότερα backlinks είχε μια ιστοσελίδα τόσο καλύτερη κατάταξη επιτύγχανε στα αποτελέσματα αναζήτησης. Το γεγονός αυτό οδήγησε αρκετούς κακόβουλους (Webmasters) να προσπαθήσουν να ξεγελάσουν τις μηχανές αναζήτησης χτίζοντας links με αθέμιτους τρόπους για να δημιουργήσουν μια τεχνητή δημοτικότητα της σελίδας τους [17].

2.1.4 Ηλεκτρονικές Δημοσκοπήσεις

Διάφορες ιστοσελίδες στο διαδίκτυο ζητούν από τους επισκέπτες τους να λάβουν μέρος σε διάφορες ηλεκτρονικές δημοσκοπήσεις ή ψηφοφορίες οι οποίες ανάλογα με το θέμα το οποίο θέτουν πολλές φορές μπορεί να διχάσουν το κοινό. Αυτό έχει ως αποτέλεσμα κακόβουλοι χρήστες οι οποίοι έχουν διαφορετική γνώμη ή έχουν οικονομικά οφέλη από την έκβαση του αποτελέσματος μια δημοσκόπησης ή ψηφοφορίας να προσπαθήσουν να αλλοιώσουν τα αποτελέσματα. Συνήθως αυτό συμβαίνει σε σελίδες που έχουν μεγάλη επισκεψιμότητα γεγονός που τις καθιστά ελκυστικές για τον επιτιθέμενο [17].

2.1.5 Επιθέσεις Λεξικού

Στις επιθέσεις λεξικού (Dictionary Attacks), στόχος είναι να αποκτηθεί η πρόσβαση σε μια ιστοσελίδα. Οι επιτιθέμενοι χρησιμοποιούν διάφορους τρόπους για να αυτοματοποιήσουν την υποβολή των αιτήσεων για login. Συνήθως όμως χρησιμοποιούν προκαθορισμένες τιμές τις οποίες στέλνουν στο Server και στη συνέχεια εξετάζουν την απόκριση (response) του Server.

Ένα απλό Brute Force Attack θα προσπαθήσει να δοκιμάσει κάθε πιθανό συνδυασμό των διαθέσιμων χαρακτήρων μέχρι να βρει το σωστό και να αποκτηθεί η ζητούμενη πρόσβαση. Οι επιτιθέμενοι έχουν πολλούς τρόπους για να μαντέψουν το Username και το Password ενός χρήστη κάνοντας συνδυασμούς και σε άλλους δικτυακούς τόπους, όπου ο χρήστης μπορεί να έχει το ίδιο αναγνωριστικό και τον ίδιο κωδικό πρόσβασης. Εάν η ιστοσελίδα WordPress περιέχει τα

προσωπικά στοιχεία του χρήστη, τα στοιχεία πληρωμής για το ηλεκτρονικό εμπόριο ή άλλα ευαίσθητα δεδομένα που συνδέονται με το λογαριασμό του τότε οι εισβολείς μπορούν να συνδεθούν με το λογαριασμό του εύκολα να τον κλέψουν. Το yahoo, AltaVista και το PayPal για ορισμένες από τις υπηρεσίες τους τα χρησιμοποιούν τα CAPTCHA για να εξασφαλίσουν ότι η χρήση τους γίνεται μόνο από ανθρώπινο δυναμικό [03, 17].

2.1.6 Worms και Spam

Τα CAPTCHA προσφέρουν μια ικανοποιητική λύση εναντίον emails Worms, κακόβουλων αυτοαναπαραγόμενων προγραμμάτων και, της ανεπιθύμητης ηλεκτρονικής αλληλογραφίας Spam [17].

Κεφάλαιο 3
CAPTCHA

Ο όρος CAPTCHA είναι ακρωνύμιο που προέρχεται από την φράση «Completely Automated Public Turing test to tell Computers and Human Apart» σε ελληνική μετάφραση *Πλήρως Αυτοματοποιημένο Δημόσιο τεστ Turing για το Διαχωρισμό Ανθρώπων και Υπολογιστών*. Το χρησιμοποιούν οι ιστοσελίδες για να προσδιορίσουν αν ένας χρήστης είναι άνθρωπος ή όχι [41].

3.1 Λόγοι Χρήσης

Το Captcha είναι ο βασικός τρόπος με τον οποίο ξεχωρίζουν οι άνθρωποι και τα ρομπότ στο διαδίκτυο. Ο μηχανισμός αυτός επιβεβαιώνει ότι αυτός που θέλει να καταχωρήσει κάποιο στοιχείο είναι άνθρωπος και αποτρέπει τα ρομπότ από το να γεμίζουν αυτοματοποιημένα μηνύματα την κάθε ιστοσελίδα (Spamming). Υπολογίζεται ότι καθημερινά, οι άνθρωποι ξοδεύουν χιλιάδες εργατοώρες για να αποδείξουν ότι δεν είναι μηχανές, λύνοντας περίπου εκατομμύρια δοκιμασίες CAPTCHA. Οι λόγοι που οι ιδιοκτήτες ιστοτόπων αναγκάζονται να υποβάλλουν τους επισκέπτες τους σε αυτή τη διαδικασία οφείλεται εκτός από το Spam και στο Black Hat SEO [05, 37, 42].

Ένα σύστημα CAPTCHA θα πρέπει να υπακούει σε τρείς βασικές αρχές [03, 37]:

- Να μπορούν να ξεπεραστούν σχετικά εύκολα από ανθρώπους.
- Η κατασκευή κάποιου αλγορίθμου που έχει την δυνατότητα να ξεπεράσει τις δοκιμές να είναι από πολύ δύσκολη έως αδύνατη.
- Να είναι εύκολο για τη μηχανή να το τεστάρει και να το βαθμολογήσει.

Υπάρχουν τρεις βασικοί τύποι CAPTCHA [42]:

- Βασισμένα στο κείμενο (Text-Based Schemes)
- Βασισμένα στον ήχο (Sound-Based Schemes)
- Βασισμένα στην εικόνα(Image-Based Schemes)

3.2 Μέθοδοι παραγωγής CAPTCHA

Η παραγωγή CAPTCHA διαδόθηκε γρήγορα στα πλαίσια μιας προσπάθειας προστασίας των online υπηρεσιών του Διαδικτύου από την κακόβουλη χρήση. Χρησιμοποιώντας ένα πρόγραμμα υπολογιστή παράγονται εικόνες που περιέχουν δοσμένο κείμενο όπως στην Εικόνα3.1, η αντίστροφη διαδικασία, δηλαδή η εξαγωγή σε μορφή κειμένου ενός αλφαριθμητικού που περιέχεται σε εικόνα (OCR – Optical Character Recognition) είναι αρκετά πιο δύσκολη διαδικασία. Μάλιστα, με σκοπό αυτή η διαδικασία να γίνει ακόμα δυσκολότερη, ένα σύστημα CAPTCHA μπορεί, ανάλογα με την υλοποίηση, να χρησιμοποιεί μια ή περισσότερες από τις παρακάτω τεχνικές [37]:

1. Χρήση τυχαίων αλφαριθμητικών αντί επιλογής από πεπερασμένο σύνολο αναγνωρίσιμων λέξεων ώστε να μην μπορούν να χρησιμοποιηθούν λεξικά κατά την διαδικασία OCR.
2. Χρήση διαφορετικής γραμματοσειράς για κάθε εικόνα που δημιουργείται ή, για ακόμα μεγαλύτερη ασφάλεια, για κάθε χαρακτήρα στην εικόνα.

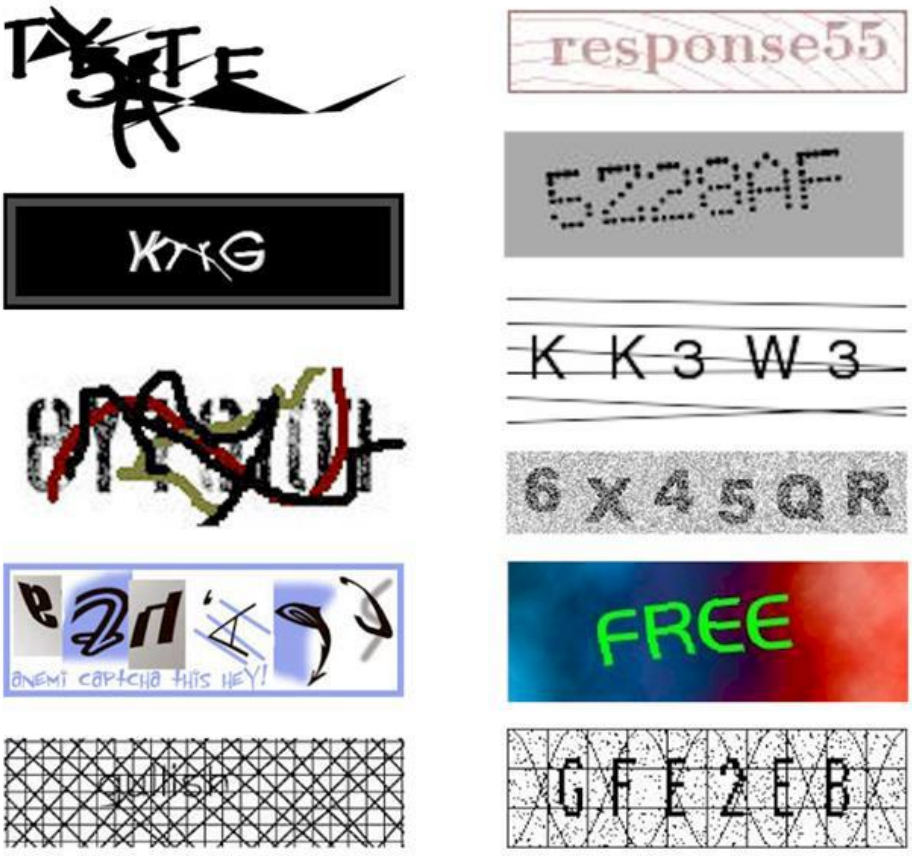
3. Χρήση διανυσματικών μετασχηματισμών σε κάθε χαρακτήρα ή στην συνολική εικόνα. Ανάλογα πάλι με την υλοποίηση, μπορεί να είναι απλοί και ευανάγνωστοι γραμμικοί μετασχηματισμοί (π.χ. περιστροφή) ή δυσκολότεροι μη-γραμμικοί (π.χ. warps και vortices).

4. Χρήση περίπλοκων μοτίβων (π.χ. θόρυβο ή πλέγμα) και χρωμάτων (gradients σε χρώματα όμοια με αυτά των γραμμάτων) στο φόντο.

5. Μερική επικάλυψη του κειμένου της εικόνας με παραμορφωμένες γραμμές strikethrough ή άλλα artifacts.

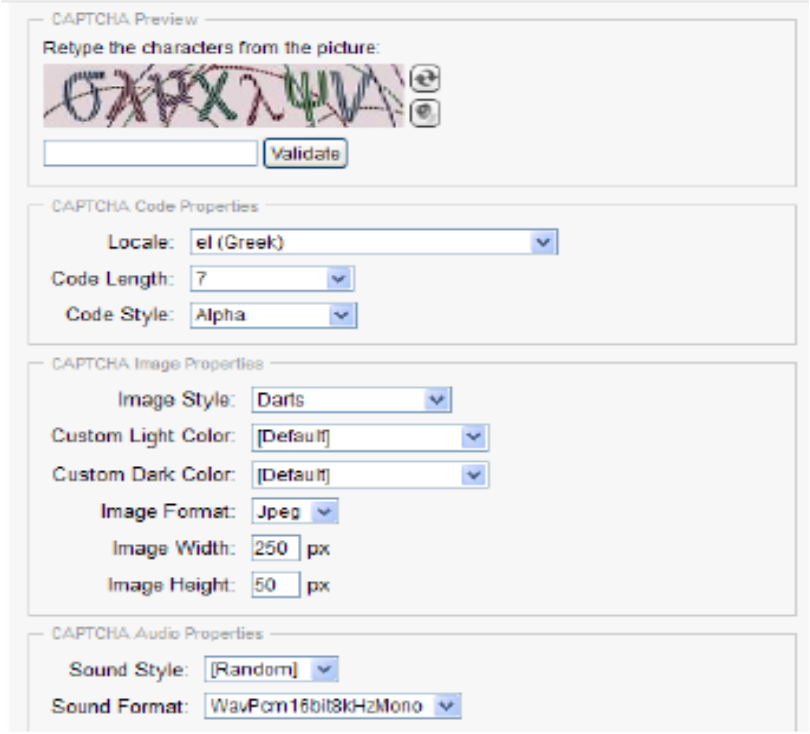
6. Υπερβολική συγκέντρωση των χαρακτήρων στην εικόνα ώστε ο κάθε χαρακτήρας να επικαλύπτει μέρος των διπλανών του.

7. Περιορισμός του χρονικού διαστήματος ισχύος της κάθε δοκιμασίας και αυτόματη αντικατάστασή της με άλλη όταν η προηγούμενη λήξει.



Εικόνα 3.1: Text based CAPTCHA

Η χρήση αυτών των τεχνικών και ο βαθμός ενσωμάτωσής τους στα διάφορα συστήματα CAPTCHAs καθορίζουν σημαντικά την ομαλή λειτουργία των υπηρεσιών τους [07]. Επιπλέον υπάρχουν εταιρίες που ασχολούνται με τη δημιουργία και πώληση CAPTCHAs προσαρμοσμένα στις απαιτήσεις του πελάτη όπως η BotDetect CAPTCHA. Στις παρακάτω εικόνες βλέπουμε ένα παράδειγμα CAPTCHA κειμένου με συγκεκριμένες παραμετροποιήσεις. Συγκεκριμένα ζητήθηκε να είναι στα Ελληνικά, με 7 χαρακτήρες, να αποτελείται μόνο από γράμματα, να έχει format jpeg και μέγεθος 250x50 [37].



Εικόνα 3.2: Παραμετροποίηση Bot Captcha

Εικόνα 3.3: Δείγματα Bot Captchas με διαφορετικές παραμετροποιήσεις

3.3 Δυσκολίες της Χρήσης CAPTCHA

Ο σκοπός δημιουργίας των CAPTCHA είναι η αντιμετώπιση των spam μηνυμάτων αλλά και αποτροπή κακόβουλων επιθέσεων. Ο στόχος του CAPTCHA είναι να επιτυγχάνεται ποσοστό αποτυχίας μικρότερο από 0.01% και το αντίστοιχο ποσοστό επιτυχίας για τον ανθρώπινο παράγοντα να είναι 90% [12].Στις σελίδες του Διαδικτύου μπορεί κάποιος να συναντήσει πολλά είδη CAPTCHA.

Τα CAPTCHA στοχεύουν στο να δείξουν ότι μια ερώτηση που η λύση της είναι απλή και εύκολη για έναν άνθρωπο δεν είναι το ίδιο εύκολη και απλή για έναν υπολογιστή ή μια μηχανή γενικότερα. Βέβαια με την πάροδο των χρόνων και την ανάπτυξη της τεχνολογίας αποκτούν όλο και περισσότερες προοπτικές και δυνατότητες αν και επικρατεί μια ασάφεια γύρω από το γεγονός τι μπορεί να λυθεί από αυτές.

3.3.1 Άτομα με Αισθητικοκινητικές Δυσκολίες

Η χρήση των CAPTCHA και ο βαθμός ενσωμάτωσής τους στα διάφορα συστήματα CAPTCHA καθορίζουν σημαντικά την ομαλή λειτουργία τους. Αυτό όμως πολλές φορές οδηγεί στη δημιουργία δυσκολιών σε χρήστες με αισθητικοκινητικές δυσκολίες, όπως για παράδειγμα σε άτομα με προβλήματα όρασης ή ακόμα και τύφλωσης ή σε άτομα με προβλήματα ακοής [13].

Η πιο ευρέως γνωστή μέθοδος επίλυσης αυτών των προβλημάτων είναι η προσφορά μιας διαφορετικής μορφής CAPTCHA, που θα απευθύνεται σε ανθρώπους με αισθητικοκινητικά προβλήματα για όπως για παράδειγμα τύπο ακουστικής ερώτησης [02].

3.3.2 Δυσκολίες Χρήσης

Οι χρήστες υπολογίζεται ότι καθημερινά ξοδεύουν συνολικά 150.000 εργατοώρες για να αποδείξουν ότι δεν είναι μηχανές, λύνοντας περίπου 60 εκατομμύρια δοκιμασίες CAPTCHA [6]. Οι λόγοι που χρησιμοποιούνται τα CAPTCHA στις διάφορες ιστοσελίδες τους δεν είναι για να κουράζουν τις επισκέπτες τους αλλά για να αποφύγουν τα προβλήματα του online Marketing, το Spam και το Black Hat SEO. Τα προβλήματα γύρω από τα Spam είναι καθαρά ένα ζήτημα που πρέπει να λυθεί από την εταιρία που παρέχει την εκάστοτε υπηρεσία και όχι να υπάρχει αυτή η απαίτηση από τον ίδιο τον χρήστη. Ένα μέσο CAPTCHA απαιτεί ένα χρονικό διάστημα περίπου 10 δευτερολέπτων ώστε να επιλυθεί, ενώ για κάποια από αυτά που έχουν δυσκολέψει πιο πολύ απαιτούνται περισσότερες από μια προσπάθειες για να επιλυθούν [05].

Υπάρχουν ορισμένες ιδιότητες που ορίζονται στην ανάπτυξη του CAPTCHA [42].

- Αυτοματοποιημένα (Automated): Τα προγράμματα υπολογιστών θα πρέπει να είναι σε θέση να παράγουν και βαθμολογούν τις δοκιμές.

- Ανοιχτά (Open): Οι υποκείμενες βάσεις δεδομένων και οι αλγόριθμοι που χρησιμοποιούνται για τη δημιουργία και βαθμολόγηση των δοκιμών πρέπει να είναι δημόσιοι, σύμφωνα με την αρχή του Kerckhoffs.

- Εύχρηστα (Usable): Οι χρήστες θα πρέπει να λύνουν εύκολα αυτές τις δοκιμές σε εύλογο χρονικό διάστημα. Η επίδραση της γλώσσας, της φυσικής θέσης, της εκπαίδευσης ή των αντιληπτικών ικανοτήτων κάθε χρήστη πρέπει να είναι ελάχιστη.

- Ασφαλής (Secure): Το πρόγραμμα που δημιουργεί τις δοκιμές θα πρέπει να είναι δύσκολο για τα μηχανήματα να τις επιλύσουν χρησιμοποιώντας οποιοδήποτε αλγόριθμο.

Οι διάφοροι τύποι συστημάτων CAPTCHA αποτελούν έναν τρόπο για ασφαλή περιήγησης των χρηστών στις υπηρεσίες των ιστοτόπων. Κανένας όμως από αυτούς δεν δίνει την δυνατότητα επιλογής του στον χρήστη. Από τη φύση τους, οι άνθρωποι έχουν μεγάλη ανταπόκριση να απαντούν σε ερωτήσεις, συγκεκριμένα ένας χρήστης εφοδιάζεται με 5 τύπους ερωτήσεων CAPTCHA [38]: Αναλυτικών, Μαθηματικών, Γενικών, Βασισμένων σε κείμενο και, βάση εικόνες. Ο χρήστης του ισότοπου πρέπει να απαντήσει εντός της καθορισμένης προθεσμίας. Το πρόβλημα το οποία γενάτε όμως είναι να βρεθεί ένας τρόπος στο να χρησιμοποιηθεί παραγωγικά ο χρόνος του χρήστη ώστε να ελαττώσει το αίσθημα ενόχλησης. Το βασικότερο παράδειγμα είναι η υπηρεσία ReCAPTCHA [22], που στόχο έχει να εμφανίσει στους χρήστες δυο λέξεις από βιβλία που βρίσκονται στο δρόμο της ψηφιοποίησης και καμία εικόνα δεν πέρασε επιτυχώς από το πρόγραμμα OCR. Η σκέψη γύρω από αυτή την υπηρεσία είναι ότι οι χρήστες δεν έχουν θέμα χρόνου και δυσκολίας σχετικά με την επίλυση του τεστ αφού με αυτόν τον τρόπο βοηθούν συνολικά..

Ωστόσο, σε μια δοκιμή μικρής κλίμακας που διεξήχθη με 20 φοιτητές τον Οκτώβριο του 2007 [37,13], παρατηρήθηκε ότι πολλοί ξένοι φοιτητές των οποίων στην μητρική γλώσσα δεν χρησιμοποιούνται τα Λατινικά στο αλφάβητο τους ήταν πολύ χειρότερο από εκείνων των οποίων η πρώτη γλώσσα βασίζεται σε λατινικό αλφάβητο. Ουσιαστικά τους ζητήθηκε να αναγνωρίσουν τις παραμορφωμένες προκλήσεις που προκαλεί το BaffleText. Ο πρώτος δυσκολεύτηκε να αναγνωρίσει, ούτε καν να μαντέψει, τα παραμορφωμένα γράμματα.

Image	Confusing characters
	Is the middle part 'd" or connected "cl"?
	Another case of "cl" or "d" confusion.
	Another case of "cl" or "d" confusion.
	Is the starting part 'm' or connected 'rn"?
	The 2nd and the 3rd character could be confused with "w".
	A real headache: is the first part "m" or "rn", the middle part "inv" or "nw"?

Εικόνα 3.4: Confusing characters in the Google CAPTCHA

Κατά συνέπεια ένας χρήστης που καλείται να χρησιμοποιήσει μια τέτοια υπηρεσία στις διάφορες ιστοσελίδες θα πρέπει οι δημιουργοί της να λαμβάνουν πολύ σοβαρά υπόψιν τους την ευχρηστία του σχεδιασμού και την χρηστικότητα διότι έχουμε παρατηρηθεί τα ακόλουθα θέματα [13]:

- Αντίθετα με την κοινή πεποίθηση, τα CAPTCHA που είναι βασισμένα σε ένα κείμενο μπορεί να είναι δύσκολο για τους αλλοδαπούς χρήστες.

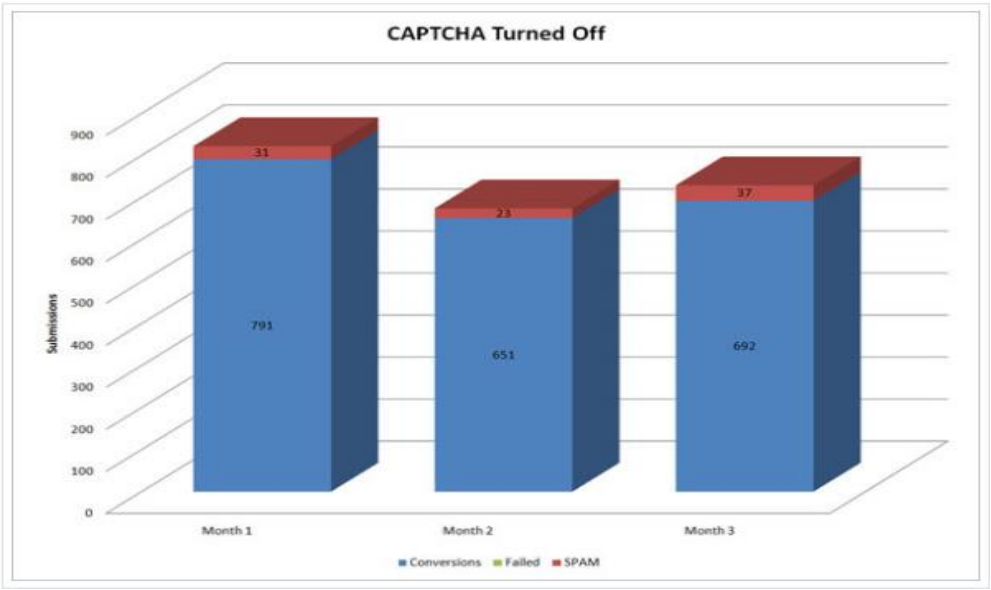
- Το μήκος των συμβολοσειρών που χρησιμοποιούνται σε ένα σχήμα ανεξάρτητα αν είναι προβλέψιμο ή όχι μπορεί να έχει σημαντικές επιπτώσεις τόσο την ασφάλεια όσο και τη χρηστικότητα.

- Η χρήση του χρώματος σε ένα CAPTCHA μπορεί να έχει αντίκτυπο στην χρηστικότητα ή στην ασφάλεια ή και στα δύο.

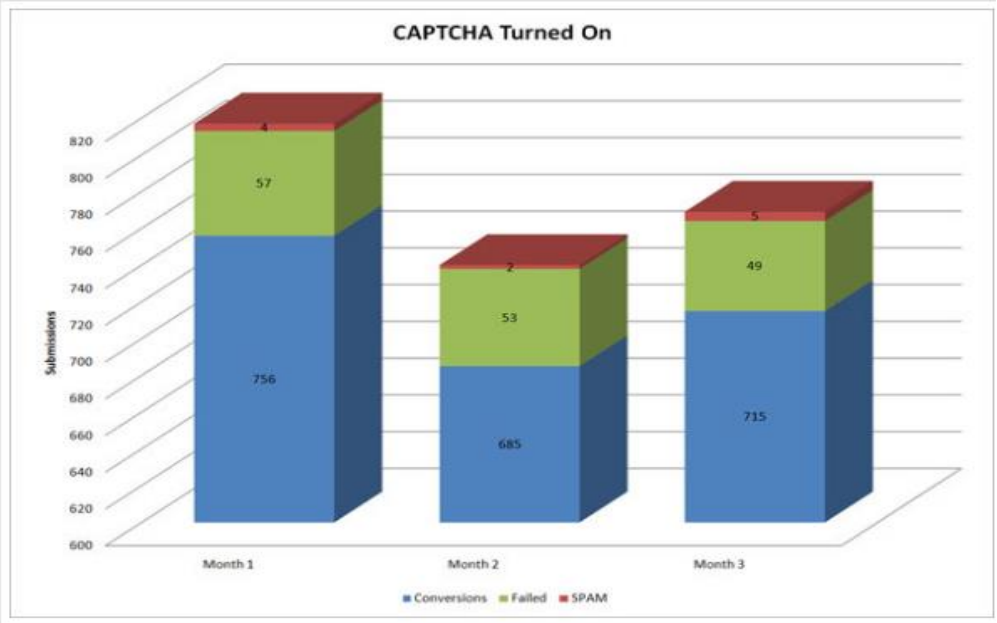
3.3.3 Ελάττωση Conversion Rate

Conversion (μετατροπή) είναι η επιθυμητή ενέργεια που εκτελεί ένας επισκέπτης πάνω σε μια ιστοσελίδα. Σε ένα OnLine κατάστημα Conversion είναι η ολοκλήρωση της αγοράς. Σε ένα site που

έχει σκοπό την προσέλκυση νέων πελατών, Conversion είναι η αποστολή της φόρμας επικοινωνίας, ή η εγγραφή στο newsletter. Το Conversion rate είναι από τα πιο σημαντικά στοιχεία για την πορεία μιας ιστοσελίδας (KPI - Key Performance Indicator) καθώς δείχνει το πόσο καλά μετατρέπει την επισκεψιμότητα σε τζίρο, νέους υποψήφιους πελάτες, εγγραφές. Το CAPTCHA σύμφωνα με μελέτη που έχει γίνει μειώνει αρκετά το conversion Rate με αποτέλεσμα να έχει αρνητική επίδραση στις Ιστοσελίδες. Η μελέτη που έγινε κατά τη διάρκεια των 6 μηνών, το ήμισυ του ιστότοπου ξεκίνησε με το CAPTCHA και το άλλο μισό ξεκίνησε χωρίς CAPTCHA [26, 28]. Τα αποτελέσματα της μελέτης παρουσιάζονται στα παρακάτω εικόνες.



Εικόνα 3.5: CAPTCHA Turned off



Εικόνα 3.6: CAPTCHA Turned on

Σύμφωνα με τα παραπάνω δεδομένα με την χρήση του CAPTCHA φαίνεται να υπήρξε μια μείωση του SPAM κατά 88%, αλλά υπήρξαν 159 αποτυχημένες μετατροπές. Αυτές οι αποτυχημένες μετατροπές θα μπορούσαν να είναι SPAM, αλλά θα μπορούσαν επίσης να είναι άνθρωποι που δεν μπορούσαν να καταλάβουν το CAPTCHA και εγκατέλειψαν την προσπάθεια. Με την ενεργοποίηση του CAPTCHA, το Spam και οι αποτυχημένες μετατροπές αντιπροσώπευαν το 7,3% όλων των μετατροπών για την περίοδο των 3 μηνών. Με την εκτός λειτουργίας του CAPTCHA, οι μετατροπές Spam αντιπροσώπευαν το 4,1% όλων των μετατροπών για την περίοδο των 3 μηνών. Αυτό πιθανότατα σημαίνει ότι με την χρήση των CAPTCHA η εταιρεία θα μπορούσε να χάσει το 3,2% όλων των μετατροπών τους. Πρόκειται για μια σημαντική τιμή καθώς αυτή είναι που προσφέρει κέρδη σε μια εταιρία είτε με τις διαφημίσεις είτε με τις πωλήσεις. Κατά συνέπεια η ευχρηστία των CAPTCHA αποτελεί έναν σημαντικό παράγοντα για την οικονομική ευρωστία των sites [26].

Κεφάλαιο 4
Τεχνικές CAPTCHA

Όλα τα αναπτυγμένα CAPTCHA βασίζονται στη στατική ταυτοποίηση ενός αντικειμένου, ενός κειμένου ή την αναγνώριση ενός αντικειμένου σε μια εικόνα. Στην συνέχεια προστέθηκε και η φωνητική λειτουργεία για την διευκόλυνση των χρηστών με προβλήματα όρασης. Τα τελευταία χρόνια ανανεώθηκαν οι μέθοδοι με εφαρμογές που στηρίζονται στην αναγνώριση της εικόνας ή γενικά κάποιας ενέργειας που μια μηχανή είναι δύσκολο να κάνει.

Στο κεφαλαίο αυτό θα γίνει μελέτη γύρω από τα είδη τεστ που υπάρχουν και θα εστιάσουμε στα κύρια χαρακτηριστικά τους καθώς και στις θετικές και αρνητικές επιδράσεις που έχουν κατά τη λειτουργία τους και τις επιθέσεις που δέχονται.

4.1 Ερωτήσεις Κειμένου (Question-Based CAPTCHA)

Προτεινόμενο το 2009, το CAPTCHA βασισμένο σε ερωτήσεις διερεύνησε διάφορες δεξιότητες ενός χρήστη μέσω μιας ερώτησης στην οποία μπορεί να απαντηθεί μόνο από άνθρωπο. Η αρχική

μορφή του CAPTCHA ήταν απλές ερωτήσεις κειμένου που ο χρήστης έπρεπε να απαντήσει. Οι ερωτήσεις έχουν ως βάση τη λογική και η λύση είναι μια απλή λέξη.

Εικόνα 4.1: Question Captcha

Αυτού του είδους το CAPTCHA αποτελεί ένα συνδυασμό OCR και Non OCR βάσης CAPTCHA. Σε αυτόν τον τύπο ένα απλό μαθηματικό πρόβλημα δημιουργείται σύμφωνα ένα προκαθορισμένο πρότυπο και στην συνέχεια εμφανίζεται στον χρήστη με τη μορφή μιας εικόνας στην οποία υπάρχει ένα ερώτημα. Για την απάντηση σε αυτό το πρόβλημα απαιτούνται τέσσερα βασικές ικανότητες: Κατανόηση του κειμένου της ερώτησης, Ανίχνευση εικόνων ερωτήσεων, Κατανόηση του προβλήματος και την Επίλυση του προβλήματος. Ένας άνθρωπος μπορεί να απαντήσει στην ερώτηση και σε πολύ σύντομο χρονικό διάστημα αλλά όχι τα προγράμματα ηλεκτρονικών υπολογιστών [15].

Η πιο πρόσφατη μορφή CAPTCHA είναι η τοποθέτηση πλήθους ερωτήσεων σε μια βάση δεδομένων και η χρήση τους από πολλές εφαρμογές συγχρόνως. Οι ερωτήσεις ανέρχονται σε εκατομμύρια δεν έχουν μια συγκεκριμένη μορφή και δεν μπορεί να επιλυθεί με ένα γενικό αλγόριθμο.

4.1.1 Μειονεκτήματα Χρήσης Ερωτήσεων Κειμένου.

Απουσία Ποικιλόγλωσσης Υποστήριξης

Το βασικότερο μειονέκτημα του CAPTCHA είναι ότι όλες οι ερωτήσεις κειμένου είναι στην αγγλική γλώσσα και έχουν καλύτερη πρόσβαση μόνο όσοι χρήστες γνωρίζουν την αγγλική. Η πλειονότητα των χρηστών ιστοσελίδων γνωρίζουν τα βασικά της αγγλικής γλώσσας και αυτομάτως και τα βασικά πεδία της κάθε φόρμας. Αυτό έχει ως αποτέλεσμα μια ερώτηση να κάνει δύσκολη την πρόσβαση σε αγγλικές ιστοσελίδες σε χρήστες με βασικές γνώσεις. Παρόλο που υπάρχει η δυνατότητα οι εφαρμογές με τέτοιου είδους ερωτήσεις να γραφούν και σε άλλες γλώσσες αντιμετωπίζουν κάποια προβλήματα [13].

Ανάγκη Μεγάλου Πλήθους Ερωτήσεων

Ο επιτιθέμενος είναι σε θέση πολλές φορές εξαιτίας των επαναλαμβανομένων ερωτήσεων να πραγματοποιήσει μια επίθεση, είτε εισάγοντας με χειροκίνητο τρόπο την λύση για κάθε ερώτηση ή και αυτόματα σε κάποια είδη κοινών ερωτήσεων. Συνήθως οι ερωτήσεις είναι στάνταρ και με έναν αυτόματο τρόπο μπορούν να χρησιμοποιηθούν στις ήδη αποθηκευμένες απαντήσεις. Κατά συνέπεια ένα μεγάλο πλήθος ερωτήσεων μπορεί να αποθαρρύνει τον επιτιθέμενο και να εγκαταλείψει τον ιστότοπο [29].

4.1.2 Τυπικές Επιθέσεις

Ευκολία Χειροκίνητων Απειλών

Η πιο σημαντική επίθεση που μπορεί να συμβεί στα CAPTCHA είναι η χειροκίνητη εισαγωγή απαντήσεων ή κανόνων για την απάντηση τέτοιων ερωτήσεων. Το πρόβλημα προκύπτει από το γεγονός ότι και οι δημιουργοί των σελίδων δεν θέλουν να παιδέψουν τους χρήστες κατά συνέπεια τα CAPTCHA να είναι εύκολα σε χρήση και να εξοικονομούν χρόνο στους χρήστες καθώς και να εκτελούνται από συσκευές που απαιτούν μικρούς υπολογιστικούς πόρους αφήνοντας όμως ένα μεγάλο κενό ασφάλειας [15].

Λειτουργία Αλγόριθμων Τεχνητής Νοημοσύνης

Με βάση μια έρευνα που έγινε πρόσφατα πρότεινε τη χρήση της μηχανής αναζήτησης *Wolfram Alpha* ως επίθεση στις ερωτήσεις κειμένου CAPTCHA. Η *Wolfram Alpha* κάνει χρήση αλγόριθμων

τεχνητής νοημοσύνης με στόχο να δώσει απαντήσεις σε ερωτήσεις που έχουν δημιουργηθεί με ανθρώπινη γλώσσα και με αυτό τον τρόπο θέλει να δώσει μια διεξοδική απάντηση σε σχέση με την ερώτηση. Η καινοτομία της επίθεσης αυτής είναι στο γεγονός ότι χρησιμοποιεί μια δωρεάν υπηρεσία του διαδικτύου [37].

Εικόνα 4.2: Η μηχανή αναζήτησης Wolfram Alpha δίνει μία επιτυχή απάντηση σε μία ερώτηση ενός τεστ CAPTCHA.

Λειτουργία Λογικών Φράσεων

Το πλήθος των επιθέσεων αποτελείται από κανόνες για την απάντηση συγκεκριμένων ερωτήσεων με βάση τη λογική χρήση εκφράσεων. Αυτού του είδους η επίθεση αντικρούεται με τη χρησιμοποίηση μεγαλύτερης ποικιλίας ερωτήσεων ώστε οι προσπάθειες για λύσεις να είναι μηδαμινές [15].

Επίθεση Laundry

Από τα πιο απλά CAPTCHA όπως είναι η απάντηση σε μια ερώτηση ή η πληκτρολόγηση ενός παραμορφωμένου κείμενο μέχρι τα πιο προηγμένα την αναγνώριση ενός αντικειμένου σε μια εικόνα, είναι ευάλωτα στην επίθεση Laundry (Laundry attack). Σε αυτό το είδος της επίθεσης ο εισβολέας δημοσιεύσει τη δοκιμή του CAPTCHA σε έναν κακόβουλο ιστότοπο και προσελκύσει επισκέπτες προκειμένου να λύσουν το πάζλ. Εκμεταλλεύεται τους ανυποψίαστους χρήστες οι οποίοι πιστεύουν ότι λύνουν το CAPTCHA προκειμένου να κάνουν χρήση της υπηρεσίας που τους ενδιαφέρει αλλά, ουσιαστικά προωθούν την λύση της δοκιμής στον επιτιθέμενο. Σε περίπτωση που ο κακόβουλος ιστότοπος δεν έχει μεγάλη δημοτικότητα μπορεί περιοδικά να ζητηθεί οι χρήστες να λύσουν παραπάνω από ένα CAPTCHA [01, 04].

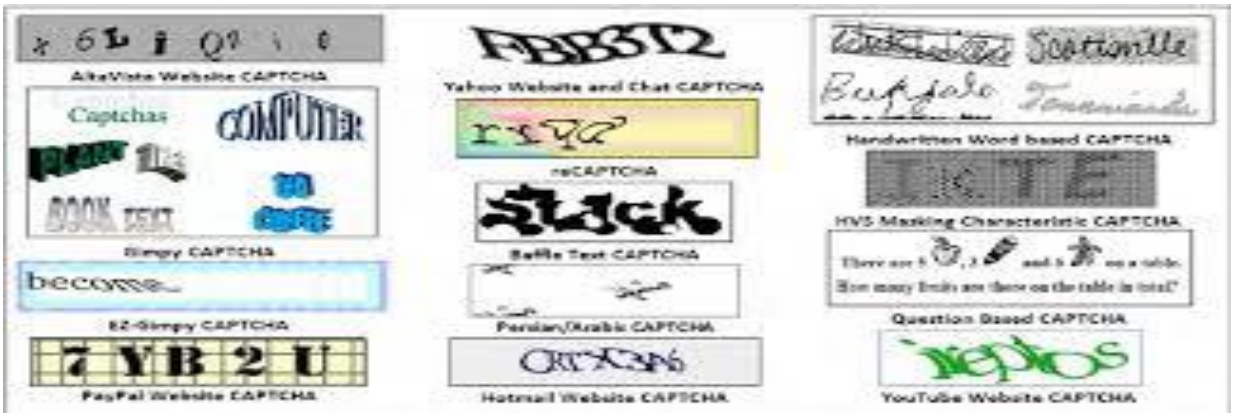
Για να περιοριστεί αυτή η μορφή επίθεσης χρησιμοποιούμε τα κινούμενα CAPTCA (Animated CAPTCHA). Δεν έχουν κάποια στατική απάντηση, επομένως ακόμα και όταν εκτίθενται σε Laundry attacks οι ανυποψίαστοι επισκέπτες θα δώσουν απαντήσεις που θα είναι άχρηστες στον επιτιθέμενο. Η λύση αυτού του CAPTCHA προϋποθέτει την κίνηση του ποντικιού, συνεπώς έστω και αν ένας εισβολέας γνωρίζει την απάντηση της δοκιμής δεν μπορεί να περάσει αυτό το τεστ χωρίς την ανθρώπινη παρέμβαση [04,10]

4.2 CAPTCHA Βασισμένα σε Κείμενο (Text-Based)

Στην παρακατω ενότητα θα αναλυθούν τα CAPTCHA κειμένου, θα παρουσιαστούν τα μειονεκτήματα από την χρήση τους καθώς και τυπικές επιθέσεις που μπορούν να δεχτούν.

4.2.1 Περιγραφή

Η πρώτη μορφή CAPTCHA που γνωρίζουμε είναι εκείνα τα πλαίσια με λέξεις που τα γράμματά τους είναι παραμορφωμένα. Η παραμόρφωση αυτή γίνεται με την περιστροφή των γραμμάτων ή των αριθμών, την αλλαγή του χρώματος και ακόμη με την εμφάνιση κάποιων γραμμών ώστε να γίνεται λίγο δυσνόητη η ανάγνωση τους.



Εικόνα 4.3: Παραδείγματα Text Based CAPTCHA

Θεωρείται η πιο γνωστή μορφή και γι' αυτό έχει αναπτυχθεί πιο πολύ. Ενώ ξεκίνησε με μικρές παραμορφώσεις κατά τη διάρκεια των χρόνων έγιναν πολλές μελέτες για να δυσκολέψουν την ανάγνωση από τους υπολογιστές [05].

4.2.2 Μειονεκτήματα

Μπερδεμένοι Χαρακτήρες

Αυτή η μορφή CAPTCHA με τη παραμόρφωση χαρακτήρων διακρίνεται για κάποιες αδυναμίες. Μια από τις πιο σημαντικές είναι οι προσπάθειες που γίνονται για να δυσκολέψουν την ανάγνωση από τους υπολογιστές, αυτό έχει ως αποτέλεσμα να εξαιρούνται από τη προσπάθεια μεγάλη μερίδα ανθρώπων. Παρά το γεγονός ότι κάποιοι χαρακτήρες έχουν πολύ διαφορετικά σχήματα, μετά την παραμόρφωση γίνεται δύσκολο να ξεχωρίσει κάποιος το έναν από τον άλλο.

Για παράδειγμα όπως το L και το I. "VV" μπορεί να μοιάζει με "W", το "cl" μπορεί να μοιάζει με "d", το "nn" μπορεί να μοιάζει με "m", Ακόμη η αλλαγή χρώματος μπορεί να κάνει δυσανάγνωστα ή θολά κάποια γράμματα [01, 13]

Περιορισμένοι Χρήστες.

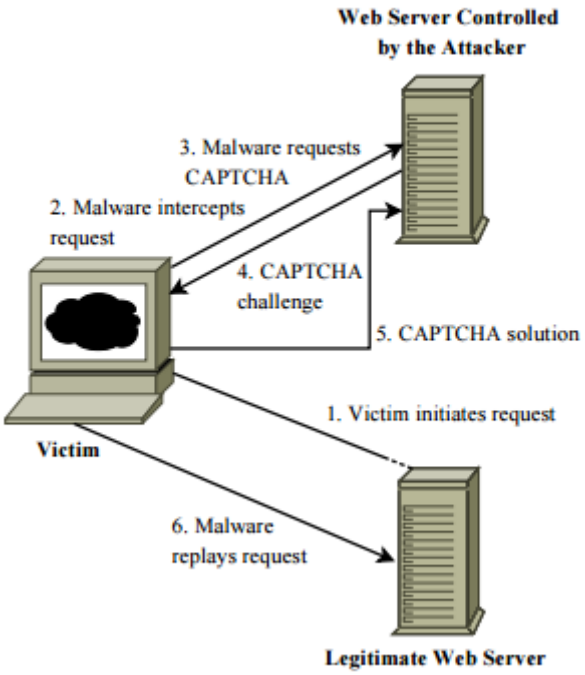
Τα CAPTCHA αυτής της μορφής στηρίζονται στο γεγονός ότι ο κάθε χρήστης που καλείται να δώσει απαντήσεις μπορεί να δει με ευκρίνεια την εικόνα, συνεπώς χρήστες που έχουν πρόβλημα με την όρασή τους έχουν πρόβλημα να περάσουν την δόκιμη. Σύμφωνα με μελέτες που έχουν γίνει άτομα τα οποία δεν έχουν στην μητρική τους γλώσσα το λατινικό αλφάβητο έχουν δυσκολία στην λύση του συγκεκριμένου τύπου CAPTCHA, επίσης παρουσιάζουν και μεγαλύτερη καθυστέρηση [13]. Μελέτη που έγινε από τον Luis von Ahn, ιδρυτής της εταιρείας ReCAPTCHA, παρατήρησε ότι η διαφορετική μητρική γλώσσα έχει διαφορετικές επιδόσεις στην αποκωδικοποίηση παραμορφωμένων λατινικών χαρακτήρων. Συγκεκριμένα η επίλυση προκλήσεων ReCAPTCHA είχε ένα μέσο ποσοστό επιτυχίας της τάξης του 97% και 93% για την στη διάρκεια της ημέρας και το βράδυ αντίστοιχα. Σύμφωνα με τις διευθύνσεις IP των αιτήσεων παροχής υπηρεσιών που είχε λάβει το ReCAPTCHA, από περισσότερους χρήστες εκτός των ΗΠΑ (π.χ. στην Ασία) ζητούν πρόσβαση σε αυτή την υπηρεσία τη νύχτα παρά στη διάρκεια της ημέρας (και ώρα ΗΠΑ) καθώς το βράδυ στις ΗΠΑ είναι ημέρα στην Ασία [37].

4.2.3. Τυπικές Επιθέσεις

Επιθέσεις Smuggling

Για να εκτελεστεί μια επιτυχημένη επίθεση Smuggling CAPTCHA, χρησιμοποιείται για παράδειγμα ένα πρόγραμμα botnet, προγράμματα που είναι ειδικά σχεδιασμένα για να εκτελούν επιθέσεις DDos και επιτρέπουν την πρόσβαση ενός εισβολέα στην λειτουργιά και τη χρήση μιας συσκευή. Ο εισβολέας έχει την δυνατότητα μετά την εγκατάσταση του κακόβουλου προγράμματος στον υπολογιστή του θύματος να παρεμποδίζει τις αλληλεπιδράσεις των χρηστών με μια ηλεκτρονική υπηρεσία (π.χ. Facebook) μπορεί να καθυστέρει την εκτέλεση τους έως ότου το θύμα επιλύσει επιτυχώς μια πρόκληση CAPTCHA. Δεν είναι απαραίτητο να δημιουργηθεί ένα νέο botnet πρόγραμμα για επιθέσεις Smuggling CAPTCHA, τα bot που ήδη προϋπάρχουν μπορούν εύκολα να επεκταθούν και να εκτελέσουν τέτοιες επιθέσεις.

Σε μια τυπική επίθεση, ο χρήστης εκτελεί πρώτα μια ενέργεια που ο εισβολέας επιθυμεί να καθυστερήσει (για παράδειγμα κλικ σε ένα κουμπί σε μια ιστοσελίδα). Το κακόβουλο λογισμικό είναι εγκατεστημένο στον κεντρικό υπολογιστή του θύματος και ενεργοποιείται μόλις το θύμα εκτελέσει μια συγκεκριμένη ενέργεια για παράδειγμα μόλις ζητήσει να κάνει μια εγγραφή σε ένα ηλεκτρονικό ταχυδρομείο, του προωθεί μια δοκιμή CAPTCHA που το ζήτα να λύσει προκειμένου να ολοκλήρωση την εγγραφή του. Για το ανυποψίαστο θύμα, φαίνεται ότι η εφαρμογή του Ιστότοπου χρησιμοποιεί έναν τυπικό μηχανισμό CAPTCHA για να προστατευτεί αλλά στην πραγματικότητα το θύμα μόλις παρείχε στον επιτιθέμενο όλες τις απαραίτητες πληροφορίες για να συνεχίσει τις κακές του αποστολές. Ουσιαστικά πρόκειται για μια Man-in-the-Mindle Attack.



Σχήμα 4.4: Επιθέσεις λαθρεμπορίου (CAPTCHA Smuggling)

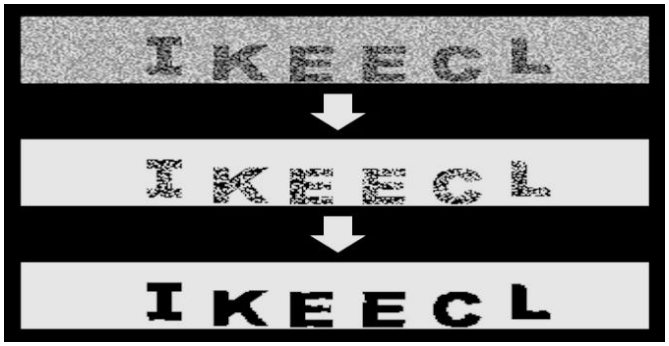
Ένα χαρακτηριστικό παράδειγμα επίθεσης Smuggling έγινε το 2009, εγκαταστήσαν σε 17 υπολογιστές εθελοντών χρηστών ένα Plugin του Firefox που περιείχε κακόβουλο λογισμικό και στην συνέχεια 5 χρήστες δημοσίευσαν στο προφίλ του Facebook τους μια ψεύτικη σελίδα με ένα βίντεο που για να ανοίξει μπορούσε μόνο με τη χρήση Firefox. Οι χρήστες του Facebook που ήθελαν να δουν το βίντεο άνοιγαν το Firefox, αυτό είχε σαν αποτέλεσμα να γίνεται αυτόματα η εγκατάσταση του Plugin. Έτσι μέσω του συγκεκριμένου Plugin επιτεύχθηκε επίθεσης Smuggling όπου στον εξυπηρετητή των εισβολέων καταγράφηκαν όλα τα αιτήματα CAPTCHA από τις ιστοσελίδες Facebook και Gmail [16].

OCR-Based CAPTCHA

Η έρευνα και η μελέτη ενός τομέα της επιστήμης των υπολογιστών έχει ως αντικείμενο την ανάλυση της εικόνας και την αναγνώριση των χαρακτήρων με την οπτική επαφή. Όπως όλα τα προγράμματα έτσι και τα CAPTCHA έχουν τις αδυναμίες τους. Η επικρατούσα εφαρμογή του CAPTCHA είναι λύση του κειμένου 2D φωτογραφιών[08], ωστόσο οι αναπτυσσόμενες τεχνολογίες τεχνητής νοημοσύνης και αναγνώρισης εικόνων επιτρέπουν στα προγράμματα ηλεκτρονικών υπολογιστών να περάσουν από αυτή τη δοκιμασία [10].

Εικόνα 4.3: Παράδειγμα OCR CAPTCHA

Η οπτική αναγνώριση χαρακτήρων (OCR, Optical Character Recognition) είναι δύσκολη εφόσον οι εικόνες των δοκιμασιών CAPTCHA έχουν κατασκευαστεί με ακριβώς τέτοιο τρόπο, ώστε να δυσκολεύουν τις συμβατικές μεθόδους OCR. Οι αλγόριθμοι OCR που χρησιμοποιούνται για «σπάσουν» δοκιμασίες CAPTCHA επικεντρώνονται σε συγκεκριμένες υλοποιήσεις (π.χ. Google Mail CAPTCHA) και δεδομένης της εφαρμογής τους θεωρούνται αποδοτικοί ακόμα και αν λύνουν μόνο ένα αρκετά χαμηλό ποσοστό δοκιμασίων.



Εικόνα 4.5: Οπτική Αναγνώριση Χαρακτήρων (OCR)

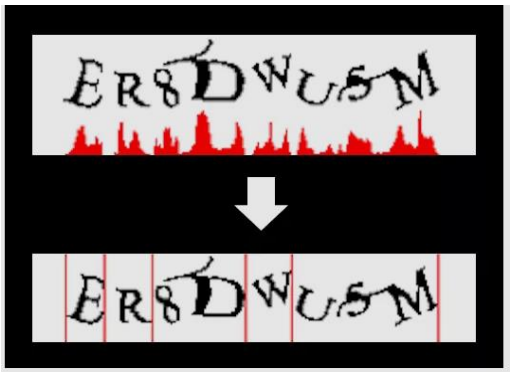
Παρά τις ειδικές διαφορές που μπορεί να έχουν ανάλογα με την υλοποίηση που στοχεύουν, όλοι αυτοί οι αλγόριθμοι δουλεύουν σε 3 κοινά στάδια:

1. Αφαίρεση της άχρηστης πληροφορίας (φόντο, θόρυβος, artifacts κλπ).

2. Κατάτμηση της εικόνας σε επιμέρους περιοχές, ώστε κάθε περιοχή να περιέχει ένα χαρακτήρα.

3. Αναγνώριση του χαρακτήρα που περιέχεται σε κάθε περιοχή.

Με την σημερινή υπολογιστική δύναμη των μηχανών και με τις διάφορες τεχνικές εκμάθησης που έχουν αναπτυχθεί για αυτές (όπως τα νευρωνικά δίκτυα), τα πρώτα στάδια υλοποιούνται αρκετά εύκολα και με τόσο μεγάλη απόδοση, ώστε οι χαρακτήρες να αναγνωρίζονται με μεγάλη επιτυχία και με ταχύτητα που μπορεί να ξεπερνά ακόμα και την ανθρώπινη [29].



Εικόνα 4.6:Οπτική Αναγνώριση Χαρακτήρων (OCR)

Μια λύση για τον περιορισμό των επιθέσεων OCR είναι η χρήση της μεθόδου Non-OCR-Based (Μη Βασισμένο σε OCR) είναι μια νέα προσέγγιση για την εφαρμογή του μηχανισμού CAPTCHA βασισμένου σε 3D Animation [06]. Έχουν την μορφή μικρών παιχνιδιών ή ζητούν από τους χρήστες να επιλέγουν εικόνες σε συγκεκριμένο θέμα μεταξύ των εικόνων διαφόρων θεμάτων. Αυτή η μέθοδος βασίζεται στην παραδοχή ότι ο άνθρωπος μπορεί να αναγνωρίσει την εικόνα 3D χαρακτήρα καλύτερα από τα bots λογισμικού οπτικής αναγνώρισης χαρακτήρων (OCR) [04, 10, 35].

Αποστολή σε Τρίτους

Η αποστολή σε τρίτους είναι μια διαδικασία η οποία λειτουργεί ακόμα και στις πιο ασφαλείς υλοποιήσεις CAPTCHA, κάνει χρήση του ανθρωπίνου δυναμικού. Ο Spammer δημιουργεί ένα bot

που αναλαμβάνει να κάνει αυτόματα όλες τις διαδικασίες που χρειάζονται (δημιουργία λογαριασμών, ανάρτηση δημοσιεύσεων και σχολίων κ.λπ.) εκτός από την λύση των δοκιμασιών CAPTCHA. Επίσης προσλαμβάνει εργάτες στους οποίους το bot ανακατευθύνει μαζικά τις δοκιμασίες CAPTCHA που συναντά κατά την λειτουργία του. Οι εργάτες λύνουν την μια δοκιμασία μετά την άλλη και να στέλνουν τις λύσεις πίσω στο bot, το οποίο τις προωθεί στους αντίστοιχους ιστότοπους ώστε να ξεπεράσει τις δοκιμασίες και να συνεχίσει απρόσκοπτα την δουλειά του. Ο τρόπος αυτός είναι και ο πιο διαδεδομένος ανάμεσα στις μεγάλες εταιρίες Spam καθώς είναι πολύ αποδοτικός και σχεδόν αδύνατον να αντιμετωπιστεί [24].

Μια παραλλαγή του παραπάνω τρόπου επίθεσης αλλά περισσότερο οικονομικός είναι ότι ο spammer αντί να προσλάβει ανθρώπους άμεσα, δημιουργεί έναν υψηλής κίνησης ιστότοπο στον οποίο οι επισκέπτες προκαλούνται να λύσουν δοκιμασίες CAPTCHA ως αντίτιμο για κάποια υπηρεσία (π.χ. για κάθε δοκιμασία που λύνει ο επισκέπτης λαμβάνει μια εικόνα). Σε αυτόν τον τρόπο ωστόσο, η ικανοποιητική απόδοση του bypassing εξαρτάται σε πολύ μεγάλο βαθμό από την κίνηση του ιστότοπου, η οποία δεν είναι ιδιαίτερα εύκολο να δημιουργηθεί [24].

CAPTCHA Redirection

Σε αυτήν την μορφή επίθεσης το bot στέλνει μια εικόνα διαστρεβλωμένη σε ένα πολυσύχναστο ιστότοπο και ο απασχολημένος χρηστής λύνει την δοκιμή CAPTCHA νομίζοντας ότι είναι απαραίτητη για να συνεχίζει να κάνει χρήση της υπηρεσίας του ιστότοπου. Χωρίς να το ξέρει έχει βοηθήσει το bot να πάρει την απάντηση από τους χαρακτήρες της εικόνας και λύσει την δοκιμή [04].

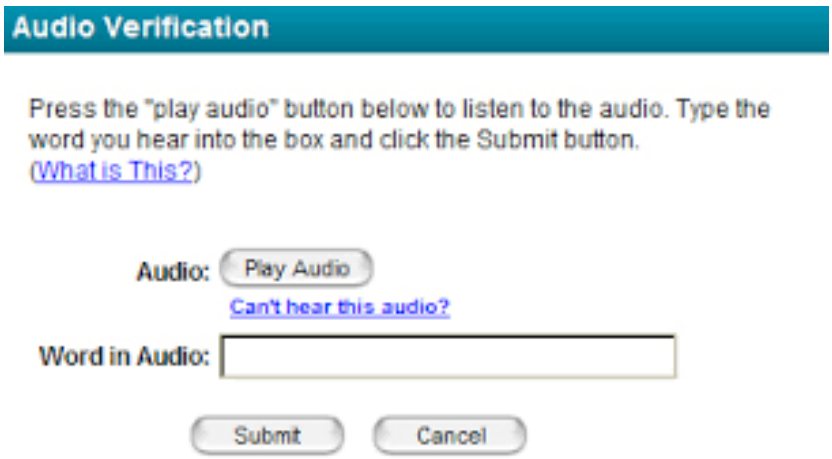
4.3 Ηχητικά CAPTCHA (AUDIO CAPTCHA)

Στην παρακατω ενότητα θα αναλυθούν τα CAPTCHA ήχου, θα παρουσιαστούν τα μειονεκτήματα από την χρήση τους καθώς και τυπικές επιθέσεις που μπορούν να δεχτούν.

4.3.1 Περιγραφή

Ένας παράγοντας που οδήγησε στη δημιουργία CAPTCHA ήχου είναι η ανάπτυξη της ψηφιοποίησης παλαιότερων εντύπων προκειμένου να αντέξουν στο χρόνο και να μπορούν να

διανεμηθούν πιο εύκολα. Τα περισσότερα CAPTCHA αποτελούνται από παραμορφωμένες εικόνες, συνήθως κείμενο, για τις οποίες ένας χρήστης πρέπει να παρέχει κάποια περιγραφή. Δυστυχώς, τα οπτικά CAPTCHA περιορίζουν την πρόσβαση στα άτομα με προβλήματα όρασης που χρησιμοποιούν το internet. Τα CAPTCHA ήχου δημιουργήθηκαν για την επίλυση του συγκεκριμένου προβλήματος προσβασιμότητας. Ωστόσο, η ασφάλεια του CAPTCHA ήχου δεν ελεγχθεί επίσημα [38]. Η λογική της λειτουργίας των Audio CAPTCHA είναι ότι γενικά αποτελούνται από ένα σύνολο λέξεων που πρέπει να προσδιοριστούν, να είναι στρωμένες πάνω σε διάφορους θορύβους πχ μουσική γεγονός που καθιστά πιο δύσκολη την διαδικασία αναγνώρισης [42].



Εικόνα 4.7: Audio Verification

4.3.2 Μειονεκτήματα

Δυσκολία Γλώσσας

Ο αρχικός λόγος δημιουργίας των Audio CAPTCHA ήταν για άτομα με ειδικές ανάγκες με στόχο να ξεπεράσουν τα προβλήματα πρόσβασης που δημιουργούσαν οι υπόλοιποι τύποι CAPTCHA στις ιστοσελίδες που επισκεπτόντουσαν. Παρ' όλα αυτά υπάρχει δυσκολία για παράδειγμα, η ηχητική έκδοση του ReCAPTCHA χρησιμοποιεί ως θόρυβο ηχητικό κλιπ στη μητρική γλώσσα των Ναβάχο, αυτή η φυσική γλώσσα χρησιμοποιείται μόνο από ένα πολύ περιορισμένο αριθμό των ανθρώπων στον κόσμο και χρησιμοποιήθηκε στο Δεύτερο Παγκόσμιο Πόλεμο ως ένα άσπαστο ραδιόφωνο κρυπτογράφησης για τον ίδιο λόγο [37]. Επίσης σε σχέση με τα CAPTCHA παραμόρφωσης που ήθελαν αναγνώριση κάποιων χαρακτήρων, τα ηχητικά θέλουν τον χρήστη να γνωρίζει τη γλώσσα του ηχητικού μιας και τα περισσότερα είναι στην αγγλική γλώσσα.

Μη Συμβατότητα

Υπάρχει πολύ μεγάλη πιθανότητα όταν τα CAPTCHA ήχου είναι ενσωματωμένα πάνω σε ιστοσελίδες να εμφανίσουν πρόβλημα ως προς την συμβατότητα του. Για παράδειγμα υπάρχουν πολλά συστήματα που απαιτούν την υποστήριξη Adobe Flash συνεπώς άτομα με προβλήματα όρασης δεν θα μπορούσαν να γνωρίζουν ότι χρειάζεται να προβούν σε κάποια εγκατάσταση ενός προγράμματος σε περίπτωση που δεν ήταν ήδη εγκατεστημένο στον υπολογιστή τους [37, 38].

Απαιτητικότητα Δημιουργίας

Τα ηχητικά κλιπ ξεχωρίζουν από το γεγονός ότι είναι αρκετά απαιτητικά ως προς τους υπολογιστικούς πόρους ακόμη και σε σχέση με την επεξεργασία εικόνων. Η χρήση έτοιμων ηχητικών κλιπ δεν είναι τόσο εύκολη στην πράξη διότι απαιτεί πολύ χρόνο και πόρους. Επίσης χρησιμοποιείται για την δημιουργία τους λογισμικό σύστασης φωνής το Computer Generated Speech Software το οποίο παρουσιάζει πρόβλημα ότι η φωνή που παράγεται δεν μοιάζει με την ανθρώπινη ούτε ως προς την προφορά ούτε ως προς τη χροιά κάνοντας πιο εύκολη τη λύση. Κατά συνέπεια γίνεται χρήση έτοιμων ηχητικών κλιπ μαζί με διάφορους θορύβους για να δυσκολέψει τη λύση αν και αυτό βαραίνει το διακομιστή [38].

4.3.3 Τυπικές Επιθέσεις

Για να επιτευχθεί επίθεση στα CAPTCHA ήχου, αρχικά παράγουμε χαρακτηριστικά από τον ήχο CAPTCHA και στην συνέχεια χρησιμοποιούμε διάφορες τεχνικές εκμάθησης μηχανών για την εκτέλεση αυτόματη αναγνώριση ομιλίας (ASR- Automated Speech Recognition) σε τμήματα του CAPTCHA. Υπάρχουν πολλές δημοφιλείς τεχνικές για την εξαγωγή χαρακτηριστικών από την ομιλία. Οι τρεις τεχνικές που χρησιμοποιούμε είναι οι συντελεστές συχνότητας mel συχνότητας (MFCC- Συντελεστές που συλλογικά συνθέτουν μια αναπαράσταση ενός βραχυπρόθεσμου ισχύος σήματος και βασίζονται στον γραμμικό μετασχηματισμό συνημίτονων), η αντιληπτική γραμμική πρόβλεψη (PLP- Τεχνική στρέβλωσης των φασμάτων για να ελαχιστοποιηθούν οι διαφορές μεταξύ των ομιλητών, διατηρώντας παράλληλα τις σημαντικές πληροφορίες ομιλίας) και το σχετικό φασματικό μετασχηματισμό-PLP (RASTA-PLP- Είναι μια ξεχωριστή τεχνική που εφαρμόζει ένα φίλτρο διέλευσης σε κάθε υποζώνη συχνοτήτων προκειμένου να εξομαλύνει τις βραχυπρόθεσμες παραλλαγές θορύβου και να απομακρύνει οποιαδήποτε σταθερή μετατόπιση που προκύπτει από στατικό φασματικό χρωματισμό στο κανάλι ομιλίας π.χ. Από μια τηλεφωνική γραμμή). Το MFCC είναι μία από τις πιο δημοφιλείς αναπαραστάσεις χαρακτηριστικών ομιλίας

που χρησιμοποιούνται. Παρόμοια με ένα γρήγορο μετασχηματισμό Fourier (FFT), το MFCC μετατρέπει ένα αρχείο ήχου σε ζώνες συχνοτήτων, αλλά (σε αντίθεση με το FFT) το MFCC χρησιμοποιεί ζώνες συχνότητας mel, οι οποίες είναι καλύτερες για την προσέγγιση της κλίμακας συχνοτήτων που ακούει ο άνθρωπος. Το PLP σχεδιάστηκε για να αποσπάσει από την ομιλία τα χαρακτηριστικά που δεν εξαρτώνται από τον ομιλητή [11].

Χαρακτηριστικό παράδειγμα επίθεσης [11] captcha ήχου έγινε το 2008 όπου χρησιμοποίησαν συνδυασμό δύο τεχνικών για να επιτευχθεί ο στόχος τους. Συγκεκριμένα χρησιμοποιώντας το PLP και μια παραλλαγή όπως το RAS TA-PLP, μπόρεσαν να εκπαιδεύσουν τους ταξινομητές τους να αναγνωρίζουν γράμματα και ψηφία ανεξάρτητα από το ποιος τους μίλησε. Σε έναν από τους τύπους CAPTCHA ήχου που εξετάστηκαν αρχικά χρησιμοποιήθηκαν πολλοί διαφορετικοί άνθρωποι για την καταγραφή των ψηφίων σε έναν, έτσι χρειάστηκαν οι τεχνικές PLP και RAS TA-PLP για να εξαχθούν τα χαρακτηριστικά που ήταν πιο χρήσιμα για την επίλυσή τους.

Σε ένα άλλο πείραμα που διεξήχθη, χρησιμοποιήθηκαν οι τεχνικές MFCC, του φάσματος σήματος PLP (PLP- SPEC), και του αντιστρόφου μετασχηματισμού Fourier του εκτιμώμενου φάσματος ενός σήματος PLP (PLP-CEPS), του φάσματος σήματος RASTA-PLP (RASTA-PLP-SPEC) και του αντιστρόφου μετασχηματισμού Fourier του εκτιμώμενου φάσματος ενός σήματος RASTA-PLP (RASTA-PLP-CEPS).Η κατασκευή του προγράμματος έγινε στην γλώσσα Matlab με την βοήθεια τριών αλγορίθμων. Ο πρώτος αλγόριθμος που χρησιμοποίησαν ήταν ο AdaBoost (προσαρμοστικός αλγόριθμος μηχανικής μάθησης). Ο δεύτερος αλγόριθμος που χρησιμοποίησαν ήταν ο SVM (Support Victim Machine Algorithm -μη παραμετρική μέθοδος που χρησιμοποιείται για ταξινόμηση και παλινδρόμηση). Ο τρίτος αλγόριθμος που χρησιμοποίησαν ήταν αυτός των K-πλησιέστερων γειτόνων (K-NN – μη παραμετρική μέθοδος που χρησιμοποιείται για την ταξινόμηση και την οπισθοδρόμηση) όπου σαν μέτρηση απόστασης χρησιμοποίησαν την ευκλείδεια απόσταση. Στα πειράματα που διεξήγαγαν προσπάθησαν να επιλύσουν τρία είδη audio CAPTCHA (Google Captca, Digg Captcha, recaptcha) με τις παραπάνω πέντε τεχνικές. Το μεγαλύτερο ποσοστό επιτυχούς ανάλυσης Google audio CAPTCHA , επιτεύχθηκε με τον αλγόριθμο SVM με ποσοστό 67% , στις μεθόδους MFCC,PLP-SPEC,PLP-CEPS και RASTA-PLP-CEPS. Το μεγαλύτερο ποσοστό επιτυχούς ανάλυσης DIGG AYDIO CAPTCHA , επιτεύχθηκε με τον αλγόριθμο SVM με ποσοστό 71% ,στις μεθόδους MFCC, PLP-CEPS και RASTA-PLP-CEPS. Τέλος το μεγαλύτερο ποσοστό επιτυχούς ανάλυσης RECAPTCHA AUDIO CAPTCHA επιτεύχθηκε με τον αλγόριθμο SVM με ποσοστό 45% , στην μέθοδο PLP-CEPS [11, 41].

Διάσπαση Ηχητικού Αποτυπώματος

Ως ηχητικό αποτύπωμα θεωρείται μια ψηφιακή αναπαράσταση ενός ηχητικού δεδομένου, που θα χρησιμοποιηθεί έπειτα για να αναγνωριστεί. Τα πρώτα CAPTCHA ήχου είχαν 26 ηχητικά αποσπάσματα. Το κάθε ένα αντιστοιχούσε σε ένα γράμμα της αλφαβήτου και το τεστ αυτό έστελνε κάποια τυχαία γράμματα και ζητούσε από το χρήστη να αναγνωρίσει τη ταυτότητά τους. Με τη χρήση αυτού του αποτυπώματος η επίθεση είναι συνήθης λόγω του ότι το κάθε αποτύπωμα που αντιστοιχεί σε κάθε γράμμα [17] είναι αρκετό για να συγκριθεί με το τεστ που πρέπει να απαντηθεί. Για να αντικρουστεί αυτή η επίθεση στην αρχή ενσωματώθηκαν κάποιοι θόρυβοι στο μήνυμα ώστε να παραμορφωθεί το αποτύπωμα.

Πειράματα έχουν αποδείξει ότι ένα Audio CAPTCHA μπορεί να σπάσει με την διάσπαση των αρχείων ήχου σε τμήματα θορύβου και λέξεων και την χρήση κατάλληλων λογισμικών [11, 38]

Αυτόματη Ταυτοποίηση Ομιλίας

Υπάρχει ένα πλήθος αλγορίθμων το λεγόμενο *Automated Speech Recognition –ASR* που μπορεί να ταυτοποιήσει μια ομιλία και να μετατρέψει το ηχητικό απόσπασμα σε κείμενο. Για να γίνει πιο δύσκολο να λυθούν τέτοια τεστ ενσωματώθηκαν τυχαίοι θόρυβοι ώστε να δημιουργήσουν σύγχυση στα προγράμματα αυτά. Η έλλειψη σε σχέση με αυτή την επίθεση έχει να κάνει με το γεγονός ότι είναι η πιο απαιτητική όσον αφορά τους υπολογιστικούς πόρους και καταναλώνει μεγάλη υπολογιστική ισχύ με αποτέλεσμα να δίνει τη δυνατότητα στον επιτιθέμενο να έχει αρκετές υποδομές για τις ενέργειές του [11].

Αποστολή Σε Τρίτους

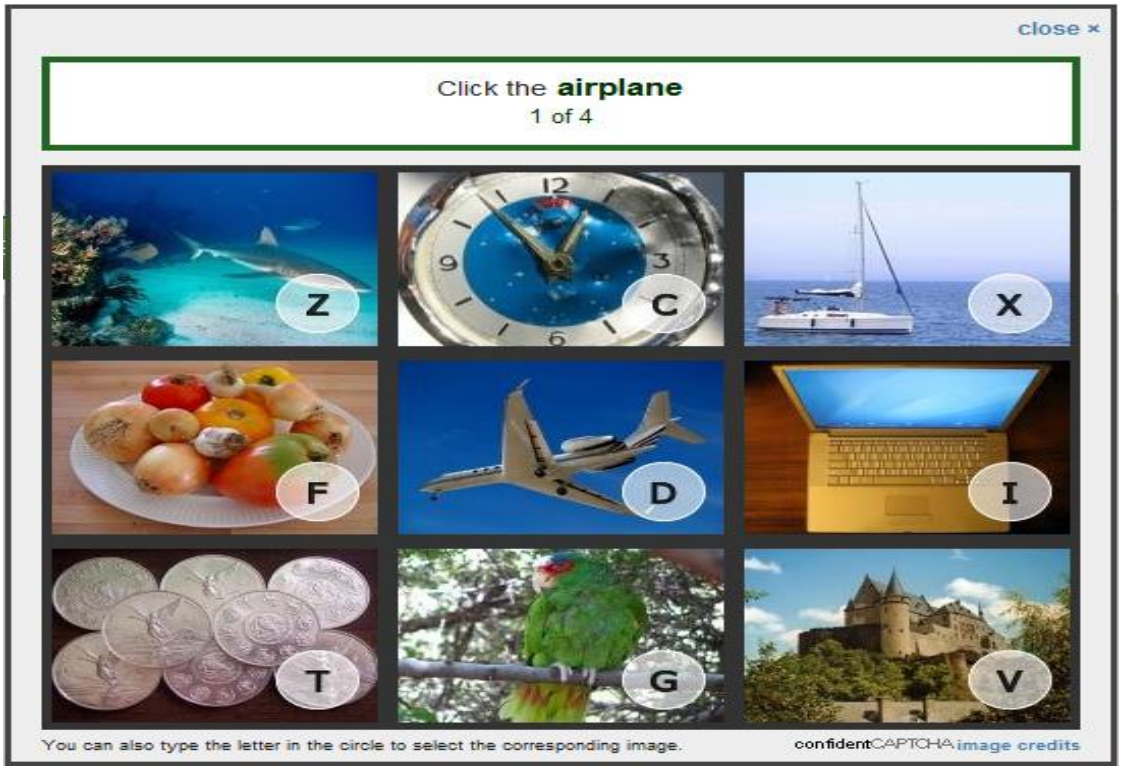
Τα CAPTCHA μπορούν να αποσταλούν σε ανθρώπους με σκοπό την λύση τους και την αποστολή της σωστής απάντησής τους. Οι ηχητικές διαδικασίες έχουν μια διαφορά σε σχέση με τα άλλα τεστ ότι θα πρέπει ο επιτιθέμενος να γνωρίζει πολύ καλά τη γλώσσα του κειμένου μιας και τα περισσότερα τεστ είναι στην αγγλική γλώσσα.

4.4 CAPTCHA Αναγνώρισης Εικόνας (Image Recognition)

Στην παρακατω ενότητα θα αναλυθούν τα CAPTCHA Εικόνας, θα παρουσιαστούν τα μειονεκτήματα από την χρήση τους καθώς και τυπικές επιθέσεις που μπορούν να δεχτούν.

4.4.1 Περιγραφή

Τα CAPTCHA τα οποία είναι βασισμένα σε εικόνα είναι αποτέλεσμα της εύρεση εναλλακτικών προσεγγίσεων στο σχεδιασμό CAPTCHA με στόχο να αντικαταστήσει το CAPTCHA κειμένου. Ζητάνε από τους χρήστες να προβούν στην αναγνώριση κάποιων εικόνων σε σύγκριση με τα προηγούμενα είδη είναι πιο εύχρηστα και πιο φιλικά ως προς του χρήστες. Η ικανότητα που έχει ο άνθρωπος να μπορεί διακρίνει τα αντικείμενα μέσα στις εικόνες είναι μια διαδικασία η οποία γίνεται φυσικά, σε αντίθεση με τους υπολογιστές για τους οποίους είναι κάτι περίπλοκο. Παρόλο αυτά με την εξέλιξη της τεχνολογίας υπάρχουν εντυπωσιακά αποτελέσματα στην <<όραση>> των υπολογιστών.



Εικόνα 4.8: Image Captcha

4.4.2 Μειονεκτήματα

Μεγάλη Χωρητικότητα Δεδομένων

Εγγύηση για την αποτελεσματικότητα του τεστ είναι ο μεγάλος όγκος εικόνων έτσι ώστε να μην μπορεί να δοθεί λύση με χειροκίνητο τρόπο. Αυτό όμως προϋποθέτει μεγάλη βιβλιοθήκη με εικόνες και μεγαλύτερο χώρο στο δίσκο του διακομιστή. Όσο αυξάνεται ο όγκος των εικόνων αυξάνονται και τα δεδομένα που θα πρέπει να κατεβάσει ο υπολογιστής για να ανοίξει μια σελίδα. Αυτό έχει ως αποτέλεσμα να αργεί να την φορτώσει και να αυξάνεται το πλήθος των δεδομένων από το διακομιστή του ιστοτόπου.

Απαγόρευση Χρήσης Ατόμων Με Ειδικές Ανάγκες

Όπως στα CAPTCHA παραμόρφωσης έτσι και εδώ οι άνθρωποι με προβλήματα όρασης θα έχουν δυσκολία να δουν τις εικόνες και άρα να λύσουν το τεστ και να δώσουν κάποια απάντηση.

4.4.3 Τυπικές Επιθέσεις

Επίθεση στο Asirra

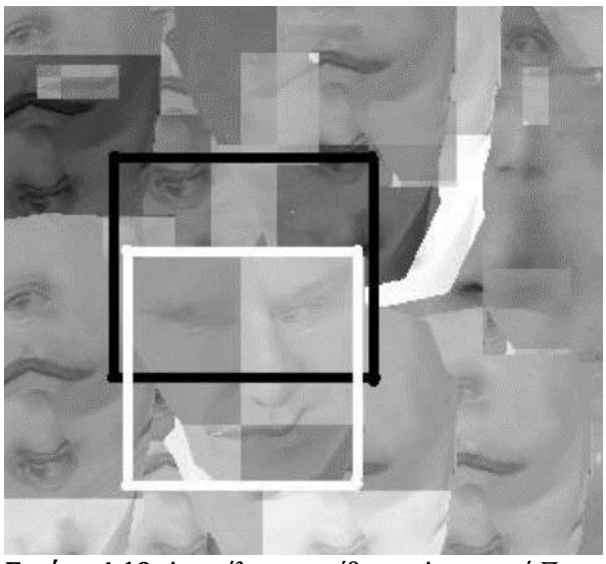
Μια συσκευή εκμάθησης επίθεσης στο Asirra (Animal Species Image Recognition) σχεδιάστηκε από τον Golle προκειμένου να πραγματοποιήσει μια επίθεση σε μια εικόνα. Το Assira είναι ένα CAPTCHA που ζητά από τους χρήστες να κατηγοριοποιήσουν φωτογραφίες που απεικονίζουν είτε γάτες είτε σκύλους. Ένα παράδειγμα εμφανίζεται στη παρακατω εικόνα. Η δύναμη της Asirra προέρχεται από μια καινοτόμο συνεργασία με το Petfinder.com, το μεγαλύτερο ιστοχώρο στον κόσμο που αφιερώνεται στην εύρεση σπιτιών για άστεγα ζώα. Το Petfinder έχει μια βάση δεδομένων με πάνω από τρία εκατομμύρια εικόνες από γάτες και σκύλους, καθεμιά από τις οποίες κατηγοριοποιείται με πολύ υψηλή ακρίβεια από εθελοντές ανθρώπους. Σε αυτή την επίθεση μια εικόνα χωρίζεται και διαιρείται σε ομοιόμορφα μπλοκ. Τα διακριτά χαρακτηριστικά που χρησιμοποιήθηκαν στην επίθεση είναι μοτίβα χρώματος του μπλοκ και πλακάκια με 5x5 υφή. Η μηχανική μάθηση σε επισημασμένα δεδομένα εκπαίδευσης παράγει έναν ταξινομητή που αγγίζει ένα ποσοστό επιτυχίας 82,7% στη διάκριση μια γάτας από ένα σκύλο που χρησιμοποιείται από το Asirra, πολύ υψηλότερο από ό, τι μια τυχαία εικασία επιτυγχάνει [09,40].Η παραπάνω μορφή επίθεσης μπορεί να χρησιμοποιηθεί σε πολλές υπάρχουσες μορφές CAPTCHA.

Εικόνα 4.9: Εικονα Assira CAPTCHA

Επίθεση στο ARTiFACIAL

Το ARTiFACIAL λειτουργεί ως εξής: Συνθέτει αυτόματα μια εικόνα με παραμορφωμένο πρόσωπο ενσωματωμένο μέσα ένα γεμάτο φόντο. Ο χρήστης καλείται να βρει πρώτα το πρόσωπο και στη συνέχεια να κάνει κλικ σε έξι σημεία συγκεκριμένα 4 κλικ στις γωνίες των ματιών και δύο κλικ γωνίες στο στόμα. Εάν ο χρήστης μπορεί να αναγνωρίσει σωστά αυτά τα σημεία, μπορούμε να συμπεράνουμε ότι είναι άνθρωπος διαφορετικά είναι μηχάνημα.

Για να επιτύχει μια επίθεση, ο επιτιθέμενος πρέπει πρώτα να εντοπίσει το πρόσωπο από το χρησιμοποιώντας έναν ανιχνευτή προσώπου και στη συνέχεια να βρει τα χαρακτηριστικά του προσώπου (π.χ. μάτια, μύτη και Στόμα) χρησιμοποιώντας έναν ανιχνευτή χαρακτηριστικών. Σε αυτή την επίθεση χρησιμοποιήθηκαν τρεις υπερσύγχρονοι ανιχνευτές προσώπου και ένας ανιχνευτή χαρακτηριστικών. Το αποτελέσματα μετά από έναν πολύ μεγάλο αριθμό επιθέσεων εμφανίζονται στην παρακάτω εικόνα. Το πρόσωπο το οποίο ανιχνεύεται οροθετείται με το μαύρο πλαίσιο ενώ η περιοχή ανίχνευσης χαρακτηριστικών στο λευκό. Παράλου τα ποσοστά επιτυχίας της ήταν πολύ χαμηλά αποτελεί μια μορφή επίθεσης εναντίον των Image CAPTCHA [42].



Εικόνα 4.10: Αποτέλεσμα επίθεσης Ανιχνευτή Προσώπου.

4.5 Διαδραστικά CAPTCHA (Interactive)

Στην παρακάτω ενότητα θα αναλυθούν τα Διαδραστικά CAPTCHA, θα παρουσιαστούν τα μειονεκτήματα από την χρήση τους καθώς και τυπικές επιθέσεις που μπορούν να δεχτούν.

4.5.1 Περιγραφή

Τα διαδραστικά CAPTCHA αποτελούν μια πιο ευχάριστη διαδικασία για τους χρήστες, οι οποίοι πρέπει να χρησιμοποιήσουν το ποντίκι του υπολογιστή για να δώσουν τη λύση. Πάντα δίνεται μια οδηγία μέσα από μια εύκολη ερώτηση που οδηγεί τον χρήστη για παράδειγμα να μεταφέρει κάποιες εικόνες σε συγκεκριμένα μέρη της φόρμας ή να επιλέξει εικόνες με το ίδιο θέμα. Αν και παίρνει περισσότερο χρόνο θεωρείται καινοτόμα ως προς τη μορφή της και ως προς τις προηγούμενες μορφές CAPTCHA.

Εικόνα 4.11: Παραδείγματα Διαδραστικών CAPTCHA.

4.5.2 Μειονεκτήματα

Φιλικό Στους Χρήστες.

Τα διαδραστικά CAPTCHA δεν είναι τόσο φιλικά ως προς όλους τους χρήστες. Τα άτομα με θέματα όρασης δεν μπορούν να λύσουν αυτά τα τεστ με μεγάλη ευκολία καθώς και άτομα τα οποία δεν έχουν μεγάλη εξοικείωση με την τεχνολογία.

Βαθμός Δυσκολίας Χρήσης

Τα τεστ αυτά θυμίζουν γρίφους που πρέπει να βρεθεί απάντηση. Η καινοτομία αυτών των τεστ αντιτίθεται σε αυτό το τομέα γιατί υπάρχει το ενδεχόμενο ο χρήστης να μην αντιληφθεί το τεστ ή τις ενέργειες που πρέπει να κάνει.

Απουσία Τεχνικού Προτύπου

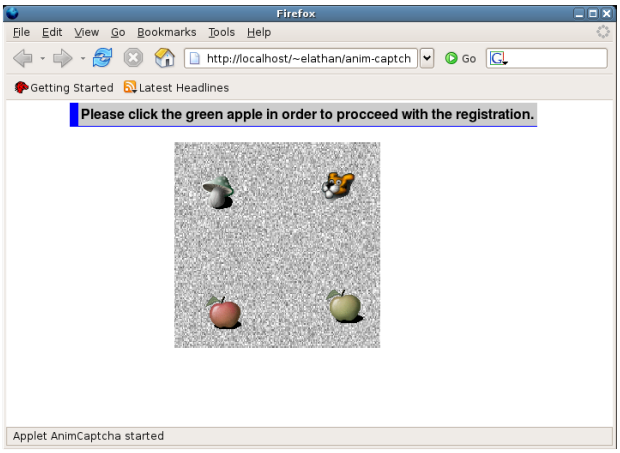
Η γλώσσα HTML δεν υλοποιεί τις τεχνικές αυτές για αυτά τα τεστ. Αυτό έχει ως αποτέλεσμα να χρειαστεί να δημιουργηθούν νέες μέθοδοι μέσω της γλώσσας JavaScript είτε με εφαρμογή για το Flash. Οι λύσεις που έχουν βρεθεί διακρίνονται για κάποιες ελλείψεις. Μολονότι οι browsers είναι σε JavaScript γλώσσα δεν έχουν τα ίδια αποτελέσματα. Το στάδιο της υποστήριξης έχει να κάνει με την έκδοση του προγράμματος. Αν και τις τελευταίες δεκαετίες διακρίνεται μια προσπάθεια για να βρεθεί ένα κοινό πρότυπο βάσει του οποίου θα λειτουργεί η γλώσσα. Ακόμη η γλώσσα JavaScript είναι ένα διάνυσμα επίθεσης για την εγκατάσταση κακόβουλου λογισμικού. Το λογισμικό Flash είναι πιο ασφαλές μιας και κάθε εφαρμογή θα λειτουργήσει χωρίς κάποιες αλλαγές και χωρίς να επηρεάζεται από το πρόγραμμα περιήγησης που χρησιμοποιεί ο χρήστης [38].

4.5.3 Τυπικές Επιθέσεις

Brute Force Attack

Τα Animated CAPTCHA (κινούμενα) είναι ιδανικά για να αποφύγουμε Laundry Επιθέσεις, ένας κακόβουλος χρήστης μπορεί να προσπαθήσει να επιτεθεί σε ένα Animated CAPTCHA μέσω της επίθεσης Brute Force Attack. Στο Animated CAPTCHA της παρακατω εικόνας ο επιταμένος μπορεί να συνεχίσει να δώσει εντολή σε ένα bot να κάνει συνέχεια κλικ στο παζλ. Υπάρχει πιθανότητα

μέσα από συνεχόμενα κλικ να πέτυχει την σωστή απάντηση. Αλώστε ο αριθμός των απαντήσεων δεν μπορεί να είναι υψηλός εξαιτίας του περιορισμένου διαστήματος του παζλ και οι χρήστες σε καμία περίπτωση δεν πρέπει να μπερδεύονται και να εγκαταλείπουν την δοκιμή [04].



Εικόνα 4.12: Παραδείγματα Animated CAPTCHA

Remote Control Attack

Στην χειροκίνητη επίθεση(Remote Control Attack) ο επιτιθέμενος μπορεί σε έναν εργασιακό χώρο να κάνει χρήση ενός υπολογιστή ως διαδικτυακό ρομπότ για παράνομες δραστηριότητες (botnet: Δίκτυο υπολογιστών ελεγχόμενο κεντρικά και που εκτελεί αυτοματοποιημένες (παράνομες) εργασίες μέσω του Διαδικτύου) και να προχωρήσει σε αυτόματη επίλυση των εικόνων απεικόνισης ασφαλείας CAPTCHA για παράνομη πρόσβαση σε δικτυακούς τόπους [04].

Μη Αποσύνδεση Session ID

Μια επίθεση η οποία θα μπορούσε να γίνει που ισχύει για όλα τα είδη των CAPTCHA είναι η εκμετάλλευση αδυναμιών που κάθε επιτιθέμενος θα μπορούσε να αξιοποιήσει. Για παράδειγμα: Η μη-αποσύνδεση της αντιστοιχίας ενός session-ID με μια συγκεκριμένη δοκιμασία CAPTCHA μετά την επιτυχημένη λύση της. Σε αυτή την περίπτωση ο επιτιθέμενος λύνει πρώτα χειροκίνητα μια δοκιμασία CAPTCHA σημειώνοντας το session-ID από ένα cookie στον browser. Στην συνέχεια προγραμματίζει το bot του να χρησιμοποιεί το cookie με το συγκεκριμένο session-ID όταν «επισκέπτεται» την σελίδα ώστε να αντιμετωπίζει πάντα την ίδια δοκιμασία CAPTCHA, της οποίας η λύση είναι γνωστή [24].

4.6 ReCaptcha

Στην παρακατω ενότητα θα αναλυθεί το RECaptcha, θα παρουσιαστούν πλεονεκτήματα και μειονεκτήματα από την χρήση του καθώς και τυπικές επιθέσεις που μπορούν να δεχτεί.

4.6.1 Περιγραφή

Το ReCAPTCHA είναι ένα σύστημα text based CAPTCHA με μια ενδιαφέρουσα καινοτομία: βοηθάει στην ψηφιοποίηση βιβλίων που περιέχουν λέξεις που αναγνωρίζονται δύσκολα από τα συμβατικά προγράμματα OCR. Συγκεκριμένα, το ReCAPTCHA είναι μια δωρεάν υπηρεσία που προστατεύει τον ιστότοπό σας από ανεπιθύμητα μηνύματα και κατάχρηση. Το ReCAPTCHA χρησιμοποιεί έναν προηγμένο μηχανισμό ανάλυσης κινδύνου και προσαρμοσμένα CAPTCHA για να αποτρέψει οποιοδήποτε αυτοματοποιημένο λογισμικό να κάνει κατάχρηση των δραστηριοτήτων σε μια ιστοσελίδα [14, 18].

Εικόνα 4.13: Παράδειγμα Re-Captcha

Όπως παρατηρούμε και στην παραπάνω εικόνα η σελίδας που χρησιμοποιεί το σύστημα ReCAPTCHA παρουσιάζει στον επισκέπτη της μια εικόνα που περιέχει δύο λέξεις. Η μια λέξη έχει επιλεχθεί τυχαία και είναι γνωστή στο σύστημα ReCAPTCHA, όπως συμβαίνει και στα υπόλοιπα συστήματα CAPTCHA. Η άλλη λέξη όμως δεν είναι γνωστή στο σύστημα καθώς πρόκειται για εικόνα που προέκυψε από ψηφιακή σάρωση κάποιου βιβλίου.

Ο διαχωρισμός ανθρώπου και μηχανής σε αυτή την περίπτωση γίνεται από την εισαγωγή της άγνωστης λέξης που ζητείται από τον χρήστη να εισάγει. Αν ο επισκέπτης λύσει με επιτυχία την δοκιμασία αυτή, τότε το σύστημα του δίνει πρόσβαση στην ανάλογη υπηρεσία και θεωρεί ότι και η δεύτερη λέξη αναγνωρίστηκε σωστά, πετυχαίνοντας έτσι την αναγνώριση μιας άγνωστης λέξης.

Η εξασφάλιση της ασφάλειας έγκειται στο γεγονός ότι οι άγνωστες λέξεις που επαναλαμβάνονται από τους χρήστες καταχωρούνται στο σύστημα ως γνωστές λέξεις και χρησιμοποιούνται πλέον ως δοκιμασίες μαζί με κάποια άλλη άγνωστη λέξη. Το ReCAPTCHA προσφέρει κάτι περισσότερο από απλή προστασία από ανεπιθύμητα μηνύματα. Κάθε φορά που επιλύονται τα CAPTCHA μας, αυτή η ανθρώπινη προσπάθεια βοηθάει στην ψηφιοποίηση του κειμένου. Αυτό με τη σειρά του βοηθά στη διατήρηση των βιβλίων, και στην επίλυση δύσκολων AI (Artificial Intelligence) προβλημάτων [14].

4.6.2 Μειονεκτήματα

Φιλικό Στους Χρήστες.

Η δυνατότητα για αναπαραγωγής ήχου είναι δύσκολη με αποτέλεσμα να αποτελεί εμπόδιο για τους χρήστες με προβλήματα όρασης, δυσλεξία και άλλες γνωστικές διαταραχές [40].

Περιορισμός Χρηστών

Χρειάζεται αρκετός χρόνος για να ολοκληρωθεί η διαδικασία λύσης ενός Re-CAPTCHA και χαμηλή Ευρωστία (Usability). Επίσης όλοι οι χρήστες δεν είναι σε θέση να αναγνωρίζουν τις λέξεις της αγγλικής γλώσσας που χρησιμοποιούνται τόσο στην οπτική όσο και στην ηχητική εκδοχή [40].

4.6.3 Τυπικές Επιθέσεις

Από την έναρξή τους, τα CAPTCHA έχουν χρησιμοποιηθεί ευρέως για να αποτρέψουν τους κακόβουλους χρήστες να εκτελέσουν οποιαδήποτε παράνομη ενέργεια. Παρ' όλα αυτά τα οικονομικά κίνητρα έχουν οδηγήσει σε μια κούρσα εξοπλισμών, όπου οι κακόβουλοι αναπτύσσουν αυτοματοποιημένες λύσεις και, με τη σειρά τους υπηρεσίες captcha. Πρόσφατη μελέτη παρουσίασε ότι και το ReCAPTCA είναι ευάλωτο σε επιθέσεις [38]. Συγκεκριμένα στόχος ήταν να ελαχιστοποιηθεί από τους χρήστες η προσπάθεια επίλυσης και να αυξηθεί η δυσκολία επίλυσης από τους υπολογιστές των CAPTCHA που χρησιμοποιούνται από το Google και το Facebook. (τα οποία δεν θα είναι απλά αναγνώριση κειμένου). Ζητήθηκε από τους χρήστες είτε να κλικάρουν μέσα σε ένα box ή να επιλέξουν ανάμεσα σε εικόνες.

(a) Before user clicks checkbox. (b) User considered human.

Εικόνα 4.14: Εικονίδιο Re-CAPTCHA

Εικόνα 4.15: Ομοιότητες Εικόνων ReCAPTCHA

Οι ερευνητές χρησιμοποίησαν έναν μεγάλο αριθμό παραγόντων, που τους συνέδεσαν για να πραγματοποιήσουν την επίθεσή τους, αξιοποιώντας προγράμματα μερικά από τα οποία θα αναφέρουμε ενδεικτικά:

- GRIS(Google Reverse Image Search)
- Clarifie
- NeuralTalk
- TDL

Το αποτέλεσμα ήταν να καταφέρουν να παρακάμψουν τα μέτρα ασφαλείας του CAPTCHA (cookies, tokens), με μηχανική μάθηση για να "μαντεύει" τη σωστή απάντηση της εικόνας CAPTCHA και με πολύ υψηλότερο βαθμό ακρίβειας από προηγούμενες μελέτες.

Με την επίλυση ενός μέσου CAPTCHA να είναι 19,2 δευτερόλεπτα. Η νέα μορφή επίθεσης έφερε αποτελέσματα καλύτερα από ότι αναμενόταν. Στο σύστημα ReCAPTCHA της Google, οι ερευνητές σημείωσαν ποσοστό επιτυχίας 70,78% σε πάνω από 2.235 CAPTCHA. Το ποσοστό επιτυχίας στο Facebook είναι 83,5% σε πάνω από 200 CAPTCHA [39].

Οι λόγοι που οδήγησαν στην επιτυχία της επίθεσης σε χαμηλότερο ποσοστό στην Google είναι επειδή κάνει χρήση φωτογραφιών χαμηλής ποιότητας και σχετίζονται μεταξύ τους. Σε αντίθεση με το Facebook που χρησιμοποιεί εικόνες με υψηλότερη ανάλυση και επιπλέον απεικονίζει αντικείμενα από διαφορετικές κατηγορίες.

Είδος CAPTCHA	**Πλεονεκτήματα**	**Μειονεκτήματα**
Question-Based	Εύκολη υλοποίηση.	Είναι πολύ εύκολο να σπάσουν.
Text-Based	Δεν λύνονται εύκολα σε αυτοματοποιημένες επιθέσεις.	Δυσκολία επίλυσης από άτομα με προβλήματα όρασης. Χρονοβόρα στην επίλυσή τους.
Audio	Εύκολη επίλυση από άτομα με ειδικές ανάγκες πχ όρασης.	Μη συμβατότητα. Δεν υποστηρίζουν πολλές γλώσσες . Απαίτηση μεγάλων Υπολογιστικών Πόρων.
Image	Αρκετά δύσκολο να επιλυθούν αυτόματα.	Δυσκολία επίλυσης από άτομα με προβλήματα όρασης.

Πίνακας 4.1: Σύνοψη Ειδών CAPTCHA

4.7 Άλλες Τεχνικές CAPTCHA

Εκτός από τα είδη των CAPTCHA που αναλύσαμε υπάρχουν και κάποια ακόμα στα οποία θα γίνει μια σύντομη αναφορά παρουσιάζοντας τα πλεονεκτήματα και τα μειονεκτήματά τους από την χρήση τους στις διάφορες ιστοσελίδες [31].

4.7.1 CAPTCHA Bot

Παρόμοιο με το ReCAPTCHA παρέχει τόσο οπτικό όσο και οπτικό CAPTCHA.

Πλεονεκτήματα

Προσφέρει στους χρήστες με προβλήματα όρασης μια χρήσιμη εναλλακτική λύση ήχου και εμφανίζει ένα σύνολο γραμμάτων σε αντίθεση με τις αγγλικές λέξεις.

Μειονεκτήματα

Μπορεί να είναι δύσκολο να ερμηνευτεί από ορισμένους χρήστες.

4.7.2 Image Recognition CAPTCHA

Μας ζητάει να προσδιορίσουμε ένα αντικείμενο σε μια εικόνα.

Πλεονεκτήματα

Δεν υπάρχουν προβλήματα ευκρίνειας και δεν χρειάζεται οι χρήστες να γνωρίζουν συγκεκριμένη γλώσσα.

Μειονεκτήματα

Δεν είναι προσβάσιμο σε χρήστες με προβλήματα όρασης. Η προσθήκη ενός εναλλακτικού κειμένου θα κάνει το CAPTCHA εύθραυστο.

4.7.3 Friends Recognition

Στηρίζεται στην κοινωνική αναγνώριση. Στους χρήστες παρουσιάζονται φωτογραφίες των φίλων τους και θα τους ζητάει να ονομάσουν το άτομο στη φωτογραφία - Social Authentication.

Πλεονεκτήματα

Το συγκεκριμένο είδος ουσιαστικά φιλτράρει τους κακόβουλους χρήστες και όχι τα αυτόματα συστήματα τα οποία μπορεί να προκαλέσουν μια επίθεση

Μειονεκτήματα

Δεν είναι εύκολη η επίλυση τους σε περίπτωση που κάποιος χρήστης δεν θυμάται τα ονόματα των απομακρυσμένων φίλων του ή τους φίλους των φίλων του.

4.7.4 User Interaction

Πρόκειται για την πιο πρόσφατη μορφή CAPTCHA που μπορεί να συναντήσει ένας χρήστης όταν επισκέπτεται μια ιστοσελίδα. Οι χρήστες καλούνται ενώ εκτελούν μια εργασία συγχρόνως να πρέπει να αλληλεπιδρούν με μια εφαρμογή για παράδειγμα να μετακινήσουν τον κέρσορα του σε συγκεκριμένο σημείο μέσα στην εφαρμογή προκειμένου να ολοκληρωθεί η εργασία.

Πλεονεκτήματα

Η επίλυσή τους στηρίζεται στον άνθρωπο και είναι αδύνατο να επιλυθούν από Virtual Intelligence bots.

Μειονεκτήματα

Απρόσιτο για άτομα με ειδικές ανάγκες .

Τεχνικές CAPTCHA	**Πλεονεκτήματα**	**Μειονεκτήματα**
Re-Captcha	Προσφέρει στους χρήστες με προβλήματα όρασης με μια εναλλακτική λύση ήχου.	Χρήστες που δεν γνωρίζουν την Αγγλική γλώσσα δυσκολεύονται στην επίλυση.
CAPTCHA Bot	Φιλικό ως προς τους χρήστες με προβλήματα όρασης.	Μπορεί να είναι δύσκολο να ερμηνευτεί από ορισμένους χρήστες.
Image Recognition CAPTCHA	Δεν υπάρχουν προβλήματα ευκρίνειας . Οι χρήστες δεν χρειάζεται να γνωρίζουν κάποια συγκεκριμένη γλώσσα.	Δεν είναι προσβάσιμο σε χρήστες με προβλήματα όρασης.
Friends Recognition	Φιλτράρει τους κακόβουλους χρήστες και όχι τα αυτόματα συστήματα.	Δεν είναι εύκολη η επίλυση τους σε περίπτωση που κάποιος χρήστης δεν θυμάται τα ονόματα των φίλων του.
User Interaction	Η επίλυση τους στηρίζεται στον άνθρωπο και είναι αδύνατο να επιλυθούν από bots.	Απρόσιτο για άτομα με ειδικές ανάγκες .

Πίνακας 4.2: Σύνοψη Τεχνικών CAPTCHA

4.8 Γλώσσες Προγραμματισμού Κατασκευής CAPTCHA.

Για να χρησιμοποιήσουμε την τεχνολογία captcha , οι ιστοσελίδες μας πρέπει να δημιουργούνται δυναμικά σε οποιαδήποτε γλώσσα προγραμματισμού. Οι δημοφιλέστερες γλώσσες προγραμματισμού για την κατασκευή της τεχνολογίας captcha είναι οι παρακάτω:

- PHP: Είναι μια γλώσσα προγραμματισμού για τη δημιουργία σελίδων web με δυναμικό περιεχόμενο. Μια σελίδα PHP περνά από επεξεργασία από ένα συμβατό διακομιστή του Παγκόσμιου Ιστού (π.χ. Apache), ώστε να παραχθεί σε πραγματικό χρόνο το τελικό περιεχόμενο, που είτε θα σταλεί στο πρόγραμμα περιήγησης των επισκεπτών σε μορφή κώδικα HTML ή θα επεξεργασθεί τις εισόδους δίχως να προβάλλει την έξοδο στο χρήστη, αλλά θα τις μεταβιβάσει σε κάποιο άλλο PHP script.

- ASP : Είναι προγραμματιστικό περιβάλλον της εταιρείας Microsoft που δημιουργήθηκε για διαδικτυακό προγραμματισμό για την δημιουργία δυναμικών ιστοσελίδων στο διαδίκτυο.

- Perl : Είναι μία πολύ δημοφιλής αντικειμενοστρεφής γλώσσα προγραμματισμού. Συνήθως ένα πρόγραμμα σε Perl εκτελείται χρησιμοποιώντας άμεσα ή έμμεσα το διερμηνέα της γλώσσας. Αυτό που διακρίνει την Perl από πολλές άλλες γλώσσες προγραμματισμού είναι το γεγονός ότι είναι διαθέσιμη για σχεδόν όλα τα λειτουργικά συστήματα.

- Python: Είναι μια υψηλού επιπέδου γλώσσα προγραμματισμού. Ο κύριος στόχος της είναι η αναγνωσιμότητα του κώδικά της και η ευκολία χρήσης της και το συντακτικό της επιτρέπει στους προγραμματιστές να εκφράσουν έννοιες σε λιγότερες γραμμές κώδικα απ' ότι θα ήταν δυνατόν σε γλώσσες όπως η C++ ή η Java. Διακρίνεται λόγω του ότι έχει πολλές βιβλιοθήκες που διευκολύνουν ιδιαίτερα αρκετές συνηθισμένες εργασίες και για την ταχύτητα εκμάθησης της.

- JSP: Είναι μια τεχνολογία που βοηθά τους προγραμματιστές λογισμικού δημιουργούν δυναμικά παραγόμενες ιστοσελίδες που βασίζονται σε HTML , XML ή άλλους τύπους εγγράφων.

- Ruby: Είναι μια δυναμική, ανακλαστική, αντικειμενοστρεφής γλώσσα προγραμματισμού γενικής χρήσης που συνδυάζει μια σύνταξη επηρεασμένη από την Perl με χαρακτηριστικά από τη Smalltalk (αντικειμενοστρεφής γλώσσα προγραμματισμού).

Όπως όλα τα λογισμικά έχουν ευπάθειες οι οποίες μπορούν να χρησιμοποιηθούν από τους κακόβουλους χρήστες, έτσι και οι ιστοσελίδες που είναι κατασκευασμένες από τις παραπάνω

γλώσσες προγραμματισμού, έχουν ευπάθειες στην τεχνολογία CAPTCHA. Χαρακτηριστικό παράδειγμα είναι το script securimage. Το Securimage είναι ένα δωρεάν ανοικτού κώδικα PHP script για τη δημιουργία σύνθετων εικόνων και κωδικών CAPTCHA, που σαν στόχο έχουν την προστασία των μορφών από το spam και την κατάχρηση. Μπορεί εύκολα να προστεθεί στις υπάρχουσες μορφές στον ιστότοπό για να παρέχει προστασία από spam bots. Το συγκεκριμένο php script για τις εκδόσεις securimage 1.0.4 και πάνω έχει ευπάθεια και αν την εκμεταλλευτεί ο κακόβουλος χρήστης μπορεί να παρακάμψει τον έλεγχο ταυτότητας και να πάρει απομακρυσμένο έλεγχο [32]. Για το συγκεκριμένο script και για την έκδοση 3.5 έχει ευπάθεια και επιτρέπει στον κακόβουλο χρήστη να κάνει επίθεση Cross Site Scripting. Συγκεκριμένα ο κακόβουλος χρήστης εκεί που πληκτρολογεί για την επαλήθευση του CAPTCHA μπορεί να γράψει κώδικα σε html ή javascript, όπου επειδή δεν φιλτράρονται τα δεδομένα εισόδου θα μπορούν να προκαλέσουν ζημιά στον ιστότοπο [33] . Κάποιες από τις ενέργειες που θα μπορούσε να επιτύχει με την επίθεση Cross Site Scripting είναι οι εξής:

- Κλοπή κωδικών/ Λογαριασμών, προσωπικών δεδομένων
- Αλλαγή ρυθμίσεων του ιστοχώρου
- Κλοπή των Cookies
- Ψεύτικη διαφήμιση [34]

Κεφάλαιο 5
Εργαλεία Επίλυσης CAPTCHA

Σε αυτή την ενότητα αρχικά θα αναφερθούν ελεύθερα λογισμικά και υπηρεσίες οι οποίες μπορούν να εκμεταλλευτούν κενά ασφαλείας των CAPTCHA και να οδηγήσουν στην επίλυσή τους. Έπειτα θα γίνει δοκιμή ορισμένων εργαλείων που <<σπάνε>> τα CAPTCHA.

5.1 Λογισμικά Επίλυσης CAPTCHA

Όλες οι μέθοδοι επίλυσης CAPTCHA και συνεπώς οι υπηρεσίες και τα λογισμικά που τα εφαρμόζουν μπορούν να χωριστούν σε δυο κατηγορίες. Στις λύσεις οπτικής αναγνώρισης χαρακτήρων (OCR) και στις ανθρώπινες υπηρεσίες επίλυσης CAPTCHA.

5.1.1 Λύσεις Οπτικής Αναγνώρισης Χαρακτήρων (OCR)

Παρακάτω παρατίθενται κάποια βασικά λογισμικά που δίνουν λύσεις στην οπτική αναγνώριση χαρακτήρων.

GSA Captcha Breaker [21]

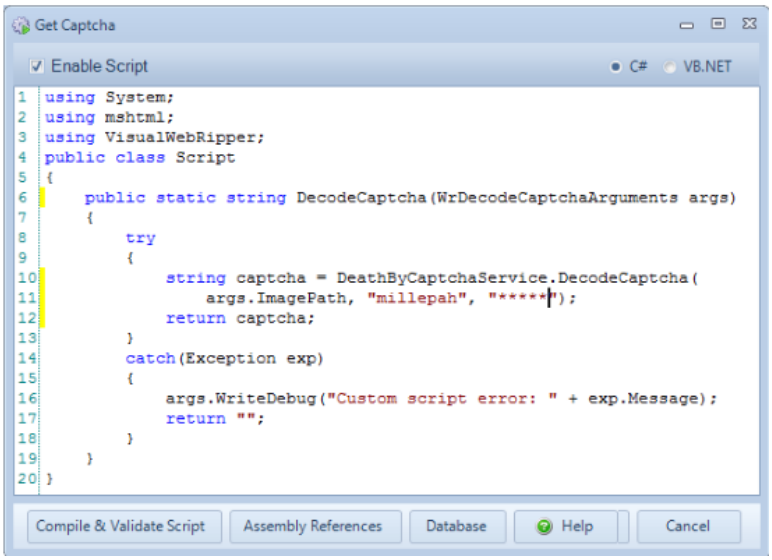
Χρησιμοποιεί αλγόριθμους οπτικής αναγνώρισης χαρακτήρων για αποκωδικοποίηση CAPTCHA. Είναι ένα αυτόνομο πρόγραμμα που λειτουργεί ανεξάρτητα από τυχόν online υπηρεσίες αναγνώρισης captcha (όπως DeathByCaptcha, BypassCaptcha και κ.λπ.). Το πρόγραμμα είναι επί πληρωμή και για να έχουμε τις υπηρεσίες του θα πρέπει να πληρώσουμε συνδρομή. Το συγκεκριμένο πρόγραμμα ακούει αιτήσεις για υπηρεσίες αναγνώρισης captcha, τις παρακολουθεί και προσπαθεί να αναγνωρίσει το CAPTCHA από μόνο του. Εάν επιτύχει, επιστρέφει την απάντηση για λογαριασμό της υπηρεσίας, εάν όχι μπορεί να την μεταβιβάσει στην υπηρεσία για περαιτέρω αναγνώριση.

Παράδειγμα: Βήματα λύσης με GSA Captcha Breaker

Έστω ότι θέλουμε να δώσουμε λύση στο CAPTCHA αυτής της σελίδας.

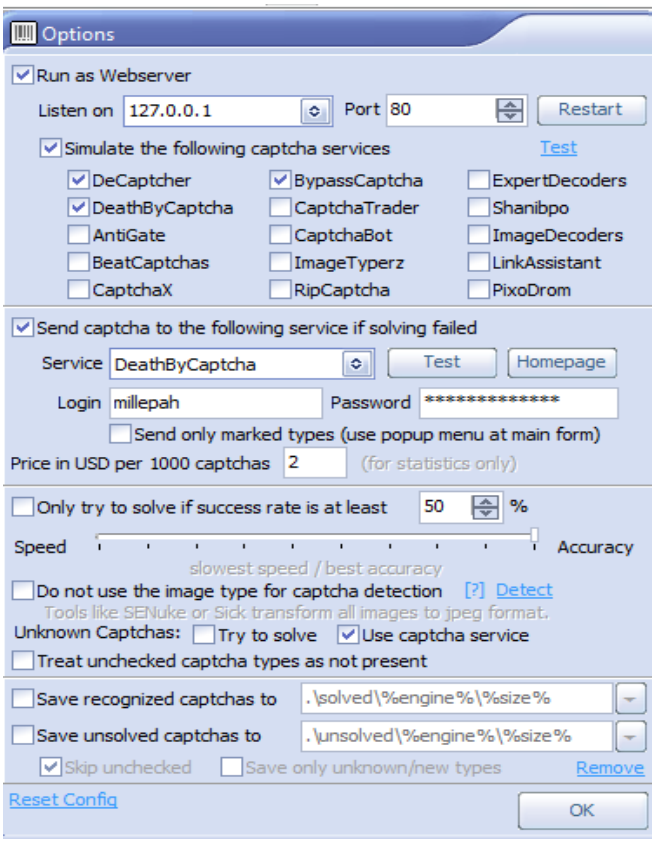
Εικόνα 5.1: Ιστοσελίδα με captcha.

Για να το κάνουμε αυτό θα πρέπει να δημιουργήσουμε ένα CAPTCHA decoding script με το Visual Web Ripper που δημιουργεί αίτημα DeathByCaptcha.



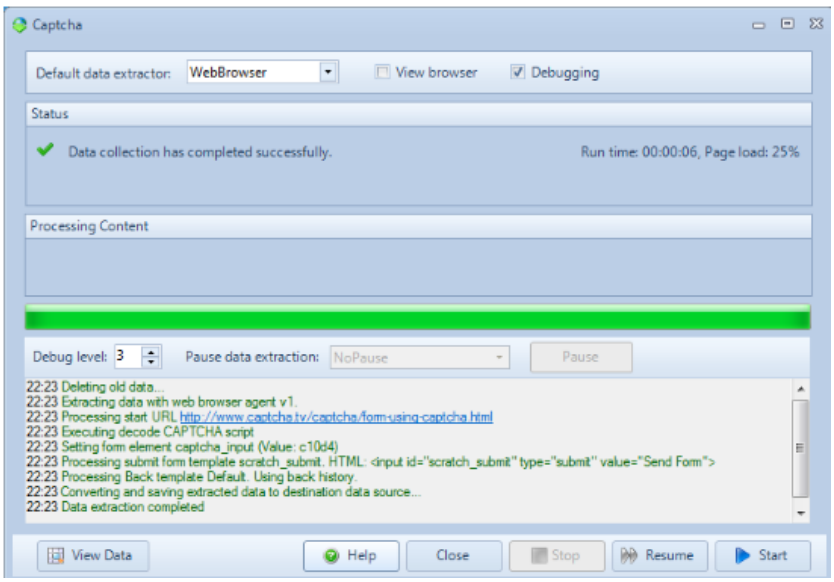
Εικόνα 5.2: CAPTCHA decoding script

Στη συνέχεια ξεκινάμε το Captcha Breaker και ρυθμίζουμε τις επιλογές του ώστε να μπορεί να τρέξει ως τοπικός διακομιστής και να στείλει μη αναγνωρισμένα CAPTCHA στην υπηρεσία DeathByCaptcha για ανθρώπινη αναγνώριση:



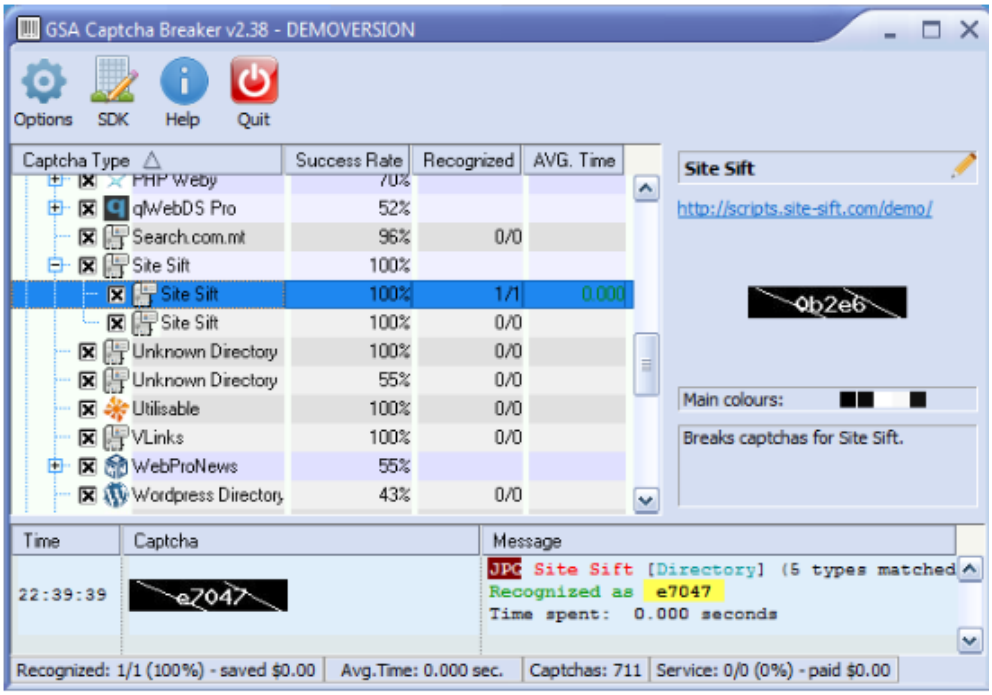
Εικόνα 5.3: Παραμετροποίηση Captcha Breaker

Μετά ξεκινάμε το Visual Web Repper και παρατηρούμε ότι το CAPTCHA έχει αυτόματα αναγνωριστεί.



Εικόνα 5.4: Παράθυρο Visual Web Repper

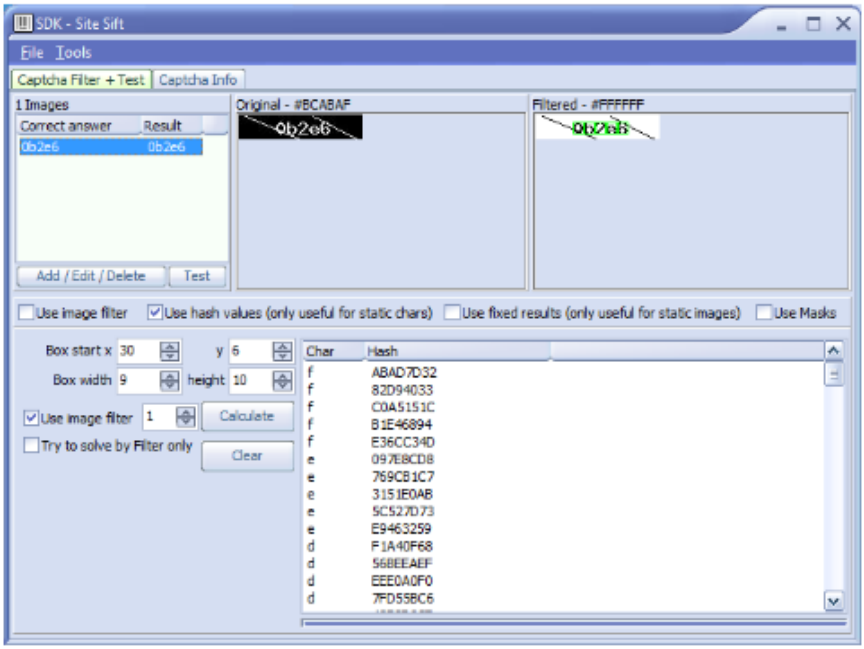
Αφού μεταβούμε στο παράθυρο Captcha Breaker, μπορούμε να δούμε ότι παρεμπόδισε το αίτημα, αναγνώρισε captcha ως τύπου "Site Sift" και το αναγνώρισε με επιτυχία:



Εικόνα 5.5: Παράθυρο Captcha Breaker

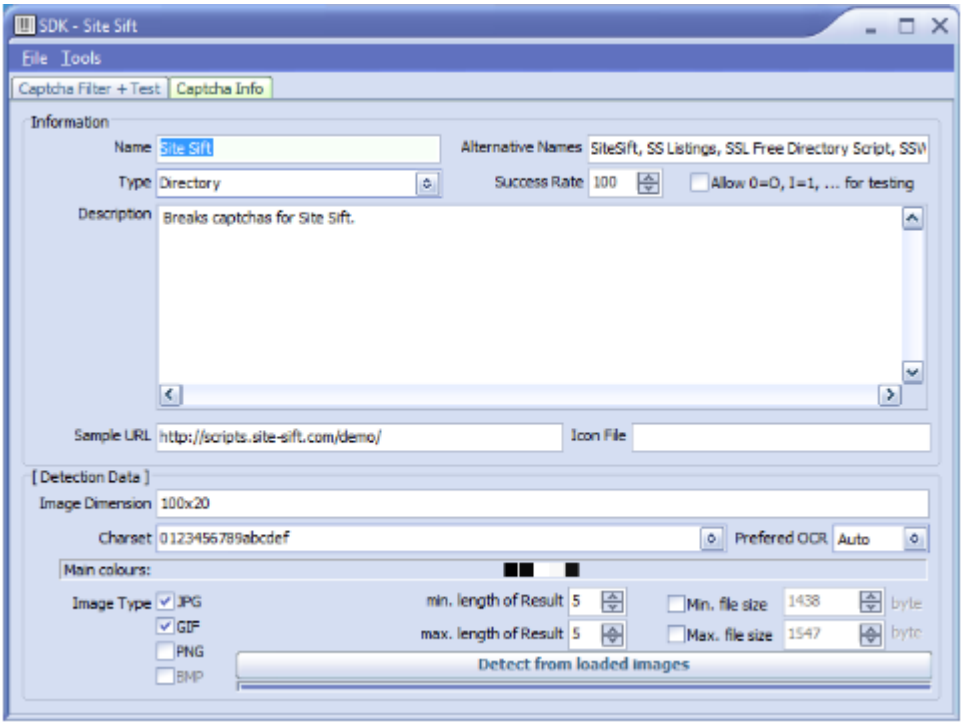
Σε περίπτωση που δεν μπορούμε να βρούμε λύση υπάρχει δυνατότητα παραμετροποίησης των αλγόριθμων επίλυσης των CAPTCHA.

Χρησιμοποιούμε τον εσωτερικό SDK editor



Εικόνα 5.6: Εσωτερικός SDK editor

Στην συνέχεια γίνεται διαμερισμός με τους υπόλοιπους χρήστες που είναι συνδεδεμένοι με την κοινότητα GSA Captcha Breaker



Εικόνα 5.7: Διαμερισμός παραμετροποίησης captcha σε χρήστες

DeCaptcher [19, 20]

Είναι ένα λογισμικό που κάνει καταγραφή της επίλυσης online υπηρεσιών που χρησιμοποιεί τεχνολογίες OCR για αναγνώριση captcha (έναντι ανθρώπινης εργασίας). Παρέχει API για περισσότερες από οκτώ γλώσσες προγραμματισμού και χρησιμοποιείται σαν εμπορικό λογισμικό [23, 24]. Οι παρεχόμενες πλατφόρμες του λογισμικού είναι οι εξής :

- C (all platforms)
- PHP (all platforms)
- Java (all platforms)
- C# (Windows, vanilla)
- C# (Windows, 32 bits)
- C# (Windows, 64 bits)

- Perl (Linux, native)
- Perl (Windows, 32 bits)
- Visual Basic 6 (Windows, 32 bits)
- VB.NET (Windows, 32 bits)
- VB.NET (Windows, 64 bits)
- VB.NET 2010 (Windows, 32 bits)
- Delphi (Windows, 32 bits)
- DLL sources (Windows, 32 bits)
- DLL sources (Windows, 64 bits)
- DLL .NET-friendly sources (Windows, 32 bits)
- DLL .NET-friendly sources (Windows, 64 bits)
- DLL .NET 2010-friendly sources (Windows, 32 bits)
- SO sources (Unix, 32 bits)
- PHP cURL-based (all platforms)
- Command line API (Windows)
- DeCapcher xRumer plugin

posted pictures history, most recent first				
#	**time (GMT)**	**picture**	**text**	**status**
1	2013-05-02 09:51:05		serve heareaw	done
2	2013-05-02 09:50:49		was negmcde	done
3	2013-05-02 09:50:38		was eveloif	done
4	2013-05-02 09:50:24		distin owillM	done
5	2013-05-02 09:49:58		described orchwi	done

Εικόνα 5.8: Δοκιμή λύσης CAPTCHA με πρόγραμμα DeCaptcher

Captcha Sniper

Είναι μια ενδιαφέρουσα εφαρμογή των Windows που παρακολουθεί τις αιτήσεις σε κοινές υπηρεσίες αναγνώρισης captcha (όπως Decaptcha, Death By Captcha, AntiCaptcha και Bypass Captcha) και επιλύει αυτόματα το captcha. Χρησιμοποιείται σαν εμπορικό λογισμικό [19].

Captcha Solver OCR

Είναι ένα απλό δωρεάν λογισμικό που μπορεί διαβάζει εικόνες με κείμενο, και επιστρέφει πίσω, το πραγματικό κείμενο. Λειτουργεί με εύκολο CAPTCHA, αλλά δεν λειτουργεί με σύνθετα CAPTCHA. Χρησιμοποιεί το πακέτο Tesseract, το οποίο είναι ο πιο ακριβές διαθέσιμος μηχανισμός OCR ανοικτής πηγής. Αυτό το λογισμικό μπορεί επίσης να μεταφράσει κάποιο κείμενο από εικόνες [25].

TesserCap

Είναι ένα δωρεάν GUI λογισμικό, που είναι εξαιρετικά ευέλικτο και διαδραστικό εργαλείο ανάλυσης CAPTCHA με τα ακόλουθα χαρακτηριστικά:

- Μια γενική μηχανή προεπεξεργασίας εικόνας που μπορεί να ρυθμιστεί σύμφωνα με τον τύπο CAPTCHA που αναλύεται.

- Tesseract ως μηχανή OCR για την ανάκτηση κειμένου από προεπεξεργασμένα CAPTCHA.

- Υποστηρίζονται διακομιστές μεσολάβησης Web και προσαρμοσμένες κεφαλίδες HTTP.

- Υποστήριξη στατιστικής ανάλυσης CAPTCHA.

- Επιλογής χαρακτήρων για τη μηχανή OCR [27].

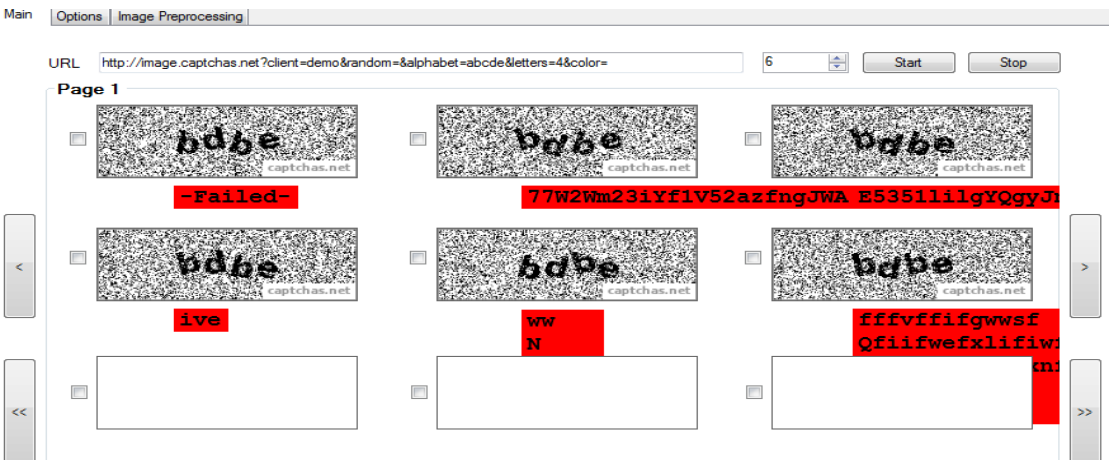
Για την εγκατάσταση του λογισμικού σε περιβάλλον Windows πρέπει πρώτα να γίνει εγκατάσταση του .net framework 4.

Παράδειγμα Λειτουργίας του Προγράμματος TesserCap

1. Στην κεντρική σελίδα του προγράμματος στο πεδίο URL βάζουμε την διεύθυνση του CAPTCHA που θέλουμε να <<σπάσουμε>>

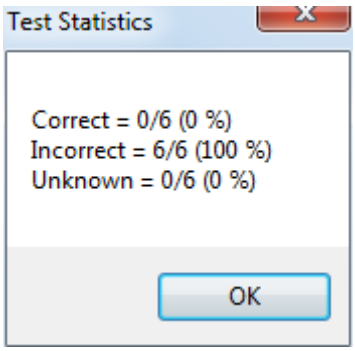
2. Στο δεύτερο πεδίο βάζουμε τον αριθμό των πιθανών λύσεων που θέλουμε να έχουμε για το συγκεκριμένο CAPTCHA.

3. Στην συνέχεια πατάμε το κουμπί Start και αρχίζει να μας εμφανίζει τις πιθανές λύσεις.



Εικόνα 5.9: Πιθανές λύσεις CAPTCHA σε συγκεκριμένο URL

Πατώντας το κουμπί «show statistics» εμφανίζει τα στατιστικά στοιχεία για την επίλυση του συγκεκριμένου captcha.



Εικόνα 5.10: Statistics

Όπως φαίνεται στο παράδειγμα δεν κατέστη δυνατό να επιλυθεί το captcha. Στην καρτέλα «options» μπορούμε να κάνουμε ρυθμίσεις για το CAPTCHA, και στην καρτέλα «image preprocessing» μπορούμε να κάνουμε ειδικές ρυθμίσεις για την εικόνα και τα χρώματα του captcha έτσι ώστε να γίνει πιο εύκολη η επίλυση του.

5.1.2 Ανθρώπινες υπηρεσίες επίλυσης προβλημάτων CAPTCHA

DeathByCaptcha

Είναι ένα υβριδικό λογισμικό του συστήματος OCR και των ανθρώπινων υπηρεσιών επίλυσης προβλημάτων CAPTCHA. Γίνεται σύνδεση με την υπηρεσία μέσω του API που είναι διαθέσιμο για C, PHP, Python, .NET C # & VB, Java, Perl, AutoIt3, iMacros κ.τ.λ. Η καθυστέρηση της αναγνώρισης του captcha είναι περίπου 10 δευτερόλεπτα.

Bypass CAPTCHA

Είναι μια υπηρεσία επίλυσης δεδομένων captcha με βάση το ανθρώπινο δυναμικό. Το εξαιρετικό χαρακτηριστικό αυτής της υπηρεσίας είναι ότι επικεντρώνεται στην εύκολη ενσωμάτωση με οποιοδήποτε λογισμικό τρίτου μέρους για να εμπλουτίσει τη λειτουργία αποκωδικοποίησης captcha με αυτόματη αναγνώριση captcha. Παρέχει API για πολλές γλώσσες όπως PHP, Python, Perl, Ruby, Java, JavaScript, C / C ++, C #, Delphi, VB.NET κ.τ.λ. .

Image Typerz

Μια ακόμα υπηρεσία CAPTCHA BYPASS χρησιμοποιώντας φθηνή ανθρώπινη εργασία. Η διαφορά στην υπηρεσία αυτή σε σχέση με άλλες παρόμοιες είναι ότι αυτή η υπηρεσία δεν επεξεργάζεται

δύσκολα αναγνωρίσιμα CAPTCHA. Χρησιμοποιείται σαν εμπορικό λογισμικό. Η υπηρεσία καθιστά εύκολη την σύνδεση μέσω πολυάριθμων APIS για .Net, PHP, Java, C / C ++, Perl, iMacros κ.τ.λ.

ExpertDecoders

Είναι ένας αυτοματοποιημένο εργαλείο λύσης captcha που βασίζεται στον άνθρωπο. Έχει αρκετά καλή ανταπόκριση και ποιότητα υπηρεσιών. Επιτρέπει στους τελικούς χρήστες, την συνεργασία με το API BypassCaptcha.com ή το De-Capcher API.

9kw.eu

Είναι μία διαδικτυακή υπηρεσία απευθύνεται σε όσους θέλουν να επιλύσουν CAPTCHA. Αρχικά πρέπει να δημιουργήσουμε έναν λογαριασμό και να συνδεθούμε με την υπηρεσία. Στην συγκεκριμένη υπηρεσία κερδίζεις πόντους ή πιστώσεις για την επίλυση των CAPTCHA που έχουν υποβληθεί μέχρι εκείνη την στιγμή. Επίσης η υπηρεσία λειτουργεί με αντίθετο τρόπο. Μπορούμε να υποβάλουμε τα δικά μας CAPTCHA που πρέπει να επιλυθούν για τα κερδίσεις πόντους. Υπάρχει επίσης διαθέσιμη επέκταση της συγκεκριμένης υπηρεσίας στο Google Chrome, για απομακρυσμένη επίλυση CAPTCHA [19].

5.2 Επεκτάσεις για αυτόματη επίλυση και παράκαμψη CAPTCHA σε προγράμματα περιήγησης ιστού.

Η πραγματικότητα της χρήσης ενός CAPTCHA είναι ότι δεν εμποδίζει πραγματικά τα bots επειδή υπάρχουν αρκετές αυτοματοποιημένες υπηρεσίες επίλυσης CAPTCHA όπως το DeathByCaptcha και περιέχουν ένα API που μπορεί να ενσωματωθεί σε οποιοδήποτε λογισμικό. Υπάρχει τρόπος που παρακάμπτει αυτόματα την επίλυση του CAPTCHA, χρησιμοποιώντας κάποιες επεκτάσεις στα προγράμματα περιήγησης ιστού. Οι συγκεκριμένες επεκτάσεις λειτουργούν με πληρωμή αγοράζοντας πιστωτικές μονάδες. Τέτοιες επεκτάσεις είναι το anticaptcha (για firefox), το CAPTCHA Be Gone (για firefox, internet explorer,google chrome) και το Rumola (για firefox, google chrome, safari) [30].

5.3 Αυτόματο εργαλείο pentesting για την παράκαμψη των CAPTCHA

Όπως αναφέρθηκε παραπάνω υπάρχουν αρκετά λογισμικά και επεκτάσεις είτε δωρεάν είτε με πληρωμή που μπορούν να παρακάμψουν τα CAPTCHA από την πλευρά των απλών χρηστών. Υπάρχει όμως ένα αυτοματοποιημένο εργαλείο που το χρησιμοποιούν οι pentesters για να κάνουν δοκιμές και να διαπιστώσουν αν παρακάμπτονται τα captcha σε κάποια ιστοσελίδα. Το εργαλείο αυτό είναι το CINtruder v0.3 (Captcha Intruder), η εγκατάσταση του γίνεται σε συστήματα Linux και θα πρέπει να είναι εγκατεστημένη και η Python.

5.3.1 Εγκατάσταση εργαλείου CINtruder [23]

Οι προαπαιτούμενες βιβλιοθήκες για να εγκατασταθεί και να τρέξει το πρόγραμμα είναι οι εξής :

- python-pycurl - Python bindings to libcurl
- python-libxml2 - Python bindings for the GNOME XML library
- python-imaging - Python Imaging Library

Βήματα εγκατάστασης:

1. Αρχικά ελέγχουμε αν έχουμε εγκαταστημένη την python δίνοντας στο terminal την εντολή python –v ή python3 –v αναλόγως την έκδοση linux που έχουμε.
2. Αν μας βγάλει την έκδοση της python προχωράμε στο επόμενο βήμα, αν δεν βγάλει κάνουμε εγκατάσταση της με την εντολή : sudo apt-get install python3
3. Στην συνέχεια δίνουμε την εντολή : sudo apt-get install python-pycurl python-libxml2 python-imaging , για να γίνει η εγκατάσταση των προαπαιτούμενων βιβλιοθηκών της python.
4. Δίνουμε την εντολή :git clone https://github.com/epsylon/cintruder για να προστεθεί το πρόγραμμα στον υπολογιστή μας.

5. Για να τρέξουμε το πρόγραμμα πρέπει να περιηγηθούμε στον φάκελο cd cintruder με την εντολή: cd cintruder και δίνοντας την εντολή ./cintruder εμφανίζεται η παρακάτω εικόνα.

Εικόνα 5.11: CINtruder v0.3

5.3.2 Παραδείγματα εντολών προγράμματος

Παρακάτω παρατίθενται παραδείγματα εντολών του προγράμματος :

Βοήθεια προγράμματος:

./cintruder --help

Εκσυγχρονισμός στην τελευταία έκδοση :

./cintruder –update

Εκκίνηση διεπαφής ιστού (GUI):

./cintruder --gui

Επίλυση CAPTCHA συγκεκριμένου αρχείου:

./cintruder --crack "inputs/captcha.gif"

Επίλυση CAPTCHA από συγκεκριμένη διεύθυνση URL:

./cintruder --crack "http://host.com/path/captcha_url"

Επίλυση CAPTCHA συγκεκριμένου αρχείου και εξαγωγή του σε αρχείο xml :

./cintruder --crack "inputs/captcha.gif" --xml "test.xml"

Επίλυση Captcha από συγκεκριμένη διεύθυνση URL, με proxy TOR και λεπτομερή αναπαραγωγή της εξόδου:

./cintruder --crack "http://host.com/path/captcha_url" --proxy="http://127.0.0.1:8118" -v

Σείρε τα CAPTCHA από συγκεκριμένη διεύθυνση URL, με proxy TOR και λεπτομερή αναπαραγωγή της εξόδου:

./cintruder --train "http://host.com/path/captcha_url" --proxy "http://127.0.0.1:8118" –v

Σείρε 50 CAPTCHA από συγκεκριμένη διεύθυνση URL, με proxy TOR:

./cintruder --track "http://host.com/path/captcha.gif" --track-num "50" --proxy "http://127.0.0.1:8118"

Κατάλογος διαθέσιμων ενοτήτων (από το "mods /"):

./cintruder --list

Εκκίνηση ενότητας OCR για να σύρετε μια συγκεκριμένη τοπική CAPTCHA:

./cintruder --train "inputs/easycaptcha.gif" --mod "module_invocation_name"

Εκκινήστε μια ενότητα OCR για να επιλύσετε ένα online captcha, λεπτομερή αναπαραγωγή της εξόδου:

./cintruder --crack "http://host.com/path/captcha_url" --mod "module_invocation_name" -v

Αντικαταστήστε την προτεινόμενη λέξη από το CIntruder μετά από σπάσιμο μιας απομακρυσμένης διεύθυνση url στις εντολές ενός άλλου εργαλείου (π.χ. "XSSer"):

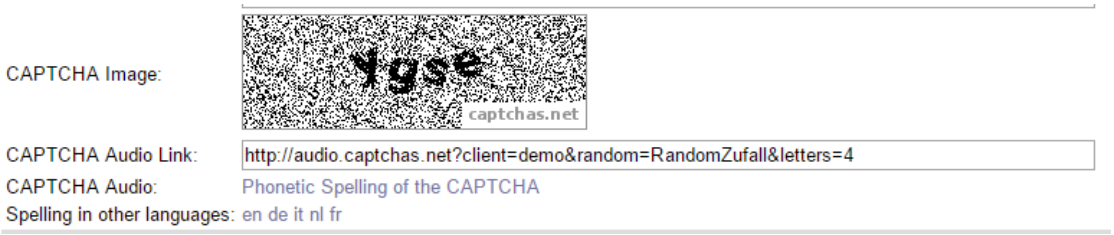
./cintruder --crack "http://host.com/path/captcha_url" --tool "xsser -u http://host.com/path/param1=foo?txtCaptcha=CINT" [7]

5.3.3 Επίλυση captcha από συγκεκριμένη διεύθυνση

Για να προβούμε στην επίλυση του captcha μιας συγκεκριμένης διεύθυνσης μέσω του προγράμματος CINtruder ακολουθούμε τα παρακάτω βήματα:

1. Ανοίγουμε την διεύθυνση που θέλουμε να επιλύσουμε ένα captcha

   Πάμε στην διεύθυνση: http://captchas.net/ όπου εμφανίζεται το παρακάτω CAPTCHA



Εικόνα 5.12: Ιστοσελίδα με CAPTCHA

2. Πατάμε με δεξί κλικ πάνω στην εικόνα του captcha και επιλέγουμε να δούμε την εικόνα και ανοίγει η εικόνα σε νέα καρτέλα.

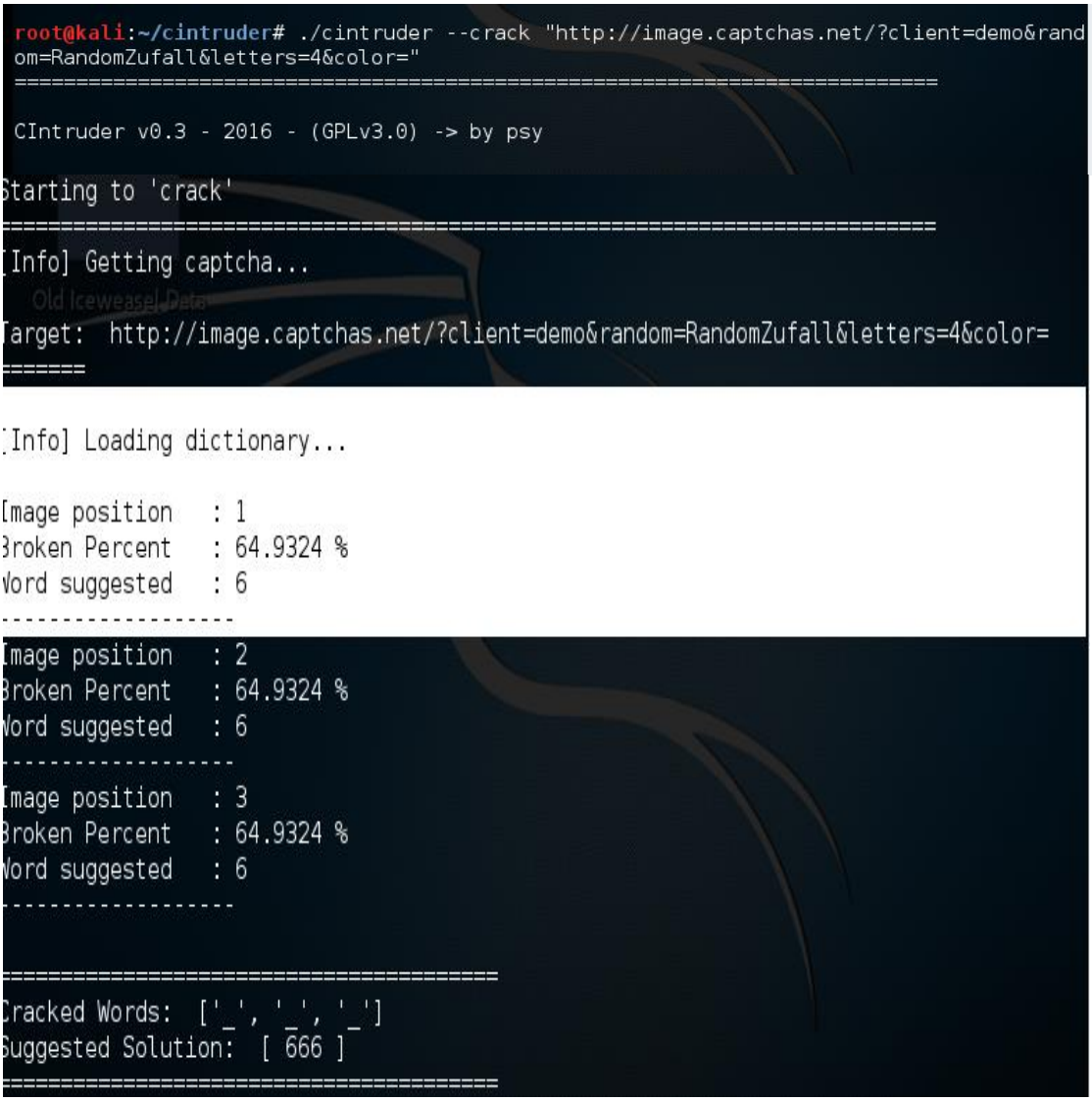
3. Από την μπάρα διευθύνσεων του browser αντιγράφουμε την διεύθυνση της εικόνας captcha.

   <http://image.captchas.net/?client=demo&random=RandomZufall&letters=4>

4. Ανοίγουμε ένα terminal και εφόσον έχουμε μπει στον φάκελο cintrunder (cd cintrunder) δίνουμε την εντολή: ./cintruder --crack –u “

   http://image.captchas.net/?client=demo&random=RandomZufall&letters=4”

Στην συνέχεια έχουμε το παρακάτω αποτέλεσμα.



Εικόνα 5.13: Επίθεση ωμής βίας σε ιστοσελίδα με CAPTCHA.

Στο παραπάνω captcha δεν ήταν δυνατή η λύση του γιατί θα πρέπει κάθε χαρακτήρα του συγκεκριμένου captcha να το προσθέσουμε στο λεξικό του προγράμματος που βρίσκεται στην διαδρομή : home/cintrunder/dictionary.

5.3.4 Επίλυση CAPTCHA σε συγκεκριμένο αρχείο εικόνας.

Για να προβούμε στην επίλυση του captcha μιας συγκεκριμένης διεύθυνσης ενός συγκεκριμένου αρχείου εικόνας μέσω του προγράμματος CINtruder ακολουθούμε τα παρακάτω βήματα:

1. Επιλέγουμε την εικόνα CAPTCHA που θέλουμε να επιλύσουμε και την αποθηκεύουμε σε ένα φάκελο τοπικά (για το παράδειγμα της επίθεσης είναι ο φάκελος Inputs).

Εικόνα 5.14: Εικόνα CAPTCHA προς επίλυση.

2. Ανοίγουμε ένα terminal και εφόσον έχουμε μπει στον φάκελο cintrunder (cd cintrunder) δίνουμε την εντολή: ./cintruder –crack "inputs/test1.gif" και εμφανίζεται το παρακατω αποτέλεσμα επιτυχούς επίθεσης σε captcha αρχείου εικόνας.

```
CIntruder v0.3 - 2016 - (GPLv3.0) -> by psy

=============================================
Starting to 'crack'
=============================================
Target:  inputs/test1.gif
========

[Info] Loading dictionary...

Image position    : 1
Broken Percent    : 100 % [+CRACKED!]
Word suggested    : p
-------------------
Image position    : 2
Broken Percent    : 100 % [+CRACKED!]
Word suggested    : 3
-------------------
Image position    : 3
Broken Percent    : 100 % [+CRACKED!]
Word suggested    : 6
-------------------
Image position    : 4
Broken Percent    : 100 % [+CRACKED!]
Word suggested    : w
-------------------


=======================================
Cracked Words:  ['p', '3', '6', 'w']
Suggested Solution:  [ p36w ]
=======================================
```

Εικόνα 5.15: Επιτυχημένη επίθεση σε αρχείο CAPTCHA με το πρόγραμμα Cintrunder.

Παρατηρούμε ότι έγινε επιτυχής επίθεση στο συγκεκριμένο captcha γιατί είχαν προστεθεί όλα τα μοτίβα των χαρακτήρων του κάθε ενός ξεχωριστά στον φάκελο dictionary.

5.3.5 Δημιουργώντας το δικό σου CAPTCHA

Ένα σύστημα CAPTCHA ανάλογα με την τεχνική που θέλει κάποιος να δημιουργήσει μπορεί να καταλήξει να είναι μια αρκετά δύσκολη και πολύπλοκη διαδικασία ανάπτυξης από άποψη γνώσεων, επένδυσης σε χρόνο ανάπτυξης και αρκετές φορές σε υπολογιστική ισχύ. Ο συνηθέστερος δρόμος είναι η χρήση μίας έτοιμης λύσης. Σε αρκετές όμως τεχνικές όπως για παράδειγμα στα Text ή Image CAPTCHA ακόμα και ένας απλός χρήστης χωρίς να χρειάζεται να έχει αρκετές γνώσεις στις γλώσσες προγραμματισμού και με την βοήθεια του διαδικτύου μπορεί να δημιουργήσει ένα τεστ. Υπάρχει η δυνατότητα κατασκευής τους μέσω οδηγιών και σεμιναρίων που μπορεί κάποιος να παρακολουθήσει, όπου αναλύεται ο τρόπος δημιουργίας τους και η ενσωμάτωση τους στις ιστοσελίδες για την προστασία των υπηρεσιών τους. Στα εργαλεία επίλυσης μέσω Ανθρώπινων Υπηρεσιών οι δημιουργοί των CATCHA έχουν την δυνατότητα να τα υποβάλουν ώστε αυτά να αξιολογηθούν, για παράδειγμα στο εργαλείο 9kw.eu .

Συνήθως χρησιμοποιούνται τεχνολογίες PHP και AJAX για την δημιουργία των CAPTCHA. Η βιβλιοθήκη captcha.php είναι πολύ φιλική προς το χρήστη και εξαιρετικά εύκολη στην ενσωμάτωση σε υπάρχουσες μορφές λόγω του απλού Application Programming Interface (API). Ο κώδικας PHP μπορεί να θέσει ερωτήματα σε βάσεις δεδομένων, να δημιουργήσει εικόνες, να διαβάσει και να γράψει αρχεία και να συνδεθεί με απομακρυσμένους υπολογιστές.

Κεφάλαιο 6
Συμπεράσματα – Επίλογος

Σύμφωνα με τα παραπάνω διαπιστώνουμε ότι οι χρήστες του Internet έρχονται καθημερινά αντιμέτωποι με όλο και περισσότερες δοκιμασίες CAPTCHA, οι οποίες εξελίσσονται προκειμένου οι υπηρεσίες που παρέχουν οι ιστοσελίδες στους επισκέπτες τους να είναι περισσότερο ασφαλή. Πάρα την εξέλιξη των τεχνικών των CAPTCHA οι επιτιθέμενοι συνεχίζουν να επιδίδονται σε ένα αγώνα προκειμένου να τις λύσουν, με κύριο στόχο κυρίως οικονομικές απολαβές.

Στην παρούσα μελέτη έγινε παρουσίαση των Τεχνικών και των Τύπων των CAPTCHA μέσα από μια αναλυτική παρουσίαση των πλεονεκτημάτων και των μειονεκτημάτων από την πλευρά της ασφάλειας. Όλες οι τεχνικές CAPTCHA ανάλογα με την υπηρεσίες της ιστοσελίδας και τους σκοπούς που θέλουν οι δημιουργοί της να πέτυχουν, παρέχουν ασφάλεια. Ιστοτόποι με μεγάλη επισκεψιμότητα είναι συνήθως πολύ ελκυστικοί σε κακόβουλους χρήστες, κατά συνέπεια οι τεχνικές CAPTCHA που επιλέγονται σε αυτή την περίπτωση διαφέρουν πολύ σε σχέση με άλλους που δεν είναι εξίσου δημοφιλής.

Στη συνέχεια έγινε χρήση και παρουσίαση ελευθέρων λογισμικών επίλυσης των CAPTCHA με στόχο να δείξουμε τον τρόπο με τον οποίο οι κακόβουλοι χρήστες μπορούν να πραγματοποιήσουν μια επιτυχημένη επίθεση σε ένα CAPTCHA. Από τη μελέτη που έγινε διαπιστώθηκε ότι δεν μπορούν να βρουν εφαρμογή σε όλες τις διευθύνσεις του ίντερνετ που χρησιμοποιούν τα

CAPTCHA και, αυτό γιατί οι περισσότερες έχουν ενημερωθεί με νέες και εξελιγμένες μορφές τους προκειμένου να θωρακιστούν απέναντι σε οποιαδήποτε επιχείρηση εισβολής.

Το Internet είναι ένας δυναμικός χώρος ο οποίος εξελίσσετε καθημερινά, η ταχύτητα με την οποία αλλάζουν τα δεδομένα είναι τεράστια. Η αλματώδης εξέλιξη της τεχνολογίας βοηθά τους κακόβουλους χρήστες να βρίσκουν συνεχώς νέους και εφευρετικούς τρόπους για να επιτεθούν στους στόχους τους. Από την ερευνά που έγινε παρατηρήθηκε ότι τόσο οι τεχνικές δημιουργίας CAPTCHA όσο και οι τεχνικές επιθέσεων σε CAPTCHA αναπτύσσονται παράλληλα και αποτελούν ανοικτό ερευνητικό πεδίο για συνεχή εξέλιξη.

Βιβλιογραφία

[01] A. Desai and P. Patadia, "Drag and drop: A better approach to CAPTCHA," Proc. INDICON 2009 - An IEEE India Counc. Conf., 2009.

[02] A. Schlaikjer, "A dual-use speech CAPTCHA: Aiding visually impaired web users while providing transcriptions of Audio Streams," LTI-CMU Tech. Rep., 2007.

[03] B. Pinkas and T. Sander, "Securing passwords against dictionary attacks," Proc. 9th ACM Conf. Comput. Commun. Secur. - CCS '02, p. 161, 2002.

[04] E. Athanasopoulos and S. Antonatos, "Enhanced CAPTCHAs : Using Animation to Tell Humans and Computers Apart," Ifip Int. Fed. Inf. Process., vol. 4237, pp. 97–108, 2006.

[05] E. Bursztein, S. Bethard, C. Fabry, J. C. Mitchell, and D. Jurafsky, "How good are humans at solving {CAPTCHA}? A large scale evaluation," Proc. 2010 IEEE Symp. Secur. Priv., pp. 399–413, 2010.

[06] G. Martinovic and Z. Krpic, "Advanced character collage captcha," Acta Polytech. Hungarica, vol. 9, no. 6, pp. 137–151, 2012.

[07] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: breaking a visual CAPTCHA," 2003 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, 2003. Proceedings., vol. 1, pp. 1–8, 2003.

[08] J. Cui, L. Wang, J. Mei, D. Zhang, X. Wang, Y. Peng, W. Zhang, "CAPTCHA design based on moving object recognition problem," Inf. Sci. Interact. Sci. (ICIS), 2010 3rd Int. Conf., pp. 158–162, 2010.

[09] J. Elson, J. R. Douceur, J. Howell, and J. Saul, "Asirra: A CAPTCHA that Exploits Interest-Aligned Manual Image Categorization," Proc. 14th ACM Conf. Comput. Commun. Secur., pp. 366–374, 2007.

[10] J. S. Cui, J. T. Mei, W. Z. Zhang, X. Wang, and D. Zhang, "A CAPTCHA implementation based on moving objects recognition problem," Proc. Int. Conf. E-bus. E-Government, ICEE 2010, pp. 1277–1280, 2010.

[11] J. Tam, J. Simsa, S. Hyde, and L. von Ahn, "Breaking Audio CAPTCHAs.," Nips, pp. 1–8, 2008.

[12] J. Yan and A. S. El Ahmad, "A low-cost attack on a Microsoft captcha," Proc. 15th ACM Conf. Comput. Commun. Secur. - CCS '08, p. 543, 2008.

[13] J. Yan and A. S. El Ahmad, "Usability of CAPTCHAs or usability issues in CAPTCHA design," Proc. 4th Symp. Usable Priv. Secur. - SOUPS '08, p. 44, 2008.

[14] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, "reCAPTCHA: Human-Based Character Recognition via Web Security Measures," Science (80-.)., vol. 321, no. 5895, pp. 1465–1468, 2008.

[15] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using Hard AI Problems for Security," pp. 294–311, 2003.

[16] M. Egele, L. Bilge, E. Kirda, S. Antipolis, and C. Kruegel, "CAPTCHA Smuggling : Hijacking Web Browsing Sessions to Create CAPTCHA Farms," Engineering, pp. 1865–1870, 1865.

[17] M. T. Nayeem, M. M. R. Akand, N. Sakib, and M. W. Ul Kabir, "Design of a Human Interaction Proof (HIP) using human cognition in contextual natural conversation," Proc. 2014 IEEE 13th Int. Conf. Cogn. Informatics Cogn. Comput. ICCI*CC 2014, no. October 2014, pp. 146–154, 2014.

[18] online: http://en.wikipedia.org/wiki/ReCAPTCHA, [20-3-2017]

[19] online: http://scraping.pro/8-best-captcha-solving-services-and-tools/, [01-05-2017]

[20] online: http://scraping.pro/captcha-breaker-review/, [01-05-2017]

[21] online: http://scraping.pro/decaptcher-review/

[22] online: http://www.google.com/recaptcha [20-3-2017]

[23] online: https://cintruder.03c8.net/#installation [01-05-2017]

[24] online: https://deltahacker.gr/captcha-introduction/ [10-03-2017]

[25] online: https://getyourbots.com/captcha-solver/ [01-05-2017]

[26] online: https://moz.com/blog/captchas-affect-on-conversion-rates [11-03-2017]

[27] online: https://www.mcafee.com/us/downloads/free-tools/tessercap.aspx [01-05-2017]

[28] online: https://www.netstudio.gr/blog/conversion-rate-site [11-03-2017]

[29] online:http://www.techi.com/blog/2010/05/are-you-human-captchas-many-ways-of-asking-the-same-question/ [4-3-2017]

[30] online:https://www.raymond.cc/blog/bypass-captcha-firefox-auto-solving-captcha-monster/ [01-05-2017]

[31] online:https://www.w3.org/WAI/GL/wiki/Captcha_Alternatives_and_thoughts[20-3-2017]

[32] online:https://www.exploit-db.com/exploits/17309/ [21-03-2017]

[33] online:https://packetstormsecurity.com/files/121575 [21-03-2017]

[34] online: https://el.wikipedia.org/wiki/Cross-site_scripting [21-03-2017]

[35] P. Baecher, N. Büscher, M. Fischlin, and B. Milde, "Breaking reCAPTCHA: A holistic approach via shape recognition," IFIP Adv. Inf. Commun. Technol., vol. 354 AICT, pp. 56–67, 2011.

[36] S. Vij and H. Rohil, "design and implementation of a user friendly text based captcha," pp. 1056–1066.

[37] S.Goloni, "Αυτόματη Αναγνώριση CAPTCHAs με Χρήση Τεχνικών ΨΕΕ", Πανεπιστήμιο Πατρών, Σχολή Θετικών Επιστήμων, Τμήμα Φυσικής, Μεταπτυχιακό τμήμα Ηλεκτρονικής και Επεξεργασίας της Πληροφορίας , Πάτρα, 2015.

[38] S.Tsixalaki, "Τεχνικές CAPTCHA", Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης, Σχολή Τεχνολογικών Εφαρμογών, Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων, Ηράκλειο, 2013.

[39] Sivakorn, Suphannee, Polakis, Jason, Keromytis, A.D"I'm not a human: Breaking the Google reCAPTCHA",Black Hat,pp.1-12,2016

[40] V. Ragavi and G. Geetha, "CAPTCHA Celebrating its Quattuordecennial – A Complete Reference," vol. 8, no. 6, pp. 340–349, 2011.

[41] X. I. E. Yue, Y. Xi, and Z. Hen, "Research on bu ild in g aud it 2 on lin e network m odes," 2006.

[42] Y. Rui and Z. Liu, "ARTiFACIAL: automated reverse turing test using FACIAL features," Multimed. Syst., pp. 8–11, 2004.