Prepoznavanje štetnih web sjedišta u sklopu kataloga www.hr

Valter Vasić, Marin Vuković

Sveučilište u Zagrebu, Fakultet Elektrotehnike i Računarstva

Sažetak – Štetna web sjedišta su sjedišta koja ugrožavaju korisnike i računala koja ih posjećuju. Važno je spriječiti širenje tih sjedišta i odvojiti ih od normalnog Interneta. Stoga u sklopu kataloga www.hr, koji je početno odredište mnogih korisnika Interenta, potrebno je prepoznati i ukloniti štetna web sjedišta. Uz analizu štetnosti u ovom radu pruža se i integrirano rješenje koje bi u implementaciji uklonilo većinu štetnih web sjedišta i sprječilo dodavanje istih u katalog.

1. Uvod

Internet u posljednjih par godina bilježi veliki rast štetnih web sjedišta koje poslužuju sadržaj štetan za posjetitelje. Jednom zaraženo računalo može prouzročiti velike probleme vlasniku računala i također zaraziti druga računala u mreži. Veliki se trud ulaže u suzbijanje takvih napada. Sjedište može izvorno biti štetno ili postati štetno zarazom postojećih normalnih sjedišta. Katalog www.hr postoji već 17 godina te se u njemu nalazi veliki broj zastarjelih web sjedišta i sjedišta koje se više ne koriste. Uz to treće strane mogu namjerno dodavati štetna web sjedišta u katalog kako bi povećale broj zaraženih računala koja bi mogle kasnije iskorištavati.

U ostatku rada dan je osvrt na prirodu problema takvih sjedišta i opasnosti koje ta sjedišta donose. Nakon pregleda opasnosti sagledavaju se mogući načini prepoznavanja tih opasnosti. U 4. poglavlju opisan je sustav koji bi omogućio smanjivanje broja štetnih web sjedišta u www.hr katalogu i smanjio mogućnost unošenja takvih sjedišta u katalog.

2. Štetna web sjedišta

Cilj je prikazati načine provjere i detekcije štetnih web sjedišta na primjeru kataloga www.hr. Osnovna prednost kataloga je baza web sjedišta koja su prijavljena i među njima potrebno je prepoznati koja su štetna. Nema potrebe za naprednim algoritmima koji će pretraživati web slijedeći poveznice na sjedištima.

Podjela štetnih web sjedišta prema uslugama:

- Web sjedišta za dijeljenje ilegalnog sadržaja - Najčešće pružaju mogućnost dohvaćanja *torrent* datoteka pomoću koji se može dohvatiti ilegalni sadržaj.
- Web sjedišta s lažnim bankarskim uslugama - Imitiraju sjedišta koje nude prave bankarske usluge, ali se nalaze na drugom web sjedištu i potiču korisnika

da unese podatke o sebi i svojim karticama kako bi to mogli iskoristiti za krađu sredstava i identiteta.

- Web sjedišta koja nude lažno uklanjanje malicioznog softvera na računalu - Sjedišta se predstavljaju kao grafičko sučelje nekog operacijskog sustava i upozoravaju korisnika da je njegovo računalo zaraženo s nekim lažnim popisom štetnog softvera (*malware)*. Kao pomoć nude dohvaćanje malicioznog softvera nudi lažno uklanjanje štetnog softvera.

2.1 Posljedice pristupa štetnim web sjedištima

Neovisno u koju kategoriju pripada štetno web sjedište ono uvijek pokušava naštetiti korisniku koji se nađe na tom web sjedištu. Ono može naštetiti korisniku na sljedeće načine:

- Pokretanje i/ili instalacija neželjenog softvera na računalu. Takav softver može omogućiti korištenje računala u razne štetne svrhe. Tada računalo postaje dio skupine računala koje napadač može upravljati, *botnet*. Sa računalima u *botnetima* mogu se izvoditi sljedeće radnje [1]:
  - Distributed Denial of Service} (DDOS) napadi na druga računala i mreže u Internetu - To su napadi onemogućavanja pristupa pojedinim računalima ili mrežama slanjem velike količine prometa na odredišta koja se želi onemogućiti.
  - Generiranje i slanje spam e-mail poruka pomoću kojih se može zaraziti druga računala ako su korisnici neoprezni.
  - Prisluškivanje i zabilježavanje prometa koji prolazi mrežom u kojoj se računalo nalazi ili snimanje prometa koje to računalo šalje i prima.
  - Zapisivanje lozinki koje korisnik računala koristi za ovjeravanje na raznim web sjedištima.
  - Širenje štetnog softvera na druga računala koja se nalaze u istoj podmreži.
  - Instalacija raznih reklamnih programa koji će ometati rad na računalu i trošiti Internet promet.
  - Masovna krađa identiteta krađom osobnih podataka korisnika računala, te uz to i krađa podataka o karticama koje služe za plaćanja putem Interneta.
- Krađa lozinke i podataka o kartici imitacijom određene usluge weba (npr. bankarske usluge).
- Pretplata na neželjene reklamne e-mail poruke.

3. Prepoznavanje štetnih web sjedišta

Mogu se razlučiti dva različita pristupa prepoznavanju web sjedišta:

- Pregledavanje koda web sjedišta.
- Korištenje baza podataka koje sadrže podatke o štetnim web sjedištima.

3.1 Pregledavanje koda web sjedišta

Za ostvarivanje ovakvog načina prepoznavanja potrebno je isprogramirati i/ili koristiti softver koji će dohvatiti cjelokupno web sjedište sa slikama i podacima te parsirati dobivene podatke u potrazi za štetnim kodom. Štetni kod se može nalaziti u sljedećim oblicima:

- U sklopu *JavaScript* (JS) koda. Uz to JS kod koji je potencijalno štetan najčešće neće biti zapisan u čitljivom obliku već pretvoren u drugi oblik. Na slici 1 vidi se normalan JS kod, dok se na slici 2 vidi isti kod pretvoren u drugi oblik, a da pritom funkcionalnost koda ostaje ista. Čim je JS kod sakriven sumnjiv je. S druge strane to se radi kako bi se izbjegla krađa i kopiranje JS koda, te prilikom proglašavanja štetnosti treba obratiti dodatnu pozornost.
- Na mjestu gdje se prikazuju reklame može se dohvaćati štetni sadržaj s drugog web sjedišta.
- U sklopu Flash animacija.
- Puno poveznica prema određenom sjedištu može značiti da to sjedište pokušava imitirati drugo web sjedište s namjerom krađe podataka. Takvo sjedište se može detektirati prebrojavanjem referenci na druga web sjedišta, ako postoji veliki broj referenci na drugo web sjedište to može upućivati na to da je sjedište štetno.

Uz to i adresa web sjedišta može odavati puno podataka o sigurnosti web sjedišta kojem se pristupa. Primjerice:

- Web sjedištu: `https://net.bank.hr` može se slobodno pristupati jer je adresa kratka i jasna i ne krije nikakve bespotrebne podatke. Uz to je zaštićena enkripcijom što se vidi iz protokola kojim se pristupa web sjedištu, `https` te se time omogućuje siguran unos povjerljivih podataka na tom web sjedištu.
- Web sjedište: `http://net.bank.hr.account-update.com` izgleda sumnjivo jer ima previše znakova, uz to da se sjedište nalazi u sklopu sasvim druge domene, `account-update.com` umjesto `pbz.hr`. Uz to koristi se

`http` protokol koji ne pruža nikakvu zaštitu podataka te će svi podaci koji se šalju na takvo sjedište biti lako čitljivi za napadača. Prilikom unosa osjetljivih i tajnih podataka na neko web sjedište prije svega mora se paziti da se sjedištu pristupa `https` protokolom, a ne nezaštićenim `http` protokolom. Ukoliko je zaštita prisutna i pravilna Internet pretraživač će to označiti u adresnoj traci.

```
function Cart(){
  this.citems=new Array();
  this.numberOfItems=0;
}
Cart.prototype.addCartItem =(citem){
  this.citems[this.numberOfItems++]
=citem;
  alert(citem.name+"added to cart");
}
```

Slika 1: Običan JavaScript kod [5]

```
function t(){
      this.v=new Array();
      this.c=0;
}
t.prototype.E=function(D){
  this.v[this.c++]=D;
  alert(D.name+"\141\144\144\145\144
  \164\157\143\141\162\164");
}
```

Slika 2: Izmijenjen JavaScript kod [5]

3.2 Korištenje baza podataka

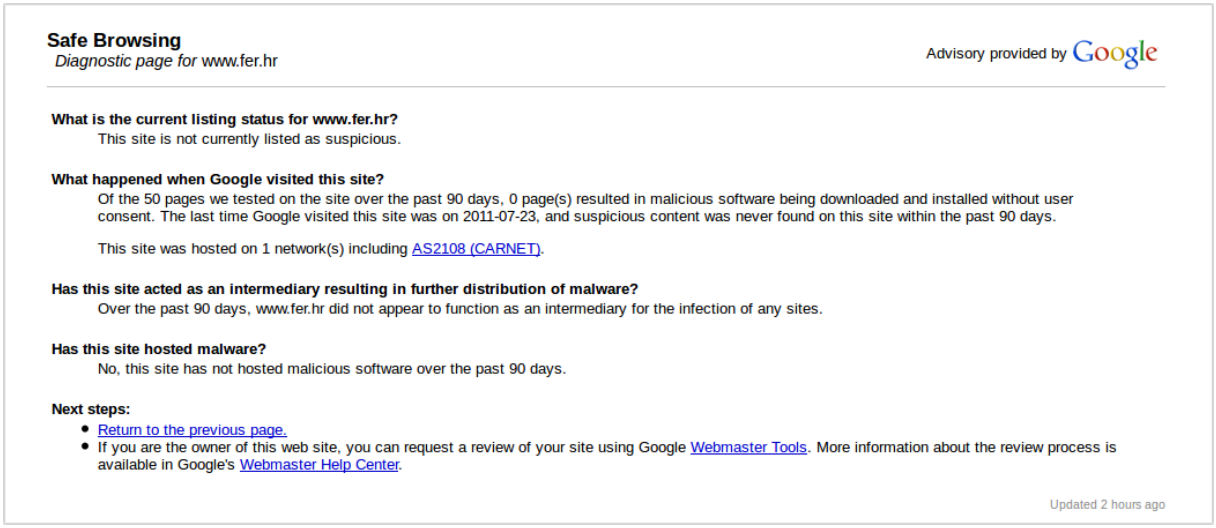
Osim korištenjem programskih rješenja za prepoznavanje štetnosti web sjedišta prilikom detekcije mogu se koristiti servisi određenih web sjedišta koje imaju projekte za suzbijanje štetnih sjedišta. Google nudi servis pomoću kojeg se može detektirati je li određeno sjedište štetno ili nije. Osim Googleovog rješenja postoje razne inicijative koje potiču detekciju i uklanjanje malicioznih web sjedišta i malicioznog softvera kao što je tvrtka StopBadware [4].

Googleov servis se koristi na sljedeći način; ako se želi provjeriti je li sjedište `www.fer.hr` štetno ili zaraženo onda se dohvati sjedište sa sljedeće poveznice:

`www.google.com/safebrowsing/diagnostic?site=www.fer.hr`

Slika 3 prikazuje rezultate za sjedište `www.fer.hr`. Sjedište je organizirano tako da odgovara na 4 osnovna pitanja:

- Je li sjedište trenutno sumnjivo?
- Što se dogodilo kad je Google posljednji put posjetio sjedište?
- Je li sjedište štetno i širi li štetan softver?
- Je li sjedište u posljednje vrijeme sadržavalo štetan softver?



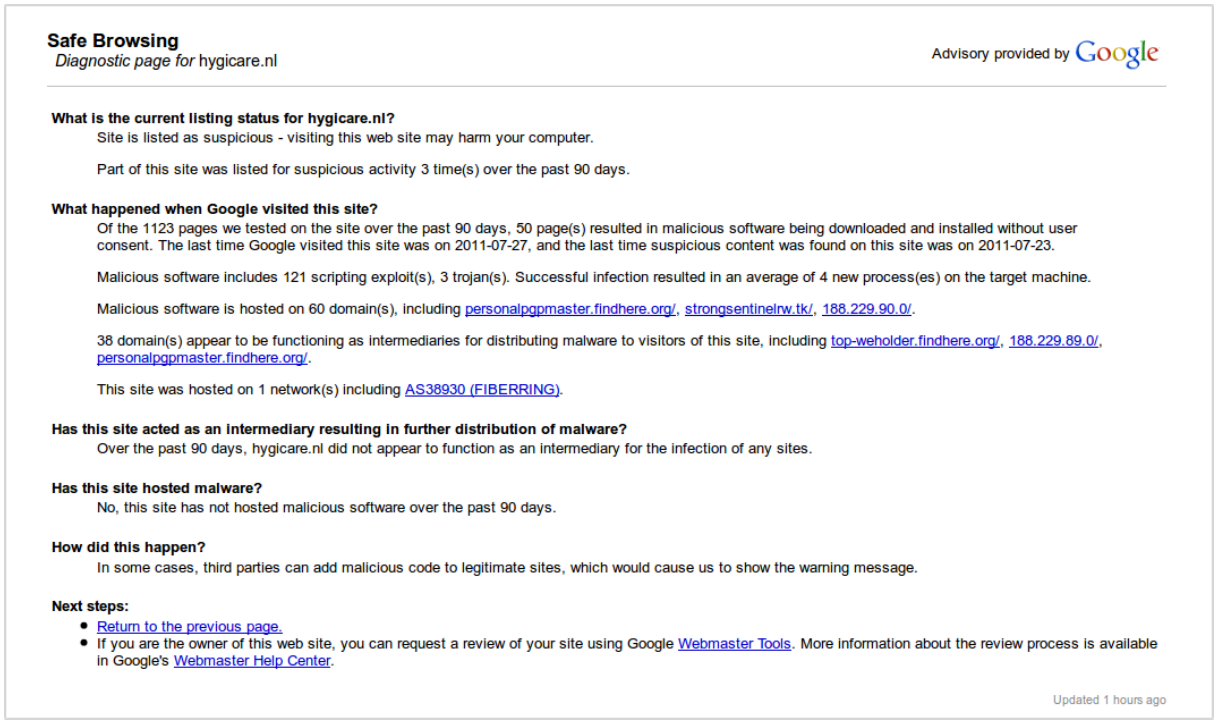
Slika 3: Rezultati upita za www.fer.hr

Na prikazu za `www.fer.hr` na slici 3 vidi se da je stanica sigurna prema Googleovim ocjenama.

Rezultati za sjedište `hygicare.nl`:

`www.google.com/safebrowsing/diagnostic?site=hygicare.nl`

su sasvim drukčiji i prikazani su na slici 4. Sjedište je trenutno prikazana kao sumnjivo i može naštetiti računalu koji ga posjeti. Uz to vidi se koje sve opasnosti sjedište sadrži i gdje se nalazi štetan softver koji je dostupan putem tog sjedišta.



Slika 4: Rezultati upita za hygicare.nl

Kako bi se automatiziralo provjeravanje sigurnosti prikaz sjedišta na slikama 3 i 4 može se parsirati i provjeriti sigurnost sjedišta. Google također nudi alat kojim se može

provjeravati sigurnost sjedišta direktno iz programskih jezika poput Pythona, Google Safe Browsing API [2].

4. Sprečavanje pristupa web sjedištima

Krajnji cilj je prepoznati koja su web sjedišta u sklopu kataloga www.hr štetna i ukloniti ih iz kataloga. Za početnu detekciju najjednostavnije je koristiti se bazom podataka koju nudi Google i pomoću njihove programske podrške provjeravati postojeće unose koje se nalaze u sklopu kataloga. Bilo bi poželjno i uvesti provjeru štetnosti prilikom prijavljivanja novih sjedišta na www.hr.

Nakon početnog postavljanja sustava za provjeru pomoću Googlea, može se raditi na istraživanju i postavljanju sustava koji bi programski prepoznao štetna sjedišta i prijavljivao ih na neki od Internet servisa koje se bave sprječavanjem štetnih sjedišta (npr. StopBadware [4]). Kao smjernice za izradu takvog programa mogu se koristiti razne upute dostupne na Internetu poput OWASP vodiča [3]. Cilj bi bio napraviti integrirani sustav koji bi koristio postojeće podatke o sumnjivosti web sjedišta, provodio aktivno prepoznavanje štetnosti te prijavljivao nova štetna web sjedišta u postojeće sustave.

Nakon što je sjedište identificirano kao štetno ono se mora ukloniti iz kataloga i postaviti u posebnu kategoriju zabranjenih sjedišta. Time se onemogućuje namjerno ponovno postavljanje štetnog sjedišta.

5. Zaključak

Radi vjerodostojnosti kataloga vrlo je važno da su sjedišta koja su sadržana u njemu vjerodostojna i ispravna. Nije prihvatljivo da sjedišta unutar kataloga sadrže ili šire štetan sadržaj jer to čini katalog opasnim za korištenje. Ukoliko je katalog nepogodan za korištenje korisnici će ga zaobilaziti. S tim ciljem provedena je analiza faktora koji utječu na štetnost web sjedišta te prepoznate opasnosti koje web sjedište može pružati. Dan je pregled načina napada i razlučeni su načini borbe protiv njih.

U konačnici preporuča se korištenje postojećih i pouzdanih baza za prepoznavanje štetnih web sjedišta te se uz to predlaže adaptacija postojećih sustava za detekciju uz mogućnost razvoja novih rješenja. Ključno je uvesti provjere štetnosti prilikom prijavljivanja sjedišta u katalog www.hr kako bi se onemogućilo namjerno podmetanje štetnih sjedišta. Korištenjem predloženog sustava uvelike bi se smanjio broj štetnih web sjedišta u sklopu kataloga www.hr.

Popis literature

1. P. Baecher, T. Holz, M. Koetter, and G. Wicherski. *Know your Enemy: Tracking Botnets*. http://www.honeynet.org/papers/bots, 2008. The Honeypot Project.
2. Google Labs. *Google Safe Browsing API*. http://code.google.com/apis/safebrowsing/.
3. M. Meucci, E. Keary, and D. Cuthbert. *OWASP Testing Guide*. http://www.owasp.org/images/5/56/OWASP Testing Guide v3.pdf, Nov. 2008. Open Web Application Security Project (OWASP).
4. StopBadware, Inc. *stopbadware.org*. http://stopbadware.org.
5. Syntropy Developments. *Example using JCE Pro 1.1*. http://www.syntropy.se/doc/jce pro/pages/example.html. Open Web Application Security Project (OWASP).