Studying the prevalence of security issues in Android

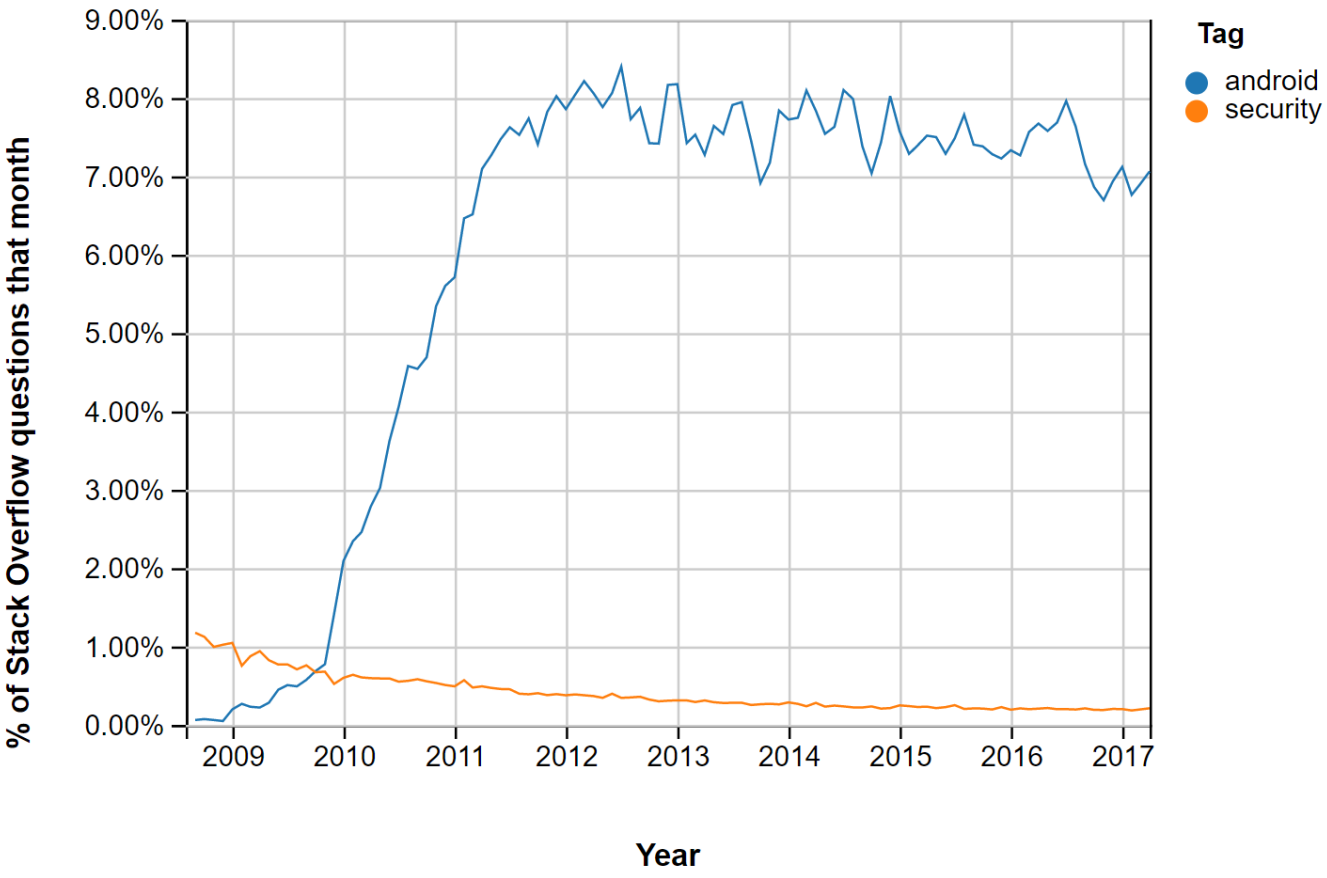
Seminar Project

Linus Schwab

Goal of the project

- Motivation: Users store a huge amount of sensitive data on their phone nowadays
- Goal: Studying Android security issues and developer involvement with security
- Approach: Mine Android issue tracker and StackOverflow questions
- Classification based on CIA
  - Confidentiality
  - Integrity
  - Availability

StackOverflow Trends



The Android Security Model

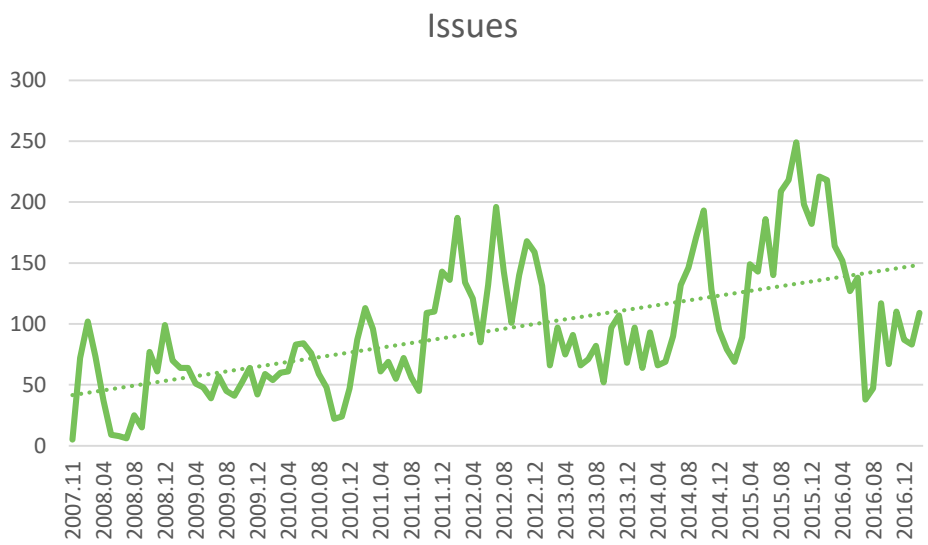
- Applications run in sandbox
- Permission-based
- Application signing and verification
- Verified boot

The Android issue tracker

- Purpose: Official way to report bugs, including security issues
- Security issues are hidden until they are fixed
- Contains lots of spam and irrelevant reports

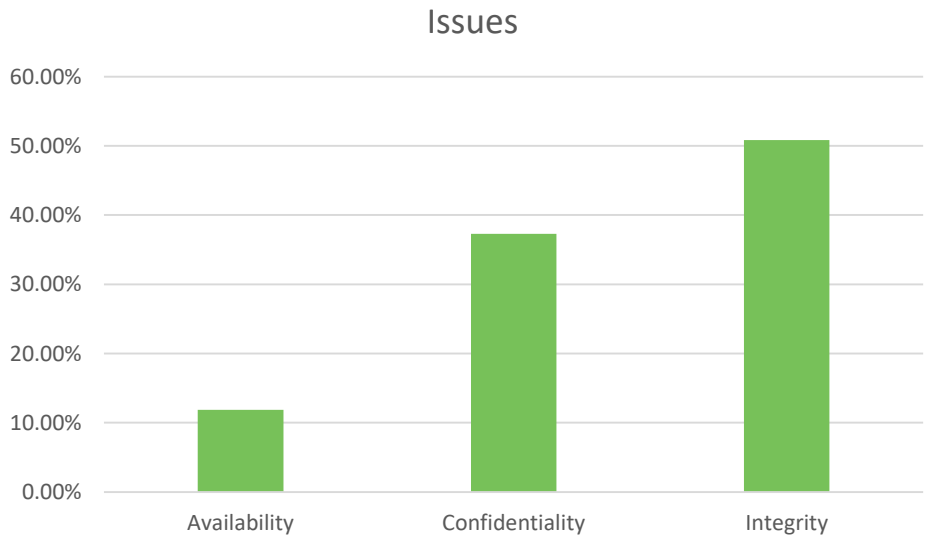
The Android issue tracker

- Mined data with Python application, stored in SQLite database
- Raw data set: 10’605 (closed) issues



Finding security issues

- Approach: build keyword list to filter data
- Manually study filtered issues
- Categorize security-related issues

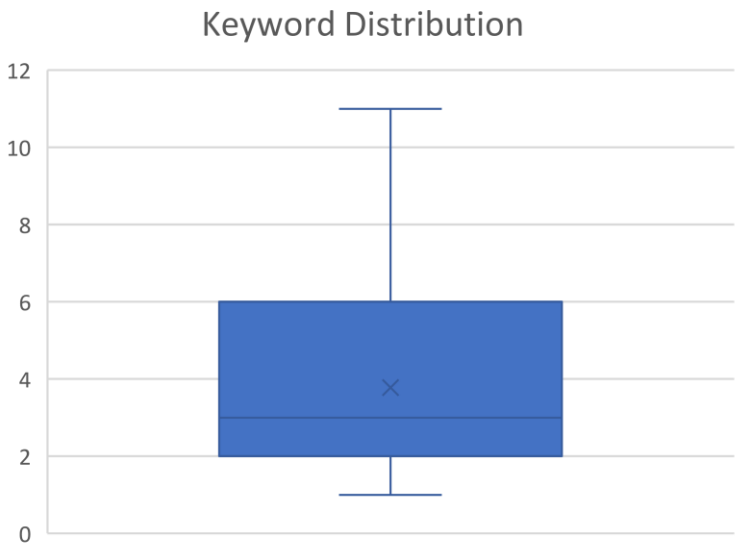


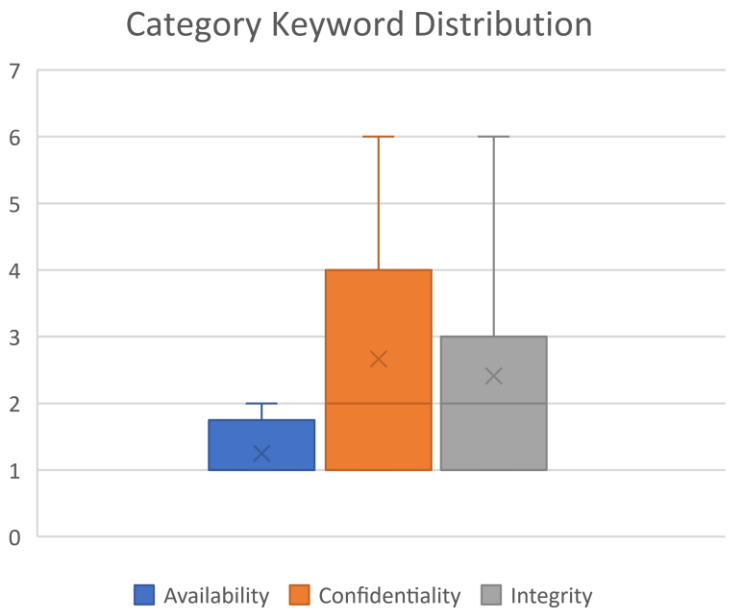
The keyword list

Availability	Confidentiality	Integrity
• availability • certification • consistent • crash • denial of service • destruction • loss • performance • reliability • reliable • safe • safety • stable • trust • validation	• access • authentication • authorization • certificate • classified • confidential • confidentiality • disclosure • encryption • grant • information disclosure • intrusion • lock • lockscreen • login • password • phishing • privacy • restrict • secret • secure • security • telephony • unauthorized • username	• attack • buffer • buffer overflow • buffer overrun • bypass • corrupt • elevation • escalation • exploit • integrity • malicious • modification • modify • overflow • overrun • permission • privilege • privilege elevation • privilege escalation • vulnerable

Based on: S. Licorish, S. MacDonell, T. Clear, “Analyzing Confidentiality and Privacy Concerns: Insights from Android Issue Logs”, EASE 2015.

Issue keyword distribution



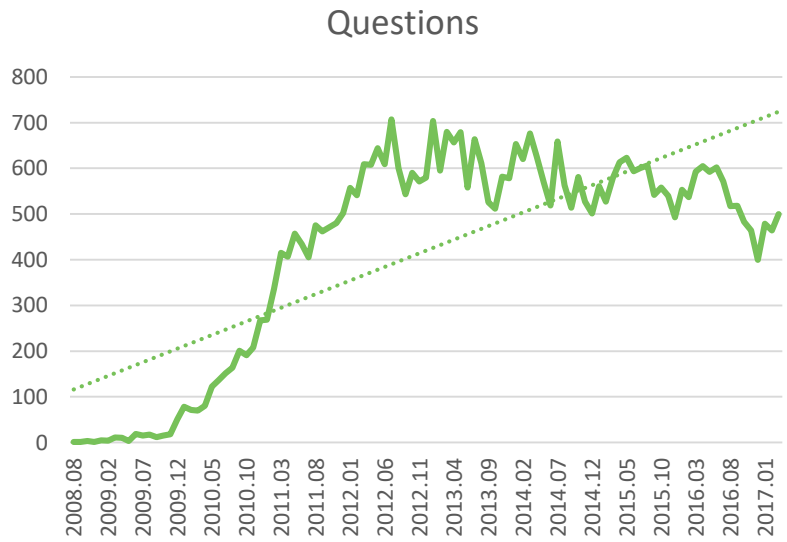


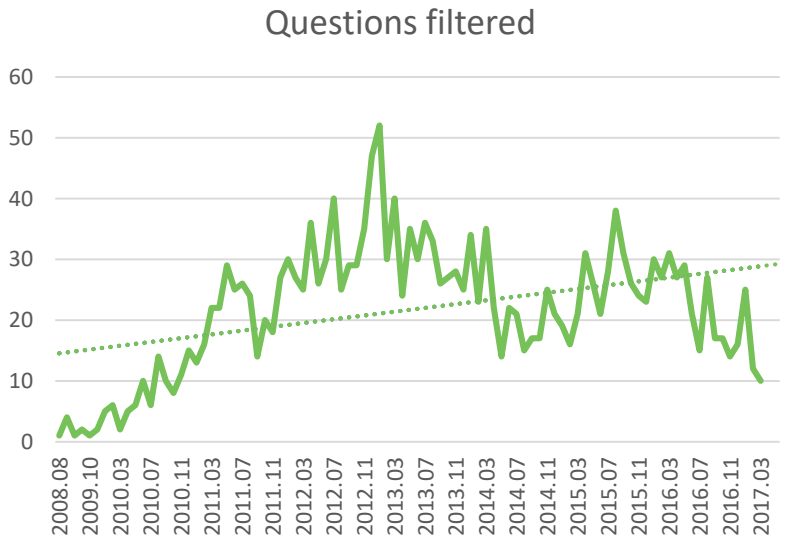
StackOverflow

- Purpose: Allows developers to ask questions about problems during software development
- Extended Python application to support the StackOverflow API
- Used keywords in combination with “Android” to search for security-related questions

StackOverflow

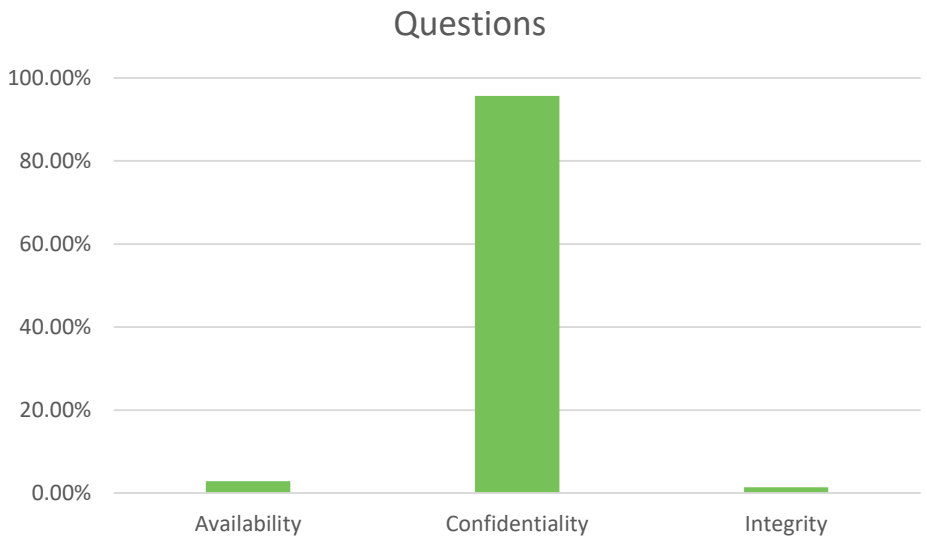
- Raw data set: 44’177 questions
- Filtered questions using score and keyword count: 1’974 questions



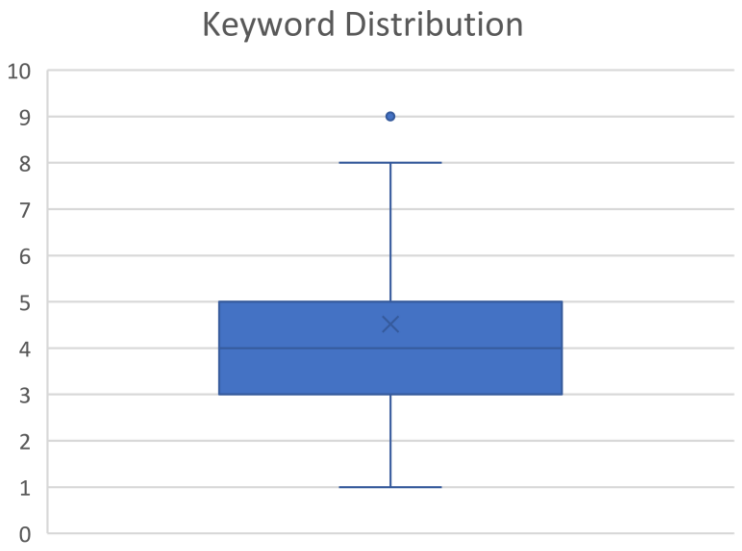


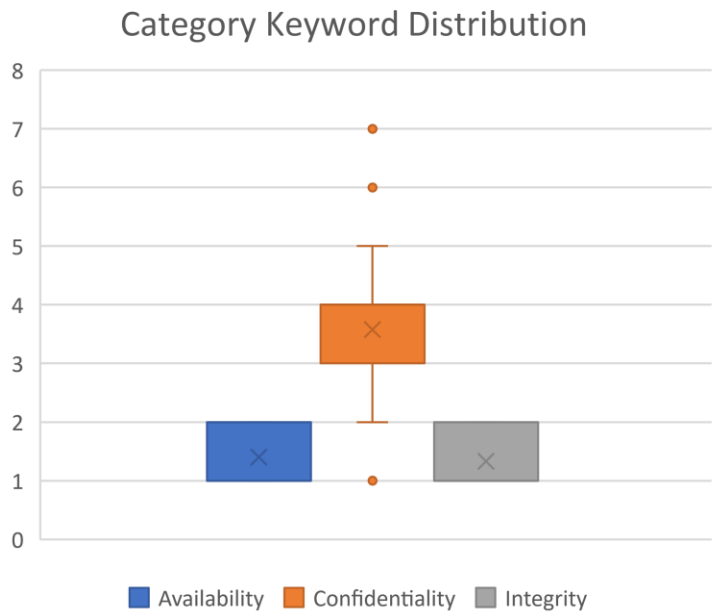
Result

- Identified and classified 300 security-related questions
- Extended keyword list



Question keyword distribution





Conclusion

- Most of the reported issues are about integrity, closely followed by confidentiality
- Developer involvement in security is mainly about confidentiality
- Observation: a lot of developers are concerned about how to securely store user credentials

Future work

- Synonyms and stemming for keywords
- Tf-idf for keyword importance
- Analyze answers in addition to title and body
- Extend the keyword list

Questions?