Finding Malicious PyPI Packages in the Wild

Frederic Baguelin

What's a malicious software package?

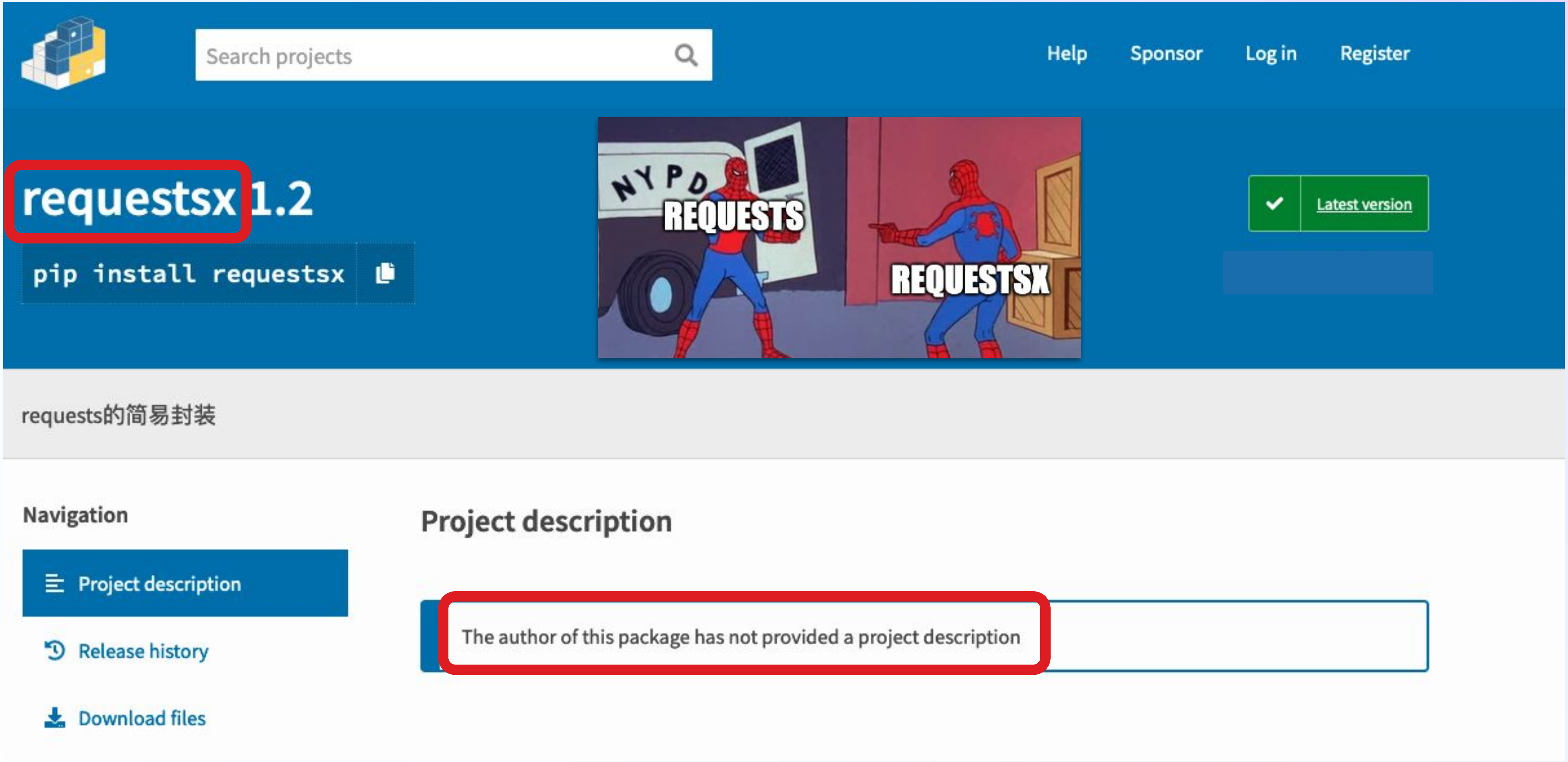
MidJourney, "malicious software package"

Malicious Software Package Attack Chain (PyPI)

Installation vector	Code execution trigger	Code execution mechanism	Objective
Typosquatting	Custom install command	Execute remote script	Steal credentials
Compromise developer PyPI account	Malicious code in __init__.py	Execute remote binary	Steal sensitive information
Compromise developer GitHub	Proxy legitimate library function	Dynamically eval code	Steal cryptocurrency
Social engineering		Backdoor condition	Mine cryptocurrency
			Trojanize system

Unmasking the Culprits
∾
Techniques for Detecting Malicious Packages

What do we want to detect?



requestsx 1.2

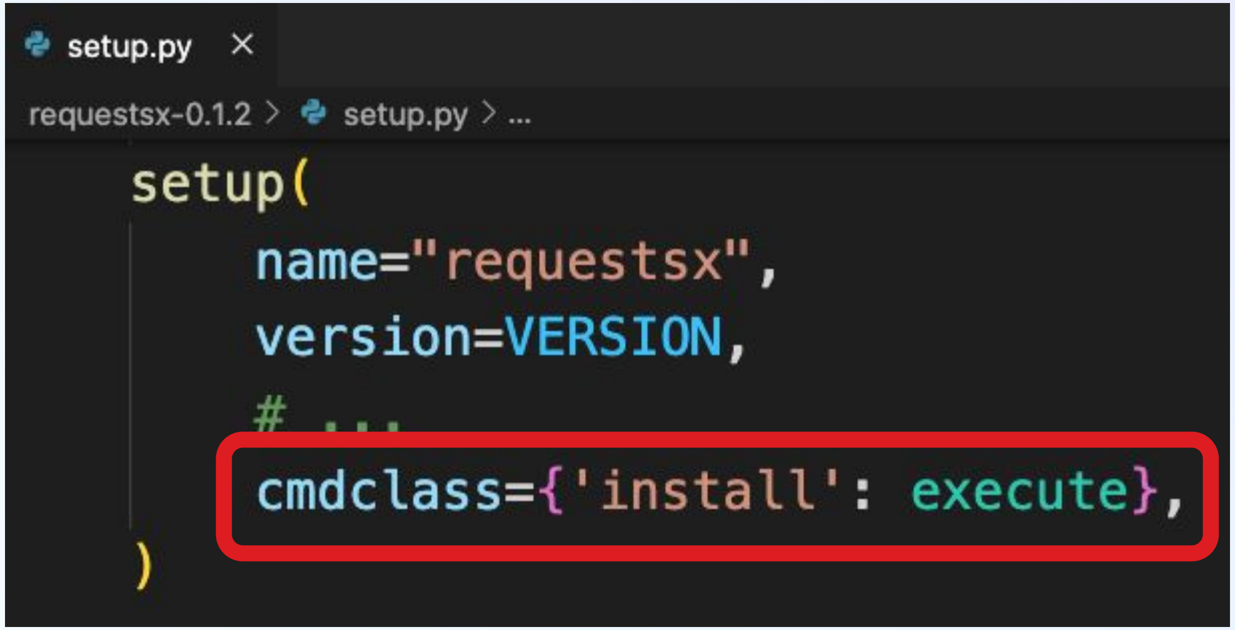
```
pip install requestsx
```

requests的简易封装

Project description

The author of this package has not provided a project description

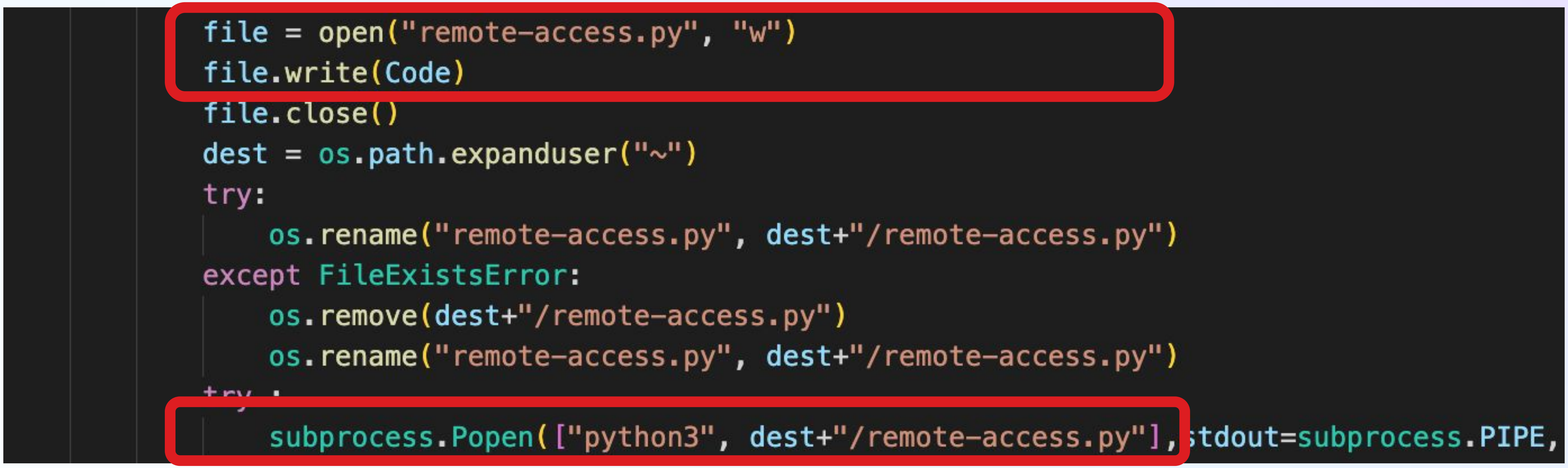
What do we want to detect?



```
setup(
    name="requestsx",
    version=VERSION,
    # ...
    cmdclass={'install': execute},
)
```

Package defining a custom setup script running automatically after "pip install"

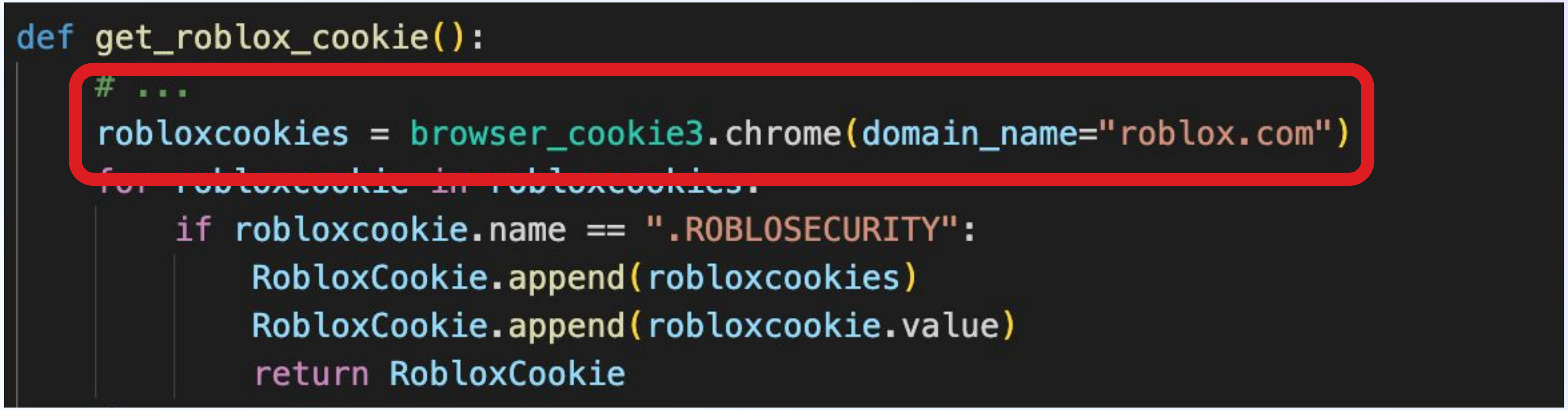
What do we want to detect?



```
file = open("remote-access.py", "w")
file.write(Code)
file.close()
dest = os.path.expanduser("~")
try:
    os.rename("remote-access.py", dest+"/remote-access.py")
except FileExistsError:
    os.remove(dest+"/remote-access.py")
    os.rename("remote-access.py", dest+"/remote-access.py")
try :
    subprocess.Popen(["python3", dest+"/remote-access.py"], stdout=subprocess.PIPE,
```

Package writing to a ".py" file and spawning a new Python process

What do we want to detect?



```
def get_roblox_cookie():
    # ...
    robloxcookies = browser_cookie3.chrome(domain_name="roblox.com")
    for robloxcookie in robloxcookies:
        if robloxcookie.name == ".ROBLOSECURITY":
            RobloxCookie.append(robloxcookies)
            RobloxCookie.append(robloxcookie.value)
            return RobloxCookie
```

Package reading browser cookies

GuardDog

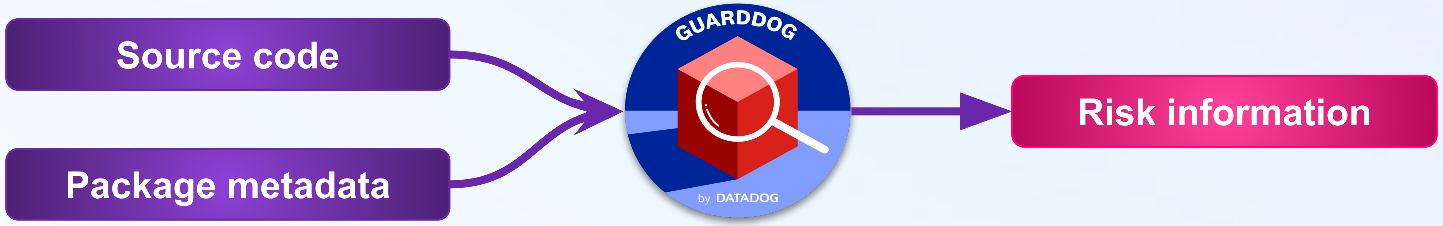
Your Open-Source Sentinel Against Malicious Packages

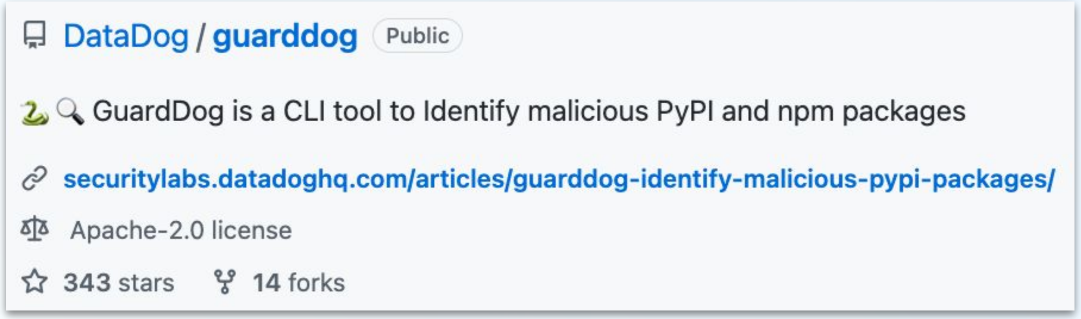
GuardDog

- ➢ Fully open source and self-contained

- ➢ Supports PyPI, npm

NEW





DataDog / **guarddog** Public

🐍🔍 GuardDog is a CLI tool to Identify malicious PyPI and npm packages

securitylabs.datadoghq.com/articles/guarddog-identify-malicious-pypi-packages/

Apache-2.0 license

☆ 343 stars 14 forks

Metadata heuristics

- ➢ Written in Python directly into GuardDog
- ➢ Based on Pypi’s metadata
- ➢ Highlight issues in packages’ health or potential installation vectors

Examples:

- ➢ Empty description on PyPI
- ➢ Name close to a popular package (typosquatting)

Source code analysis with Semgrep

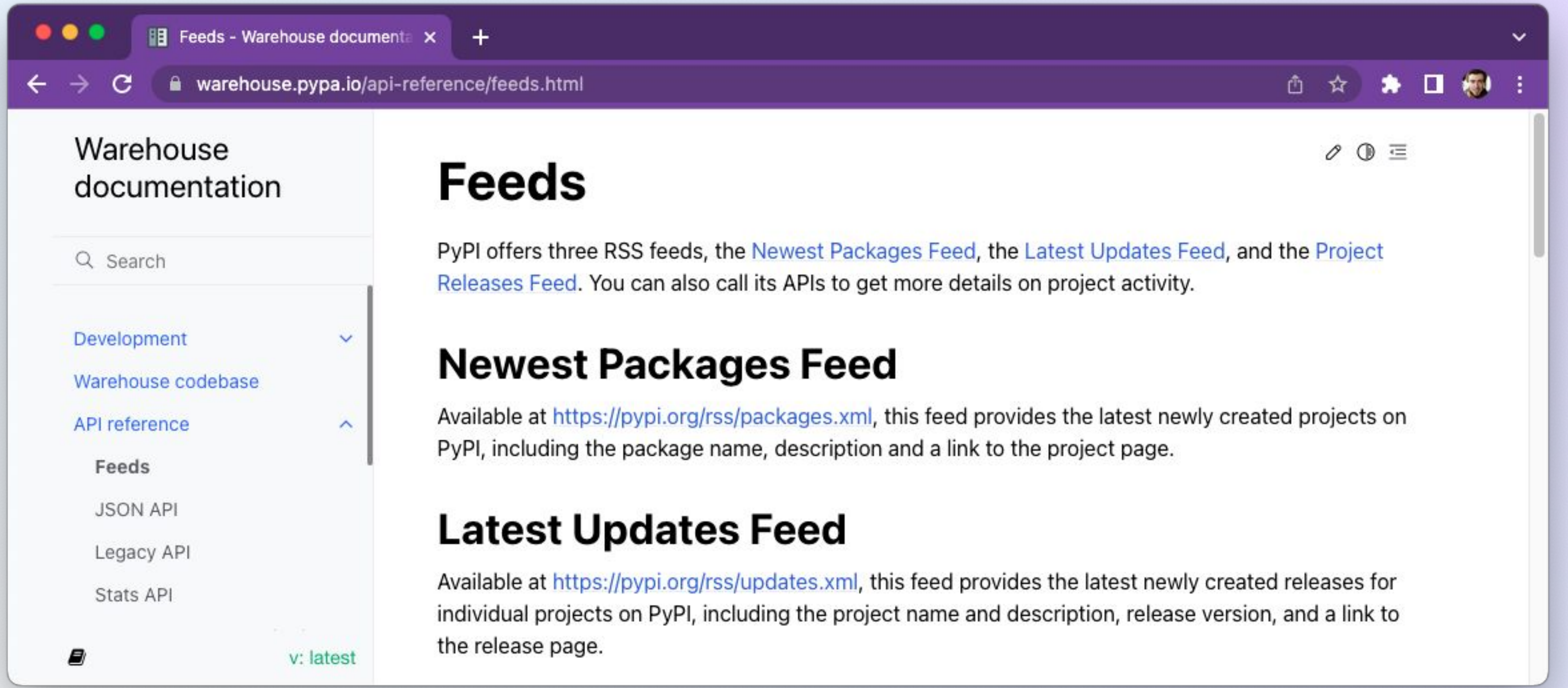
- Open Source static code analysis tool
- Traditionally used to find vulnerabilities
- Provides taint-tracking capabilities
- Transparently packaged within GuardDog

Semgrep rules

```
rules:
 - id: code-execution
   message: This package is executing OS commands in the setup.py file
   patterns:
     - pattern: subprocess.Popen(...)
     - pattern: os.system(...)
   paths:
     include:
       - "*/setup.py"
```

Continuously Scanning PyPI

PyPI RSS feeds



Feeds

PyPI offers three RSS feeds, the Newest Packages Feed, the Latest Updates Feed, and the Project Releases Feed. You can also call its APIs to get more details on project activity.

Newest Packages Feed

Available at https://pypi.org/rss/packages.xml, this feed provides the latest newly created projects on PyPI, including the package name, description and a link to the project page.

Latest Updates Feed

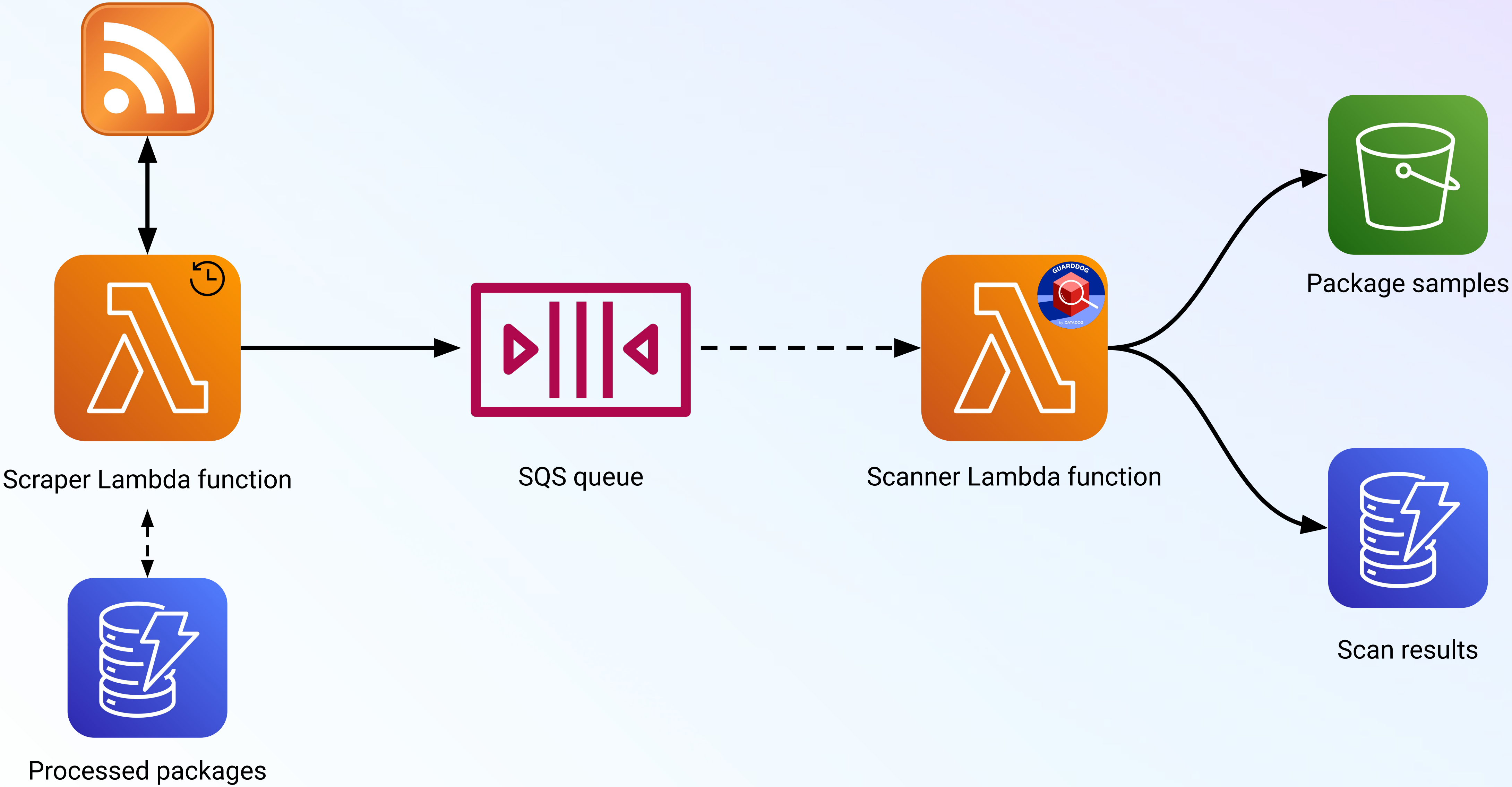
Available at https://pypi.org/rss/updates.xml, this feed provides the latest newly created releases for individual projects on PyPI, including the project name and description, release version, and a link to the release page.

PyPI RSS feeds

```
<rss version="2.0">
<channel>
 <title>PyPI newest packages</title>
 <link>https://pypi.org/</link>
 <description>Newest packages registered at the Python Package Index</description>
 <language>en</language>

 <item>
  <title>astro-toolbox added to PyPI</title>
  <link>https://pypi.org/project/astro-toolbox/</link>
  <guid>https://pypi.org/project/astro-toolbox/</guid>
  <description>Toolbox for observational astronomy</description>
  <pubDate>Wed, 15 Mar 2023 10:44:00 GMT</pubDate>
 </item>
```

Implementation in AWS



Scraper Lambda function

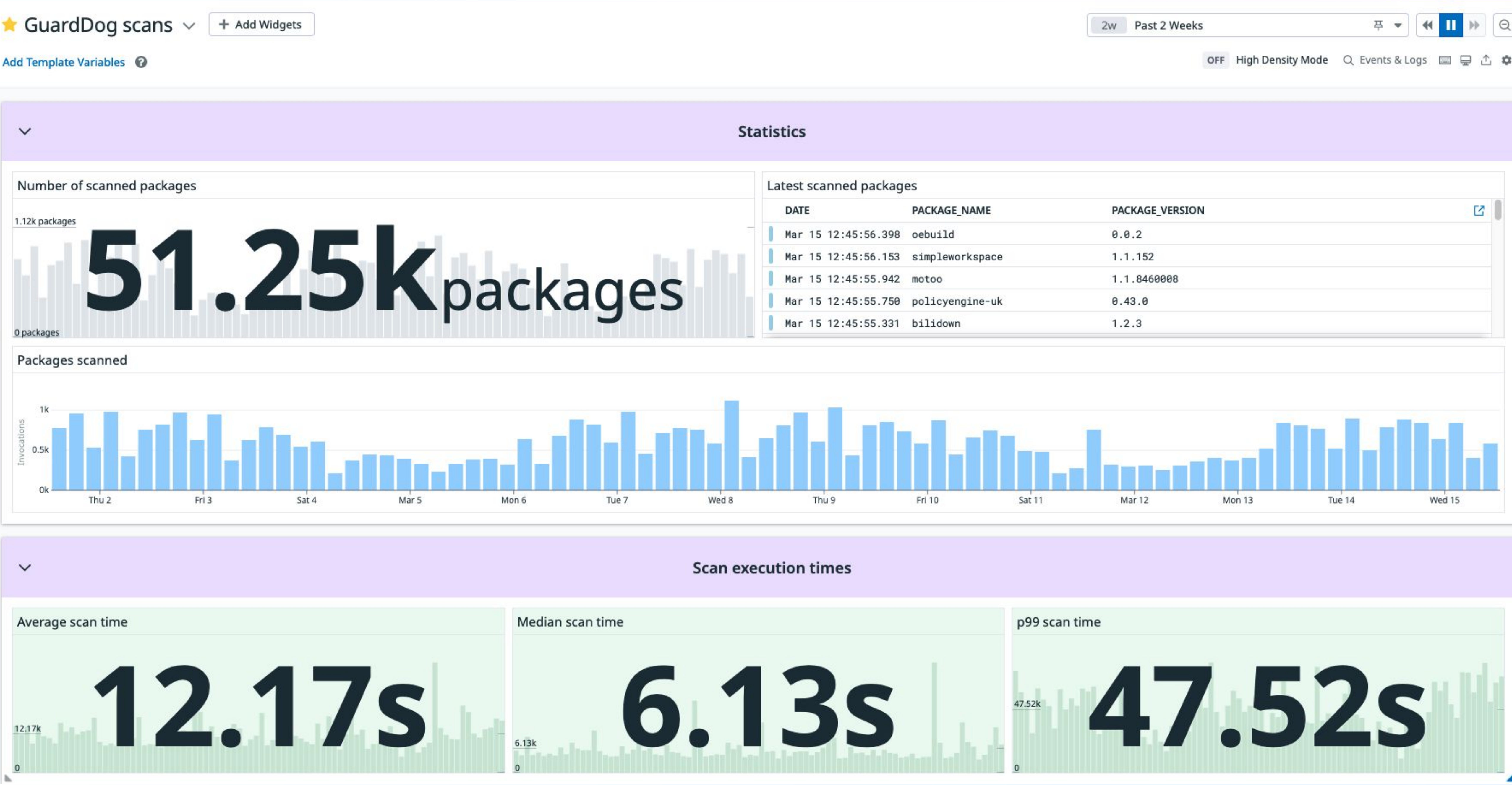
SQS queue

Scanner Lambda function

Package samples

Scan results

Processed packages



Statistics

Number of scanned packages

51.25k packages

Latest scanned packages

DATE	PACKAGE_NAME	PACKAGE_VERSION
Mar 15 12:45:56.398	oebuild	0.0.2
Mar 15 12:45:56.153	simpleworkspace	1.1.152
Mar 15 12:45:55.942	motoo	1.1.8460008
Mar 15 12:45:55.750	policyengine-uk	0.43.0
Mar 15 12:45:55.331	bilidown	1.2.3

Packages scanned

Scan execution times

Average scan time

12.17s

Median scan time

6.13s

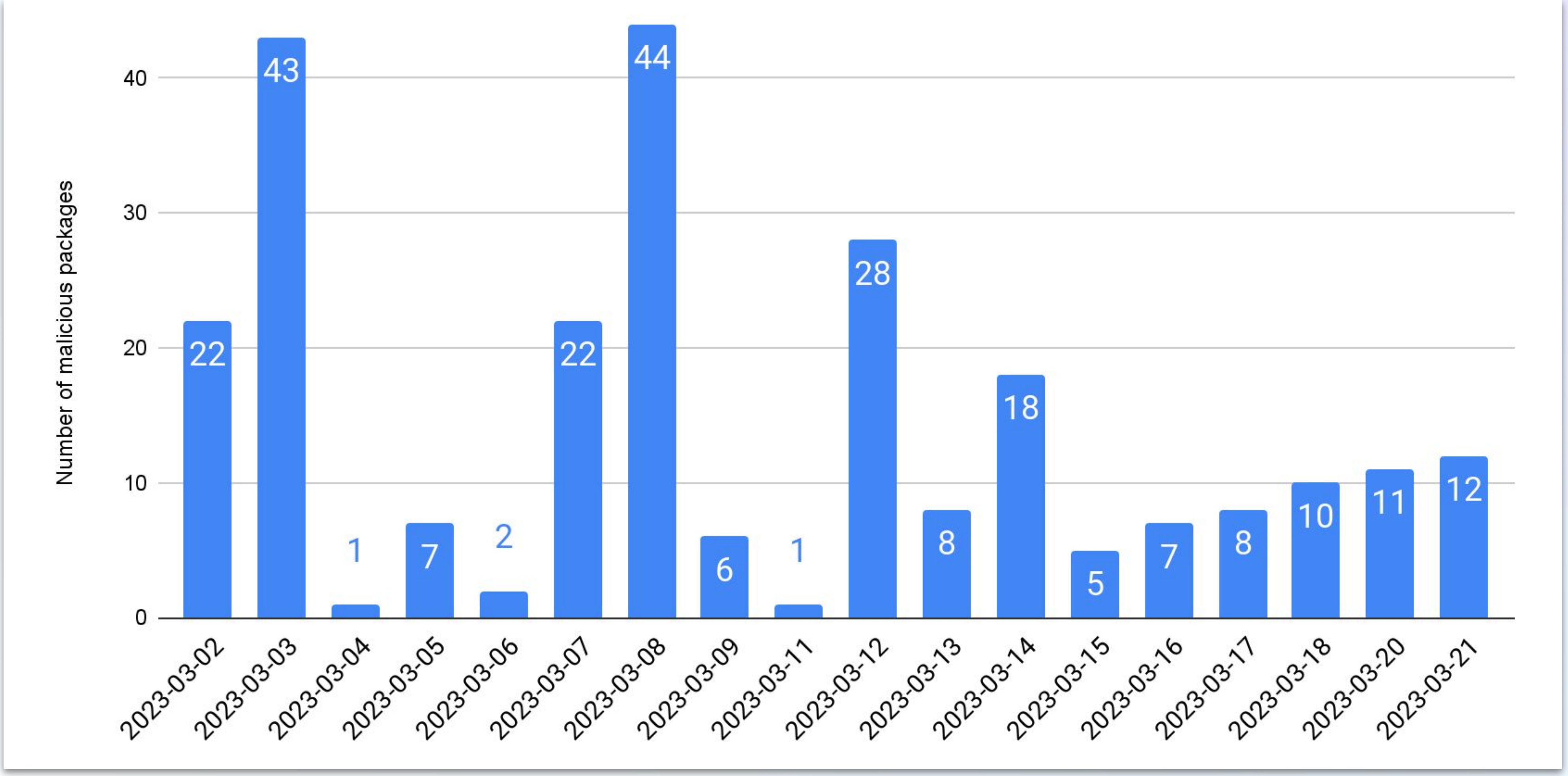
p99 scan time

47.52s

Malicious PyPI Packages in the Wild

Results

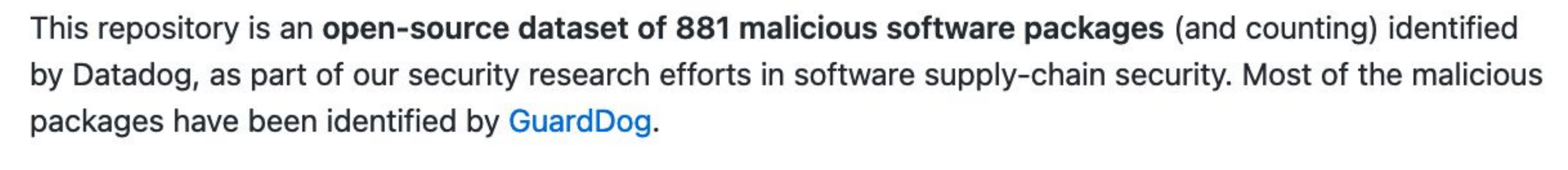
- ➢ Hundreds of malicious packages, continuous stream



Results

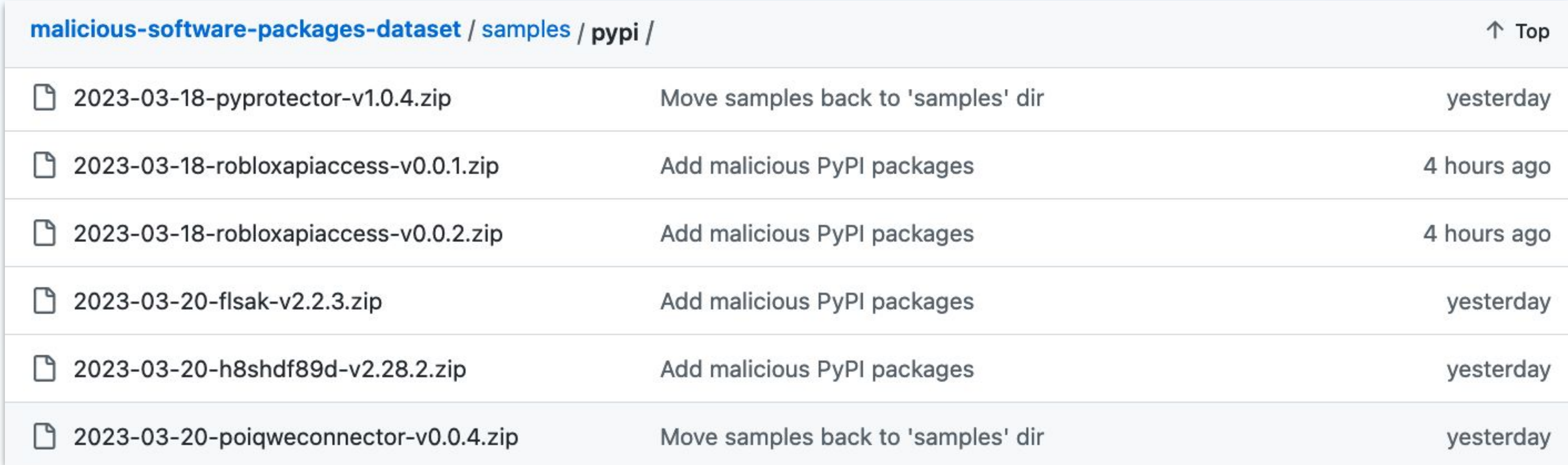
- 3k-5k packages scanned per day
- From script kiddies to elaborated backdoors
- Challenging to analyze due to many similar packages

A new open-source dataset



This repository is an **open-source dataset of 881 malicious software packages** (and counting) identified by Datadog, as part of our security research efforts in software supply-chain security. Most of the malicious packages have been identified by GuardDog.

https://github.com/datadog/malicious-software-packages-dataset



malicious-software-packages-dataset / samples / **pypi** /

2023-03-18-pyprotector-v1.0.4.zip	Move samples back to 'samples' dir	yesterday
2023-03-18-robloxapiaccess-v0.0.1.zip	Add malicious PyPI packages	4 hours ago
2023-03-18-robloxapiaccess-v0.0.2.zip	Add malicious PyPI packages	4 hours ago
2023-03-20-flsak-v2.2.3.zip	Add malicious PyPI packages	yesterday
2023-03-20-h8shdf89d-v2.28.2.zip	Add malicious PyPI packages	yesterday
2023-03-20-poiqweconnector-v0.0.4.zip	Move samples back to 'samples' dir	yesterday

Everyone is invited!

- We would love contributions on GuardDog
  - "good first issues"
  - fix false negatives/positives
  - implement new features!

- Explore the dataset!
  - Help us expand it!
    - Not by creating malware please

Thank you

dtdg.co/guarddog-insomnihack