Part1: A **CW2** Doruk Taneli, Rehan Shah, Tanisha Mandre

Question 1 - The adb devices command gives a list of all the Android devices and emulators attached to the computer. The output contains an adb created identifier for the device (which also shows the port number the device is running on, for example - "emulator-5556") and device state information (with three possible values - offline, device and no device).

Question 2 - The IMEI number of a device cannot be changed, so an ad library could track a user for years using the IMEI number. Another concern is that telecom providers can use the IMEI number to stop a lost or stolen device from accessing their network, so if the IMEI number falls into the hands of an attacker he/she could restrict user access to the network. The introduction of the Advertising ID solves some of these issues. The Ad ID is also a unique string that is used to identify a device, but the difference is that the Ad ID can be reset by the user, giving him/her a lot more control over the data associated with his/her Ad ID.

Question 3 - On trying to read the user's contacts, Android throws a fatal security exception which causes the app to crash. Since neither our ad library nor our app declare the READ_CONTACTS or WRITE_CONTACTS permissions in their Manifest, Android does not allow the app to access contacts. We can handle this by adding the permissions stated above to the Manifest for the ad library. We can also do a permission check in ad library code to make sure we only access the contacts if we have the permission to do so.

Part1: B

Question 1 - The permissions declared in the host app were changed to add a request for the READ_CALENDAR & READ_CALL_LOG permissions.

Callyzer (`https://play.google.com/store/apps/details?id=com.websoptimization.callyzerpro&hl=en_GB&gl=US`) is a real world app that organises call log data and requests the READ_CALL_LOG permission. Google Tasks - (`https://play.google.com/store/apps/details?id=com.google.android.apps.tasks&hl=en`) is a real world app that manages tasks from the calendar and requests the READ_CALENDAR permission.

Question 2 - Data collected:

- **IP Address**: Sensitive as it can be used to track browsing activity, approximate location tracking(eg.upto city/country) & can help someone to to remotely take over a device or impersonate your identity. This information is useful to advertising networks as it can help track your browsing activity which is indicative of your behavior patterns, products you're interested in, etc.
- **Calendar event data**: Sensitive as it contains information about your work, study, commitments, etc., can indicate your location at different times of the day, contains records of people you meet with, sometimes including their contact information as well. This information can be used by advertisers, fot example if you have information on your calendar regarding a doctors appointment for an ultrasound, it might indicate you are pregnant and require baby care products.
- **Call logs**: Sensitive as personal phone numbers and Names of people that can be used for promotional or malicious purposes and records of who you interact with can be accessed. This information could potentially help advertisers understand the services you might be looking for, eg. if you call someone with a number stored as 'Electrician' or 'Plumber', etc. this could also indicate need for their services and promote personalised advertising related to them.

All this information is considered sensitive and should not be accessible to advertising networks as it reveals a large part of your personal life. If advertising libraries are allowed access to this information, the user must be asked and must choose to reveal this information if they want it to be used to improve the quality of advertisements they see.

Question 3 - The IP address is retrieved through the Inet address, got from the Network Interface. For accessing calendar events and call logs, a similar algorithm/strategy was chosen which uses Cursor to loop through the various entries and retireve the information. Calendar logs requires the device to have an account logged into Google Calendar and have some calendar events. For the purposes of the coursework, a dummy account was created and can be accessed at ***advancedcompcw@gmail.com, password: Computer20!***. Call logs require call logs to be present. Having call logs from contacts allows retreival of information like Name of the caller.

Part2: A

Question 1 - We checked the online android ad documentation, for example:
`https://developers.google.com/admob/android/banner`
and saw that the ad files are in <com.google.android.gms.ads.AdView> directory. We were then able to log "Hello Malvertising!" in the constructor of the AdView.smali file in this directory.

Question 2 - To embed our malicious code in smali format, we decompiled the application obtained in Part1 using apktool to see how our java injections translates into smali code. Then we made respective changes to the AdView.smali.

Part2: B

Question 1 - We weren't able to embed the rest of the malicious code as it was much more complex than just logging, so we decided to use the app given for part1 to do the rest of injections and steal sensitive user information.

Question 2 - Same data was collected as Part1B. The logs have been copied over to Part2_malad.txt where the Tag indicates the information being collected followed by the timestamp and followed by the information itself.

Eg. I/LOCATION: (4622): 2020/11/24 19:10:07; longitude:-0.11809166666666666;latitude:51.509863333333335

Question 3 - When we try to co-install both apps, we get the error: INSTALL-FAILED-ALREADY-EXISTS. This is happening because Android uses package names as an ID to check for app uniqueness. The package names of both apps are the same, so we can't install both apps at the same time.

Part2: C

We wrote the script for the given base.apk for part2, but then decided to use another app for parts 2A and 2B, so the script doesn't work for Part2_repackaged.apk. To run the repackaging script, put the following files provided in Part2_source.zip: "script.py", "base.apk", and "my-release-key.keystore" in the same folder. Make sure you have python, apktool and adb installed in your environment, and run the script. For example in Windows 10: Go to said folder using terminal and run >python script.py

To be able to co-install the original version and the repackaged version at the same time, we need to change the package name in AndroidManifest.xml, which android uses to check for app uniqueness. The script decompiles base.apk, changes the package name by appending "amended." to the beginning of the package name, builds a new apk called "amended.apk", signs it using the keystore, and installs it to the device by calling >adb install amended.apk. You can view the script in Part2_source.zip

Appendix

Expanded Part1: B/ Part 2:B

Question 1 - The permissions declared in the host app were changed to add a request to the following permissions:

1. READ_CALENDAR - Allows the app to read calendar event details including event Name, Location, Date, Time, Description, etc.
2. READ_CALL_LOG - Allows the app to read the call logs including Name of caller, Number of the caller, Duration of call, etc.

Many real world apps request such permissions from a user. some examples are as follows:

1. Callyzer - Analysis Call Data - `https://play.google.com/store/apps/details?id=com.websoptimization.callyzerpro&hl=en_GB&gl=US` - This app helps you analyse your call logs in a statistical fashion. It also provides other features including backing up call log data, exporting call logs to excel/CSV files, analyse most frequent caller details, etc.
2. Google Tasks - `https://play.google.com/store/apps/details?id=com.google.android.apps.tasks&hl=en` - This app helps you manage your tasks in an efficient way and accesses your calendar data to help achieve this. It uses the calendar events to convert them into tasks to be maintained on a single task manager. It integrates tasks from other platforms including Gmail.

Question 2 - Data collected:

1. **IP Address**:

   Why is it sensitive data?

   (a) An IP address is sensitive data as knowledge of it can enable someone to track your browsing activity.

   (b) An IP address can also be used to track general location (not precise address). To provide an example of how this might be harmful, if there is a situation where your IP address is known and say the location of your city can be tracked, a person can be aware if you go for a vacation to another city. In this case someone could use this information to try and rob your house for example, knowing that you will not be there.

   (c) Knowledge of an IP Address can allow someone to hack into your device remotely by trying to brute force an attack to access a connection from the tens of thousands of ports available for your IP address. Once a connection is established they can steal control your device, steal data, etc.

   (d) IP address could help someone impersonate your identity

   Do you think this type of information could be useful to advertising networks?

   This information is very useful to advertising networks as it can help track your browsing activity which is indicative of your behavior patterns, products you're interested in, etc.

2. **Calendar event data**:

   Why is this sensitive data?

   (a) Contains information about your work, study, commitments, etc. which is sensitive information/

   (b) Can indicate your location at different times of the day.

   (c) Contains records of people you meet with, sometimes including their contact information as well.

   Do you think this type of information could be useful to advertising networks?

   This information might be useful to advertisers depending on the amount of information your calendar contains. For example if you have information on your calendar regarding a doctors appointment, they can get an idea of what products you might be interested in if it contains details indicative of why you need the visit. For example a ultrasound could indicate that you might be pregnant and would be interested in those kind of advertisements, if the doctor is an ENT doctor then it might indicate you could be in need of some fly medication, etc..

3. **Call logs**:

   Why is this sensitive data?

   (a) Personal phone numbers and Names of people that can be used for promotional or malicious purposes.

   (b) Records of who you interact with

   Do you think this type of information could be useful to advertising networks?

   This information could potentially help advertisers in some way. Call logs might sometimes indicate sleep patters, for example if you spend 3 hours on call from midnight-3am it might indicate that you need something to help you sleep. If contacts are stored with descriptive names for example ’Electrician’ or ’Plumber’, etc. this could also indicate need of specific services and promote personalised advertising related to them.