Suggesting Secure Implementation to Vulnerable Code Snippets on Stackoverflow.

ABSTRACT

1 INTRODUCTION

≪**Mazharul 1.0:** This is working..≫

```
  private byte[] encrypt(byte[] raw, byte[] clear) {
    ...
    Cipher cipher = Cipher.getInstance("AES");}
    // Cipher cipher = Cipher.getInstance("AES/CBC/PKCS5Padding");
    ....
    return encrypted
  }

```

Listing 1: A real code snippet taken from Stackoverflow. I want to build a tool which after analyzing the code snippet will highlight the part of the code that is insecure and suggest an alternative secure implementation as showed in the figure.

In this project I want build a static analysis tool which will achieve the following.

- Analyze the code snippets from Stackoverflow for identifying out which part of the code is vulnerable by showing warning signs to highlight that part of the code to the developer.
- When developer clicks on the warning sign a secure implementation while be shown to the developers. In case of failure of building generating a secure implementation, the tool will show insightful/helpful messages explaing why this part of the code is flaged as insecure.

The problem is interesting for two reasons.

Difficulty of writing crypto code securely. Writing/implementing cypto code securely is a diffculut task for programmers. Any potential bug in crypto code can lead to serious vulnerablities open for attackers. Even so unlike other code, crypto code can be insecure even if it works perfectly on traditional test-suite's input/output which is used only to prove the implementation correctness of the program.

Online platforms roles in spreading insecure code. Online programming discussion platforms such as Stack Overflow have a rich source of ready to use code snippets for software developers. It is the defac-to place where developers go to find solutions of their problems and turn to the community for answers to their problems. Insecure code snippets found on Stackoverflow itself is not a serious problem. However, Fischer et al. has shown that developers have a tendency to directly copy paste code form Stack Overflow [1]. Therefore there are chances that any insecure code snippets posted on Stackoverflow can potentially find it way into production level code. To make matters worse, Meng et al. [2] has showed that many accepted answers on Stackoverflow have seriously insecure code and often-times given by users having high reputation. This adds to the problem copy pasting vulnerable code from online platform and furthermore increases the chances of the insecure code snippet being trickled down to production level code. Unfortunately there is not state-of-the-art tool to analyzie if a code posted by developer on Stackoverflow is secure or not. In absense of such tool, Stackoverflow is potentially contributing as a major source of vulnerability in production level code.

A static tool which can identify which part of the code snippet is insecure and suggest secure alternatives can help stopping the flow of insecure code from Stackoverflow to production level code.

2 METHODOLOGY

2.1 Code Repair

- Syntax error: illegal characters (&gt;, &lt;, &amp;, ..., ""), match brackets, @Overide,
- Missing class and method names: wrapped the code snippets inside a class, called methods without class, package names.
- importing common crpto classes.

2.2 Converting code snippets to IR

- Have used PPA.
- Talk about Jimple / AST.

≪**Mazharul 2.0:** Add a number on how many code snippets you have been able to convert here..≫

2.3 Insecure Patterns

≪**Mazharul 2.0:** How did we come up with these insecure patterns? Say why this is not secure and show one line of example.≫

2.3.1 *AES default encryption mode ECB .*

2.3.2 *Insecure cryptographic hash.*

2.3.3 *Presence of AllHostNameVerifier.*

2.3.4 *Abuse of X509TrustManager Verifier Interface.*

2.3.5 *Absence of performing hostname verification.*

2.3.6 *Weak key length.*

2.3.7 *Static/constant/predictable keys/IV .*

2.3.8 *Turning of CSRF protection .*

2.3.9 *Explicitly seeding random number.*

2.4 Identifying insecure patterns

3 RESULTS

3.1 Data collection

We used dataset from two previous sources [1, 2]. Both of these dataset contain code snippets posted on Stackoverflow. Fisher et al. crawled 1,161 code snippets posted on Stackoverflow related to Andrioid Securuity [1]. They considered a code snippet related to Android security if the code snippets makes API calls to one of the security services such as Java cryptography, Java secure Communications, public key infrastructure X.509 certificates, and Java authentication - authorization services. The popular crypto libraries used by Andriod developers such as Bouncy Castle, SpongyCastle, Apache TLS/SSL, keyczar, jasypt, and GNU Crypto were also included.

Meng et al. extracted 503 code snippets from 22,195 Stackoverflow posts by filtering the posts based on votes, duplications, and absence of code snipeets [2]. In total our study is baded on the dataset by combining these two. Our dataset contains 1,664 code snippets. The timeline of these code snippets are from 2008-2017. ≪**Mazharul 3.0:** add some more info and some statistics≫

4 LIMITATIONS

5 DISCUSSIONS

6 RELATED WORK

7 FUTURE WORK AND CONCLUSIONS

REFERENCES

[1] Felix Fischer, Konstantin Böttinger, Huang Xiao, Christian Stransky, Yasemin Acar, Michael Backes, and Sascha Fahl. 2017. Stack overflow considered harmful? the impact of copy&paste on android application security. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 121–136.

[2] Na Meng, Stefan Nagy, Danfeng Yao, Wenjie Zhuang, and Gustavo Arango Argoty. 2018. Secure coding practices in java: Challenges and vulnerabilities. In *Proceedings of the 40th International Conference on Software Engineering*. 372–383.