

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

# **Sybil napadi u društvenim mrežama i zaštita od njih**

*Antun Razum*

Voditelj: *prof. dr. sc. Siniša Srbljić*

Zagreb, svibanj 2016.

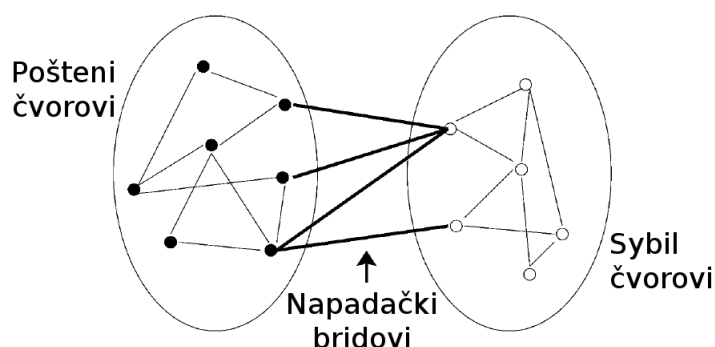
# SADRŽAJ

<b>1. Uvod</b>	<b>1</b>
<b>2. Sybil napadi</b>	<b>2</b>
2.1. Povijest i motivacija . . . . .	2
2.1.1. Središnji autoritet . . . . .	2
2.1.2. Izazovi u decentraliziranim pristupima . . . . .	3
2.2. Pojmovi i definicije . . . . .	3
2.2.1. Model sustava . . . . .	3
2.2.2. Slučajne šetnje i vrijeme miješanja . . . . .	4
<b>3. Obrana od sybil napada</b>	<b>5</b>
3.1. Pretpostavke . . . . .	5
3.2. Algoritam . . . . .	6
3.2.1. Algoritam za identifikaciju sybil čvorova . . . . .	6
<b>4. Zaključak</b>	<b>8</b>
<b>5. Literatura</b>	<b>9</b>

# 1. Uvod

*Sybil* napadi (engl. *Sybil attacks*) su dobro poznata vrsta napada u distribuiranim sustavima poput senzorskih i *peer-to-peer* mreža. U osnovnom obliku ovog napada napadač stori velik broj lažnih identiteta preko kojih utječe na ponašanje napadnutog sustava. Broj identiteta koji napadač može stvoriti ovisi o napadačevim resursima kao što su propusnost mreže, memorija i računarna moć.

## 2. Sybil napadi



Slika 2.1: Napadački bridovi u grafu društvene mreže

### 2.1. Povijest i motivacija

#### 2.1.1. Središnji autoritet

Sybil napadi mogu se lagano kontrolirati pomoću pouzdanog središnjeg autoriteta koji izdaje i provjerava podatke jedinstvene stvarnom čovjeku. Na primjer, ako sustav zahtijeva registraciju korisnika pomoću broja osobne iskaznice ili vozačke dozvole, onda prepreka za sybil napade postaje puno viša. Autoritet također može zahtijevati i plaćanje za registraciju. Nažalost, postoje mnogi slučajevi gdje takvi sustavi nisu poželjni. Na primjer, može biti teško izabrati jedinstveni autoritet kome mogu vjerovati svi korisnici na svijetu. Nadalje, taj središnji autoritet može lako postati točka zatajenja (engl. *single point of failure*) – meta *denial-of-service* napada ili usko grlo u izvođenju sustava. Na poslijetku, plaćanje registracije ili traženje osjetljivih informacija prilikom iste odbilo bi velik broj korisnika. Jedini način da se ovi problemi izbjegnu je distribuiranost sustava za provjeru. [4]

### 2.1.2. Izazovi u decentraliziranim pristupima

Obrana od sybil napada bez puzdanog središnjeg autoriteta puno je teža. Puno decentraliziranih sustava pokušavalo se boriti protiv sybil napada povezivanjem identiteta korisnika s IP adresom. No, napadačima još tada korištenjem određenih metoda nije bio problem ukrasti i iskoristiti veći broj IP adresa s kojih mogu djelovati. [2] Osim krađe IP adresa, napadač može preuzeti veći broj korisničkih računala tako stvarajući *botnet* sačinjen od tisuća računala diljem svijeta. Mnogi drugi decentralizirani načini obrane također su se pokazali neuspješnima. Primjerice zagonetke koje zahtijevaju ljudski napor, kao što je *CAPTCHA*, napadač može upotrijebiti na svojoj stranici tražeći od korisnika njegove stranice da ih riješe. Sve to vodi do zaključka kako se obrana od napada ne može temeljiti na jednokratnim provjerama prilikom registracije korisnika na sustav, već na temeljitoj analizi grafa društvene mreže te ispitivanju na raznih algoritamskih svojstava grafa. [4]

## 2.2. Pojmovi i definicije

### 2.2.1. Model sustava

Društvena mreža promatra se kao neusmjereni beztežinski graf  $G = (V, E)$ . Čvorovi predstavljaju korisnike mreže, a bridovi određenu vrstu odnosa između korisnika. Vrijedi  $|V| = n$ ,  $V = v_1, v_2, \dots, v_n$  te  $|E| = m$ , gdje je  $e_{ij} \in E$  tj.  $v_i \rightarrow v_j$  ako je  $v_i \in V$  susjedan  $v_j \in V$ , za  $1 \leq i, j \leq n$ . Matrica  $A = [a_{ij}]^{n \times n}$  naziva se matrica susjedstva gdje je  $a_{ij} = 1$  ako je  $e_{ij} \in E$  i  $a_{ij} = 0$  ako nije. Matrica  $P = [p_{ij}]^{n \times n}$  naziva se matrica prijelaza (engl. *transition matrix*)

$$p_{ij} = \begin{cases} \frac{1}{\deg(v_i)} & e_{ij} \in E \\ 0 & \text{inače} \end{cases} \quad (2.1)$$

gdje je  $\deg(v_i)$  stupanj čvora  $v_i$ . Skup susjeda čvora  $v_i$  označava se s  $N(v_i)$  tako da je  $\deg(v_i) = |N(v_i)|$ .

U mreži se nalazi određeni broj *poštenih čvorova* (engl. *honest nodes*), svaki sa svojim identitetom, koji predstavljaju stvarne korisnike mreže. Također postoji jedan ili više napadača u mreži, svaki s određenim brojem lažnih identiteta koji su u grafu predstavljeni *sybil čvorovima*. Bridovi između sybil čvorova i poštenih čvorova nazivaju se *napadački bridovi* (engl. *attack edges*). *Sybil regija* (engl. *sybil region*) sastoji se od svih sybil čvorova, a *poštena regija* (engl. *honest region*) sastoji se od svih po-

štenih čvorova. Sve sybil čvorove kontrolira napadač. Stoga, napadač može stvoriti proizvoljan broj bridova unutar sybil regije.

### 2.2.2. Slučajne šetnje i vrijeme miješanja

Prijelaz između dva čvora u grafu može se prikazati markovljevim lancem (ML) koji predstavlja slučajnu šetnju kroz graf  $G$ . *Slučajna šetnja* duljine  $w$  preko  $G$  je niz čvorova u  $G$  koji počinje čvorom  $v_i$  i završava čvorom  $v_t$ . ML je ergodičan (engl. *ergodic*) ako je ireducibilan i aperiodičan, što znači da ima stacionarnu distribuciju  $\pi$  i distribucija poslije slučajne šetnje duljine  $w$  konvergira prema  $\pi$  kada  $w \rightarrow \infty$ . Stacionarna distribucija ML je distribucija koja je invarijantna s obzirom na matricu prijelaza  $P$  tj.  $\pi P = \pi$ . *Vrijeme miješanja* (engl. *mixing time*) ML definira se kao najmanja duljina slučajne šetnje kojom se postiže stacionarna distribucija. Formalno i preciznije, ovu definiciju možemo izreći na sljedeći način:

$$T(\epsilon) = \max_i \min\{t : |\pi - \pi^{(i)} P^t|_1 < \epsilon\} \quad (2.2)$$

gdje je  $\pi$  stacionarna distribucija,  $\pi^{(i)}$  početna distribucija koncentrirana na čvoru  $v_i$ ,  $P^t$  matrica prijelaza nakon  $t$  koraka, a  $|\cdot|_1$  ukupna varijacijska udaljenost, definirana kao  $\frac{1}{2} \sum_j |\pi_j - \pi_j^{(i)}|$ . Za ML kažemo da je *brzo miješajući* (engl. *rapidly mixing, fast mixing*) ako je  $T(\epsilon) = O(\log n)$ . [1]

*Teorem o stacionarnoj distribuciji:* Za neusmjereni beztežinski graf  $G$ , stacionarna distribucija ML preko  $G$  je vektor  $\pi = [\pi_{v_i}]$  gdje je  $\pi_{v_i} = \frac{\deg(v_i)}{2m}$  tj.  $\pi = [\frac{\deg(v_1)}{2m} \frac{\deg(v_2)}{2m} \dots \frac{\deg(v_n)}{2m}]$ .

Uz pomoć prethodnog teorema može se dokazati da je vrijeme miješanja u korelaciji s povezanošću grafa. Dobro povezani grafovi imaju malo vrijeme miješanja, dok slabo povezani grafovi imaju veće vrijeme miješanja. [1]

## 3. Obrana od sybil napada

### 3.1. Pretpostavke

Algoritmi za obranu od sybil napada temelje se na sljedećim pretpostavkama za koje se pokazalo da vrijede u stvarnim društvenim mrežama [3]:

**Poštena regija je brzo miješajuća.** Kao što je definirano u prethodnom poglavlju, to znači da vrijedi  $T(\epsilon) = O(\log n)$ , za funkciju vremena miješanja definiranu prema jednačbi 2.2. Slučajne šetnje u brzo miješajućim grafovima općenito brzo konvergiraju ka stacionarnoj distribuciji. [3] Pokazano je da su stvarne društvene mreže brzo miješajuće. [5]

**Jedan poznat pošten čvor.** U grafu postoji barem jedan poznati pošten čvor. Iz tog čvora kreće potraga za sybil čvorovima.

**Administratoru je poznata topologija društvene mreže.** Ovo povlači tvrdnju da je algoritam centraliziran. Budući da su sve današnje društvene mreže pod centraliziranom kontrolom, može se pretpostaviti da su administratori tih mreža ti koji brinu za sigurnost sustava pa tako i obranu od sybil napada.

**Veličina sybil regije nije usporediva s veličinom poštene regije.** Uz današnje veličine društvenih mreža (npr. Facebook preko 1,6 milijardi), može se pretpostaviti da je napadaču nemoguće stvoriti toliko sybil identiteta. To je tim više otežano uobičajenim postupcima provjere korisnika kao što su potvrda e-mail adrese, unos osobnih informacija te rješavanje CAPTCHA-e.

**Broj napadačkih bridova je ograničen.** Kao posljedica, kada napadač stvori veći broj sybil čvorova, postojat će neproporcionalno mali *rez* (reznici skup) između poštene i sybil regije. Postojanje malog reza narušava svojstvo brzog miješanja: miješanje između poštenih čvorova je brzo, dok je miješanje između poštenih i sybil čvorova sporo.

## 3.2. Algoritam

Cjelokupni algoritam sastoji se od tri komponente: algoritma za identifikaciju sybil čvorova, algoritma za detekciju sybil grupa te dva pristupa ograničavanju broja napadačkih bridova. Tri se komponente mogu koristiti u konjunkciji za što bolju obranu. [3]

### 3.2.1. Algoritam za identifikaciju sybil čvorova

Algoritam za identifikaciju sybil čvorova prima graf društvene mreže  $G(V, E)$ , poznati poštteni čvor  $h$  i *osumnjičeni čvor* (engl. *suspect node*)  $u$  kao ulaz, a vraća je li čvor  $u$  uistinu sybil čvor ili ne. Algoritam se temelji na slučajnim šetnjama. Slučajna šetnja na grafu definirana je nizom prijelaza između čvorova grafa  $G$ . Ako je šetnja u nekom koraku na čvoru  $v_i$ , onda je vjerojatnost da će se šetnja nastaviti preko brida  $v_i \rightarrow v_j$  jednaka  $\frac{1}{deg(v_i)}$  prema jednadžbi 2.1.

Intuicija algoritma je sljedeća: budući da je rez između poštene i sybil regije mali, slučajne šetnje koje polaze od sybil čvorova u pravilu će zapeti u sybil regiji. Također, zbog pretpostavke o neusporedivosti veličine sybil regije i poštene regije, broj čvorova prijeđenih slučajnim šetnjama iz poštenih čvorova bit će veći od broja čvorova prijeđenih šetnjama iz sybil čvorova sve dok su šetnje dovoljno dugačke i dok je broj šetnji dovoljno velik da se ta razlika istakne. Poradi daljnjih objašnjenja, broj prijelaza skupa slučajnih šetnji preko određenog čvora definira se kao *frekvencija* tog čvora. Valja primijetiti da neki čvor može biti prijeđen više puta u jednoj šetnji.

Algoritam za identifikaciju sybil čvorova sastoji se od dvije faze opisane u algoritmima 3.1 i 3.2. Prva faza prima graf  $G$  i  $h$  kao ulaze i vraća pragove koje koristi druga faza kako bi identificirala sybil čvorove. Prva se faza treba pozvati samo jednom za cijeli graf društvene mreže. Kao što je pokazano u algoritmu 3.1, algoritam prvo napravi  $f$  kratkih slučajnih šetnji duljine  $l_s = \log n$  iz poznatog poštenog čvora  $h$ .  $f$  konačnih čvorova ravna se prema stacionarnoj distribuciji poštene regije zbog pretpostavke o brzom miješanju poštene regije. Konačni su čvorovi stoga poštteni s velikom vjerojatnošću. [4] Poslije ovoga poznati poštteni čvor  $h$  i  $f$  konačnih čvorova koriste se kao *čvorovi sudci* od kojih algoritam postavlja kriterije za identifikaciju sybil čvorova. Valja primijetiti da mogućnost pojavljivanja sybil čvorova među čvorovima sudcima ne utječe na učinkovitost algoritma zbog njihovog veoma ograničenog broja. Počevši od najmanje duljine  $l_{min}$  do najveće  $l_{max}$  s razlikama od 100 za svaku duljinu  $l$ , algoritam radi  $R$  (od 1,000 do 2,000) slučajnih šetnji polazeći od svakog čvora sudca.



Pritom broji čvorove čija frekvencija nije manja od praga  $t$  (koji mala konstatna, npr. 5). Algoritam pronade  $f+1$  takvu vrijednost za svaku duljinu  $l$ . Zatim izračuna srednju vrijednost i standardnu devijaciju tih vrijednosti i ispiše trojku  $(l, srVr, stdDev)$ .

---

**Algoritam 3.1:** Računanje pragova

---

```

1 function pretprocesiranje( $G, h$ ) begin
2    $J = \{h\};$ 
3   for  $i = 1$  to  $f$  do
4     Napravi slučajnu šetnju duljine  $l_s = \log n$  od čvora  $h$ ;
5      $J = J \cup \{\text{posljenji čvor slučajne šetnje}\};$ 
6    $l = l_{min};$ 
7   while  $l \leq l_{max}$  do
8     forall  $i \in J$  do
9       Napravi  $R$  slučajnih šetnji duljine  $l$  iz čvora  $i$ ;
10       $n_i = \text{broj čvorova s frekvencijom manjom od } t;$ 
11       $\text{ispis}(l, srVr(\{n_i : i \in J\}), stdDev(\{n_i : i \in J\}));$ 
12       $l = l + 100;$ 

```

---

Kao što je pokazano u algoritmu 3.2, kako bi se odredilo je li čvor  $u$  sybil čvor, algoritam prvo napravi  $R$  slučajnih šetnji početne duljine  $l = l_0$  iz čvora  $u$ .  $l_0$  je veći ili jednak duljini  $l_{min}$  korištenoj u algoritmu 3.1. Algoritam zatim uspoređuje broj čvorova čija frekvencija nije manja od  $t$  sa srednjom vrijednosti  $srVr$  iz trojke  $(l, srVr, stdDev)$  ispisane u algoritmu 3.1. Ako razlika broja prebrojenih čvorova i srednje vrijednosti veća od  $stdDev \cdot \alpha$  ( $\alpha = 20$ ), čvor  $u$  smatra se sybil čvorom i algoritam završava. Inače, algoritam udvostručuje  $l$  i ponavlja proces sve dok  $l$  nije veća od  $l_{max}$ . Ako tada  $u$  i dalje nije identificiran kao sybil čvor, on se tada smatra poštenim i algoritam se završava.

---

**Algoritam 3.2:** Identifikacija sybil čvorova

---

## **4. Zaključak**

## 5. Literatura

- [1] A. Mohaisen, N. Hopper, i Y. Kim. Keep your friends close: Incorporating trust into social network-based sybil defenses. U *INFOCOM, 2011 Proceedings IEEE*, stranice 1943–1951, April 2011. doi: 10.1109/INFCOM.2011.5934998.
- [2] Anirudh Ramachandran i Nick Feamster. Understanding the network-level behavior of spammers. *SIGCOMM Comput. Commun. Rev.*, 36(4):291–302, Kolovoz 2006. ISSN 0146-4833. doi: 10.1145/1151659.1159947. URL <http://doi.acm.org/10.1145/1151659.1159947>.
- [3] W. Wei, F. Xu, C. C. Tan, i Q. Li. Sybildefender: A defense mechanism for sybil attacks in large social networks. *IEEE Transactions on Parallel and Distributed Systems*, 24(12):2492–2502, Dec 2013. ISSN 1045-9219. doi: 10.1109/TPDS.2013.9.
- [4] H. Yu, M. Kaminsky, P. B. Gibbons, i A. D. Flaxman. Sybilguard: Defending against sybil attacks via social networks. *IEEE/ACM Transactions on Networking*, 16(3):576–589, June 2008. ISSN 1063-6692. doi: 10.1109/TNET.2008.923723.
- [5] H. Yu, P. B. Gibbons, M. Kaminsky, i F. Xiao. Sybillimit: A near-optimal social network defense against sybil attacks. *IEEE/ACM Transactions on Networking*, 18(3):885–898, June 2010. ISSN 1063-6692. doi: 10.1109/TNET.2009.2034047.

## **Sybil napadi u društvenim mrežama i zaštita od njih**

### **Sažetak**

Sybil napad je napad kojim se pokušava srušiti sustav reputacije stvaranjem lažnih identiteta u peer-to-peer mrežama koji djeluju na sličan način. Društvene mreže su vrlo česta meta sybil napada. Zaštita od sybil napada na društvenim mrežama temelji se na algoritamskim svojstvima grafova društvenih mreža putem kojih se računa razina povjerenja koja se može pridijeliti proizvoljnom čvoru grafa.

**Ključne riječi:** društvene mreže, sybil napad, teorija grafova, sigurnost podataka, identitet