Java Cryptography Architecture

SOPHIE PFISTER
Context

- Kühne et al. (2017)

and their parameters. We have implemented a compiler that translates a CryptSL ruleset into a context- and flow-sensitive demand-driven static analysis. The analysis automatically checks a given Java or Android app for violations of the CryptSL-encoded rules.

We empirically evaluated our ruleset through analyzing 10,001 Android apps. Our results show that misuse of cryptographic APIs is still widespread, with 96% of apps containing at least one misuse. However, we observed fewer of the misuses that were reported in previous work.
Context

• Cryptography libraries lack usability.
• API misuse leads to security vulnerabilities.

→ unsafe applications!
Research Questions

1. What **issues** do programmers face when implementing symmetric encryption using the JCA library?

2. What **security risks** are present on Stack Overflow referring to the implementation of symmetric encryption using JCA library?

3. To what extent are these **linked to** missing or inadequate **documentation**?
Methodology

Sampling
- 150 Stack Overflow threads targeting symmetric encryption in JCA
- 50% «newest» / 50% «most popular»

Analysis 1
- identify issues
- qualitative content analysis + frequency analysis
Analysis 1

- Summarizing
- Classification 1: Technical Aspects
- Classification 2: Requirements
### Analysis 1: Technical Aspects

- Cipher Object Instantiation
- Parameters Generation
- Cipher Object Initialization
- Transformation
- Parameter Transmission
- Dependency Encryptor - Decryptor
Analysis 1: Technical Aspects

<table>
<thead>
<tr>
<th>Category</th>
<th>Issues</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>197</td>
<td>(90%)</td>
</tr>
<tr>
<td>Generation of Algorithm Parameters:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Key Derivation</td>
<td>36</td>
<td>(16.5%)</td>
</tr>
<tr>
<td>Cipher Object Instantiation:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Encryption Mode &amp; Padding</td>
<td>40</td>
<td>(18.3%)</td>
</tr>
</tbody>
</table>


Analysis 1: Requirements

- Use Cases (Functional Requirements)
- Performance
- Space
- Reliability
- Interoperability
- Portability
- Security
Records Classified Regarding Requirements (N = 219)

- Security: 46
- Portability: 12
- Interoperability: 7
- Functional Requirements: 5
- Reliability: 5
- Performance: 4
- Space: 2
Methodology

Sampling
- 150 Stack Overflow threads targeting symmetric encryption in JCA
- 50% «newest» / 50% «most popular»

Analysis 1
- identify issues
- qualitative content analysis + frequency analysis

Analysis 2
- security risks
- “check list”
Analysis 2: Check List

- Deriving rules from existing sets
  - CRYLOGGER tool for dynamic security risk tracking
  - CogniCrypt compiler for static security risk tracking
Analysis 2: Security Rules

- Cipher Object Instantiation
- Parameters Generation
- Cipher Object Initialization
- Parameter Transmission
Analysis 2: Checklist

4 Checklists:

• Question – Code
• Question – Text
• Answer – Code
• Answer – Text
Analysis 2: Results

- 150 question posts and 84 answer posts
- 331 security risks

<table>
<thead>
<tr>
<th></th>
<th>Code</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>249</td>
<td>38</td>
</tr>
<tr>
<td>Answer</td>
<td>35</td>
<td>9</td>
</tr>
</tbody>
</table>
Cipher Object Instantiation: 137
Generation of Algorithm Parameters: 97
Cipher Object Initialization: 12
Transmission of Algorithm Parameters: 3
Methodology

Sampling
- 150 Stack Overflow threads targeting symmetric encryption in JCA
- 50% «newest» / 50% «most popular»

Analysis 1
- identify issues
- qualitative content analysis + frequency analysis

Analysis 2
- security risks
- “check list”

Documentation
- derive questions
- seek answers in documentation
Questions

• based on results from former analyses
• How many threads are targeted by a question? → Priority
• 2 Lists:
  • «General Questions»
  • «Questions to Documentation»
Questions - to Documentation

<table>
<thead>
<tr>
<th>Question</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>What happens if I do not specify the IV although it is required?</td>
<td>9</td>
</tr>
<tr>
<td>How can I derive a key from a password?</td>
<td>7</td>
</tr>
<tr>
<td>What is the default value if I do not specify padding?</td>
<td>6</td>
</tr>
<tr>
<td>What kind of parameters do I have to pass to the decryption methods (update / doFinal)?</td>
<td>6</td>
</tr>
<tr>
<td>Which of the provided key derivation functions are standardized?</td>
<td>6</td>
</tr>
</tbody>
</table>

- default behavior
- (password based) key derivation
- method overloads
## Questions - General

<table>
<thead>
<tr>
<th>Question</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which encryption modes are safe?</td>
<td>113</td>
</tr>
<tr>
<td>What are security requirements for the key?</td>
<td>43</td>
</tr>
<tr>
<td>What requirements must the IV / nonce meet to be safe?</td>
<td>37</td>
</tr>
<tr>
<td>Which input data and specifications must be equal for encryption and decryption?</td>
<td>27</td>
</tr>
<tr>
<td>Which symmetric encryption algorithms are safe to use?</td>
<td>24</td>
</tr>
<tr>
<td>What are requirements for safe password based key derivation?</td>
<td>15</td>
</tr>
</tbody>
</table>

- security related
- dependency encryptor - decryptor
# Questions - General

<table>
<thead>
<tr>
<th>Question</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>What encryption modes require padding?</td>
<td>9</td>
</tr>
<tr>
<td>Which encryption modes do require an IV or nonce?</td>
<td>6</td>
</tr>
<tr>
<td>What is the required size for an IV?</td>
<td>5</td>
</tr>
<tr>
<td>What key sizes are supported by AES?</td>
<td>4</td>
</tr>
<tr>
<td>What does an IV do during encryption?</td>
<td>4</td>
</tr>
</tbody>
</table>

- encryption mode
- initialization vector
- key
Questions

- Total: 66 Questions
- 10 questions unanswered
- 13 answers are unclear, incomplete, or misleading

“Advanced Encryption Standard as specified by NIST in FIPS 197. Also known as the Rijndael algorithm by Joan Daemen and Vincent Rijmen, AES is a 128-bit block cipher supporting keys of 128, 192, and 256 bits.”
Unanswered Questions

• How can I specify PKCS#7 padding in Java?
• What properties does AES-256 require?
• Which symmetric algorithms are safe?
General Observations

• examples are not working (incomplete)
• platform / provider issue
  → decreases portability
Conclusion
What **issues** do programmers face ...?

**Tasks / Properties:**
- (password based) key derivation
- key storage / transmission
- initialization vector
- encryption mode

**API Related Issues**
- platforms and providers
- overloaded methods, `update(...)` vs. `doFinal(...)`

**Programmer Related Issues**
- lack of (domain) knowledge
What **security risks** are present...?

- unsafe encryption mode (ECB, CBC)
- static values for key and initialization values
- unsafe key derivation procedure
... issues **linked to ... documentation**?

- 65% of questions are answered by documentation
- “higher priority” questions are answered

Yet, the quality could be improved by
- providing working code examples
- linking trusted resources
- adding security hints / warnings
Conductive Thoughts

• Do programmers read documentation?
• API usability is very complex.
• Implementing cryptography requires expertise.
Summary

Methodology
- Sampling
  - 150 Stack Overflow threads targeting symmetric encryption in ICA
    - 50% newest / 50% most popular
- Analysis 1
  - Identify issues
  - Qualitative content analysis + frequency analysis
- Analysis 2
  - Security risks
  - “Check list”
- Documentation
  - Derive questions
  - Seek answers in documentation

Analysis 1: Technical Aspects

<table>
<thead>
<tr>
<th>Total</th>
<th>197 issues (90%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation of Algorithm Parameters:</td>
<td></td>
</tr>
<tr>
<td>- Key Derivation</td>
<td>36 issues (16.5%)</td>
</tr>
<tr>
<td>Cipher Object Instantiation:</td>
<td></td>
</tr>
<tr>
<td>- Encryption Mode &amp; Padding</td>
<td>40 issues (18.3%)</td>
</tr>
</tbody>
</table>

Questions - General

<table>
<thead>
<tr>
<th>Question</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which encryption modes are safe?</td>
<td>313</td>
</tr>
<tr>
<td>Which key derivation functions are safe?</td>
<td>40</td>
</tr>
<tr>
<td>What requirements must an IV / nonce / salt meet to be safe?</td>
<td>37</td>
</tr>
<tr>
<td>Which input data and specifications must be equal for encryption and decryption?</td>
<td>27</td>
</tr>
<tr>
<td>Which symmetric encryption algorithms are safe to use?</td>
<td>24</td>
</tr>
<tr>
<td>What are requirements for a safe password based key derivation function?</td>
<td>15</td>
</tr>
</tbody>
</table>