An Empirical Study on the Effectiveness of Security Code Review

Anne Edmundson, Brian Holtkamp, Emanuel Rivera, Matthew Finifter, Adrian Mettler, David Wagner

Introduction

- As of March 2012, over 644 million websites
- 70-90% of all websites have at least one serious vulnerability
- Web application layer is the most commonly exploited layer for attackers


NBC
MySQL
The New York Times
ON THE WEB
zynga
ORACLE
AMERICAN EXPRESS
The Washington Post
EA
ELECTRONIC ARTS
NASA
NOKIA
SONY
Sun
microsystems
at&t
PBS
THE WALL STREET JOURNAL.
github
SOCIAL CODING
twitter
LinkedIn

Review Techniques

Manual Code Review

- Advantages
  - Less false negatives and positives
  - Insight into code design/ quality
- Disadvantages
  - Slow
  - Expensive

Static Analysis Tools

- Advantages
  - Fast
  - Less expensive
- Disadvantages
  - More false negatives and positives
  - Missing security requirements

Goals: Effectiveness

- What fraction of the vulnerabilities can we expect to be found by a single security reviewer?
- Are some reviewers significantly more effective than others?
- How much variation is there between reviewers?

Optimal Number of Reviewers

- Will multiple independent code reviewers be significantly more effective than a single reviewer?
- If so, how much more effective?
- How many reviewers are needed to find most or all of the bugs in a web application?

Goals: Predicting Effectiveness

- Can we predict how effective a code reviewer will be, based upon their background?

Methodology

Vulnerability Type: Cross-Site Scripting (XSS)
Vulnerability Location: /foo/bar.php; Line 5
Vulnerability Description: Allows any external user regardless of privileges browsing the page to inject malicious JavaScript that can be reflected to any other users browsing the same page.
Impact: Could run malicious code on any visitor to the same webpage.
Steps to Exploit:

1) Load the page
2) Click on the text field
3) Insert malicious JavaScript code, such as <script>alert(1);</script>
4) Submit the form
5) Reload the page to receive the attack

Anchor CMS

ANCHOR CMS

(5 vulnerabilities)

TEST CMS

(7 vulnerabilities)

Recruitment

- Recruited web developers with PHP experience
- Administered a pre-screen test with questions on web security

Reviewer's Task

- Report vulnerabilities
- Complete survey

Vulnerability Type: Cross-Site Scripting (XSS)
Vulnerability Location: /foo/bar.php; Line 5
Vulnerability Description: Allows any external user regardless of privileges browsing the page to inject malicious JavaScript that can be reflected to any other users browsing the same page.
Impact: Could run malicious code on any visitor to the same webpage.
Steps to Exploit:

1) Load the page
2) Click on the text field
3) Insert malicious JavaScript code, such as <script>alert(1);</script>
4) Submit the form
5) Reload the page to receive the attack

Limitations

- Population: contract developers
- We injected artificial vulnerabilities
- No guarantee of manual review
- We only tested one small codebase

Vulnerability Classification

Valid

- Exploitable vulnerability
- Correct description of exploit

Invalid

- Not exploitable/not a vulnerability

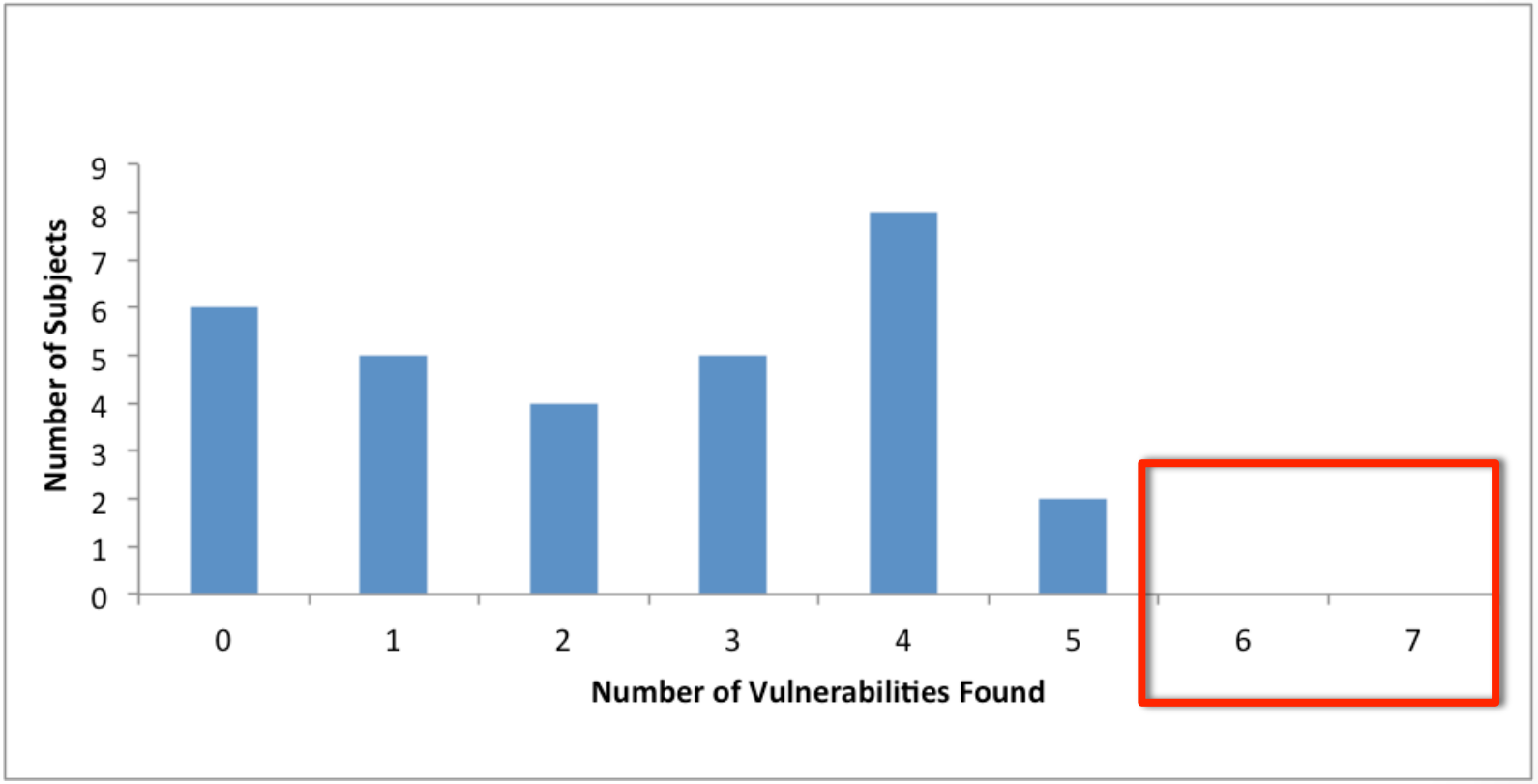
Weakness

- Configuration issues or poor coding practices

Out of Scope

- Administrative interface
- Duplicates

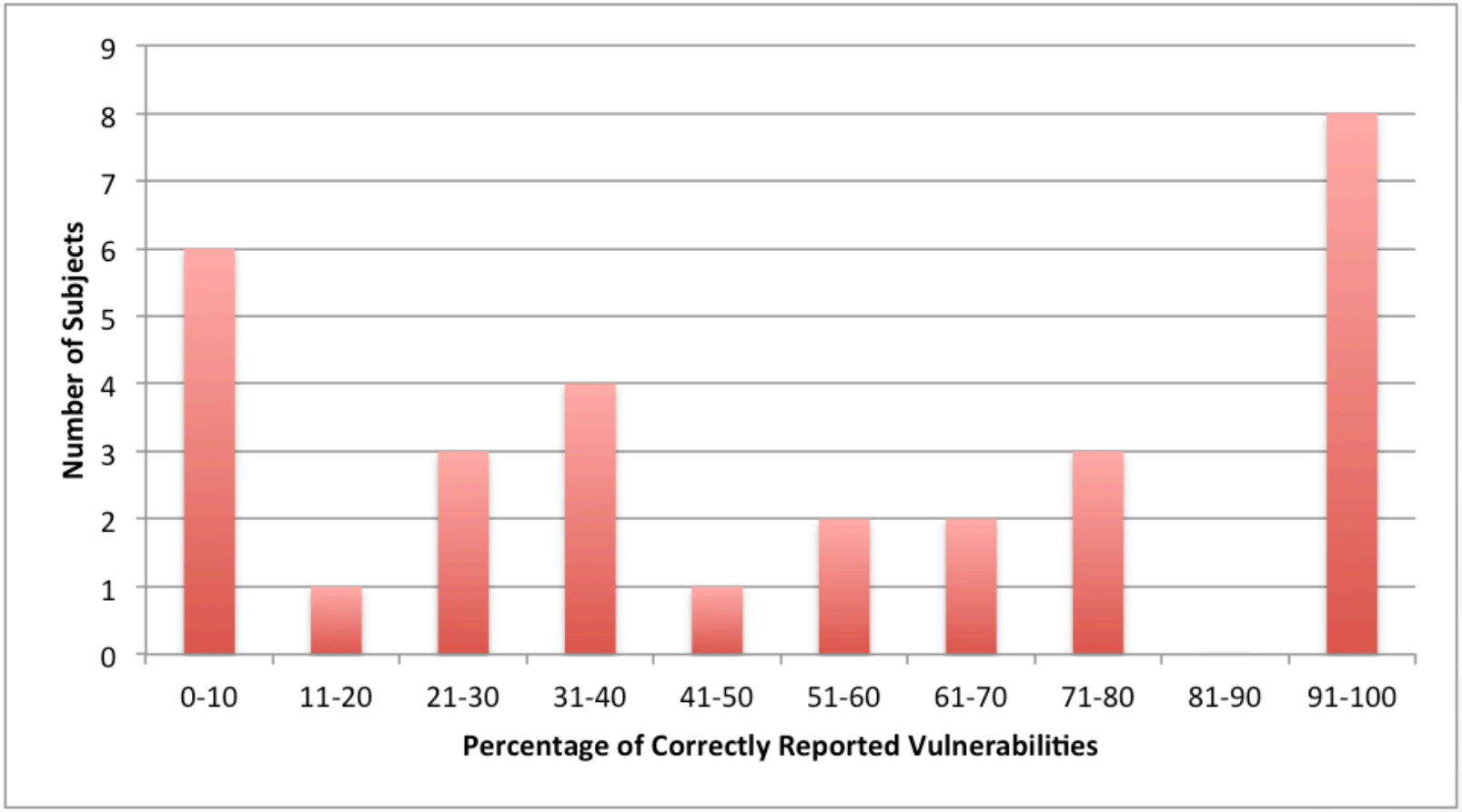
Results: Reviewer Effectiveness



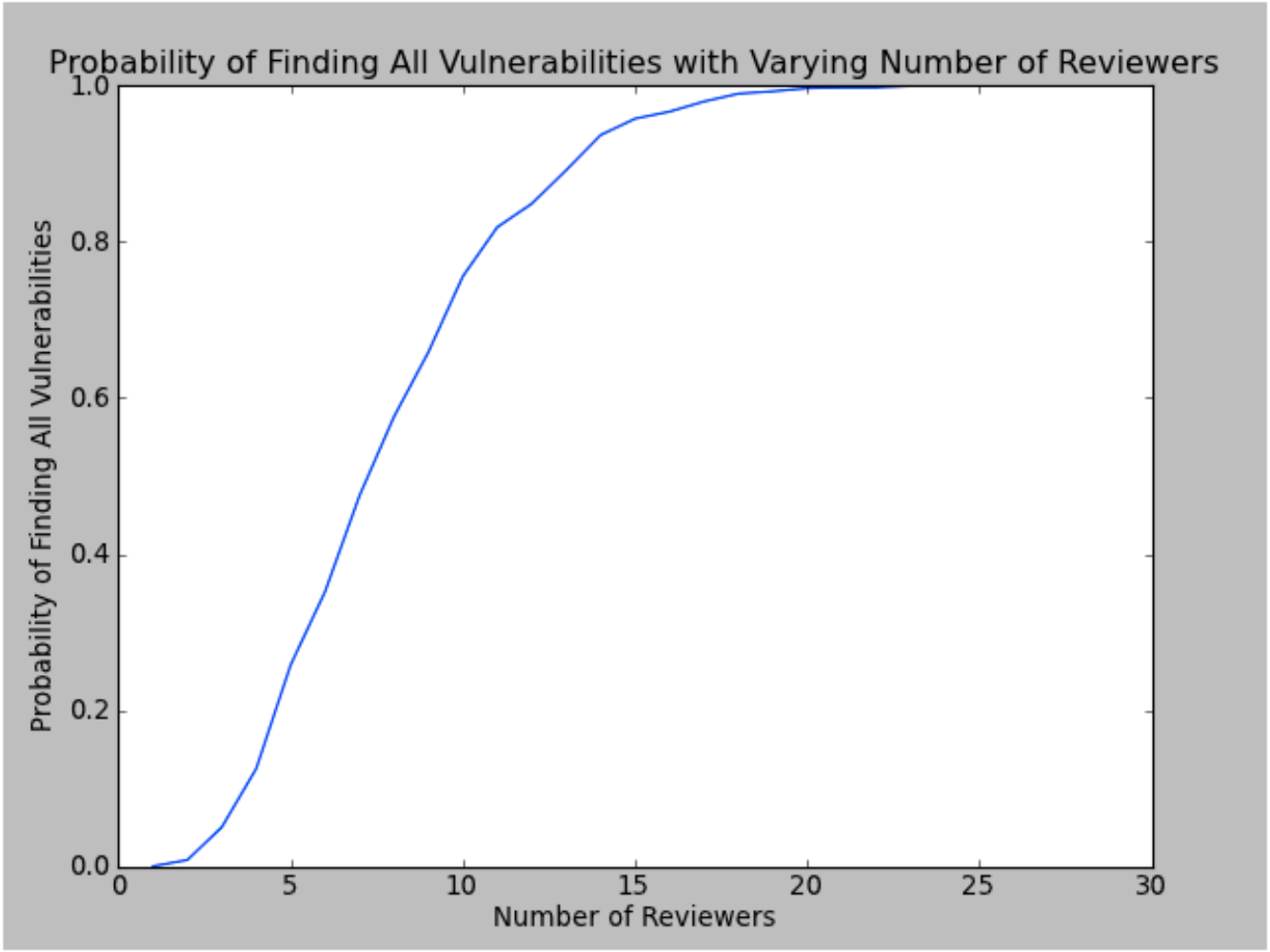
Results: Reviewer Effectiveness

Cross-Site Scripting 1	37%
Cross-Site Scripting 2	73%
Cross-Site Scripting 3	20%
Cross-Site Scripting 4	30%
SQL Injection 1	37%
SQL Injection 2	20%
Cross-Site Request Forgery	17%

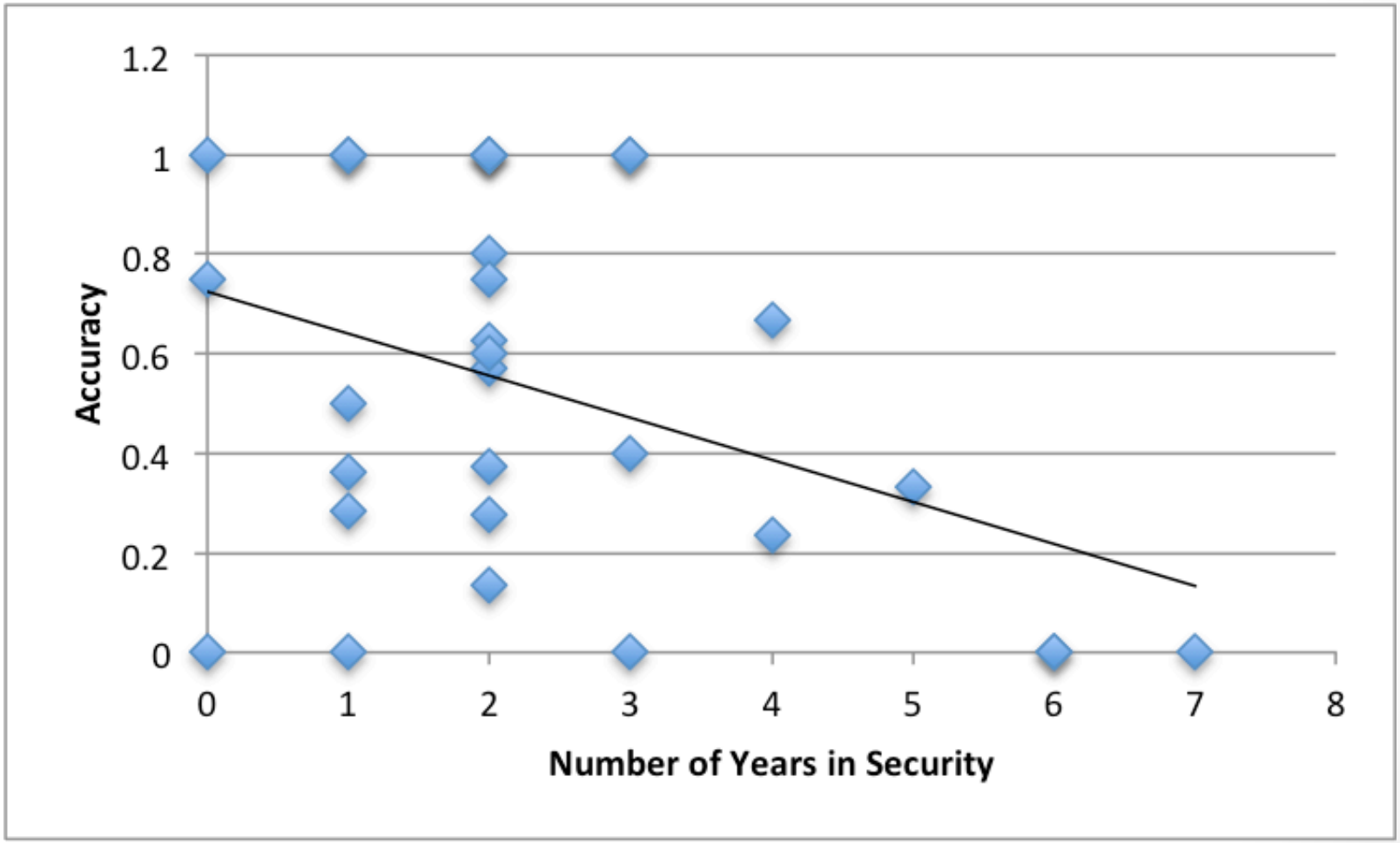
Results: Reviewer Variability



How many reviewers do I need?



Does experience help?



Can we predict effectiveness from the reviewer's background?

No

Conclusion

- Prior experience and education were not useful in predicting how well a subject was able to complete the code review

- Overall effectiveness was low: 20% found no true vulnerabilities, no developer found more than 5 vulnerabilities

Future Work

- Do these results apply to other populations?
- How does a pre-screen test correlate to effectiveness?
- How does the effectiveness of manual security review compare to that of other techniques, such as static analysis tools?

Questions?