IDENTIFYING SKYPE TRAFFIC USING NETFLOW DATA

Patrik Šebeň
Master Degree Programme (2), FIT BUT
E-mail: xseben00@stud.fit.vutbr.cz

Supervised by: Matěj Grégr
E-mail: igregr@fit.vutbr.cz

Abstract: Netflow data are mainly used for statistics, but can also be used to identify the Skype client communication. This paper deals with patterns in the Skype protocol and how to find these patterns in NetFlow data. Described patterns can lead to identification nodes and supernodes in the Skype network.

Keywords: skype, netflow, supernode, udp probe, ipflow

1 ÚVOD

Existuje niekoľko metód zaoberajúcich sa identifikáciou Skype komunikácie. Väčšina z nich sa snaží na základe doterajších znalostí o Skype protokole skúmať obsah zachytených paketov a identifikovať ho. Tento prístup je však efektívne realizovateľný iba v menších sieťach.

Iný a pomerne nový prístup sa snaží identifikovať účastníkov zo zachytených NetFlow dát. Keďže množstvo získaných informácií je obmedzené (chýba obsah paketu), je tento prístup menej presný, avšak realizovateľný aj v sieťach väčších rozmerov.

2 ARCHITEKTÚRA SKYPE SIETE

Hlavnou výhodou Skype siete je jej architektúra peer-to-peer, ktorá na rozdiel od iných alternatív umožnuje minimalizovať náklady na jej fungovanie. Skype sieť pozostáva z nasledujúcich prvkov [2]:

- *Skype nóda* je každý bežný užívateľ, u ktorého je spustená aplikácia a je pripojený do siete.
- *Skype login server* je jeden z mála centralizovaných prvkov siete Skype, ktorý má na starosti prihlasovanie užívateľov do siete.
- *Skype supernóda* je obyčajná Skype nóda povýšená do stavu supernódy po splnení určitých predpokladov (dostupný výpočetný výkon, rýchlosť linky, dlhšia doba bežiacej aplikácie bez prerušenia, verejná IP adresa). V tomto režime môže užívateľom, ktorí nie su schopní nadviazať navzájom spojenie (NAT, firewall), slúžiť ako sprostredkovateľ komunikácie. Taktiež môže pracovať ako login server a uľahčiť tým prácu centralizovaným prvkom siete. Skype ma niekoľko vlastných supernód, ktoré sú využívané pri novej inštalácií klienta a registrácií užívateľa do siete.
- *Skype gateway* je hraničný bod siete Skype, ktorý smeruje hovory do klasickej telefónnej siete.

3 METÓDY DETEKCIE ÚČASTNÍKOV SKYPE SIETE POMOCOU DÁT NETFLOW

Sledovaním NetFlow dát bolo objavených niekoľko podobností vo veľkosti a počte prenášaných signalizačných paketov [1], z ktorých je možné detekovať jednotlivých účastníkov Skype komunikácie. Zaujímajú nás hlavne dva druhy komunikácie, čo je UDP probe a TCP handshake.

3.1 UDP PROBE

Jedná sa o UDP komunikáciu medzi klientom a supernódou po spustení aplikácie pri prihlasovaní do siete. Rozlišujeme dve varianty UDP probe, a to krátku a dlhú. Krátka používa dve správy, dotaz a odpoveď, zatiaľ čo dlhá má správ viac. Pri spustení klienta je jednou z variant kontaktovaných niekoľko supernód zo zoznamu klienta. Nie je zrejmé, podľa čoho sa varianta vyberá. Pri oboch typoch nás veľkosť poslednej správy od supernódy informuje o výsledku. Pokiaľ je správa o veľkosti 18 bajtov, znamená to, že nás daná nóda akceptuje a môžeme pristúpiť k nadviazaniu TCP spojenia. V prípade, že je veľkost 26, 51 alebo 53 bajtov, chápeme túto odpoveď ako odmietnutie a daná nóda už nebude ďalej kontaktovaná. Krátka UDP probe je odosielaná aj po nadviazaní TCP spojenia, aby sa klient uistil, že má stále dostupné supernódy v prípade zlyhania aktuálnej. Vzhľadom na počet a charakter správ, je táto komunikácia ideálna na detekciu klientov a supernód.

3.2 TCP HANDSHAKE

Po výbere supernódy môžeme pristúpiť k nadviazaniu TCP spojenia na celú reláciu alebo do zlyhania supernódy. V prípade neúspechu komunikácie na porte, ktorý máme uložený v lokálnom zázname, skúša klient komunikovať na porte 80, prípadne 443. Pokiaľ sa nepodarí nadviazať spojenie ani na týchto portoch, je zahájená komunikácia s inou supernódou, ktorá nám kladne reagovala na UDP probe. Pri tejto komunikácií už nemáme tak presnú informáciu o veľkosti paketov ako v predchádzajúcom prípade. Pri komunikácií na rozdielnych portoch sú isté odlišnosti medzi veľkosťou a počtom správ. Avšak všetky pakety majú príznak TCP PSH (PSH=1) a posledné správy majú zhodnú veľkosť nezávisle na použitom porte.

3.3 DETEKČNÝ ALGORITMUS

Vyššie spomenuté metódy detekcie supernódy je možné aplikovať nad NetFlow dátami. Vychádzame z prebraného algoritmu [3], ktorý bol testovaný nad dátami IPFIX získaných zo siete SWITCH (The Swiss Education and Research Network). Vzhľadom na oneskorenie pri získavaní záznamov (až 4 hodiny) a detekcií supernód, nebolo možné presnejšie overiť, či sa skutočne jednalo o supernódu (zariadenie mohlo byť už vypnuté). Tomuto sa pri použití NetFlow dát z univerzitnej siete VUT snažíme vyhnúť.

Algoritmus (viz Algoritmus 1) funguje nasledovne :

- Vyberie záznam f z NetFlow dát (flowstream).
- Ak je to UDP komunikácia, algoritmus sa snaží detekovať krátke a dlhé UDP probe (hodnoty 46, 85 boli získané na základe veľkosti IP a UDP hlavičiek a samotnej probe).
  - V prípade úspechu sú IP adresy (sn - supernóda, cl - klient) a port (q, 80, 443) uložené do zoznamu potenciálnych klientov Skype siete (acklist).
- Ak je to TCP komunikácia, algoritmus sa snaží detekovať TCP handshake na základe počtu správ v toku a prítomného PSH príznaku
  - V prípade nájdenia odpovedajúceho toku je porovnavaná adresa zdroja, cieľa a port so zoznamom potenciálnych klientov a pri zhode sme detekovali klienta a supernódu.

Algoritmus 1: detekcia Skype nód a supernód

```
acklist ← 0, supernodes ← 0, clients ← 0
for all f ∈ flowstream do
  if protocol(f) = UDP and port_dest ≥ 1024 then
    if <packets(f), bytes(f)> ∈ {<1,46>,<2,85>} then
      sn ← address_source(f), q ← port_source(f), cl ← address_dest(f)
      acklist ← acklist ∪ <sn,{q,80,443},cl>
    end if
  else
    if protocol(f) = TCP and packets(f) ≥ 3 and PSH ∈ flags(f) then
      sn ← address_dest(f), q ← port_dest(f), cl ← address_source(f)
      if <sn,sp,cl> ∈ acklist then
        supernodes ← supernodes ∪ sn
        clients ← clients ∪ cl
      end if
    end if
  end if
end for
```

Analýzou algoritmu d'alej vidíme, že sme pri UDP komunikácii vylúčili porty nižšie ako 1024, aby nedošlo k nesprávnemu ohodnoteniu klienta ako užívateľa Skype siete, keďže na nižších portoch komunikuje napríklad DNS protokol, ktorého dotaz by mohol byť zle vyhodnotený. Záznamy v ackliste je potrebné pravidelne obnovovať a po určitej dobe odstraňovať staré. Algoritmus má však aj svoje obmedzenia. Flow data by nemali byť vzorkované, keďže k úspešnej detekcii potrebujeme presné poradie, počet a veľkosť správ.

4 ZÁVER

Tento príspevok sa čitateľovi snažil priblížiť fungovanie Skype protokolu a metódu detekcie účastníkov siete na základe NetFlow dát. Bolo popísaných niekoľko základných operácií, ktoré je možné detekovať v zachytených dátach, a identifikovať tak jednotlivé nódy a supernódy. Cieľom práce je implementovať, vhodne upraviť prezentovaný algoritmus pre prostredie pateřní sítě VUT a získať tak informácie o počte aktívnych Skype supernód a nód a taktiež overiť jeho presnosť detekcie. Aktuálne sa nám podarilo v hodinovom zázname o veľkosti približne 2 GB detekovať supernódy a pokročili sme do fázy validácie výsledkov a optimalizácie algoritmu.

REFERENCE

[1] Adami, D.; Callegari, C.; Giordano, S.; aj.: A Real-Time Algorithm for Skype Traffic Detection and Classification. In *Smart Spaces and Next Generation Wired/Wireless Networking*, *Lecture Notes in Computer Science*, ročník 5764, editace S. Balandin; D. Moltchanov; Y. Koucheryavy, Springer Berlin / Heidelberg, 2009, ISBN 978-3-642-04188-4, s. 168–179, 10.1007/978-3-642-04190-7_16.

[2] Baset, S. A.; Schulzrinne, H. G.: An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, april 2006, ISSN 0743-166X, s. 1 –11.

[3] Trammell, B.; Boschi, E.; Procissi, G.; aj.: Identifying Skype Traffic in a Large-Scale Flow Data Repository. In *TMA*, *Lecture Notes in Computer Science*, ročník 6613, editace J. Domingo-Pascual; Y. Shavitt; S. Uhlig, Springer, 2011, ISBN 978-3-642-20304-6, s. 72–85.