Filipe Barroso

Mentor: Ming Chow

Comp 116: Final Project Outline

12/12/2014

Privacy Concerns: Analysis of music applications on Android

Abstract

In the world of computer security, a man-in-the-middle attack is not something you want to be on the wrong end of. It can allow an attacker to gain access to your personal information (i.e. credit card info, account numbers, or even your pokerhand[1]) as it is sent to and from the intended server. However, what if we could use it for good? What if we could use it to figure out exactly where our personal information is going when we use our favorite mobile applications?

In this article, I will use mitmproxy[2] to do exactly that. I will leverage its ability to conduct a man-in-the-middle attack in order to analyze HTTP/S traffic to some of the most popular music applications that are available today on android mobile devices. What results may surprise you.

Introduction

Love them or hate them we all have a smartphone or know someone else that does. They are becoming a fabric of our daily lives no matter where we live or who we are. Research shows that smartphone usage is continuing to grow at a rapid pace. In 2014, emarketer.com[3] expects there to be 1.75 billion smartphone users worldwide, up from 1.13 billion in 2012. That’s a 55% increase in two years.

People from all walks of life are using them in their daily lives for email, social media, online banking, or to simply order takeout. No matter who you talk to, everyone seems to have a favorite app of theirs.   Just look at lifehacker.com[4], and you'll see a list of what they consider to be the most essential applications for android.

With smartphones being so prevalent and seeming to make everyone's life easier and more enjoyable, what's there not to like about them? Well, like most things, not everything is as picture perfect as it seems. In this case, the elephant in the room here is people's privacy and the personal data they are storing &/or sharing with third-party vendors, often-times without the owner's knowledge.

As this information begins to pile up, it can start to paint a picture of who you are as a person as well as your daily schedule and behaviors. This can have undesired effects.   For example, if someone had access to your GPS location on a daily basis, they could track where you begin your days, where you go, and then where you return at night. This information can later be used to determine your home address, work address, and your favorite hangouts. If this were to fall into the wrong hands, it could provide a home invader with a blueprint of your schedule. The home invader would then know the perfect time to strike.

Needless to say that is a bad thing. Along with this example, there are countless others I could provide where giving personal information to the wrong people can be a bad thing. That being said, in this article, I will focus strictly on the privacy concerns as it relates to some of the more popular music applications available on android. What results is an investigative study into some of the information that Pandora & Spotify disclose on your behalf without your knowledge.

To the Community

Security & privacy on mobile devices is like the “Wild Wild West” and not many people truly understand how they relate to the applications they use on a daily basis. However, most people don't even think twice about using them. If you were asked to share your credit card or bank information with a random application or stranger most people could see the potential ill effects of this and refuse to do so. However, oftentimes, it’s not that blatantly obvious. Either there is a lack of transparency from the application regarding the information it collects on your behalf or it’s simply not that black and white.

One study shows that 53% of people are neutral or not at all concerned with apps accessing their personal information[5]. When you add in the folks that are slightly concerned we get 83% of mobile users. This is an alarming amount. To me that says 83% of people do not think there is anything wrong with the current state of affairs when it comes to sharing their personal data on their mobile devices or do not care to do anything about it. This leaves the remaining 17% concerned individuals in the minority. Whether or not they actually try do something about it is another story.

The reason I chose to investigate music applications in this paper is because they are so prevalent and almost everyone uses them on their smart phones. The thought is, if personal information is being leaked by something that is as commonplace as music apps then what other types of personal information may be getting leaked when we use our other mobile applications on a daily basis.

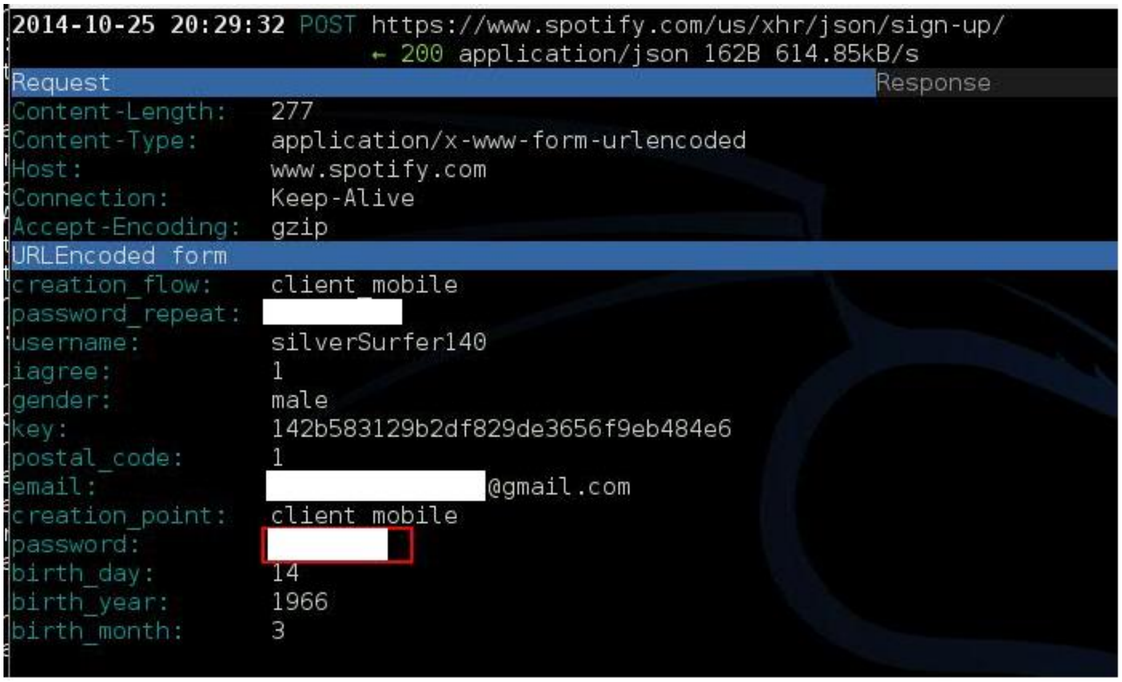
The hope is that the information presented here within will help educate the mobile end user on the types of personal information that may be exposed when using seemingly harmless music apps on a daily basis. This will hopefully make them question the other apps they are using and the types of personal data those apps, in turn, may be sending out on their behalf. Whether they decide to use those applications afterwards is up to them. The important part is that they consider it in order to make a more informed decision.

Application Analysis

To perform the privacy analysis on these applications, I decided to use mitmproxy[2]. When configured properly, this meant that the requests from Pandora or Spotify would first pass through my proxy (e.g. mitmproxy) prior to reaching their respective servers and vice versa. In other words, we can simply listen in on the conversation and see what type of information is passed back and forth.

SPOTIFY

For Spotify, it did not take me long before I found something very alarming. Since I did not have an existing account, I needed to create one. To my dismay, upon creating an account, I uncovered a lot of personal information being sent to the server in clear text. The most alarming of the personal information was my username & password being sent to the server in clear text (see Figure 1). For those that do not know, this is a big no-no in the security realm, as that is something that can be easily viewed by a hacker on a compromised network (think hotel Wi-Fi, Starbucks Wi-Fi or any other free Wi-Fi for that matter).



```
2014-10-25 20:29:32 POST https://www.spotify.com/us/xhr/json/sign-up/
                         ← 200 application/json 162B 614.85kB/s
Request                                                    Response
Content-Length:     277
Content-Type:       application/x-www-form-urlencoded
Host:               www.spotify.com
Connection:         Keep-Alive
Accept-Encoding:    gzip
URLEncoded form
creation_flow:      client_mobile
password_repeat:    [redacted]
username:           silverSurfer140
iagree:             1
gender:             male
key:                142b583129b2df829de3656f9eb484e6
postal_code:        1
email:              [redacted]@gmail.com
creation_point:     client_mobile
password:           [redacted]
birth_day:          14
birth_year:         1966
birth_month:        3
```

Figure 1. Username & password sent in clear

Given my username and password, an attacker could simply login to my Spotify account at a later point and view all of the information I have listed there within including: Payment history, Payment settings (i.e. credit card info, PayPal), subscription, offline devices, profile (i.e. mobile phone, gender, birth date, zip code etc.) as well as sign me out of all devices and/or change my password to lock me out.

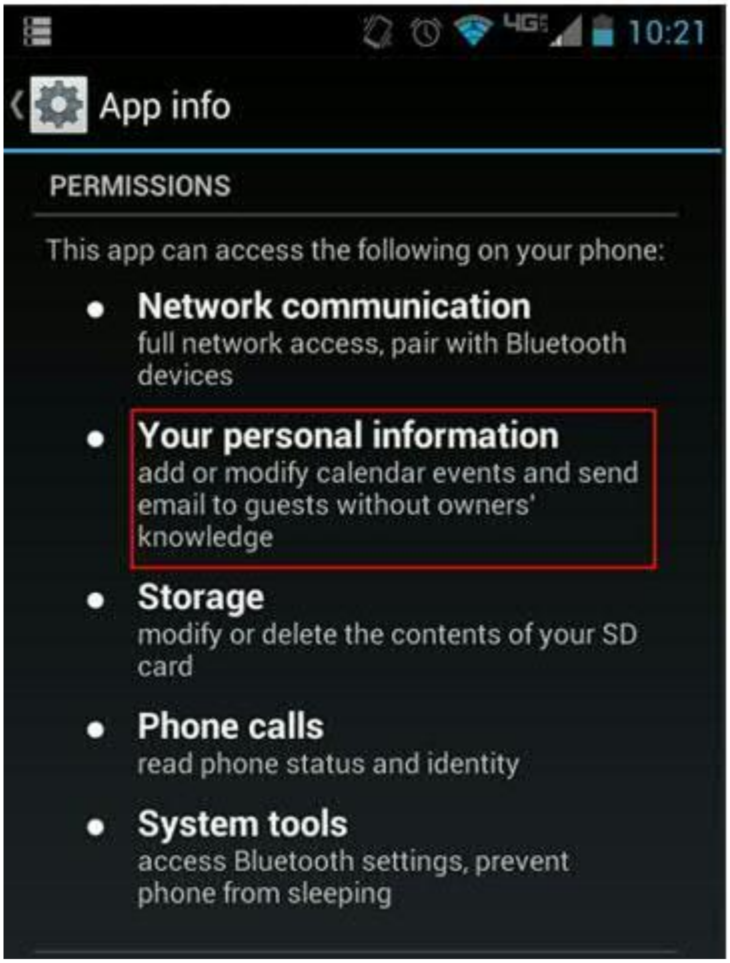
You don’t have to be a rocket scientist to figure out that is bad news. Furthermore, if I was like many people who use the same username & password for multiple accounts like email, banking, etc. I’ve effectively given them the keys to the kingdom as the same username and password used here would work on those sites as well.

In addition to username and password, the following information was also sent as part of an account creation: gender, postal code, email, & birth day. You may be asking yourself, “why does Spotify need to know my gender, postal code, birthday etc.?” That is a very good question and something we will be touching upon shortly.

PANDORA

Since the goal of this article was to investigate privacy in music applications rather than perform a point for point comparison of them, I decided to take on a different strategy for Pandora. For starters, I decided to take a look at the application's permissions. These are the things people usually ignore when they download a new app from the Play Store.

In looking at the permissions, there were definitely a few red flags that shot up immediately. Pandora has the ability to modify or delete the contents of my SD card. It is able to read my phone status and identity. It has full network access, meaning it could send and receive HTTP/S requests at will, and last but certainly not least it is able to ***add or modify calendar events and send email to guests without owners' knowledge***. Feel free to read that last permission again. This would seem to fall into that blatantly obvious category we discussed earlier. How many people feel comfortable giving any application, let alone a music application, the ability to add or modify calendar events and send email to guests without their knowledge?



10:21

App info

PERMISSIONS

This app can access the following on your phone:

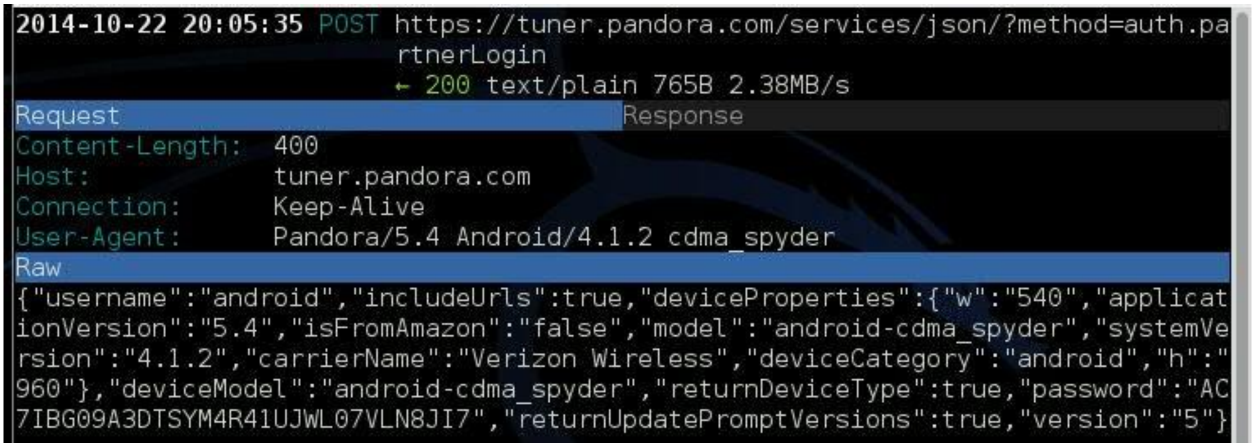
- **Network communication**
  full network access, pair with Bluetooth devices
- **Your personal information**
  add or modify calendar events and send email to guests without owners' knowledge
- **Storage**
  modify or delete the contents of your SD card
- **Phone calls**
  read phone status and identity
- **System tools**
  access Bluetooth settings, prevent phone from sleeping

Figure 2: Permissions granted to Pandora when users install the app

Once I got over that initial shock, I began my mitmproxy analysis on Pandora. Since I already had a user account, I decided to skip the account creation step and simply log in to a pre-existing account. After all, this is probably the most common use case since users only need to create an account once. I listened to my favorite station for a few hours and gave thumbs up to a few songs I liked. The thumbs down always seemed harder to come by, but in the name of science I sprinkled in a few for good measure.

Once I felt I had a good sample size, I decided to dig in and analyze the resulting mitmproxy output. To my relief, my password was not sent to the server in clear-text like it had been during the previous analysis. Instead it appeared to be encrypted making it non-human readable. While this was a positive sign, there were other pieces of information being sent that made me wary.

I noticed that they were capturing quite a bit of information about the type of device I was using. Things like the type of smartphone (i.e android), the phone model, system version, carrier name, the height and width of my smartphone screen amongst other things (See Figure 3). One could argue the display size may help with the visual layout of the application, but why on earth do they need to know my phone model number or phone carrier, especially when the app is just playing music? This was all sent as part of the login request.

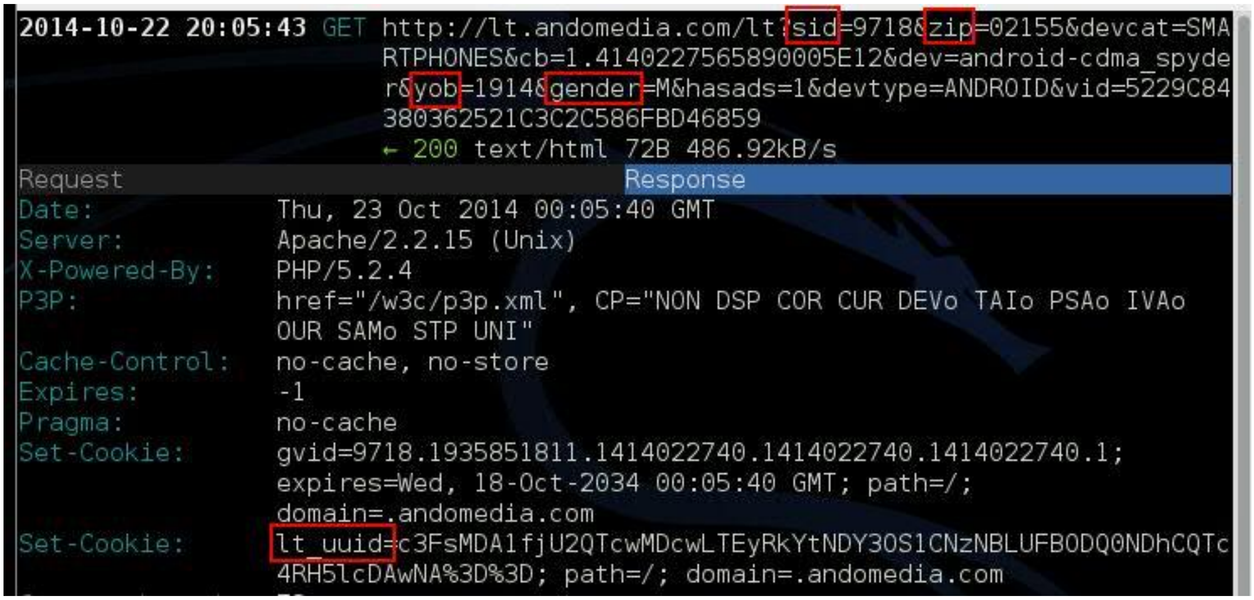


```
2014-10-22 20:05:35 POST https://tuner.pandora.com/services/json/?method=auth.pa
                         rtnerLogin
                         ← 200 text/plain 765B 2.38MB/s
Request                                        Response
Content-Length:   400
Host:             tuner.pandora.com
Connection:       Keep-Alive
User-Agent:       Pandora/5.4 Android/4.1.2 cdma_spyder
Raw
{"username":"android","includeUrls":true,"deviceProperties":{"w":"540","applicat
ionVersion":"5.4","isFromAmazon":"false","model":"android-cdma_spyder","systemVe
rsion":"4.1.2","carrierName":"Verizon Wireless","deviceCategory":"android","h":"
960"},"deviceModel":"android-cdma_spyder","returnDeviceType":true,"password":"AC
7IBG09A3DTSYM4R41UJWL07VLN8JI7","returnUpdatePromptVersions":true,"version":"5"}
```

Figure 3: Info collected by Pandora during login.

After every song, I noticed my zip code, year of birth (yob), gender and other id values being sent to third party advertisement companies like lt.andomedia.com (See Figure 4). Does this sound familiar? It should, because some of those values were also being captured by Spotify.

The sid value appeared to correspond to the song I was listening to at the time. However, the vid value still remains a mystery to me. At any rate, it appeared to set an additional Universal Unique Identifier (lt_uuid) value in the response so that lt.andomedia.com could further track my interactions.



```
2014-10-22 20:05:43 GET http://lt.andomedia.com/lt?sid=9718&zip=02155&devcat=SMA
                        RTPHONES&cb=1.4140227565890005E12&dev=android-cdma_spyde
                        r&yob=1914&gender=M&hasads=1&devtype=ANDROID&vid=5229C84
                        380362521C3C2C586FBD46859
                        ← 200 text/html 72B 486.92kB/s
Request                                         Response
Date:                Thu, 23 Oct 2014 00:05:40 GMT
Server:              Apache/2.2.15 (Unix)
X-Powered-By:        PHP/5.2.4
P3P:                 href="/w3c/p3p.xml", CP="NON DSP COR CUR DEVo TAIo PSAo IVAo
                     OUR SAMo STP UNI"
Cache-Control:       no-cache, no-store
Expires:             -1
Pragma:              no-cache
Set-Cookie:          gvid=9718.1935851811.1414022740.1414022740.1414022740.1;
                     expires=Wed, 18-Oct-2034 00:05:40 GMT; path=/;
                     domain=.andomedia.com
Set-Cookie:          lt_uuid=c3FsMDA1fjU2QTcwMDcwLTEyRkYtNDY3OS1CNzNBLUFBODQ0NDhCQTc
                     4RH5lcDAwNA%3D%3D; path=/; domain=.andomedia.com
```

Figure 4: Pandora sending info to 3rd party advertisement companies

Doing a simple google search on lt.andomedia.com revealed that the URL belonged to a company by the name of Ando Media. In turn, Ando Media falls under the umbrella of a much larger advertisement company called Triton Digital[6]. Triton Digital boasts such clients as Pandora, iHeartMedia, Spotify, CBSRadio, ClearChannel, npr digital services and many others. It's really a who's who of radio and the music streaming industry. So, if Pandora & Spotify are tracking this info, you can rest assure those other apps are as well.

Their website greets customers with the banner *"Welcome to Triton Digital. We are dedicated to connecting audio, audience, and advertising into the next **multi-billion dollar marketplace**."* So, if you were wondering how much your personal data is worth, rest assured its worth quite a lot!

Veracode, a Gartner leader in Application security, performed a similar analysis[7] of Pandora in 2011. In addition to Ando Media, their team also found that personal info was being sent to the following Advertisement libraries: AdMarvel, AdMob, comScore, Google.Ads, and Medialets. If all this seems to be a breech in consumer privacy you're not the only one that thinks so. In 2011, Federal

Prosecutors in New Jersey began an investigation[8] into certain mobile applications, including Pandora, for violating the Computer Fraud and Abuse Act in 2011.

Conclusion

Now it starts to make a little more sense as to why Spotify and Pandora are capturing your gender, zip code, date of birth amongst other things. They simply pass it along to advertisement companies. It becomes a win-win situation for both the music application and the advertisement company. The music application companies get a kickback to pay the bills while the advertisement companies get your personal data to use in their multi-billion dollar marketplace.

While advertisement companies are notorious for data mining and doing whatever it takes to get an edge, they are not the only ones. Retail stores are now tapping into smartphones to track user movements[9]. So does that mean we should all ditch our smartphones and go back to using flip phones w/o data plans? Of course not. It just means that as consumers, we need to be more aware of what we download on our phones and what information we provide to certain applications. As this investigative report shows, some applications will use anything and everything possible to collect as much information as they can about you so you do not need to provide any more than you have to.

As part of the supporting material for my report, I will be providing a separate video that outlines certain precautions and steps[10] that consumers can take in order to protect their privacy and prevent some of the above privacy breeches from occurring. In the meantime, we should all be vigilant regarding the applications we download and use on a day to day basis and always be mindful of the types of data it could be collecting and sending out on your behalf. After all, if music applications are collecting this much of your personal data just imagine what the other applications on your phone are collecting and sending out on your behalf.

References

[1] http://tuftsdev.github.io/DefenseOfTheDarkArts/lecture_notes/2600vol24no1.pdf

[2] http://mitmproxy.org/doc/

[3] http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536

[4] http://lifehacker.com/lifehacker-pack-for-android-our-list-of-the-essential-819094535

[5] http://insights.wired.com/profiles/blogs/mobile-privacy-lock-down?xg_source=activity#axzz3HRyf11Yk

[6] http://www.tritondigital.com/

[7] http://www.veracode.com/blog/2011/04/mobile-apps-invading-your-privacy/

[8] http://www.wsj.com/news/articles/SB10001424052748703806304576242923804770968?mg=reno64-wsj&url=http%3A%2F%2Fonline.wsj.com%2Farticle%2FSB10001424052748703806304576242923804770968.html

[9] http://lifehacker.com/how-retail-stores-track-you-using-your-smartphone-and-827512308

[10] http://lifehacker.com/how-you-leak-your-privacy-every-day-and-how-to-stop-1547653862