Identification of the Impacts of Code Changes on the Security of Software

Moataz Abdelkhalek
Iowa State University, moataz@iastate.edu

Ameerah-Muhsina Jamil
Iowa State University, amjamil@iastate.edu

Lotfi ben Othmane
Iowa State University, othmanel@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/ece_conf

Part of the Electrical and Computer Engineering Commons, Information Security Commons, and the Software Engineering Commons

Recommended Citation
https://lib.dr.iastate.edu/ece_conf/78

This Conference Proceeding is brought to you for free and open access by the Electrical and Computer Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering Conference Papers, Posters and Presentations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Identification of the Impacts of Code Changes on the Security of Software

Abstract
Companies develop their software in versions and iterations. Ensuring the security of each additional version using code review is costly and time consuming. This paper investigates automated tracing of the impacts of code changes on the security of a given software. To this end, we use call graphs to model the software code, and security assurance cases to model the security requirements of the software. Then we relate assurance case elements to code through the entry point methods of the software, creating a map of monitored security functions. This mapping allows to evaluate the security requirements that are affected by code changes. The approach is implemented in a set of tools and evaluated using three open-source ERP/E-commerce software applications. The limited evaluation showed that the approach is effective in identifying the impacts of code changes on the security of the software. The approach promises to considerably reduce the security assessment time of the subsequent releases and iterations of software, keeping the initial security state throughout the software lifetime.

Keywords
Software security, security assurance case, incremental development, code change impact

Disciplines
Electrical and Computer Engineering | Information Security | Software Engineering

Comments

Rights
© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Identification of the Impacts of Code Changes on the Security of Software

Moataz Abdelkhalek, Ameerah-Muhsina Jamil, and Lotfi ben Othmane

Iowa State University, Ames, IA, USA

Abstract—Companies develop their software in versions and iterations. Ensuring the security of each additional version using code review is costly and time consuming. This paper investigates automated tracing of the impacts of code changes on the security of a given software. To this end, we use call graphs to model the software code, and security assurance cases to model the security requirements of the software. Then we relate assurance case elements to code through the entry point methods of the software, creating a map of monitored security functions. This mapping allows to evaluate the security requirements that are affected by code changes. The approach is implemented in a set of tools and evaluated using three open-source ERP/E-commerce software applications. The limited evaluation showed that the approach is effective in identifying the impacts of code changes on the security of the software. The approach promises to considerably reduce the security assessment time of the subsequent releases and iterations of software, keeping the initial security state throughout the software lifetime.

I. Introduction

The economic impacts of cyber-security have led companies to adopt secure software development practices [1], including security assurance: ensuring the software fulfills its security requirements. The common approach to identify code-level vulnerabilities, such as buffer overflow, is to use code analysis techniques [2], which can parse large software in reasonable time. Logic flaws, such as broken authentication and broken access control, which account for 50% of the vulnerabilities [3], [4] cannot be detected by code analysis techniques.

Assume that we have a login function that uses the writeToFile method from class changedClass of Listing 1 and assume that a developer changes the value of PermissionState used in the FileIOPermission from Read to Unrestricted. The common security assessment tools do not detect the impacts of this code change because the change impacts the logic, not the structure. The security experts may not detect the problem using manual review of the security-critical code chunks, the login function code, because the change is not in the login function itself, but in one of its dependencies. Failure to detect the impact of this code change can lead to serious privilege escalation vulnerabilities.

Code changes are changes to the software code that may introduce new functionality to the software, fix existing defects, or adapt the software to changes in its environment, cf. [5]. These changes can break the security of the software. Currently, there are two common approaches to identify code changes that impact the security of a software. The first approach requires the senior developers to discuss the security impacts of each change request before making the change. This approach is frequently used, especially for open source software although it is time consuming [4]. The second approach is to have annotations in the code itself to mark the security aspects related to the given code portion. The annotations warn the developers who change the code of these components about the implication of these changes on the security of their software [6].

The used approaches are time-consuming and prone to omission and commission errors. First, they consider security aspects in a conceptual way and not as mechanisms. For instance, they consider the impact of code changes on access control and not how access control may fail due to code change. Second, the approaches rely on the understanding of the experts of the interactions between the software components; that is, on the architecture view the developers make based on their experience with the code of the software.

This paper investigates automated tracing of the impacts of code changes on the security of the given software. To this end, we (1) model the software code using the call graph technique, which represents the calling relationships between the functions/methods of the software, (2) model the security assurance of the software using security assurance cases, and (3) relate assurance case elements to the code through code entry points, which compose the attack surface of software.

The contributions of the paper are:

1) Propose an approach to trace the impact of code

Listing 1. An example of risky code change.
B. Call Graphs

A program call graph is a compile-time data structure that represents the run-time calling relationships among a program’s methods or functions [12], [13]. Each node of the graph represents a method/function and each edge that links two nodes represents a call relationship between these two methods/functions. Call graphs are commonly used in inter-procedural static analysis including compilers, verification tools and program analysis tools [14]. Programs often include pointers and conditions for methods/functions calls. The call graph tools consider such program complexities to ensure accurate construction of call graphs from programs [12].

Figure 1 shows an example of a simplified small section of a call graph obtained from analyzing one of our test applications. We observe in this example that method doPrivileged() is called by methods SystemClassLoader(), Updater(), and NSProviderCreator() and calls methods LoaderAction.run() and InetAddress.run().

III. Related Work

The common approaches for software security assurance focus on ensuring that the security properties are enforced at each version of the software. They have a significant cost associated with their use in time and resources [15], [4]. This led to the idea to investigate the evolution of security requirements and design of a software, with the assumption that the code of the software represents the updated requirements and design [16], Jurjens et al. [15], for example, extended the UML models with security annotations, called UMLSec, and investigated the use of design models to support the multiple evolution paths and the incremental verification process of the evolving models.

Bosu and Carver conducted an empirical evaluation of peer-code review of Open Source Software (OSS) to detect security vulnerabilities [17]. They developed a Java application to mine the text related to the security flaws from the code review comments and stored the mining results in a database. They query the database using specific keywords and inspect results manually. They found that code review indeed helps in identifying and removing security vulnerabilities. In another work, Bosu et. al. analyzed the security vulnerabilities that could be discovered by code review, identified the characteristics of vulnerable code, and the developers that wrote it [18]. They found that the common types of identified vulnerabilities are buffer overflow and cross-site scripting. In addition, they found that the number of line changes also affects the likelihood of vulnerabilities and modified files have higher likelihood to have vulnerabilities than the new ones.

Rapid Release Change (RCC) had attracted researchers in a way that frequent releases of new features are likely to increase the number of vulnerabilities and the discovery rate. Clark et al. [19] investigated the relation between RCC adapted by Firefox and the different security vulnerabilities. Firefox, which adapted RCC, releases a new version every 6 weeks. The authors found that vulnerabilities are disclosed in newer code
as often as the older code. The authors also found that frequent releases increase the time needed by attackers to learn the software code, which forces them to adapt their tools to cope with the changes.

Vanciu and Abi-Antoun mentioned in [6] that half of the security vulnerabilities are caused by the flaws in the architecture of the system. They proposed Scoria, a semi-automated approach to find architectural flaws that can be used during Architectural Risk Analysis (ARA). The main features of Scoria are the use of static analysis to extract from the code, the use of queries to assign security properties to objects and edges, and the use of dataflow to trace code from edges that the constraints highlight that potential architectural flaws exist. The results of the study prove Scoria ability to detect security vulnerabilities that cause architectural flaws. However, the focus was only on detecting information disclosure and tampering vulnerabilities.

Othmane et al. [20] proposed integrating security re-assurance into the agile software development processes to ensure the security of the developed software with each iteration. In addition, they demonstrated the use of the technique to iteratively develop security features that fulfill their security requirements [21]. The process helps, for example, to identify customer change requests that conflict with the security requirements of the iteration. In a subsequent work, they analyzed the impacts of incremental secure software development for an existing open source software, Zen Cart [4].

The existing solutions for security assurance in the literature require full reassessment or are ad-hoc, e.g., using keywords search, which can be ineffective. This paper proposes tracing the code changes to the security requirements that they affect, if any. This should reduce the cost of security re-assurance and maintain the security of the software throughout its lifetime.

IV. Approach

The approach hypothesizes that code changes could be traced to the security requirements. To identify the impacts of code changes on the security of software, we model (1) the software code, (2) the security requirements of the software and (3) the relationship between the software code and its security requirements. The entry points are the source of data that flow into software [22]. Thus, security requirements are strongly related to the entry points of the software, since attacks happen by sending or receiving data to the system [23] using channels, such as open ports, connection sockets, and input forms. Hence, if code changes could be tracked to entry points, having a code path from the given entry point to the changed code portion, would indicate if the change is security-related. A security-related change can then be flagged for intensive reviewing.

Several techniques are used to implement the proposed approach. First, we represent software code with call graphs, which represent the calling relationships between the functions/methods of the software. We used the open-source static code analysis tool T. J. Watson Libraries for Analysis (WALA) [24] to generate the call graph of a given software. For software analysis, WALA was chosen because it (1) produces low overhead in performance and time, (2) supports several programming languages including Java and JavaScript and (3) can be extended to support more languages.

Second, we model the security requirements of the software using security assurance cases. The assurance cases are used at the design phase of the software to specify graphically the security requirements of the software and choose the security mechanisms that address them. In addition, in the testing phase, it can be used in the security assessments of the software to verify the security of the software.

Third, the assurance case elements are manually related to the corresponding methods of the software. This mapping requires expertise knowledge of the software and its security requirements. In this phase, developers and security administrators should be involved to ensure correct and accurate mapping.

Fourth, as developers use version control systems such as Git [25] to trace code changes; to ensure alignment in our toolset; we used integration with the Git versioning system to track and detect code changes. Then, these changes were mapped to corresponding functions and classes.

These four main steps allow us to detect and map the code changes to the assurance cases that reflect the security requirements of the given software. As depicted by Algorithm 1, the DetectSuspiciousChanges mechanism monitors the storage of the software for new versions using Git, detects the code changes, and identifies the methods/functions that the changes

<table>
<thead>
<tr>
<th>Algorithm 1 Identify code changes that impact the security of the software</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
</tr>
<tr>
<td><strong>Input</strong></td>
</tr>
<tr>
<td><strong>Input</strong></td>
</tr>
<tr>
<td><strong>Output</strong></td>
</tr>
<tr>
<td><strong>procedure</strong></td>
</tr>
<tr>
<td>ChangeList ← IdentifyChangedFunctions()</td>
</tr>
<tr>
<td>TopNode ← getMonitoredNodes(FunctionAssuranceCaseMap)</td>
</tr>
<tr>
<td>while ChangeList ≠ ∅ do</td>
</tr>
<tr>
<td>changedNode ← ChangeList.pop()</td>
</tr>
<tr>
<td>PathsList ← getPathList(changedNode,TopNode)</td>
</tr>
<tr>
<td>if (PathsList ≠ ∅) then</td>
</tr>
<tr>
<td>SuspiciousCodeChanges ← SuspiciousCodeChanges ∪ changedNode</td>
</tr>
<tr>
<td>Return SuspiciousCodeChanges</td>
</tr>
<tr>
<td>End</td>
</tr>
</tbody>
</table>
belong to. Then, it uses the call graph to identify the paths that link the identified methods/functions to the ones associated with the security assurance claims of the software. Changes related to assurance cases are flagged as suspicious changes that require comprehensive review to ensure they do not break the security requirements of the software.

V. SYSTEM OVERVIEW - SECURE CODE CHANGE ANALYSIS TOOLSET

We developed a set of tools to assist developers and security experts to create call graphs, design security assurance cases, and map the security claims and evidence parts of the assurance case to the related entry point methods of the given software. Figure 2 shows the interactions between the components of the approach, as discussed in the previous section.

We developed a graphical tool to design security assurance cases for the software [26] as an Eclipse plug-in. The tool supports adding and removing security claims, arguments or rational, evidence sets, and strategy nodes to a security assurance case, storing the assurance case in an XML format for further use. In addition, we developed a module that creates the call graph of the given software using WALA [24] and produces a call graph file in the DOT format that can be later parsed using graphical tools. With the help of the graphical call graphs, the previously developed assurance case elements could be associated to corresponding methods/functions of the software marking them as monitored nodes in the call graph. To automatically detect changes, we also developed a module that continuously queries the software Git repository to identify the delta cumulative changes of any two versions of a given software.

Finally, we developed a Code Tracking Algorithm that searches all the possible paths between the detected changed nodes and the monitored nodes; the ones associated with the security assurance cases. Then, the changed methods included in the identified paths are flagged as suspicious and the claims associated with the related monitored methods are flagged as potentially broken and should be reviewed manually.

VI. EVALUATION OF THE PROPOSED APPROACH

We selected three software applications to evaluate our approach, which are: Compiere,[2] Shopizer[3], and OpenZ. The used selection criteria are: (1) written in Java, (2) be an open-source security-sensitive software, (3) completeness of the software source code, (4) having recent and regular updates, and (5) have high frequency of downloads. Table II provides the details of the selected software applications.

We applied the workflow of Figure 2 for the three software applications. In the first phase, the initiation phase, we created the security assurance cases of the software based on the Payment Card Industry (PCI) [28] and the OWASP top 10 requirements [29]. These choices were made because the chosen software applications are financial and the original software security requirements were not available. Figure 3 shows the security assurance case for the safe authentication and session management of Compiere application. Next, we created

---

2https://sourceforge.net/projects/compiere/
3https://github.com/shopizer-ecommerce/shopizer
4https://sourceforge.net/projects/openz/
TABLE I
List of selected open-source software.

<table>
<thead>
<tr>
<th>Application</th>
<th>Lines of Code</th>
<th>Function Calls</th>
<th>Classes</th>
<th>Avg Methods/Class</th>
<th>Last Updated</th>
<th>Last version</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenZ</td>
<td>172793</td>
<td>846364</td>
<td>3456</td>
<td>19</td>
<td>Sep 2018</td>
<td>3.5.0</td>
</tr>
<tr>
<td>Compiere limited version</td>
<td>22059</td>
<td>4257</td>
<td>136</td>
<td>7</td>
<td>Nov 2016</td>
<td>1.3.0</td>
</tr>
<tr>
<td>Compiere Full version + Oracle XE</td>
<td>649571</td>
<td>154917</td>
<td>2289</td>
<td>15</td>
<td>Mar 2017</td>
<td>Community Edition 3.5</td>
</tr>
<tr>
<td>Shopizer</td>
<td>106858</td>
<td>26705</td>
<td>961</td>
<td>6</td>
<td>Dec 2018</td>
<td>2.3.0</td>
</tr>
</tbody>
</table>

![Fig. 3. Assurance Case for session management in Compiere application](image)

The call graphs for each of the software using software analysis module. Table II provides the performance of the code analysis and call graph generation module. After that, we associated the methods to their related security claims. Table III gives a simplified map that we developed for the session management claims for Compiere application. This phase resulted into a set of security-sensitive methods that the system needs to monitor.

The second phase of the workflow, as depicted by Figure 2, consists of capturing the impacts of code changes on the security of the software. To increase the number of test-cases, we applied the analysis on a large number of changes using different versions of the software applications, in addition to, intentionally injected code changes. Then, we applied the identification of code changes tool and the code tracking algorithm to identify whether the code changes are associated with any of the security-sensitive entry-point methods.

The results showed that the toolset could detect all the security related changes successfully. In addition, the overhead of using the toolset was very low, as it processes software that has a call graph with up-to 500 nodes in 1 seconds, up-to 1000 nodes in 20-50 seconds, and up-to 2500 nodes in 5 minutes. The call graph generated with the tool was filtered using keywords to select the security-sensitive portions of the code, to improve the performance of the analysis.

Although, the evaluation of the approach is limited to three software applications, we experimented with a good number of test-cases, which showed that the approach is effective in identifying the impacts of code changes on the security of the software. The approach promises to considerably reduce the security assessment time of the subsequent releases and iterations of software, keeping the initial security state throughout the software lifetime. We intend to improve the usability of the toolset and extensively evaluate it using more commonly used security-sensitive software.

![Fig. 4. Performance measurement of the call graph generation tool](image)

TABLE II
Performance measurement of the call graph generation tool.

<table>
<thead>
<tr>
<th>Application</th>
<th>No. of nodes</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenZ</td>
<td>846364</td>
<td>9 min 36 sec</td>
</tr>
<tr>
<td>Compiere limited version</td>
<td>4257</td>
<td>11 sec</td>
</tr>
<tr>
<td>Compiere Full version + Oracle XE</td>
<td>154917</td>
<td>2 min 14 sec</td>
</tr>
<tr>
<td>Shopizer</td>
<td>26705</td>
<td>20 sec</td>
</tr>
</tbody>
</table>
We developed a toolset that trace the impacts of code changes on the security of a given software and evaluated it using three open-source software applications. We found that the proposed approach successfully identifies the security impacts of code changes and proves our hypothesis. It needs, though, a thorough evaluation with commonly used software to assess its capabilities and limitations. The approach would help to considerably reduce the security code review time and the time that developers spend discussing the impacts of the proposed code changes on the security of the software.

ACKNOWLEDGMENT

The authors thank Mike Johnson and Jim McClurg from John Deere for the thorough discussions along the execution of the research. This project is funded by a grant from John Deere.

REFERENCES


TABLE III

<table>
<thead>
<tr>
<th>Assurance Case Element</th>
<th>Associated Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>User authentication credentials are protected in storage</td>
<td>util.Login.checkPermission()</td>
</tr>
<tr>
<td>No weak credentials used</td>
<td>security.AccessControl()</td>
</tr>
<tr>
<td>Strong session management</td>
<td>process.SessionStartAll()init(), process.SessionEndAll()main()</td>
</tr>
<tr>
<td>Secure transmission of Passwords, session IDs, and other credentials</td>
<td>security.PermissionsHash.add(), db.TestConnection()</td>
</tr>
<tr>
<td>Protection against automated attacks such as brute forcing</td>
<td>apps.Process.Ind(), process.SvrProcess(), processFactAckReset()</td>
</tr>
<tr>
<td>Strong effective registration, credential recovery and forgot password processes</td>
<td>DB.Oracle(), model.Registration()</td>
</tr>
</tbody>
</table>

VII. Conclusion

We developed a toolset that trace the impacts of code changes on the security of a given software and evaluated it using three open-source software applications. We found that the proposed approach successfully identifies the security impacts of code changes and proves our hypothesis. It needs, though, a thorough evaluation with commonly used software to assess its capabilities and limitations. The approach would help to considerably reduce the security code review time and the time that developers spend discussing the impacts of the proposed code changes on the security of the software.