Analysis of Authentication Code available on GitHub

BSc Thesis
1st Presentation

13 April 2021

Adrian Jörg
Existing Password Policies...
...may decrease security

Mohammad Ghafari
@ghafarii

Hey @LinkedIn, it seems you have two password policies (a bug?) in place. A password can include "many sequential numbers" when signing up, but not when resetting a password.

10:08 vorm. · 10. Apr. 2020 · Twitter Web App
Existing problems

- Fewer updates for duplicated code
- Less readable and understandable code
- Fewer test cases
- Impaired functionality
Where can we find code about policies?

GitHub!
Initial Exploration

Collected Features:

**Password Policy**
- Length
- Pattern

**Inconsistencies in Policies**
- Duplication
- Impaired Functionality

**File- & Functionnames**
- Sign Up
- Login
- Password Reset

**Libraries**
- Authentication
- Validation
- Encryption

**Metadata**
- Forks & Stars
- First & Last commit
Initial Results

Total: 90 Repositories

- matched criteria: 47%
- policy validation duplication: 10%
- multiple policies: 1%
- impaired functionality: 2%
- unrelated: 40%
Example

Sidenote: Multiple instances of bad hashing algorithms, possibilities of sql injections, usage of old libraries etc
How can we detect broken policies on GitHub?

CodeQL Queries!
CodeQL Queries

Free semantic code analysis engine by GitHub

codeql.github.com or lgtm.com

Supports

C/C++, C#, Go, Java, JavaScript, Typescript, Python

and many frameworks (e.g. ASP.NET, express...)
import javascript

from Function f

where f.getName().matches("$\%\%\%\%$")

select f.getName(), f.getParameter(0)
Current state

Password Policy
- Length
- Pattern

Inconsistencies in Policies
- Duplication
- Impaired Functionality

Filenames
- Sign Up
- Login
- Password Reset

Libraries
- Authentication
- Validation
- Encryption

Metadata
- Forks & Stars
- First & Last commit
Detector Example (1/3)

<table>
<thead>
<tr>
<th>Function</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handling Request(s)</td>
<td>contains</td>
</tr>
<tr>
<td></td>
<td>Sensitive Data</td>
</tr>
<tr>
<td></td>
<td>Data is a Password</td>
</tr>
</tbody>
</table>
Detector Example (2/3)

Function
Handling Request(s) calls

Function
Validating Policy

contains

Variable
Sensitive Data
Data is a Password

Library Validation
Pure JS Validation
Detector Example (3/3)

### Function Handling Request(s)
- Variable:Sensitive Data
- Data is a Password

### Function Validating Policy
- Library Validation
- Pure JS Validation

<table>
<thead>
<tr>
<th># of Results</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No password policy</td>
</tr>
<tr>
<td>1</td>
<td>Only one password policy</td>
</tr>
<tr>
<td>2+</td>
<td>Interesting Case</td>
</tr>
</tbody>
</table>
Interesting Case

Multiple Policies OR

```javascript
set passwordSchema = { required | min:4 }; set emailSchema = { required | email }; set nameSchema = { required }

function validateSignup(user) {
   return
   passwordSchema.validate(user.password) &&
   emailSchema.validate(user.email) &&
   nameSchema.validate(user.name)
}

function validateLogin() {
   return
   passwordSchema.validate(user.password) &&
   emailSchema.validate(user.email) &&
```
Next Steps

Implement the recognition of more validation libraries

Filter the results which use the same validation codebase in different functions

Apply to larger base of repositories

Make the query available to GitHub for automatic reviews
Summary

Initial Results

Total: 100 Repositories
- matched criteria: 45%
- policy violation duplication: 10%
- multiple policies: 1%
- impaired functionality: 2%

CodeQL Queries

Semantic code analysis engine by Github
codeql.github.com or lgtm.com

Supports
- C/C++, C#, Go, Java, Javascript,
  Typescript, Python
  and many frameworks (e.g. ASP.NET,
  express...)

Detector Example (3/3)

Next Steps

- Implement the recognition of more validation libraries
- Filter the results which use the same validation codebase in different functions
- Apply to larger base of repositories
  Make the query available to GitHub for automatic reviews