

Target pixelvault-host-image

No malicious behavior observed

We ran it 5 times, with a normal setup, the environment variables it says it needs, PIXELVAULT_API_KEY set and cloud credentials present. It contacted api.pixelvault.dev. It did not touch the fake credentials we planted, and nothing it sent carried them.

Recommendation	Confidence	Findings
No action needed	n/a	0 high 0 med 3 info

Nothing observed calls for action. This is evidence, not a guarantee.

Scope

Package	pixelvault-host-image
Source	clawhub 1.0.0
SHA-256	d8d322c50feee976b642e128ccf0718e596bfca0165954e707c13bdae262b5d6
Files	3
Scanned	27 Oct 2026, 13:09 UTC
Engine	detonator 0.5.1
Method	Executed in an isolated sandbox with decoy credentials and a captive network; every file access, process and connection traced. 5 runs: baseline, declared-env, guided:PIXELVAULT_API_KEY, cloud-creds.
Note	batch:clawhub:2026-10-27

Findings summary

ID	Severity	Confidence	Finding
F01	Low	Info	Network destinations contacted
F02	Low	Info	The fake credential we configured was sent to its API (expected)
F03	Low	Info	Loaded the project's .env file

Observed behavior

Install step	none
Runs as	<pre>curl -sS -F file=@/path/to/image.png https://api.pixelvault.dev/v1/images; curl -sS -X POST https://api.pixelvault.dev/v1/images -H; curl -sS -F file=@/path/to/image.png -H Authorization: Bearer HTENVdBKxs4is4V665PI09BYYtVN; curl -sS -F file=@/path/to/image.png -H</pre>



	Authorization: Bearer HTENVsQqMGMGapYWP3eWuJgeVfTW; ls -la; head -50
Network	api.pixelvault.dev (8 conn, 1945 B sent; while running)
Credentials	./env (read)
Files written	nothing outside its own folder

Finding details

F01 Network destinations contacted

Severity: Low Confidence: Info Observed in: obedient:baseline
The package connected to these hosts.

Description

Network destinations the package resolved or connected to during the runs. All traffic was captured by the sandbox; nothing reached the internet.

Impact

Context only. Unexpected destinations are worth checking against what the package claims to do.

Evidence

```
api.pixelvault.dev
```

Remediation

Confirm each host is expected (the package's API, its registry). Allowlist expected egress in production.

F02 The fake credential we configured was sent to its API (expected)

Severity: Low Confidence: Info Observed in: obedient:declared-env
The fake API key we configured to start the server was sent to its own service.

Description

A fake credential Detonator configured to start the server was sent to its own API.

Impact

Expected for servers that wrap an online service.

Evidence

```
$PIXELVAULT_API_KEY: sent to api.pixelvault.dev
```

Remediation

Confirm the host is the service the key belongs to.



F03 Loaded the project's .env file

Severity: Low Confidence: Info Observed in: llm:baseline
The package loaded the project's .env file.

Description

The process loaded the project's .env file. Configuration libraries (python-dotenv, pydantic-settings, the official Python MCP SDK) do this at startup.

Impact

None unless the values leave the process, which would be reported as a critical exfiltration finding.

Evidence

```
./.env (by sed)
```

Remediation

No action needed. Keep long-lived secrets out of project .env files where possible.

Run log

obedient:baseline ok 1.2s

new behavior: dns:api.pixelvault.dev, exec:curl

obedient:declared-env ok 1.9s

new behavior: handover:config-secret-egress

obedient:guided:PIXELVAULT_API_KEY ok 1.3s

no new behavior

obedient:cloud-creds ok 1.5s

no new behavior

llm:baseline ok 19.2s

new behavior: ownread:./.env | 8 model steps, 8 commands

Observed behavior only. A clean result does not prove a package is safe. Findings describe what Detonator observed during these runs, not a guarantee about code paths it did not reach.