An Analysis Of Security Discussions In Stack Overflow

Mitchell Drummer

Stack Overflow Introduction

- Stack Overflow was launched in 2008
- Place for developers to ask and answer questions about coding
- User asks a question, then selects one answer as the “accepted” one

Reputation System

- Users post answers, questions, or comments
- Users vote (upvote or downvote) posts
- Votes impact reputation of the one who posted them
  - Answers upvoted earn 10 points
  - Questions upvoted earn 5 points
  - Comments upvoted earn 2 points

Privileges Earned Through Reputation

- A user with 15 points can vote up posts
- A user with 50 points can comment on others’ posts
- A user with 125 points can vote down posts
- Users with at least 20K points are considered “trusted users”, and can edit or delete other people’s posts

Outline

- I wanted to learn more about security personally, that’s why I did this presentation
- Security discussions concerning Java library security
- Understanding the security community on SO
- Security of code snippets included in SO posts
- Conclusion

Secure Coding Practices In Java: Challenges and Vulnerabilities

- Written by Na Meng et al in 2018
- Goal: find developers’ programming obstacles, and insecure coding practices

Methods

- Analyzed 503 Java security posts on Stack Overflow
- Collected 22,195 posts, filtered them down to 503
- Filtered by removing unrelated answers, and ones without code
- Categorized into groups on next page

Post Distribution Across Categories

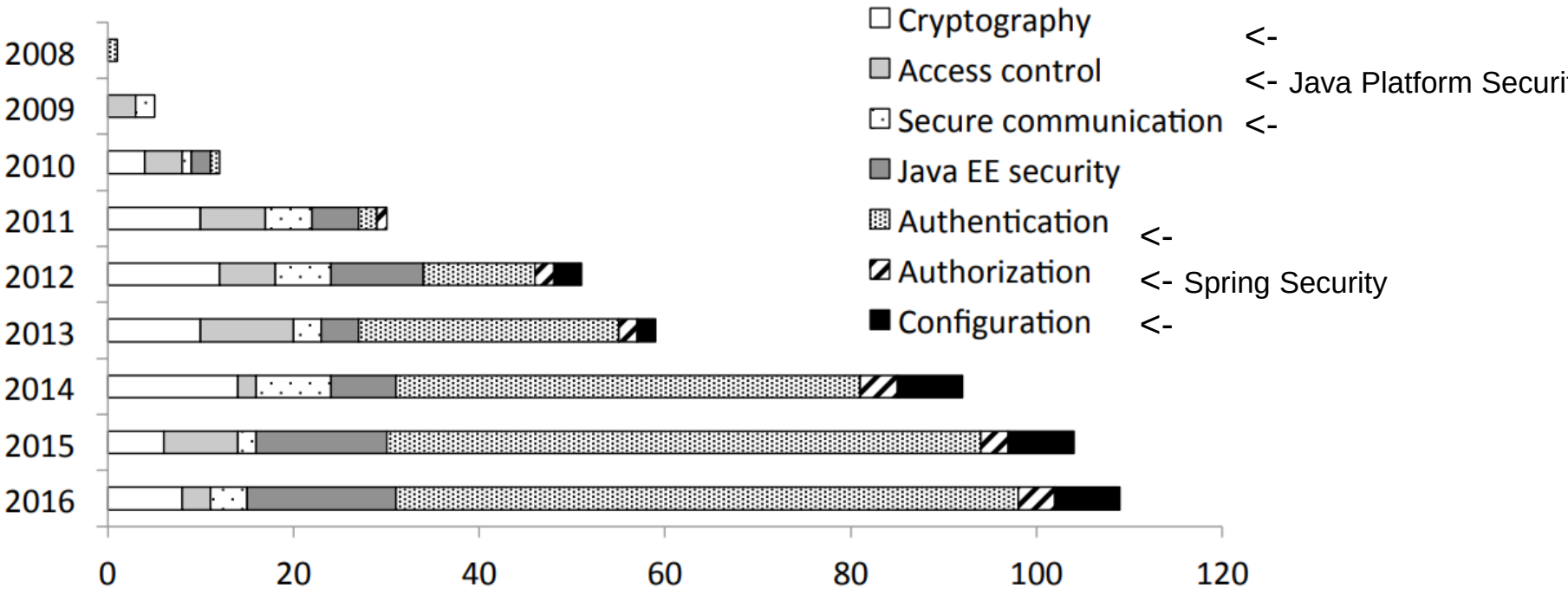


Figure 3: The post distribution during 2008-2016

Results

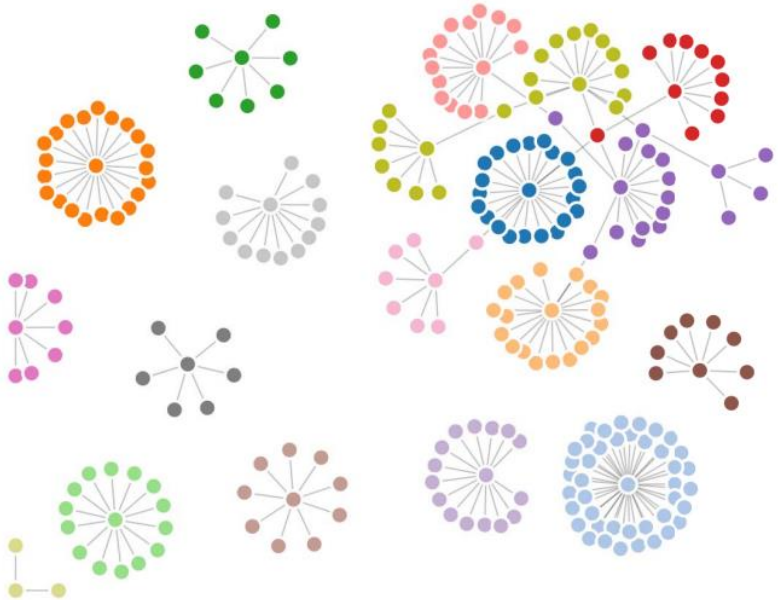
- 17 posts contained insecure code - 3%
- These 17 had a view count of 622,922 as of August 2017
- Spring Security specifically had many issues due to incomplete documentation, more than half the posts
- 9/10 SSL/TLS posts were about bypassing it
- 5/12 CSRF posts were about disabling it
- Misleading indicators: accepted answers, answers' upvotes, and responders' high reputation

An Anatomy Of Stack Overflow Posts

- Written by Tamara Lopez et al. in 2019
- “How do developers on Stack Overflow engage with one another when dealing with issues related to security?”

Methods

- Top 20 security questions, and their comments on the site were analyzed
- Top 20 selected by number of upvotes
- 250 Stack Overflow users made 364 comments
- 197 left a single comment
- 32 left two comments
- 10 left three comments
- 11 left more than three comments
- Low interconnectivity: most kept to one post



- A question is represented by the center dot in a cluster
- Each dot on the outside represents a user commenting once or more
- Lines between two different colored dots indicate a user commenting in both questions

Results

- This illustrates that a majority of commenters only answered a few questions about security
- 75% of users had answered fewer than 19 questions about security
- Two users answered 143 and 146 questions about security
- Not really a community specifically focused on security

How Reliable is the Crowdsourced Knowledge of Security Implementation?

- Written by Mengsu Chen et al. in 2019
- “In-depth investigation of the popularity of both secure and insecure coding suggestions on SO, and the community activities around them”

Methods

- Collected 3121 posts with code snippets
- 42.2% of snippets were insecure
- Some code snippets were duplicated
- Made 953 clone groups
- 587 duplicated secure clone groups
- 326 duplicated insecure clone groups
- 34.2% of groups were insecure

Results

- Insecure posts were more popular
  - More upvotes, views, and comments
- 34% of posts by trusted users were insecure
- No correlation between how many times a snippet was copied and security
- Insecure posts dominate the SSL/TLS category

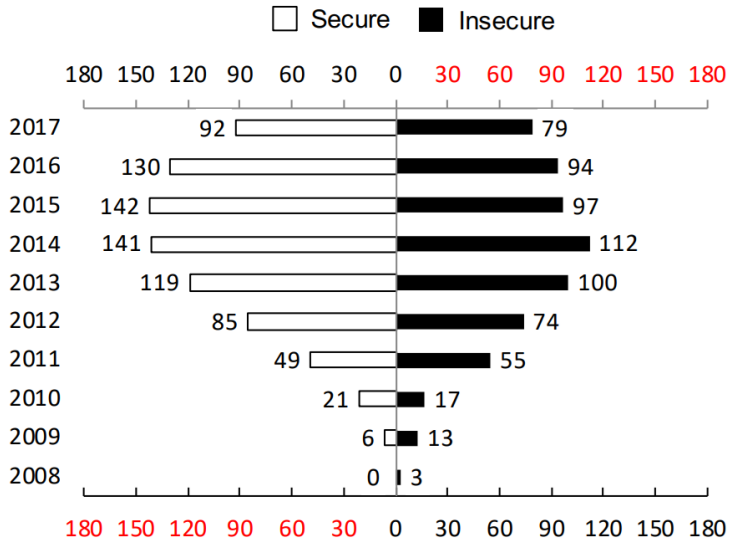


Fig. 4: The distribution of posts over during 2008-2017

- Insecure posts are more common the older they are on average

Recap

- Out of 503 posts, 17 insecure posts were present - 3%
- There is not a clear community around security
- Out of 3121 posts, 1319 were insecure - 42.2%

Overall Findings

- Code included in posts are not always secure
- SO posts are not updated as security evolves
- SSL is insecure and outdated - switch to TLS

My Advice Based On What I Learned

- Checking security is harder than checking if it runs
- Elsewhere, just having code work means your answer is acceptable
- Always keep security in mind while coding
- Many instances of “trust all certificates” or disabling authentication

Questions?

[1] Na Meng et al. “Secure Coding Practices in Java: Challenges and Vulnerabilities”. In: Proceedings of the 40th International Conference on Software Engineering. ICSE’18. Gothenburg, Sweden: Association for Computing Machinery, 2018, pp. 372–383. isbn: 9781450356381. Doi: 10.1145/3180155.3180201. url: https://doi-org.ezproxy.morris.umn.edu/10.1145/3180155.3180201

[2]TamaraLopez et al. “An Anatomy of Security Conversations in Stack Overflow”. In: Proceedings of the 41st International Conference on Software Engineering:Software Engineering in Society. ICSE-SEIS ’19. Montreal, Quebec, Canada: IEEE Press, 2019, pp. 31–40. doi: 10 . 1109 / ICSE - SEIS . 2019 . 00012. Url: https://doi-org.ezproxy.morris.umn.edu/10.1109/ICSE-SEIS.2019.00012

[3] Tamara Lopez et al. “An Investigation of Security Conversations in Stack Overflow: Perceptions of Security and Community Involvement”. In: Proceedings of the 1st International Workshop on Security Awareness from Design to Deployment. SEAD ’18. Gothenburg, Sweden: Association for Computing Machinery, 2018, pp. 26–32. isbn: 9781450357272. doi: 10.1145/3194707.3194713. url: https://doi-org.ezproxy.morris.umn.edu/10.1145/3194707.3194713