Integracija Web API reCAPTCHA tehnologije na Web-stranicu

Zoran Veličković, Zoran Milivojević
Visoka tehnička škola strukovnih studija Niš
Niš, Srbija
zoran.velickovic@vtsnis.edu.rs, zoran.milivojevic@vtsnis.edu.rs

Sažetak— **U ovom radu je prikazana integracija zaštite Web-stranice od napada malicioznog automatizovanog softvera na HTML forme. Maliciozni softver imitira rad čoveka na HTML formi i time izaziva probleme u funkcionisanju Web-stranice. Zbog toga je izuzetno važno sprečiti automatizovani softver da pristupi HTML formama. U praktičnom delu rada je integrisana besplatna Guglova tehnologija - reCAPTCHA kojom se ustanovljava da li je formi pristupio čovek ili automatizovani softver. Za integraciju reCAPTCHA-e na Web-stranicu koristi se Web API razvijen od strane Gugla. Analiza rezultata reCAPTCHA testa obavlja se na serverskoj strani Web-aplikacije korišćenjem PHP skripta. Jednostavna integracija i pouzdan test reCAPTCHA tehnologiju preporučuju za korišćenje u visokopouzdanim Web-stranicama za zaštitu od napada automatizovanog softvera.**

Ključne riječi - Softverski robot. reCAPTCHA. Tjuringov test. Web API.

I. UVOD

Sloboda pristupu Web sadržajima je osnova popularnosti Interneta kod korisnika. Korisnici ne samo da mogu slobodno pristupiti Web sadržajima već ih sami mogu kreirati i postavljati na mreži. Postavljeni sadržaji su potom globalno dostupni. Pored pristupa korisnicima, savremeni Web omogućava i pristup automatizovanim softverima. Ova vrsta komunikacije na Webu je poznata kao M2M (engl. *Machine to machine*) komunikacija čiji će udeo u globalnim komunikacijama sa 37% u 2017. porasti na 51% u 2022. godini. Očekuje se da će tada biti oko 14.6 milijardi M2M konekcija [1]. Intenzivno se razvijaju M2M aplikacije koje se odnose na pametna brojila, video nadzor, nadgledanje zdravstvene zaštite, prevoz i praćenje paketa.

Pristup Web-u bez ograničenja je podložan potencijalnim zloupotrebama zlonamernih korisnika ili napadu automatizovanim softverom. U ovom radu se razmatra problem kako sprečiti pristup Web sadržajima automatizovanom softveru ako je Web-stranica dizajnirana za pristup ljudi. Potencijalne opasnosti od pristupa automatizovanog softvera Web-stranici dizajniranoj za ljude kreću od bezazlenih do izuzetno opasnih. Neke od opasnosti su prijem i slanje neželjene e-pošte (engl. *spam*), kreiranje lažnih korisničkih naloga, generisanje lažnih Web sadržaja, napadi uskraćivanjem usluge DoS (engl. *Denial of Service*) i slično [2].

Autorizacija korisnika, kao često korišćena tehnika za davanje određenih pristupnih prava, ne rešava problem potencijalne zloupotrebe. Naime, i registrovani korisnik se može ponašati zlonamerno. Problem kod otvorenog pristupa Web-u nisu neautorizovani već automatizovani korisnici – softverski roboti.

Softverski roboti (engl. *bots*) – "botovi" predstavljaju automatizovani softver kojim se pristupni računar maskira kao čovek u cilju interakcije sa Web sadržajima. Neprestana interakcija bota sa Web-stranicom može usporiti njen odziv i beskorisno potrošiti raspoloživu procesorsku snagu servera. Zahtevi pravih korisnika neće biti opsluženi i to dovodi do njihovog nezadovoljstva i napuštanja Web-stranice.

Evidentno je potrebno razviti tehnike za razlikovanje korisnika od automatizovanog softvera kako bi se sprečile potencijalne zloupotrebe. Tjuringov (engl. *Turing*) test je klasičan primer uočavanja razlika u ljudskim i računarskim aktivnostima [3]. Ovaj test je formulisao Alan Tjuring (engl. *Alan Turing*) 1950. godine, a osnovna ideja se sastoji u postavljanju niza istovetnih pitanja i čoveku i mašini. Prema ovom testu, ako ispitivač nije u stanju da razlikuje odgovore čoveka od odgovora mašine, smatraće se da je mašina (softver) dostigla nivo čovečije inteligencije i načina razmišljanja. Iako su originalni testovi izvođeni u formi razmene poruka – četa (engl. *chat*), podrazumeva se da su reči govornog jezika mera inteligencije čoveka. Iako mašine mogu uspešno imitirati ljudski govor, imaju poteškoća oko razumevanja i davanja smislenih odgovora. Zapravo, savremeni Tjuringov test se može shvatiti mnogo šire od puke razmene poruka.

CAPTCHA (engl. *Completely Automated Public Turing test to tell Computers and Humans Apart*) je računarska tehnika kreirana sa ciljem diferencijacije čovečijih od automatizovanih aktivnosti [4]. CAPTCHA je najčešće korišćena tehnika za unapređenje bezbednost Web-stranica od automatizovanog napada. Ovom tehnologijom se sprečava zloupotreba mrežnih usluga, kao što su prijem neželjene e-pošte ili napad softverskim robotom.

Kako bi se Web-stranice odbranile od neželjenog automatizovanog mrežnog napada, neophodno je integrisati tehnologije koje prave jasnu razliku između ponašanja bota i čoveka [5].

II. PREGLED CAPTCHA TEHNOLOGIJA

Postojeće CAPTCHA tehnologije se prema formi postavljanja upita mogu svrstati u četiri osnovne kategorije:

- CAPTCHA bazirana na tekstu [6],
- CAPTCHA na bazi slika [7],
- Audio bazirana CAPTCHA [8],
- Video bazirana CAPTCHA [9].

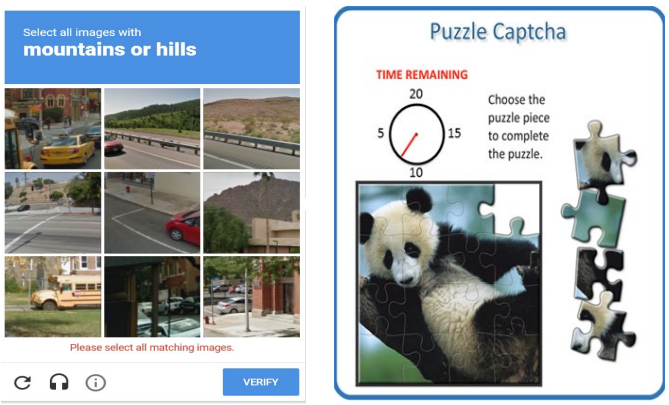
Kod CAPTCHA zasnovane na tekstu se od korisnika zahteva prepoznavanje niza preklopljenih ili zakrivljenih karaktera. Prepoznavanje zakrivljenog teksta se može dodatno otežati dodavanjem različitih pozadina tekstu. CAPTCHA zasnovana na tekstu se do skora najčešće koristila u praksi zahvaljujući njenoj jednostavnosti. Pouzdanost klasifikacije korisnika je bila zadovoljavajuća sve do pojave savremenih rešenja prepoznavanja zasnovanih na veštačkoj inteligenciji AI (engl. A*rtificial Intelligence*). Na sl. 1a) prikazan je primer CAPTCHA-e sa zakrivljenim slovima, dok je na sl. 1b) prikazan primer kombinovan sa dodavanjem pozadine.

Kod CAPTCHA-e baziranoj na slici se traži od korisnika da odabere jednu ili više slika sa specifičnim semantičkim sadržajem. Korisniku se nudi niz slika koje on treba da svrsta prema traženom kriterijumu. Različite varijacije informacija na slikama ne predstavljaju veliki problem za čoveka, dok je to ozbiljan problem za automatizovani softver. Naprednije verzije CAPTCHA-e zasnovane na slikama se kreiraju u formi slagalica ili pronalaženju semantičke oblasti na slici. CAPTCHA-e bazirane na slikama su pouzdanije u odnosu na CAPTCHA-e zasnovane na tekstu. Na sl. 2 prikazane su dve najčešće korišćene CAPTCHA-e zasnovane na slikama a) selekcija slika sa zadatim kontekstom i b) rešavanje slikovne slagalice za određeno vreme. CAPTCHA bazirana na audiju, zahteva prepoznavanje glasovnih sadržaja u deliću zvučnog uzorka. Ova kategorija CAPTCHA-e se često kombinuje sa prethodno pomenutim realizacijama.

CAPTCHA-a na bazi video zapisa traži od korisnika da dovrše logički niz na bazi prikazanog video sadržaja. Jedna verzija video CAPTCHA-e može biti zasnovana na prepoznavanju reklame na koju se odnosi video. Ako korisnik odabere pravu opciju, može se pretpostaviti da je korisnik čovek a ne bot. Ova kategorija CAPTCHA-e je najsloženiji i zahteva mnogo više vremena da korisnici ispravno reše zadatak. Iako je veoma pouzdan, sporost u odlučivanju je razlog zbog čega se retko koristi. Na Sl. 3 prikazana je jedna verzija video CAPTCHA-e gde korisnik treba da u videu pronađe traženu informaciju.

a) b)

Slika 1. CAPTCHA bazirana na tekstu sa a) zakrivljenim tekstom b) dodavanje pozadine zakrivljenom tekstu [7]



a) b)

Slika 2. CAPTCHA-e bazirana na slikama gde korisnik treba da a) selektuje slike sa planinama ili brdima b) reši slikovnu slagalicu za određeno vreme [10].

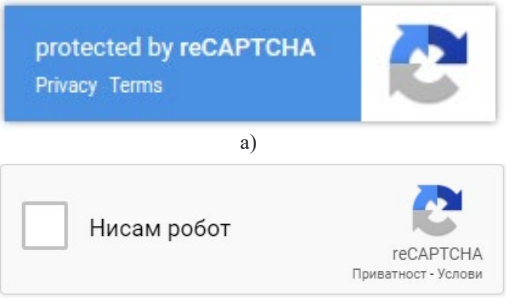
Slika 3. CAPTCHA bazirana na videu gde korisnik treba da u videu pronađe traženi kod [9].

III. GOOGLE WEB API: RECAPTCHA

A. *Izbor verzije i forme reCAPTCHA-e*

Guglova (engl. *Google*) verzija CAPTCHA-e je poznata pod imenom reCAPTCHA [11]. Gugl je ovu tehnologiju realizovao u formi API-a (engl. *Aplication Programming Interface*) kao besplatni servis koji pomaže u zaštiti Web lokacija od neželjene pošte i zlonamernog automatizovanog softvera. Dodavanjem Guglove reCPATHA-a na Web lokaciju može se blokirati automatizovani softver i istovremeno pomoći pravim korisnicima da jednostavno pristupe sadržajima Web-stranice. Usluga reCAPTCHA koju nudi Gugl, može se naći na mnogim popularnim Web-stranicama. Gugl je kreirao nekoliko generacija reCAPTCHA-e, a trenutno se mogu izabrati verzije reCAPTCHA v2 ili reCAPTCHA v3.

Verzija reCAPTCHA v2 ima dve forme. Nevidljiva forma podrazumeva oznaku da je Web-stranica zaštićena reCAPTCHA-om bez potrebe korisnika za ikakvom interakcijom (sl. 4a). Provera se obavlja kada korisnik klikne na postojeće dugme na formi Web-stranice ili se može indirektno pozvati preko JavaScript API poziva. Za integraciju je potreban povratni poziv JavaScripta kada je verifikacija reCAPTCHA-e završena.



a)

b)

Sika 4. Guglovi vidžeti reCAPTCHA-e a) nevidljive forme b) sa poljem za potvrdu.

Samo za najsumnjiviji mrežni saobraćaj će se ponuditi korisniku da reši dodatne zadatke.

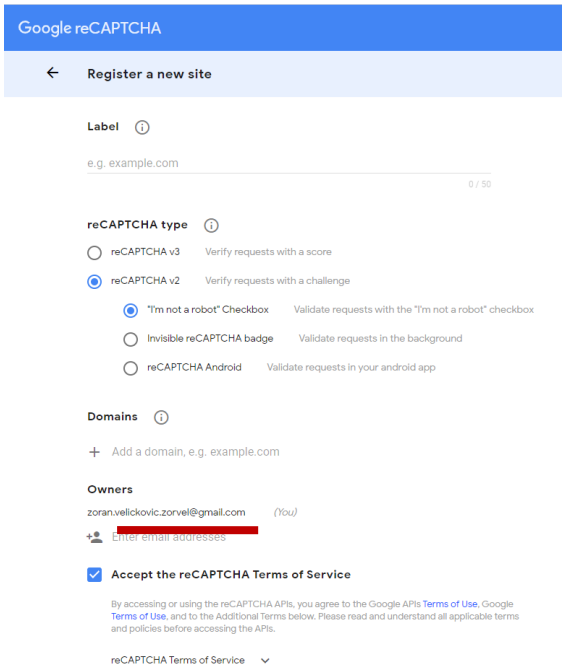
Verzija reCAPTCHA v3 omogućava proveru legitimnosti interakcija bez bilo kakve intervencije korisnika. To je JavaScript API koji vraća ocenu na osnovu koje se mogu planirati dalje aktivnosti na Web-stranici. Na osnovu rezultata, mogu se zahtevati dodatne aktivnosti u procesu autentifikacije korisnika ili ga prihvatiti kao pravog korisnika. Na sl. 4 prikazani su Guglovi vidžeti reCAPTCHA-e za a) nevidljive forme b) forma sa poljem za potvrdu. Kod verzije reCAPTCHA v2 sa poljem za potvrdu, ako korisnik potvrdi polje, biće izazvan da se utvrdi da li se zaista radi o čoveku. Ovo verzija je najjednostavnija za integraciju na Web-stranicu i zahteva samo dve linije HTML koda. Ovim kodom se zapravo postavlja polje za potvrdu na Web-stranici. U ovom radu je za integrisanje provere korisnika korišćena upravo ova verzija.

B. *Registracija servisa*

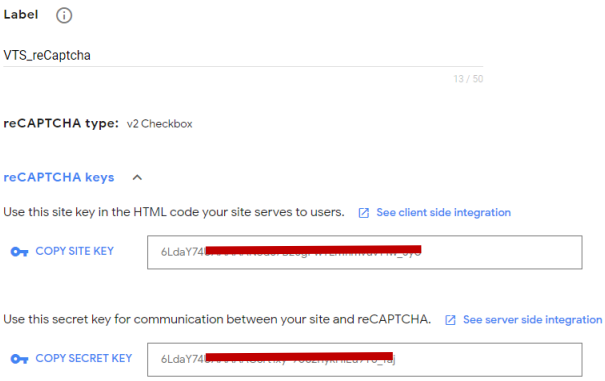
Da bi se mogao koristiti ovaj besplatan Guglov API, neophodno je prethodno obaviti registraciju Web-stranice na kojoj će se koristiti reCAPTCHA-a [11]. Da bi se obavila registracija Web-stranice koja koristi reCAPTCHA-u potrebno je imati važeći Guglov nalog.

Uobičajeni način za korišćenje Guglovih usluga u formi API-a zasnovan je na konceptu API ključeva. API ključ se zahteva za svaku Web lokaciju na kojoj se koristi reCAPTCHA. Prijavljivanje Web-stranice na koju se integriše reCAPTCHA (sl. 5) se obavlja na Guglovoj stranici za programere [11].

Za korišćenje reCAPTCHA-e na Guglovoj stranici se kreira par ključeva: *site key* i *secret key*. „*Site key*“ se koristi pri pozivanju reCAPTCHA usluge sa zaštićene Web-stranice, dok se „*secret key*“ koristi za autorizaciju komunikacije Web aplikacije i Guglovog reCAPTCHA servera kako bi se verifikovao korisnički odgovor. „*Secret key*“, kako mu i samo ime kaže, treba čuvati na bezbednom mestu jer je jedinstven i kreiran samo za registrovanu Web-stranicu. Prvi korak integracije reCAPTCHA-e u Web-stranicu je registracija i selekcija tipa reCAPTCHA-e koji će se koristiti (v2 ili v3). Potom, treba registrovati ovlašćene domene ili naziv projekta. Nakon prihvatanja uslova pod kojima se usluge obavlja, treba kliknuti na dugme *Registation* da bi se generisali jedinstveni API ključevi (sl. 6).



Slika 5. Registraciona stranica i izbor verzije reCAPTCHA-e.



Slika 6. Par API ključeva dobijenih u procesu registracije.

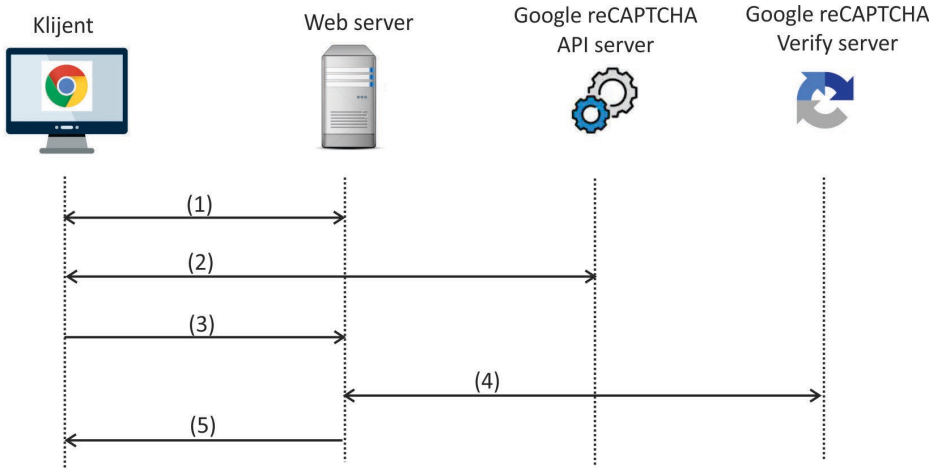
IV. Integracija reCAPTCHA-e na Web-stranicu

Guglova reCAPTCHA tehnologija je bazirana na klijent-server arhitekturi. Zbog toga, integracija ove tehnologije zahteva intervenciju i na klijentskoj i na serverskoj strani Web aplikacije. U nastavku će biti prikazane bazične programske sekvence koje treba integrisati u HTML kod na klijentskom delu aplikacije, odnosno u PHP kod, na serverskom delu aplikacije.

U procesu verifikacije korisnika učestvuju sledeći entiteti: Klijentski računar, Web server – server aplikacije, Google reCAPTCHA API server i Google reCAPTCHA verify server. Proces verifikacije je prikazan na sl. 7, a opisan je u sledećih 5 koraka:

Korak (1): Klijent zahteva Web-stranicu od Web servera i renderuje stranicu sa integrisanom reCAPTCHA-om.

Korak (2): Da bi kompletirao prikaz, Web čitač šalje “*site key*” Google reCAPTCHA API serveru koji odgovara reCAPTCHA slikom i identifikacionim tokenom.



Slika 7. Guglov reCAPTCHA API: Proces verifikacije.

Korak (3): Klijentsko rešenje zadatka reCAPTCHA-e se šalju Web serveru.

Korak (4): Klijentsko rešenje reCAPTCHA-e i "*sicret key*" se dostavljaju Google reCAPTCHA Verify serveru na proveru. Posle provere rešenja, reCAPTCHA Verify server šalje Web serveru aplikacije odgovor u formi JSON-a.

Korak (5): Ako je odgovor korektan, korisniku se dozvoljava pristup ostalim sadržajima na Web-stranici, u suprotnom se šalje na ponovnu proveru.

Na klijenskoj strani se u Web-stranicu integriše reCAPTCHA vidžet sa poljem za potvrdu. Slika vidžeta i odgovarajuće metode JavaScript-a se dobavljaju sa Google reCAPTCHA Web API servera (Korak 2). U ovom radu je HTML forma na klijentskoj strani dizajnirana u Butstrep studiju (engl. *Bootstrep Studio*) [12]. Butstrep studio je razvojno okruženje u kome se kreiraju atraktivne Web-stranice uz podršku kaskadnim opisima stilova i JavaScriptu. Klase potrebne za rad sa reCAPTCHA tehnologijom se učitavaju iz JS biblioteke *api.js* koja je uključena u HTML kod na sledeći način.

```
<script src="https://www.google.com/recaptcha/api.js"
async defer></script>
```

Sama HTML forma se sastoji od niza `<div>` elemenata kojim se definiše struktura stranice. Klasom `g-recaptcha` se integriše reCAPTCHA u HTML kod klijentskog dela aplikacije. Ova oznaka treba da ima važeći atribut `sitekey` dobijen prilikom registracije servisa. Primer koda je dat u nastavku.

```
<div class="g-recaptcha" data-
sitekey="6LdaY74UAAAAANod87B28gPwT******uvT1w_3yU"></div>
```

Odgovor klijenta na postavljeni reCAPTCHA test se potom šalje na serverski deo Web aplikacije koji zatim kontaktira Goole reCAPTCHA Verify server. Serverski deo Web aplikacije je realizovan u programskom jeziku PHP (engl. *Hypertext Preprocessor*). Programska sekvenca za kreiranje upita i analiza odgovora Verify servera je data u nastavku.

```
$response = file_get_contents($url . "?secret=" .
$privatekey . "&response=" . $_POST['g-recaptcha-
response'] . "&remoteip=" . $_SERVER['REMOTE_ADDR']);
```

Upit sadrži neophodne podatke za kontaktiranje Verify servera. Goole reCAPTCHA Verify server svoj odgovor pakuje u formi JSON objekta sa sledećom strukturom.

```
{
   "success": true|false,          // uspešnost
   "challenge_ts": timestamp,      // vremenska oznaka
   "hostname": string,             // ime hosta
   "error-codes": [...]            // opconalni niz
}
```

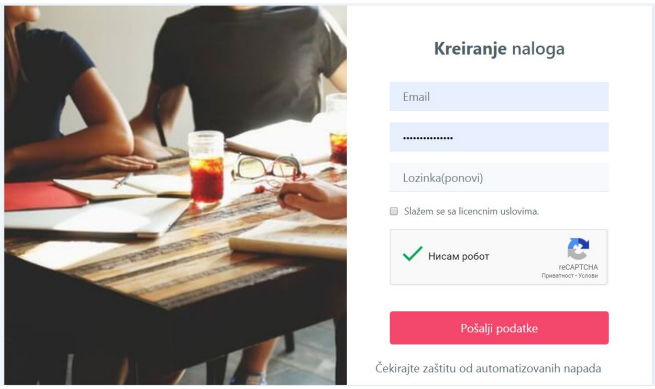
Programski kod na serverskoj strani Web aplikacije analizira odgovor Verify server-a. Ako je procena Guglovog API-a da se radi o automatizovanom softveru, ne šalje se pozitivan odgovor klijentu. U suprotnom, šalje se pozitivan odgovor i dozvoljava se kreiranje korisničkog naloga. U tom slučaju polje „`succes`“ ima vrednost „`true`“.

Iz strukture JSON objekta se vidi da se u odgovoru Verify servera nalaze i dodatne informacije vezane za vreme upućivanja zahteva. Vreme prijema zahteva je formatirano prema ISO-u: `yyyy-MM-dd'T'HH:mm:ssZZ`. Vraćaju se i podaci vezani za ime Web-stranice u formi stringa za koju je verifikacija obavljena. Opciono se može vratiti i niz grešaka. Po potrebi, na osnovu ovih informacija se može reagovati u programskom kodu.

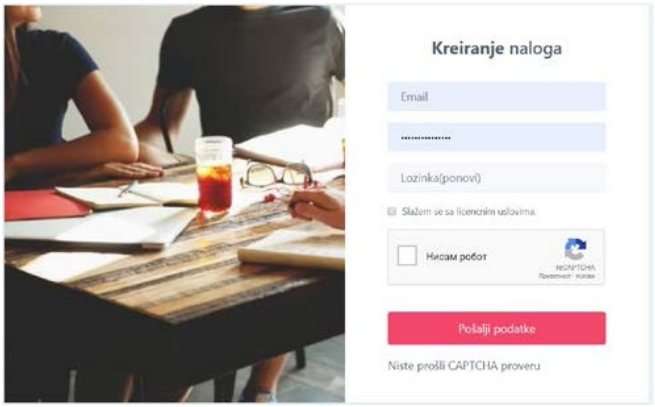
Na sl. 8. su prikazani izgledi Web forme za kreiranje naloga sa integrisanom reCAPTCHA tehnologijom. Pre provere korisnika, potrebno je čekirati polje za potvrdu "Нисма робот" kako bi se inicirao Guglov reCAPTCHA test.

Na sl. 8a) prikazan je slučaj kada je obavljena uspešna provera korisnika. U tom slučaju, podaci sa prijavne forme su poslati Web serveru aplikacije na dalju obradu. Istovremeno je korisniku prikazana poruka da je uspešno prošao Guglov reCAPTCHA test.

Na sl. 8b) prikazan je slučaj kada korisnik nije čekirao polje za potvrdu "Нисма робот", a pokušao je slanje podataka iz prijavne forme Web serveru aplikacije. Google reCAPTCHA Verify server vraća odgovor u kome se konstatuje da reCAPTCHA test nije obavljen i slanje podataka sa forme je onemogućeno. PHP skript Web aplikacije analizira odgovor i ispisuje poruku korisniku da nije prošao Tjuringov test. Takođe, negativan odgovor se dobija ako nije uspostavljena veza sa Google reCAPTCHA veify serverom u određenom



a)



b)

Slika 8. Izgled Web forme za kreiranje naloga sa integrisanom reCAPTCHA tehnologijom kada je a) test uspešno prošao b) provera neuspešna.

vremenskom prozoru. Ovde je implementirana reCAPTCHA verzije 2 koja prati kretanje miša i navigacije na Web-stranici.Kod primenjene verzije reCAPTCHA-a korisnik mora potvrditi polje za potvrdu, a potom se šalje zahtev za proveru Guglovom reCAPTCHA serveru. Testiranjem integrisane tehnologije sve potencijalno kritične situacije su uspešno rešene.

Na osnovu sprovedenih testova može se zaključiti da je Guglova reCAPTCHA tehnologija uspešno klasifikovala korisnike i može se preporučiti za zaštitu Web-stranica od automatizovanih napada.

V. ZAKLJUČAK

Na početku je Internet dizajniran da omogući slobodan pristup Web sadržajima. Pored interakcije Web sadržaja i ljudi, savremeni Web podrazumeva i međusobnu interakciju automatizovanih softvera – botova. Maliciozni botovi mogu izazvati različite probleme u funkcionisanju Web-stranica, a neki od njih su nelegalno slanje i prijem e-pošte, kreiranje lažnih naloga, postavljanje lažnih sadržaja i uskraćivanje usluge DoS napadom. Da bi se zaštitila Web-stranica od napada automatizovanim softverom, u ovom radu se koristi Tjuringov test kojim se uočava razlika između postupaka čoveka i softvera - mašine.

U ovom radu je prikazana integracija besplatne Guglove tehnologije za obavljanje Tjuringovog testa - reCAPTCHA. Za integraciju reCAPTCHA-e na Web-stranici korišćen je Web API razvijen od strane Gugla. U radu je prikazan deo HTML koda kojim se reCAPTCHA sa poljem za potvrdu integriše na Web-stranicu. Takođe, prikazan je deo PHP koda kojim se rezultat obavljenog Tjuringovog testa analizira na serverskoj strani Web-stranice.

Jednostavna integracija Web API-a i pouzdan test reCAPTCHA obezbeđuju sigurnu zaštitu Web-stranice od zlonamernog automatizovanog softvera. Guglova reCAPTCHA tehnologija se preporučuje za korišćenje u visokopouzdanim Web-stranicama. Može se očekivati da će u (bliskoj) budućnosti primenom AI (engl. *Artificial Intelligence*) tehnologija svakako rešiti danas aktuelni Tjuringovi testovi. Kao odgovor na ove izazove očekuje se pojava novih alternativnih CAPTCHA tehnologija [13].

LITERATURA

[1] White paper, "Cisco Visual Networking, Index: Forecast and Trends, 2017–2022", 2019.

[2] P. Larsen, A. Homescu, S. Brunthaler, M. Franz, "SoK: Automated Software Diversity", IEEE Symposium on Security and Privacy, 2014.

[3] https://searchenterpriseai.techtarget.com/definition/Turing-test

[4] http://captcha.net/

[5] N. Roshanbin, J. Miller „A survey and analysis of current captcha approaches", Journal of Web Engineering, Vol. 12, No.1&2, pp. 1 - 40, 2013.

[6] J. Chen, X. Luo, Y. Guo, Y. Zhang, D. Gong, „Survey on Breaking Technique of Text-Based CAPTCHA", Security and Communication Networks, vol. 2017.

[7] https://www.letsnurture.com/blog/8-widely-used-captcha-examples.html

[8] J. Tam, S. Hyde, J. Simsa, L. Von Ahn, "Breaking audio CAPTCHAs" NIPS 2008, pp. 1625 – 1632, 2008.

[9] A. Nadaph, J. Shaikh, N. Bodhe, H. Pingale at all, "Video CAPTCHA – Design Based on Moving Object Recognition", Int. Jour. of Inn. Research in Comp. and Comm. Eng., Vol. 4, Issue 4, April 2016.

[10] T. Ahmed, K. A. Tushar, S. I. Nova, Md. Mahbubur, "RahmanSimple, Robust & User Friendly CAPTCHA 'InstaCap' for Web Security", Int. Jour. of Hybrid Inf. Technology, Vol.9, No.1 (2016), pp. 163-182, http://dx.doi.org/10.14257/ijhit.2016.9.1.15

[11] https://developers.google.com/

[12] https://bootstrapstudio.io/

[13] W3C Working Group, "Alternatives to Visual Turing Tests on the Web", 2019.

ABSTRACT

This paper describes the protection of the Website against malicious automated software attacks. Google's free reCAPTCHA technology to perform Turing tests that detect differences between human and machine behavior has been implemented. A Web API developed by Google was used on the client side to integrate reCAPTCHA. The results of the performed reCAPTCHA test is analyzed by PHP script on the server side of the Web application. Easy-to-integrate and highly reliable test reCAPTCHA technology is recommended for use in highly reliable Web sites for protection against automated software.

INTEGRATION OF THE WEB API reCAPTCHA TECHNOLOGY INTO THE WEBSITE

Zoran Veličković, Zoran Milivojević