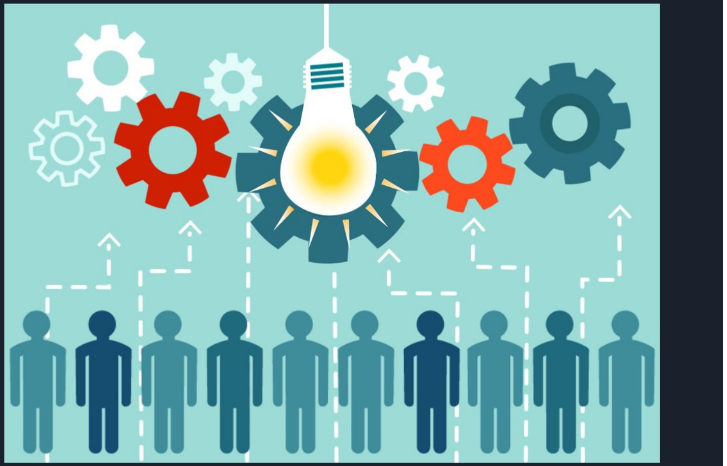
Security Flaws of Crowdsourced Knowledge

Will Anwara, Jacob Summers, Ana Haines,
Ben Nguyen, Jacob Thomas

What is Crowdsourcing?

- Crowdsourcing: Obtaining work, information or opinions from a large group of people who submit data
  - Posting answers on Stack Overflow, voting on and reviewing questions
  - Traffic apps encourage users to report accidents and incidents so that other app users are aware
- Code makes its way into many popular applications that millions rely on



Pros to Crowdsourcing

Scheduling

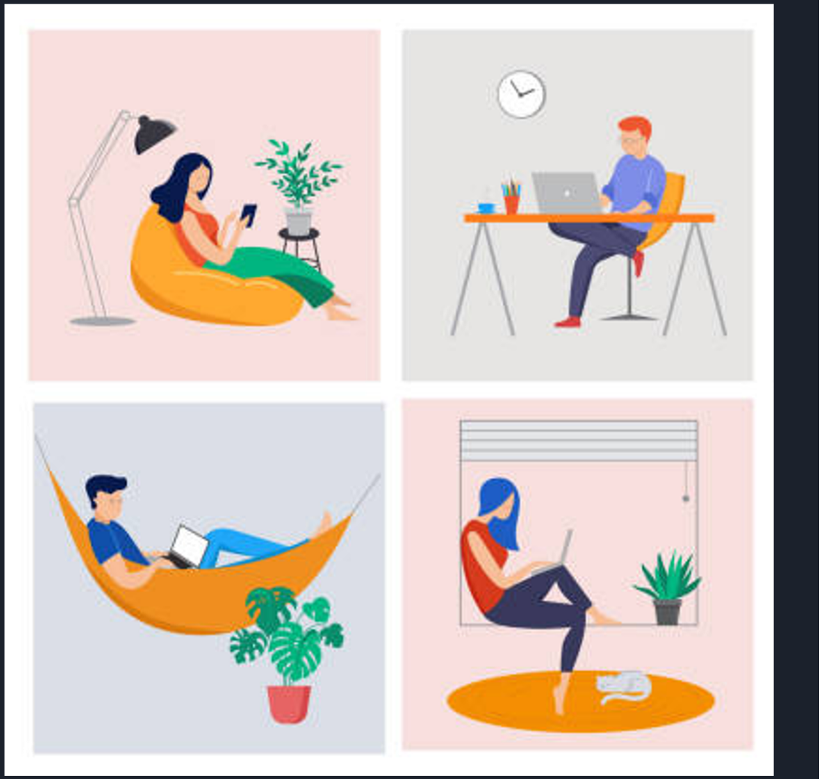
- Flexibility, different time zones, public incentives

Cost

- Reduce traditional costs (employee hiring, office space, etc.)

Quality

- Democracy as a filter
- Increased variety of talent = innovation



Case: Stack Overflow

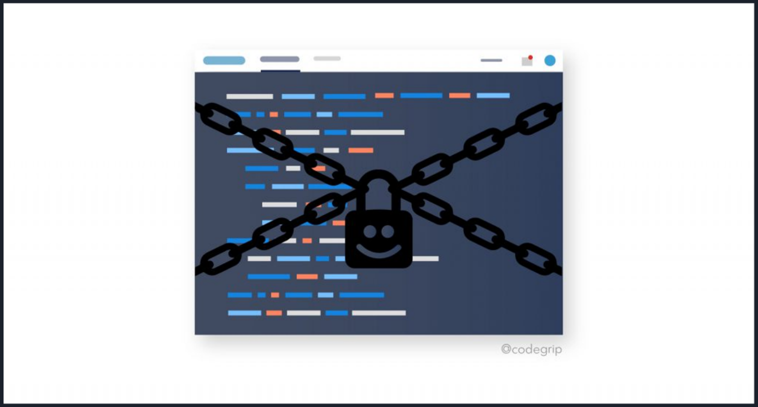
stack overflow

- Most popular site for developers to discuss coding problems and get feedback and answers

- Recent research shows that SO code commonly contains vulnerabilities related to security

- Many answers were unsafe and were ridden with security vulnerabilities

Secure vs. Insecure Statistics

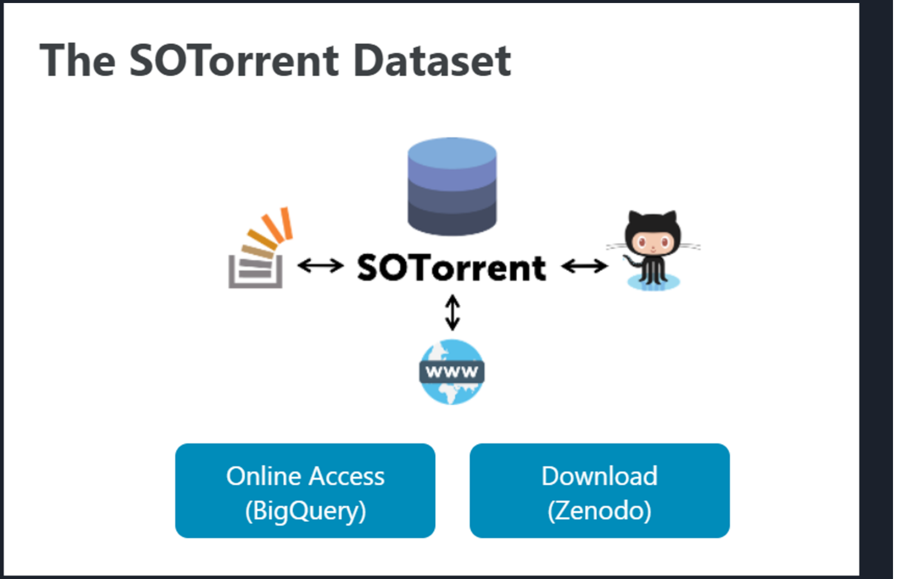


Statistics taken from https://people.cs.vt.edu/nm8247/publications/ICSE-Main-293.pdf

- A study in which over 950 different groups of SO posts was conducted to compare secure and insecure advice
  - Out of the answer posts in the different groups, 785 were considered secure, and 644 were considered insecure
  - Insecure posts were viewed more (36,508 views vs. 18,713)
  - Insecure posts reviewed higher scores (14 vs. 5)
  - Insecure posts were found to be duplicated more often on average (3.8 vs. 3.0)
- 34% of the code posts provided were deemed insecure

How Often are Vulnerabilities Solved?

- A group of researchers studied the SOTorrent data set which contained over 10 years worth of Stack Overflow data
- Data was narrowed down to 69 vulnerable snippets
- Traces of these vulnerable snippets were found in about 2800 live, published projects
- The owners of the repositories that contained the original snippets were contacted, some answered
- Unfixed solutions still floating around



The SOTorrent Dataset

Types of Insecure Solutions

1. Improper Check for Unusual or Exceptional Conditions
2. Improper input validation
3. Unchecked return value
4. Use of obsolete function
5. Uncontrolled memory allocation
6. Improper neutralization of null byte or null character
7. Use of externally controlled format string
8. Null pointer deference

(Donovan, 2019)

Prevention of Insecure Solutions

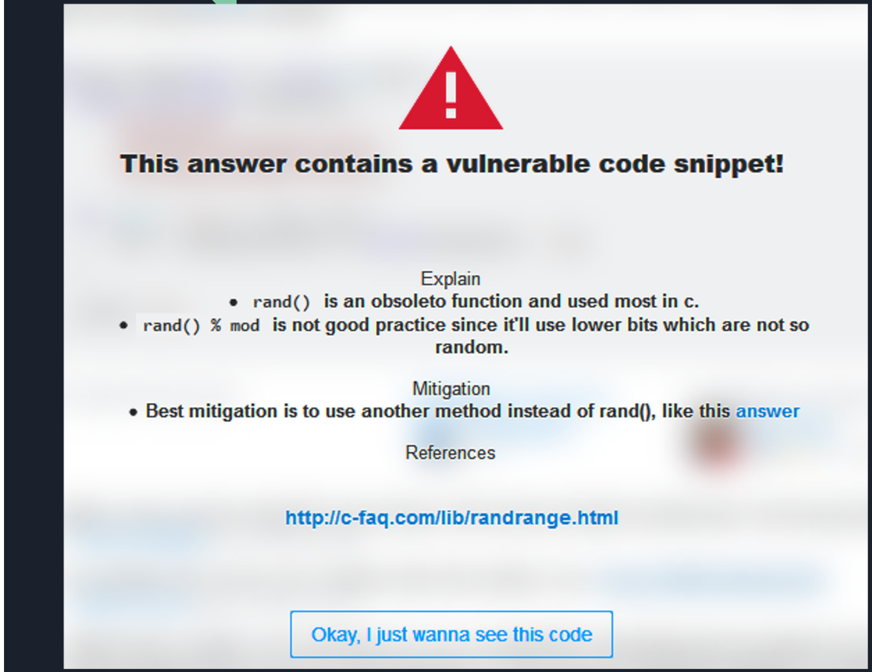
- Ensuring all conditions and exceptions are accounted for
  - Testing for low resource conditions
  - Reviewing cross language programs thoroughly
  - Ensuring data types are expected
  - Throwing exceptions if expectations are not met
- Using whitelists & blacklists to determine valid inputs
- Double checking for antiquated functions
- Excluding null inputs
- Ensuring string inputs are developer (not user) controlled

If you find any issues with solutions on a crowdsourced platform, make sure to identify it and share information about it

Conclusion

- Crowdsourcing websites have a lack of regulation
- Insecure advice acquires more views and upvotes
- Lack of incentives for reporting these solutions.

Conclusion Cont.



This answer contains a vulnerable code snippet!

Explain

- `rand()` is an obsolete function and used most in c.
- `rand() % mod` is not good practice since it'll use lower bits which are not so random.

Mitigation

- Best mitigation is to use another method instead of rand(), like this answer

References

http://c-faq.com/lib/randrange.html

- No systems designed in place to prevent malicious use of code.
- What crowdsourced websites can do
  - Provide incentives for their users
  - Implement static checkers that check for vulnerable code.

Question

- In the case of stack overflow, do the pros outweigh the security vulnerabilities when using code to develop products?
- If websites similar to stack overflow were to shut down due to these flaws how negatively would it impact the tech environment?

Sources

https://stackoverflow.blog/2019/12/02/preventing-the-top-security-weaknesses-found-in-stack-overflow-code-snippets/

https://www.researchgate.net/publication/260338589_Researching_Crowdsourcing_Software_Development_Perspectives_and_Concerns

https://reader.elsevier.com/reader/sd/pii/S1877050916000119?token=3D353212C017DA033F997A54F4C0A4802B8E243FFA83F470FB36387442EBAD7EBCCBA817094FE1AE5598C0CABB31DEEA&originRegion=us-east-1&originCreation=20220424190448

https://people.cs.vt.edu/nm8247/publications/ICSE-Main-293.pdf