

Guard Your AI+Software Supply Chain with Sectrend



Wenhui Jin
Cybersecurity Expert, Sectrend



Table of CONTENTS

- Sectrend Stories
- Contribution to Global Open Source Community
- OSS Security and Compliance Empowered by CleanSource SCA
- AI is Reshaping SCA 2.0
- We are Embracing the Future

A decorative graphic consisting of a solid blue circle with the white number "01" inside. This circle is surrounded by a dashed white circle and a solid blue arc that partially encircles it.

Sectrend Stories



Sectrend is a leading provider of AI + software supply chain governance tools and services.

We specialize in building AI + DevSecOps tool chain and are dedicated to providing world-class application security products.



Founded:

June 2021

FTE: 80

(70 SWEs)

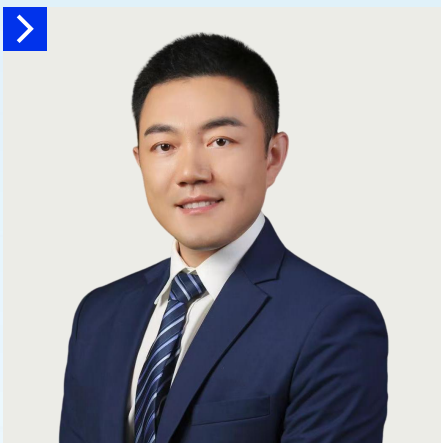
R&D team are mostly from Alibaba, Huawei, ZTE, OPPO, etc.

HQ:

Shanghai,
China

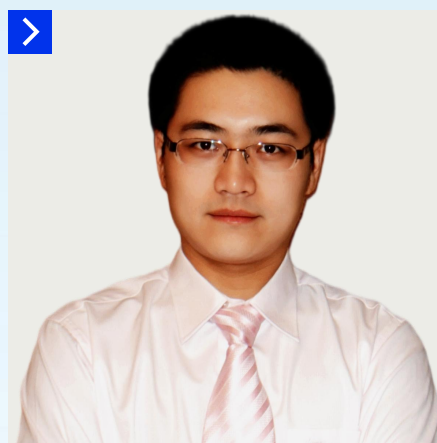
Offices:

Beijing,
Shenzhen



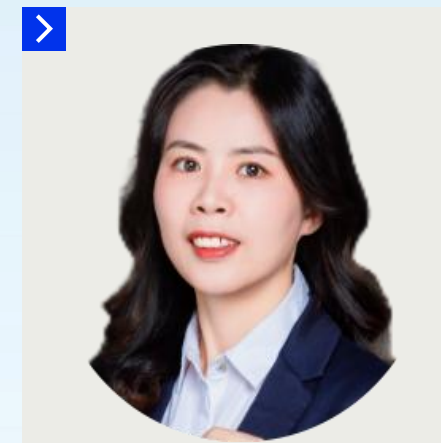
Alex Xue | CEO, Founder

- Former Checkmarx Greater China GM
- Former Greater China Business Leader of SIG at Synopsys
- Leader of OpenAtom DevOps Group
- Council Member of KAIYUANSHE
- Draft China DevSecOps standards



King Gao | SVP, Co-founder

- China's Top Open Source Expert
- Former Huawei Open Source Governance Expert
- Extensive experience in open source governance, compliance, and supply chain security across international and domestic initiatives
- Actively engaged in open source communities and contributed to industry standard development



Xianman Zhu | Dir. of Arch.

- Former ZTE Head of Open Source Governance
- Former Lead of Open Source Governance Implementation & Practice at OPPO
- Expert in R&D, Product Development, and Open Source IP Strategy



Safeguard Multi-Dimensional AI + SW Supply Chain



01 CleanSource SW Composition Analysis

Generate complete
Software Bill of Materials (SBOM)

Identify known vulnerabilities in
3rd party components

Manage security and
compliance risks in OSS

Source code and binary
(Powered by CleanBinary)

02 PureStream AI Risk Management

Generate complete
AI Bill of Materials (AI BOM)

Address privacy, copyright, and
content security risks in AIGC

Promote AIGC
compliance and audit

Enhance AIGC transparency in
critical infrastructure sectors

03 CleanCode Static Analysis

Identify critical code
vulnerabilities and weaknesses

Standards Compliance
(MISRA, AUTOSAR, 21434, etc.)

OWASP Top 10 & CWE Top 25

Ensure code quality and security

04 Consulting Open Source Governance

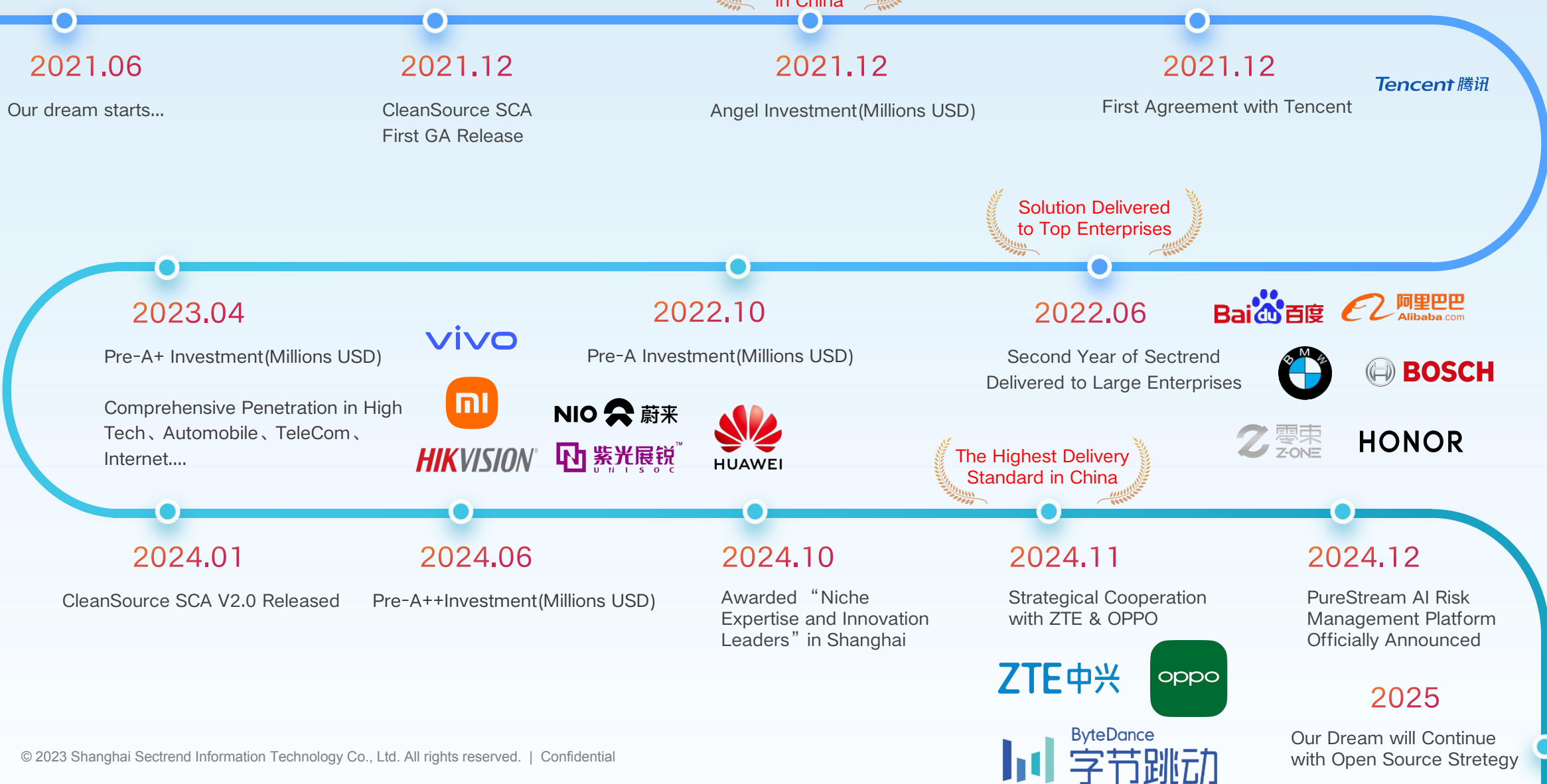
Assist in building
open source governance teams

Help establish open source
management systems

Assist in embedding supply
chain security and compliance
tools into R&D systems

Goals:
Security and Compliance
lead to Client Success

Our Story Starts from 2021.06



A decorative graphic consisting of a solid blue circle with the white number "02" inside. This circle is surrounded by a dashed blue circle above it and a solid blue arc to its right.

Contribution to the Global Open Source Community



Linux Foundation

- Member of Linux Foundation
- Member of OpenSSF

OpenChain

- Member of OpenChain and is actively promoting OpenChain adoption in China.
- Participated in the development of the SPDX 3.0 standard and previously incubated projects such as AI BOM maintainer and gopi.

Local Organization

- Lead the security and devops SIG in OpenAtom Foundation
- Help expand and enhance the security and compliance capabilities of openAtom Foundation

Open Source Community

- CHAOSS Maintainer
- Organize the first CHAOSS meeting in China
- Actively involve in openEuler, openHarmony and openAnolis, help them continuously build the capability

Standard Drafting - ISO / China

- ISO 18974
- 《Cybersecurity technology Software supply chain security requirements》 GB/T 43698-2024 - The first software supply chain standard in China

Academia

- Collaborate with universities on new tech research and talent acquisition

Fudan University
Shanghai Jiaotong University
East China Normal University

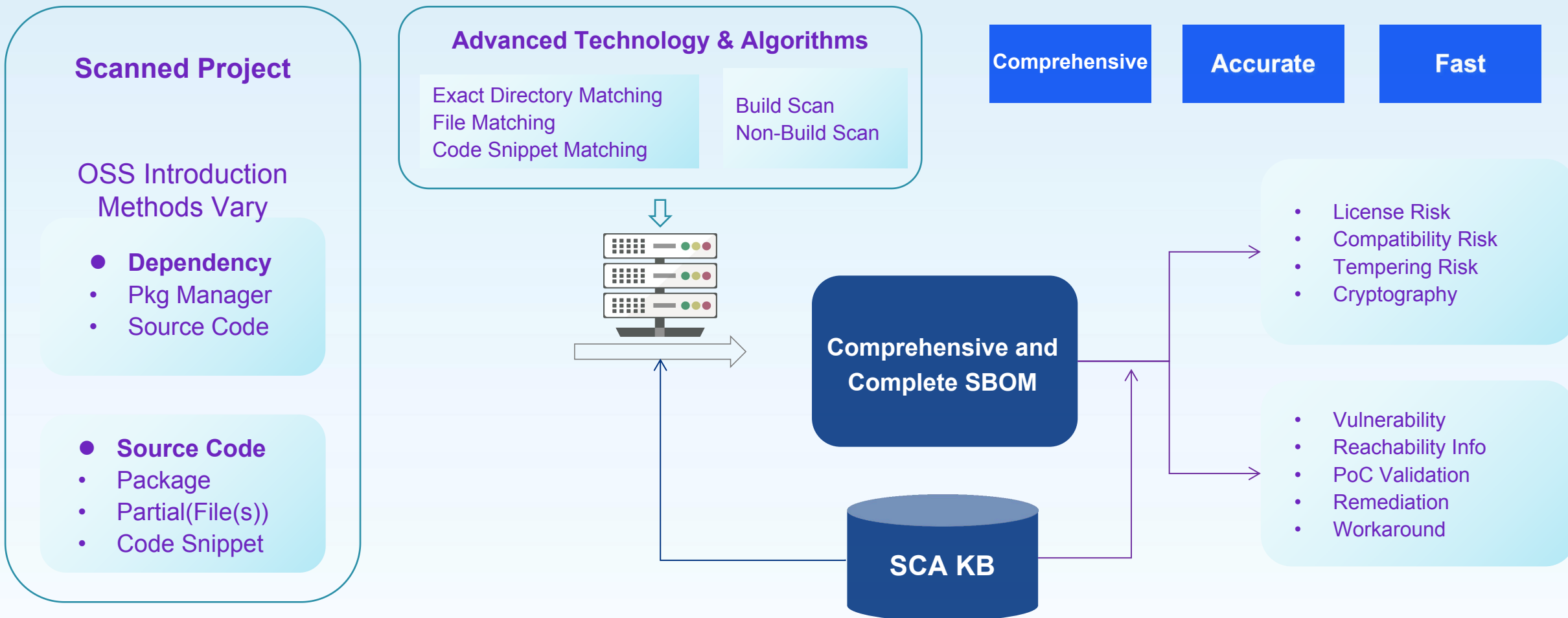


Our Open Source Future Starts from NOW



03

OSS Security and Compliance Empowered by CleanSource SCA





Compliance Risk Identification & Reporting

License Risk

- Dynamic risk levels based on distribution scenarios, usage and license terms to facilitate timely identification of high-risk OSS

License Compatibility Risk

- Top 110*110 License Matrix
- Risk/no risk/uncertain at a glance, reducing the workload of legal team

License Tempering Risk

- Reduce OSS compliance risk and improve regulatory efficiency by identifying files with tempered license or copyright

Notice File Generation

- Notice file export in different formats to empower open source compliance obligations

SBOM Generation

- Enables traceability and assists in quickly locating vulnerable OSS
- Meet customer or regulatory requirements

Cryptography Identification

- Identify cryptographic algorithms used in open source components to help organizations achieve export control compliance



Up-to-date Vulnerability DB with Comprehensive and Accurate Info Helps Quickly Locate, Fix and Monitor Vulnerability



Vulnerability

- Risk Level
- Affected Entity
- Vulnerability Incident

Remediation

- Clear Function Call Relationship
- Exploitable PoC
- Fix Suggestion(Component/File)
- Patch/Affected Code Snippet
- Vulnerability Reachability

Awareness

- Real Time Updated Database
- Vulnerability Alerts to Improve Awareness
- Continuous Monitoring Based on SBOM

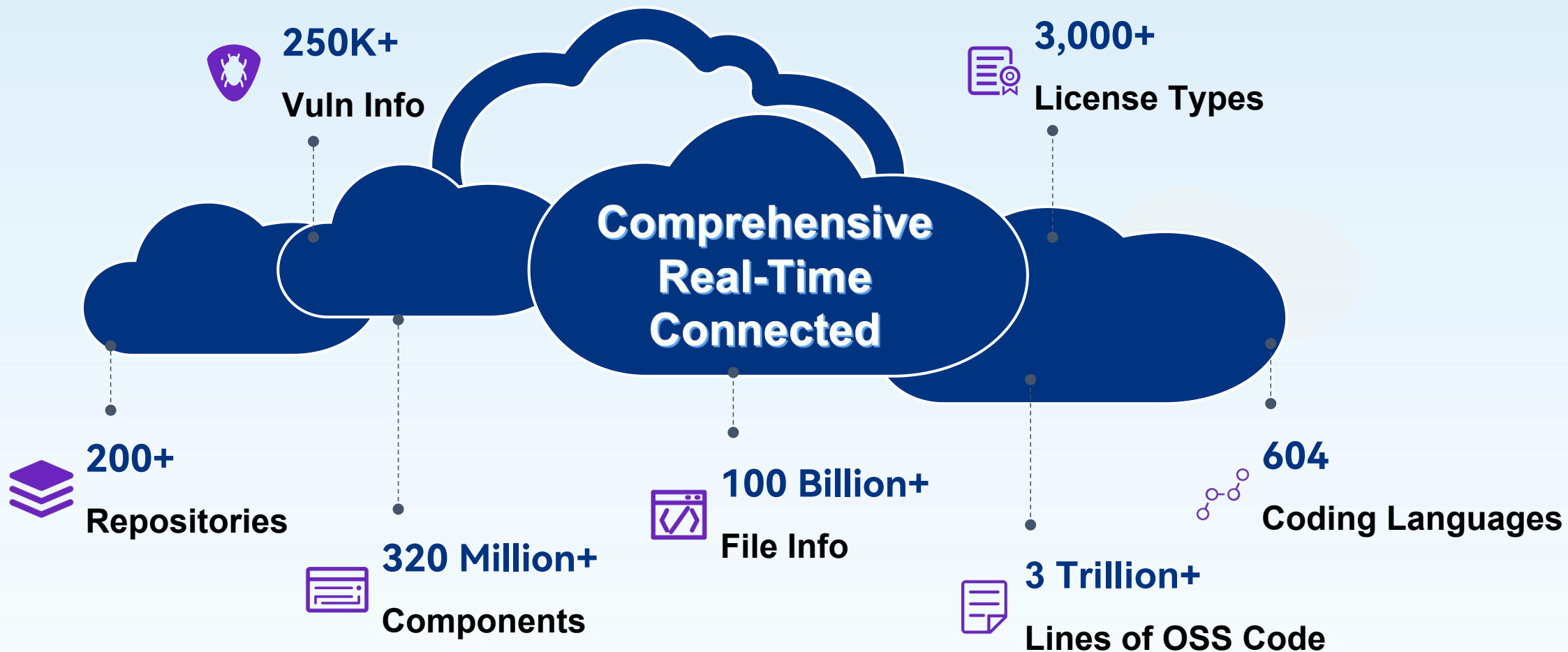
End-to-End DevOps Integration



...more to come in each release



Comprehensive OSS Component Database



Best Practice – CleanSource SCA Helps Xiaomi TOC with a Eased Software Supply Chain Management

Pain Points

Solutions

Achievements

Cross BU Collaboration:
Phone, Auto, IoT, Cloud

OSS Introduced via Different Ways



Lack of Guidance on the Best Practices

Super Large Files with OS Failed or Speed Limited

Code Exposure & Cross Border Data Exchange

Deep integration into development platform with API and DevOps plugins

All forms of OSS identification including component, code snippet, dependency, etc

Performed project decomposition with streaming scan to reduce resource consumption and balance scan efficiency(400GB)

CleanSource SCA(On-Prem)

Expert consulting services on open source governance to ensure efficient deployment of tools

Protected Software Supply Chain from the Bottom Up

Improved OSS Security and Compliance

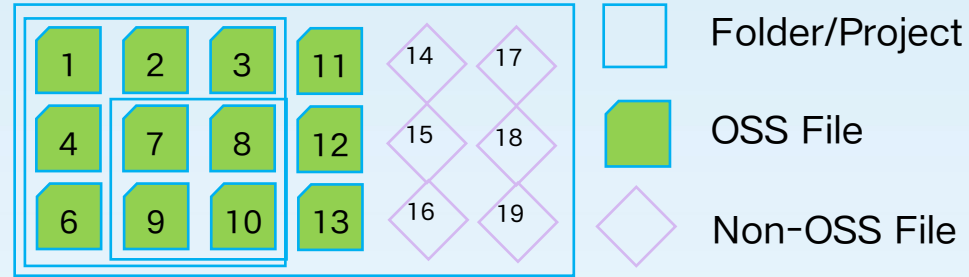
Reduced Intellectual Property Risk

Integrated into Development Process

Enhanced Brand Image

Tech Highlight 1: Exact Directory

Input



Before

After

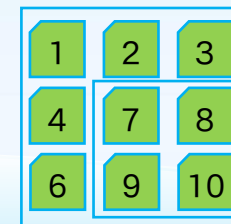
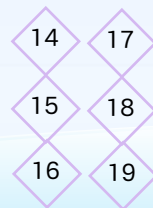
File Match

Code Snippet Match

Exact Directory

File

Snippet



With exact directory scan, we can get:

- More converged and accurate results
- Inter-components lineage relationship
 - 30%+ speed improvement



Tech Highlight 2: Component Identification System



WHAT

- Leveraging AI capabilities to merge different distributions of the same component into a unified CID.
- This process involves building a knowledge graph to perform knowledge inference on attributes such as licenses, vulnerabilities, and EOL.

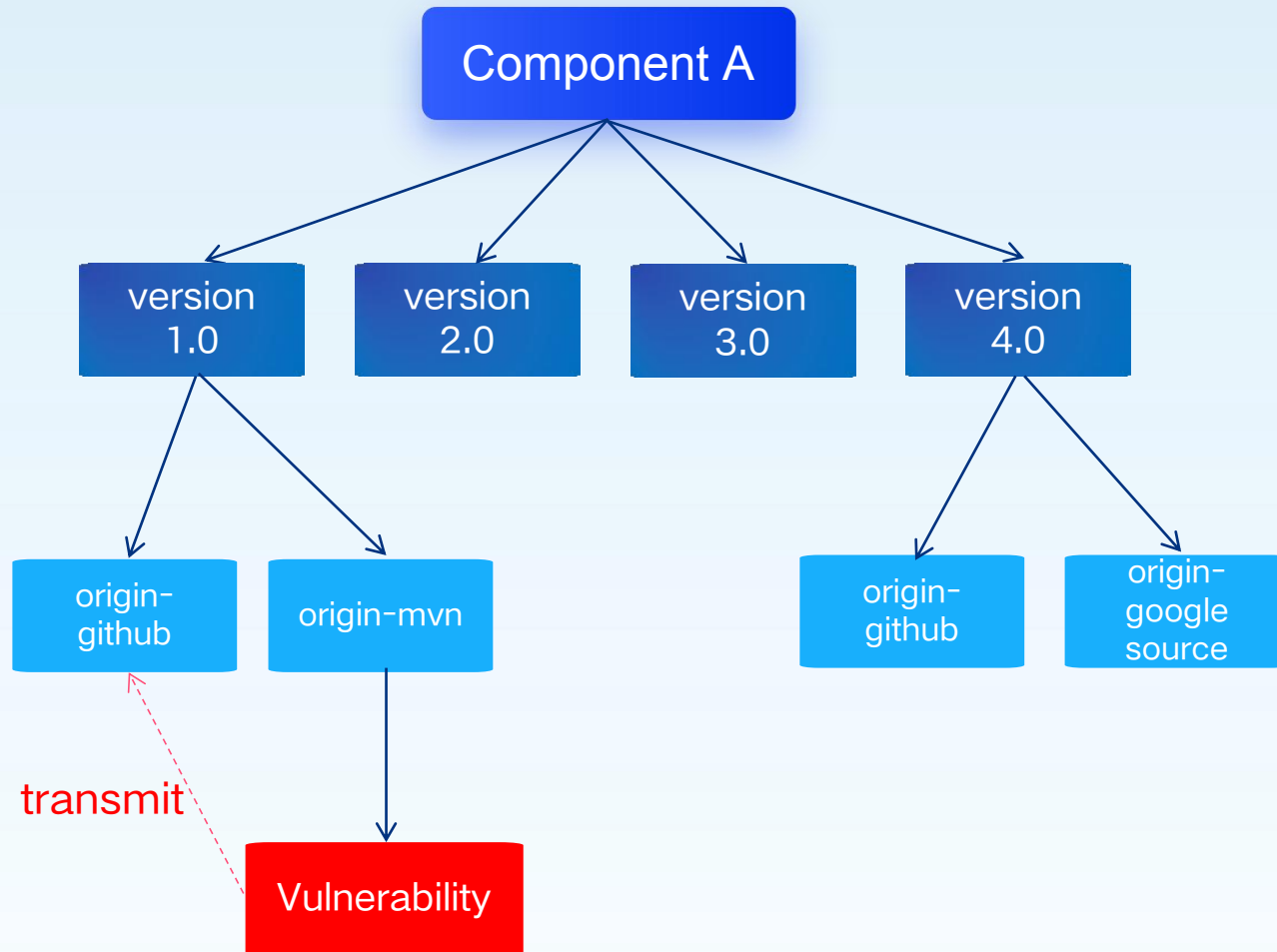
VALUE

Accuracy

Fast

Efficiency

Real Time

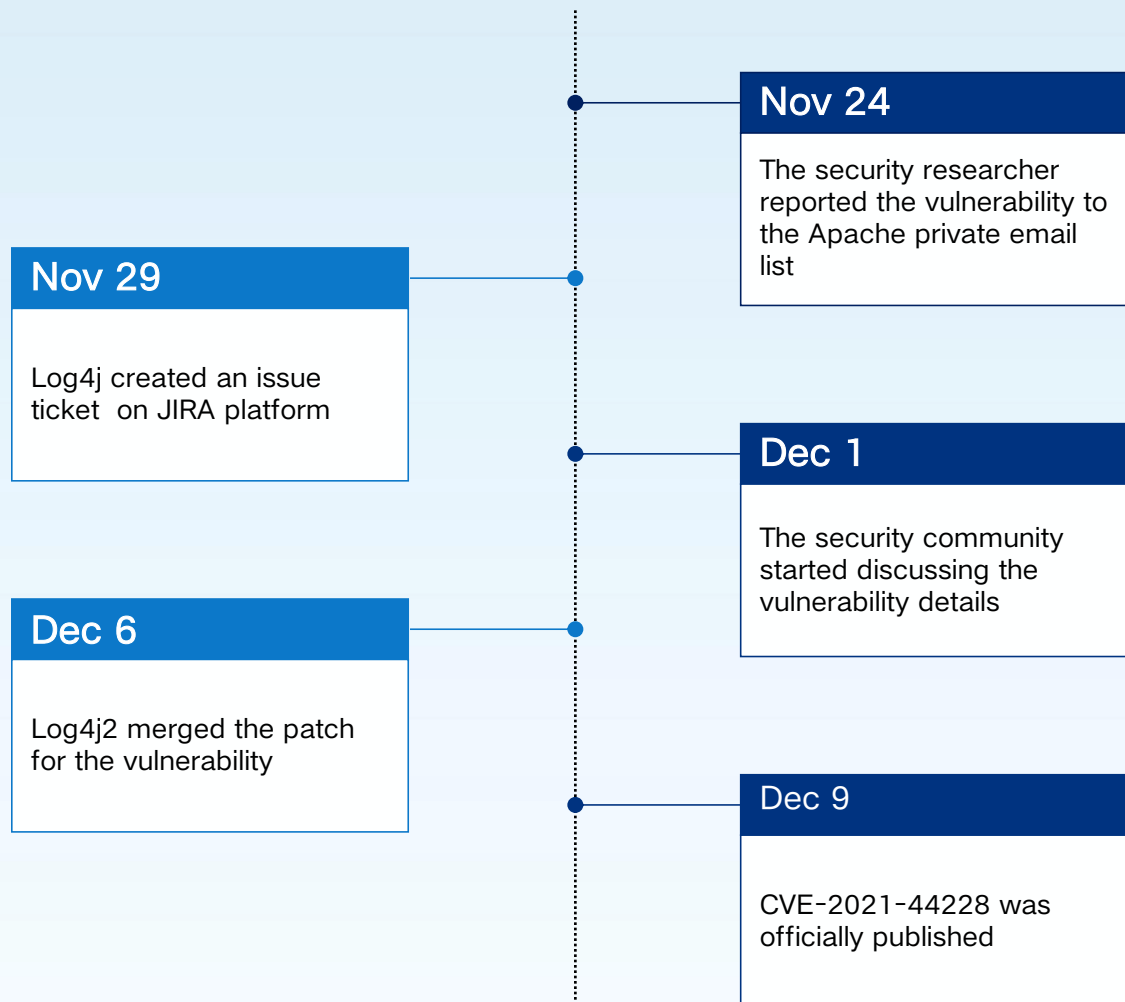




Tech Highlight 3: 0-day Vulnerability Intelligence

Identify potential vulnerabilities before disclosure

- Some open-source vulnerabilities may not be assigned a CVE number
- Relying solely on CVEs or community security advisories lacks timeliness; by monitoring publicly available information in open-source repositories, potential CVEs can be identified in advance
- Researchers or developers may inadvertently disclose vulnerability details in open-source communities, such as GitHub Issues





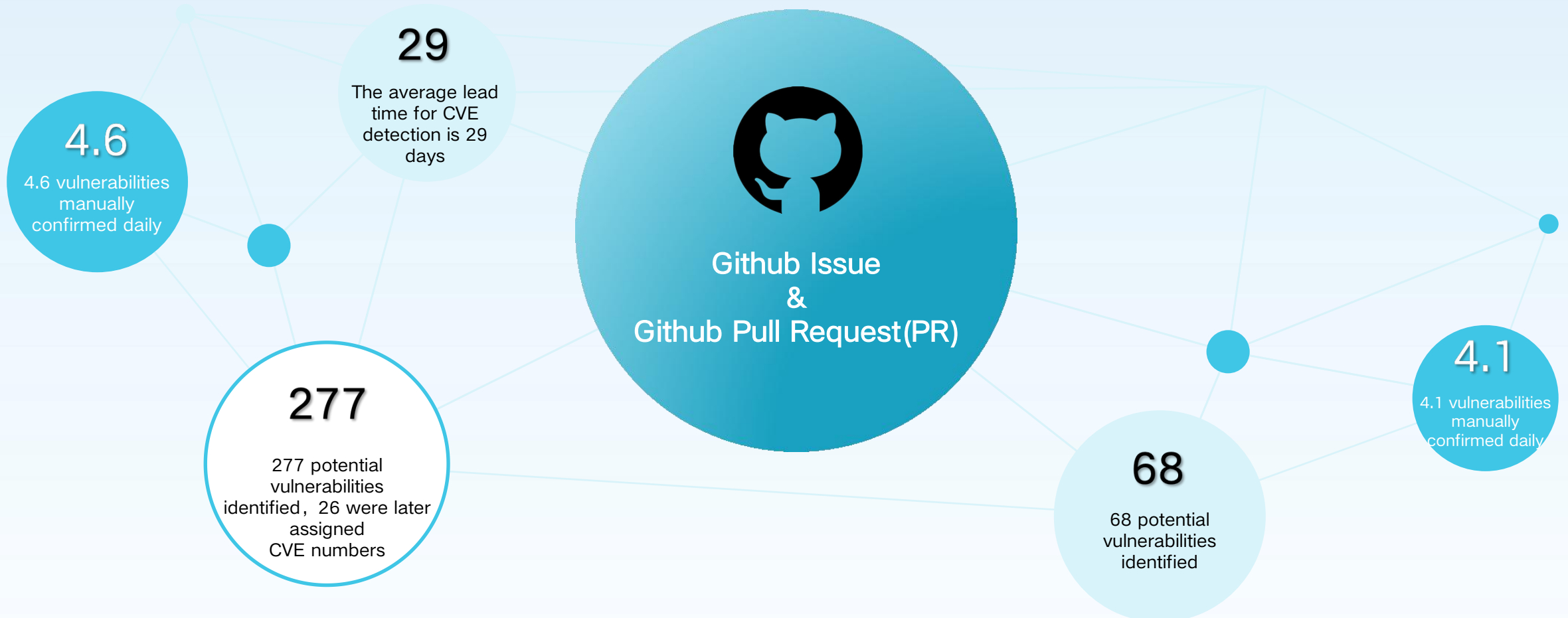
Tech Highlight 3: 0-day Vulnerability Intelligence

Github Issue

Monitored GitHub issues from Nov 2024 - Jan 2025

Github Pull Request(PR)

Monitored GitHub PRs from Feb 2024 - Now





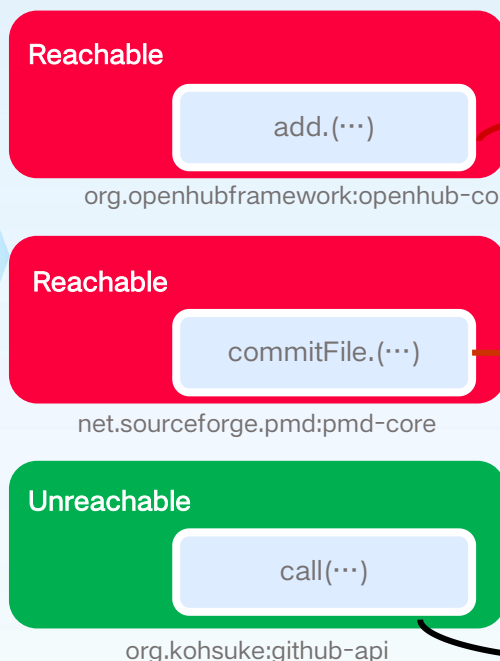
Tech Highlight 4: Vulnerability Reachability

Enhanced Vulnerability Analysis, Accuracy & Fix Efficiency

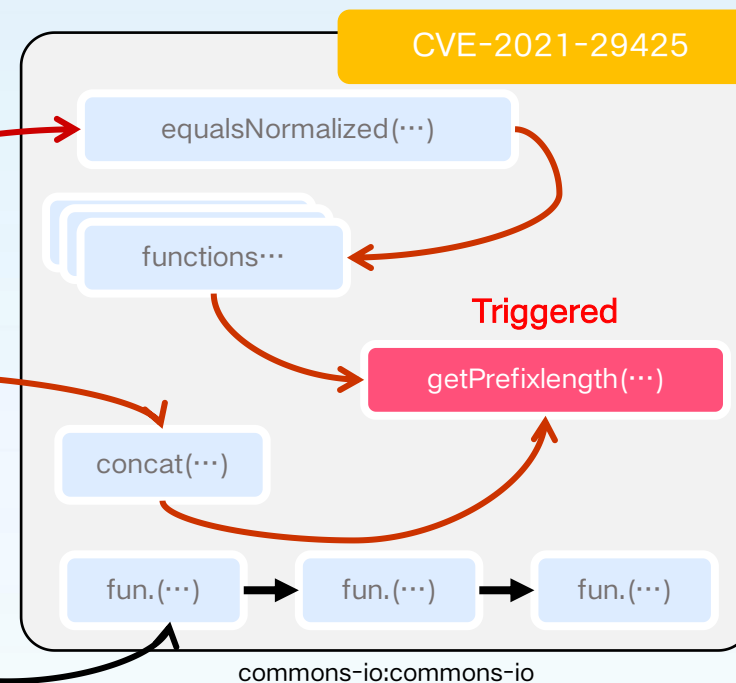
Function-level vulnerability reachability is of critical importance. It not only helps reduce security risks in the software supply chain but also enhances the speed and efficiency of responding to and addressing security vulnerabilities.



Downstream Project



OSS Component (with Vulnerability)



xstream-CVE-2021-29505 / xstream-CVE-2021-29505 / defaultVersion

Distribution: External | Stage: Developing | Project Licenses 1

Unconfirmed... 0 | License 2 | Total OSS Files 0 | Cryptography 0 | Vulnerability 32 | Alert 0 | License Risk 0 | Self-researc... 100.00%

Component | Source | **Vulnerability** | Policy | Sensitive Information | Report | History

Q Please Inp... | Please Input C...
Please Select Risk Level

xstream 1.4.15
30 Known Vulnerability | Remediation Suggestion: Upgrade to version 1.4.21

commons-collections 3.1
Comp... [https://mvnrepository.com/...](https://mvnrepository.com/)
Vulner... 1 1

Risk Level Please Select ... | CVE Please Input... | Vulnerability Reachability All | CWE Please Input...
Fix Status Please Select ...

Set Fix Status in Batch | 1 Items Selected

xstream 1.4.15
Comp... [https://mvnrepository.com/...](https://mvnrepository.com/)
Vulner... 7 22 1

<input type="checkbox"/>	Vulnerability	Level	Vulnera... <i>i</i>	CVE	CNNVD	Release Date	Operate
<input type="checkbox"/>	XStream 代码问题漏洞	High	Reachable	CVE-2021-21349	CNNVD-202103-1236	2021-03-23 08:15	Fix Status
<input type="checkbox"/>	XStream 资源管理错误漏洞	High	Reachable	CVE-2021-21348	CNNVD-202103-1238	2021-03-23 08:15	Fix Status
<input type="checkbox"/>	XStream 缓冲区错误漏洞	High	Unreachable	CVE-2022-40151	CNNVD-202209-1234	2022-09-16 18:15	Fix Status
<input checked="" type="checkbox"/>	XStream 代码问题漏洞	High	Reachable	CVE-2021-29505	CNNVD-202105-1981	2021-05-29 05:15	Fix Status
<input type="checkbox"/>	XStream 安全漏洞	Medium	Unreachable	CVE-2021-39140	CNNVD-202108-1900	2021-08-24 03:15	Fix Status

Dashboard

Tasks

Elements

Policies

Wiki

Report

System

Manage

High XStream 代码问题漏洞

CVE-2021-29505 | CNNVD-202105-1981 | Release Date: 2021-05-29 05:15 | Update At: 2023-11-07 11:32

File Path

src/main/java/Main.java

src.../Main.java

Line

Local Source Code

66

```
50         <short>0</short> \n" +
51         <boolean>>false</boolean> \n" +
52         </java.rmi.server.RemoteObject> \n" +
53         </registry> \n" +
54         <host>127.0.0.1</host> \n" +
55         <port>1099</port> \n" +
56         </ctx> \n" +
57         </candidates> \n" +
58         </aliases> \n" +
59         </nullIter> \n" +
60         </sm> \n" +
61         </message> \n" +
62         </value> \n" +
63         </javax.naming.ldap.Rdn_-RdnEntry> \n" +
64         </java.util.PriorityQueue> \n" +
65         "</java.util.PriorityQueue> ";
66         xStream.fromXML(xml);
67         //CommonsBeanutils1 cb = new CommonsBeanutils1();
68         //String xml = xStream.toXML(cb.getObject("cmd.exe /c echo 111>99999999.txt"));
69         System.out.println(xml);
70     }
71 }
```


A decorative graphic consisting of a solid blue circle with the white number "04" inside. This circle is surrounded by a dashed blue circle and a solid blue arc that partially encircles it.

AI is Reshaping SCA 2.0



AIGC is Reshaping the World and the Future SCA

Commercial use prohibited; only permitted for research, study, and similar purposes. Certain scenarios are explicitly prohibited.

Lawsuit

Mandate to maintain copyright, license, and similar provisions

comply with corresponding ethical and legal requirements.

Requires a statement on the product or website

Prohibited from using the output to train other models.

Model License Restrictions

Model License Risk

Model License Compliance

Waste of Resources

Goodwill Damage

Modification statement is required

Specific requirements for model naming

Attribution notice is required (attribution notice)

Restriction on Commercial Scale: If the licensee's product or service exceeds a certain threshold of monthly active users (e.g., >700 million), they must apply for a license from the original licensor.

Data Leak

Personal privacy, trade secrets, and similar matters.

Copyright Violation

Some training data may not have been legally authorized and could potentially involve copyright-protected data.

Content Security

Resulting in the product being taken down or facing regulatory penalties.

Product(PureStream) Main Features

PureStream clearly lists dataset elements and their sources, helping trace origins, mitigate legal risks, and ensure compliance. It details dataset names, risky content ratios, licenses, copyrights, and sensitive personal information.

PureStream enables precise traceability of models, displaying a list of local model names along with corresponding matched model names, types, official distribution sources, code repositories, licenses, official links, methods, use cases, model architectures, base models, datasets, applicable scopes, Commit IDs, release dates, and more.

PureStream enables model searches by source and type, helping users understand model features, performance, and usage details. Its knowledge base displays model names, types, sources, licenses, architectures, datasets, scopes, and revision records (Commit ID & release date).

Multi-dimensional corpus comprehensive detection

Dataset	Risky Content Ratio	Risky Content Count	License Count
CulturaX	0.0051%	64	4
The Stack	0.0032%	60	2
Cirrussearch wiki	0.0022%	32	1
peS2o	0.0009%	31	1

Model risk detection

Model Name	Status	Size	Creator	Creation Time	Build Time
test model 1	扫描完成	388.88 MB	admin	2024-12-12 16:00:00	2024-12-12 16:00:00
test model 2	扫描完成	388.88 MB	admin	2024-12-12 16:00:00	2024-12-12 16:00:00
test model 3	扫描完成	388.88 MB	admin	2024-12-12 16:00:00	2024-12-12 16:00:00
test model 4	扫描完成	388.88 MB	admin	2024-12-12 16:00:00	2024-12-12 16:00:00
test model 5	扫描完成	388.88 MB	admin	2024-12-12 16:00:00	2024-12-12 16:00:00
test model 6	扫描完成	388.88 MB	admin	2024-12-12 16:00:00	2024-12-12 16:00:00
test model 7	扫描完成	388.88 MB	admin	2024-12-12 16:00:00	2024-12-12 16:00:00

Model Knowledge Base Selection

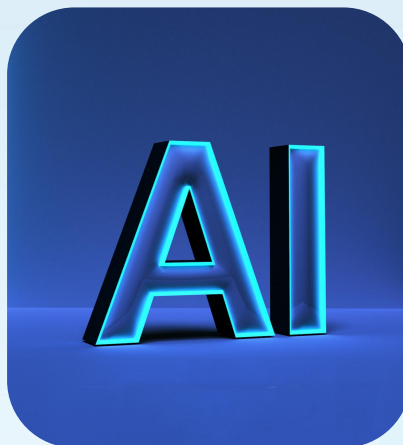
Source	Model Name	Type
HuggingFace	stable-diffusion-v1-4	diffusers
Multimodal	Meta-Llama-3-8B	transformers
Computer Vision	bloom	transformers
Natural Language Processing	Mixtral-8x7B-Instruct-v0.1	transformers
Audio	Llama-2-7b-chat-hf	transformers



Technology Advantages

LLM Recognition ModelCard

Using pre-trained large models' unsupervised learning or named entity recognition (NER) capabilities, key entities and relationships in ModelCards (e.g., model-dataset usage or version-performance metrics) are identified and integrated into a complete, visually structured format for user-friendly understanding.



Multi-Source Dataset Collection

PureStream collects datasets from diverse sources, supporting around 20 mainstream datasets like Wikipedia, Zhihu, and CC100. It detects AIGC-generated content, assessing risks such as copyright issues, privacy concerns, and content safety.



Model Risk Identification

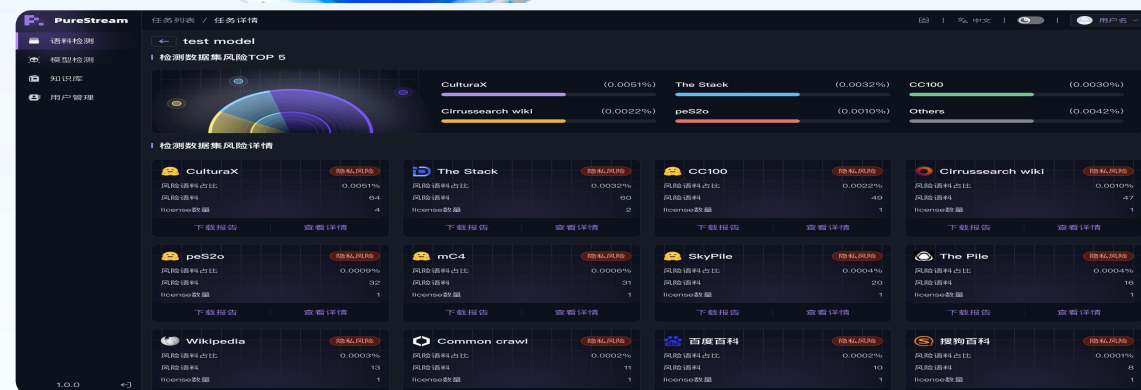
PureStream supports model detection traceability and assists in model selection. It currently includes 5 model sources and enables retrieval of over 1,000 model types. By constructing a model derivative relationship graph, it establishes communities for high-imitation models and untrusted models, analyzes model licenses and copyright risks, and traces official links, code repositories, and download addresses of the used models.

Dimensionality Reduction and Compression

To protect user data, before uploading a model for detection, users can use PureStream's in-house dimensionality reduction and compression engine to locally process the target into an irreversible encrypted fingerprint. The model host remains offline, and the platform never accesses the source code, ensuring data security.

Complex Multi-Scenario Model Detection

PureStream uses feature extraction to analyze fused layers, determining original feature dimensions and representations. It leverages feature space similarity, calculating distances or mapping features to align detection techniques with fused or segmented models. By comparing statistical characteristics of original and modified model parameters, it identifies optimal detection strategies.





Super Large Scale KnowledgeBase(PB-Scale)



Data Set

CulturaX

The Stack

CC100

mC4

Cirrussearch
wiki

peS2o

SkyPile

The Pile

Wikipedia

Common crawl

Baidu Baike

Sougou Baike

Zhihu

Quora

wanjuan

C4

Wudao

OLMO

project
gutenberg

openwebtext



Data Label

Dataset

Name, License, Confidentiality Level, Sensitive Information, Data Type, Purpose, Language, Size, etc.

File

Name, URL, Language, Copyright Information, Sensitive Information Label, Compressed File Location, Positioning, etc.

Paragraph

Paragraph Positioning, Sensitive Information Label



Data Format

parquet

jsonl

zip

rar

zstd

zst

Two small squares, one light blue and one teal, stacked vertically in the top left corner.

Thank you!

Feel free to contact us...



King Gao

kun.gao@sectrend.com.cn

Feng Wang (Director of Marketing) feng.wang@sectrend.com.cn

**Beyond Security, more than
open
source.
Thank you!**