YOU GET WHERE YOU’RE LOOKING FOR
THE IMPACT OF INFORMATION SOURCES ON CODE SECURITY

Yasemin Acar^, Michael Backes*, Sascha Fahl#, Doowon Kim, Michelle Mazurek, Christian Stransky^  
U. of Maryland; ^Leibniz U. Hannover; *CISPA, Saarland U; #Ruhr-U. Bochum
Has this happened to you?
That doesn’t seem right ....

- Answer suggests to trust all certs
  - Many real apps [Fahl+ 2012]

- Some interviewees: pasted from internet
Stack Overflow considered insecure

- “Everyone knows” copy-paste from the internet is bad for security
  - Particularly for “amateur” app devs?

- Can we measure this empirically?
- How does it contrast with official docs?
- What do real devs do?
Online developer survey

- Sent 50k invites, collected from Play
  - 295 valid responses

- Strategy for help with security/permissions

- General use of programming resources
Where do you look up ...

<table>
<thead>
<tr>
<th>Category</th>
<th>Books</th>
<th>Official Android docs</th>
<th>Search engines</th>
<th>Stack Overflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>69%</td>
<td>25%</td>
<td>75%</td>
<td>27.5%</td>
</tr>
<tr>
<td>Encryption</td>
<td>62%</td>
<td>10%</td>
<td>60%</td>
<td>30%</td>
</tr>
<tr>
<td>HTTPS</td>
<td>45%</td>
<td>20%</td>
<td>50%</td>
<td>45%</td>
</tr>
<tr>
<td>Permissions</td>
<td>58%</td>
<td>35%</td>
<td>25%</td>
<td>60%</td>
</tr>
</tbody>
</table>

69% Stack overflow, 62% search engines, 27.5% official
Next, a lab study

- Complete four short programming tasks
  - Designed to have secure/insecure solutions
- Resources constrained by condition:
  - Official docs, Stack Overflow, book, free choice
- Exit interview

- Not primed for security or privacy!
Skeleton app, emulator
Task 1: Secure networking

- Convert HTTP to HTTPS
  - In presence of X.509 cert error

- Sample secure solution:
  - Accept only this cert

- Sample insecure solution:
  - Accept all certs

http://5zin.com/certificate-of-authenticity-template.html
Task 2: Inter-component comms

• Given a service, limit access to only apps from same developer

• Sample secure solution:
  • Define a “signature” permission

• Sample insecure solution:
  • Export publicly
Task 3: Secure storage

• Store user ID and password locally

• Sample secure solution:
  • Private shared preference

• Sample insecure solution:
  • Public on SD card

http://www.routercheck.com/administrator-password/
Task 4: Least permissions

• Dial a customer-support phone number

• Sample secure solution:
  • Dial but don’t call

• Sample insecure solution:
  • Call (extra permission)
Evaluation

- **Correctness**: Does it compile and work?

- **Security**: If it works, was solution secure?
  - Coded manually in predefined categories

- **Self-reported sentiment**
  - Security thinking
  - Correctness and usefulness of resources
Recruitment

• In/around 3 universities, U.S. and Germany
  • Email, flyers, craigslist, developer forums

• 1+ Android course or 1+ yrs pro

• Pass basic Android knowledge questions
Participants

- 54 total
- 13 or 14 per condition
- 12 U.S., 42 Germany
- Ages 18-40; median 25
- 46 men, 8 women
- 14 professional, 40 non-professional
Demographics: lab vs. online

Many similarities; Lab had more formal education
Resource was easy to use

Free choice was easiest; book was worst
Books, official docs considered most correct
Security thinking

- Observed via think-aloud:
  - 16% thought about it
  - 5% said they ignored it for study / time

- Self-reported: 60% thought about it

- No significant difference in conditions
Functional correctness

- SO (67%) and Book (66%) performed best
- Official (40%) performed worst
  - Significantly worse than SO
Functionality by task

- Easiest: Least permissions (87%)
- Hardest: Secure networking (33.3%)
But what about security?

SO worst (51%), Official best (86%) (significant)
Security by task

- Storage: 100% of functional solutions secure
- Networking: Only 39%
Professionals vs. students

- More functional
- But not significantly more secure!
Lookup behavior

- Official: scrolling, clicking internal links
- Stack Overflow: many search resets
- Free choice:
  - Everyone used official, all but one used SO
  - One picked up a book!
  - Results closest to SO
A closer look at Stack Overflow

- Collected via browser history
- 149 unique pages, 41 relevant
- 20 with code snippets
  - 7 only secure, 10 only insecure, 3 both
  - 3 insecure have warnings
So now what?

- If you want functional, secure code:
- Cut off the internet, give your devs a book!
Real takeaways

- Stack Overflow: **quick, functional** solutions
  - Official docs don’t
- But, it’s **less secure** than official or books

- We need resources that integrate both!
  - Add a security rating to influence upvote?
  - Integrate Q&A into official docs?
  - Use SO to identify trouble spots, provide code snippets in the official docs?
Acknowledgements

• Sven Bugiel, Marten Oltrogge, Andrew Lui, Yichen Qian for helping with the study

• Jennifer DeSimone and Joseph Smith at UMD IRB for smoothing international collaboration

• BMBF, CISPA, and NIST for funding
Real takeaways

• Stack Overflow: **quick, functional** solutions
  • Official docs don’t
• But, it’s **less secure** than official or books

• We need resources that integrate both!
  • Add a security rating to influence upvote?
  • Integrate Q&A into official docs?
  • Use SO to identify trouble spots, provide code snippets in the official docs?