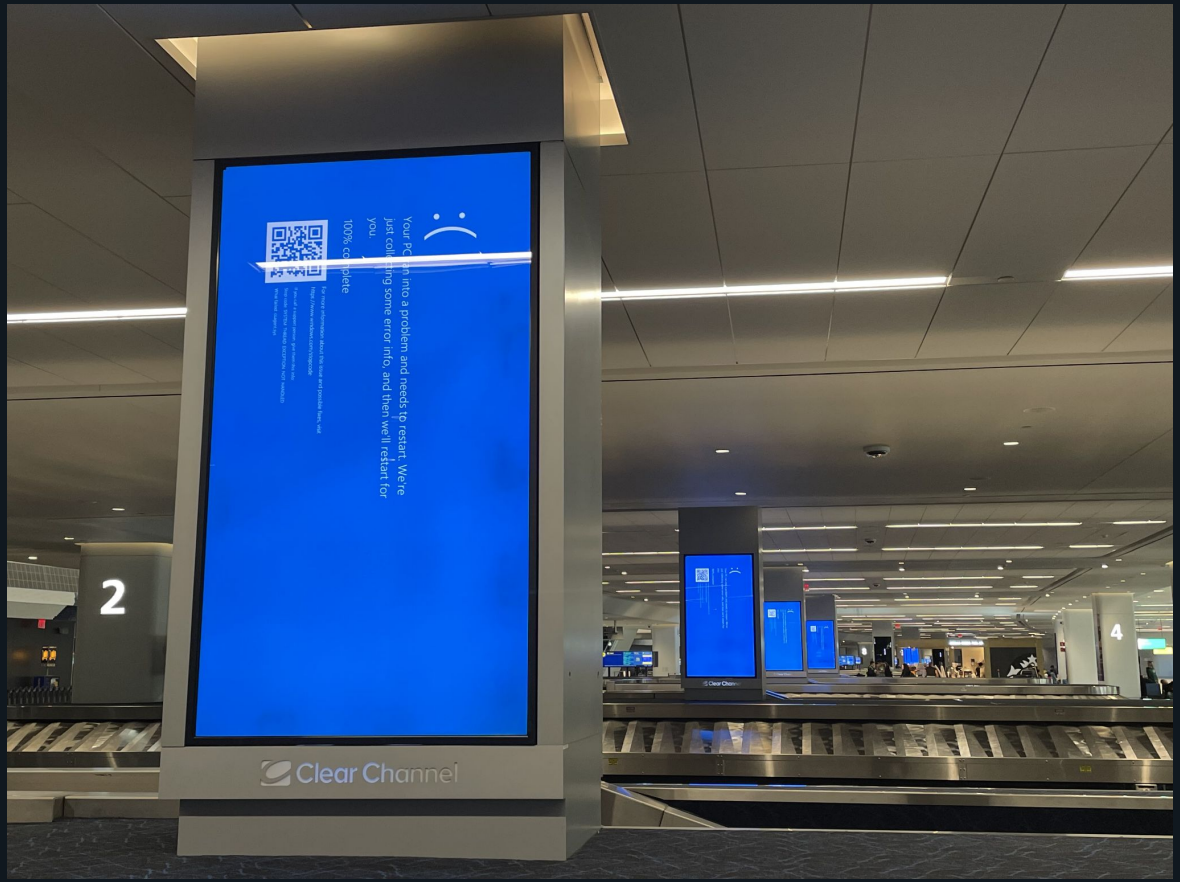
Buttercup and DARPA's AI Cyber Challenge



Source: Wikipedia

BUTTERCUP: World-Class Vulnerability Discovery and Patching

$3 Million Award Winner

28 Vulnerabilities

20 CWE Categories

90%+ Accuracy

👁️ **Watch it work**
Live demo

🧠 **Learn how**
AI + Security

🚀 **Use it**
Open source

Buttercup Demo

🎯 Target
Modified libpng

⚡ Goal
Find and patch a bug

📊 Scale
Laptop to cluster

Keys to success

Accuracy

Versatility

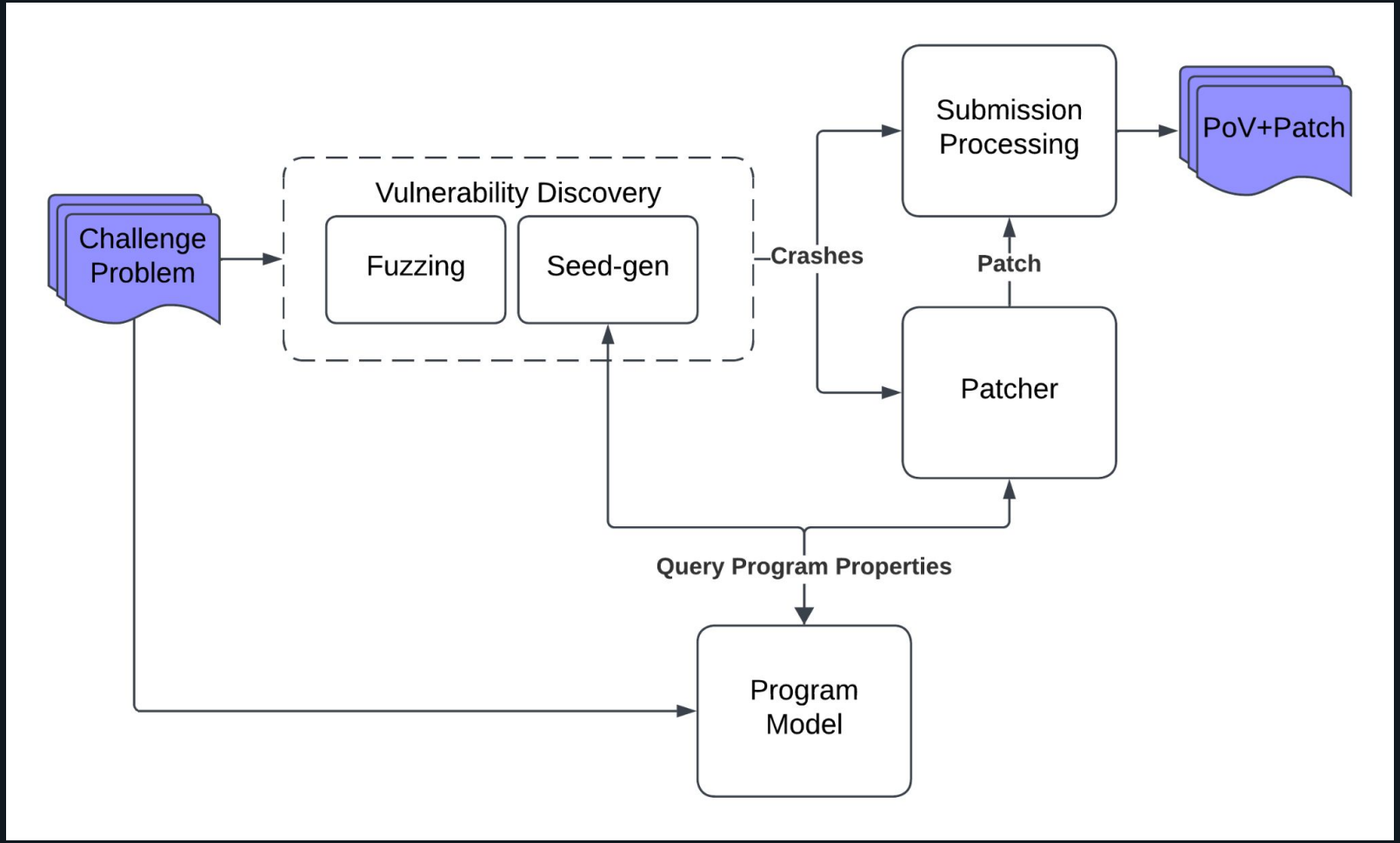
Reliability

Team	LLM spend	Compute spend	Total spend	Cost per point
Team Atlanta	$29.4k	$73.9k	$103.3k	$263
Trail of Bits	**$21.1k**	**$18.5k**	**$39.6k**	**$181**
Theori	$11.5k	$20.3k	$31.8k	$151
fuzzing_brain	$12.2k	$63.2k	$75.4k	$490
Shellphish	$2.9k	$54.9k	$57.8k	$425
42-b3yond-6ug	$1.1k	$38.7k	$39.8k	$379
LACROSSE	$631	$7.1k	$7.2k	$751

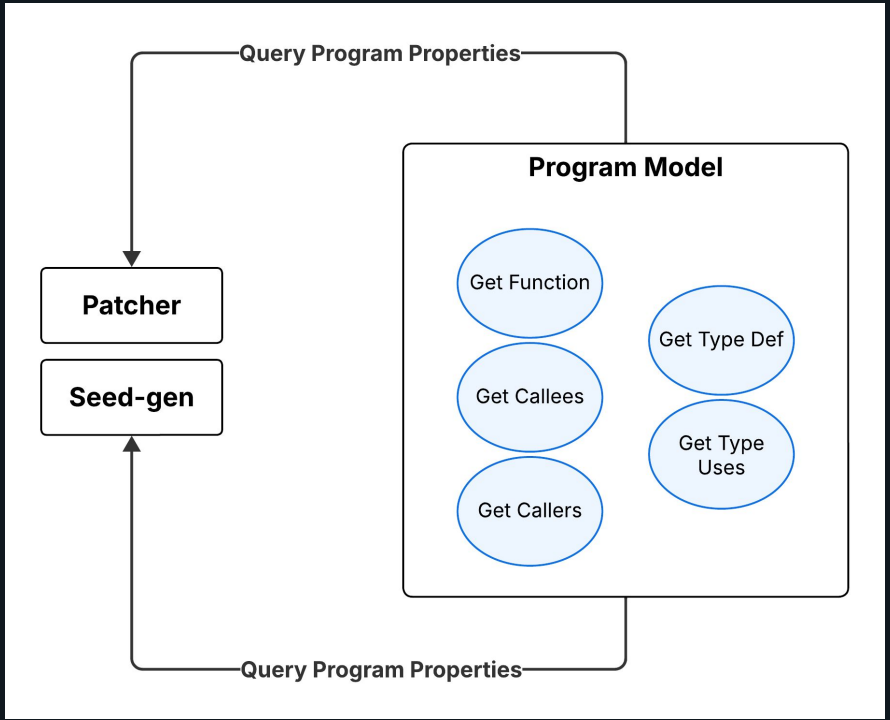
Results!

2nd in LLM Spend
6th in Compute Spend
2nd in $ per Point

Buttercup System Design



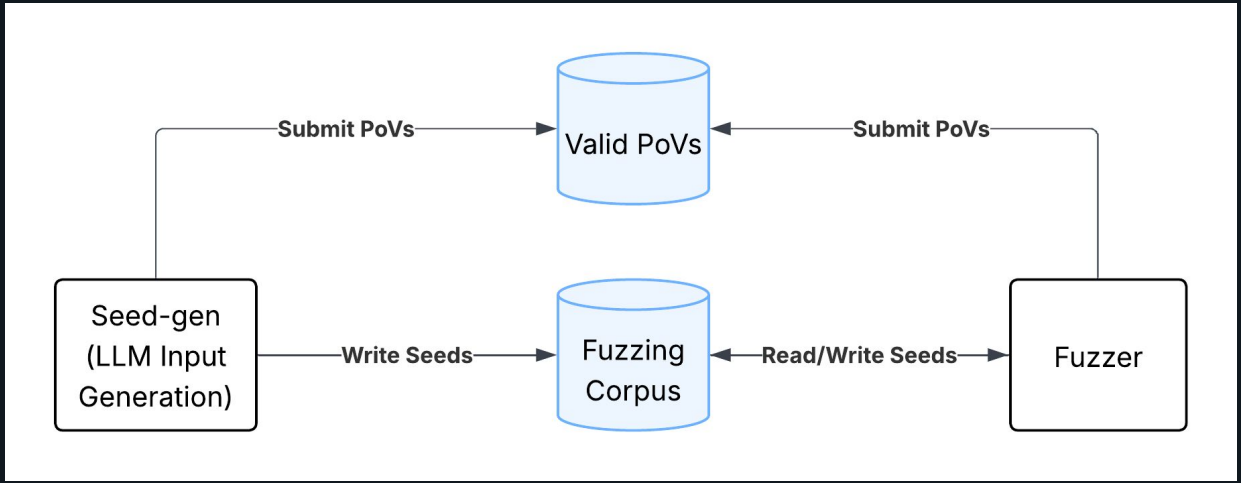
Program Model



- Constructs program model using CodeQuery + Tree-sitter
- Supports querying program properties (functions & types)
- Used by LLM components

Vulnerability Discovery

Approach: Combine fuzzing and LLM input generation



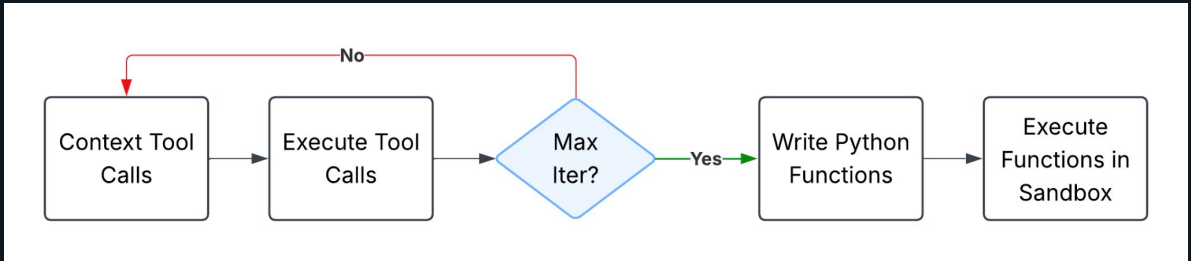
Note: PoVs stands for Proofs of Vulnerability

Fuzzing

- **Standard OSS-Fuzz fuzzers:**
  - LibFuzzer for C/C++
  - Jazzer for Java
- **Fuzzer bots sample active harnesses to run short fuzz campaigns**
- **Merger bots save inputs which improve coverage**
  - Merge a fuzzer bot's local corpus to the shared corpus

Seed-Gen: Enhancing Fuzzing

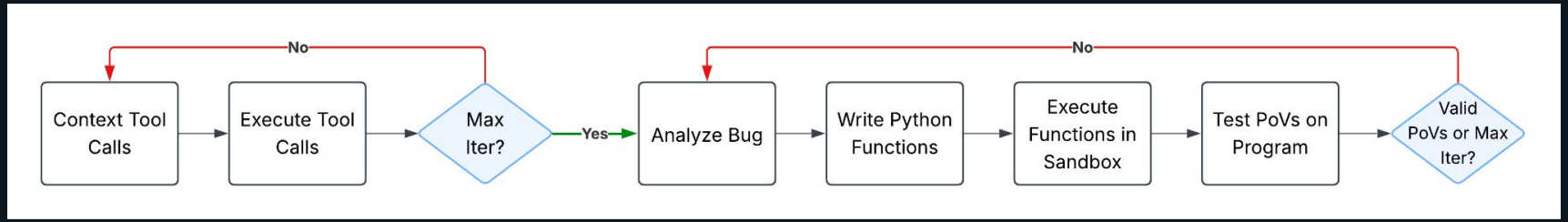
- Generate inputs with LLM to support the fuzzer
- **Initialize Task:** Bootstrap fuzzer with initial seed inputs that exercise harness
- **Explore Task:** Increase coverage for a target function
  - Sample function with low coverage
  - Generate reaching inputs



Initialize and Explore Tasks

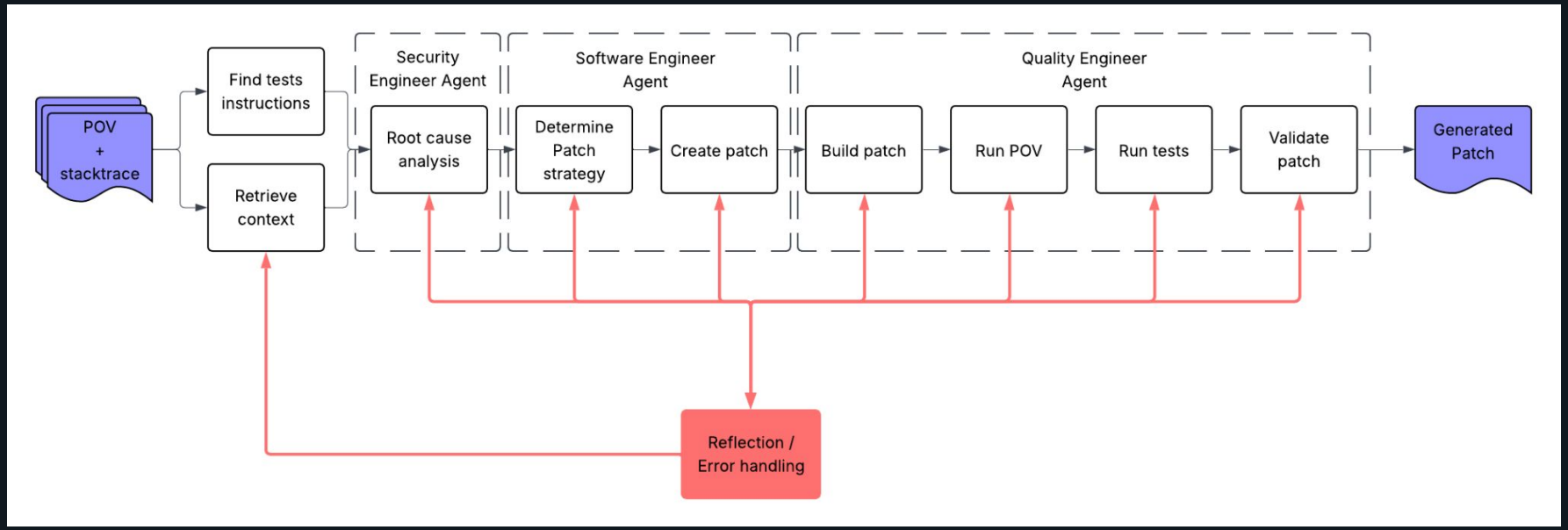
Seed-Gen: Vulnerability Discovery

- **Independently find bugs with an LLM agent**
- **Vuln discovery task: Identify and validate vulnerabilities to create PoVs**
  - Similar to fuzzing: find a triggering input

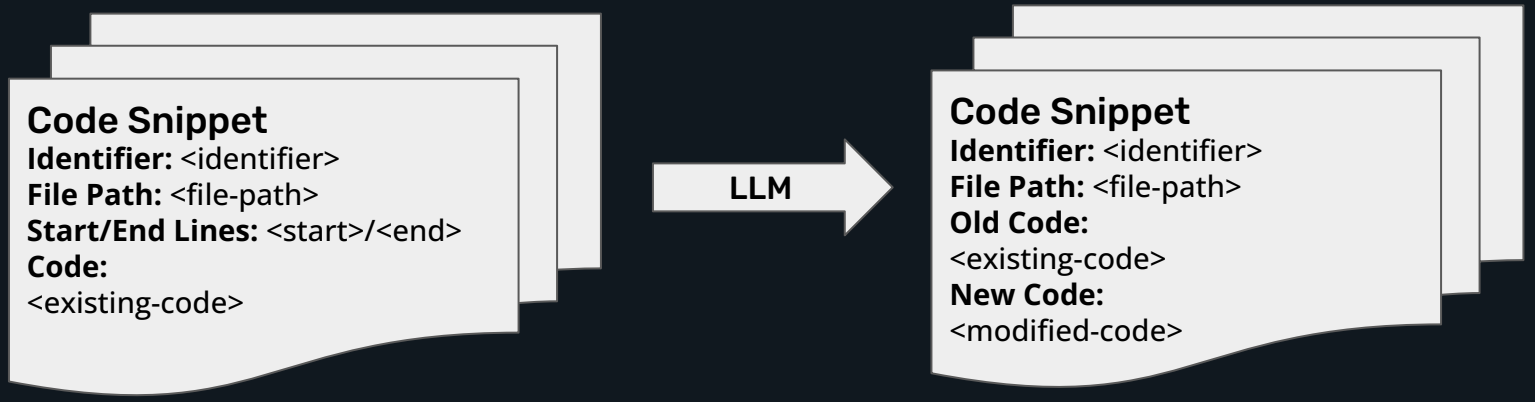


Vulnerability Discovery Task

Patcher


POV
+
stacktrace
Find tests instructions
Retrieve context
Security Engineer Agent
Root cause analysis
Software Engineer Agent
Determine Patch strategy
Create patch
Quality Engineer Agent
Build patch
Run POV
Run tests
Validate patch
Generated Patch
Reflection / Error handling

Patch Creation



Building a Reliable AI System

LLM Testing

- **Experiments**
  - Determine empirically if a change improves performance
- **Observability**
  - Understand what the LLM is doing

Classic Testing

- **Unit, component, and system integration tests**
- **Weekend competition dry runs**

What did we learn about AI?

AI can automate software security tasks–when applied correctly:

- **What would a security researcher need? The LLM probably needs it too!**
  - Tools: the right context is essential
  - Opportunities to reason and test ideas
- **Validate the LLM output with ground truth as much as possible**
  - Reliable bug oracle for seed-gen PoVs
  - Rigorous process to validate LLM patches
- **Give LLMs well-defined problems**
  - Only select problems where it makes sense to use an LLM

Buttercup: Beyond the Competition

Making it accessible

Runs on your laptop, scales to clusters

Expanding language support

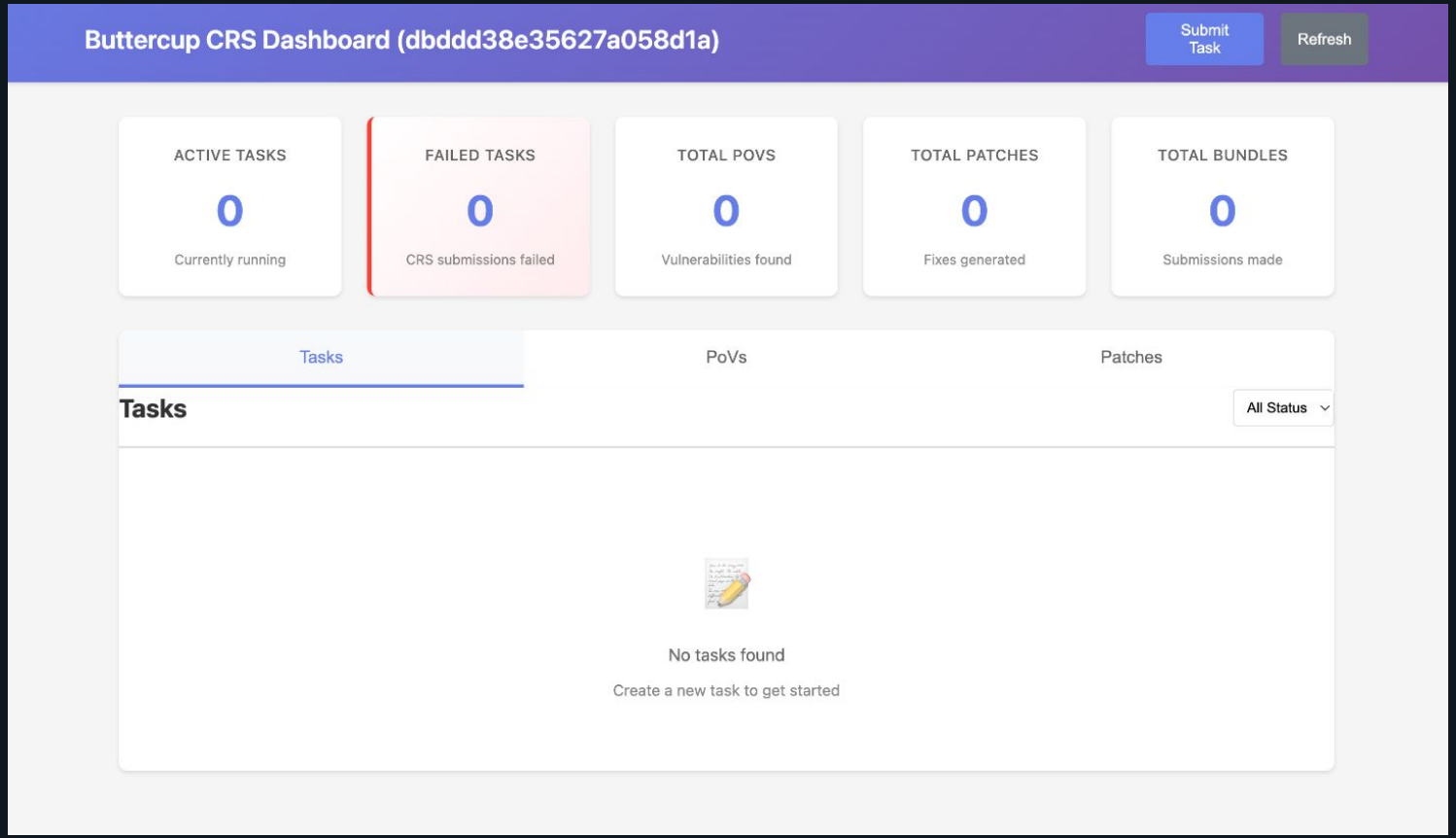
Competition: C/Java **Now:** C++ **Roadmap:** Golang, Rust, Python

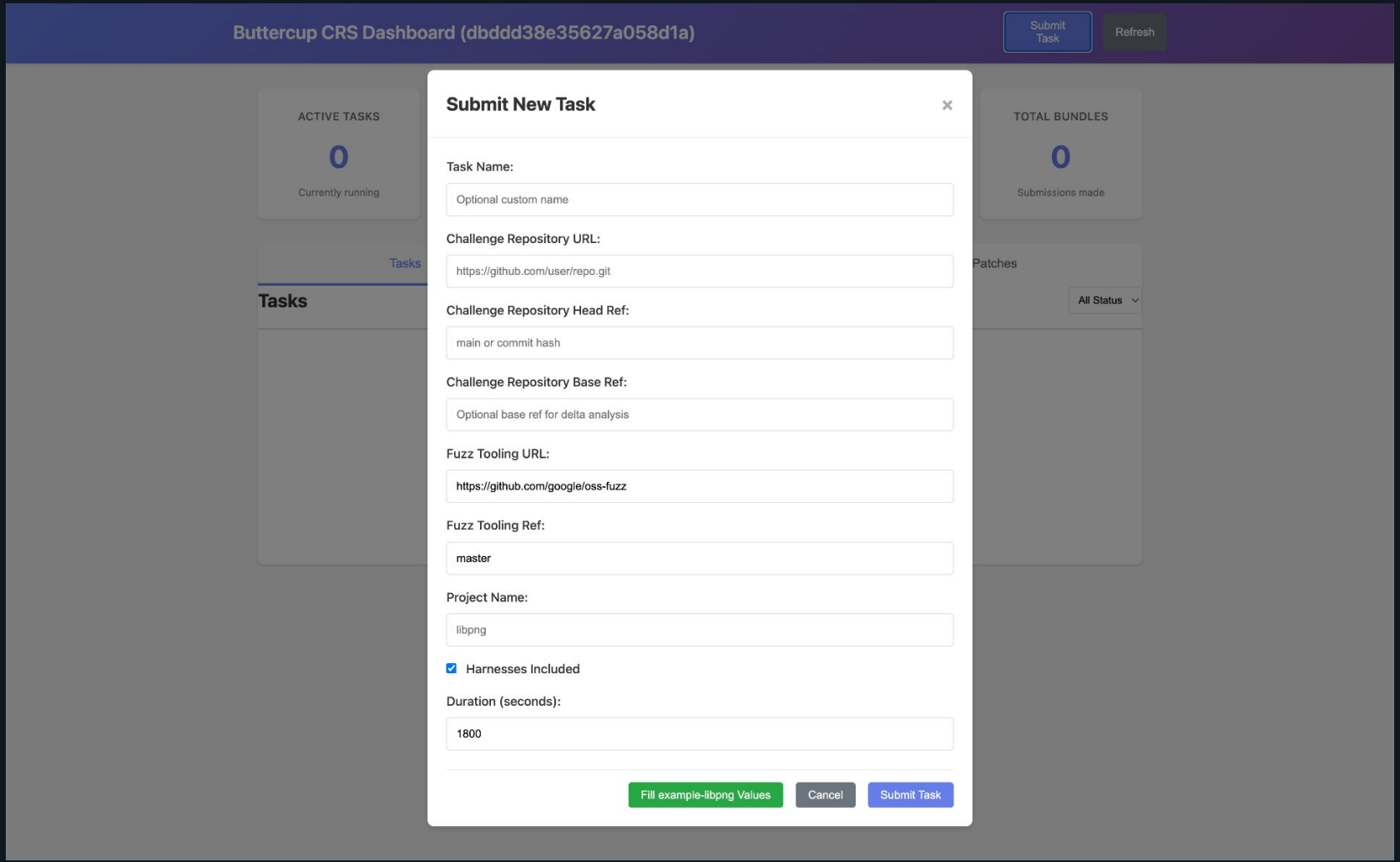
Pushing capabilities forward

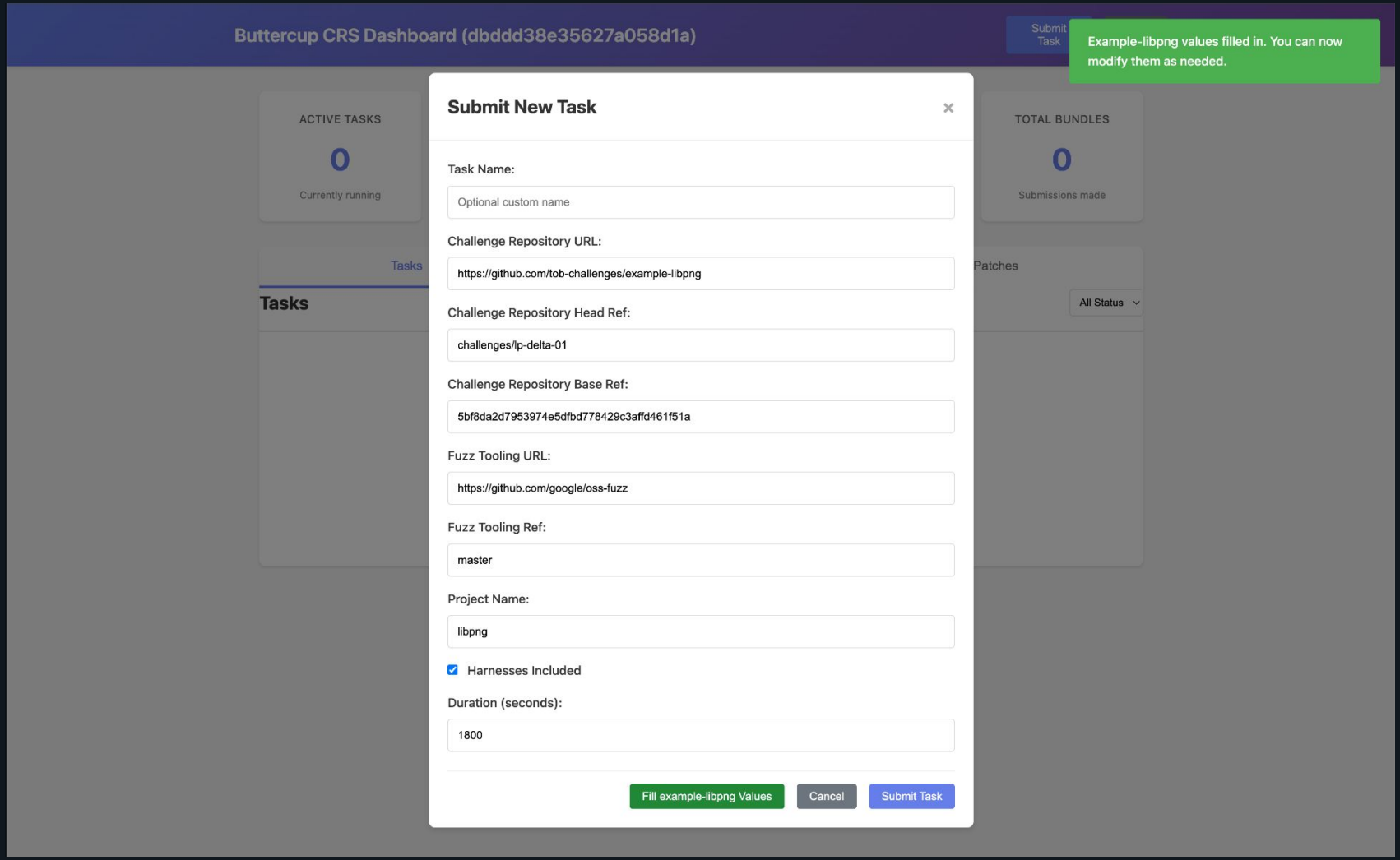
Gemini models • Advanced agentic workflows • Expanded agent tooling

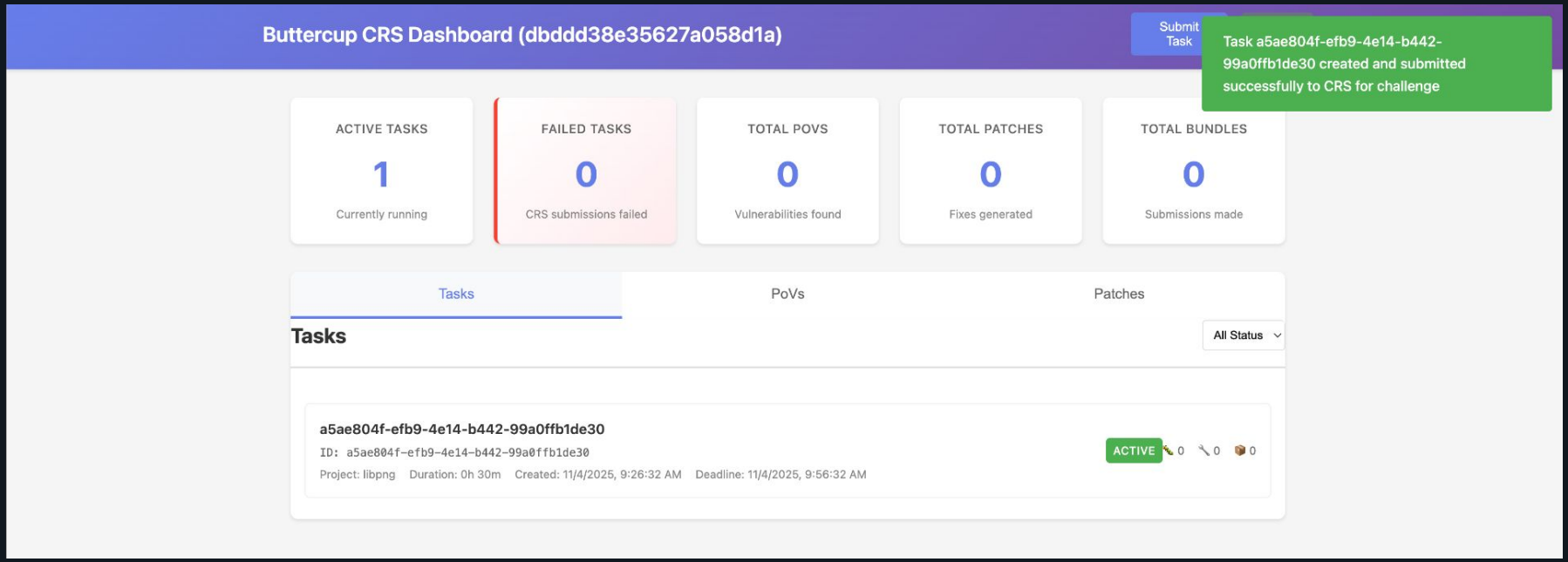
Checking the Demo

Let’s check the Buttercup UI!


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Refresh
ACTIVE TASKS
0
Currently running
FAILED TASKS
0
CRS submissions failed
TOTAL POVS
0
Vulnerabilities found
TOTAL PATCHES
0
Fixes generated
TOTAL BUNDLES
0
Submissions made
Tasks
PoVs
Patches
Tasks
All Status
No tasks found
Create a new task to get started


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Refresh
ACTIVE TASKS
0
Currently running
TOTAL BUNDLES
0
Submissions made
Tasks
Patches
Tasks
All Status
Submit New Task
Task Name:
Optional custom name
Challenge Repository URL:
https://github.com/user/repo.git
Challenge Repository Head Ref:
main or commit hash
Challenge Repository Base Ref:
Optional base ref for delta analysis
Fuzz Tooling URL:
https://github.com/google/oss-fuzz
Fuzz Tooling Ref:
master
Project Name:
libpng
Harnesses Included
Duration (seconds):
1800
Fill example-libpng Values
Cancel
Submit Task


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Example-libpng values filled in. You can now modify them as needed.
ACTIVE TASKS
0
Currently running
TOTAL BUNDLES
0
Submissions made
Tasks
Patches
Tasks
All Status
Submit New Task
Task Name:
Optional custom name
Challenge Repository URL:
https://github.com/tob-challenges/example-libpng
Challenge Repository Head Ref:
challenges/lp-delta-01
Challenge Repository Base Ref:
5bf8da2d7953974e5dfbd778429c3affd461f51a
Fuzz Tooling URL:
https://github.com/google/oss-fuzz
Fuzz Tooling Ref:
master
Project Name:
libpng
Harnesses Included
Duration (seconds):
1800
Fill example-libpng Values
Cancel
Submit Task


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Task a5ae804f-efb9-4e14-b442-99a0ffb1de30 created and submitted successfully to CRS for challenge
ACTIVE TASKS
1
Currently running
FAILED TASKS
0
CRS submissions failed
TOTAL POVS
0
Vulnerabilities found
TOTAL PATCHES
0
Fixes generated
TOTAL BUNDLES
0
Submissions made
Tasks
PoVs
Patches
Tasks
All Status
a5ae804f-efb9-4e14-b442-99a0ffb1de30
ID: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Project: libpng Duration: 0h 30m Created: 11/4/2025, 9:26:32 AM Deadline: 11/4/2025, 9:56:32 AM
ACTIVE 0 0 0

Buttercup CRS Dashboard (dbddd38e35627a058d1a)

Submit Task

Refresh

Task: a5ae804f-efb9-4e14-b442-99a0ffb1de30

Task Information

Name:	a5ae804f-efb9-4e14-b442-99a0ffb1de30
ID:	a5ae804f-efb9-4e14-b442-99a0ffb1de30
Project:	libpng
Status:	ACTIVE
Created:	11/4/2025, 9:26:32 AM
Deadline:	11/4/2025, 9:56:32 AM

PoVs (Vulnerabilities) (0)

No povs (vulnerabilities) found

Patches (0)

No patches found

Bundles (0)

No bundles found

Task: a5ae804f-efb9-4e14-b442-99a0ffb1de30

Task Information

Name:	a5ae804f-efb9-4e14-b442-99a0ffb1de30
ID:	a5ae804f-efb9-4e14-b442-99a0ffb1de30
Project:	libpng
Status:	ACTIVE
Created:	11/4/2025, 9:26:32 AM
Deadline:	11/4/2025, 9:56:32 AM

PoVs (Vulnerabilities) (1)

126030c7-3e0c-47d7-9389-21d01942f258 — 11/4/2025, 9:30:50 AM — Download

Type: Binary Data
Size: 133 bytes
Hex Preview:

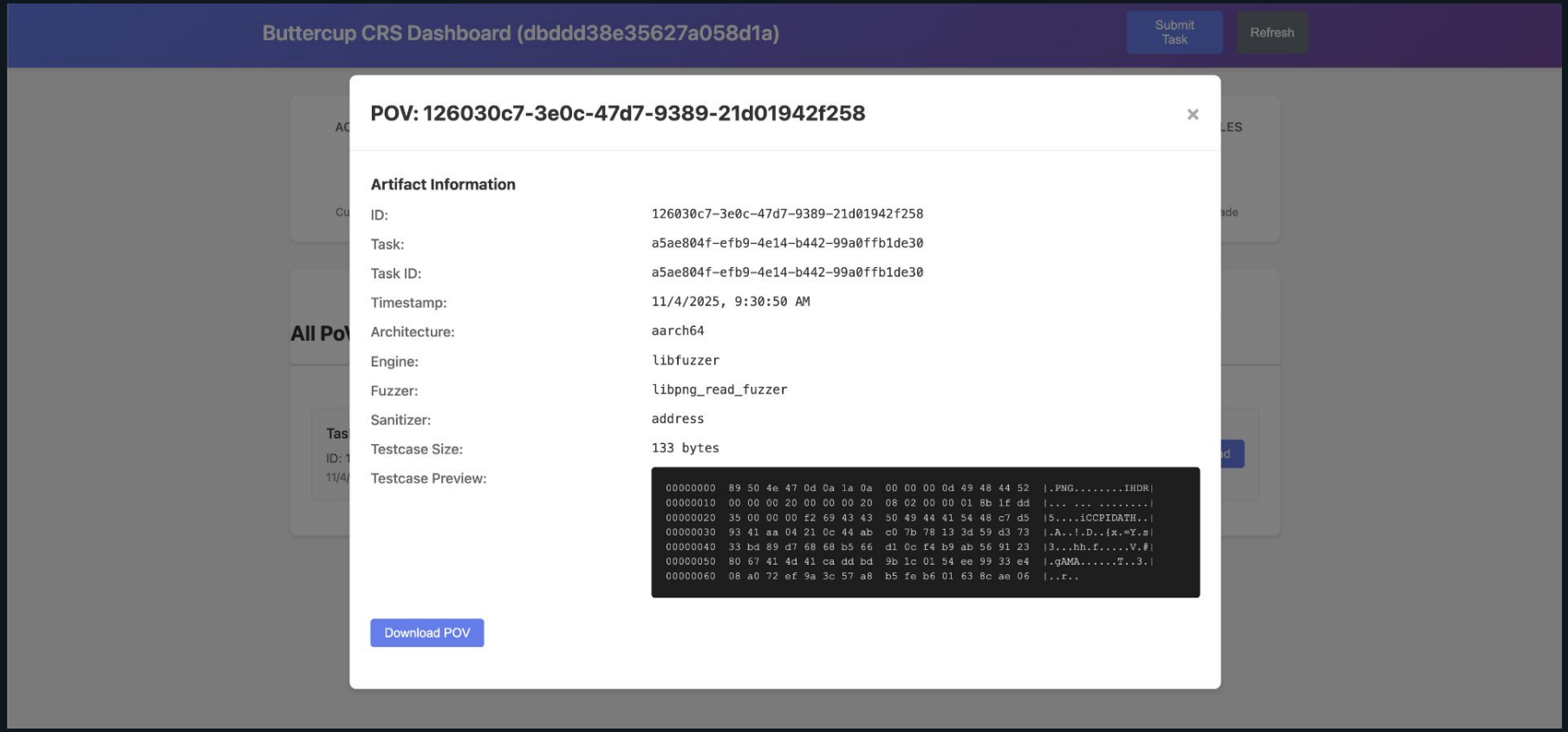
```
00000000  89 50 4e 47 0d 0a 1a 0a  00 00 00 0d 49 48 44 52  |.PNG........IHDR|
00000010  00 00 00 20 00 00 00 20  08 02 00 00 01 8b 1f dd  |... ... ........|
00000020  35 00 00 00 f2 69 43 43  50 49 44 41 54 48 c7 d5  |5....iCCPIDATH..|
00000030  93 41 aa 04 21 0c 44 ab  c0 7b 78 13 3d 59 d3 73  |.A..!.D..{x.=Y.s|
00000040  33 bd 89 d7 68 68 b5 66  d1 0c f4 b9 ab 56 91 23  |3...hh.f.....V.#|
00000050  80 67 41 4d 41 ca dd bd  9b 1c 01 54 ee 99 33 e4  |.gAMA......T..3.|
00000060  08 a0 72 ef 9a 3c 57 a8  b5 fe b6 01 63 8c ae 06  |..r..<W.....c...|
```

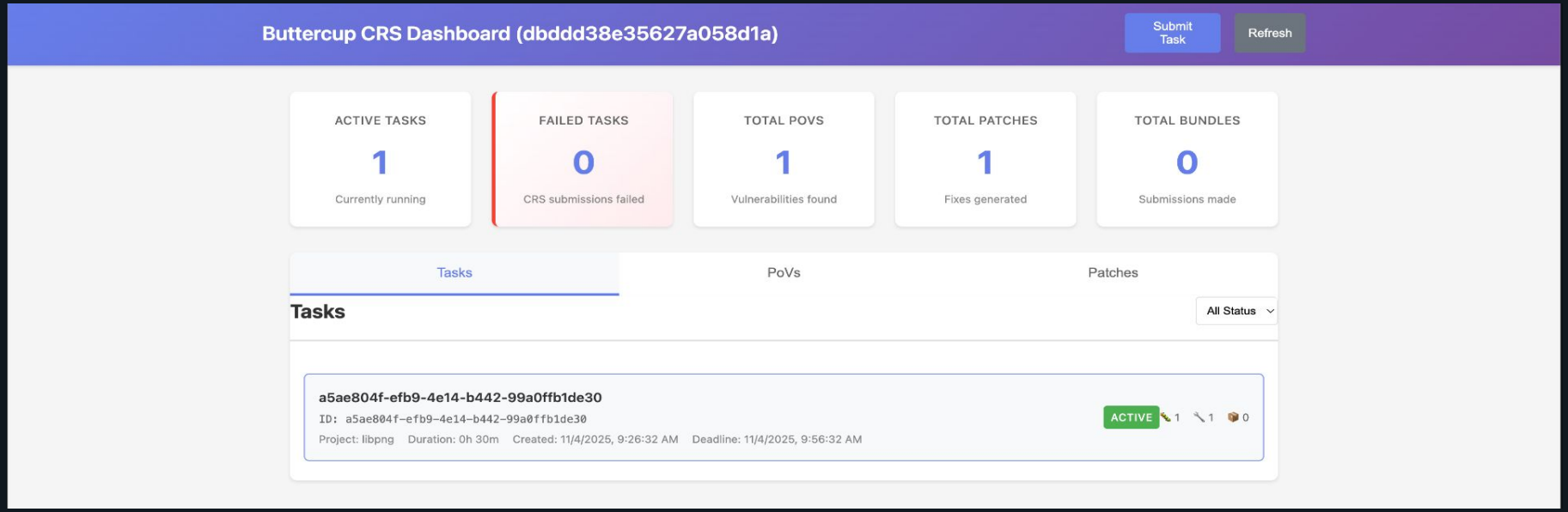
Patches (0)

No patches found

Bundles (0)

No bundles found


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Refresh
POV: 126030c7-3e0c-47d7-9389-21d01942f258
Artifact Information
ID: 126030c7-3e0c-47d7-9389-21d01942f258
Task: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Task ID: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Timestamp: 11/4/2025, 9:30:50 AM
Architecture: aarch64
Engine: libfuzzer
Fuzzer: libpng_read_fuzzer
Sanitizer: address
Testcase Size: 133 bytes
Testcase Preview:
00000000 89 50 4e 47 0d 0a 1a 0a 00 00 00 0d 49 48 44 52 |.PNG........IHDR|
00000010 00 00 00 20 00 00 00 20 08 02 00 00 01 8b 1f dd |... ... .|
00000020 35 00 00 00 f2 69 43 43 50 49 44 41 54 48 c7 d5 |5....iCCPIDATH..|
00000030 93 41 aa 04 21 0c 44 ab c0 7b 78 13 3d 59 d3 73 |.A..!.D..{x.=Y.s|
00000040 33 bd 89 d7 68 68 b5 66 d1 0c f4 b9 ab 56 91 23 |3...hh.f.....V.#|
00000050 80 67 41 4d 41 ca dd bd 9b 1c 01 54 ee 99 33 e4 |.gAMA......T..3.|
00000060 08 a0 72 ef 9a 3c 57 a8 b5 fe b6 01 63 8c ae 06 |..r..
Download POV


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Refresh
ACTIVE TASKS
1
Currently running
FAILED TASKS
0
CRS submissions failed
TOTAL POVS
1
Vulnerabilities found
TOTAL PATCHES
1
Fixes generated
TOTAL BUNDLES
0
Submissions made
Tasks
PoVs
Patches
Tasks
All Status
a5ae804f-efb9-4e14-b442-99a0ffb1de30
ID: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Project: libpng Duration: 0h 30m Created: 11/4/2025, 9:26:32 AM Deadline: 11/4/2025, 9:56:32 AM
ACTIVE 1 1 0

Buttercup CRS Dashboard (dbddd38e35627a058d1a)

Submit Task | Refresh

Task: a5ae804f-efb9-4e14-b442-99a0ffb1de30

Task Information

Name:	a5ae804f-efb9-4e14-b442-99a0ffb1de30
ID:	a5ae804f-efb9-4e14-b442-99a0ffb1de30
Project:	libpng
Status:	ACTIVE
Created:	11/4/2025, 9:26:32 AM
Deadline:	11/4/2025, 9:56:32 AM

PoVs (Vulnerabilities) (1)

126030c7-3e0c-47d7-9389-21d01942f258 — 11/4/2025, 9:30:50 AM — Download

Type: Binary Data
Size: 133 bytes
Hex Preview:

```
00000000  89 50 4e 47 0d 0a 1a 0a  00 00 00 0d 49 48 44 52  |.PNG........IHDR|
00000010  00 00 00 20 00 00 00 20  08 02 00 00 01 8b 1f dd  |... ... ........|
00000020  35 00 00 00 f2 69 43 43  50 49 44 41 54 48 c7 d5  |5....iCCPIDATH..|
00000030  93 41 aa 04 21 0c 44 ab  c0 7b 78 13 3d 59 d3 73  |.A..!.D..{x.=Y.s|
00000040  33 bd 89 d7 68 68 b5 66  d1 0c f4 b9 ab 56 91 23  |3...hh.f.....V.#|
00000050  80 67 41 4d 41 ca dd bd  9b 1c 01 54 ee 99 33 e4  |.gAMA......T..3.|
00000060  08 a0 72 ef 9a 3c 57 a8  b5 fe b6 01 63 8c ae 06  |..r..<W.....c...|
```

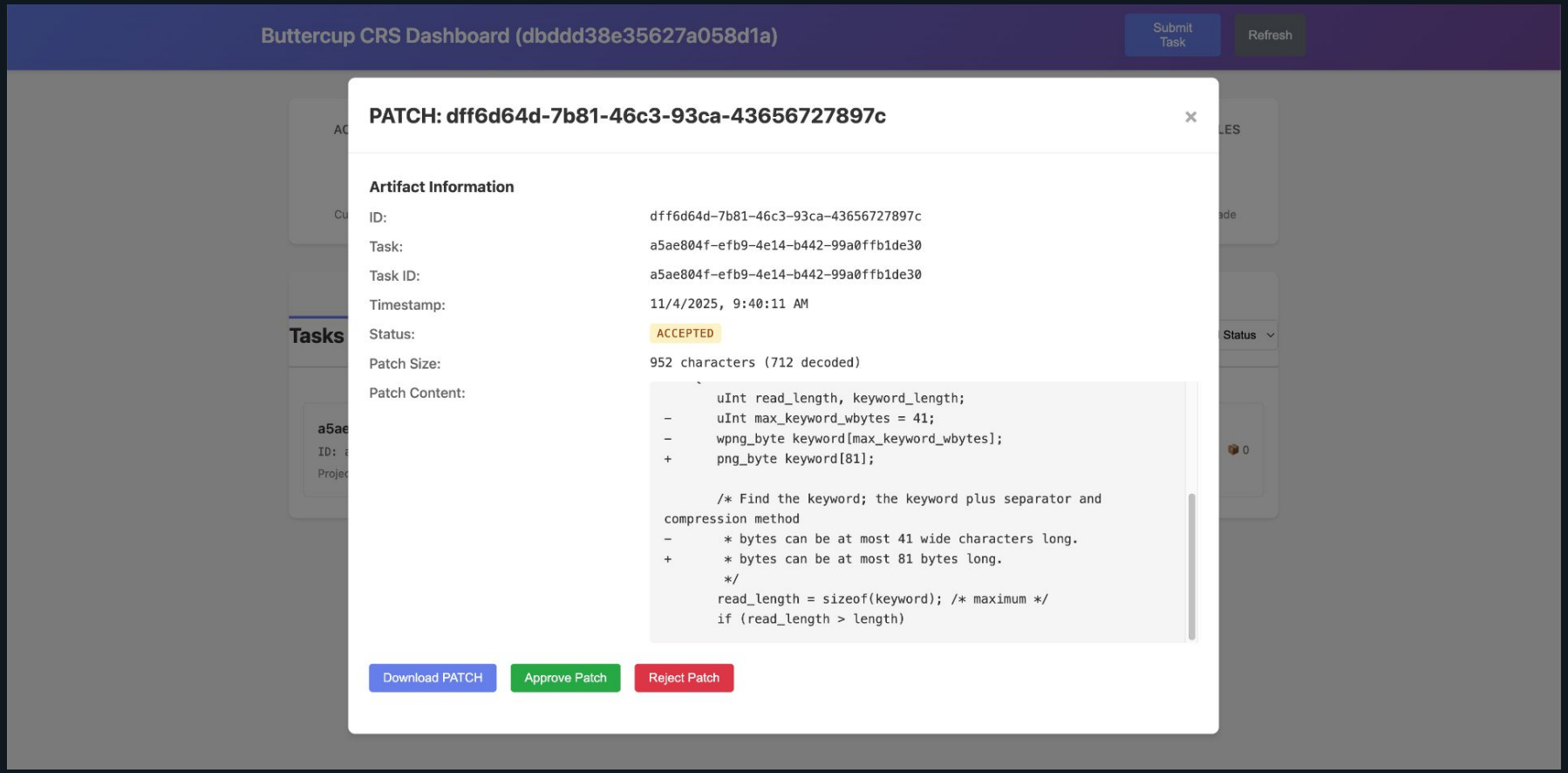
Patches (1)

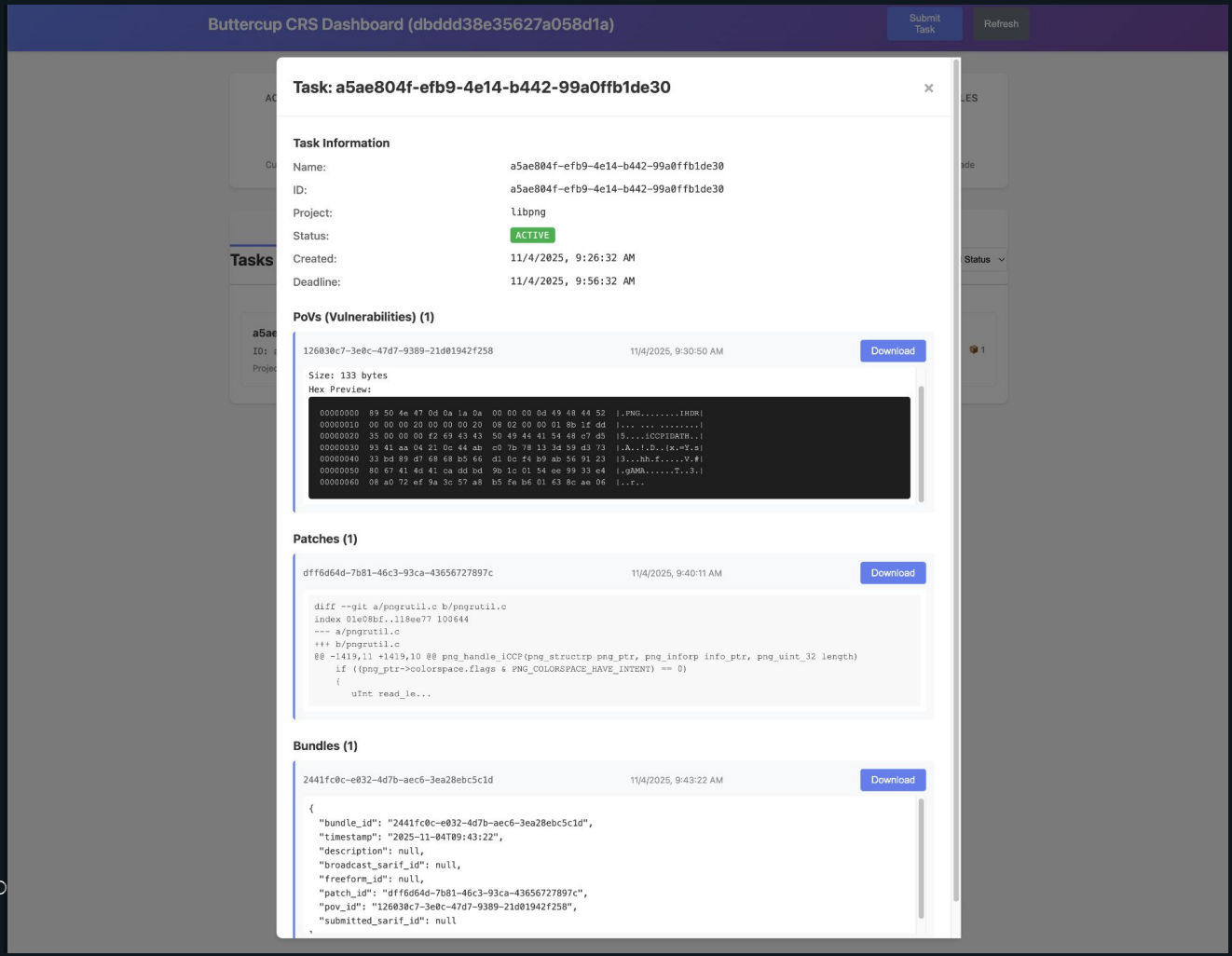
dff6d64d-7b81-46c3-93ca-43656727897c — 11/4/2025, 9:40:11 AM — Download

```
diff --git a/pngrutil.c b/pngrutil.c
index 01e08bf..118ee77 100644
--- a/pngrutil.c
+++ b/pngrutil.c
@@ -1419,11 +1419,10 @@ png_handle_iCCP(png_structrp png_ptr, png_inforp info_ptr, png_uint_32 length)
    if ((png_ptr->colorspace.flags & PNG_COLORSPACE_HAVE_INTENT) == 0)
    {
       uInt read_le...
```

Bundles (0)

No bundles found


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Refresh
PATCH: dff6d64d-7b81-46c3-93ca-43656727897c
Artifact Information
ID: dff6d64d-7b81-46c3-93ca-43656727897c
Task: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Task ID: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Timestamp: 11/4/2025, 9:40:11 AM
Status: ACCEPTED
Patch Size: 952 characters (712 decoded)
Patch Content:
uInt read_length, keyword_length;
- uInt max_keyword_wbytes = 41;
- wpng_byte keyword[max_keyword_wbytes];
+ png_byte keyword[81];
/* Find the keyword; the keyword plus separator and compression method
- * bytes can be at most 41 wide characters long.
+ * bytes can be at most 81 bytes long.
*/
read_length = sizeof(keyword); /* maximum */
if (read_length > length)
Download PATCH
Approve Patch
Reject Patch
Tasks
Status


Buttercup CRS Dashboard (dbddd38e35627a058d1a)
Submit Task
Refresh
Task: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Task Information
Name: a5ae804f-efb9-4e14-b442-99a0ffb1de30
ID: a5ae804f-efb9-4e14-b442-99a0ffb1de30
Project: libpng
Status: ACTIVE
Created: 11/4/2025, 9:26:32 AM
Deadline: 11/4/2025, 9:56:32 AM
PoVs (Vulnerabilities) (1)
126030c7-3e0c-47d7-9389-21d01942f258
11/4/2025, 9:30:50 AM
Download
Size: 133 bytes
Hex Preview:
00000000 89 50 4e 47 0d 0a 1a 0a 00 00 00 0d 49 48 44 52 |.PNG........IHDR|
00000010 00 00 00 20 00 00 00 20 08 02 00 00 01 8b 1f dd |... ... .|
00000020 35 00 00 00 f2 69 43 43 50 49 44 41 54 48 c7 d5 |5....iCCPIDATH..|
00000030 93 41 aa 04 21 0c 44 ab c0 7b 78 13 3d 59 d3 73 |.A..!.D..{x.=Y.s|
00000040 33 bd 89 d7 68 68 b5 66 d1 0c f4 b9 ab 56 91 23 |3...hh.f.....V.#|
00000050 80 67 41 4d 41 ca dd bd 9b 1c 01 54 ee 99 33 e4 |.gAMA......T..3.|
00000060 08 a0 72 ef 9a 3c 57 a8 b5 fe b6 01 63 8c ae 06 |..r..
Patches (1)
dff6d64d-7b81-46c3-93ca-43656727897c
11/4/2025, 9:40:11 AM
Download
diff --git a/pngrutil.c b/pngrutil.c
index 01e08bf..118ee77 100644
--- a/pngrutil.c
+++ b/pngrutil.c
@@ -1419,11 +1419,10 @@ png_handle_iCCP(png_structrp png_ptr, png_inforp info_ptr, png_uint_32 length)
if ((png_ptr->colorspace.flags & PNG_COLORSPACE_HAVE_INTENT) == 0)
{
uInt read_le...
Bundles (1)
2441fc0c-e032-4d7b-aec6-3ea28ebc5c1d
11/4/2025, 9:43:22 AM
Download
{
"bundle_id": "2441fc0c-e032-4d7b-aec6-3ea28ebc5c1d",
"timestamp": "2025-11-04T09:43:22",
"description": null,
"broadcast_sarif_id": null,
"freeform_id": null,
"patch_id": "dff6d64d-7b81-46c3-93ca-43656727897c",
"pov_id": "126030c7-3e0c-47d7-9389-21d01942f258",
"submitted_sarif_id": null

How to Use Buttercup

1.

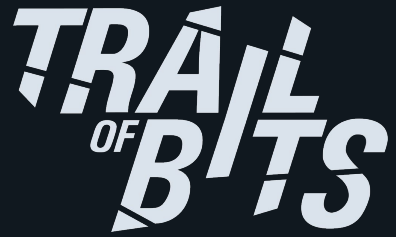
```
$ make setup-local
<api-key>, next, next, …
$ make deploy
$ make web-ui
```

2. https://github.com/org/your-oss-fuzz-compatible-project

3. **Vulnerability exposing input**
Patch that remediates the vulnerability

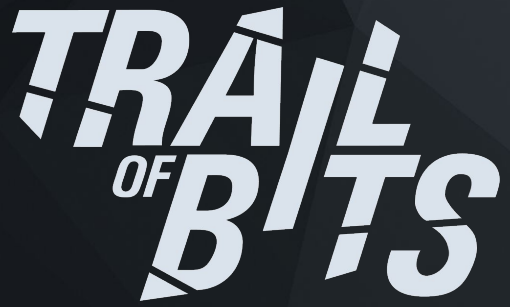
Try out Buttercup:

github.com/trailofbits/buttercup



Website: trailofbits.com

Blog: blog.trailofbits.com


TRAIL
OF
BITS