Kaj se dogaja z našimi podatki?

Kaj se dogaja z našimi podatki?

Jure Brenčič Jazbec,
Timotej Gregorič

5. letnik

Peter Krebelj, mag. posl. ved

Računalništvo in informatika

Raziskovalna naloga

2019/2020

Srednja šola tehniških strok Šiška

Povzetek

Splošno znano je, da dandanes človek ni nikoli več sam. Na vsakem koraku ga spremljajo sateliti, najrazličnejši strežniki in seveda ljudje, ki stojijo za vsem tem. V nalogi je opisano raziskovanje. Ali je to, da smo ljudje pod nadzorom kaj počnemo na mobilnih napravah 24 ur na dan, 7 dni v tednu, resnično ali ne.

Zelo veliko virov pravi, da je ta trditev resnična, a mnogi temu močno nasprotujejo. Pravi odgovor bomo pridobili na samo en način. S preizkusom.

Ključne besede: podatki, pametni telefon, Andorid, Huawei

Vsebina

Povzetek 3
Kazalo slik 5
1. Uvod 6
2. Teoretičen del 6
2.1. Zbiranje informacij 6
2.2. Postavljene hipoteze 7
2.3. Raziskave o pametnih telefonih 7
3. Eksperimentalni del 8
5.1. Priprava naprav 8
5.2. Izbira programske opreme 9
5.2.1. Wireshark 9
5.2.2. PfSense 10
5.2.3. TCPDump 11
5.3 Rezultati testiranja 12
5.3.1 Stanje naprave v mirovanju in aktivnem delovanju 14
4. Razprava 18
5. Zaključek 18
6. Zahvale 19
Viri 19

Kazalo slik

Slika 1: Zajeti podatki v programu Wireshark 10
Slika 2: Spletni vmesnik PfSense 11
Slika 3: Primer izvoženih podatkov 13
Slika 4: Število paketov/10min v odvisnosti z uro 15
Slika 5: Število paketov/1min v odvisnosti z uro 16
Slika 6: Časi in podatki o kontaktiranih strežnikih 17

1. Uvod

Razvoj tehnologije je danes hiter, ob tem pa podjetja potrebujejo čim več povratnih informacij, ki jim pomagajo pri izboljševanju izdelkov in storitev. S sprejemom pravil uporabe nekega izdelka ali storitve podamo soglasje, da se pošiljajo anonimni podatku proizvajalcu npr. Microsoft, Google in drugi. S tem podamo soglasje za del anonimiziranih podatkov, ki se zbirajo, pošiljajo in obdelujejo. Naju je zanimal predvsem del, kjer soglasja ne podamo in se podatki vseeno zbirajo in pošiljajo. Odločila sva se, da se lotiva pregleda delovanja mobilnega telefona in delovanja v ozadju. Ključna tukaj sta mobilna sistema Android in iOS, ki sta nameščena na večini današnjih naprav. Seveda dodatno zanimanje ustvarja tudi trgovinska vojna, ki jo zaznamo predvsem v obliki obtoževanja med ZDA in Kitajsko. Ker je ključno, kateri podatki in koliko le-teh se zbira in pošilja, je to tudi postalo izhodišče najine raziskovalne naloge, kjer se bova usmerila predvsem v to, da preveriva katere podatke naprave pošiljajo v ozadju.

Zastavila sva si cilje, da bova:

- Preverila obstoječo literaturo na to temo.
- Ugotovila kaj telefon počne v stanju mirovanja.
- Poskušala preveriti razliko med dnevnim in nočnim delovanjem naprave v stanju mirovanja. Kakšna je razlika med delovanjem podnevi in ponoči.
- Analizirala bova ali se v času mirovanja, posodabljanja ali v drugih fazah delovanja na telefonu prenašajo kateri drugi podatki.

2. Teoretičen del

2.1. Zbiranje informacij

Za začetek naloge, sva si zadala vprašanje, kako sploh se ljudje počutijo v današnjem tehnološkem svetu. Da bi dobila vsaj približen odgovor oz. vpogled na to vprašanje, sva se lotila prebiranja člankov in že obstoječih

raziskav na temo pošiljanja osebnih podatkov. Ko sva izvedela v katero smer približno morava gledati, sva zbrala nekaj starejših telefonov in jih testirala s požarnim zidom PfSense. Preverjala sva procese v ozadju, ki se seveda izvajajo dan in noč, povezavo na internet (kam, kdaj, kako in kaj).

Ob vsemu temu sva lahko postavila tudi hipoteze, ki so bile glavne smernice po katerih sva se lahko ravnala.

2.2. Postavljene hipoteze

1. Pametni telefoni imajo vgrajen program oz. algoritem, ki spremlja uporabnikovo početje in to posreduje proizvajalcu oz. nekim zunanjim strežnikom.

2. Pošiljanje podatkov poteka ponoči, ob določenem času in ne redno.

3. Podatki so kriptirani, vendar tudi osebni, s čimer je kršena pravica o varnosti osebnih podatkov.

2.3. Raziskave o pametnih telefonih

Po prebranih in pregledanih člankih različnih medijev in raziskav, sva ugotovila, da je bil to kar nekaj časa zelo velik trend. V času vsega tega dogajanja, so namreč na svetovni trg s pametnimi telefoni prišli proizvajalci iz države Kitajske. Nekaj ljudi je opazilo nenavadna dogajanja v ozadju telefona na določen časovni interval in tako je prišlo do začetka raziskav. Mediji so v tem obdobju zelo hitro začeli objavljati članke o takšnih in drugačnih zlorabah osebnih podatkov, ki naj bi se pošiljali na neznane strežnike locirane na Kitajskem. Po pregledu spletnih strani nekaj Kitajskih proizvajalcev kot so Cubot, Xiaomi, Redmi, Oppo, One Plus in še nekaj podobnih, sva ugotovila, da skoraj nobeden nima navedenih kakršnih koli

obvestil in opozoril do česa vse imajo uradno dostop. V primerjavi z bolj znanimi proizvajalci, ki imajo navedeno uporabo osebnih podatkov in vseh podobnih stvari, ti hitro-rastoči, sveži proizvajalci niso imeli ničesar, kar pomeni, če pridobivajo osebne podatke kršijo pravice uporabnika.

Kako so Kitajski proizvajalci prodrli na trg? Odgovor na to vprašanje je zelo preprost. Sama izdelava ni tako draga in ob pomisleku lahko v telefon vgradijo opremo, ki se na začetku lahko kosa s cenovno boljšimi telefoni, vendar čez čas ta moč nekako upade. In zaradi nizke cene ter odličnih specifikacij vsak človek, ki ni zvest kakšni znamki, hitro ugotovi, da je to veliko boljša stvar. A kaj hitro se lahko izkaže, da temu ni tako.

Po nekaj raziskavah je bilo ugotovljeno, da naj bi telefoni pošiljali informacije o uporabniku vsakih 72 ur, v nočnih urah oz. v času, ko telefon miruje, na neznane strežnike v Šanghaju. Dostopali naj bi do klicev, sporočil, brskanj po internetu, uporabe aplikacij ipd. Tovrstne raziskave so se pojavile na spletnem portalu XDA Developers. Takih objav je zelo veliko z najrazličnejšimi naslovi in temami. Izpostavljeni pa so predvsem telefoni Kitajskih proizvajalcev, na katere se tudi osredotočava.

Kljub temu veliko podjetij in medijev te trditve in raziskave še vedno zanika, saj naj bi bile le te ponarejene in napačno dokazane. To sva si zadala kot največjo nalogo in sicer poskusila sva priti čim bližje tem raziskavam in trditvam, da vidiva ali je vse to resnično ali ne.

3. Eksperimentalni del

5.1. Priprava naprav

Vsa testiranja so potekala na enem pametnem telefonu in sicer Huawei P6-U06 z nameščenim operacijskim sistemom Android verzije 4.4.2., ker je ravno ob izbruhu šknadaloznih novic večina telefonov oprabljala Android verzije 4. Naprava je bila ponastavljena na tovarniške nastavitve in ni imela nobenih drugih aplikacij, ki bi bile nameščene s strani uporabnika poleg standardnih prednameščenih programov.

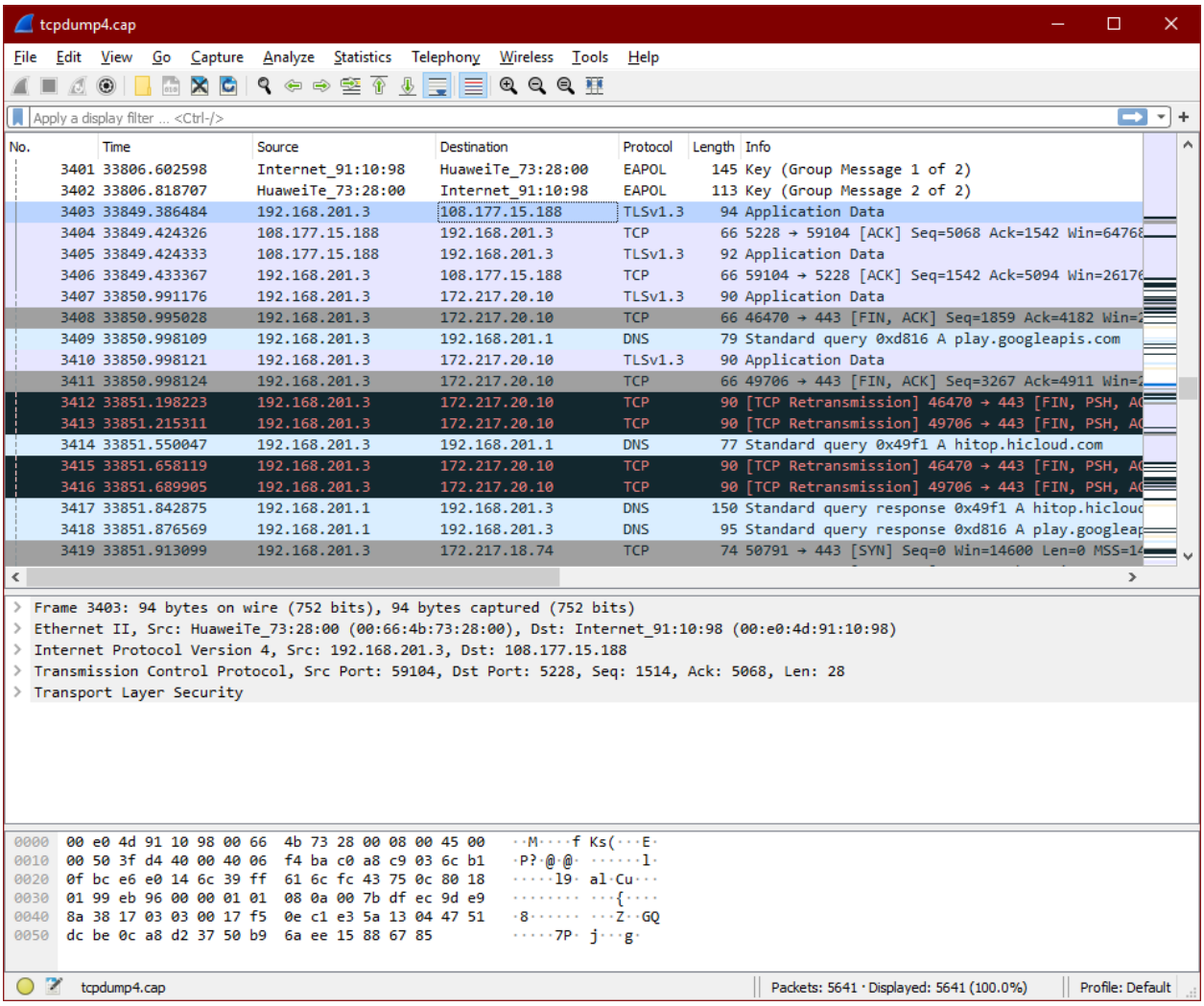
5.2. Izbira programske opreme

Za izvedbo celotne naloge je bilo seveda potrebno poiskati tudi primerno programsko opremo, ki spremlja delovanje telefona in dostop do interneta. Sprva sva poizkusila program Wireshark, ki je orodje za beleženje prenosa podatkov in lahko deluje ves čas v ozadju. Drugo orodje, ki je bilo veliko bolj uporabno ampak malo zahtevnejše je bilo PfSense. To je požarni zid, katerega se namesti na strežnik in deluje kot samosvoj operacijski sistem. Znotraj le-tega se nastavi vse potrebno kot so omrežje, požarni zid, pravila, zajemanje podatkov ipd.

5.2.1. Wireshark

Wireshark je odprtokodni program, ki analizira omrežne protokole in prenos podatkov. Je zelo uporaben in prilagodljiv ter nezahteven za sisteme na katerih deluje. Pri samem analiziranju in zajemanju podatkov, se lahko določi tudi točne parametre, po katerih želimo pretok spremljati. Pri najini nalogi, sva izbrala samo parameter IP naslov telefona in s pomočjo tega spremljala samo tiste podatke, ki jih telefon pošlje ali pa so mu namenjeni.

V Wiresharku sva pregledovala dobljene podatke, ki so se zapisovali med zajemanjem delovanja telefona. Prav tako je program že sam izrisal določene grafe, glede na podane informacije.

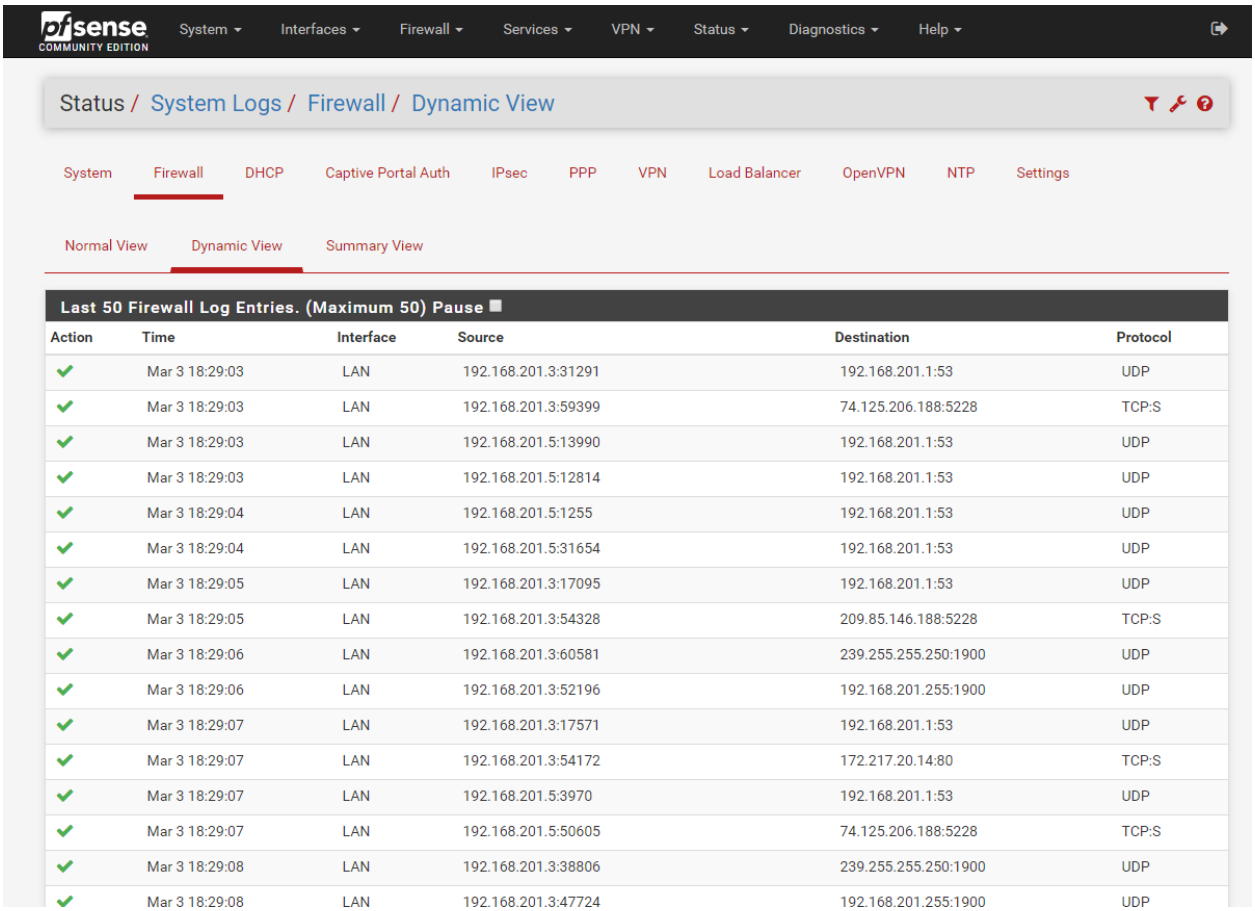


Slika 1: Zajeti podatki v programu Wireshark

5.2.2. PfSense

PfSense je brezplačni požarni zid, katerega se lahko namesti na katerokoli napravo, ki bo delovala kot strežnik oziroma požarni zid. Namestitev in uporaba sta preprosti, saj je operacijski sistem osnovan na že obstoječem FreeBSD.

V najinem primeru, sva operacijski sistem namestila na virtualni strežnik, ki je deloval kot požarni zid za omrežje v katerem se je nahajal telefon in miroval.



Slika 2: Spletni vmesnik PfSense

5.2.3. TCPDump

TCPDump je programsko orodje, ki beleži vse podatke na mrežni kartici in jih shrani v datoteko, katero lahko potem odpremo s programom Wireshark ali obdelamo s Python knjižnico Scapy.

5.3 Rezultati testiranja

Že po enem dnevu je bilo dostopanj do svetovnega spletna in različnih strežnikov ogromno, prav tako pa obratno. Rezultate sva pregledovala tako, da sva izpiske iz programa PfSense pregledala, zapisovala različne strežnike, kje se nahajajo in kako pogosto se je to dostopanje izvajalo.

Da sva si le-to olajšala, sva napisala preprosti skripti v programskem jeziku Python, ki sta poiskali naslov IP danega strežnika kamor oz. od so podatki prišli in s pomočjo spletne strani ipinfo.io pridobil naziv podjetja, ki je lastnik tega naslova in kje se strežnik nahaja.

Python skripta za izvoz različnih IP naslovov iz 4 datotek, ki jih zapiše v eno datoteko:

```
from scapy.all import *

a = []
for fileI in range(1, 5):
    print(fileI)
    file = sniff(offline="tcpdump{}.cap".format(fileI))
    for i in file:
        if type(i.payload) is IP:
            if i.payload.dst not in a:
                a.append(i.payload.dst)
            if i.payload.src not in a:
                a.append(i.payload.src)
with open("ipji.txt", "w") as i:
    i.write(', '.join(a))
```

Python skripta za pridobivanje podatkov o IP naslovu, ki jih pridobimo s spletne strani ipinfo.io v obliki json podatkovnega tipa. V kodi je uporabljen tudi token, ki je za uporabo te spletne strani obvezen, vendar je vezan na privatni uporabniški račun.

```
import http.client
import json

with open("ipji.txt", 'r') as array:
    ipji = array.read().split(', ')
token = "☺ "
connection = http.client.HTTPConnection('ipinfo.io')
a = []
for i in ipji:
    connection.request("GET", "/{}?token={}".format(i, token))
    response = connection.getresponse()
    data = json.load(response)
    a.append(data)
with open("ipinfo.json", "w") as output:
    json.dump(a, output)
```

Primer dobljenega rezultata iz strani ipinfo.io, v json obliki:

```
"ip": "2.18.69.204",
"hostname": "a2-18-69-204.deploy.static.akamaitechnologies.com",
"city": "Cambridge",
"region": "Massachusetts",
"country": "US",
"loc": "42.3620,-71.0830",
"org": "AS16625 Akamai Technologies, Inc.",
"postal": "02142",
"timezone": "America/New_York"
```

Slika 3: Primer izvoženih podatkov

Večinski delež strežnikov v najinem novem seznamu so bili strežniki, čigar lastnik je podjetje Google. Strežniki se nahajajo v Ameriki, vendar jih je tudi več. Na seznamu se pojavi kar 36 različnih IP naslovov, ki so locirani v Ameriki in seveda je lastnik le-teh Google. Zelo pogosta zadeva je bila tudi Akami Technologies inc. To podjetje je v najinem primeru obratovalo samo z dvema strežnikoma 4-krat na dan v zamiku 6 ur. En strežnik je bil aktiven ob 5. in 17., drugi pa ob 11. in 23. uri, oba strežnika pa sta v Ameriki,

Cambridge. Med celotnim testiranjem, pa se je vmes pojavil tudi en strežnik v lasti istega podjetja vendar lociran v Amsterdamu. Razlog za to nama ni znan.

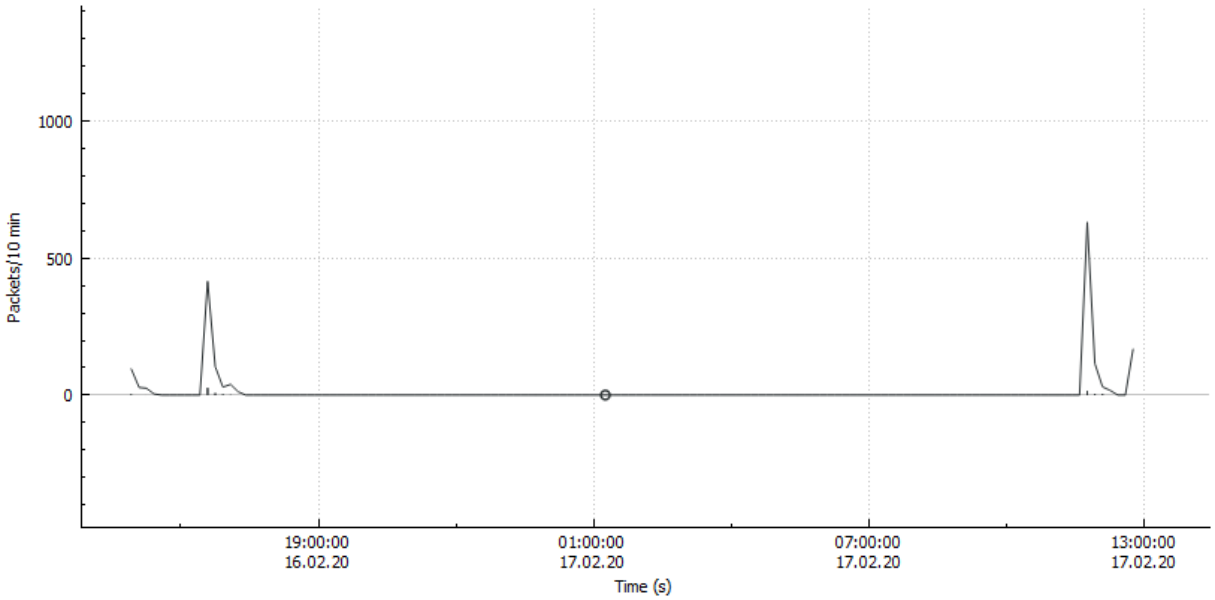
Po brskanju po spletu sva našla, da podjetje Akamai skrbi za internetno varnost in prav tako oblačno shranjevanje za proizvajalca Huawei. Podjetje naj bi po nekaterih trditvah tudi pomagalo zbirati podatke preko svojih strežnikov po svetu, ampak to nikoli ni bilo potrjeno zato bi tudi bolj težko kaj dokazala.

Poleg odobrenih oz. posredovanih internetnih sporočil, pa so tu tudi tista neuspešna in zavrnjena. Največkrat se je to pokazalo kot neuspešno iskanje spletnega naslova preko storitve DNS. Naprava je iskala najrazličnejše in morda tudi celo malo nenavadne naslove, ki so po preverjanju na prvi pogled videti kot strani, ki delujejo kot vaba za prevare.

5.3.1 Stanje naprave v mirovanju in aktivnem delovanju

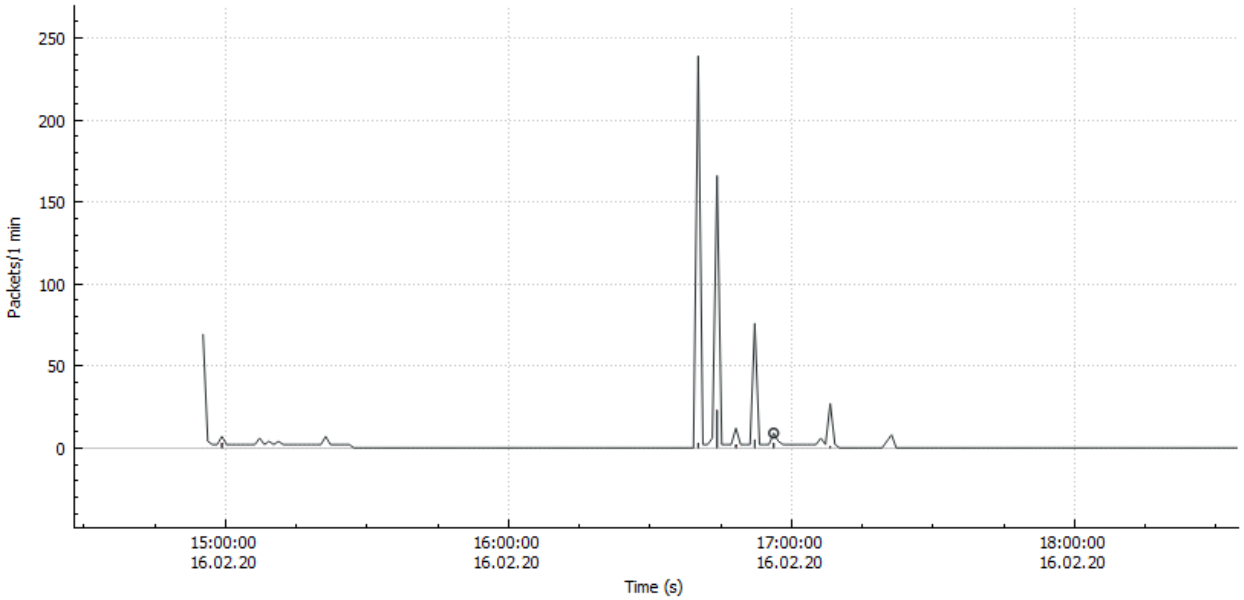
Kot že povedano, sva imela namen tudi raziskati kako se naprava obnaša v svojem stanju mirovanja in delovanja. Splošno znano je, da ko naprava deluje zagotovo pošilja več informacij kakor pa, če miruje. Čeprav je naprava večino časa mirovala, sva lahko zaznala posamezne aktivnosti, ki so odstopale od pričakovane neaktivnosti naprave. Kot pomoč prikazu le-tega sva uporabila program Wireshark, ki lahko izriše grafe glede na aktivnost telefona in omrežja. Za izris grafa sva uporabila število paketov na minuto v odvisnosti od ure in datuma oziroma za manjši graf z večjim časovnim intervalom število paketov na 10 minut v odvisnosti od ure in datuma.

Na spodnjem grafu je prikazano število paketov v 10 minutah v odvisnosti z datumom in uro. Kakor je razvidno je telefon večino časa miroval razen okoli 17. In 12. ure. Prenesenih je bilo nekje med 300 in 700 paketov v samo 10 minutah, kar pomeni, da je bila takrat aktivnost kar zelo povečana. Da to lahko potrdiva je tudi iz grafa zelo razvidno kako aktivnost v trenutku naraste.



Slika 4: Število paketov/10min v odvisnosti z uro

Za lažjo in malo natančnejšo predstavo je spodaj še slika kjer so na ordinatni osi prikazani preneseni paketi na minuto. Vidimo lahko, da je vseeno tudi nekaj vmesnega delovanja telefona, vendar nikjer ni aktivnost tako visoka kot že v izpostavljenih točkah.



Slika 5: Število paketov/1min v odvisnosti z uro

V tem času, torej 16.2.2020 okoli 17. ure je bila večina paketov prenesena med podjetjem Google in telefonom preko strežnikov na Madžarskem, Irski in Ameriki. Vmes pa se je kot že povedano pojavil tudi prejemnik Akamai Technologies. Zaradi povečanega števila prenesenih paketov, bi lahko rekli, da je najin sum vsaj malce utemeljen. Podatki se pošiljajo časovno načrtovano na točno določene IP naslove strežnikov. Zaradi kriptiranih podatkov je sicer skoraj nemogoče pogledati, za izmenjavo katerih podatkov točno tukaj je šlo, ampak zagotovo niso bile samo navadne posodobitve. Na spodnji sliki je možno razbrati, da je bila povezava s strežnikom vzpostavljena samo enkrat. Po 6 urah se je to ponovilo še enkrat, vendar z drugim strežnikom in čez 12 ur je bil isti strežnik zopet kontaktiran.

```
Tue Feb 16 14:37:59 2020, 172.217.18.74, AS15169 Google LLC, US, Mountain View
Tue Feb 16 14:38:02 2020, 172.217.18.74, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:53 2020, 216.58.214.196, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:53 2020, 172.217.20.10, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:53 2020, 172.217.19.110, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:53 2020, 216.58.214.196, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:53 2020, 172.217.20.10, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:54 2020, 64.233.166.188, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:54 2020, 172.217.16.106, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:10:54 2020, 184.51.9.204, AS16625 Akamai Technologies, Inc., US, Cambridge
Tue Feb 16 17:10:56 2020, 172.217.16.106, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:13:18 2020, 172.217.20.10, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:13:19 2020, 172.217.20.10, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:36:20 2020, 172.217.19.106, AS15169 Google LLC, US, Mountain View
Tue Feb 16 17:36:21 2020, 172.217.18.74, AS15169 Google LLC, US, Mountain View
Tue Feb 16 18:41:35 2020, 74.125.71.188, AS15169 Google LLC, US, Mountain View
Tue Feb 16 18:41:35 2020, 172.217.19.110, AS15169 Google LLC, US, Mountain View
Tue Feb 16 18:41:38 2020, 172.217.19.106, AS15169 Google LLC, US, Mountain View
Tue Feb 16 18:41:40 2020, 172.217.18.74, AS15169 Google LLC, US, Mountain View
Tue Feb 16 18:41:44 2020, 172.217.19.106, AS15169 Google LLC, US, Mountain View
Tue Feb 16 18:41:46 2020, 172.217.16.106, AS15169 Google LLC, US, Mountain View
Tue Feb 16 18:41:47 2020, 74.125.140.113, AS15169 Google LLC, US, Mountain View
```

Slika 6: Časi in podatki o kontaktiranih strežnikih

4. Razprava

Vseh hipotez, ki sva si jih postavila, na žalost ne moreva potrditi zaradi premajhnega števila dokazov. Potrjujeva hipotezo, da telefon določene podatke pošilja na določene strežnike in sicer strežnike podjetja Akamai Technologies.

Delno lahko potrdiva hipotezo, da se pošiljajo podatki na določen časovni zamik, vendar se to ne dogaja zmeraj ponoči pač pa na vsakih 6 ur.

In še zadnja hipoteza, da so v paketih kriptirani osebni podatki. Hipotezo ne moreva niti potrditi niti zavreči. Ker lahko zaznava aktivnosti, kjer pa zaradi šifriranja podatkov ne vidiva, ne moreva preveriti vsebnost poslanih/prejetih podatkov. S tem pa ne moreva hipoteze niti zavreči, niti potrditi.

5. Zaključek

Naloga je bila zabavna in poučna. Oba sva se kaj hitro zavedla, da ne bo tako enostavno kot je na prvi pogled, saj je že v eni sami sekundi tako veliko paketov, da je analiziranje vsega tega skoraj da mučno.

Vmes sva si tudi pomagala s svojim znanjem iz programiranja in delček analize rešila tudi s pomočjo le tega. V nadaljnje bova preizkusila še nekaj naprav in morda prideva vsemu temu do dna.

Ker sva skozi nalogo dokazala, da se aktivnosti v ozadju izvajajo, lahko le podava prijazen nasvet uporabniku, da naj pazi na uporabo mobilne naprave in podatke, ki jih tej napravi zaupa. Vsekakor pa še vedno ostaja pomislek, kako smotrn je nakup naprave manj znanega proizvajalca, običajno kitajskega proizvajalca, kjer ne vemo, kdo je pravzprav odgovoren za ravnanje z našimi podatki in kaj z njimi počnejo.

6. Zahvale

Rada bi se zahvalila mentorju Petru Kreblju in sošolcema Davidu Paniču in Jakobu Vadnjalu.

Viri

Aashishvanand. (14. 12 2019). *XDA Developers*. Pridobljeno iz XDA Developers - Data collection without consent: https://forum.xda-developers.com/oneplus-5/how-to/data-collection-consent-t3688082

IpInfo. (20. 2 2020). *IpInfo*. Pridobljeno iz IpInfo: https://ipinfo.io/

PfSense. (20. 2 2020). *PfSense*. Pridobljeno iz PfSense: https://www.pfsense.org/

Scapy. (25. 1 2020). *Scapy*. Pridobljeno iz Scapy: https://scapy.net/

Shirohige4. (19. 1 2020). *XDA Developers*. Pridobljeno iz XDA Developers - Oxegyn OS collecting data: https://forum.xda-developers.com/oneplus-5/how-to/oxegynos-secret-data-collection-logging-t3686732

TcpDump. (11. 10 2019). *TcpDump*. Pridobljeno iz TcpDump: https://www.tcpdump.org/

Wireshark. (20. 2 2020). *Wireshark*. Pridobljeno iz Wireshark: https://www.wireshark.org/