WIKIPEDIA

CAPTCHA

Da Wikipedia, l'enciclopedia libera.

Un captcha di "6138B" ottenuto utilizzando uno sfondo confuso, scritto con font diversi, di colori vari e non allineati

Con l'acronimo inglese **CAPTCHA** (pronuncia: [ˈkæp.tʃə]) si denota nell'ambito dell'informatica un test fatto di una o più domande e risposte per determinare se l'utente sia un umano (e non un computer o, più precisamente, un bot).

L'acronimo, che si pretende derivato dall'inglese "***C**ompletely **A**utomated **P**ublic **T**uring-test-to-tell **C**omputers and **H**umans **A**part*" ("Test di Turing pubblico e completamente automatico per distinguere computer e umani"), in effetti riproduce foneticamente l'espressione colloquiale *"Caught you!" (Ti ho beccato!)*. Il termine è stato coniato nel 2000 da Luis von Ahn, Manuel Blum e Nicholas J. Hopper della Università Carnegie Mellon e da John Langford della IBM. Come di consueto nella terminologia informatica, il termine inglese viene utilizzato anche nella terminologia informatica italiana.

Un test *CAPTCHA* tipicamente utilizzato è quello in cui si richiede all'utente di scrivere quali siano le lettere o i numeri presenti in una sequenza, che appare distorta o offuscata sullo schermo.

Dal momento che il test viene gestito da un computer, mentre il test originale di Turing viene gestito da un umano, a volte si descrive il test *CAPTCHA* come un test di Turing inverso; questa è però una definizione fuorviante, perché potrebbe indicare anche un test di Turing in cui entrambi i partecipanti tentano di provare che non sono umani infatti da gergo letterale *tu sei un robot*.

Indice

Descrizione
- Origini
- Applicazioni
- Caratteristiche
- Accessibilità

Contromisure

reCaptcha

CAPTCHA nella cultura di Internet

Note

Voci correlate

Altri progetti

Collegamenti esterni

Descrizione

Origini

I *CAPTCHA* sono stati sviluppati per la prima volta nel 1997 dal settore ricerca e sviluppo di AltaVista capitanato da Andrei Broder, per impedire ai bot di aggiungere URL al loro motore di ricerca. Broder e colleghi cercavano di creare immagini resistenti agli attacchi degli OCR e così consultarono il manuale degli scanner della Brother, su cui erano indicate tutte le caratteristiche che un testo deve avere per poter essere riconosciuto dallo scanner: caratteri ben definiti e lineari, mancanza di differenze tra i font utilizzati nel testo, sfondo omogeneo e ben distinguibile dal testo e così via. Applicando al contrario queste indicazioni, si riuscì a ottenere la peggiore situazione possibile, ossia un testo la cui scansione sarebbe stata molto difficile: caratteri storti, font diversi, colori del testo simili a quello dello sfondo o accorgimenti simili. Broder sostenne che l'introduzione di tale tecnologia avesse ridotto lo spam di oltre il 95%.

In modo indipendente dal team di AltaVista, Luis von Ahn e Manuel Blum realizzarono e diffusero nel 2000 l'idea del test *CAPTCHA*, intendendo con ciò qualunque tipo di programma che fosse in grado di distinguere tra persone e computer. Loro inventarono vari tipi di test, compresi i primi a ricevere un'ampia diffusione grazie all'uso da parte di Yahoo!

Applicazioni

I *CAPTCHA* sono utilizzati per impedire che i bot utilizzino determinati servizi, come i forum, la registrazione presso siti web, la scrittura di commenti e in generale tutto ciò che potrebbe essere usato per creare spam o per violare la sicurezza con operazioni di hacking come il brute force. Questo tipo di test è stato utilizzato anche per contrastare lo spam generato da bot, obbligando il mittente di un messaggio e-mail non conosciuto dal destinatario a superare un test *CAPTCHA* prima di consentire la consegna del messaggio.

Caratteristiche

Per definizione i test *CAPTCHA* sono completamente automatici e non richiedono di norma interventi umani per la somministrazione o la manutenzione, con indubbi vantaggi in termini di costi e affidabilità.

Gli algoritmi utilizzati per realizzare i test vengono spesso divulgati al pubblico, anche se in molti casi sono protetti da brevetto. Tale politica di trasparenza è tesa a dimostrare il fatto che la sicurezza del metodo non risiede nella conoscenza di un algoritmo segreto (che potrebbe essere ricavata con tecniche di reverse engineering o in modo fraudolento); al contrario, per 'rompere' l'algoritmo è necessario risolvere un problema classificato come 'hard' nel campo dell'intelligenza artificiale.

Non è obbligatorio ricorrere a tecniche visive: qualunque problema di intelligenza artificiale che abbia lo stesso grado di complessità, ad esempio il riconoscimento vocale, è adatto a fare da base per un test di questo tipo. Alcune implementazioni consentono all'utente di scegliere in alternativa un test basato su tecniche auditive, anche se tale approccio ha subito uno sviluppo più lento e non è detto che possieda lo stesso grado di efficacia di quello visivo. Inoltre, è possibile ricorrere ad altri tipi di verifiche che richiedano un'attività di comprensione testuale, quali la risposta a una domanda o a un quiz logico, il seguire delle specifiche istruzioni per creare una password ecc. Anche in questo caso i dati sulla resistenza di tali tecniche alle contromisure sono scarsi.

Una promettente tecnica che si sta sviluppando negli ultimi anni impiega dei test basati sul riconoscimento di una faccia all'interno di un'immagine familiare. Per questo tipo di *CAPTCHA* si parla di *RTT based on faces recognition*. In letteratura allo stato attuale sono stati implementati soltanto due metodi basati su questo tipo di *CAPTCHA*: l'ARTiFACIAL[1] e un *CAPTCHA* basato sul riconoscimento facciale.[2][3]

Accessibilità

L'uso di test *CAPTCHA* basati sulla lettura di testi o altre attività legate alla percezione visiva impedisce o limita fortemente l'accesso alle risorse protette agli utenti con problemi di vista e, poiché tali test sono progettati specificamente per non essere leggibili da strumenti automatici, i normali ausili tecnologici usati dagli utenti ciechi o

ipovedenti non sono in grado di interpretarli; ma anche gli utenti daltonici possono non essere in grado di superare il test. L'uso dei test *CAPTCHA*, generalmente legato alle fasi iniziali di accesso o registrazione ai siti e talvolta ripetuto per ogni accesso, può costituire una discriminazione nei confronti di tali utenti disabili tale per cui in alcuni ordinamenti esso costituisce una violazione delle norme di legge.

Nelle nuove generazioni di *CAPTCHA*, create per resistere ai più sofisticati programmi di riconoscimento di testi, può diventare abbastanza complicato, se non impossibile, riuscire a riconoscere il testo da parte di molti utenti, anche nel pieno possesso della propria capacità visiva.

Il W3C ha redatto un rapporto in cui vengono evidenziati alcuni dei problemi di accessibilità legati all'uso di tali tecniche.[4]

Contromisure

Dopo l'uso massiccio di *CAPTCHA*, sono state scoperte alcune contromisure che permettono agli spammer di superare i test.

Software intelligenti sono oggi in grado di risolvere *CAPTCHA* di varie tipologie.[5]

Greg Mori e Jitendra Malik hanno presentato nel 2003 uno studio[6] che illustra come aggirare uno dei sistemi più diffusi per realizzare test *CAPTCHA*, EZ-Gimpy; tale approccio si è rivelato efficace nel 92% dei casi. Nei confronti del sistema Gimpy, più sofisticato ma meno diffuso, l'efficacia del metodo scende al 33%. Al momento non è però noto se tale algoritmo sia stato implementato al di fuori del contesto della ricerca.

Sono stati creati anche alcuni programmi per cercare una soluzione ripetutamente e altri per riconoscere i caratteri scritti, utilizzando tecniche apposite e non quelle standard degli OCR. Progetti come PWNtcha[7] hanno fatto grandi passi in avanti, contribuendo alla generale migrazione verso *CAPTCHA* sempre più difficili.

Un altro metodo per superare un *CAPTCHA* è sfruttare sessioni in cui il test è già stato superato, salvando i test per poi creare un archivio di soluzioni.

Ma il metodo più efficace è quello di utilizzare un essere umano per risolvere il *CAPTCHA*: è infatti possibile affidare a persone pagate il compito di risolvere i *CAPTCHA*. Il già citato documento del W3C[4] afferma che un operatore può facilmente risolvere centinaia di test *CAPTCHA* in un'ora.

Questa possibile soluzione necessiterebbe di un investimento economico che non sempre è giustificato, ma è stato scoperto un metodo più a buon mercato per ottenere gli stessi risultati: lo spammer utilizza a tal fine un sito Internet con un servizio a cui gli utenti umani chiedono l'accesso, che può essere un forum ma anche una collezione di immagini pornografiche. Così, quando un utente chiede di accedere, gli viene proposto un *CAPTCHA* ottenuto dal sito esterno che lo spammer vuole attaccare: il test viene quindi risolto dall'utente, che ottiene in cambio una remunerazione che per lo spammer ha un costo trascurabile, mentre il sistema "ricicla" la soluzione del test per superare la barriera del sito bersaglio.

reCaptcha

I test *CAPTCHA* hanno avuto degli utilizzi secondari non legati unicamente all'eliminazione dello spam: il più noto riguarda il riconoscimento di testi contenuti in libri antichi e si chiama *reCaptcha*. Molte biblioteche stanno provvedendo a convertire in digitale le loro collezioni di antichi testi (anche manoscritti); questa conversione viene ottenuta tramite la digitalizzazione delle pagine e la loro successiva analisi tramite un programma OCR, che analizza le immagini delle pagine ed estrae il testo in esse contenuto. I programmi OCR però interpretano con difficoltà le lettere sbiadite e le pagine ingiallite dei testi antichi e quando non sono in grado di riconoscere con certezza un testo necessitano di un intervento umano, che rallenta il processo e innalza il costo della digitalizzazione.

Ricercatori della Università Carnegie Mellon hanno deciso di utilizzare i sistemi *CAPTCHA* per interpretare le parole dubbie individuate dai programmi OCR. Quando due sistemi OCR identificano in modo diverso una parola, questa viene associata a una parola nota e inviata a un utente che deve superare un test *CAPTCHA* per accedere a un servizio. Si presuppone che se un utente riesce ad individuare correttamente la parola nota, allora individuerà con elevata probabilità anche la parola ignota. Quando tre utenti danno la stessa risposta, il sistema archivia la parola come corretta. Questo sistema ha permesso di convertire 440 milioni di parole con un'accuratezza del 99%. Ad agosto 2008, questo sistema convertiva 4 milioni di parole al giorno.[8] Il progetto in seguito è diventato una *startup company* che nel settembre del 2009 è stata acquisita da Google, il quale ha avviato una procedura di scansione di decine di milioni di libri immagazzinati in centinaia di librerie sparse per il pianeta e intende sfruttare il progetto *reCaptcha* per correggere gli errori derivati dalla scansione OCR dei testi.[9] Il *reCaptcha* può essere assimilato alla categoria dei giochi con uno scopo (GWAP).

CAPTCHA nella cultura di Internet

Uno dei fenomeni di Internet nati su 4chan riguarda proprio il *CAPTCHA*. Esso si riferisce ad un codice in cui si leggeva "Inglip Summoned": ne è scaturita la finta leggenda che un dio oscuro, tale Inglip, sia ritornato sulla terra per trascinarla nell'oscurità. Sono presenti inoltre diversi video su YouTube in cui Inglip darebbe ordini ai suoi adepti, sempre tramite degli stravaganti e spesso incomprensibili codici *CAPTCHA*.[10]

Note

1. **^** http://research.microsoft.com/en-us/um/people/yongrui/ps/mmsj04hip.pdf
2. **^** Free Face Recognition Captcha Downloads: Luxand FaceSDK by Luxand Development, Luxand Blink! Pro by Luxand Development and More (http://www.fileguru.com/apps/face_recognition_captcha)
3. **^** IEEE Xplore - Abstract Page (http://ieeexplore.ieee.org/iel5/10670/33674/01602255.pdf)
4. **^** *a* *b* (EN) Matt May, *Inaccessibility of Visually-Oriented Anti-Robot Tests*, su *W3C Working Group Note*, 23 novembre 2005. URL consultato il 12 luglio 2011.
5. **^** Startup americana crea un software per risolvere il Captcha tramite intelligenza artificiale (http://it.cesarnews.com/a/403/startup-americana-crea-un-software-per-risolvere-il-captcha-tramite-intelligenza-artificiale)
6. **^** (EN) Greg Mori, Jitendra Malik, *Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA* (PDF), su *cs.sfu.ca*. URL consultato il 12 luglio 2011.
7. **^** PWNtcha – Caca Labs (http://sam.zoy.org/pwntcha/)
8. **^** *I testi antichi hanno un futuro "Li salverà un metodo antispam"*, Repubblica.it, 19 agosto 2008. URL consultato il 19 agosto 2008.
9. **^** *Google acquista reCaptcha*, MaCityNet.it, 16 settembre 2009. URL consultato il 16 settembre 2009.
10. **^** *Inglipedia*, su *inglipnomicon.wikia.com*, 16 settembre 2009. URL consultato il 5 giugno 2011.

Voci correlate

- Gioco con uno scopo
- Test di Turing

Altri progetti

- Wikiquote contiene citazioni di o su **CAPTCHA**
- Wikimedia Commons (https://commons.wikimedia.org/wiki/?uselang=it) contiene immagini o altri file su **CAPTCHA (https://commons.wikimedia.org/wiki/Category:Captcha?uselang=it)**

Collegamenti esterni

- (EN) *The Captcha Project*, su *captcha.net*.

- (EN) Inaccessibility of Visually-Oriented Anti-Robot Tests: Problems and Alternatives (http://www.w3.org/TR/turingtest/) (raccomandazione del W3C)
- (EN) Storia del test *captcha* (https://web.archive.org/web/20060110192706/http://www2.parc.com/istl/projects/captcha/history.htm) (Xerox PARC)
- (EN) Il brevetto USA 6.195.698 (http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=6195698.PN.&OS=PN/6195698&RS=PN/6195698) (nota: il termine *captcha* non era ancora stato inventato all'epoca)

Estratto da "https://it.wikipedia.org/w/index.php?title=CAPTCHA&oldid=98612643"

Questa pagina è stata modificata per l'ultima volta il 21 lug 2018 alle 02:33.

Il testo è disponibile secondo la licenza Creative Commons Attribuzione-Condividi allo stesso modo; possono applicarsi condizioni ulteriori. Vedi le condizioni d'uso per i dettagli.