Chief Investigator: Richard Greene
School of Science / Edith Cowan University
Edith Cowan University
270 Joondalup Drive
JOONDALUP WA 6027
Phone: +61 437708729
Email: rgreene0@our.ecu.edu.au

Participant Information Letter

Project title: Software supply chain protection using statistical analysis of development behaviour
Approval Number: 2021-03052-GREENE
Principal Investigator: Richard Greene, Mike Johnstone (Supervisor)

An invitation to participate in research

Your project has been selected to participate in a project titled "Software supply chain protection using statistical analysis of development behaviour" which seeks to investigate if it is possible to classify GitHub code commits that fall outside normal development patterns. Once classified as an outlier commit, this may indicate that the commit requires additional security. You are being notified because you may have previously submitted code to one of the selected GitHub repositories under investigation. The project code you submitted, including the metadata will be forked and used in the creation of a predictive statistical model. This research project is being undertaken as part of the requirements of a PhD at Edith Cowan University, Western Australia.

Please read this information carefully. Ask questions about anything that you do not understand or want to know more about. Before deciding to opt-out, you might want to talk about it with a relative or friend.

If you decide you wish to opt-out of taking part in this research project, you can contact the researcher directly via email on rgreene0@our.ecu.edu.au before the **8th May 2022**, providing your GitHub username and stating that you do not wish to participate. By failing to respond you are telling us that you:

- Understand what you have read;
- Consent to your commit meta-data being used in the research project;
- Consent to the use of your commit meta-data as described.

What is this project about?

This project aims to determine if it is possible to create a statistical model that can classify the normal workflow of a developer committing updates to open-source projects. By tracking the GIT commit meta-data recorded when code is uploaded to GitHub during the development process, patterns of practice for developers may be determined over time. This may include working at specific times of the day, working in specific folders of the project, etc. Classifying changes using the commit cycle may allow for changes which fall outside this normal pattern, to be flagged as requiring additional security review. In this way it is hoped, impersonation of developers or the deliberate introduction of malicious changes will be reduced. The project is funded by the researcher independently and without the need for any external investment or stakeholder.

What does my participation involve?

Your participation in this research project will not involve any direct activities as the content has already been generated and recorded within the GIT repository. As per the GitHub terms of service section 5 “Any User-Generated Content you post publicly, including issues, comments, and contributions to other Users' repositories, may be viewed by others. By setting your repositories to be viewed publicly, you agree to allow others to view and "fork" your repositories (this means that others may make their own copies of Content from your repositories in repositories they control)”.

The research activities will involve forking the current project and analyzing the code changes which you and the other contributors may have made. All data indicating which user made the change or the project involved will be anonymized.

Do I have to take part in this research project?

Your participation in this research project is assumed from Section 7 of the GitHub Acceptable Use Policy. If, however you do not wish to take part, you do not have to and can opt-out. You are free to withdraw from the project at any time before **8th May 2022**, by informing the researcher via email to rgreene0@our.ecu.edu.au, providing your GitHub username and stating that you do not wish to participate. Withdrawal will not be possible after this date, however as all data will be anonymized, it will not be possible to identify which if any of your data was used, or identify the project on which the data was captured.

Your decision to opt-out will not affect your relationship with the research team.

Your privacy

By taking part you consent to the research team collecting and using GIT commit meta-data about you for the research project. Any information obtained in connection with this research project that can identify you will remain confidential. All the information will be de-identified, this will be accomplished by using a one-way hashing algorithm for the GitHub user identifier and the project name. Your information will only be used for the purpose of this research project and will only be disclosed as required by law.

It is anticipated that the results of this research project will be published and/or presented in a variety of forums. In any publication and/or presentation, information will be provided in such a way that you cannot be identified, except where requested for specific reasons at which point you will be asked to provide written consent.

In accordance with relevant Australian and/or Western Australian privacy and other relevant laws, you have the right to request access to the information about you that is collected and stored by the research team. You also have the right to request that any information with which you disagree be corrected. Please inform the research team member named at the end of this letter if you would like to access your information.

All data collected will be kept in accordance with ECU’s Data Management Policy. Electronic data will be stored on a secure Microsoft SharePoint site provisioned by ECU’s IT Services. All records will be stored as required in ECU’s Records Management Policy. The data will be retained for 7 years and destroyed, if appropriate at the end of the retention period. Data will be de-identified when stored and at the end of the retention period, the data will be destroyed, if appropriate under the State Records Act.

Possible Benefits

This research may not provide benefit to you personally but may provide benefits for people that use the open-source project you have been involved with, in the future. This research may assist others avoiding malicious software.

Possible Risks and Risk Management Plan

There are no known risks to participating in this research project.

What happens when this research study stops?

We intend to publish our results in research journals and present them at research conferences locally and nationally. Your GitHub Username or any identifying information will not be included in any of the publications or presentations.

Has this research been approved?

This research project has received the approval of Edith Cowan University's Human Research Ethics Committee, in accordance with the National Health and Medical Research Council's *National Statement on Ethical Conduct in Human Research 2007 (Updated 2018)*. The approval number is 2021-03052-GREENE.

Contacts

If you would like to discuss any aspect of this project, please contact the following people.

Chief Investigator
Richard Greene
Student
Edith Cowan University
E: rgreene0@our.ecu.edu.au

Supervisor
Mike Johnstone
Professor
Edith Cowan University
P: +61 8 6304 6615
F: +61 8 9370 6100
E: m.johnstone@ecu.edu.au

If you have any concerns or complaints about the research project and wish to talk to an independent person, you may contact:

Independent Person
Research Ethics Support Officer
Edith Cowan University
P: +61 6304 2170
E: research.ethics@ecu.edu.au

If you do not wish to participate in this research, please email the researcher on **rgreene0@our.ecu.edu.au** before the **8th May 2022**.

Sincerely,

Richard Greene

Richard Greene

Chief Investigator