2024 第三届 CCF 区块链技术与创新应用竞赛测试报告

CryBlock：区块链密码应用安全分析平台

目 录

1. 测试概述

本测试报告针对 CryBlock 区块链密码应用安全分析平台开展测试，主要针对 CryBlock 平台在分析区块链密码应用中各类安全漏洞的能力开展研究，研究 CryBlock 在检测真实世界区块链密码应用安全问题过程中的精确率（Precision）与召回率（Recall），以检验 CryBlock 是否能高精确率、高召回率地检出真实世界密码应用中的各类安全漏洞，并为链上密码应用创新实践提供有效的安全保障。

2. 功能测试

2.1 密码应用数据集

为充分测试 CryBlock 在检测密码应用安全问题方面的有效性，本功能测试过程收集了一个包含 25,745 个真实世界与密码应用有关的智能合约的大规模数据集。为收集真实世界密码应用数据集，本测试过程首先重放了从以太坊主链从区块 1 到区块 15,500,000 的 1,704,224,022 个历史以太坊交易（从 2015 年 7 月到 2022 年 9 月），分析各个交易是否涉及密码学操作，并记录进行这些密码学操作的智能合约。在历史交易重放期间，共识别出 426,296 个与密码应用相关的合约。之后，本测试过程查询 Etherscan 以收集这些合约的公开可用的源代码和 ABI 信息。结果显示，共有 25,745 个具有可用源代码和 ABI 信息的密码应用相关智能合约。在这些智能合约中，83.6%的智能合约拥有超过 10 个历史交易，反映出该测试数据集中合约对于真实世界密码操作的代表性。

基于收集到的密码应用数据集，测试过程通过 CryBlock 对数据集开展大规模分析，以测试 CryBlock 在检测安全问题方面的精确率和召回率。所有实验均在一台配备两个 Intel Xeon(R) Platinum 8352V CPU、512GB RAM 并运行 Ubuntu 22.04.2 LTS 的机器上进行。测试过程共花费 408.0 小时分析了 25,745 个合约，平均每个合约的执行时间为 57.1 秒。在测试过程中，CryBlock 检测

到 5,847 个（22.7%）真实世界密码应用至少存在一个安全漏洞。

2.2 精确率测试

为评估 CryBlock 在检测每种漏洞类型时的精确度，测试过程手动分析了工具在大规模实验中所检出的漏洞。具体地，为便于手动分析，测试过程为每种漏洞类型随机采样了一定数量的漏洞。每种漏洞类型的样本大小经过特定选择，以达到 95%的置信水平和 10 的置信区间。之后，测试过程将标记检出的漏洞合约划分为真阳性（TP）或假阳性（FP）。下表中展示了每种漏洞类型的真阳性、假阳性数量及计算出的精确度。最终，计算工具的总体精确度为这些精确度的加权平均数，权重为每种漏洞的数量。结果显示，工具的总体精确度为 95.4%，表明 CryBlock 可以精确地检出真实世界中密码应用智能合约中存在的安全问题。

表 1：密码应用漏洞的检测精确度

漏洞类型	检测数	抽样数	真阳性	假阳性	精确度
SSR	151	59	59	0	100.0%
CSR	2,536	93	89	4	95.7%
SF	274	71	69	2	97.2%
SM	1,803	91	89	2	97.8%
ISV	24	20	17	3	85.0%
MR	122	54	48	6	88.9%
MF	33	25	23	2	92.0%
HC	89	46	43	3	93.5%
WR	2,626	93	87	6	93.5%

2.2 召回率测试

为评估 CryBlock 召回率，本测试过程对于数据集进行采样，并构建了一个带有真阳性（True Positive，TP）和假阴性（False Negative， FP）标注的密码应用数据集。具体地，其首先从大规模密码应用数据集中随机抽取了一些智能合约并手动注释它们。具体来说，测试过程从 25,745 个合约中随机抽取了 96 个合约，以达到 10 的置信区间和 95% 的置信水平。之后，按照与精确度评估相同的标记过程手动分析这些抽样合约。总计，我们在 96 个真实密码应用中发现了 34 个漏洞。比较这些手动标记和工具给出的结果后，我们发现工具对于这些合约报告了 31 个真阳性，1 个假阳性，和 3 个假阴性，CryBlock 的召回率为 91.2%。

2.3 真实世界影响评估

在展示工具有效性的同时，我们的大规模实验也首次对现实世界智能合约密码应用中的安全漏洞进行了深入观察，展示了安全漏洞对于真实世界密码应用的普遍影响。

表 2 的第一列展示了每种安全漏洞的密码应用比例。在 9 个漏洞类型中，WR、CSR 和 SM 是最常见的安全漏洞，分别出现在 2,626（10.20%）、2,536（9.85%）和 1,803（7.00%）个已分析的密码应用中。尽管其余六种漏洞类型较少见（出现在约 1%或更少的应用中），但受其影响的密码应用总数仍然相当可观。具有这些漏洞的密码应用即表明相关密码应用的实现偏离了密码学实现领域所推荐的最佳实践。尽管漏洞可能不会直接导致安全问题，但它们会削弱合约的可维护性，并增加未来安全漏洞的风险。例如，当只有一个密码应用验证特定授权者的签名时，CSR 漏洞可能不会立即引发安全问题。然而，如果系统发展到多个应用开始使用同一授权者的签名来管理敏感操作，此漏洞可能直接导致跨合约签名重放攻击。

表 2：密码应用漏洞的统计指标

类型	比例(%)	代码行数	函数数	以太余额	交易数
SSR	0.59%	1369.5	33.8	6.0	8,163
CSR	9.85%	1487.1	30.3	9.3	37,958
SF	1.06%	1466.2	28.5	12.8	76,188
SM	7.00%	1238.4	27.6	5.4	64,906
ISV	0.09%	783.9	26.6	0.1	30,669
MR	0.47%	1747.5	41.9	4.2	1,299
MF	0.13%	1585.1	38.9	5.1	1,544
HC	0.35%	1551.5	29.7	0.8	6,868
WR	10.20%	1352.9	30.7	6.3	3,469
总计	**22.71%**	**1396.8**	**30.3**	**7.1**	**22,930**

同时，相关结果表明，在超过 5,000 个真实世界密码应用存在至少一个安全漏洞，表现出真实世界密码创新应用普遍存在的安全问题。同时，这些具有安全漏洞的密码应用已经吸引了大量真实世界用户与资金，展示出密码应用安全漏洞广泛的影响范围与潜在危害，突出了利用 CryBlock 等分析平台对于链上密码创新应用开展安全分析的重要性。

3. 测试总结

综上，本测试报告针对 CryBlock 开展系统功能测试，相关结果表明，CryBlock

可以高精确率、高召回率地检测出智能合约密码应用中的安全漏洞。CryBlock 实现了完备的密码应用安全分析功能，可以提供高精确率、高召回率、高可用性的在线智能合约密码应用安全检测服务，有效支撑智能合约密码应用的部署前安全检测，保障真实世界蓬勃发展的链上密码创新应用的安全性与可靠性。