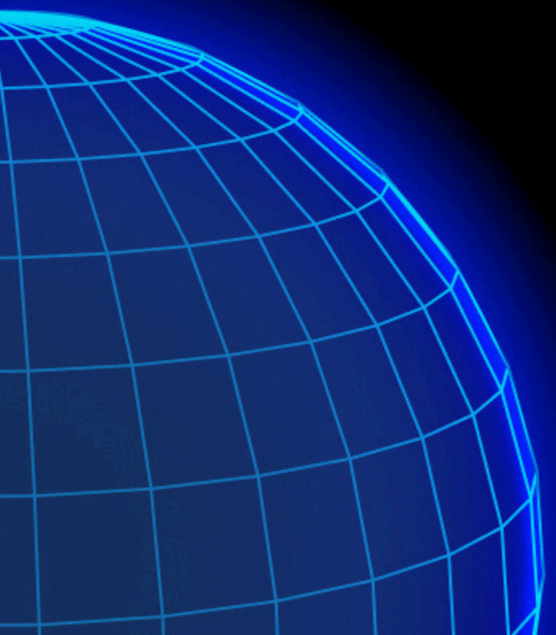
ONESAVIE

AI-Driven Smart Contract Vulnerability Detection

講者: Alice, Daky



ONESAVIE

Who am I

- Product Development Lead @OneSavie Lab
- Member @DeFiHackLabs
- Focus on AI & Web3 Development

Daky

https://x.com/ga013077

https://medium.com/@ga013077

ONESAVIE

Who am I

- Security Research Engineer @OneSavie Lab
- OP and White Hat @DeFiHackLabs
- Audit Team member @TaiChi Audit Group
- Focus on Web3 Security Research

DEFIHACKLABS
Let's make web3 more secure!

Alice

@AliceHsu_kou

@AliceHsu_Hsu

ONESAVIE

Agenda

01 Background

02 About Bastet

- Introduction
- Methodology

03 Conclusion

Background

What is DeFi ?

- Decentralized Finance(DeFi)
- Including lending, derivatives, decentralized exchanges, asset management, stablecoin, insurance, borrowing, etc.

Motivation: Driving Industry Security

ONESAVIE

- DeFi applications are becoming more diverse, with increasing security awareness and service demand. Together, we aim to promote industry development through the power of the community.

- Expecting to optimize the process from test to audit.

As a white hat in competitive auditing platforms ...

- When I see a new codebase, many attack vectors come to mind.
- However, many common bugs still frequently appear in codebases, even up until now.

The goal of code review

ONESAVIE

- When we conduct code reviews, our goal is to focus on enhancing security coverage.
- Identifying all problematic code snippets, and writing detailed reports for these bugs.

They are just like your BFF

ONESAVIE

Type	Checklist
AMM	hardcoded slippage? deadline check? arbitrary calls? ...
Lending	Self-liquidation? Can prevent liquidation? User can repay? ...
EIP Compliant	EIP-4626 Rounding issues ? ERC-5095 return check? EIP-2981 able to pay correct amounts of royalties

The execution of project is time-limited

- Within limited resources, the most important one is time.
- Focus on high-quality findings.

Test -> Audit


Development
Test
Incident Response
Audit
Continuous Monitoring

Common Smart Contract Analysis Tool

ONESAVIE

- Mostly static analysis tools
- Rules are maintained using AST or Regex

Automated Findings / Publicly Known Issues

The 4naly3er report can be found here.

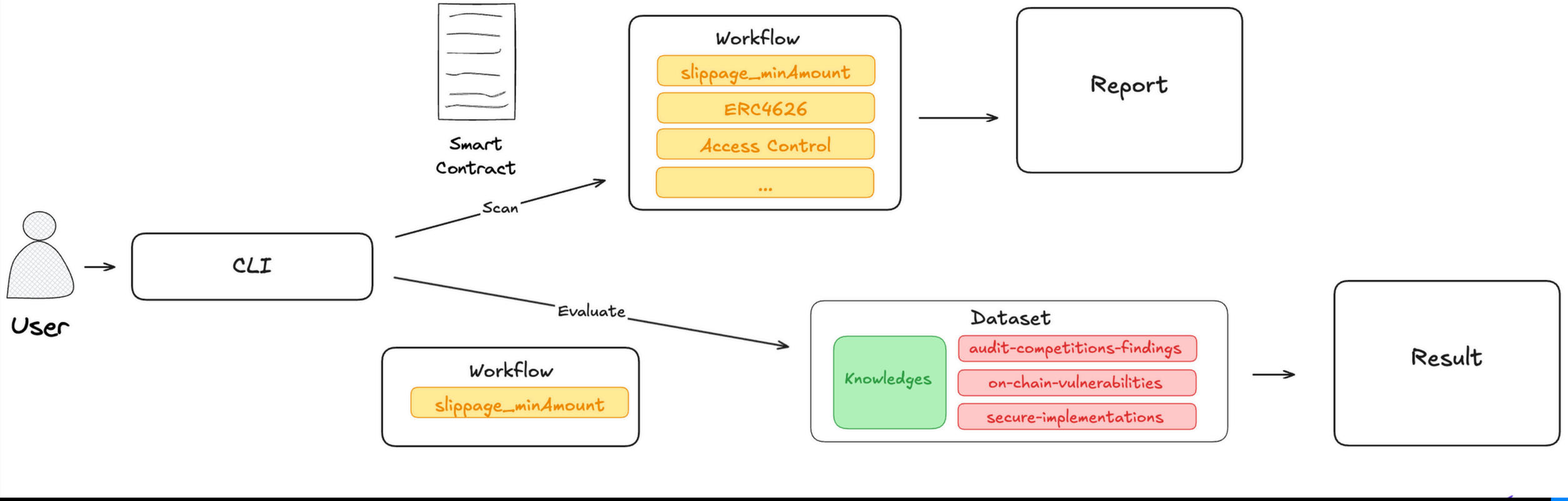
About Bastet

Introduction

ONESAVIE

- This open source project Bastet focus on two parts:
  - Vulnerability dataset + AI Audit benchmark tool
  - AI Vulnerability identification process

User Scenario



How to Select Targets ?

- The vulnerability selection focuses on common issues that are hard to maintain with static analyzers.
- Designing processes to enhance AI's accuracy in detecting vulnerabilities.

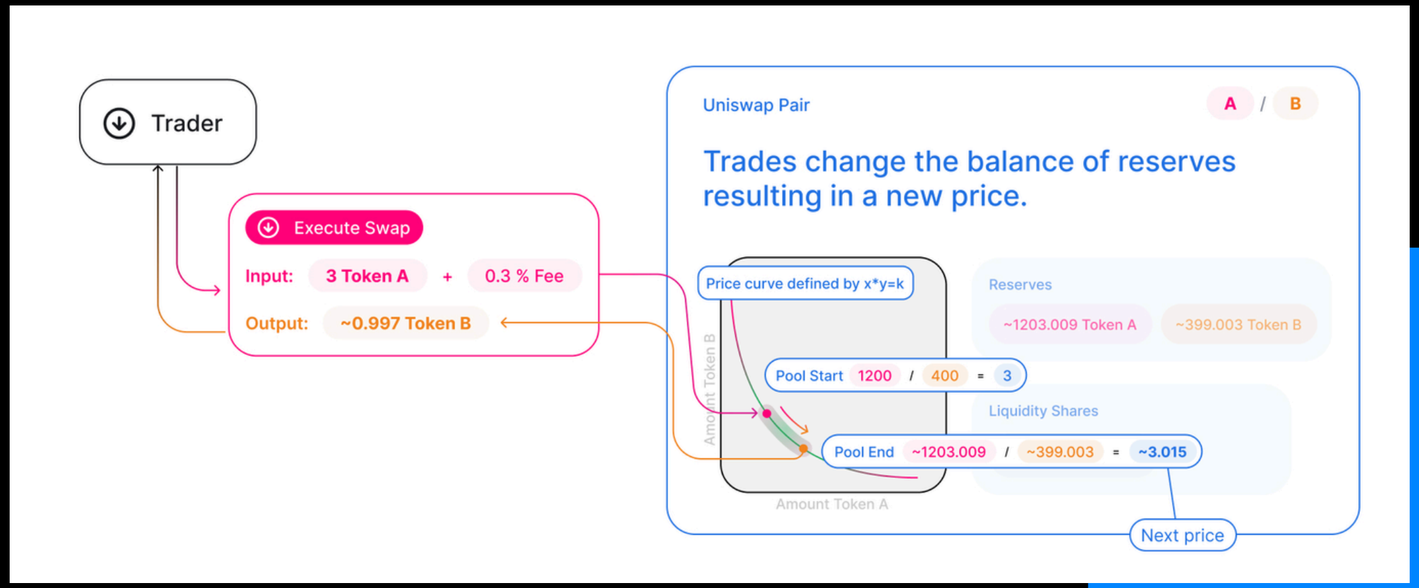
Methodology

ONESAVIE

- What we need to prepare
  - Knowledges
  - Prompt Engineering Technique
  - Validation

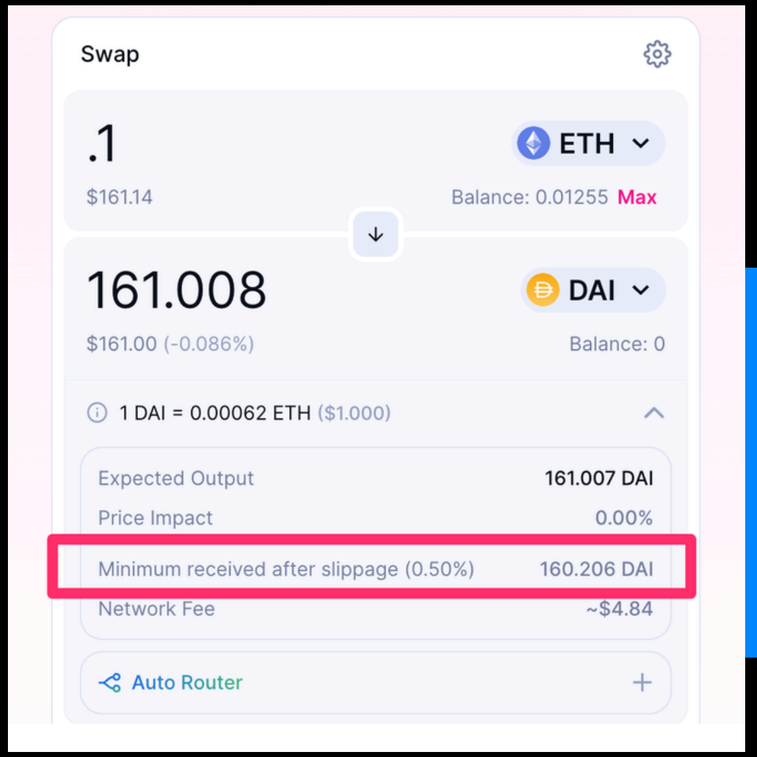
Knowledges - Uniswap

- Uniswap is an automated liquidity protocol powered by a constant product formula (x*y=k).



Knowledges - Slippage

- Slippage refers to the difference between the **expected price** of a transaction and the **actual price** at which the transaction is executed.



Knowledges - Slippage

ONESAVIE

- If there is no slippage check, MEV can monitor transactions and perform a **sandwich attack** before and after the transaction, profiting from the price fluctuations.

How to perform detection using static analyzers?

ONESAVIE

01 Use **findAll** to search for specific functions like **addLiquidity.**

02 Find the location of parameters like **minAmountOut** in the function.

03 Write corresponding conditional checks for each type of improper slippage setting scenario.

What are the challenges of doing this?

- There are variations in the function names for swaps and slippage parameters across different protocols.
- The same issues cannot be detected in different programming languages.

What method are we using?

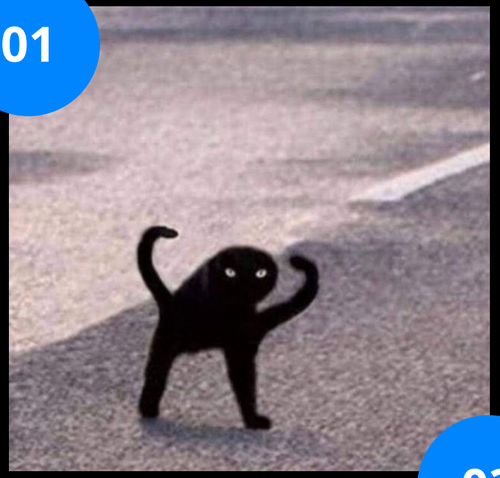
- Example, human brain learns to recognize cats

- Knowledge Base:
  - A cat's appearance typically features **soft fur** and **a sleek, agile body**.
  - They have **four legs**, with **sharp claws** that are perfect for climbing and hunting.
  - Cats' **ears are upright**, and their **noses are small and sensitive**.
  - A cat's **tail is long and flexible**, often used to maintain balance.



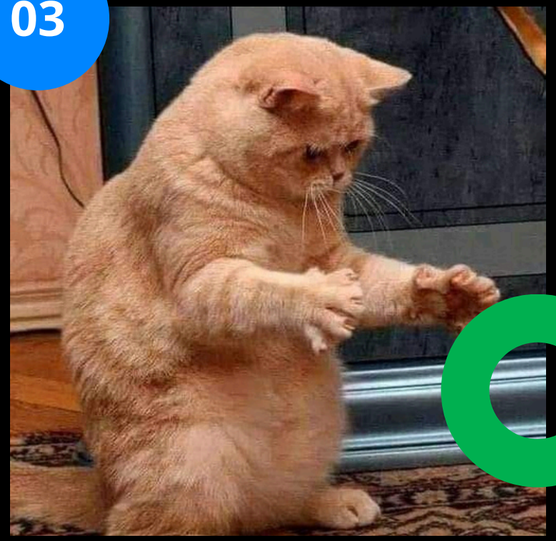

AI Might Think ...

01



02

03



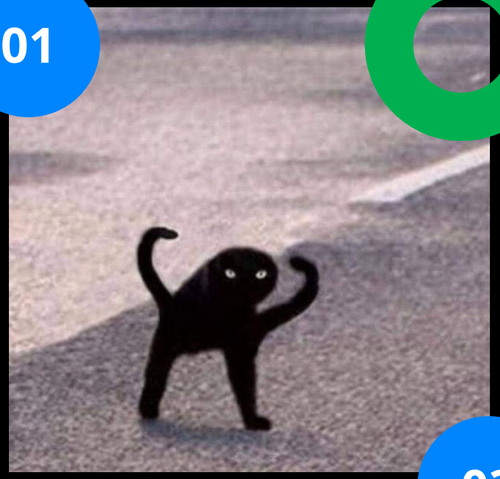
04

05

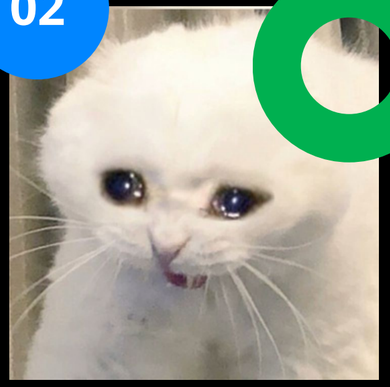
06

The Answer

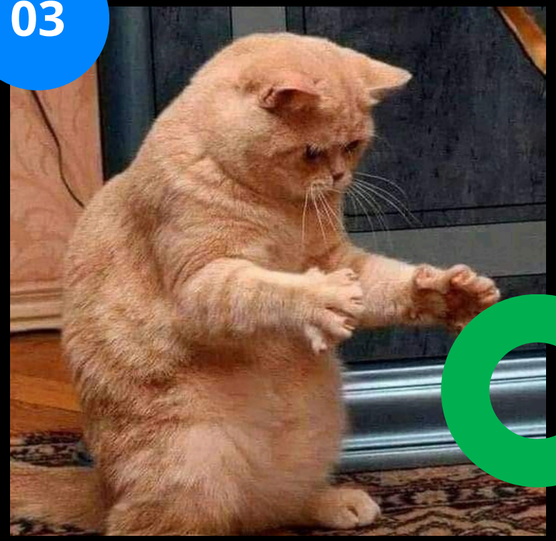
01



02



03



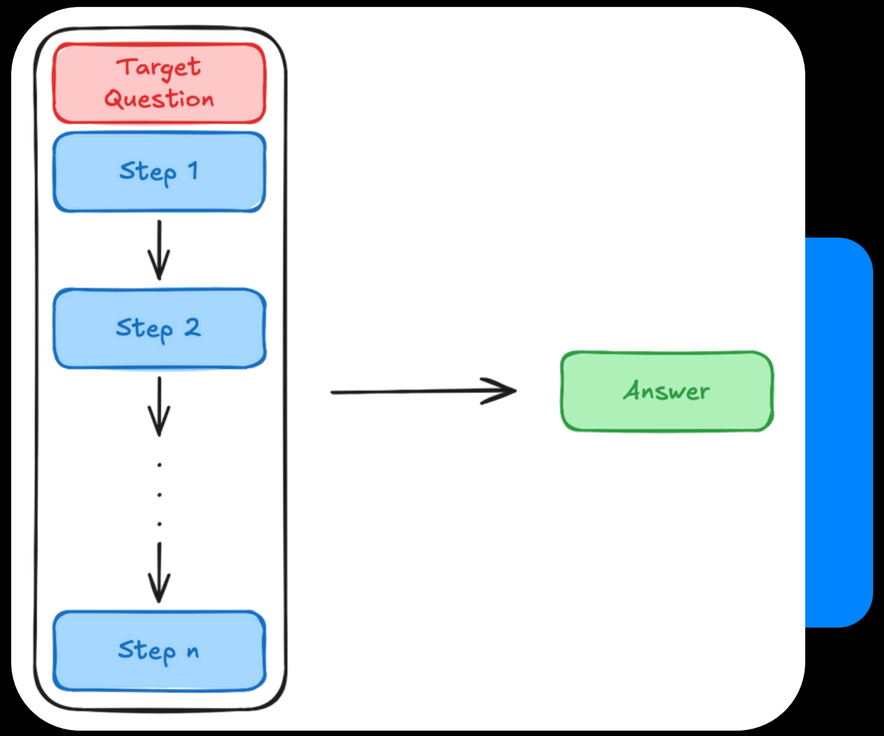
04

05

06

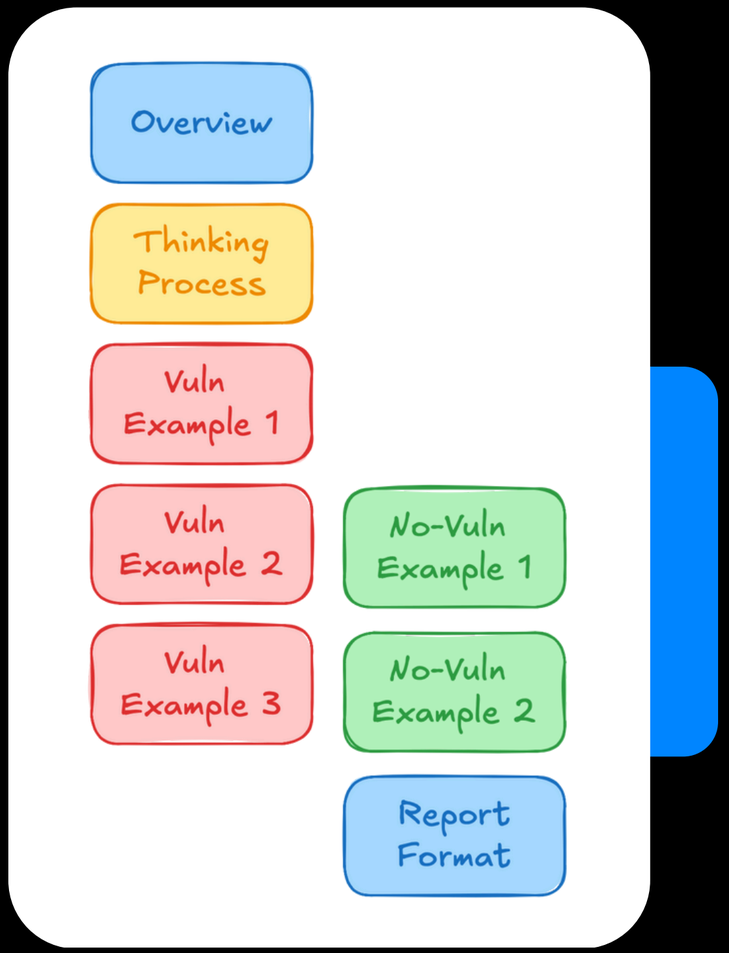
Prompt Engineering Technique – CoT

- By using the CoT method, the reasoning ability of large language models can be improved.
- By breaking down the reasoning process, we help the model think step-by-step, with each step closely tied to the root causes of vulnerabilities.



Prompt Engineering Technique – CoT

- The LLM analyzes each function according to the specified steps.
- The LLM details the thought process in the examples.
- Generate results based on the report format.



Bastet's Output

- Summary
- Severity
- Vulnerability Details
  - Description
  - Code Snippet
- Recommendation

Bastet Security Report

Missing Slippage Protection for Token Swap

Severity: High

Vulnerability Details

- **File Name:** `IbbtcVaultZap.sol`
- **Function Name:** `deposit`
- **Description:**
  The function uses `add_liquidity` from the `CURVE_IBBTC_DEPOSIT_ZAP` contract without specifying a minimum amount of LP tokens to receive.
  This exposes the user to potential slippage, which could result in receiving fewer tokens than expected due to price fluctuations.

Code Snippet

```
uint256 vaultDepositAmount = ICurveZap(CURVE_IBBTC_DEPOSIT_ZAP).add_liquidity(
    CURVE_IBBTC_METAPOOL,
    depositAmounts,
    0,
    address(this)
);
```

Recommendation

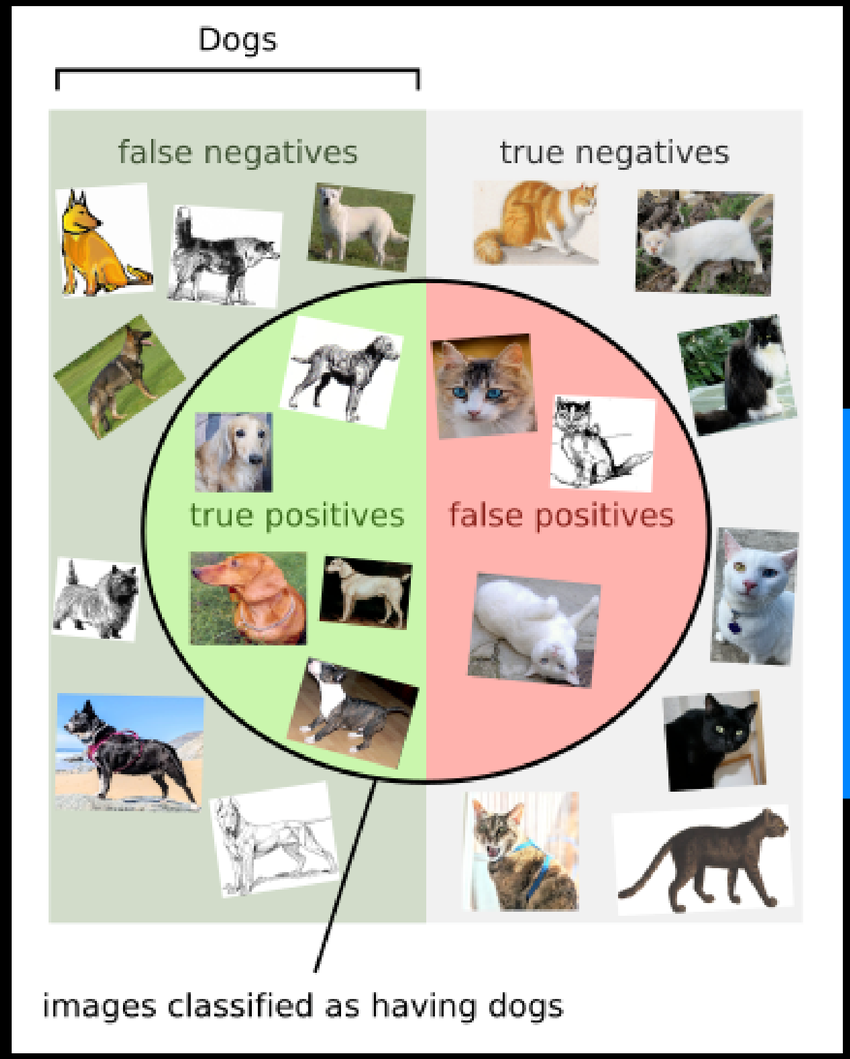
Introduce a parameter for the minimum amount of LP tokens to receive (e.g., _minAmountOut) and use this parameter in the add_liquidity function to ensure slippage protection.

Test Results - example

- Accuracy
  - How to calculate?

$$\text{Accuracy} = \frac{\text{correct classifications}}{\text{total classifications}}$$

$$= \frac{TP + TN}{TP + TN + FP + FN}$$



Bastet Eval Results

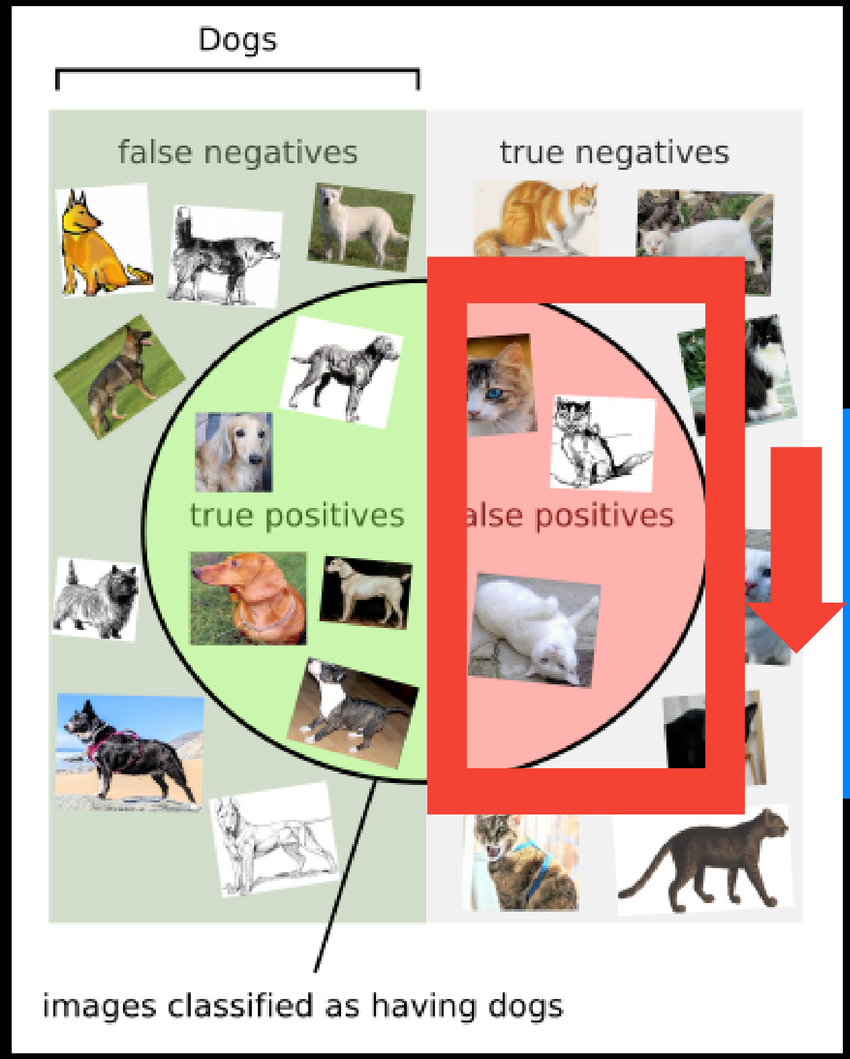
- The evaluation result of minAmountOut workflow is
  - True Positive: 12
  - True Negative: 27
  - False Positive: 2
  - False Negative: 5

- Accuracy: (12+27)/46 = 84.7%

Metric	Value
True Positive	12
True Negative	27
False Positive	2
False Negative	5

Bastet Test Results

- High Accuracy is good
- Less False Positive (FP) is good
- Save Developer and Researcher's Time



Can Bastet One Day Replace Human Researchers?

ONESAVIE

- The goal of Bastet is to assist researchers to focus on higher level issues such as systemic risks, logic errors, integration issues, and more.
- As a tool, it will undoubtedly become stronger over time.

What Does Bastet Need To Become Stronger?

- Data
- Knowledge Base
- Methodology

What if the vulnerabilities that are not in the knowledge base?

ONESAVIE

- Bastet is highly skilled at finding details and implementation errors in code.
- We are also working closely with the community.
- Bastet has already identified **valid** findings in every audit, including **high** and **medium** severity issues.

Conclusion

Takeaway

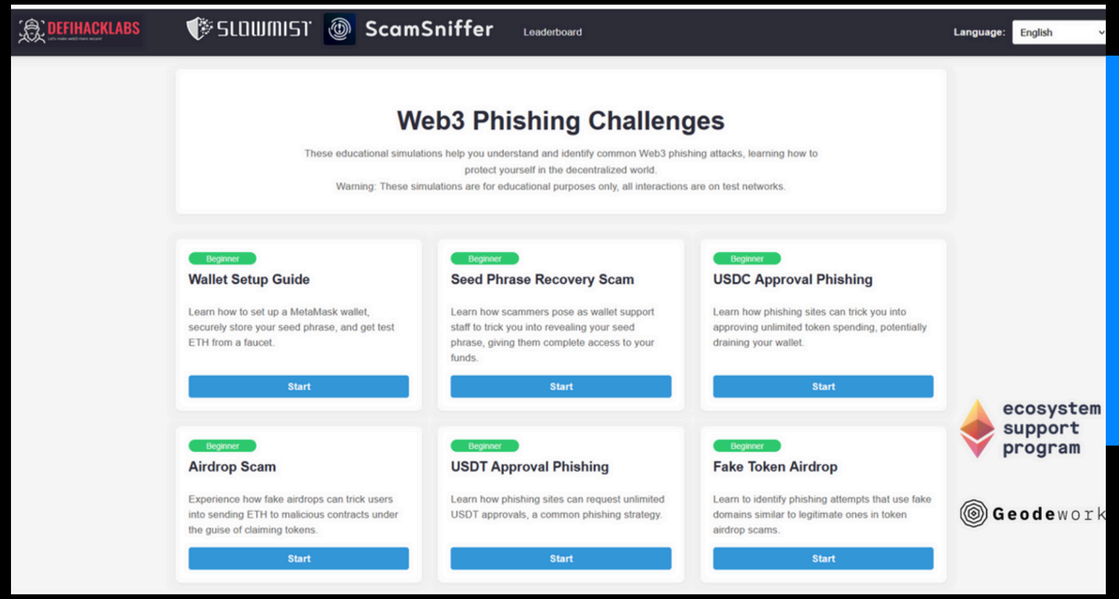
ONESAVIE

- Bastet features:
  - Dataset + Evaluation
  - AI workflows

DeFiHackLabs Time

ONESAVIE

- PhishQuest: Interactive Web3 Anti-Phishing Training Platform
- Founding Contributors: DeFiHackLabs、SlowMist、ScamSniffer
- Supported by: Ethereum Foundation Grant
- Current Status: Prototype
- Features:
  - Gamification Design
  - 30+ Challenges
  - Community Collaboration
- Release: Before July



ONESAVIE

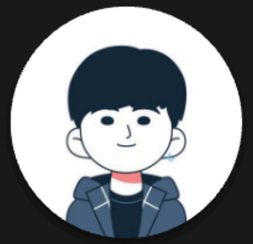
Special Thanks

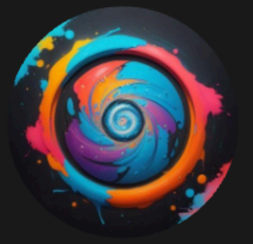
We are inspired by the spirit of community, open source, and selflessness, and hope to gather collective intelligence.

Together Strong!!!

DEFIHACKLABS

Let's make web3 more secure!

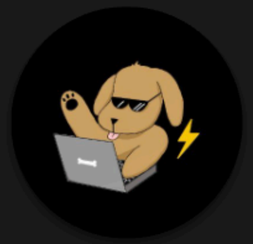



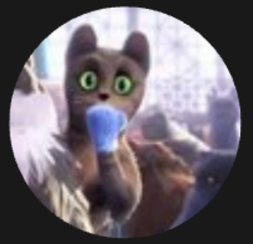






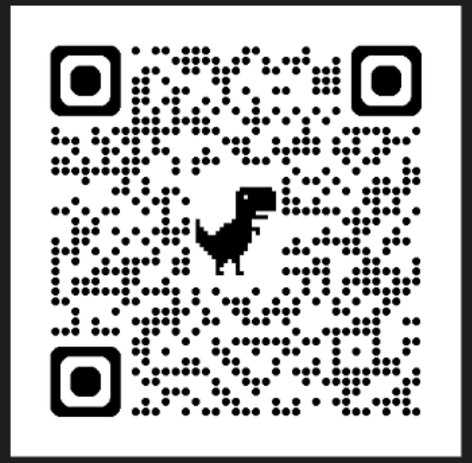




DEFIHACKLABS

Let's make web3 more secure!



OneSavie
Lab


CYBERSEC 2025
TEAM CYBERSECURITY
APRIL 15-17 TaiNEX 2, Taipei
VISIT US @ Booth No.
C 230
1st Floor, TaiNEX 2
Cymetrics
VULCAN