Java Cryptography Uses in the Wild

Mohammadreza Hazhirpasand, Mohammad Ghafari, Oscar Nierstrasz

University of Bern, Switzerland

A crypto API misuse happens due to an insecure

algorithm name, key length, padding mode, or method call.

Crypto API misuse

- MessageDigest (to produce hash)

MD5 algorithm has 128 bit length

Possible collisions

Available dictionaries with over 1 billions hashes

MessageDigest - Official documentation

Note that if a given implementation is not cloneable, it is still possible to compute intermediate digests by instantiating several instances, if the number of digests is known in advance.

Note that this class is abstract and extends from `MessageDigestSpi` for historical reasons. Application developers should only take notice of the methods defined in this `MessageDigest` class; all the methods in the superclass are intended for cryptographic service providers who wish to supply their own implementations of message digest algorithms.

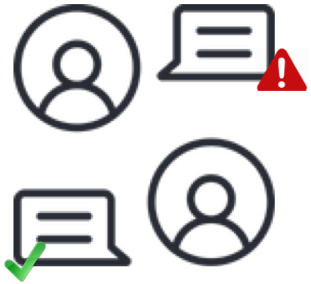
Every implementation of the Java platform is required to support the following standard `MessageDigest` algorithms:

- `MD5`
- `SHA-1`
- `SHA-256`

These algorithms are described in the MessageDigest section of the Java Security Standard Algorithm Names Specification. Consult the release documentation for your implementation to see if any other algorithms are supported.

Contribution

Crypto Dataset - CryptoMine



Developer feedback

CryptoMine - Pipeline

Search and filter

Clone and compile

Analyze

Manual analysis and store

1. Find Java projects
2. GitHub code search API to find crypto uses
3. Exclude forked projects

CryptoMine - Pipeline

Search and filter

Clone and compile

Analyze

Manual analysis and store

1. Maven build tool (skipping running test cases)
2. Exclude projects with failed dependencies

CryptoMine - Pipeline

Search and filter

Clone and compile

Analyze

Manual analysis and store

1. Run the CogniCrypt static analysis tool
2. Abort analyses with more than 15 minutes

CryptoMine - Pipeline

Search and filter

Clone and compile

Analyze

Manual analysis and store

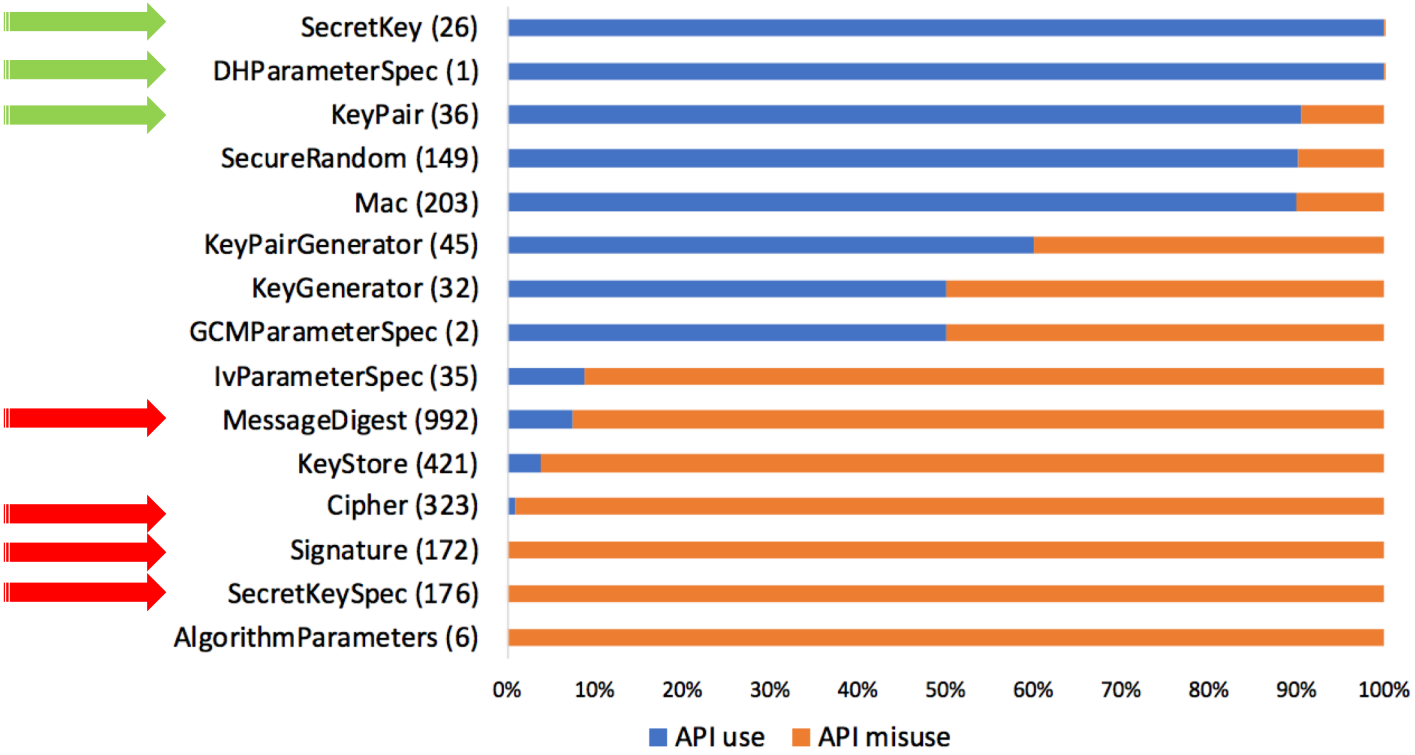
1. Manually check 1280 records of CryptoMine (48% of the dataset)
2. Store in a CSV file

CryptoMine - Structure

1. Project URL
2. Star count
3. Fork count
4. Creation date
5. Updated date
6. Last visited
7. File path
8. State of use
9. API name
10. Line number
11. User method
12. Misuse type
13. Misuse description
14. Manual check (Accepted, Rejected, Unvalidated)

74 records (Rejected)

CryptoMine – API use vs misuse



Developer feedback

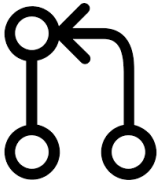
Personal repository

Will fix later

Not maintained anymore

Uncertainty

Refer to other libraries



Pull request

Consult documentation

Disagreement and context

Developer feedback

Personal repository

- Underestimate the impact of such issues on those who rely on online examples
- No concern about security when a program is being used on a very small scale.

The project is created for internal use and no issue will be addressed

Developer feedback

Will fix later

- Developers often underestimate the impact of a crypto misuse.

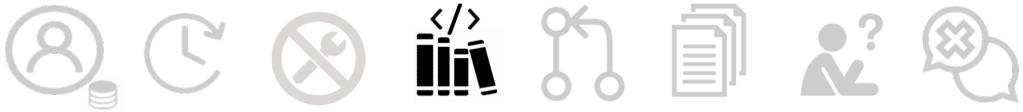
The misuses do not affect the functionality of the program

Developer feedback

Not maintained

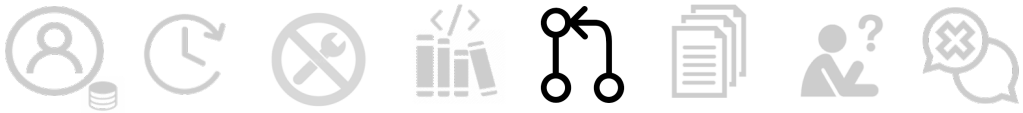
- As long as the code is available online, novice developers may rely on open-source projects irrespective of how active the projects are.

Developer feedback



- Developers seem not to be concerned about security risks associated with external libraries.

Developer feedback



- There is a risk that developers who lack security knowledge blindly accept security-related pull requests

> *I'm not sure if I understand the problem. I am not a cryptologist.*

Developer feedback

documentation

- Developers have confidence in official documentation, but security concerns are mainly absent in such resources.

MD5 is still supported by java according to the Java documentation.

Developer feedback

Uncertainty

- Developer uncertainty is mainly related to the right method call or the secure algorithm name to pass

how the misuses can be exploited in real life.

Developer feedback

Disagreement and context

- 45 repositories mainly argued that their context is not related to security.
- The use of crypto APIs to produce hashes was the most common non-security related usage

SHA1 was used only to generate a single hash for the entire contents of a folder

MD5 to track if the template source has been changed or not.

One contributor discussed that SHA1 is still usable regardless of the existing collision vulnerability

Summary

Pipeline

Search and filter

Clone and compile

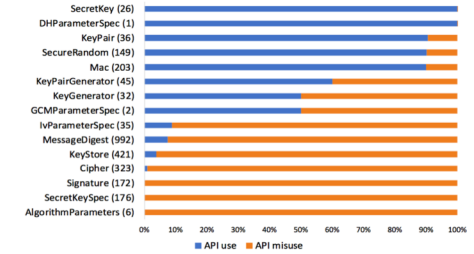
Analyse

Parse, inform and store

CryptoMine

1. Project URL
2. Star count
3. Fork count
4. Creation date
5. Updated date
6. Last visited
7. File path
8. State of use
9. API name
10. Line number
11. User method
12. Misuse type
13. Misuse description
14. Manual check (Accepted, Rejected, Unvalidated)

CryptoMine – API use vs misuse



Developer feedback

Personal repository

Will fix later

Not maintained anymore

Uncertainty

Refer to other libraries

Pull request

Consult documentation

Disagreement and context