


Code Metrics For Predicting Risk Levels of Android Applications

Akond A. Rahman

North Carolina State University, arahman@ncsu.edu

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/ccerp>

 Part of the [Information Security Commons](#), [Software Engineering Commons](#), and the [Technology and Innovation Commons](#)

Rahman, Akond A., "Code Metrics For Predicting Risk Levels of Android Applications" (2016). *KSU Proceedings on Cybersecurity Education, Research and Practice*. 1.

<https://digitalcommons.kennesaw.edu/ccerp/2016/Student/1>

This Event is brought to you for free and open access by the Conferences, Workshops, and Lectures at DigitalCommons@Kennesaw State University. It has been accepted for inclusion in KSU Proceedings on Cybersecurity Education, Research and Practice by an authorized administrator of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

Abstract

Android applications pose security and privacy risks for end-users. Early prediction of risk levels that are associated with Android applications can help Android developers in releasing less risky applications to end-users. Researchers have showed how code metrics can be used as early predictors of failure prone software components. Whether or not code metrics can be used to predict risk levels of Android applications requires systematic exploration. The goal of this paper is to aid Android application developers in assessing the risk associated with developed Android applications by identifying code metrics that can be used as predictors to predict two levels of risk for Android applications. In this exploratory research study the author has investigated if code metrics can be used to predict two levels of risk for Android applications. The author has used a dataset of 4416 Android applications that also included the applications' 21 code metrics. By applying logistic regression, the author observes two of the 21 code metrics can predict risk levels significantly. These code metrics are functional complexity and number of directories. Empirical findings from this exploratory study suggest that with the use of proper prediction techniques, code metrics might be used as predictors for Android risk scores successfully.

Disciplines

Information Security | Software Engineering | Technology and Innovation

SUMMARY

Android applications can pose security and privacy risks for end-users. Early prediction of risk levels that are associated with Android applications can help Android developers in releasing less risky applications to end-users. Researchers have showed how code metrics can be used as early predictors of failure prone software components. Whether or not code metrics can be used to predict risk levels of Android applications requires systematic exploration. The goal of this paper is to aid Android application developers in assessing the risk associated with developed Android applications by identifying code metrics that can be used as predictors to predict two levels of risk for Android applications. This research paper focuses on answering the following research question to achieve this goal: RQ: What code metrics can be used as significant predictors to predict different levels of risk for Android applications?

In this paper the author has evaluated how code metrics such as number of lines, and McCabe's complexity can be used as predictors to predict multiple level of risk for Android applications. In this exploratory research study the author has investigated if code metrics can be used to predict two levels of risk for Android applications. The author has used a dataset of 4416 Android applications that also included the applications' 21 code metrics. By applying logistic regression, the author observes two of the 21 code metrics can predict risk levels significantly. These code metrics are functional complexity and number of directories.

In this exploratory research study the author has observed when all code metrics are combined in a logistic regression model, two specific code metrics are significant predictors. The author acknowledges that one type of regression model is not sufficient to predict multiple levels of risk for Android applications. The author observes the opportunity for future research in this direction, e.g. using other techniques such as principal component analysis to determine the significant predictors. Furthermore, the author observes the scope of applying a wide range of statistical learners to correctly predict the levels of risk. Sampling techniques such as over and under sampling can also be used to evaluate the performance of such prediction models.

Empirical findings from this exploratory study suggest that with the use of proper prediction techniques, code metrics might be used as predictors for Android risk scores successfully. The author observes the opportunity of applying statistical learning techniques on the used dataset as well as other data sources.