AUTOMATYKA • 2008 • Tom 12 • Zeszyt 2

Maciej Laskowski*

Analiza jakości użytkowej najpopularniejszych implementacji testów CAPTCHA

1. Wprowadzenie

W przeciągu ostatniego dziesięciolecia Internet stał się najpopularniejszym medium w historii ludzkości. Zgodnie z ideą *Web 2.0*, środek ciężkości tworzenia treści dostępnych w ogólnoświatowej sieci został przesunięty w kierunku użytkownika – serwisy internetowe przestały jedynie udostępniać określone usługi, stając się elementami pośrednimi w akcie wymiany wiedzy pomiędzy swoimi użytkownikami. Łatwość i prostota umieszczenia *contentu* w Internecie spowodowały jednocześnie pojawienie się potrzeby opracowania i wdrożenia techniki zabezpieczającej przed nadmiernym i niechcianym użyciem systemu – zautomatyzowanym korzystaniem z udostępnianych usług (np. masowym zakładaniem kont emailowych w celach rozsyłania spamu) czy też publikowaniem niechcianych treści.

Jedną z najpopularniejszych idei zabezpieczenia serwisu internetowego można sprowadzić do przeprowadzenia testu Turinga – należy odróżnić akcje wykonywane przez człowieka (które są dozwolone) od tych, które są wykonywane przez program komputerowy (których wykonanie powinno być zabronione) [1]. Metoda ta została zaimplementowana m.in. w technologii CAPTCHA opracowanej w roku 2000 przez naukowców z Carnegie Mellon University we współpracy z IBM.

2. CAPTCHA

CAPTCHA (*Completely Automated Public Turing Test To Tell Computers and Humans Apart*) jest testem typu wyzwanie-odpowiedź (*challenge-response*). Aby wykonać określoną operację w systemie (np. wysłać formularz), użytkownik musi rozwiązać prosty test, który jest wygenerowany i oceniany przez komputer pełniący rolę serwera. W najbardziej 'klasycznej' formie [2] sprowadza się to do wprowadzenia przez użytkownika kodu (zazwyczaj w formie ciągu kilku znaków bądź jednego wyrazu) umieszczonego na obrazku-tokenie. Metoda ta oparta jest na następujących założeniach (za: [3]):

– token użyty w teście jest prosty do odczytania przez człowieka;
– odczytanie tokenu przez komputer jest bardzo trudne, bądź też niemożliwe;

* Instytut Informatyki, Wydział Elektrotechniki i Informatyki, Politechnika Lubelska w Lublinie

- token generowany jest automatycznie przez system i przez ten sam system może zostać oceniony;
- prawdopodobieństwo błędu – uznania człowieka za maszynę, bądź odwrotnie jest bardzo niskie;
- test jest odporny na atak metodą słownikową;
- człowiek będzie w stanie odczytać kod z tokenu bez względu na język jakim się posługuje oraz uwarunkowania kulturowe;
- odporność systemu na zautomatyzowane próby ataku nie wynika z nieznajomości systemu zabezpieczeń przez atakującego – przykładowo pole typu *checkbox* 'Nie jestem botem' w formularzu rejestracyjnym może w pewien bardzo ograniczony sposób pełnić rolę testu służącego do odróżniania użytkowników od komputerów, nie jest jednak testem Turinga – w przypadku zaimplementowania obsługi tego pola w algorytmie programu atakującego test ten przestaje pełnić swoją funkcję. Nie jest także w żaden sposób generowane automatycznie przez system.

Należy zauważyć, że wbrew swojej nazwie test CAPTCHA jest tak naprawdę odwróconym testem Turinga, gdyż jest przeprowadzany przez maszynę, zaś skierowany jest do ludzi.

2.1. Zalety i wady testów CAPTCHA

Pierwsze wdrożenia testów CAPTCHA (dla systemu dodawania nowych stron do bazy wyszukiwarki AltaVista) charakteryzowały się znaczącym ograniczeniem (do 95%, za: [4], [1]) ilości danych wprowadzanych przez boty. Jednak jak dowodzą ostatnie badania ([4, 5, 6]), wraz z rozwojem technologii wykorzystujących algorytmy rozpoznawania znaków OCR (*Optical Character Recognition*) skuteczność łamania niektórych implementacji testów CAPTCHA (wykorzystujących tokeny generowane w oparciu o proste zniekształcenia grafiki, takie jak rozmycie czy dodanie szumu) może nawet zbliżyć się do 100%! Należy zauważyć, że implementacje charakteryzujące się już kilkuprocentowym prawdopodobieństwem złamania nie spełniają swojej zabezpieczającej roli.

Odpowiedzią twórców testów CAPTCHA jest wykorzystywanie coraz to bardziej skomplikowanych przekształceń graficznych w celu wygenerowania tokenu, takich jak niski kontrast, zmniejszenie rozmiaru znaków kodu czy przekształcenia nieliniowe. Według założeń programistów ma to spowodować zwiększenie poziomu trudności odczytu kodu przez automaty. Jak jednak dowodzą badania firmy Microsoft [6], większość przekształceń graficznych wykorzystanych do wygenerowania tokenu nie miała przełożenia na skuteczność działania istniejących obecnie algorytmów OCR użytych do odczytu kodu.

Główną wadą testów CAPTCHA pozostaje jednak brak dostępności testów opartych na tokenach dla niektórych grup użytkowników: m.in. osób niewidomych i z wadami wzroku (np. z zaburzeniem widzenia barw), czy używających przeglądarek tekstowych. Rozpoznanie kodu nie jest więc z punktu widzenia systemu izomorficzne z człowieczeństwem. Wraz ze zwiększeniem liczby przekształceń graficznych używanych do wygenerowania tokenu pojawia się również pytanie o wzrost stopnia skomplikowania odczytu kodu przez użytkowników nie posiadających wad wzroku uniemożliwiających skorzystanie z systemu zabezpieczonego testem CAPTCHA.

3. Badania jakości użytkowej wybranych implementacji testów CAPTCHA

W celu zbadania jakości użytkowej testów CAPTCHA wybrano dziesięć serwisów internetowych, z których pobrano tokeny wygenerowane przy użyciu przekształceń graficznych różnego typu. Kody znajdujące się na tokenach zostały uprzednio sprawdzone, tzn. udało się w sposób pozytywny wykonać operację zabezpieczoną określonym kodem.

Użytkowników biorących udział w badaniu podzielono na dwie szesnastoosobowe grupy. Grupa I otrzymała zestaw tokenów w formie niezmodyfikowanej, zaś grupa II otrzymała ten sam zestaw zmodyfikowany losowo za pomocą aplikacji Colorblind [7], co pozwoliło na zasymulowanie kilku najpopularniejszych zaburzeń w widzeniu barw. Uczestniczący w badaniu mieli za zadanie odczytać kod znajdujący się na tokenie. Podczas badania mierzono dwa parametry:

– poprawność odczytania kodu (w skali 0-1, 1 – odczyt poprawny, 0 – odczyt niepoprawny),
– subiektywną ocenę trudności odczytania kodu (w skali 0-10, 0 – bez trudności, 10 – ekstremalnie trudne).

W trakcie badania nie mierzono czasu odczytu poszczególnych kodów, przyjęto jedynie ograniczenie całkowitego czasu badania do 15 minut.

3.1. Wyniki badań

Wyniki badań zostały przedstawione w tabelach 1 – 10.

Tabela 1
Wyniki badań jakości użytkowej testu CAPTCHA – test 1

Token	
Źródło	yahoo.com
Poprawny kod	2re4eg
Cechy tokenu	czcionka o stałym rozmiarze i kolorze, zniekształcenie obrazu, obrót fragmentów obrazu pod różnym kątem, dodane linie utrudniające wykrycie krawędzi znaków, jednolite tło, duży kontrast
% poprawnych wyników dla I grupy badawczej	62,5%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	2,75
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	75%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	4,75
% poprawnych wyników ogółem	68,75%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	3,75
Uwagi	dodane linie mogą zlewać się ze znakami tokena, 'tworząc' inne znaki

Tabela 2
Wyniki badań jakości użytkowej testu CAPTCHA – test 2

Token	
Źródło	sms.orange.pl
Poprawny kod	elity
Cechy tokenu	czcionka o stałym rozmiarze i kolorze, niewielkie przesunięcie części znaków, dodane w miarę jednolite tło, duży kontrast
% poprawnych wyników dla I grupy badawczej	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,06
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,0
% poprawnych wyników ogółem	100%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	1,03

Tabela 3
Wyniki badań jakości użytkowej testu CAPTCHA – test 3

Token	
Źródło	o2.pl
Poprawny kod	za433
Cechy tokenu	czcionka o różnych rozmiarach i kolorach, wszystkie znaki obrócone o różny kąt, dodane niejednolite kolorowe tło, wysoki kontrast
% poprawnych wyników dla I grupy badawczej	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,13
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,13
% poprawnych wyników ogółem	100%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	1,13

Tabela 4
Wyniki badań jakości użytkowej testu CAPTCHA – test 4

Token	
Źródło	google.com
Poprawny kod	submer
Cechy tokenu	czcionka o jednakowym rozmiarze i kolorze, deformacja części obrazu, jednolite tło, pochylenie poszczególnych znaków tekstu o zbliżony kąt
% poprawnych wyników dla I grupy badawczej	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,06
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	2,13
% poprawnych wyników ogółem	100%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	1,6

Tabela 5
Wyniki badań jakości użytkowej testu CAPTCHA – test 5

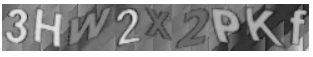
Token	
Źródło	myspace.com
Poprawny kod	3hw2x2pkf
Cechy tokenu	czcionka o niejednolitym rozmiarze i kolorze, różnokolorowe, niejednolite tło, stosunkowo niewielki kontrast
% poprawnych wyników dla I grupy badawczej	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,0
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	62,5%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,13
% poprawnych wyników ogółem	81,25%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	1,06

Tabela 6
Wyniki badań jakości użytkowej testu CAPTCHA – test 6

Token	
Źródło	www.jeans.com.ua/sms
Poprawny kod	xlr
Cechy tokenu	czcionka o niejednolitym rozmiarze i kolorze, jednolite tło, znaki zachodzą na siebie (pseudotrójwymiar), specyficzna czcionka
% poprawnych wyników dla I grupy badawczej	87,5%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	2,38
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	75%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	3,75
% poprawnych wyników ogółem	81,25%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	3,06

Tabela 7
Wyniki badań jakości użytkowej testu CAPTCHA – test 7

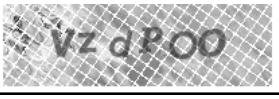
Token	
Źródło	mail9.com
Poprawny kod	jlvzdpoo
Cechy tokenu	znaki umieszczone w losowym rozstrzeleniu, relatywnie niejednolite tło, nieduży kontrast, losowe przesunięcie wybranych znaków
% poprawnych wyników dla I grupy badawczej	0%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,5
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	0%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	2,75
% poprawnych wyników ogółem	0%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	1,625

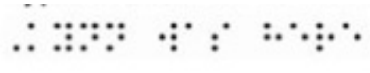
Tabela 8
Wyniki badań jakości użytkowej testu CAPTCHA – test 8

Token	
Źródło	di.com.pl
Poprawny kod	888852
Cechy tokenu	znaki tokenu zapisane jednokolorową czcionką o identycznym rozmiarze, duże rozstrzelenie znaków, w tle umieszczone mniejsze znaki w innym kolorze
% poprawnych wyników dla I grupy badawczej	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	1,88
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	100%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	3,63
% poprawnych wyników ogółem	100%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	2,75

Tabela 9
Wyniki badań jakości użytkowej testu CAPTCHA – test 9

Token	
Źródło	hemmy.net
Poprawny kod	3wcre4
Cechy tokenu	znaki tokenu, zapisane różnokolorową czcionką o identycznym rozmiarze, zostały zastąpione znakami autorskiego alfabetu, relatywnie jednolite tło, relatywnie duży kontrast
% poprawnych wyników dla I grupy badawczej	75%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	6,38
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	18,75%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	6,75
% poprawnych wyników ogółem	46,875%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	6,56
Uwagi	Niektóre znaki autorskiego alfabetu różniły się między sobą tylko i wyłącznie kolorem – w przypadku osób z zaburzeniem widzenia barw znaki te mogły wydawać się identyczne, co znacząco utrudniało dokonanie podstawienia

Tabela 10
Wyniki badań jakości użytkowej testu CAPTCHA - test 10

Token	
Źródło	crookedbrains.net
Poprawny kod	uynnwashere
Cechy tokenu	zastąpienie alfabetu łacińskiego alfabetem specjalnym – w tym przypadku alfabetem Braille'a, jednolity kolor i rozmiar czcionki, jednolite tło, wysoki kontrast
% poprawnych wyników dla I grupy badawczej	0%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	8,63
% poprawnych wyników dla II grupy badawczej (symulacja zaburzenia widzenia barw)	0%
Średnia subiektywna ocena trudności odczytu kodu w I grupie badawczej	10
% poprawnych wyników ogółem	0%
Średnia subiektywna ocena trudności odczytu kodu dla obu grup badawczych	9,31
Uwagi	poprawne odczytanie tokenu wymaga znajomości alfabetu Braille'a – ogranicza to znacznie potencjalną liczbę użytkowników

3.2. Analiza i interpretacja wyników badań

Jak dowiodły testy nr 2, 3 oraz 4, użytkownicy z obydwu grup nie mieli problemów z odczytaniem tokenów o wysokim kontraście, wygenerowanych za pomocą bardzo prostych przekształceń graficznych, takich jak rozmycie, dodanie jednostajnego szumu, czy prostego tła. Należy zauważyć, że niektóre proste przekształcenia, takie jak dodanie elementów utrudniających rozpoznanie kształtu liter przez algorytmy OCR mogą, ze względu na losowość generowanych efektów, w znaczący sposób utrudnić (bądź nawet uniemożliwić) prawidłowe odczytanie tokenu, czego dowodem mogą być wyniki testu nr 1. Również losowe przesunięcia znaków mogą w szczególnych przypadkach uniemożliwić poprawne odczytanie kodu – żaden z badanych nie zauważył litery 'l' umieszczonej na tokenie z testu 7.

Ważnym elementem, który ma wpływ na poziom trudności odczytu tokenu przez użytkownika jest dobór odpowiedniej palety wykorzystywanych barw. Do wygenerowania tokenu używanego przez serwis myspace.com (test nr 5) wykorzystano zbyt ubogą paletę kolorów. Dowiodły tego wyniki testu – ponad 1/3 użytkowników, którzy otrzymali token zmodyfikowany przez aplikację Colorblind nie była w stanie prawidłowo odczytać kodu.

Ciekawą techniką, która ma zabezpieczyć tokeny przez możliwością odczytu przez oprogramowanie wykorzystujące algorytmy OCR, jest umieszczanie znaków kodu w przestrzeni pseudotrójwymiarowej. Jednak jak wynika z analizy wyników testu nr 6, takie rozmieszczenie znaków, połączone z użyciem specyficznej czcionki może błędne zasugerować użytkownikowi, że znaki kodu należy odczytywać nie od lewej do prawej, tylko przestrzennie, z góry na dół.

Znacznie częściej stosowanym rozwiązaniem jest generowanie tokenu wykorzystującego mieszanie znaków o różnym charakterze – kod stanowią tylko znaki jednego typu (np. liczby), zaś pozostałe znaki umieszczane są jako tło mające na celu oszukać program próbujący złamać to zabezpieczenie. Jak dowodzą wyniki testu nr 8, metoda ta nie sprawia użytkownikom większego problemu – obydwie grupy badawcze uzyskały 100% poprawność odczytu kodu przy stosunkowo niewielkim wzroście poziomu trudności.

Rozwiązaniem, które zdobywa ostatnio popularność [1] jest zastąpienie znaków alfabetu używanego języka innym alfabetem. Tokeny wykorzystane w testach 9 i 10 reprezentują dwojakie podejście twórców – w pierwszym przypadku jest to alfabet autorski, objaśniony na samym tokenie. Niewątpliwą zaletą tego rozwiązania jest fakt, że użytkownik nie musi znać wcześniej tego alfabetu. Jest to rozwiązanie zwiększające jednak poziom trudności testu CAPTCHA, szczególnie w przypadku użycia niejednoznacznych symboli bądź wąskiej palety barw. Do wygenerowania tokenu można również użyć istniejącego już alfabetu. Powoduje to jednak zawężenie grupy potencjalnych użytkowników systemu do osób znających tenże alfabet. Test 10 świetnie ilustruje ten przypadek – poszczególne znaki kodu zostały zastąpione swoimi odpowiednikami z alfabetu Braille'a, który nie był znany żadnej z badanych osób. Należy również podkreślić, że token użyty w tym teście paradoksalnie nie stanowiłby żadnego zabezpieczenia przez programami do rozpoznawania znaków, o ile tylko miałyby one zaimplementowaną obsługę alfabetu Braille'a.

3.3. Istniejące alternatywy dla CAPTCHA

Omawiając testy CAPTCHA należy zastanowić się nad możliwymi alternatywami dla istniejących rozwiązań. System oparty na generowaniu tokenów z kodami jest z założenia niedostępny dla osób niewidomych lub posiadających poważne wady wzroku. Poprzez losowość niektórych stosowanych przekształceń graficznych tokeny mogą być również nieczytelne dla użytkowników, którzy na co dzień nie mają problemów ze wzrokiem. Czym więc można byłoby je zastąpić? Jednym z proponowanych rozwiązań jest stosowany m.in. przez firmę Yahoo system alternatywnych tokenów dźwiękowych – kod jest odczytywany użytkownikowi, bądź tożsamość użytkownika jest potwierdzana drogą telefoniczną. Do wad tego rozwiązania należą jednak ciągle duże koszty (szczególnie w przypadku aktywacji telefonicznej), a także bariera językowa. Innym interesującym rozwiązaniem jest rozwijany przez Microsoft system Assira [8] – użytkownik musi wybrać z zestawu obrazków te, które spełniają określony warunek (przykładowo ze zbioru zdjęć kotów i psów musi wybrać tylko te, które przedstawiają koty). Wielu twórców serwisów internetowych wy-

korzystuje do ich zabezpieczenia testy MAPTCHA (*Mathematical CAPTCHA*) – wymagające od użytkownika rozwiązania prostych działać matematycznych. W3 Consortium proponuje również rozwiązania oparte o tzw. '*common sense puzzle logic*', czyli o zestaw prostych pytań i odpowiedzi (np. 'jakiego koloru jest niebo?') [9]. Należy jednak zauważyć, że to ostatnie rozwiązanie nie spełnia założeń systemu CAPTCHA, gdyż nie jest automatycznie generowane przez system.

4. Wnioski

Analizując wyniki przeprowadzonych testów należy zauważyć, że dobrze zaprojektowane testy CAPTCHA, oparte o przemyślany system generowania tokenów, nie muszą być wcale znaczącym utrudnieniem dla użytkownika, przy jednoczesnym zapewnieniu systemowi odpowiedniego poziomu ochrony. Jako przykład w literaturze [1, 4], wymieniany jest najczęściej generator tokenów firmy Google [10], co potwierdzają wyniki testu nr 4. Ostatnio jednak pojawiły się głosy, iż generator ten został złamany [11].

Systemy oparte o tokeny dźwiękowe wydają się być interesującą i przyszłościową alternatywą, podobnie jak system Assira. Warto jednak zauważyć, że z proponowanych obecnie alternatywnych rozwiązań tylko tokeny dźwiękowe umożliwiają skorzystanie z systemu osobom niewidomym.

Literatura

[1] http://en.wikipedia.org/wiki/Captcha.
[2] Lesiński K., *CAPTCHA – jak odróżnić złe od gorszych*. http://pornel.net/captcha.
[3] *History of CAPTCHA*, http://www2.parc.com/istl/projects/captcha/history.htm.
[4] Golański A., *CAPTCHA: Odróżniamy maszyny od ludzi*. http://webhosting.pl/layout/set/print/content/view/full/2806.
[5] Hocevar S., *PWNtcha*. http://sam.zoy.org/pwntcha.
[6] Chellapila K. et al., *Computers beat Humans at Single Character Recognition in Reading based Human Interaction Proofs (HIPs)*. Microsoft Research, 2005.
[7] http://colorfilter.wickline.org.
[8] http://research.microsoft.com/asirra.
[9] May M., *Inaccessibility of CAPTCHA. Alternatives to Visual Turing Tests on the Web*. w3 Working Group Note, 2005.
[10] http://google.com.
[11] *Google's CAPTCHA busted in recent spammer tactics*. Websense Security Labs Threat Blog, http://www.websense.com/securitylabs/blog/blog.php?BlogID=174.