Data Article

Data on security requirements in open-source software projects

Wentao Wang\textsuperscript{a}, Kavya Reddy Mahakala\textsuperscript{a}, Arushi Gupta\textsuperscript{a}, Nesrin Hussein\textsuperscript{a}, Yinglin Wang\textsuperscript{b}

\textsuperscript{a} Department of Electrical Engineering and Computer Science, University of Cincinnati, USA
\textsuperscript{b} Department of Computer Science and Technology, Shanghai University of Finance and Economics, China

A R T I C L E   I N F O

Article history:
Received 7 November 2018
Received in revised form 5 December 2018
Accepted 10 December 2018

Keywords:
Security requirements
Open-source software

A B S T R A C T

The datasets presented in this article are related the research paper entitled “A Linear Classifier Approach for Identifying Security Requirements in Open Source Software Development” Wang et al. (2018) [1]. This article describes requirements collected from three open-source software (OSS) projects and labels of security requirements. The datasets are made available to support automated security requirements analyzing tools development as well as tools’ evaluation.

© 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications Table

<table>
<thead>
<tr>
<th>Subject area</th>
<th>Computer: Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>More specific subject area</td>
<td>Requirements Engineering</td>
</tr>
<tr>
<td>Type of data</td>
<td>Figures, spreadsheets, and text file</td>
</tr>
<tr>
<td>How data was acquired</td>
<td>Extracted by using REST API and manually analysis</td>
</tr>
<tr>
<td>Data format</td>
<td>Filtered and analyzed</td>
</tr>
<tr>
<td>Experimental factors</td>
<td>Two researchers collected requirements from three open-source software projects, systematically analyzing all requirements, and manually labeled all security requirements.</td>
</tr>
</tbody>
</table>
## Experimental features

### Data source location

University of Cincinnati, Cincinnati, OH, USA

### Data accessibility

The original data are stored in projects' issue tracking systems, following are URLs of those issue tracking systems:

- **Drools**: [https://issues.jboss.org/projects/DROOLS/](https://issues.jboss.org/projects/DROOLS/)
- **GeoServer**: [https://osgeo-org.atlassian.net/projects/GEOS/](https://osgeo-org.atlassian.net/projects/GEOS/)

## Related research article


## Value of the data

- The goal of publishing our datasets is to support automated security assessment tools building as well as evaluation. Our datasets can be used in following research topics.

- Security requirements identification/classification tool building: Our datasets record security requirements from three OSS projects as well as non-security requirements. They can be used to help conduct textual information based automated tools for security requirements identification.

- Security requirements tracing/identification approach evaluation: We label security requirements in three OSS projects. This labeled data can be used to evaluate different security requirements tracing and identification approaches.

## 1. Data

There are three datasets provided with this article: Axis2, Drools, and GeoServer. In each project, we have a spreadsheet called “Requirements.xlsx” which records all requirements in this project. For each requirement, we record its id, title, descriptions, and an URL which link to its original data saved in issue tracking systems. We also saved all security requirements of this dataset in a text file named as “SecReq.txt”. Whether a requirement is a security requirement or not is judged by two security experts. We briefly describe the data collecting and judging process in the next section.

See also Figs. 1 and 2.

## 2. Experimental design, materials and methods

Original requirements of three OSS projects are saved in projects' JIRA issue trace systems. JIRA REST API [2] is used to download requirements to the local database. Fig. 1 presents a subset of requirements we collected from AXIS2. For each requirement, four types of information are recorded in our datasets. The first column “JIRA_KEY” records requirements' IDs which can be used to search and trace security requirements in the issue track system. Columns “Title” and “Description” parts record textual information of each requirement. We also provide URL of each requirement in which researchers can collect more information about this requirement for their own studies.

Two researchers with strong security background (i.e., they have at least one year working experience on the security requirements), then, systematically judge whether a specific requirement is a security requirement. If both of them have the same judgment (i.e., this requirement is a security requirement), attribute “JIRA_KEY” of this requirement is saved in the result TXT file. Otherwise, a joint meeting will be conducted to make final decisions. According to Cohen's Kappa analysis, two analysts achieved almost perfect agreement. Fig. 2 presents a sub-list of security requirements in
AXIS2. The full lists of security requirements in three OSS projects can be found in Supplementary material files.

Funding source

The work is funded in part by the U.S. National Science Foundation (Award CCF 1350487) and the National Natural Science Foundation of China (Fund No. 61375053).

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.12.029.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.12.029.
References
