



Secrets At Sea

Hunting Exposed Code & Container Registries



Gaetan FERRY & Guillaume VALADON

Introduction

\$ whoami



Guillaume

Cybersecurity Researcher

editor-in-chief of the **MISC**
magazine

Scapy maintainer

previously at **Quarkslab**,
ANSSI...



Gaetan

Cybersecurity Researcher

former researcher @**Sonar**

Synacktiv red teamer for 7
years

Public & Private Leaks

Leaking secrets is very easy

hardcoding secrets is easier than handling them safely!
private things will go public, **PoCs will go to production...**
developers leak in personal projects

35%

Private repos leak

5%

For public repos

Closer to production means leakier

secrets are mostly needed in production
production > container > artifacts > source code

Public leaks are the worst but **private leaks are bad too**



Research Key Questions

Are there **publicly accessible registries**?

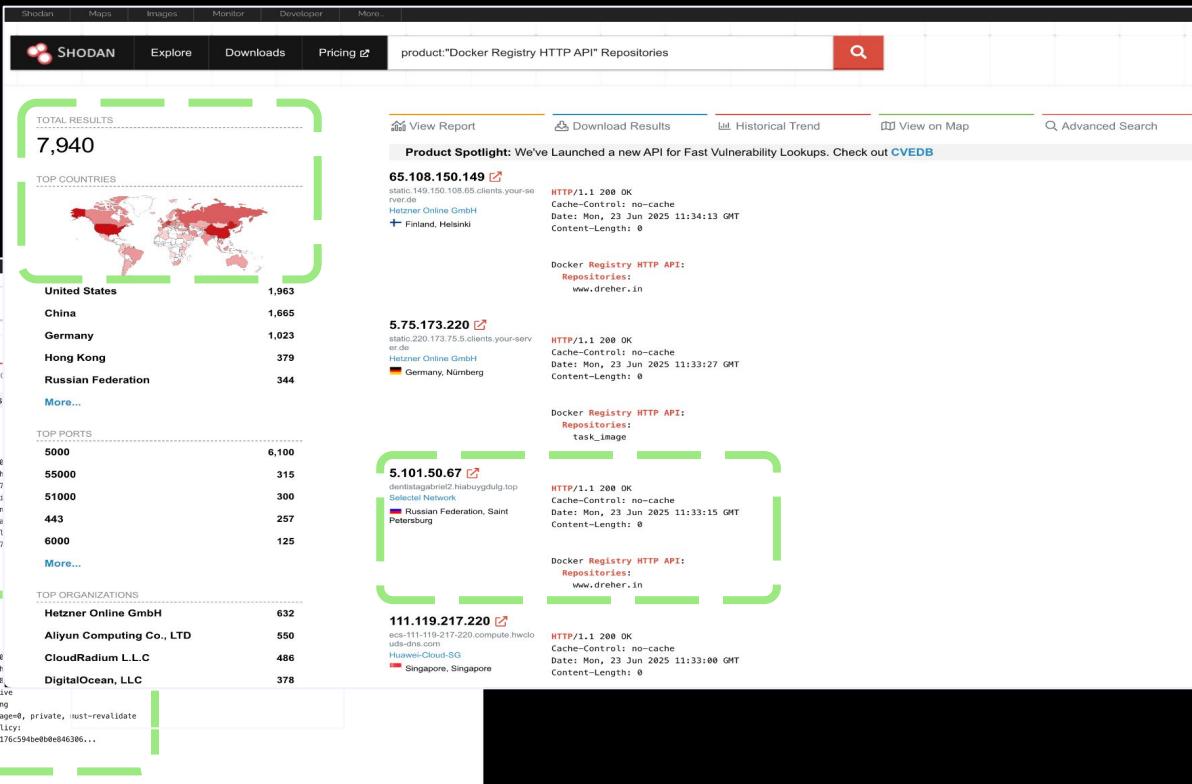
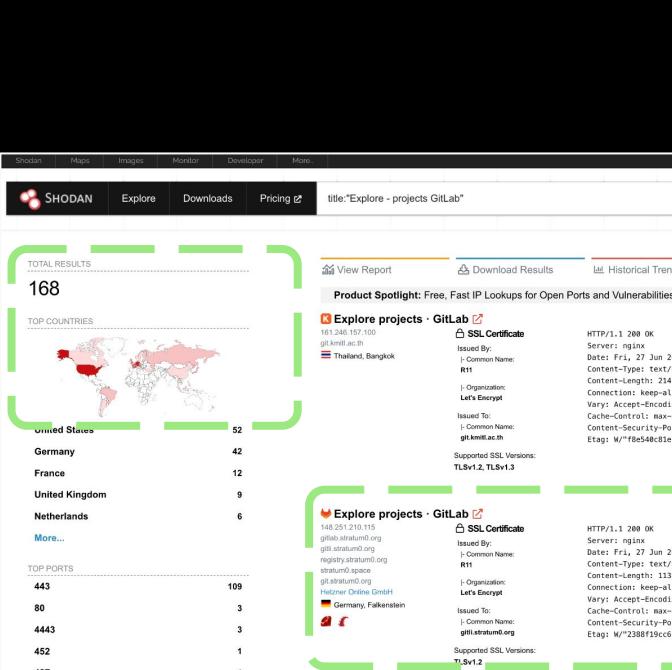
Do they contain **private repositories**?

Do they **leak secrets**?

Who owns them?



Research Short Answer



Research Previous Work - www.dreher.in

The screenshot shows a website with a dark background featuring a photograph of a circuit board on the left. The main navigation bar at the top includes links for Home, Services, Press, Contact, Blog, and a menu icon. A "Follow me on Twitter" button is also present. The central content area contains a blog post titled "Ethical Considerations in Security Research". The post discusses the researcher's approach to ethical dilemmas, mentioning the use of ASCII animations and DNS lookups. Below the post is a terminal window showing a Docker command: `> docker run --rm registry.drehsec.tk/www.dreher.in:latest`. At the bottom of the page, there is a footer section with the text "Internet connection." and a link to "www.dreher.in your friendly IT-Security helpers".

Ethical Considerations in Security Research

During my research, I grappled with ethical dilemmas. Backdooring available images would provide tangible proof of vulnerability but would be unethical and illegal. Instead, I opted to upload my own image to vulnerable targets, which displayed an ASCII animation and performed a DNS lookup to a canary domain when executed. This approach, though controversial, aimed to raise awareness without causing harm.

If you want to enjoy the experience yourself, here you go:

```
> docker run --rm registry.drehsec.tk/www.dreher.in:latest
```

Proudly presented by 
www.dreher.in your friendly IT-Security helpers

Internet connection.

02

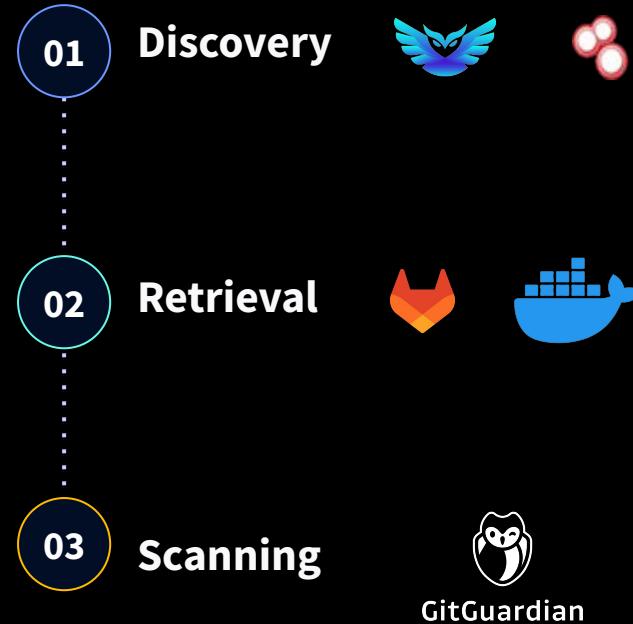
Methodology

Discovering exposed registries ourselves.

From One to Three

Three simple and common steps.

Looking for vulnerabilities at scale is often splitted into three distinct phases.



Discovery Phase Examples



Product queries

Docker Registry

GitLab Self-Managed



Specific HTTP fingerprinting

X-gitlab-meta HTTP header

well-known HTTP titles



SHODAN



Keywords

docker

gitlab

registry

harbor



MERKLEMAP

Retrieval Phase Examples

Various TCP ports

443, **80**, 8099, 8080...

GET /explore

list public repositories

might be authn !

GET /api/v4/projects

more reliable

Various TCP ports

443, **80**, 5000...

GET /v2/_catalog

list repositories

git clone



docker pull



Docker Registry Over Plain HTTP

Not supported by common tools

must build a custom tool based on the Registry API.

```
$ docker pull 192.0.2.28:5000/pts2025:latest
Error response from daemon: Get "https://192.0.2.28:5000/v2/": http: server gave
HTTP response to HTTPS client
```

Inspecting images

<https://github.com/containers/skopeo>

Registry Endpoints

`/v2/<repository>/tags/list`

`/v2/<repository>/manifest/<tag>`

`/v2/<repository>/blobs/<blob>`

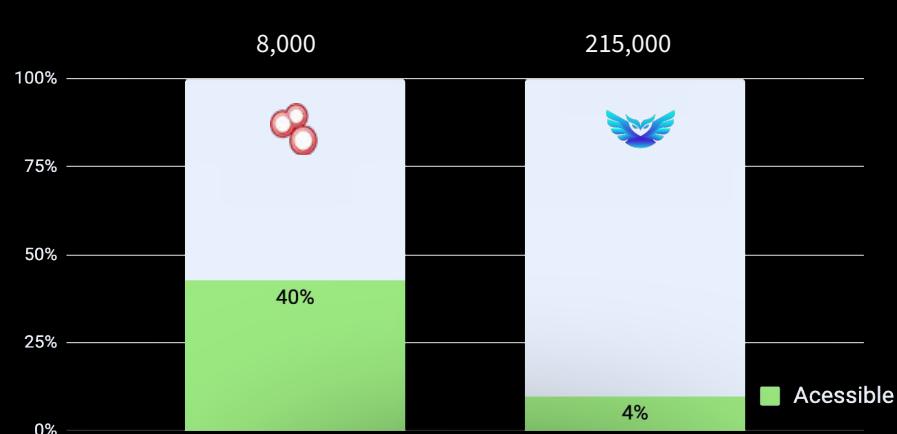
Methodology

Discovery Phase Results



Key Findings

39k registries
11k public
67k git repositories
~2 TB



Key Findings

4500 registries
75k repositories
225k Docker images

04

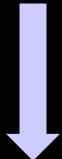
Secrets Analysis

Charts & Numbers.

Secrets Exposures

Instances

3,983 leak secrets - 10 %



Repositories

7,871 leak secrets - 12%

Instances

2,269 leak secrets - 60%



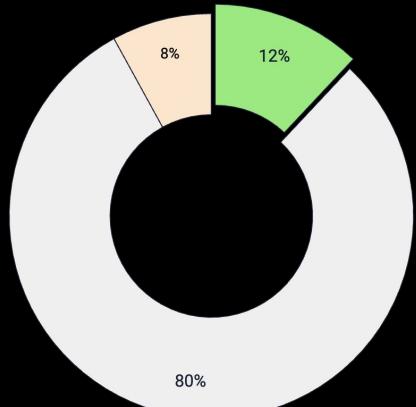
Images

41,200 leak secrets - 18%



Secrets Validity

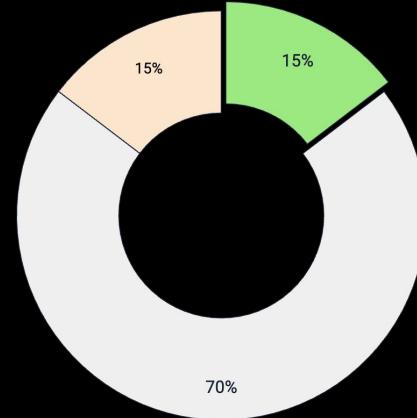
- Valid
- Cannot Check
- Not Valid



Total secrets: 57,000
Specifics: 20,000
Generics: 37,000



- Valid
- Cannot Check
- Not Valid



Total secrets: 23,000
Specifics: 9,000
Generics: 14,000

Secret Types Breakdown

VCS



930

2% valid



130

40% valid

Cloud



1200

47% valid



800

60% valid

Data Storage



3100

4% valid

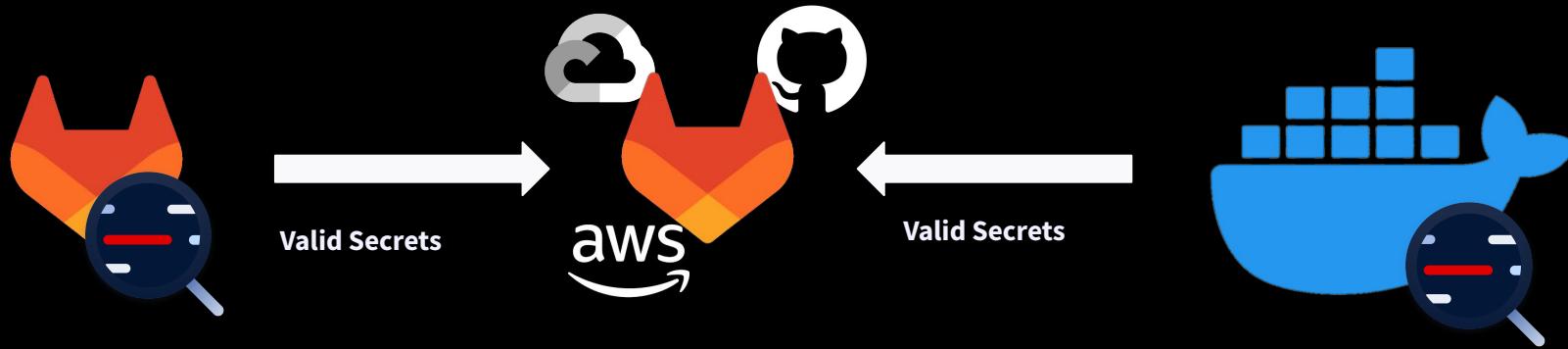


1000

32% valid

Docker: less secrets, more valids - It's **closer to production!**

Secrets Russian Dolls



Publicly exposed leaks lead to private repositories, code, and
other secrets being exposed.



Name Dropping



Well, not today.

Notable Industries Leaking Repositories

Public entities (Govs)

Consulting co.

Software development

Software development

Video game industry

Universities

Legal services



Beyond Secret Leaks - Private Code

Lots of hints the exposed data should sometimes not be public

80% failed check DB creds -> DB hosts are internal!

Confidentiality mentions

Might include juicy information for a RedTeam

67k GitLab repos revealed 300k+ email addresses (2k .gov)
(Open?) Source code to look for vulnerabilities in



But also: S3 bucket names, SaaS application URL, domain names,...

*This software is the confidential
and proprietary information of
REDACTED.*

Takeaways

transient leaks

18% of the Docker registries not accessible anymore
most secrets are still valid

GitLab servers badly protected by a frontend

GitLab API's web_url provides a host, the host is IP whitelisted
But the IP we found is different **and not filtered!?** 😱

hardcoding secrets in private repositories is **also a bad practice**

private repositories might turn public,
or are indeed public

Takeaways

Self hosting **to better**



...

(or else? DockerHub quotas?)

acy...

control security and privacy.

Thank you

Question Time 🔥



Gaetan FERRY & Guillaume VALADON

GitGuardian

Docker Images Retrieval Methodology

Based on GitGuardian R&D

latest 5 tags per repository

ADD & COPY layers up to 45MB

RUN layers up to 5MB

deduplication based on layers names

Informed trade-off to **limit the amount of data** to be downloaded and scanned.