Comparison of TikTok, Facebook, and Twitter
Penetrum LLC
07/14/2020

Overview:

We have been contacted by a lot of media outlets due to our recent release of our TikTok research. They all seem to want to know the same things;

“What is the difference in the data collections between TikTok and their other larger competitors?”

Until now we have had to answer that we didn’t know because we hadn’t done the research to determine whether or not TikTok collects more information than the others. So, we decided to perform static analysis on the following APK’s:

- Facebook mobile app APK v272.0.0.50.125
- Facebook messenger mobile app APK v266.0.0.16.117
- Twitter mobile app APK v8.52.0
- TikTok mobile app APK v16.0.42

The goal of the static analysis was to gather the differences between the aforementioned APK files. From there, we could perform a comparison (diff) with each one against TikTok itself. This would provide us with the differences in data collection between the applications. All of our research will be publicly available on our website at https://penetrum.com/research.

Facebook and TikTok:

The first thing we did was perform a comparison against Facebook and TikTok.

APP INFORMATION

	Failed - Failed	Failed - Failed
File name	TikTok_v16.0.42_apkpure.com.apk	Facebook_v272.0.0.50.125_apkpure.com.apk
MD5	0905dca66330eed1b796f917550d41a3	9d31a235371403ea6e370689a664513c
Size	102.38MB	54.3MB
Certificate	Subject: C=CN, ST=Beijing, L=Beijing, O=ByteDance, OU=ByteDance, CN=Micro Cao	Subject: C=US, ST=CA, L=Palo Alto, O=Facebook Mobile, OU=Facebook, CN=Facebook Corporation

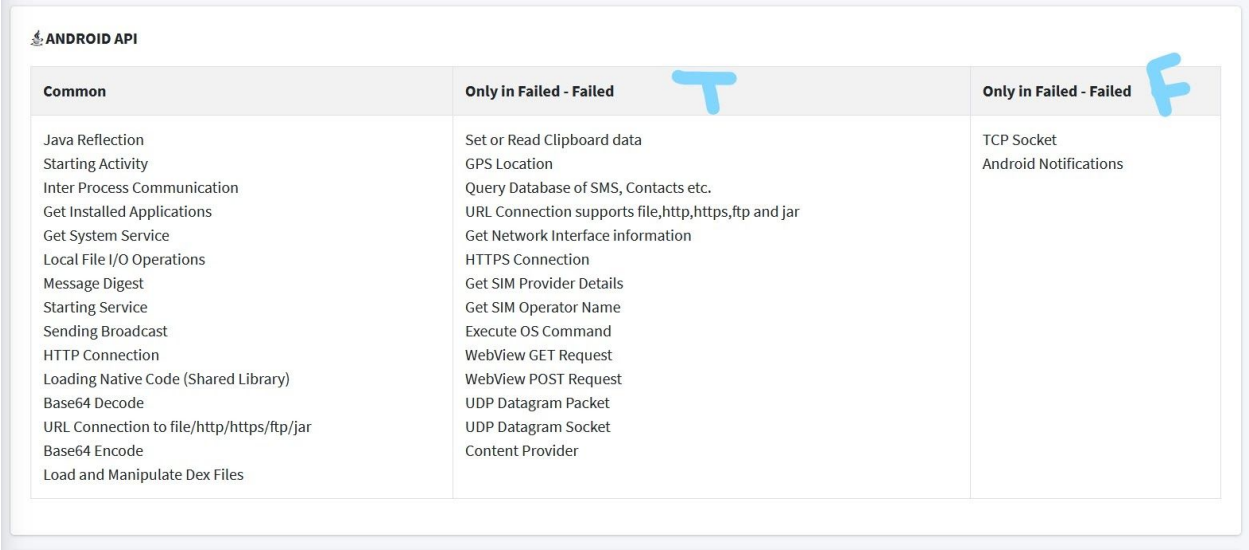
The first thing we would like you to take into consideration is the size difference between both applications. While Facebook sits at 54.3MB TikTok is almost double the size at a massive 102.38MB. We assume that the reason for this is that TikTok does a lot more on the client side than Facebook does. If you notice both applications have “Failed - Failed” on them, what this means is that the obfuscation inside of the code was too much for the parser to handle and it wasn’t able to pull some of the information out of the APK files. The obfuscation done in both applications is extreme, even going as far as to try and stop disassembly of the APK itself. In the below picture you see the differences between each APK TikTok is marked with a “T” and Facebook is marked with an “F”:

APKiD

	ANTI-VM	COMPILER	OBFUSCATOR	PACKER	DROPPER	MANIPULATOR	ANTI-ASSEMBLY	ANTI-DEBUG	ABNORMAL PATTERN
Common	SIM operator check Build.BOARD check Build.MODEL check network operator name check Build.MANUFACTURER check Build.FINGERPRINT check Build.PRODUCT check Build.TAGS check possible Build.SERIAL check	dx					illegal class name		
Failed - Failed	Build.HARDWARE check		unreadable field names unreadable method names						
Failed - Failed	ro.kernel.qemu check	unknown (please file detection issue!)						Debug.isDebuggerConnected() check	

As you can tell above, there are multiple actions performed to check whether they are being run inside of a sandbox or inside of a virtual machine. We would also like you to take notice of the anti-assembly section of the above table. What this means is that the developers purposely put illegal characters inside of a class name to prevent full disassembly of the application. This is actually a measure taken by a lot of malware to prevent reverse engineering, or make it harder for the engineer to find out what's actually going on. You can read more about malware obfuscation here: https://securityboulevard.com/2020/02/what-is-malware-obfuscation/.

Having said that, we were still able to extract a lot of the data collection out of both applications:



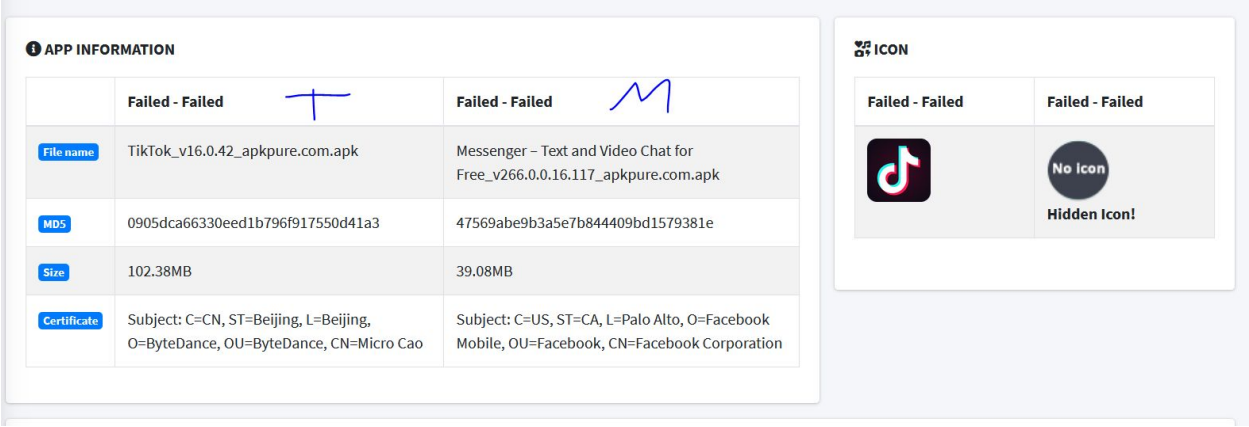
ANDROID API

Common	Only in Failed - Failed	Only in Failed - Failed
Java Reflection	Set or Read Clipboard data	TCP Socket
Starting Activity	GPS Location	Android Notifications
Inter Process Communication	Query Database of SMS, Contacts etc.	
Get Installed Applications	URL Connection supports file,http,https,ftp and jar	
Get System Service	Get Network Interface information	
Local File I/O Operations	HTTPS Connection	
Message Digest	Get SIM Provider Details	
Starting Service	Get SIM Operator Name	
Sending Broadcast	Execute OS Command	
HTTP Connection	WebView GET Request	
Loading Native Code (Shared Library)	WebView POST Request	
Base64 Decode	UDP Datagram Packet	
URL Connection to file/http/https/ftp/jar	UDP Datagram Socket	
Base64 Encode	Content Provider	
Load and Manipulate Dex Files		

In the above picture you can see which API each application uses. The above picture states that TikTok uses more API's than Facebook does. However, as mentioned earlier, there is a possibility that Facebook's heavy obfuscation did not allow us to pull everything. With that in mind, you can download the Facebook code that we were able to extract at https://penetrum.com/research. From the current given context, we can safely assume that not only does TikTok access more API's than Facebook does, but they also do more on the data collection side than Facebook's application does. Everything from getting the SIM provider to querying the SMS contacts.

Facebook Messenger and TikTok:

I'm sure we've all heard the whole "Facebook messenger is what collects the data" statement that floats around the internet. After conducting our Facebook analysis, we were extremely surprised about how little information the Facebook application was collecting. We felt as though we were missing something and that there was more to it. We decided that we would also perform static analysis on Facebook Messenger to get a clearer understanding of what's really going on:



APP INFORMATION

	Failed - Failed	Failed - Failed
File name	TikTok_v16.0.42_apkpure.com.apk	Messenger – Text and Video Chat for Free_v266.0.0.16.117_apkpure.com.apk
MD5	0905dca66330eed1b796f917550d41a3	47569abe9b3a5e7b844409bd1579381e
Size	102.38MB	39.08MB
Certificate	Subject: C=CN, ST=Beijing, L=Beijing, O=ByteDance, OU=ByteDance, CN=Micro Cao	Subject: C=US, ST=CA, L=Palo Alto, O=Facebook Mobile, OU=Facebook, CN=Facebook Corporation

ICON

Failed - Failed	Failed - Failed
	No Icon Hidden Icon!

Once again we would like you to take notice on the size differences in the applications. TikTok once again has a significantly bigger APK than Facebook does. Also, as both failed again, meaning that there is heavy obfuscation in Facebook Messenger as well. We were able to extract the same amount of information from Facebook Messenger as we were from Facebook:

	ANTI-VM	COMPILER	OBFUSCATOR	PACKER	DROPPER	MANIPULATOR	ANTI-ASSEMBLY	ANTI-DEBUG	ABNORMAL PATTERN
Common	Build.HARDWARE check Build.PRODUCT check Build.BOARD check SIM operator check Build.MANUFACTURER check Build.MODEL check possible Build.SERIAL check network operator name check Build.FINGERPRINT check Build.TAGS check						illegal class name		
Failed - Failed		dx	unreadable field names unreadable method names						
Failed - Failed	ro.kernel.qemu check	unknown (please file detection issue!)						Debug.isDebuggerConnected() check	

However, there was an issue with the way Facebook Messenger was compiled, which was different from TikTok's DX compiler. Most of the information is the same as Facebook. They take steps to prevent disassembly, as well as take steps to prevent you from running the application in a virtual machine or sandbox.

ANDROID API

Common	Only in Failed - Failed	Only in Failed - Failed
Java Reflection Starting Activity Inter Process Communication Get Installed Applications Get System Service Set or Read Clipboard data Loading Native Code (Shared Library) Local File I/O Operations Message Digest Starting Service GPS Location Sending Broadcast HTTP Connection URL Connection supports file,http,https,ftp and jar HTTPS Connection Base64 Decode Crypto Execute OS Command Content Provider Base64 Encode TCP Socket URL Connection to file/http/https/ftp/jar Android Notifications	Query Database of SMS, Contacts etc. Get Network Interface information Get SIM Provider Details Get SIM Operator Name Get WiFi Details WebView JavaScript Interface WebView GET Request WebView POST Request UDP Datagram Packet UDP Datagram Socket Load and Manipulate Dex Files	

As you can tell, Facebook Messenger collects noticeably more data than the Facebook application does, going as far as to “set or read clipboard” information. We cannot as this time determine if the messenger application pulls the clipboard information. We do however feel that it is entirely possible to do so. We at Penetrum feel that it is safe to assume that TikTok is collecting more information from it’s users than Facebook is. Just think about that for a second. Facebook, the mother of all beasts, collects less information than TikTok does.

Twitter and TikTok:

APP INFORMATION

	Failed - Failed	com.twitter.android - 8.52.0-release.00
File name	TikTok_v16.0.42_apkpure.com.apk	twitter-8-52-0-release-00.apk
MD5	0905dca66330eed1b796f917550d41a3	c1afb7003211206516eb3233b38419db
Size	102.38MB	52.52MB
Certificate	Subject: C=CN, ST=Beijing, L=Beijing, O=ByteDance, OU=ByteDance, CN=Micro Cao	Subject: C=US, ST=CA, L=San Francisco, O=Twitter, Inc., OU=Mobile, CN=Leland Rechis

ICON

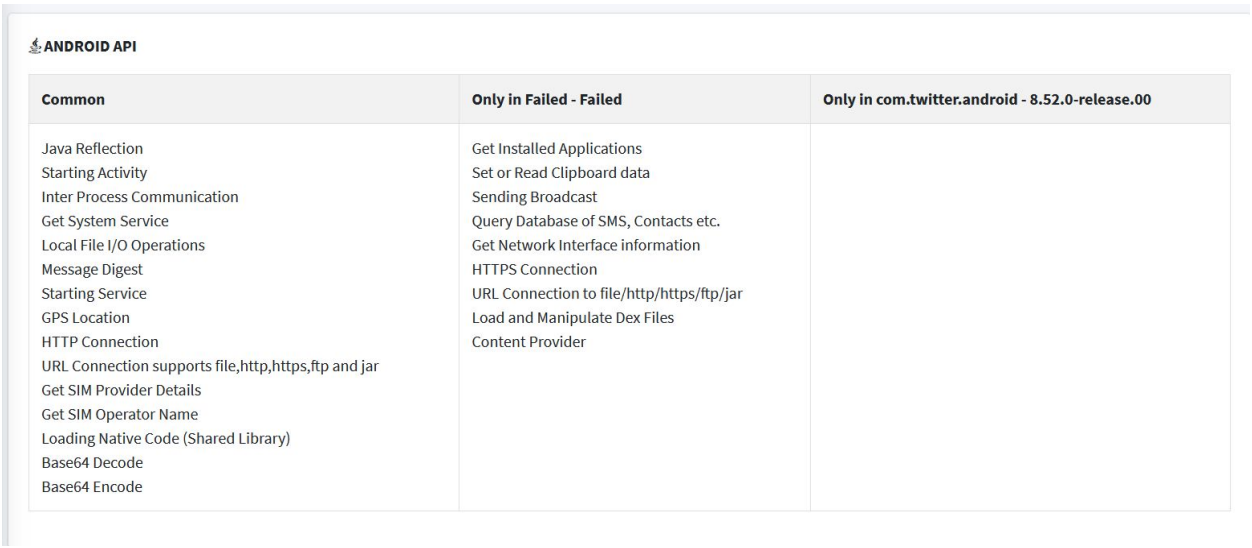
Failed - Failed	com.twitter.android - 8.52.0-release.00
	

Out of all the applications we did static analysis on. Twitter is the only one who's obfuscation wasn't so extreme that we were able to fully extract the application's information. Once again, TikTok is significantly larger than Twitter's 52.52MB APK file. Since Twitter's obfuscation isn't done to an extreme manner we were able to get a lot of the information out of it, but for the sake of this whitepaper we will only be performing analysis on the same information as Facebook, and Facebook Messenger. You can of course download all the information at https://penetrum.com/research. Having said that, continuing on we can see the differences in the protection taken from each applications:

APKiD

	ANTI-VM	COMPILER	OBFUSCATOR	PACKER	DROPPER	MANIPULATOR	ANTI-ASSEMBLY	ANTI-DEBUG
Common	Build.FINGERPRINT check Build.MODEL check network operator name check Build.MANUFACTURER check Build.PRODUCT check SIM operator check Build.TAGS check Build.HARDWARE check	dx						
Failed - Failed	possible Build.SERIAL check Build.BOARD check		unreadable field names unreadable method names				illegal class name	
com.twitter.android - 8.52.0-release.00								Debug.isDebuggerConnected() check

In the above image, you can see that Twitter takes basic precautions against virtual machine running, and also checks for a debugger. They do not attempt to prevent reverse engineering of the application, and do not use an obfuscator on their APK file.



ANDROID API

Common	Only in Failed - Failed	Only in com.twitter.android - 8.52.0-release.00
Java Reflection	Get Installed Applications	
Starting Activity	Set or Read Clipboard data	
Inter Process Communication	Sending Broadcast	
Get System Service	Query Database of SMS, Contacts etc.	
Local File I/O Operations	Get Network Interface information	
Message Digest	HTTPS Connection	
Starting Service	URL Connection to file/http/https/ftp/jar	
GPS Location	Load and Manipulate Dex Files	
HTTP Connection	Content Provider	
URL Connection supports file,http,https,ftp and jar		
Get SIM Provider Details		
Get SIM Operator Name		
Loading Native Code (Shared Library)		
Base64 Decode		
Base64 Encode		

Once again, TikTok is collecting significantly more information from their end users than Twitter is. We would also like to note that Twitter does not access the users clipboard, they do not query your SMS, nor do they gather the applications that are installed on your phone. Penetrum feels that we can safely say that TikTok is gathering much more data than Twitter is.

Conclusion:

So, what can we take away from the comparisons done? Well, for one we can safely assume that TikTok is gathering more data than both Facebook and Twitter. Mind you, we are not talking about the legality of data collected, some or all of the data collected may be used legally by the company depending on the privacy policy and agreements made when signing up to the application, we are not lawyers and will leave the legal issues up to those who are qualified or want to speculate. In conclusion, from what we have seen digging through the apps, TikTok simply takes more information than the other applications.