Finding Security Issues in (Open Source) Software Repositories

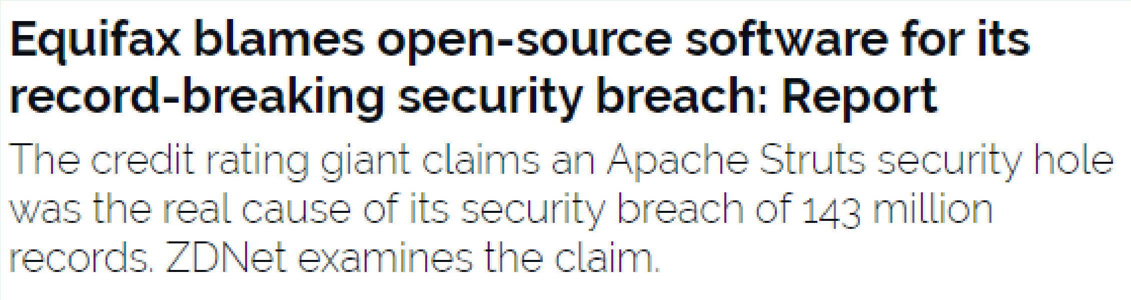
By

Zer Jun Eng

Supervised By

Dr. Achim Brucker

Motivation

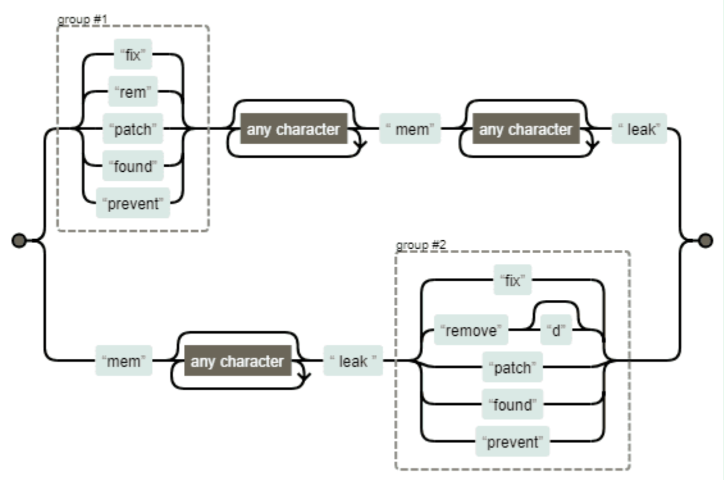


Equifax blames open-source software for its record-breaking security breach: Report

The credit rating giant claims an Apache Struts security hole was the real cause of its security breach of 143 million records. ZDNet examines the claim.

- Open source software are widely used as third-party components in both free and commercial projects
- Not all security vulnerabilities are published in CVE format
- Details of security vulnerability patches are not always publicly disclosed

Commit Message Matching



Regular Expressions

- Match the commit message with regular expression patterns
- Each vulnerability type has its own regular expresssion pattern

Vulnerable Code Searching

Files Changed

- Search for vulnerable code using static analysis
- Flawfinder for C/C++ source code
- Bandit for Python source code
- RegEx boundary search for Java source code

Results

- 1,514,726 commits analysed in 52 different repositories
- 780,725 potential vulnerability-fixing commits found
- On manual evaluation, 120 out of 271 commits were found to be positive
- True positive rate: **44.28%**

Conclusion

- Static analysis techniques are effective against common vulnerabilities
- Finding hidden security vulnerabilities is hard, especially when the security vulnerability is related to very specific part of code
- To improve the true positive rate, additional effort is required to refine the regular expression patterns and improve the analysis