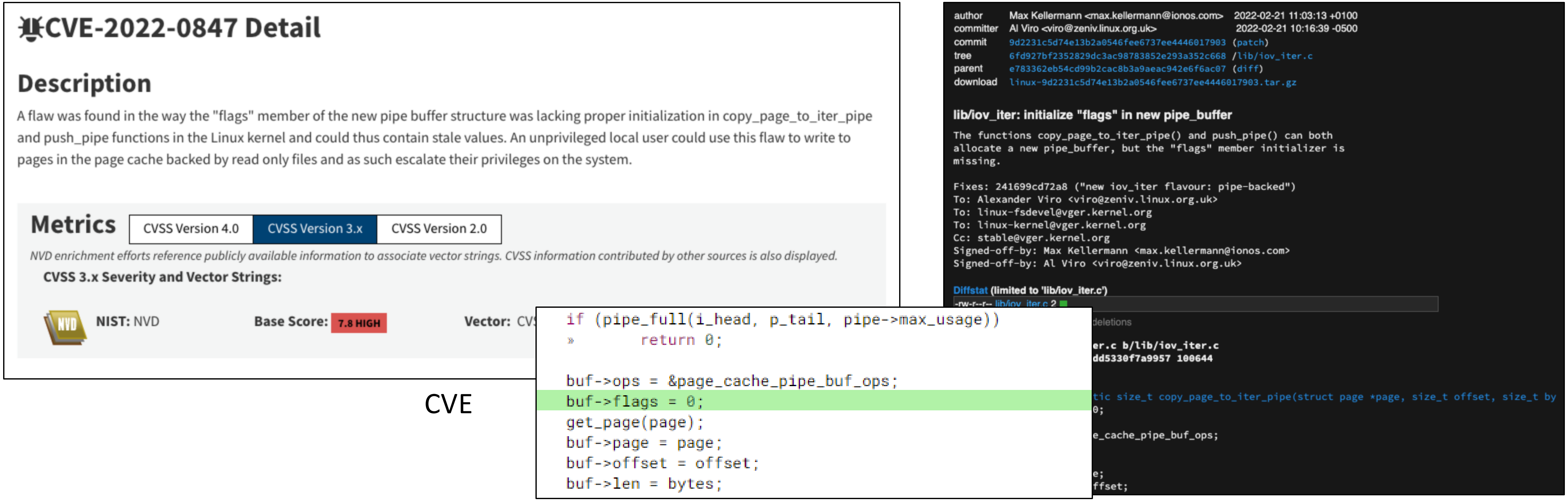
VulZoo: A Comprehensive Vulnerability Intelligence Dataset

Bonan Ruan, Jiahao Liu, Weibo Zhao, Zhenkai Liang

ASE Tool Demo, October 2024
Sacramento, California, USA

NUS
National University
of Singapore

Vulnerability Profile: Scattered Data



CVE-2022-0847 Detail

Description

A flaw was found in the way the "flags" member of the new pipe buffer structure was lacking proper initialization in copy_page_to_iter_pipe and push_pipe functions in the Linux kernel and could thus contain stale values. An unprivileged local user could use this flaw to write to pages in the page cache backed by read only files and as such escalate their privileges on the system.

Metrics

NVD enrichment efforts reference publicly available information to associate vector strings. CVSS information contributed by other sources is also displayed.

CVSS 3.x Severity and Vector Strings:

NIST: NVD **Base Score:** 7.8 HIGH **Vector:** CV

CVE

```
if (pipe_full(i_head, p_tail, pipe->max_usage))
»       return 0;

buf->ops = &page_cache_pipe_buf_ops;
buf->flags = 0;
get_page(page);
buf->page = page;
buf->offset = offset;
buf->len = bytes;
```

Source Code

```
author     Max Kellermann <max.kellermann@ionos.com>   2022-02-21 11:03:13 +0100
committer  Al Viro <viro@zeniv.linux.org.uk>             2022-02-21 10:16:39 -0500
commit     9d2231c5d74e13b2a0546fee6737ee4446017903 (patch)
tree       6fd927bf2352829dc3ac98783852e293a352c668 /lib/iov_iter.c
parent     e783362eb54cd99b2cac8b3a9aeac942e6f6ac07 (diff)
download   linux-9d2231c5d74e13b2a0546fee6737ee4446017903.tar.gz
```

lib/iov_iter: initialize "flags" in new pipe_buffer

```
The functions copy_page_to_iter_pipe() and push_pipe() can both
allocate a new pipe_buffer, but the "flags" member initializer is
missing.

Fixes: 241699cd72a8 ("new iov_iter flavour: pipe-backed")
To: Alexander Viro <viro@zeniv.linux.org.uk>
To: linux-fsdevel@vger.kernel.org
To: linux-kernel@vger.kernel.org
Cc: stable@vger.kernel.org
Signed-off-by: Max Kellermann <max.kellermann@ionos.com>
Signed-off-by: Al Viro <viro@zeniv.linux.org.uk>
```

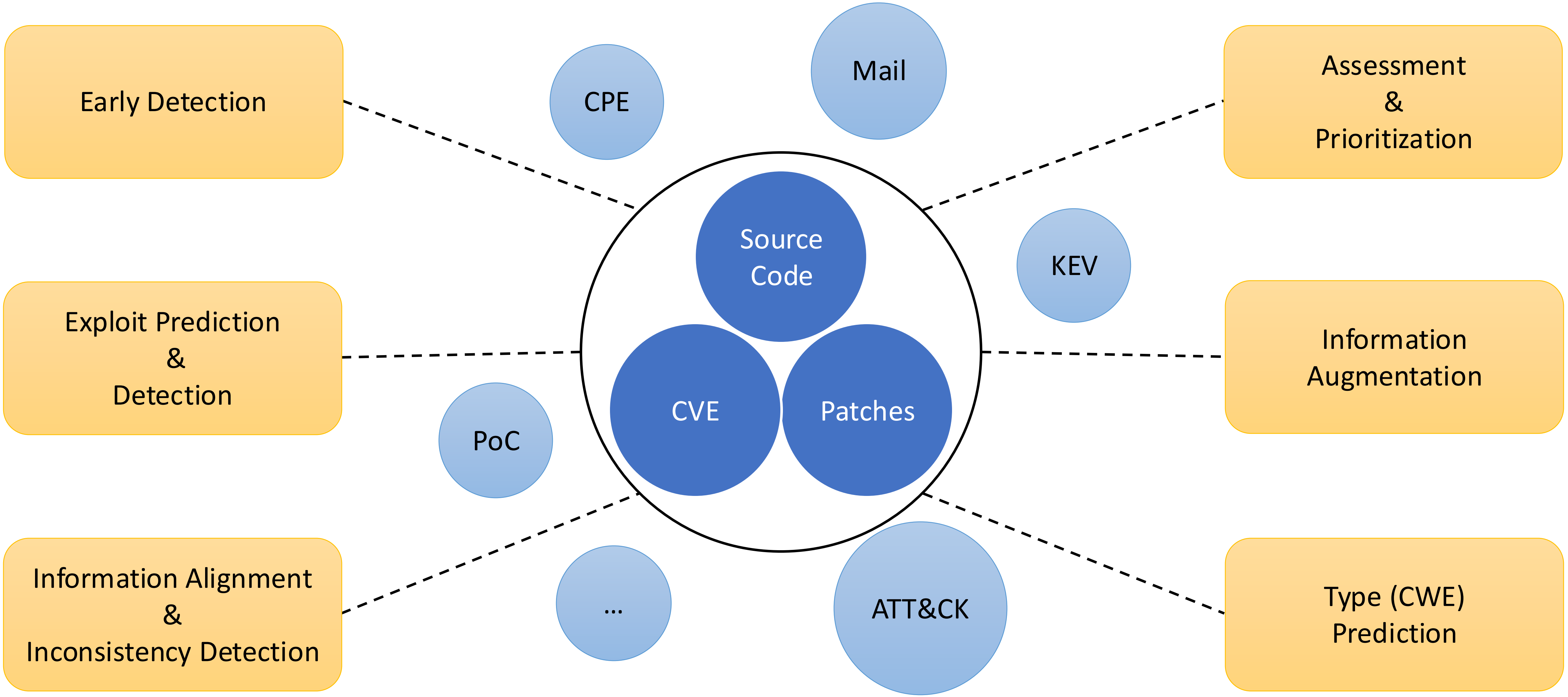
Diffstat (limited to 'lib/iov_iter.c')

```
-rw-r--r-- lib/iov_iter.c 2 ■
deletions

er.c b/lib/iov_iter.c
dd5330f7a9957 100644

tic size_t copy_page_to_iter_pipe(struct page *page, size_t offset, size_t by
0;

e_cache_pipe_buf_ops;


e;
ffset;
```

Patch

- Vulnerability-related information is scattered and segmented
- For example, CVE-2022-0847, a Linux kernel vulnerability, is detailed in CVE database, source code, patch, and more.

Underutilized Vulnerability Intelligence



Early Detection

Exploit Prediction & Detection

Information Alignment & Inconsistency Detection

Assessment & Prioritization

Information Augmentation

Type (CWE) Prediction

Underutilized Vulnerability Intelligence

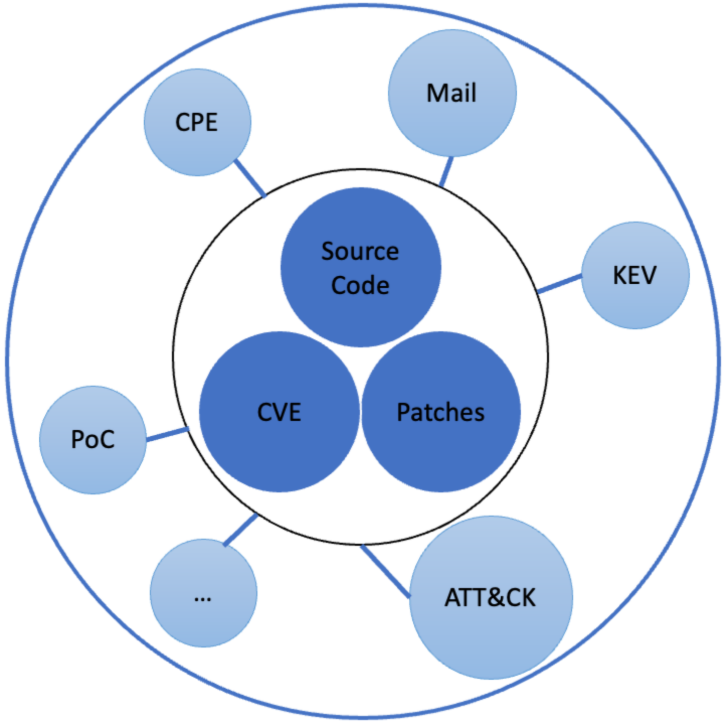
- Observation: Vulnerability data has not yet been fully utilized!
- Limited data adopted in existing studies:
  - CVE information databases (MITRE CVE, NVD, OSV, ...)
  - Vulnerable source code in software projects
  - Patches derived from software project commits
- Motivating Questions:
  - How much diverse vulnerability intelligence on earth is available?
  - Can we **profile vulnerabilities** better than what is provided by CVE information, source code, and patches?
    - As a comprehensive dataset derived from domain knowledge?

Towards A Comprehensive Dataset

Existing Dataset
MITRE CVE
NVD
OSV
PatchDB
CVEfixes
Vulnerable Code Snippets
…

- Limited Scope
- Different Specializations
- Separated
- Scattered

VulZoo: a comprehensive vulnerability intelligence dataset

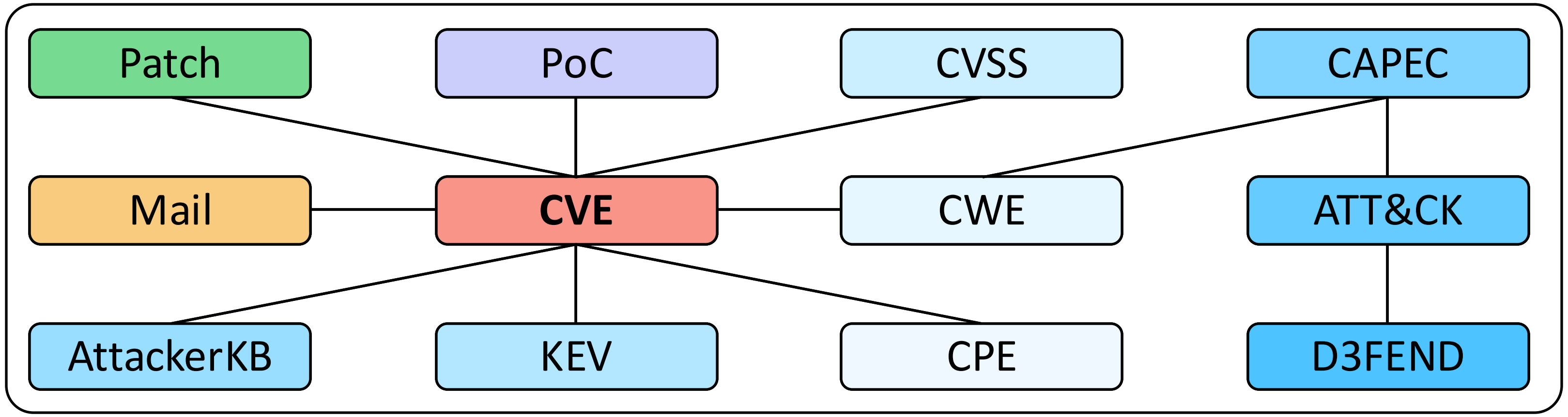


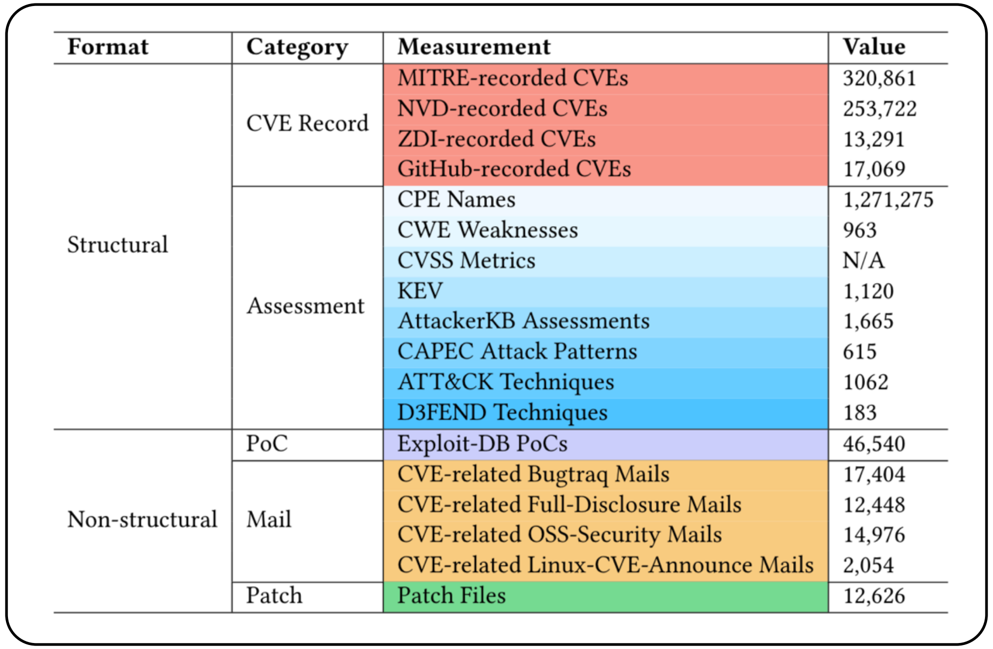
More Detailed Vulnerability Profile

VulZoo can

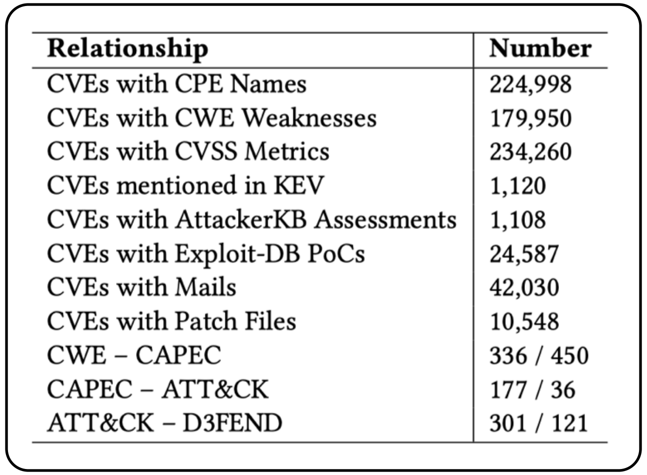
- connect the scattered data
- comprehensively profile vulnerabilities

Overview of VulZoo





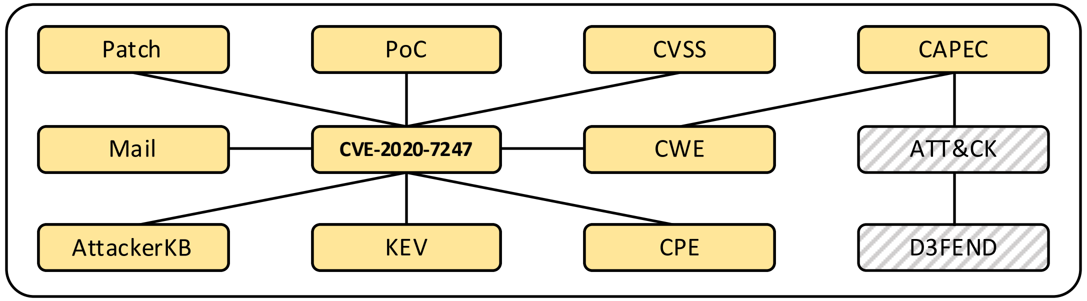
Format	Category	Measurement	Value
Structural	CVE Record	MITRE-recorded CVEs	320,861
		NVD-recorded CVEs	253,722
		ZDI-recorded CVEs	13,291
		GitHub-recorded CVEs	17,069
	Assessment	CPE Names	1,271,275
		CWE Weaknesses	963
		CVSS Metrics	N/A
		KEV	1,120
		AttackerKB Assessments	1,665
		CAPEC Attack Patterns	615
		ATT&CK Techniques	1062
		D3FEND Techniques	183
Non-structural	PoC	Exploit-DB PoCs	46,540
	Mail	CVE-related Bugtraq Mails	17,404
		CVE-related Full-Disclosure Mails	12,448
		CVE-related OSS-Security Mails	14,976
		CVE-related Linux-CVE-Announce Mails	2,054
	Patch	Patch Files	12,626



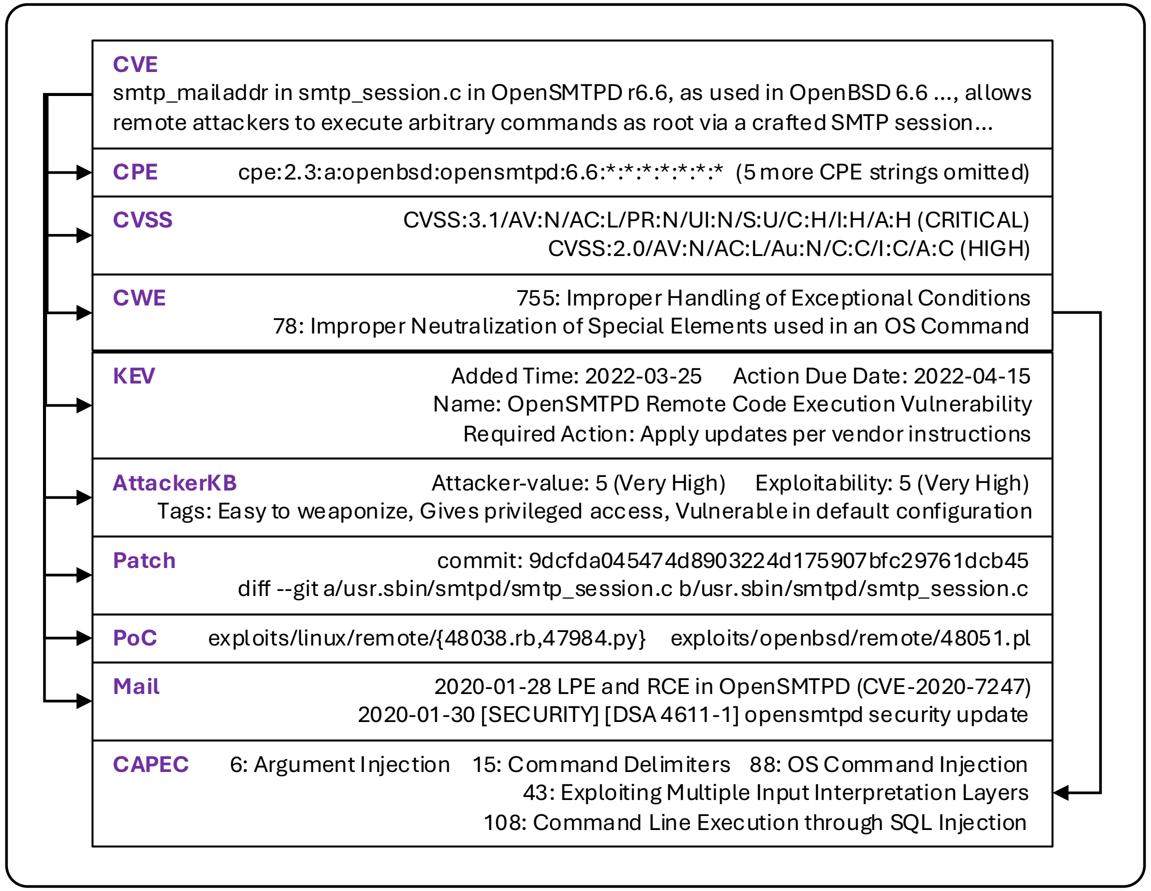
Relationship	Number
CVEs with CPE Names	224,998
CVEs with CWE Weaknesses	179,950
CVEs with CVSS Metrics	234,260
CVEs mentioned in KEV	1,120
CVEs with AttackerKB Assessments	1,108
CVEs with Exploit-DB PoCs	24,587
CVEs with Mails	42,030
CVEs with Patch Files	10,548
CWE – CAPEC	336 / 450
CAPEC – ATT&CK	177 / 36
ATT&CK – D3FEND	301 / 121

n_1* - *n_2 denotes bidirectional relationships

Example: CVE-2020-7247



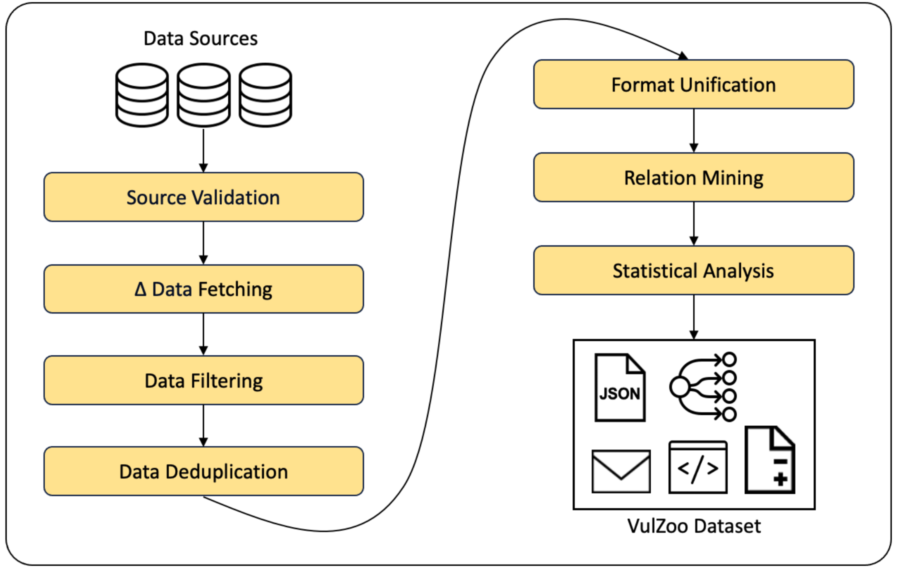
Detailed View



CVE	smtp_mailaddr in smtp_session.c in OpenSMTPD r6.6, as used in OpenBSD 6.6 ..., allows remote attackers to execute arbitrary commands as root via a crafted SMTP session...
CPE	cpe:2.3:a:openbsd:opensmtpd:6.6:*:*:*:*:*:*:* (5 more CPE strings omitted)
CVSS	CVSS:3.1/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H (CRITICAL) CVSS:2.0/AV:N/AC:L/Au:N/C:C/I:C/A:C (HIGH)
CWE	755: Improper Handling of Exceptional Conditions 78: Improper Neutralization of Special Elements used in an OS Command
KEV	Added Time: 2022-03-25 Action Due Date: 2022-04-15 Name: OpenSMTPD Remote Code Execution Vulnerability Required Action: Apply updates per vendor instructions
AttackerKB	Attacker-value: 5 (Very High) Exploitability: 5 (Very High) Tags: Easy to weaponize, Gives privileged access, Vulnerable in default configuration
Patch	commit: 9dcfda045474d8903224d175907bfc29761dcb45 diff --git a/usr.sbin/smtpd/smtp_session.c b/usr.sbin/smtpd/smtp_session.c
PoC	exploits/linux/remote/{48038.rb,47984.py} exploits/openbsd/remote/48051.pl
Mail	2020-01-28 LPE and RCE in OpenSMTPD (CVE-2020-7247) 2020-01-30 [SECURITY] [DSA 4611-1] opensmtpd security update
CAPEC	6: Argument Injection 15: Command Delimiters 88: OS Command Injection 43: Exploiting Multiple Input Interpretation Layers 108: Command Line Execution through SQL Injection

Profile Overview

Data Sources and Collection


Data Sources

Source Validation

Δ Data Fetching

Data Filtering

Data Deduplication

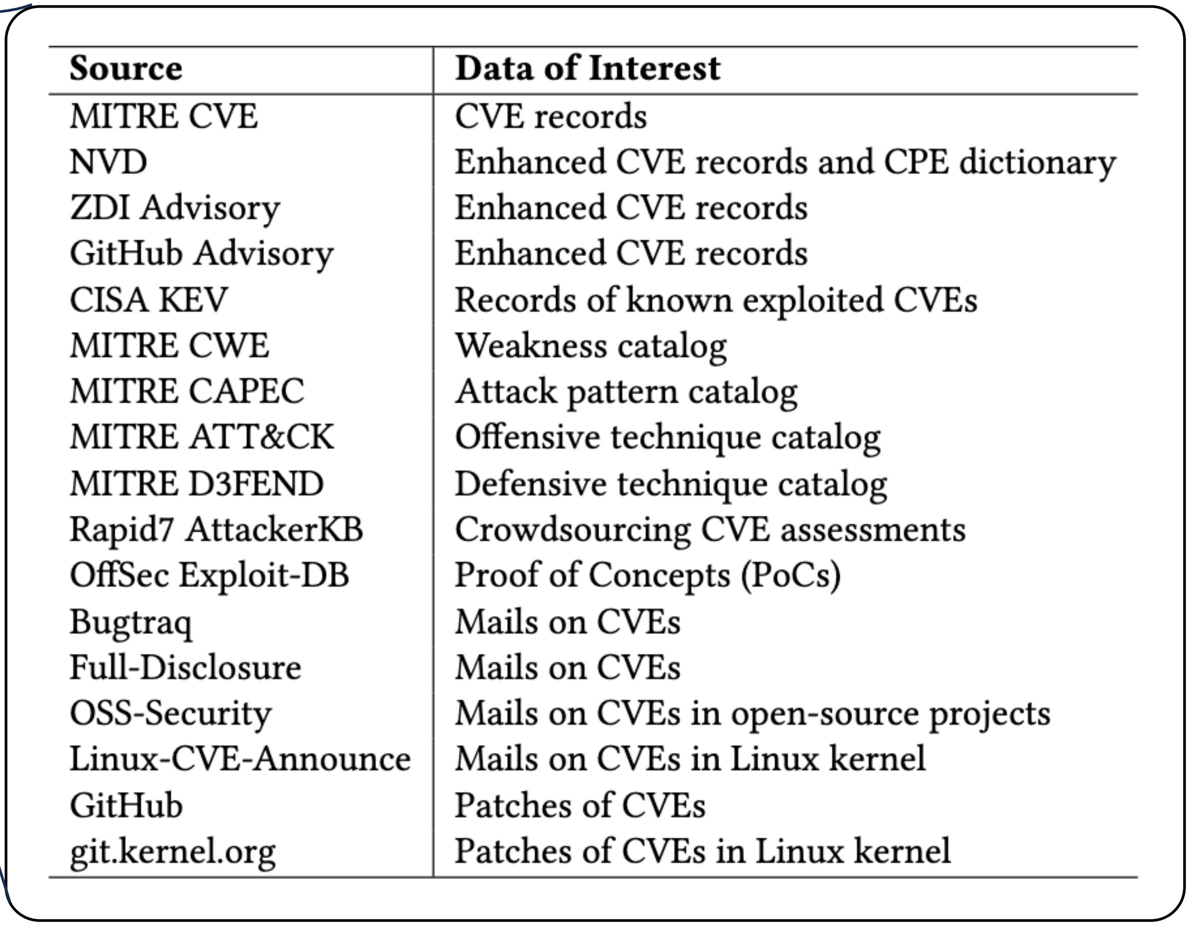
Format Unification

Relation Mining

Statistical Analysis

VulZoo Dataset

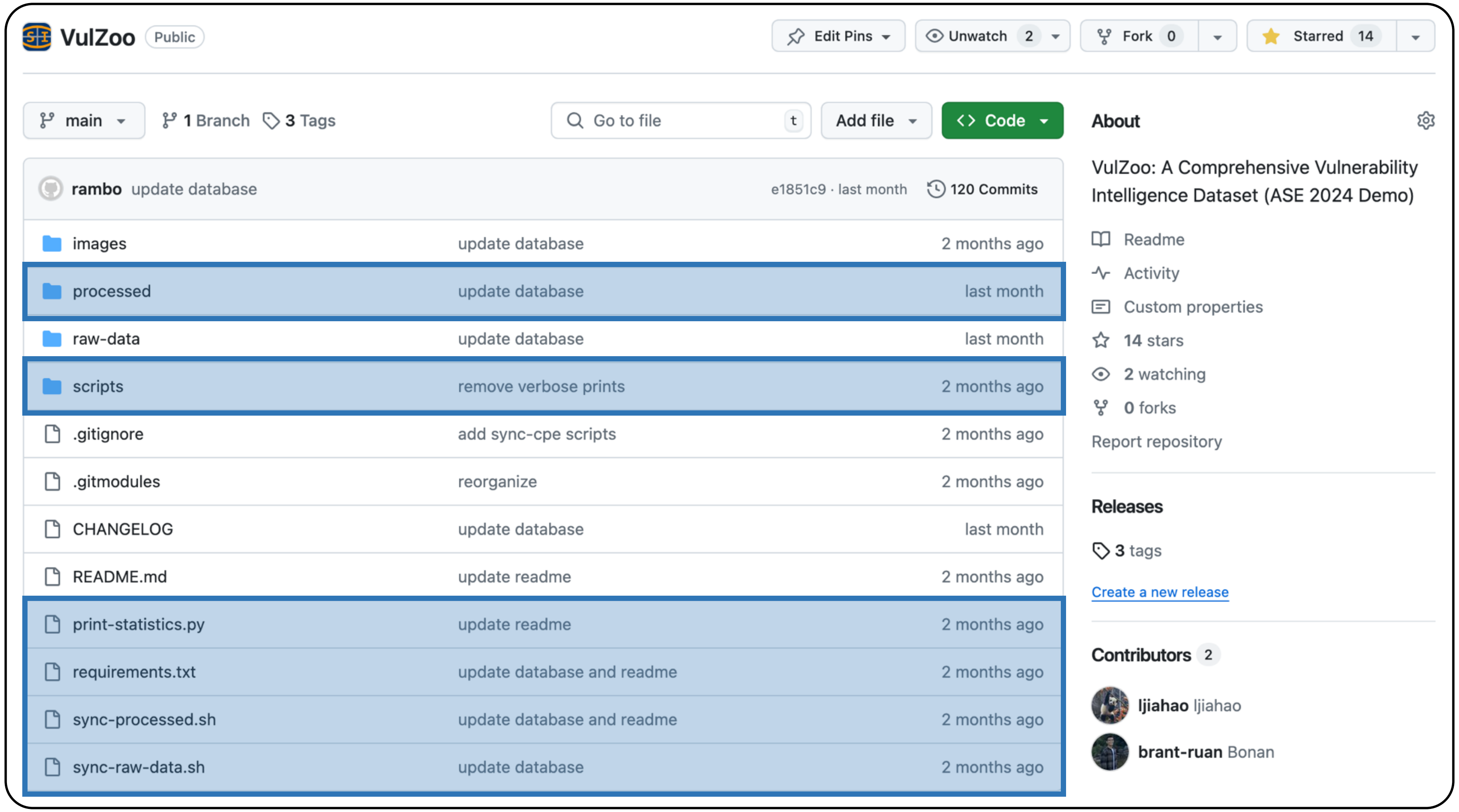
Collection Process



Source	Data of Interest
MITRE CVE	CVE records
NVD	Enhanced CVE records and CPE dictionary
ZDI Advisory	Enhanced CVE records
GitHub Advisory	Enhanced CVE records
CISA KEV	Records of known exploited CVEs
MITRE CWE	Weakness catalog
MITRE CAPEC	Attack pattern catalog
MITRE ATT&CK	Offensive technique catalog
MITRE D3FEND	Defensive technique catalog
Rapid7 AttackerKB	Crowdsourcing CVE assessments
OffSec Exploit-DB	Proof of Concepts (PoCs)
Bugtraq	Mails on CVEs
Full-Disclosure	Mails on CVEs
OSS-Security	Mails on CVEs in open-source projects
Linux-CVE-Announce	Mails on CVEs in Linux kernel
GitHub	Patches of CVEs
git.kernel.org	Patches of CVEs in Linux kernel

17 Data Sources

GitHub Repository



VulZoo Public

Edit Pins · Unwatch 2 · Fork 0 · Starred 14

main · 1 Branch · 3 Tags · Go to file t · Add file · Code

rambo update database e1851c9 · last month 120 Commits

images	update database	2 months ago
processed	update database	last month
raw-data	update database	last month
scripts	remove verbose prints	2 months ago
.gitignore	add sync-cpe scripts	2 months ago
.gitmodules	reorganize	2 months ago
CHANGELOG	update database	last month
README.md	update readme	2 months ago
print-statistics.py	update readme	2 months ago
requirements.txt	update database and readme	2 months ago
sync-processed.sh	update database and readme	2 months ago
sync-raw-data.sh	update database	2 months ago

About

VulZoo: A Comprehensive Vulnerability Intelligence Dataset (ASE 2024 Demo)

- Readme
- Activity
- Custom properties
- 14 stars
- 2 watching
- 0 forks

Report repository

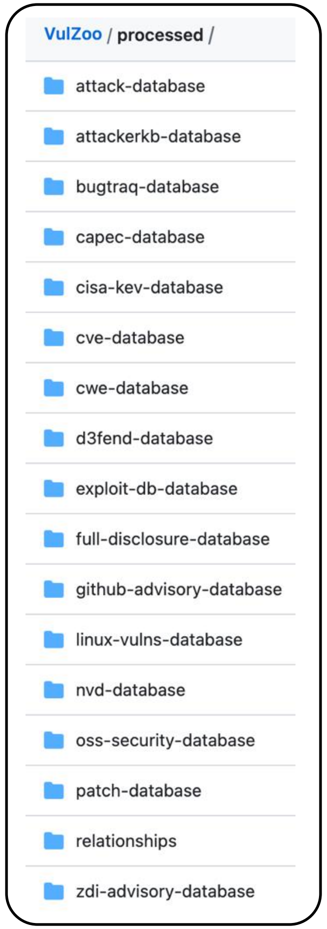
Releases

3 tags

Create a new release

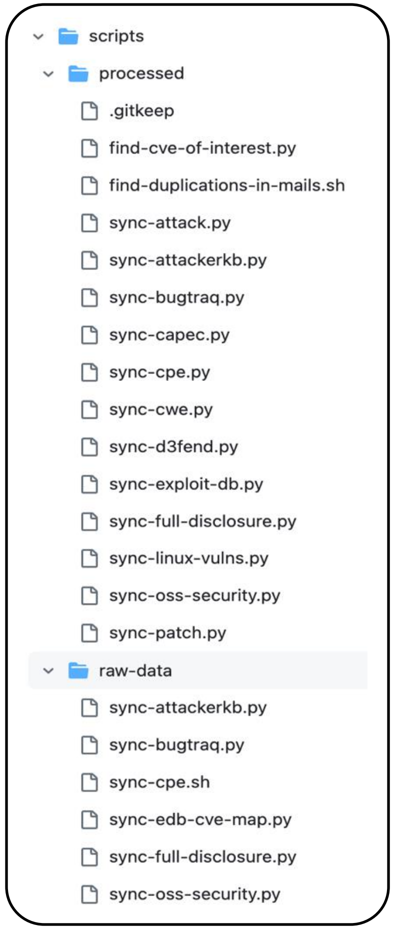
Contributors 2

- ljiahao ljiahao
- brant-ruan Bonan



VulZoo / processed /

- attack-database
- attackerkb-database
- bugtraq-database
- capec-database
- cisa-kev-database
- cve-database
- cwe-database
- d3fend-database
- exploit-db-database
- full-disclosure-database
- github-advisory-database
- linux-vulns-database
- nvd-database
- oss-security-database
- patch-database
- relationships
- zdi-advisory-database



- scripts
  - processed
    - .gitkeep
    - find-cve-of-interest.py
    - find-duplications-in-mails.sh
    - sync-attack.py
    - sync-attackerkb.py
    - sync-bugtraq.py
    - sync-capec.py
    - sync-cpe.py
    - sync-cwe.py
    - sync-d3fend.py
    - sync-exploit-db.py
    - sync-full-disclosure.py
    - sync-linux-vulns.py
    - sync-oss-security.py
    - sync-patch.py
  - raw-data
    - sync-attackerkb.py
    - sync-bugtraq.py
    - sync-cpe.sh
    - sync-edb-cve-map.py
    - sync-full-disclosure.py
    - sync-oss-security.py

URL: https://github.com/NUS-Curiosity/VulZoo

Ongoing Efforts

- Overall Goal:
  - Gain better intelligence of cybersecurity vulnerabilities
  - Understand vulnerabilities and attacks through experimentation
- Kernel Vulnerability Reproduction:
  - KernJC: Automated Vulnerable Environment Generation for Linux Kernel Vulnerabilities (RAID 2024)
- Infrastructure emulation and attack recreation
  - Audit log dataset
  - Network traffic
  - Application traces

National
Cybersecurity R&D
Laboratory

Conclusion

- Vulnerability Intelligence Dataset
  - ✓ Heterogeneous vulnerability information
  - ✓ Profile vulnerabilities better
- Utility Scripts
  - ✓ Data synchronization
  - ✓ Data cleaning
  - ✓ Relationship mining
  - ✓ Statistics generation
- Application Scenarios
  - ✓ Assessment & prioritization
  - ✓ Information alignment
  - ✓ Information augmentation
  - ✓ ...

https://github.com/NUS-Curiosity/VulZoo

Use and feedback are welcome!

Thank you!

r-bonan@comp.nus.edu.sg