CodePecker

安 全 解 决 方案

CodePecker 源代码缺陷分析系统 V7.0
产品白皮书

北京酷德啄木鸟信息技术有限公司

2025-01

目录

第一章 概述

CodePecker 软件供应链安全静态分析系统是一款专业级源代码安全审计工具，旨在帮助企业和开发团队在软件开发生命周期中及时发现并修复安全漏洞。该系统采用先进的静态分析技术，能够在不运行代码的情况下深度分析源代码，识别潜在的安全风险、代码缺陷和合规性问题。

系统支持多种主流编程语言，覆盖全面的安全检测标准，提供精准的缺陷定位和详细的修复建议，帮助企业提升代码质量和安全防护能力。

第二章 背景

随着数字化转型的深入推进，软件已成为各行各业的核心资产。然而，软件安全问题日益突出，代码层面的安全漏洞成为攻击者的主要入口。据统计，超过 70%的安全事件源于应用程序层面的漏洞，其中大部分可以通过静态代码分析在开发阶段被发现和修复。

传统的安全测试方法往往在软件开发生命周期后期进行，不仅成本高昂，而且修复难度大。因此，迫切需要一种能够在开发早期发现安全问题的解决方案。静态应用安全测试（SAST）技术应运而生，它能够在代码编写阶段就识别潜在的安全风险，从根本上提升软件安全性。

CodePecker 软件供应链安全静态分析系统正是基于这一需求而研发，致力于为企业提供全方位的代码安全保障。

第三章 技术架构

3.1. 架构概述

CodePecker 软件供应链安全静态分析系统采用分层架构设计，遵循高内聚、低耦合的原则，确保系统的可扩展性、可维护性和高性能。整个架构分为六个核心层次，每个层次承担特定的职责，通过标准接口进行通信。

3.2. 各模块能力与技术逻辑分析

3.2.1. 用户交互层

用户交互层是系统与用户直接交互的界面，提供多样化的访问方式。

模块名称	核心能力	技术逻辑
Web 管理界面	• 项目管理与配置 • 检测任务监控 • 结果查看与分析 • 系统管理功能	采用前后端分离架构，前端使用 Vue.js+Element UI，后端提供 RESTful API。支持实时任务状态更新和结果可视化展示。
IDE 插件	• 实时代码检测 • 缺陷即时提示 • 快速修复建议 • 开发环境集成	基于 LSP（Language Server Protocol）协议，支持 IntelliJ IDEA、Eclipse、VS Code 等主流 IDE。采用增量分析技术，减少开发中断。
REST API 接口	• 自动化任务调度 • 检测结果获取 • 系统状态查询 • 第三方集成	基于 OpenAPI 规范，提供完整的 API 文档。支持 OAuth2.0 认证和 API 密钥管理，确保接口安全性。

3.2.2. 应用服务层

模块名称	核心能力	技术逻辑
项目管理	• 项目创建与配置 • 代码仓库集成 • 检测策略管理 • 团队协作支持	采用领域驱动设计(DDD)，通过项目聚合根管理项目生命周期。支持 Git/SVN 等版本控制系统自动同步。
任务调度	• 并发任务管理 • 优先级调度 • 资源负载均衡 • 失败重试机制	基于消息队列实现异步任务处理，支持任务优先级设置和资源隔离。采用分布式锁确保任务一致性。
用户管理	• 多租户支持 • 角色权限控制 • 操作审计跟踪	基于 RBAC 权限模型，支持细粒度权限控制。集成 LDAP/AD 实现统一身份认

	• SSO 集成	证。
规则管理	• 检测规则配置 • 规则模板管理 • 自定义规则支持 • 规则版本控制	采用规则引擎架构，支持 DSL 规则定义。提供图形化规则编辑器，支持规则测试和验证。

应用服务层是系统的业务逻辑核心，负责协调各个模块完成具体的业务功能。

3.2.3. 检测引擎层

检测引擎层是系统的核心技术层，负责源代码的深度分析和安全缺陷检测。

模块名称	核心能力	技术逻辑
语法分析器	• 多语言词法分析 • 语法树构建 • 语法错误检测 • 代码结构解析	基于 ANTLR 等解析器生成器，支持 28 种编程语言的语法解析。构建统一的抽象语法树(AST)表示。
语义分析器	• 符号表管理 • 类型系统分析 • 作用域解析 • 语义错误检测	在 AST 基础上进行语义分析，建立完整的符号引用关系。支持跨文件、跨模块的语义关联分析。
数据流分析	• 变量值跟踪 • 污点传播分析 • 数据依赖分析 • 敏感信息追踪	采用静态单赋值(SSA)形式，构建数据流图(DFG)。通过污点分析技术识别安全漏洞。
控制流分析	• 程序流程图构建 • 路径可达性分析 • 循环结构识别 • 异常处理分析	构建控制流图(CFG)，分析程序执行路径。支持路径敏感的分析，提高检测精度。

3.2.4. 数据存储层

数据存储层负责系统的数据持久化，确保数据的安全性和可靠性。

模块名称	核心能力	技术逻辑

模块名称	核心能力	技术逻辑
项目数据库	• 项目元数据存储 • 检测配置管理 • 版本信息维护 • 团队协作数据	采用关系型数据库(MySQL/PostgreSQL)，支持事务处理和复杂查询。使用 ORM 框架进行数据访问。
规则知识库	• 检测规则存储 • 漏洞模式库 • 修复建议库 • 合规标准库	使用图数据库存储规则依赖关系，支持规则的快速检索和匹配。提供规则版本管理和更新机制。
检测结果库	• 缺陷详情存储 • 分析结果缓存 • 历史数据归档 • 统计信息计算	采用时序数据库存储检测结果，支持大规模数据的快速写入和查询。实现数据分区和索引优化。
审计日志库	• 操作日志记录 • 系统运行监控 • 安全事件追踪 • 性能指标收集	使用 ELK 技术栈进行日志管理，支持实时日志分析和告警。提供完整的审计追踪能力。

3.2.5. 集成接口层

集成接口层提供与外部系统的对接能力，支持 DevOps 流程的自动化集成。

模块名称	核心能力	技术逻辑
代码仓库接口	• Git/SVN 集成 • 自动代码拉取 • 分支标签管理 • Webhook 支持	实现版本控制系统的通用接口，支持多种认证方式。提供增量代码识别和自动触发检测。
CI/CD 接口	• Jenkins 集成 • 流水线插件 • 质量门禁 • 构建阻断	基于 Jenkins Pipeline 插件机制，提供质量门禁功能。支持检测结果与构建状态的联动。
缺陷管理接口	• Jira/禅道集成 • 缺陷自动创建 • 状态同步更新 • 工作流定制	实现通用的缺陷跟踪系统接口，支持缺陷模板定制和字段映射。提供双向同步能力。

模块名称	核心能力	技术逻辑
第三方 API	• 开放 API 服务 • Webhook 回调 • 数据导出接口 • 扩展插件机制	基于 RESTful 设计原则，提供完整的 API 文档和 SDK。支持插件化扩展，便于功能定制。

3.2.6. 基础设施层

基础设施层提供系统运行的基础环境支持，确保系统的高可用性和可扩展性。

模块名称	核心能力	技术逻辑
Docker 容器	• 容器化部署 • 资源隔离 • 快速扩缩容 • 环境一致性	基于 Docker 和 Kubernetes 实现容器编排，支持微服务架构。提供健康检查和自动恢复机制。
国产化平台	• 国产 CPU 支持 • 国产 OS 适配 • 国产数据库兼容 • 信创环境认证	支持鲲鹏、海光、飞腾等国产 CPU，兼容统信 UOS、麒麟 OS 等国产操作系统。
分布式集群	• 负载均衡 • 故障转移 • 数据同步 • 弹性扩展	采用主从复制和分片技术，支持水平扩展。实现智能负载均衡和故障自动切换。
监控告警	• 性能监控 • 资源使用率 • 异常检测 • 自动告警	集成 Prometheus+Grafana 监控体系，提供全方位的系统监控和可视化仪表板。

3.3. 技术架构核心特点

3.3.1. 模块化设计

各层模块职责清晰，支持独立升级和扩展，降低系统维护复杂度。每个模块都提供标准化的接口，便于系统集成和功能扩展。

3.3.2. 高性能分析

采用增量分析、并行计算等技术，支持百万行代码级别的快速分析。检测引擎支持分布式部署，可以水平扩展以处理大规模代码库。

3.3.3. 高可用性

通过分布式部署、负载均衡、故障转移等机制，确保系统 7×24 小时稳定运行。采用多副本数据存储，防止单点故障。

3.3.4. 可扩展性

支持水平扩展和垂直扩展，可根据业务需求灵活调整系统规模。模块化的架构设计使得新功能可以快速集成到现有系统中。

3.3.5. 安全性

多层次安全防护，包括身份认证、权限控制、数据加密、安全审计等。所有敏感操作都有完整的审计日志，支持安全事件追溯。

3.3.6. 国产化兼容

全面支持国产化软硬件环境，满足信创要求。已在多个国产化平台上完成测试和认证，确保在关键信息基础设施中的可靠运行。

3.4. 技术架构优势

全面的检测能力：支持 28 种编程语言，覆盖国内外主流安全编码标准；

精准的缺陷定位：通过数据流和控制流分析，提供精确的缺陷定位和修复建议；

灵活的部署方式：支持物理机、虚拟机、容器等多种部署方式，适应不同环境需求；

强大的集成能力：提供丰富的 API 接口和插件，支持与现有开发工具链无缝集成；

卓越的性能表现：采用先进的静态分析技术，确保在大规模代码库上的高效分析；

本技术架构为 CodePecker 软件供应链安全静态分析系统提供了坚实的技术基础，确保了系统在功能性、性能、可靠性和可扩展性等方面的卓越表现。

第四章 能力介绍

4.1. 多语言支持

系统支持 28 种主流编程语言的代码安全检测，包括但不限于：

C、C++、C#、Java、Python、Go、HTML、JSP、PHP、JavaScript、XML、SQL、COBOL、Objective-C、RUBY、Swift、Shell、KOTLIN、Fortran、DART、Lua、Erlang、VUE、CSS、TYPESCRIPT、IAC、ARKTS、OCAML、SCALA 等。

4.2. 合规性检测

系统支持多项国内外安全编码标准和规范，包括：

CERT 安全编码标准、GB/T 38674-2020、GB/T 34944-2017、GJB 8114-2013、GJB 5369-2005、GB/T 34943-2017、GB/T 34946-2017 等国家标准，以及 OWASP Top10 2021/2017、CWE/SANS TOP25 等行业标准。

4.3. 漏洞检测能力

系统能够检测各类常见安全漏洞，包括：

代码注入、跨站脚本(XSS)、输入验证、危险函数、代码质量、API 误用、密码管理、异常处理等常见安全缺陷问题，全面覆盖 CWE/SANS TOP 25 和 OWASP TOP 10 等权威漏洞列表。

4.4. 敏感信息检测

系统具备敏感信息识别能力，可检测代码中的：

商业凭证、身份信息、加密密钥、非对称私钥、API 密钥、银行卡号和访问 Token 等敏感数据，并支持用户自定义敏感信息检测规则。

4.5. 数据合规检测

针对数据安全法规要求，系统提供：

- 自动识别和标记个人身份信息、账号、卡号等敏感数据处理点
- 分析展示敏感数据流向，包括日志输出、数据库写入、发送到第三方
- 提供从数据合规缺陷源到问题爆发点的图形化跟踪流

4.6. 部署与集成

系统支持多种部署方式和集成能力：

支持分布式部署、Docker 容器化部署，兼容 Redhat、CentOS、统信、银河麒麟、欧拉等操作系统环境，支持海光、鲲鹏、飞腾等国产 CPU 以及 TiDB 等国产数据库。

支持 GitLab、GitHub、Gitee、SVN 等代码仓库集成，可与 Jira、禅道等缺陷管理系统对接，并支持 Jenkins 等 CI/CD 平台集成。

第五章 功能清单

系统提供全面的功能模块，涵盖代码审计的全流程需求：

功能	能力技术	模块	详细描述
检测分析功能	多语言检测能力	28 种编程语言检测	支持 C、C++、C#、Java、Python、Go、HTML、JSP、PHP、JavaScript、XML、SQL、COBOL、Objective-C、RUBY、Swift、Shell、KOTLIN、Fortran、DART、Lua、Erlang、VUE、CSS、TYPESCRIPT、IAC、ARKTS、OCAML、SCALA 等 28 种编程语言检测能力

功能	能力技术	模块	详细描述
		框架支持	支持Spring、Spring Boot、Spring JDBC、Spring MVC、MyBatis、Ibatis、Spring Data JPA、Spring Web Flow、Spring Security、Struts1、Struts2、Hibernate、Fastjson、Log4j、ApacheAxis1、Qt4、Qt5、Vue、Nodejs、Jquery Laravel、Django、Flask 等主流开发框架的代码检测
	检测技术能力	语法及语义分析	采用语法及语义分析、数据流、控制流、污点传播分析等静态分析技术
		无编译检测	扫描不需要依赖具体的编译器和开发环境，无需用户预编译，用户可直接提交源代码进行检测
		字节码分析	支持基于 JAVA 字节码的代码缺陷分析，可分析 jar、war 包
	安全标准合规	CERT 安全编码标准	支持 CERT 安全编码标准的合规性检测
		CWE/SANS TOP 25	覆盖 CWE/SANS TOP 25 2023、CWE/SANS TOP 25 2019、漏洞缺陷列表
		OWASP TOP 10	支持 OWASP TOP 10 2021、OWASP TOP 10 2017 等版本检测
		GB/T 系列标准	支持 GB/T 38674-2020、GB/T 34944-2017、GB/T 34943-2017、GB/T 34946-2017、等国家标准
		军工标准	支持 GJB 8114-2013、GJB 5369-2005 等军工标准
漏洞检测功能	安全漏洞及代码质量检	注入类漏洞	支持代码注入、SQL 注入、命令注入等注入类漏洞检测

功能	能力技术	模块	详细描述
	测	跨站脚本	支持跨站脚本(XSS)漏洞检测
		输入验证	支持输入验证不充分导致的安全问题检测
		危险函数	检测代码中使用的危险函数
		API 误用	检测 API 的错误使用方式
		密码管理	检测密码管理相关的安全问题
		异常处理	检测异常处理不当导致的安全隐患
	敏感信息检测	商业凭证检测	可检测代码文件、配置文件等中的商业凭证信息
		身份信息检测	检测个人身份信息泄露风险
		加密密钥检测	检测硬编码的加密密钥、非对称私钥
		API 密钥检测	检测 API 密钥泄露风险
		银行卡号检测	检测银行卡号等金融敏感信息
		访问 Token 检测	检测访问 Token 泄露风险
	数据合规检测	敏感数据处理点识别	自动识别和标记敏感数据处理点，包括个人身份信息、账号、卡号等
		数据流向分析	支持分析展示敏感数据流向，包括日志输出、数据库写入、发送到第三方
		数据流分析	支持数据流分析以确保全面检测潜在的数据安全问题
		图形化跟踪流	提供从数据合规缺陷源到问题爆发点的图形化跟踪流
检测管理功能	检测对象管理	多种格式支持	支持多种格式的检测对象上传，包括源代码文件、文件夹、压缩包等，支持的压缩包格式包

功能	能力技术	模块	详细描述
			括.zip、.tar、.gz、.7z 等格式
		版本管理集成	具备检测版本管理功能，且具备与不同版本管理工具的集成其中包括 GitLab、SVN、CVS、FTP、共享目录等
		代码库直接检测	支持从代码仓库获取代码进行检测，支持 GitLab、SVN、FTP、CVS、共享目录等
	检测任务管理	多任务并发	支持多任务同时检测，可以一键发起多任务的同时检测
		数据检索	支持检索功能，包括通过任务名称、缺陷名称关键字、检测时间等条件进行任务检索
		任务启停控制	支持停止扫描任务，停止后可重启任务，可通过任务的启停调整任务优先级
		任务暂停恢复	支持暂停扫描任务，暂停后可在断点继续扫描，不需要重新发起扫描任务
		异常处理	检测任务资源占用过多，出现拥堵时，或检测任务陷入死循环等异常情况时，将自动停止任务并输出异常报告，并对异常点进行定位
		增量扫描	支持全量分析及增量分析，可针对改动代码进行增量分析
		定时任务	支持配置单次检测和持续检测，持续检测支持配置每天、每周和每月定时发起检测
		分支标签检测	创建 Git 项目支持拉取分支、标签和 commit id 进行扫描的功能，支持指定路径下的文件或文

功能	能力技术	模块	详细描述
			件夹进行扫描
		UI 前端展示定制	可对任务、项目等前端界面的风格、标识等内容进行切换和修改，可实现列表与模块之间的切换，并可对标识内容增加或减少。
	检测规则管理	自定义检测规则	支持添加自定义检测规则，黑、白名单等规则策略
		规则模板管理	提供代码检测规则自定义功能，管理人员能通过此功能新增、修改或去除代码检测规则
		规则集合管理	支持自定义缺陷规则集合，只分析指定的代码缺陷类型
缺陷管理功能	缺陷信息展示	完整缺陷信息	提供缺陷问题的相关信息，包含缺陷名称、问题描述、问题等级、问题定位、问题示例、修复建议、正确代码示例等
		缺陷定位	缺陷检测结果支持定位到代码路径、代码文件、代码行号
		污点轨迹图	支持查看缺陷详情及污点轨迹图，可定位到具体的代码行，便于污点溯源和分析
	缺陷审计功能	结果审计	检测结果支持人工进行审计，可以对有问题的检测结果进行更正，可把缺陷更改为误报、忽略等状态
		审计携带	支持审计携带功能，可以将同一个项目的审计信息自动携带到下一次扫描结果中，无需人工重复审计
		批量审计	根据审计需求，支持逐个审计，支持批量审计

功能	能力技术	模块	详细描述
		大模型审计	基于消费级显卡算力，提供 AI 审计能力，能够根据源代码缺陷给出针对性的修复建议
		智能学习	能够基于历史代码审计信息学习识别有效缺陷，支持自动化代码审计，提高审计效率
	缺陷统计分析	Top10 统计	支持对缺陷类型 Top10 进行图形化统计展示
		多维度统计	支持以项目、源码语言、缺陷类型维度对进行检测结果数据的统计
		结果对比	支持对项目多次检测结果的对比分析
		趋势分析	支持统计项目数、缺陷总数、缺陷密度、代码重复率和圈复杂度等信息
报告输出功能	报告内容	基本统计信息	报表内容包含缺陷名称任务耗时、任务结果、问题描述、问题定位、问题等级、问题类型、问题示例、修复建议等
		任务信息	包含任务发起人、任务开始时间、任务结束时间、任务配置等信息
		项目信息	包含文件数量、代码数量、开发语言、缺陷类型、缺陷风险级别、缺陷数量等信息
	报告输出格式	多种格式导出	支持报表导出和报表模板自定义，导出的格式类型包括 html、pdf、word、excel、xml 等
		自定义内容	支持用户按风险等级等条件自定义导出报告内容，支持对报告中的缺陷概览、缺陷级别、缺陷状态、修复建议、代码片段、对比

功能	能力技术	模块	详细描述
			结果和审计信息等内容进行选择并导出
	在线查看	报表在线查看	支持报表在线查看，可在线查看缺陷统计信息及缺陷清单及缺陷详细定位，如果是多次检测，可实现对缺陷的趋势分析报表。
		实时查看	支持检测结果实时查看和分析
系统管理功能	用户权限管理	多用户支持	支持多用户提交代码检测任务，多用户查看检测结果，不限制用户数量
		权限控制	提供用户权限管理功能，按照角色分配用户系统的权限；支持多个角色之间数据隔离，不同角色间数据无法越权查看
		部门项目管理	支持多部门、多项目的团队级代码审计管理功能
	系统配置	白名单管理	支持 IP 白名单配置功能
		历史数据清理	提供历史数据的管理功能，支持系统垃圾数据清理功能
	日志管理	登录行为日志	记录用户登录行为日志，支持登录行为审计
		操作使用日志	记录用户操作使用日志，支持操作使用审计
		系统监控	提供系统监控功能，支持实时监控服务器和引擎内存、CPU、磁盘和数据库的使用情况
集成接口功能	API 接口	任务管理接口	提供任务 API 接口，可通过任务 API 接口实现任务下发（包括上传被测软件包、对接代码库获取源代码的方式）、任务暂停、任务重启、任务删除、报表查看、报

功能	能力技术	模块	详细描述
			表下载、检测状态查询等功能
		数据导出接口	数据导出 API 接口主要为了方便集成第三方平台集中查看检测结果信息，可提供标准接口规范，通过 API 接口实现检测结果导出等，具备可自定义的导出数据格式、字段。
		第三方集成	提供开放 API 接口，其中包括项目管理接口、任务管理接口、缺陷详细信息接口、日志数据接口等，提供接口规范文档，方便与第三方系统进行集成
	开发工具集成	主流 IDE 支持	支持 Eclipse、IntelliJ IDEA、VSCode 等插件发起检测
		实时检测	通过提供 IDE 插件，协助开发人员开展扫描检测工作，完成扫描工作后展示检测结果并辅助开发人员定位和分析问题，并提供修复建议
	DevOps 集成	Jenkins 集成	支持 Jenkins Pipeline 等 DevOps 平台发起检测，支持与 Jenkins 等平台进行对接，集成至 CI 流水线中
		构建工具集成	支持 Maven 构建管理工具集成
		缺陷系统集成	支持检测结果与 Bugzilla、Jira、禅道等 Bug 管理系统进行集成，支持与 Jira、禅道项目缺陷跟踪系统对接
部署运维功能	部署方式	单独部署	支持独立部署使用
		虚拟化部署	支持虚拟化部署使用
		Docker 支持	支持 docker 容器化部署

功能	能力技术	模块	详细描述
		集群部署	支持分布式部署，提高系统性能及检测能力
	国产化支持	国产 OS 支持	支持 Redhat、CentOS、统信、银河麒麟、欧拉（openEuler）等环境部署
		国产 CPU 支持	支持海光、鲲鹏、飞腾等国产 CPU 部署
		国产数据库支持	支持 TiDB 等国产数据库部署
	性能与扩展	大规模代码支持	支持分析百万行级别的源代码（硬件环境：CPU 32 核及以上，CPU 主频 2.0GHz 及以上，内存 128GB 及以上）
		并发性能	支持多用户并发操作，不限制检测次数，不限制检测项目数量
		弹性扩缩容	可根据硬件配置及检测引擎数量对检测并发数进行弹性扩缩容
其他特色功能	代码质量分析	代码统计	能够从源码文件大小和代码行数、安全缺陷数、圈复杂度、代码相似度等方面量化和定位源代码相关问题
		质量指标	支持按文件数、代码行数、注释行数、空白行数和各语言代码行数对源码信息进行统计
	组件安全分析	组件识别	支持在代码审计过程中识别出代码中引用的第三方开源组件及其漏洞（CNNVD 和 CVE）
		风险评估	支持对开源组件进行风险评估，并提供修复建议
		组件统计	支持第三方组件统计分析功能，支持按组件名称进行检索，支持查看组件影响明细和导出组件报

功能	能力技术	模块	详细描述
			告功能
	唯一标识与验证	Hash 值生成	对被检测源代码进行唯一 hash 值标识，可在界面及检测报告中查看源代码文件 hash 值
		代码验证	通过 hash 值确保代码的完整性和一致性

第六章 我们的优势

6.1. 技术领先性

采用先进的静态分析技术，结合语法语义分析、数据流分析、控制流分析和污点传播分析等多种技术手段，确保检测的准确性和全面性。

6.2. 全面的检测能力

多语言支持：支持 28 种主流编程语言，满足企业多样化技术栈需求。

标准合规：全面覆盖国内外主流安全编码标准和行业规范。

6.3. 高效的部署与集成

灵活部署：支持物理机、虚拟机、容器等多种部署方式，兼容国产化环境。

无缝集成：提供丰富的 API 接口和插件，轻松融入现有开发流程。

6.4. 准确的检测结果

精确定位：提供详细的缺陷信息，包括缺陷名称、问题描述、问题等级、问题定位、问题示例、修复建议等。

低误报率：通过智能分析和人工审计相结合，有效降低误报率。

第七章 附录：系统性能指标

系统具备卓越的性能表现，能够满足大规模企业级应用需求：

- 支持分析百万行级别的源代码（硬件环境：CPU 32 核及以上，CPU 主频 2.0GHz 及以上，内存 128GB 及以上）
- 支持多用户并发操作，不限制检测次数、项目数量和用户数量
- 支持根据硬件配置和检测引擎数量对检测并发数进行弹性扩缩容