Dirk Fox

Captcha

Hintergrund

„Captcha" ist ein Akronym für ***C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part* – also ein öffentlicher Turing-Test, um Menschen und Maschinen voneinander zu unterscheiden [1].

Als Turing-Test wird ein im Jahr 1950 von Alan Turing publiziertes Verfahren bezeichnet, mit dem er feststellen wollte, ob ein Computer eine dem Menschen vergleichbare Denkleistung erbringen kann [2]. Seine Idee: Ein Tester stellt an einen Computer und einen Menschen, die sich in separaten geschlossenen Räumen befinden, Fragen, die diese beantworten müssen. Aus den Antworten versucht er abzuleiten, ob die Antwort vom Computer oder vom Menschen stammt.

Alan Turing prognostizierte, dass es einem Tester im Jahr 2000 nur noch mit einer höchstens 70%igen Wahrscheinlichkeit gelingen würde, Mensch und Maschine korrekt zu unterscheiden.

Funktionsweise

Als Captcha werden in der Praxis zumeist zufällig gewählte Folgen von Buchstaben und Ziffern als verzerrte oder durch einen „störenden" Hintergrund entfremdete Grafiken angezeigt, aus denen der Benutzer die eingebettete Zeichenfolge herauslesen muss. Bei der Gestaltung der Grafiken wird versucht, es Programmen mit automatischer Zeichenerkennung möglichst schwer zu machen.

Manchmal wird als Captcha auch eine (einfache) Rechenaufgabe als verzerrte Grafik angezeigt, die vom Benutzer zu lösen ist. Oder es sind auf Fotos abgebildete Tier- oder Pflanzenarten zu identifizieren. Manchmal kommen auch Audio- und Video-Captchas zum Einsatz.

Die „Lebensdauer" eines Captchas ist in der Regel begrenzt, oft auf weniger als 60 Sekunden. Erfolgt die (richtige) Antwort nicht in dem vorgegebenen Zeitintervall, wird ein neues Captcha angezeigt.

Einsatz

Captchas werden zumeist verwendet, um den Missbrauch von Foren, Blogs, Online-Shops, Kontaktformularen etc. durch Spam-Robots zu unterbinden. Denn Kommentarmöglichkeiten auf Webseiten werden gerne für Werbezwecke oder die Verbreitung politischer Meinungen „genutzt" – meist unter Verwendung von Tools, die die Suche nach Formularen und den Texteintrag automatisiert vornehmen. Da solche Zweckentfremdungen reguläre Nutzer eines Forums oder Leser eines Blogs schnell vertreiben, lassen viele Betreiber Eingaben durch ein Captcha bestätigen, um möglichst sicher zu gehen, dass der Eintrag von einem menschlichen Nutzer stammt.

Bei einigen Anwendungen dienen Captchas dazu, die Geschwindigkeit der Datenabfrage zu begrenzen, um automatisierte Massenabfragen zu unterbinden, die in kurzer Zeit den gesamten Inhalt der dahinter liegenden Datenbank „auslesen". Und bei Online-Umfragen lässt sich mit Captchas eine Ergebnisverfälschung durch Mehrfachteilnehmer zumindest erschweren.

Captchas werden auch von einigen Online-Banking-Lösungen verwendet, um trojanischen Pferden den Kontozugriff mit der abgefangenen PIN eines Nutzers zu erschweren. Bei anderen Anwendungen werden sie statt Login-Sperren (nach Fehleingaben) eingesetzt, indem sie schnelle lexikalische Passwort-Attacken ausbremsen.

Sicherheit

Die Sicherheit eines Captchas – genauer: die Zuverlässigkeit, dass ein Captcha nur von einem menschlichen Benutzer in der gegebenen Zeit gelöst werden kann – hängt im Wesentlichen davon ab, ob das dem Captcha zu Grunde liegende AI-Problem (*Artificial Intelligence*) für einen Computeralgorithmus (in vernünftiger Zeit) lösbar ist oder nicht.

Es hat sich gezeigt, dass die häufig verwendeten einfachen Text- und Ziffern-Captchas heutigen Zeichenerkennungsalgorithmen nicht (mehr) gewachsen sind [3]. Zwar sind in der Praxis oft noch Verbesserungen beim eingesetzten Captcha-Verfahren möglich (zufällige Länge der Folge, wechselnde Zeichengrößen, sich überschneidende oder durchgestrichene Zeichen); damit steigt jedoch auch die Schwierigkeit für einen menschlichen Nutzer, die Zeichenfolge in kurzer Zeit richtig zu erkennen. Wesentlich zuverlässiger sind Captchas, die die Fähigkeiten des Menschen zur Bilderkennung besser nutzen, wie z. B. die Suche nach einem zusammengehörigen Bildpaar aus mehreren Fotografien.

Die mit einem Captcha erreichbare Sicherheit kann auch aus anderen Gründen begrenzt sein: Ist z. B. die Zahl der verschiedenen vorausberechneten Captchas zu klein, kann ein automatischer Angreifer so lange neue Captchas anfordern, bis eines erscheint, dessen Lösung er bereits einmal (möglicherweise mit Zeitüberschreitung) gefunden hat. Es wurden auch schon trojanische Pferde entdeckt, die das Captcha auf eine „eigene" Webesite (mit z. B. einem Porno-Angebot) kopieren – und es dort von einem Besucher lösen lassen.

Fazit

Captchas können automatisierten Angreifern die missbräuchliche Nutzung von Eingabefeldern und Web-Formularen erschweren, bieten jedoch keinen verlässlichen Schutz. Denn die rasanten Fortschritte in Sprach- und Bilderkennung ermöglichen es Angreifern, immer schwierigere Aufgaben mit vertretbarem Aufwand ebenso schnell und zuverlässig wie ein Mensch zu lösen. Bisher ist keine für ein Captcha nutzbare Aufgabenstellung bekannt, die den Turing-Test 100%ig besteht.

Literatur

[1] Wikipedia: *Captcha*. http://de.wikipedia.org/wiki/captcha

[2] Alan Turing: *Computing Machinery and Intelligence*. Mind 59, Nr. 236, Oktober 1950, S. 433–460. http://www.loebner.net/Prizef/TuringArticle.html

[3] Elie Bursztein, Matthieu Martin, John C. Mitchell: *Text-based CAPTCHA Strengths and Weaknesses*. ACM Computer and Communication Security 2011 (CSS 2011). http://cdn.ly.tl/publications/text-based-captcha-strengths-and-weaknesses.pdf