Vysoké učení technické v Brně
Fakulta informačních technologií

Projekt Tarzan

Analýza mobilní komunikace

1. září 2017 Jan Havlík (xhavli43@stud.fit.vutbr.cz)

1 Úvod

Cílem této technické zprávy je popsat nástroj, kterým lze detekovat mobilní komunikaci na lokální síti. Nástroj používá offline i online mód pro detekci toků, které generují mobilní zařízení. Detekce zařízení je prováděna pomocí informací z konfiguračnho souboru, který se dá dle potřeby rozšířit. Součástí nástroje je také výpis statistických informací ke každému zařízení do formátu CSV, který se dá následně zpracovat.

2 Prostředí

Z důvodu kompatibility byla zvolena verze Pythonu `3.5.4` a ve skriptu je použita externí knihovna `scapy` [1].
Skript je plně funkční na serveru `merlin.fit.vutbr.cz`.

2.1 Použití

Skript nabízí režim online a offline analýzy:

- **Online analýza**
  `python3 mal.py -f <filename>`, kde
  `<filename>` je cesta k souboru ve formátu `pcap` či `pcapng`.

- **Offline analýza**
  `python3 mal.py -l`

  Pro online analýzu je potřeba mít práva uživatele `root`.

Pro správnou funkčnost programu je nutné, aby adresář obsahoval také **konfigurační soubor** `config.json` ve formátu `json`[2].
Formát tohoto souboru je popsán v sekci 4.

3 Způsoby detekce

Zde jsou popsány metody, které používá skript pro detekci toku mobilních dat. Obecně se jedná o metody použitelné na lokální síti, ale některé, například `User-Agent` (3.1) lze použít i pro veřejné sítě.

3.1 User-Agent

Pole `User-Agent` se nachází v nekterých paketech zapouzdřených v protokolu `HTTP`. Jedná se o řetězec, který může obsahovat informace o typu aplikace, operačním systému, výrobci nebo modelu daného zařízení, které používá `HTTP` protokol. Postupně rozšiřujeme údaje týkající se detekovaného zařízení o nové řetězce `User-Agent`, které obsahují nové informace a obsahují klíčová slova z konfiguračního souboru.

[1] `http://www.secdev.org/projects/scapy/`
[2] `https://cs.wikipedia.org/wiki/JavaScript_Object_Notation`

Příklad řetězce User-Agent v HTTP protokolu:

```
Mozilla/5.0 (Linux; Android 6.0.1; SM-A510F Build/MMB29K) AppleWebKit/537.36 (KHTML,
                like Gecko) Chrome/58.0.3029.83 Mobile Safari/537.36
```

Zde vidíme, že se jedná o operační systém Android verze 6.0.1 a model telefonu je Android A5. Do konfiguračního souboru jsou pak přidány záznamy o operačním systému nebo typu zařízení, které chceme detekovat.

3.2 DNS query

Další možností je hledat v DNS dotazech na záznamy typu A (případně AAAA) na specifická doménová jména. Různé mobilní operační systémy a jejich verze používájí různá doménová jména pro Captive portál[1], kam se mobilní zařízení připojují při volbě připojení pomocí Wi-Fi.

Příkladem takového doménového jména může být connectivitycheck.gstatic.com, což je adresa Captive portálu pro verze Android 6.0.0 a vyšší.

DNS záznamy také lze použít pro rezoluci IP adres, se kterými dané mobilní zařízení komunikuje. Toho můžeme využít například pro výpis statistik z komunikace, kde vypíšeme koncové body komunikace ve formě URL.

3.3 Organizationally Unique Identifier

Organizationally unique identifier, zkráceně OUI je 24-bitové číslo, které identifikuje výrobce zařízení. Je to prvních 24 bitů z MAC adresy, která slouží k adresování na linkové vrstvě. Pro stejný název výrobce může však existovat více OUI záznamů, je tedy potřeba ověřit, zda se jedná o OUI patřící mobilním zařízením daného výrobce.

Pro ověření příslušnosti OUI k mobilnímu zařízení jsem použil databázi Fingerbank (https://fingerbank.org/), která obsahuje informace o digitální stopě zařízení.

3.4 DHCP Host Name

Jestliže nedetekujeme žádné toky mobilních dat, můžeme také použít DHCP Host Name, které se nachází v DHCP requestu. Tato položka nám může říci, že se jedná o mobilní zařízení, nevíme však nic jiného. Je pak třeba sledovat daný tok a případné informce doplnit. Doplnění informací pak probíhá pomocí filtrace podle MAC adresy daného zařízení.

[1] https://en.wikipedia.org/wiki/Captive_portal

4 Konfigurační soubor

Konfigurační soubor ve formátu `JSON`, ze kterého skript čerpá informace pro detekci mobilních zařízení se skládá z jednotlivých složek popsaných v sekci 3. Konfigurační soubor lze nadále rozšiřovat podle znalostní báze uživatele (například nové mobilní `OUI` nebo změna url pro `Captive portal`). Záznamy v konfiguračním souboru pro jednotlivá zařízení vypadají následovně:

- **User-Agent**

```
"userAgent":[
  {
    "str":[
      "Android",
      "iPad",
      "iPhone"
    ]
  }
]
```

- **DNS query**

```
"dns":[
  {
    "qry":"connectivitycheck.gstatic.com.",
    "os": "Android",
    "version": "~6.0.0"
  }
]
```

- **Organizationally Unique Identifier**

```
"oui":[
  {
    "mac":"70:28:8b:ff:ff:ff",
    "os":"Android",
    "version":"?",
    "type":"Mobile"
  }
]
```

- **DHCP Host Name**

```
"dhcpHostName":[
  {
    "str":[
      "android",
      "iPad",
      "iPhone"
    ]
  }
]
```

5 Zhodnocení a testování

Pro testování jsem vytvořil dataset, který obsahuje komunikaci více typů zařízení s různými operačními systémy a jejich verzemi. Dataset byl vytvořen pomocí programu Wireshark při zachytávání paketů mobilních zařízení, která byla připojena na hotspot na stanici, která obsahovala program Wireshark.
Dataset se skládá z následujících souborů:

Název souboru	**Zařízení**	**IPv4 adresa**	**Typ**	**OS**	**Verze**
a5_android_mob.pcapng	Android A5	10.42.0.156	Mobilní	Android	6.0.0
huawei_android_tab.pcapng	Huawei MediaPad T1	10.42.0.230	Tablet	Android	4.4.4
ipad_ios_tab.pcapng	iPad	10.42.0.11	Tablet	OS X	Neznámá
iphone_ios_mob.pcapng	iPhone	10.42.0.24	Mobilní	OS X	Neznámá

Tabulka 1: Seznam souborů, ze kterých se skládá dataset

U souborů `a5_android_mob.pcapng` a `huawei_android_tab.pcapng` se jedná o zachycení šifrované komunikace se službami Facebook a YouTube. Zde se objevují různé protokoly z aplikační vrstvy jako je například `QUIC` (jedná se o multiplexovaný UDP stream od společnosti Google [1]).
Zbylé soubory obsahují komunikaci více zařízení naráz a protokoly z aplikační vrstvy specifické pro zařízení s operačním systémem OS X (například protokol RTP naslouchající na portu číslo 16385, který je používaný službou Apple iChat pro přenos audia a videa [2])

Nástroj je schopen detekovat mobilní zařízení podle parametrů zmíněných v sekci 3. Na standardní výstup vypisuje relevatní informace k detekovaným zařízením a společně s nimi vypisuje také statistické informace. Tyto statistické informace se dělí do tří pro nás důležitých kategorií:

- **S kým zařízení komunikuje (v obou směrech)**
- **Jaké protokoly při tom používá**
- **Kolik dat při komunikaci zařízení přeneslo a získalo**

Stejné informace jsou pak zaneseny do souboru v `csv` formátu pro další možné zpracování. Příklad výstupu skriptu může být následující:

```
New mobile device detected!
********************
Operating system: OS X
Type of device: Tablet
Device addressing: 10.42.0.11 (64:20:0c:0f:66:d8)
User-Agent: Mozilla/5.0 (iPad; CPU OS 9_3_5 like Mac OS X) AppleWebKit/601.1.46
Accept-Language: cs-cz
DHCP Host Name: iPad
```

Pokud se výpis položek `User-Agent` nebo `Accept-Language` opakuje, je to z toho důvodu, že skript obejvil nové informace, které doposud o daném zařízení neměl a rozšířil tak jednotlivé

[1] `https://www.chromium.org/quic`
[2] `https://en.wikipedia.org/wiki/List_of_TCP_and_UDP_port_numbers`

záznamy.
Výpis statistik k danému zařízení na standardní výstup může vypadat například takto:

```
*** DEVICE COMMUNICATION (Endpoint to Device) ***
10.42.0.1 -> 10.42.0.11 (10687 bytes)
        Protocol: DNS (53), Data: 10687 bytes
17.253.57.205 -> 10.42.0.11 (1215 bytes)
        Protocol: HTTP (80), Data: 1215 bytes
104.125.59.189 -> 10.42.0.11 (9802 bytes)
        Protocol: TLS/SSL/QUIC (443), Data: 7354 bytes
        Protocol: HTTP (80), Data: 2448 bytes

*** DEVICE COMMUNICATION (Device to Endpoint) ***
10.42.0.11 -> 17.248.147.21 (4248 bytes)
        Protocol: TLS/SSL/QUIC (443), Data: 4248 bytes
10.42.0.11 -> 77.75.79.33 (10363 bytes)
        Protocol: TLS/SSL/QUIC (443), Data: 6765 bytes
        Protocol: HTTP (80), Data: 3598 bytes
10.42.0.11 -> 185.175.85.39 (5980 bytes)
        Protocol: HTTP (80), Data: 5980 bytes
```

6 Závěr

Nástroj slouží pro detekci mobilní komunikace na lokální síti buď ze souboru, nebo pomocí živého zachytávání paketů. Pro správnou detekci zařízení je nutná přítomnost konfiguračního souboru ve formátu `JSON`, který je nadále rozšiřitelný v případě dostupnosti nových informací. Po detekci daného zařízení jsou informace o tomto zařízení vypsány na standardní výstup (`stdout`) a zároveň zaneseny do souboru ve formátu `CSV` pro lepší automatické zpracování. Společně s informacemi o zařízení se vypisují statistické informace o tom, s jakými adresami zařízení komunikovalo, kolik dat přeneslo, či přijalo a jaké při tom použilo protokoly. U některých zařízení se ve výpisu objevily různé IP adresy pro stejnou MAC adresu. Příčinou této anomálie je pravděpodobně způsob vytváření datasetů, kdy pro zachycení paketů mobilního zařízení byl vytvořen hotspot na stanici, která obsahovala program Wireshark a na tento hotspot se mobilní zařízení připojila a následně byla zachytávána jejich data.