**Final Project Report**

**Authors:**

Emaje Hall

Wyatt Babb

Xochitl Ryan

**Question of Interest:**

How does rebooting your system impact a ransomware infection?

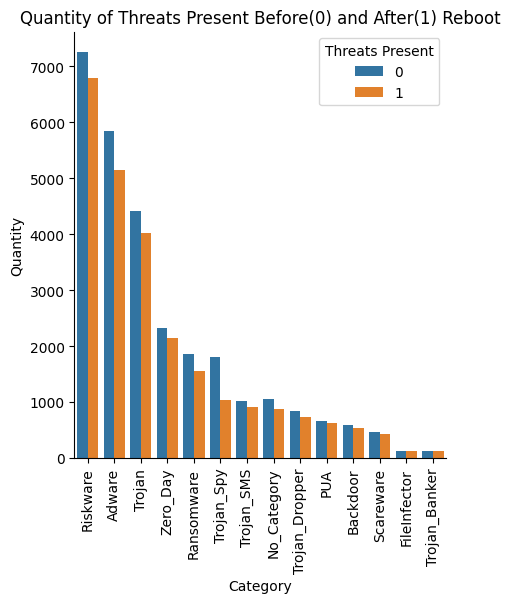
**Background:**

Android devices have dominated the global market since inception. Despite iOS popularity and dominance in the US, Android still retained 70.77% of the global market as of August 2023 (<https://finance.yahoo.com/news/ios-vs-android-market-share-135251641.html>). A captive audience of this proportion provides substantial opportunity for both hackers and criminals to infect, steal, and threaten large populations. One such method used is ransomware, a particular kind of malware that “[kidnaps] the victim’s data by encrypting it and threatening to deny access to it until a ransom is paid” (Young & Yung). Using the CCCS-CIS-AndMal2020 dataset provided by the University of New Brunswick allows for dynamic analysis measuring the change in ransomware execution before and after a system reboot. The dataset is “[comprised of] 12 Android malware categories and 147 malware families that are dynamically analyzed in an emulated environment.” Such parameters in the original data allow researchers to record the processes of malware that may attempt to conceal itself upon initial infection (David Sean Keyes et al.).

For ransomware in particular, this involves dividing each type of ransomware into one of 6 family’s, koler, lockscreen, congur, masnu, jisut, and slocker, and recording each process in emulation before and after the reboot. Each of the 6 family’s corresponds to a different style of execution the ransomware can perform. The following analysis looks at three different features, extracted from all 6 family’s, more intimately to measure how a system reboot affects ransomware execution. Total Processes is the count of each process that occurs in total, Battery Wakelock is a measure of each of the malware’s requests to keep the device awake, and Total API Sessions is a measure of how many different APIs are active over the course of emulation.

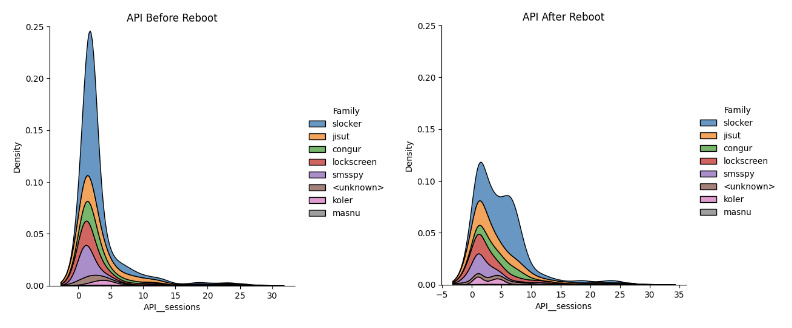
**Visualizations with source code and descriptions:**

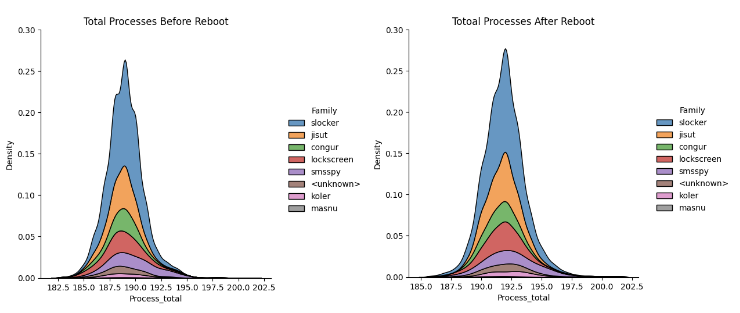
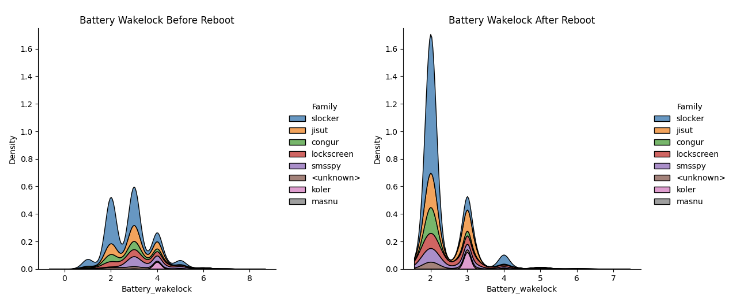
This visualization describes the overall number of threats to the system detected before and after the reboot



**Analysis (beyond plotting and summarizing the source data) or model, with source code and explanations:**

Given the dataset, we were mostly limited to count data and chose to use stacked kernel density estimate (KDE) plots to visualize the three features of the ransomware families: API sessions, battery wakelock, and total processes. Preference was given to KDE plots since they are easily interpretable in comparison to often-cluttered histograms. Likewise, Seaborn's visualization library was selected for its ease-of-use. By using the hue parameter of Seaborn’s distribution plot function, we could quickly visualize and compare the behavior of various ransomware families. With count data, it was important to observe the frequency of API sessions as ransomware generally relies on a system’s API to execute its tasks. Similarly, an abnormal increase in total processes would help indicate the presence of ransomware. Given the nature of this malware, it was also presumed that battery wakelock could aid in preventing ransomware processes from being interrupted – considering that some families do more than restrict access to files, but encrypt them as well. This method of analysis was thought to aid in understanding how ransomware is affected by a reboot process.





**Conclusion:**

The analysis of our data showed that in most comparisons of the feature categories there were less activities performed by the ransomware after the reboot of the system, this general trend can be seen in the visualization of the total dataset. One example of this trend in the extracted features is shown in the first figure of App Programming Interface interactions. This trend could suggest that the effectiveness of the malware is negatively impacted by a reboot of the system. However, our analysis of the Battery Wakelock feature category displayed direct opposition to this trend, with more access to battery wakelock and services by malware after the reboot than had occurred beforehand. Additionally, the Total Processes feature category showed little to no impact from the reboot of the system. Altogether this would suggest that the interactions between the malware and the system generally decreasing after a reboot is due to the ransomware already completing the majority of its processes before the reboot, not the reboot being able to stop them from occurring at all. Further, this analysis suggests that rebooting the system is more detrimental to the system in the case of a ransomware infection, as a reboot allows the malware to complete processes that otherwise would have remained dormant.

**Possible Sources of Bias:**

Addressing some sources of bias in this data, there is possibly a bias introduced to the data via the ML method of utilizing classifiers to classify malware categories and families. It is disclosed in the source of our data set that, within the ransomware data, 50 samples of the congur family were incorrectly classified as a part of the slocker family. (David Sean Keyes et al.) This is important to keep in mind in our analysis. However, it is also important to note that there were 998 total slocker family samples and 252 total congur family samples taken.

References:

Simoiu, C., Goel, S., Gates , C., & Bonneau, J. (n.d.). “I was told to buy a software or lose my ... - Stanford University. https://web.stanford.edu/~csimoiu/doc/2019\_ransomware\_SOUPS.pdf

David Sean Keyes, Beiqi Li, Gurdip Kaur, Arash Habibi Lashkari, Francois Gagnon, Frederic Massicotte, "EntropLyzer: Android Malware Classification and Characterization Using Entropy Analysis of Dynamic Characteristics", Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS), IEEE, Canada, ON, McMaster University, 2021

Abir Rahali, Arash Habibi Lashkari, Gurdip Kaur, Laya Taheri, Francois Gagnon, and Frédéric Massicotte, "DIDroid: Android Malware Classification and Characterization Using Deep Image Learning", 10th International Conference on Communication and Network Security (ICCNS2020), Pages 70–82, Tokyo, Japan, November 2020

*UNB*. University of New Brunswick est.1785. (n.d.). https://www.unb.ca/cic/datasets/andmal2020.html

A. L. Young and M. Yung, "On Ransomware and Envisioning the Enemy of Tomorrow," in Computer, vol. 50, no. 11, pp. 82-85, November 2017, doi: 10.1109/MC.2017.4041366.

<https://github.com/xochi7l/hello-world/blob/main/Final_Project_Analysis.ipynb>