Android Mischief Dataset

stratosphereips.org/blog/2020/11/10/android-mischief-rats-dataset

Kamila Babayeva

November 18, 2020

Introduction

A Remote Access Trojan (RAT) is a type of malware that allows the attacker (client) to gain control of the target's device (server) to remotely control it. RATs are one of the most important threats nowadays since they are used as part of most attacks, from APTs to Ransomware. It is not an easy task to detect RATs in the network traffic, especially when it comes to Android RATs in phones. Why? The main problem is that there are no easy ways to look at the network traffic on our mobile devices. Our phones are much harder to protect than our computers. Even in cases where there are external network traffic analyzers, there are no good RAT detectors. To approach the problem of the lack of Android RATs detection in the network traffic, we want to help the community by creating the **Android Mischief Dataset**, which contains network traffic from mobile phones infected with real and working Android RATs.

The Android Mischief Dataset is part of the Civilsphere Project (https://www.civilsphereproject.org/), which aims to protect the civil society at risk by understanding how the attacks work and how we can stop them. Check the webpage for more information.

Download the version 1 of the dataset from here: https://mcfp.felk.cvut.cz/publicDatasets/Android-Mischief-Dataset/ (short link https://bit.ly/AndroidMischief)

The Android Mischief Dataset

The Android Mischief Dataset is a dataset of network traffic from mobile phones infected with Android RATs. Its goal is to offer the community a dataset to learn and analyze the network behaviour of RATs, in order to propose new detections to protect our devices. The current version of the dataset includes 7 packet captures from 7 executed Android RATs. The Android Mischief Dataset was done in the Stratosphere Laboratory, Czech Technical University in Prague.

Execution Methodology

To create this dataset, we followed a methodology for each of the RATs. The methodology consists of the following 4 steps: (i) Installation, (ii) Execution, (iii) Traffic Capture, and (iv) Dataset Logging

1. **Installation.** This step consists of searching for the code of the RAT on the Internet, downloading it, installing an appropriate virtual machine for execution of the RAT’s controller, including all the library requirements on the virtual machine (e.g .NET Framework, JRE), and finally preparing the physical phone or phone virtual emulator as a victim to infect.

2. **Execution.** In this step we execute the downloaded RAT in these steps. First, use the Builder app in the Windows VM to create and build a new APK file. Second, start the RAT Controller in the Windows VM so it is ready to receive victims. Third, send the APK to the phone

3. **Traffic Capture.** When performing actions in the controller and the server, we capture the network traffic using our own VPN server, or in case of Android virtual emulator, we can use the computer network interface.

4. **Dataset Logging.** When performing actions in the client and the server, we also write a log file of the performed actions and take screenshots for each action in the Controller and the phone. As a result, each RAT in the dataset includes an APK file, a log file, screenshots files, a pcap file and a README.md.

Dataset files for each executed RAT

Each RAT of the dataset contains the following files:

1. **README.md** - This file is the generic description of the execution, containing the name of the executed RAT, details of the RAT execution environment, details of the pcap (client’s IP and server’s IP, time of start of the infection).

2. **APK** - The APK file generated by the RAT’s builder. Be aware that the APK was built for our own servers, so it can not be used in a real attack.

3. **log** - very detailed and specific time log of all the actions performed in the client and the server during the experiment, e.g “2020-08-11 10:20:21 controller: execute command ‘Take Photo - Back Camera’”. The purpose of this log is to let the researchers match the actions with the packets in the pcap.

4. **pcap** - network traffic of the whole infection. Sometimes captured on the host computer running the controller VM, sometimes using the Emergency VPN software.

5. **screenshots** - a folder with screenshots of the mobile device and controller while performing the actions on the client and the server.

Executed RATs and Download

The first version of the Android Mischief Dataset, v1, includes the following 7 RATs: Android Tester v6.4.6, DroidJack v4.4, HawkShaw, SpyMax v2.0, AndroRAT, Saefko Attack Systems v4.9 and AhMyth.

The Android Mischief Dataset can be downloaded in two ways, as one zip file containing all the RATs together, or it can be downloaded each RAT individually.

Download the whole Android Mischief Dataset all together as one zip file from here: https://mcfp.felk.cvut.cz/publicDatasets/Android-Mischief-Dataset/

To download each RAT execution individually:, use these links

RAT01 - Android Tester v.6.4.6 [download here]

RAT02 - DroidJack v4.4 [download here]

RAT03 - HawkShaw [download here]

RAT04 - SpyMAX v2.0 [download here]

RAT05 - AndroRAT [download here]

RAT06 - Saefko Attack Systems v4.9 [still not available]

RAT07 - AhMyth [download here]

Citation

If you are using this dataset for your research, please reference it as “Stratosphere Laboratory. Android Mischief Dataset v1. November 18th. Kamila Babayeva. https://www.stratosphereips.org/android-mischief-dataset”

Contacts

if you have any questions or you want the source code of RATs and their requirements, do not hesitate to contact kamifai14@gmail.com