Security Metrics - Review of group 2

Reviewed by group 14

Summary

This research investigates malicious apps in two popular chinese app stores, Baidu and 360. In specific, it focuses on children between 8 and 15 as a target group of these kind of attacks. At first, a set of ideal metrics is determined in order to assess the current threat landscape and the risks for the users of such app stores. Next, some metrics, related to the user's risk and the incident response life cycle, that are used in practice are stated. In the methodology, the processing of the data set is described and what metrics are derived from it. Lastly, in the evaluation, the two app stores are compared regarding the amount of malicious apps per category and the danger level emitted by these apps. It was found that the risk of downloading a malicious app in the Baidu app store was lower than in the 360 app store.

Strengths

- Assumptions: We are very happy to see which assumptions you made. They are clearly stated stated and this is useful to better understand what you did in de rest of the paper.
- Ideal Metrics: The ideal metrics are useful and clearly stated. You gave enough motivation on the metrics you choose. However we would like to see some more out of the box thinking for the ideal metrics. We would have liked see some examples that might not be measurable (to easily). For example think of that you would like to know who exploited the safety issues in the app.
- Metrics in Practice: These are again clearly stated. As well you clearly state that you made use of publications. However we would prefer a writing style more looking like a literature review. "For example Kikuchi et. al. (2016) state…"
- Data Cleaning and processing steps: It is good that this is included so we can see what you did with the data. This makes it easier as well to repeat a comparable research on new data.
- Limitations: It is nice to see that you critically assessed your own work and show what points could be improved in future work.

Major issues

- Assumptions: Assuming that all apps older than 50 days are malicious is too general. While this is mentioned in the limitations section, next to getting a lot of false positive using this method, the app's age could have been used as a security metric as well. While incomplete, the inclusion of malware databases and other kinds of open source intelligence as metrics for malice of apps possibly would have been a good alternative for this assumption. Besides this it would be nice to see why you choose

50 days. Maybe there is some literature on which you made this decision. Please elaborate on why 50 days is a ‘good’ choice.

- Victim/Actor + Security issues: 8-15 y.o. children is indeed a very interesting target group. However, it is not mentioned what specific security issues the target group is exposed to. We would like to see some motivation on why you chose this target group and why they are so interesting indeed. Furthermore, the target group is not really being taken into account in the sections about metrics except for focusing on gaming related apps in later sections.
- Metric: There is no weighting of how malicious the apps are in defining how dangerous a platform is. We would like to advise you to introduce some kind of weighting in the calculation of the dangerousness. You could also think about different weighting for different actors. For example, loss of data due to a malicious app might be a bigger problem for an entrepreneur than for a child.

Minor issues

- Introduction: It would be nice to combine this with assumptions, actor and security issue in a continuous story. This might be a good opportunity to introduce the data you are going to use as well.
- Actors + Metrics in practice: Who is going to use these metrics and for what purpose? Some of the metrics, like incident response and recovery times, may be difficult to assess for the companies behind the app stores and government actors, as these issues are not always made public.
- Evaluation: Comparing subcategories of gaming related apps is possibly too specific to make accurate conclusions, since there might not be any preference of subcategory on the side of the criminal, and metrics about the malice of apps may be not sufficiently accurate, as mentioned above. It may also be difficult to make use of this data as you cannot realistically tell the parents and children to be wary of, for instance, racing games. However, a comparison of major app categories could give interesting insights into the target groups and attack methods of criminals.