My ZIP isn't your ZIP: Identifying and Exploiting Semantic Gaps Between ZIP Parsers

Yufan You[1], Jianjun Chen[1,2,✉], Qi Wang[1], Haixin Duan[1,2]

[1]Tsinghua University [2]Zhongguancun Laboratory

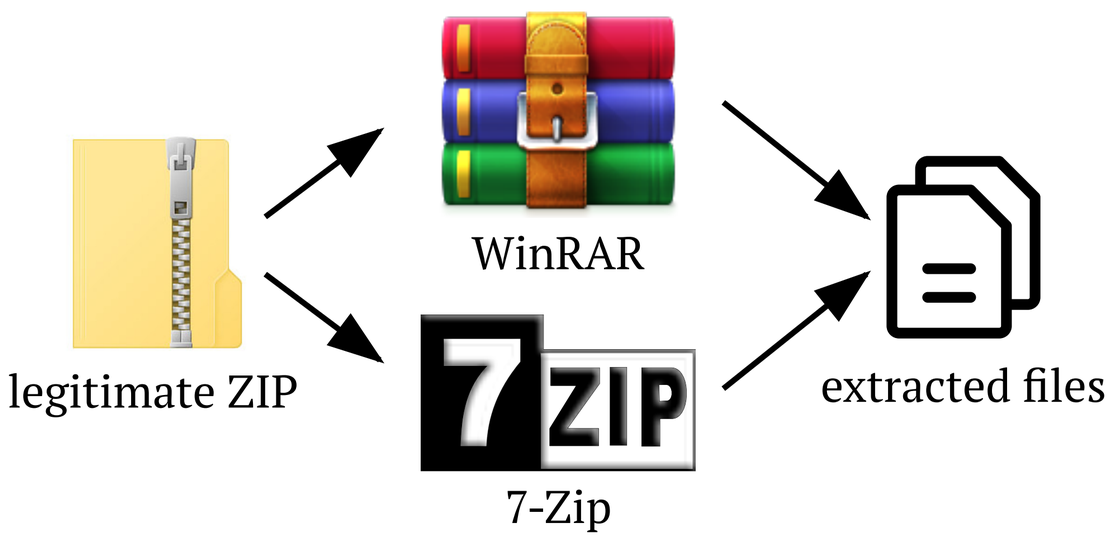
✉Corresponding author: jianjun@tsinghua.edu.cn

USENIX

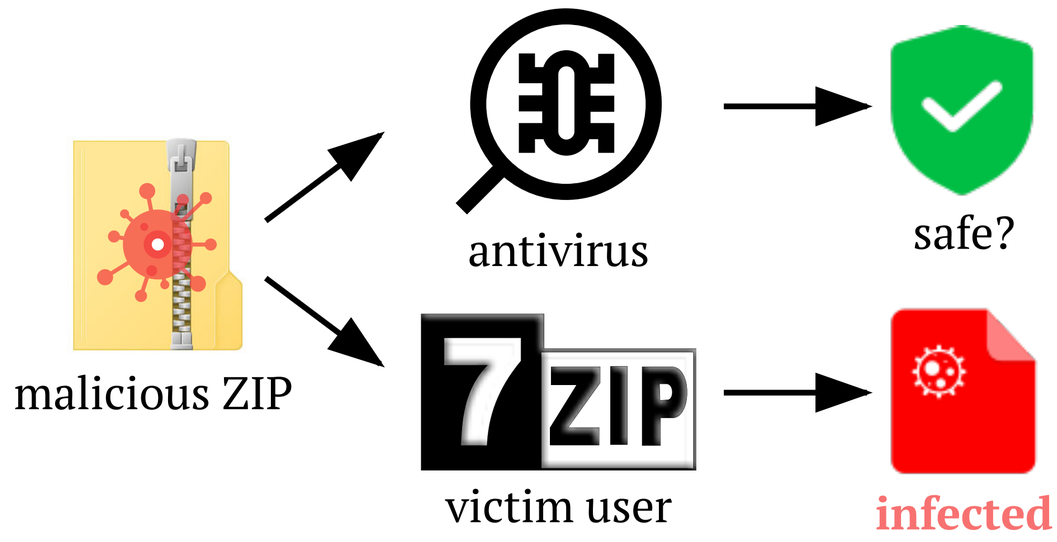
34TH USENIX SECURITY SYMPOSIUM

Motivation

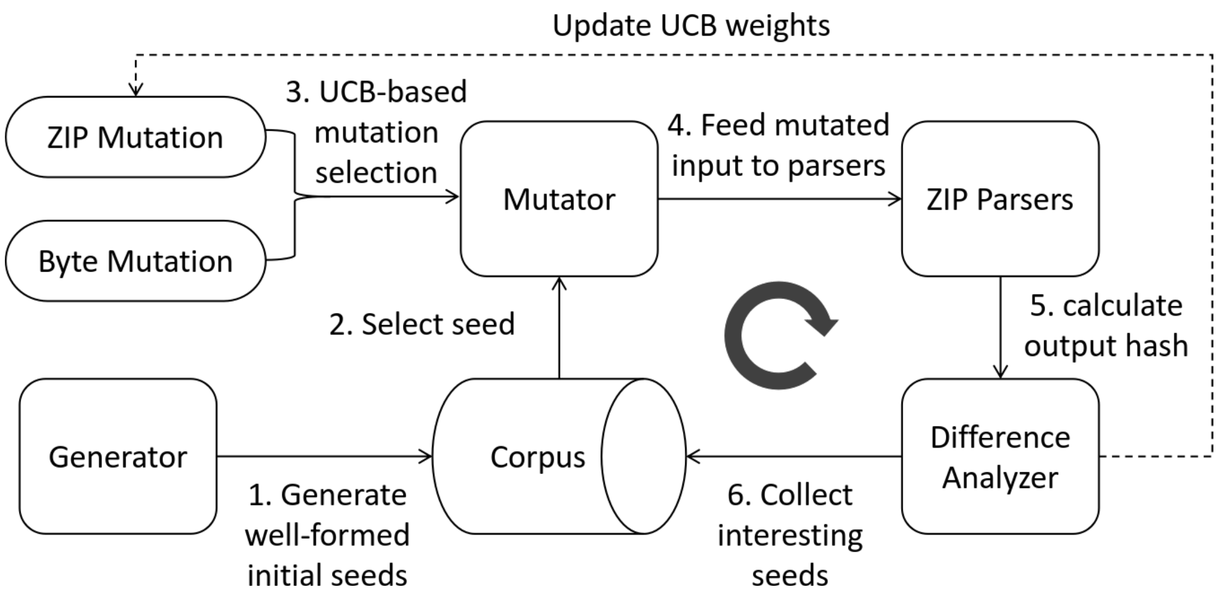
same ZIP file *multiple* parsers ***same*** contents?



same ZIP file *multiple* parsers ***different*** contents!



Workflow



Results

- 50 parsers across 19 programming languages
- 1221 out of 1225 pairs are inconsistent
- 14 types of ambiguities in three categories
  - Redundant Metadata
  - File Path Processing
  - ZIP Structure Positioning
- Detailed analysis, construction, and measurement

Exploitations

- Secure Email Gateway Bypass
- Office Document Content Spoofing
- LibreOffice Document Signature Forgery
- Spring Boot Nested JAR Signature Forgery
- VS Code Extension ID Impersonation

Responsible Disclosure

Bug Bounty Rewards

Gmail

Coremail

ZOHO

CVEs

GO spring boot

LibreOffice The Document Foundation

Other Acknowledgments

Mitigation

- Use the same parser
- Normalize the ZIP file
- Identify ambiguous patterns in ZIP files
- Incorporate different parsing logics
- Fix unique parsing behaviors
- On-access scanning
- Better file format design

Artifact

ARTIFACT EVALUATED usenix ASSOCIATION AVAILABLE

ARTIFACT EVALUATED usenix ASSOCIATION FUNCTIONAL

ARTIFACT EVALUATED usenix ASSOCIATION REPRODUCED

GitHub Repo

https://github.com/ouuan/ZipDiff

Differential Fuzzer ZIPDIFF & Ambiguity Constructions

> *be conservative in what you send, be ~~liberal~~ strict in what you accept*
>
> — *Postel's law (corrected)*