Quarks Flow: your security analysis hub

Marion Guthmuller, *Engineering Manager, Quarks Flow*

Fred Raynal, *CEO, Entropy Generator*

Quarkslab

INITIAL PROBLEM TO SOLVE

How to make sure that files received, sent or transiting through my organization are safe without exposing any data?

KEY USES CASES

- **Malware hunting**: Manual file analysis
- **Self-service malware detection**: users submit their files
- **Automate file analysis**: Security & SecDevOps tools submit files via API
- **Incident Response**: IR teams analyze files in bulk and at scale after a breach
- …

INITIAL PROBLEM SOLVED

How to manage the analyses?
Several analyses at the same place?
Distributed system vs centralized?

Analyzing files

=

How fast can you extract the files? How much space to store them?
For how long?

1) **Obtain** the files

2) **Send** the files to whom will analyze it

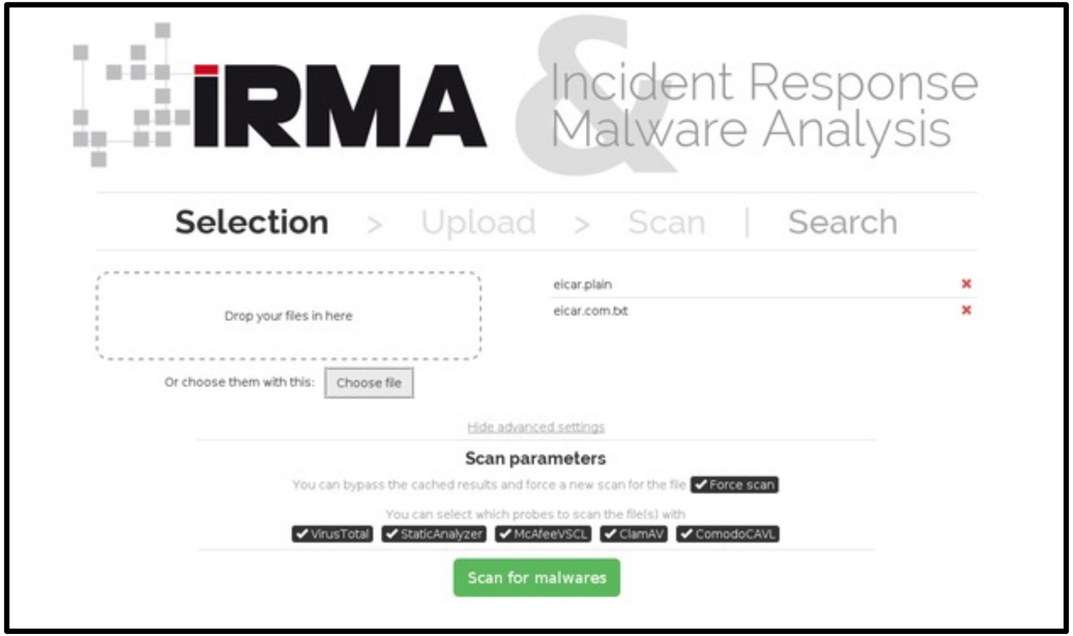
3) Do all the required **analyses**

4) Gather and store the **reports**

How to send the files?
To whom?
In what order?
« Interesting » probes?

How should the results be searchable?
Post-processing?
How to display the results?

DEMO #1: IRMA



Or choose them with this:

eicar.plain
eicar.com.txt

Scan parameters

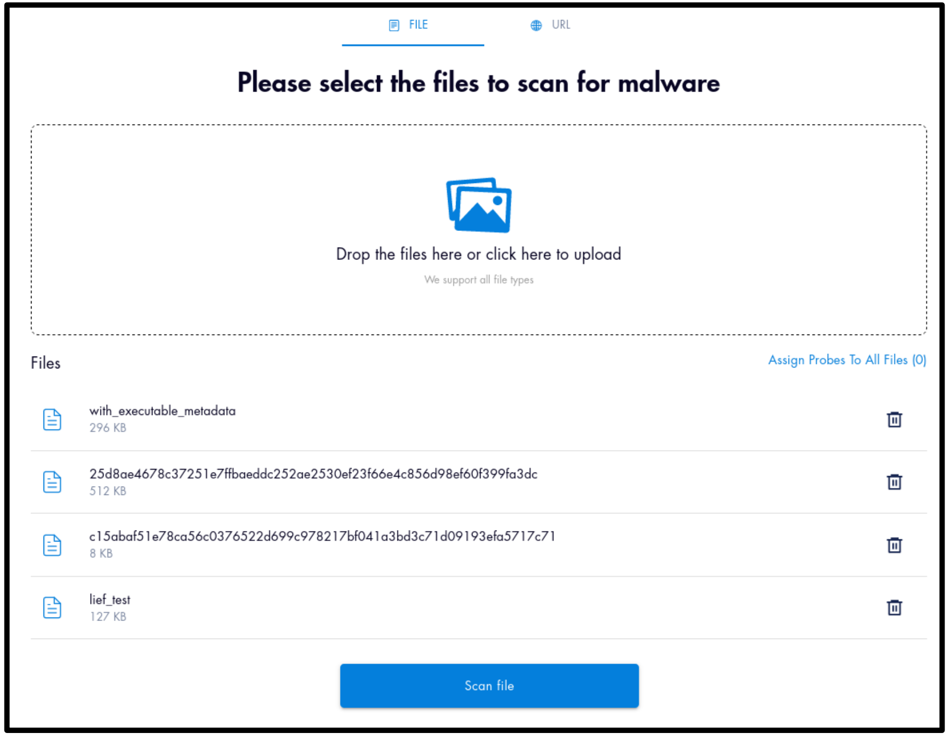
You can bypass the cached results and force a new scan for the file ✓ Force scan

You can select which probes to scan the file(s) with

✓ VirusTotal ✓ StaticAnalyzer ✓ McAfeeVSCL ✓ ClamAV ✓ ComodoCAVL

IRMA v1 (2015)

DEMO #2: QFLOW



Please select the files to scan for malware

Drop the files here or click here to upload

We support all file types

Files

Assign Probes To All Files (0)

with_executable_metadata
296 KB

25d8ae4678c37251e7ffbaeddc252ae2530ef23f66e4c856d98ef60f399fa3dc
512 KB

c15abaf51e78ca56c0376522d699c978217bf041a3bd3c71d09193efa5717c71
8 KB

lief_test
127 KB

QFlow v1 (2022)

FROM IRMA TO QFLOW

BABY IRMA: VIRTUALIZATION AS SUPPORT

Goal: scaling probes

Pros

- Easily "manually" clone probes to scale
- Fine tuning for performances (I/O, RAM,…)

Cons

- Hypervisors are dependent on the host OS and hardware
- No "on demand" configuration

Bash scripting powered

(Install time: 1 day – a lot of manual configuration in scripts)

TODDLER IRMA (2015): DEVOPS ON THE RISE

Goal: Improving installation

Pros

- Build and configure vm automatically

Cons

- Hypervisors are dependent on the host OS and hardware
- No “on demand” configuration
- More packages dependencies

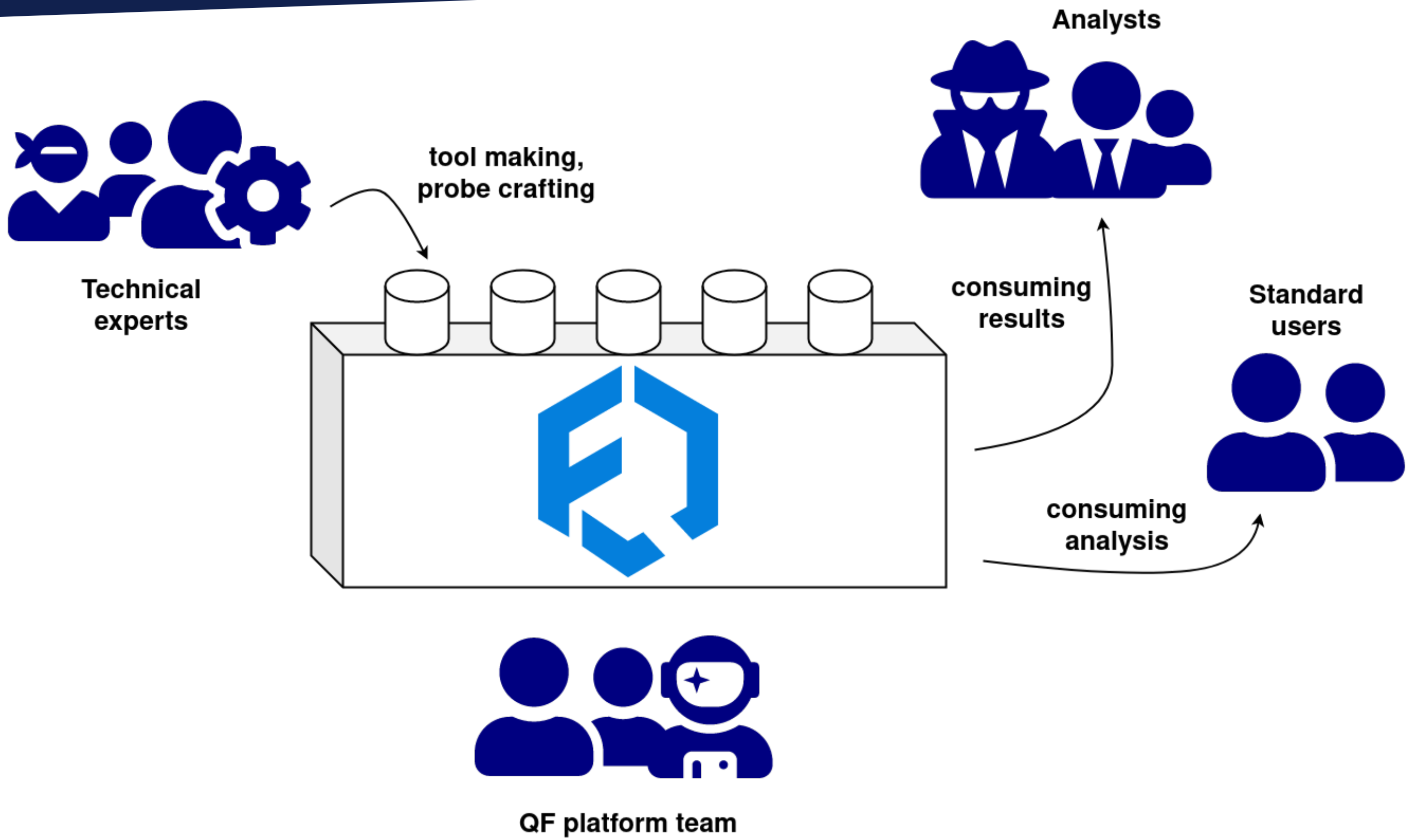
Ansible to deploy, Vagrant to manage (VM)

(Install time: still 1 day – less manual configuration, more external dependencies)

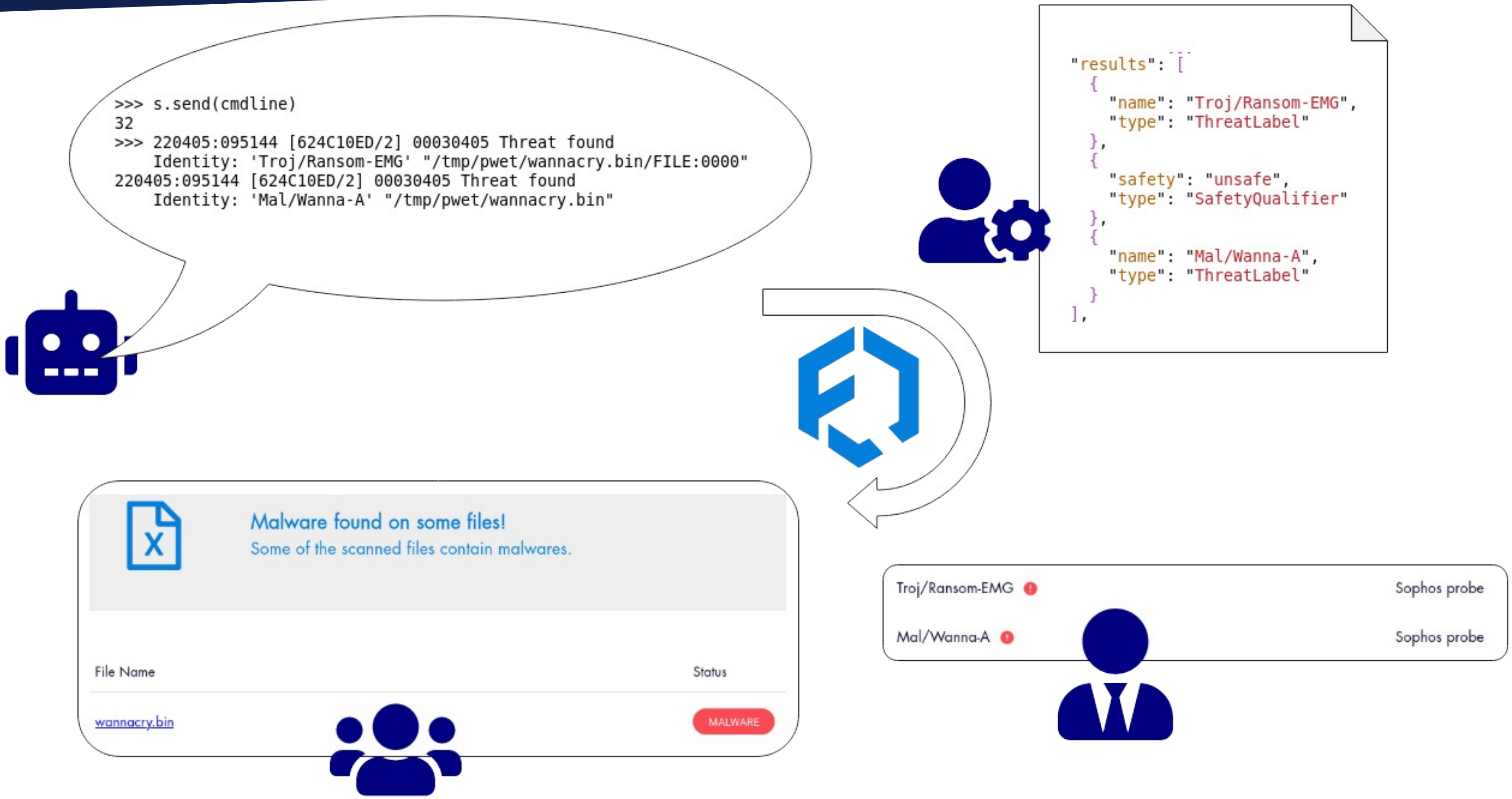
Product

Build, ship, run a platform: the actors

DIFFERENT ACTORS



USERS: DIFFERENT EXPECTATIONS



```
>>> s.send(cmdline)
32
>>> 220405:095144 [624C10ED/2] 00030405 Threat found
    Identity: 'Troj/Ransom-EMG' "/tmp/pwet/wannacry.bin/FILE:0000"
220405:095144 [624C10ED/2] 00030405 Threat found
    Identity: 'Mal/Wanna-A' "/tmp/pwet/wannacry.bin"
```

```
                 ...
"results": [
  {
    "name": "Troj/Ransom-EMG",
    "type": "ThreatLabel"
  },
  {
    "safety": "unsafe",
    "type": "SafetyQualifier"
  },
  {
    "name": "Mal/Wanna-A",
    "type": "ThreatLabel"
  }
],
```

Malware found on some files!

Some of the scanned files contain malwares.

File Name	Status
wannacry.bin	MALWARE

Troj/Ransom-EMG ●	Sophos probe
Mal/Wanna-A ●	Sophos probe

BUILD A PLATFORM: WHO?

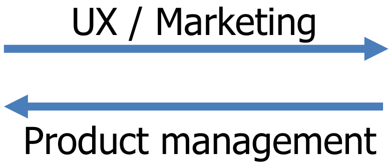
Building a platform: more than a tech challenge!!

Builders

- Developers: code the backend, frontend…
- Job: the platform needs people with experience in security
- Infra: operate the run

Users

- Define personas for your users
- Tests both the platform with people matching your personnas

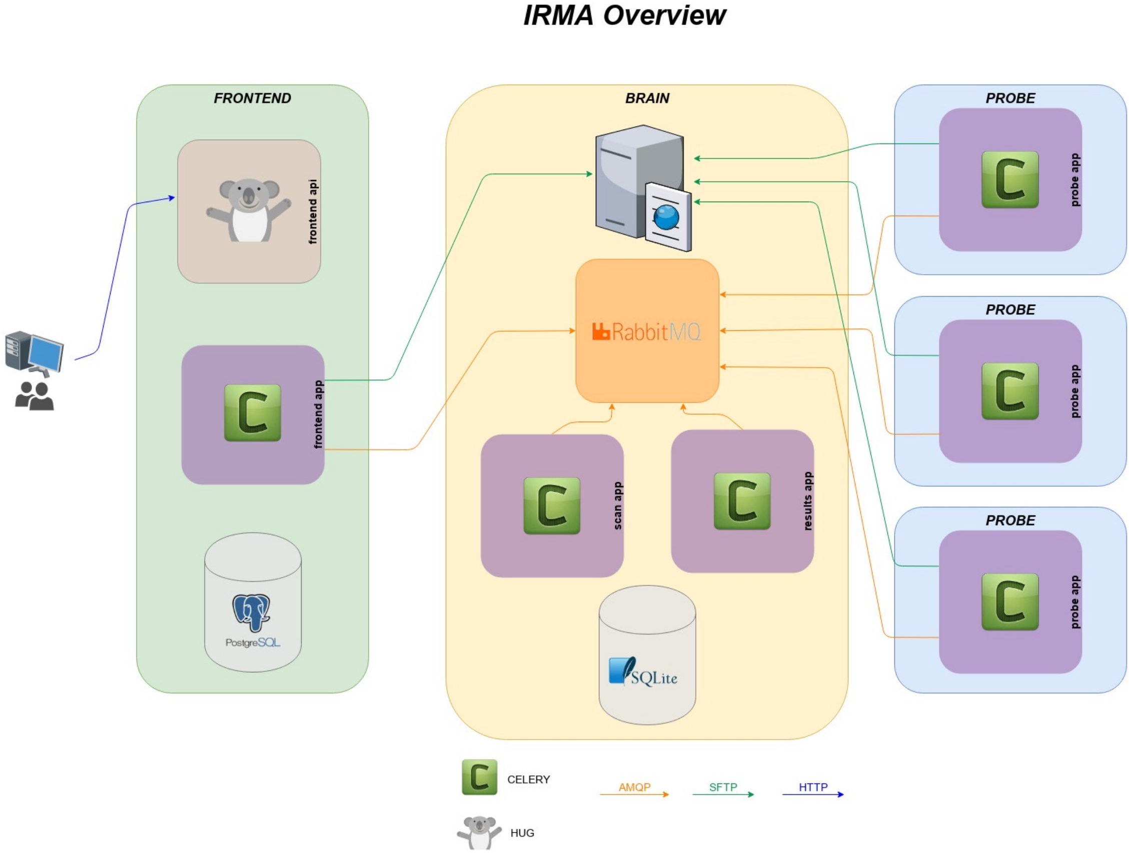


UX / Marketing

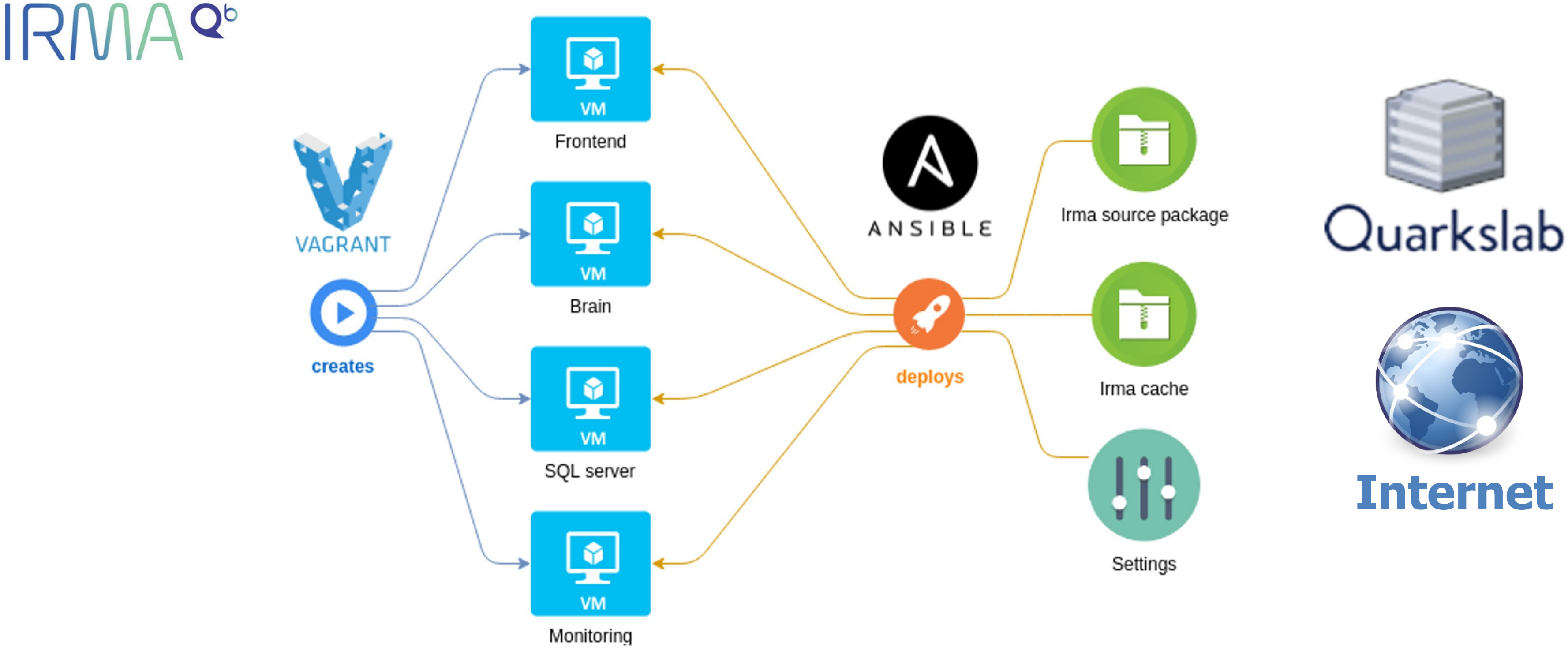
Product management

Build and Ship vs. Build or Ship

IRMA ARCHITECTURE



IRMA INSTALLATION



ADDING A PROBE IN IRMA

1. Create new VM
2. Relaunch complete installation

Install == update

CHALLENGES ADDRESSED BY QUARKS FLOW

Scalability

- VMs (probes) can be cloned depending the number of files to analyze
- Disks must be fast, enough RAM to avoid swapping
- System configuration on demand not easy

Extensibility

- Adding a probe to IRMA requires to create a VM and relaunch the complete installation

Deployment flexibility

- VMs virtualize at hardware level, hypervisor required, full OS installation
- Installation done « on site » fetching pieces from our servers AND the Internet

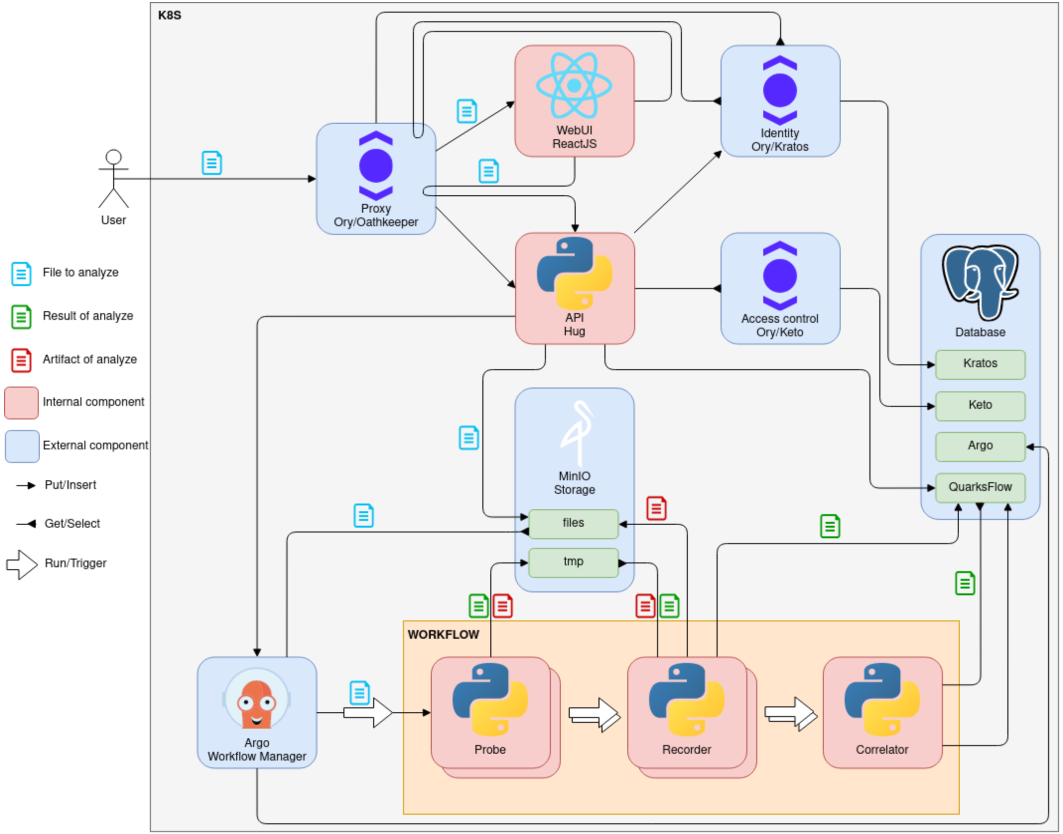
Observability

- Single VM for the core part with several services

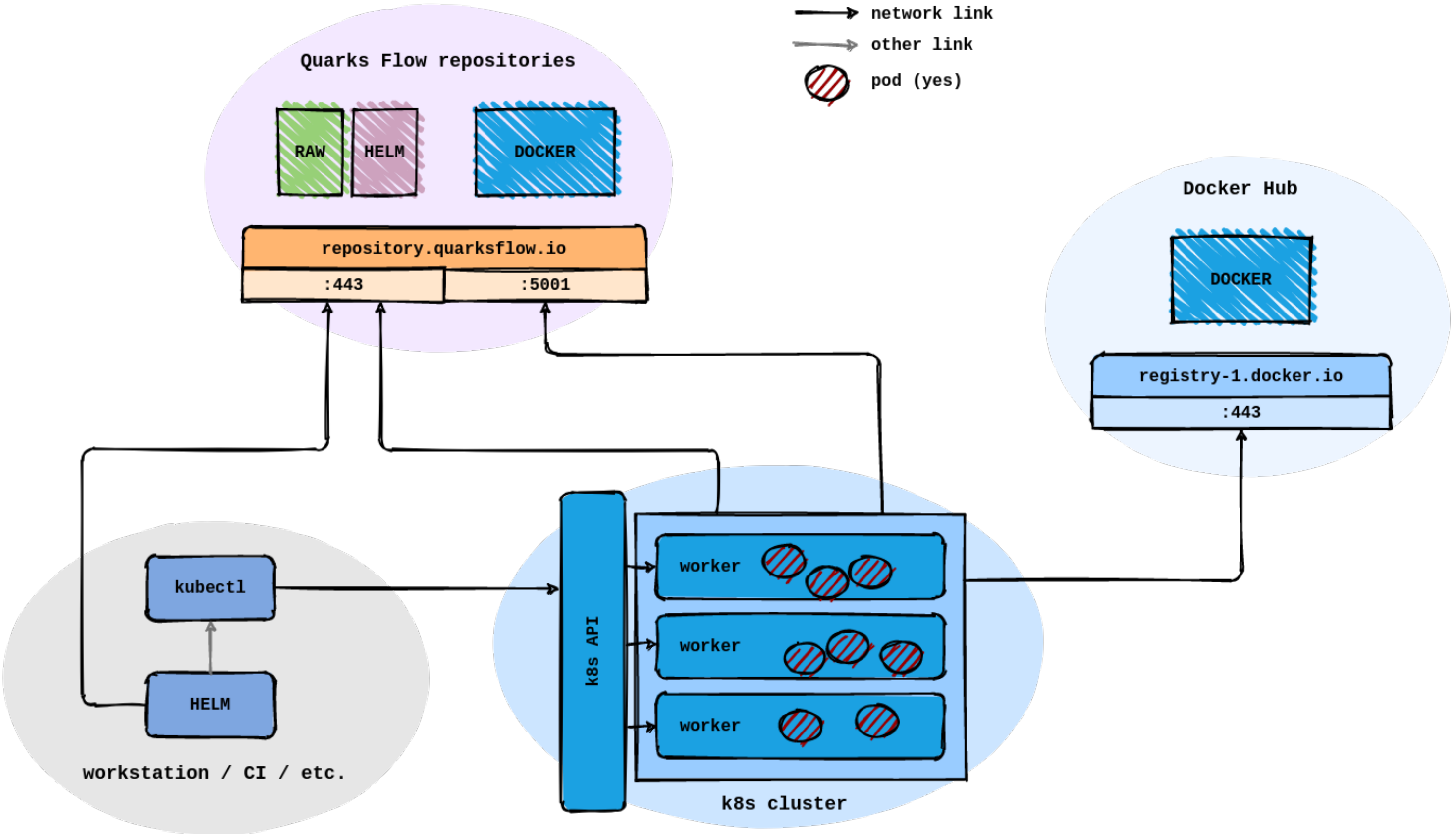
QUARKS FLOW IS THE NEW IRMA

- Started in 2020
- A complete rethinking of IRMA: new architecture, new technologies, new interface
- Based on **Docker** and **Kubernetes**

QUARKS FLOW ARCHITECTURE

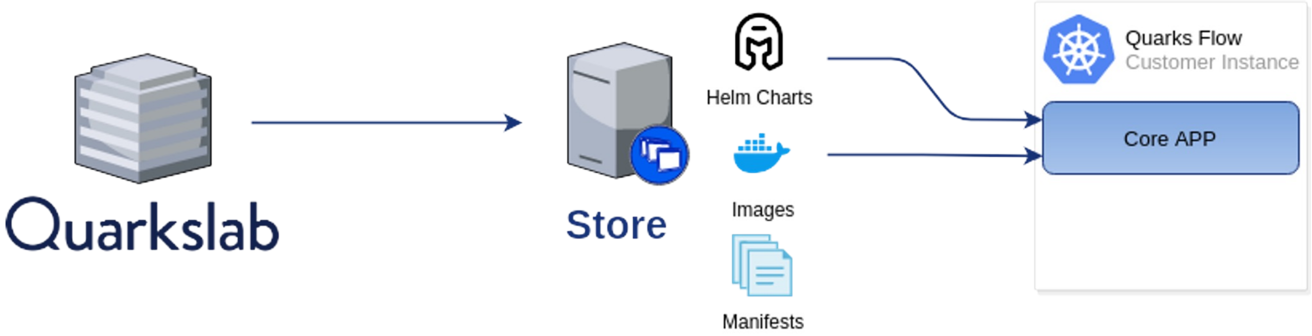


QUARKS FLOW SHIP



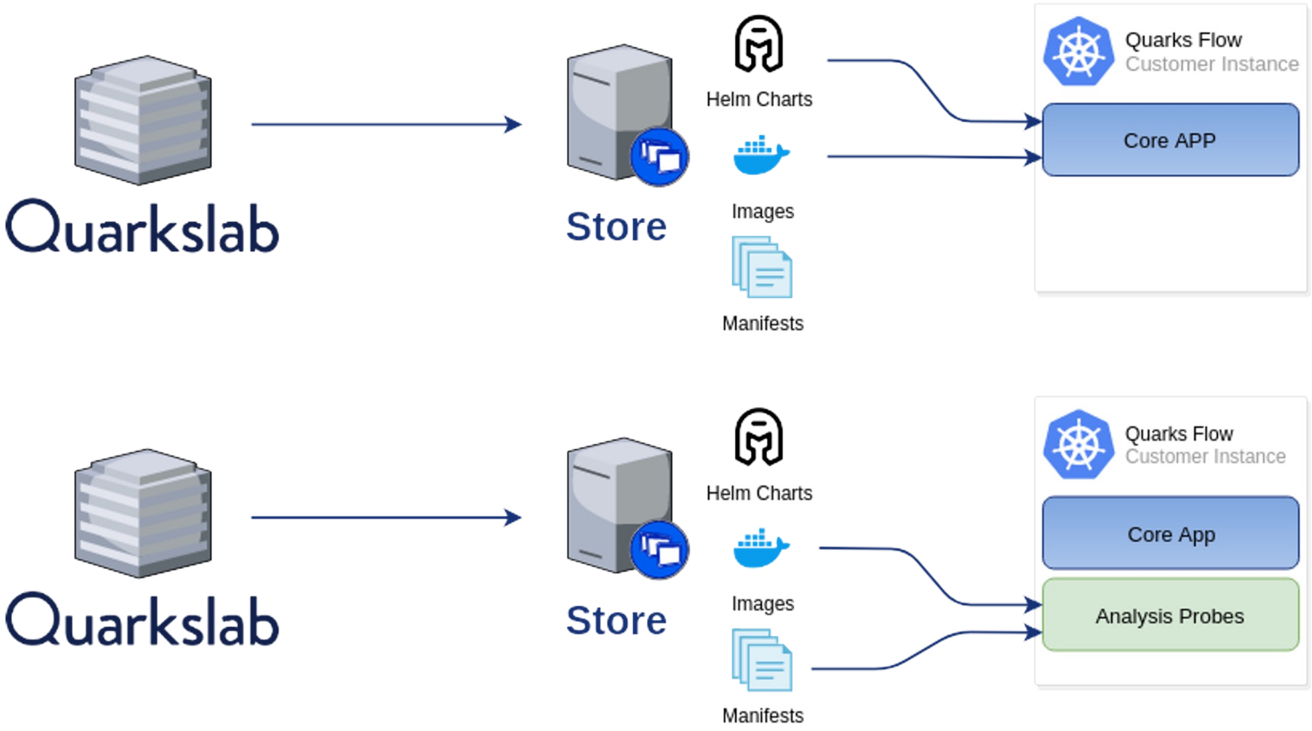
PACKAGING: INSTALLATION THROUGH HELM

Quarks Flow



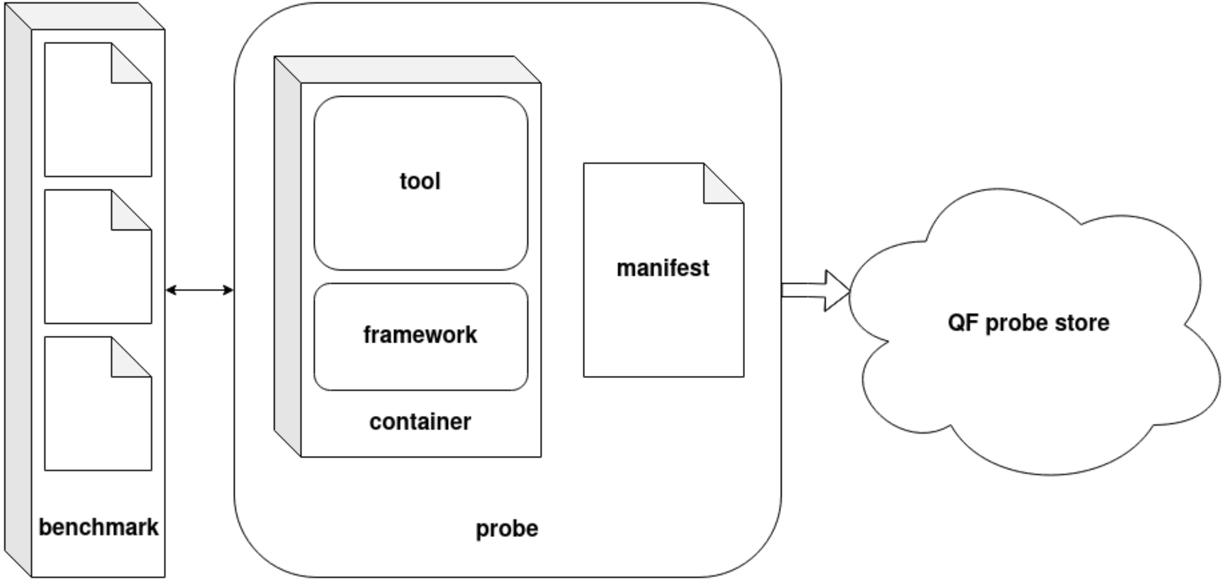
All CORE & PROBES components
are installed from our server

PACKAGING: INCREASE DETECTION WITH NEW PROBES



DRAW ME A PROBE

Quarks Flow



CHALLENGES ADDRESSED BY QUARKS FLOW

Scalability

- **Resources optimization** with Containers and k8s
- k8s **Auto-scaling**
- Dedicated nodes according to services

Extensibility

- Probes catalog
- Dual catalogs (Qb vs. Customer)

Deployment flexibility

- Packaging with **helm charts**
- Automated deployment **on-premise, airgap mode, cloud**

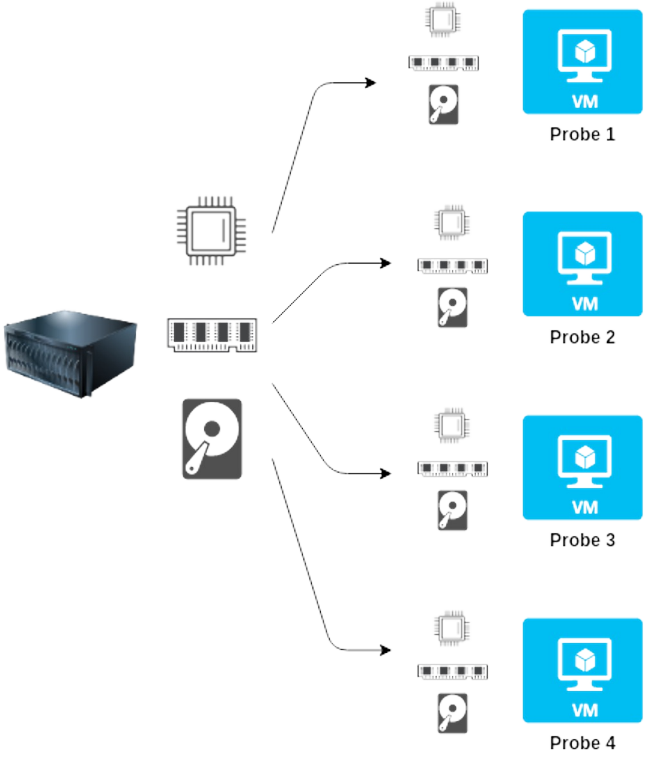
Observability

- **Micro-services**
- k8s tools + external apps such as Jaeger, Fluentd, Prometheus, Grafana or the ELK stack

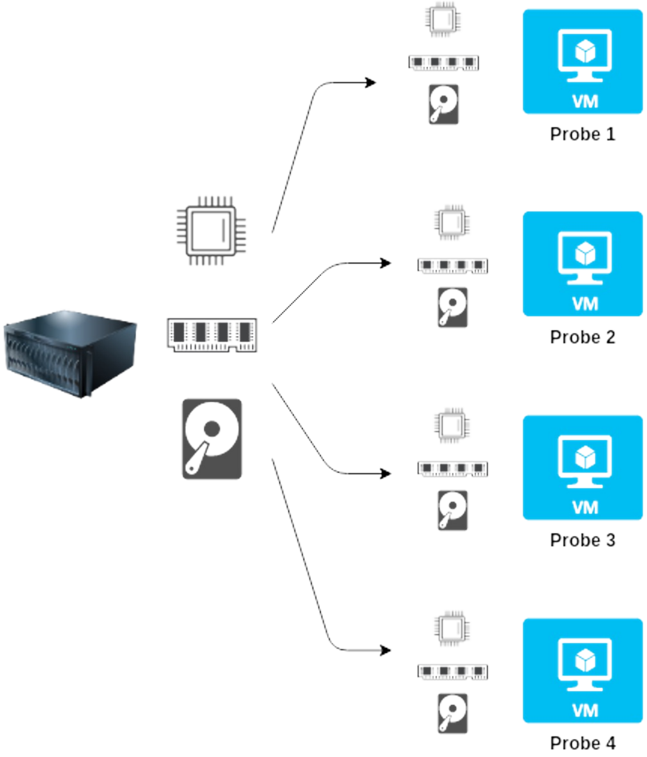
Run (and Scale)

SCALING WITH VMS

IRMA



SCALING WITH VMS



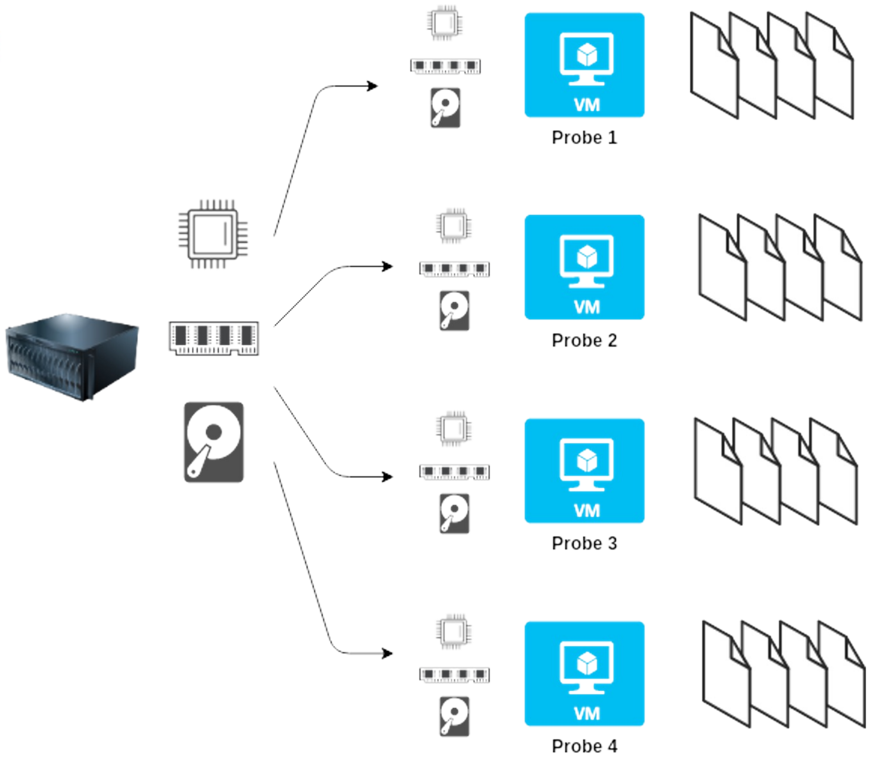
Process 1 file every **4 seconds**

Process 1 file every **2 seconds**

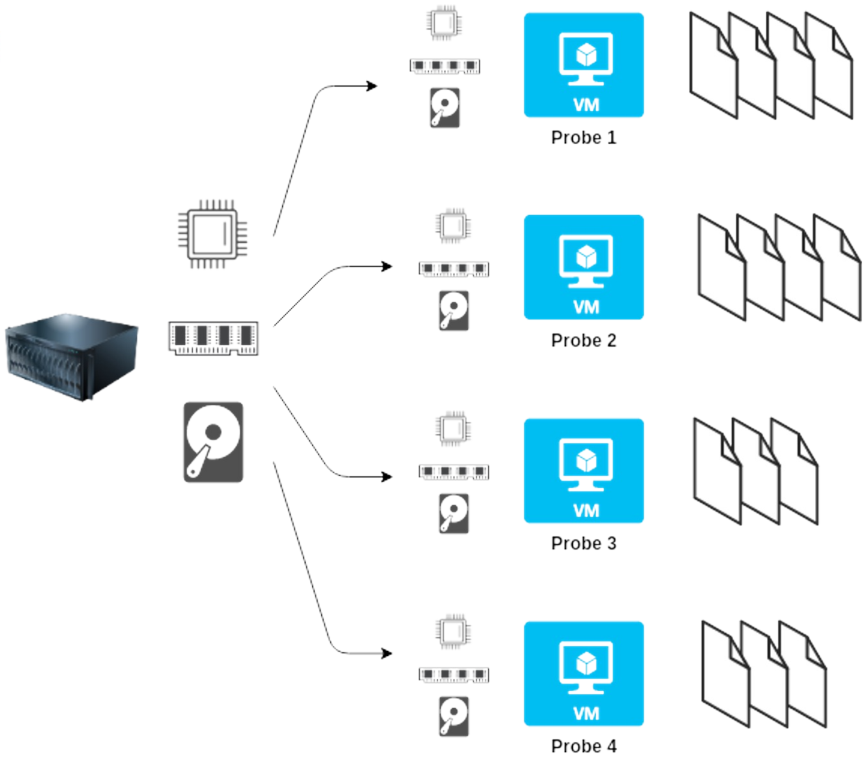
Process 1 file per **second**

Process 1 file per **second**

SCALING WITH VMS (T=0)

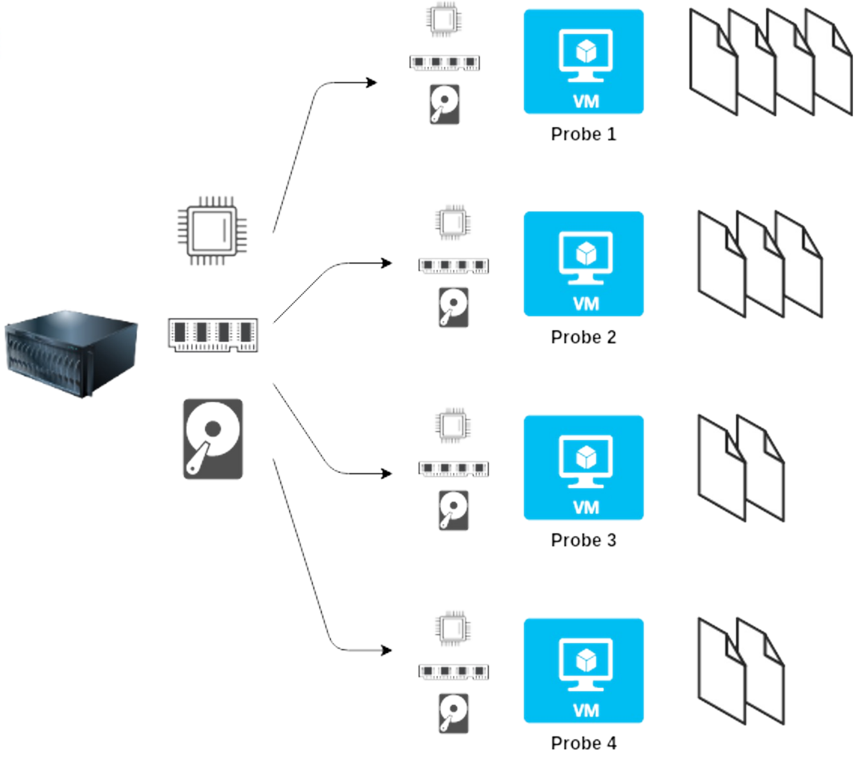


SCALING WITH VMS (T=1)

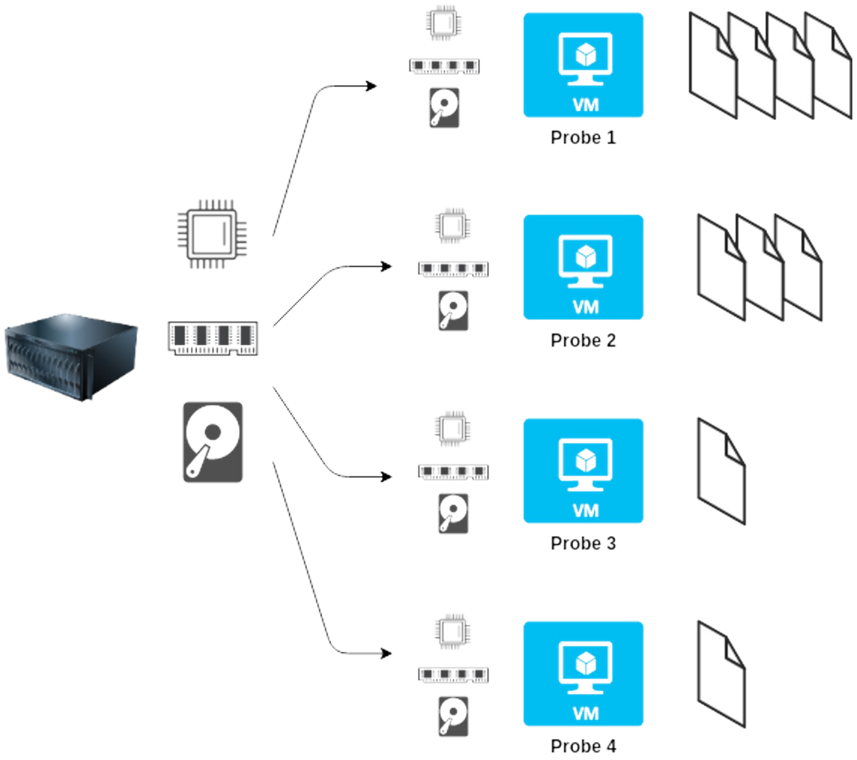


SCALING WITH VMS (T=2)

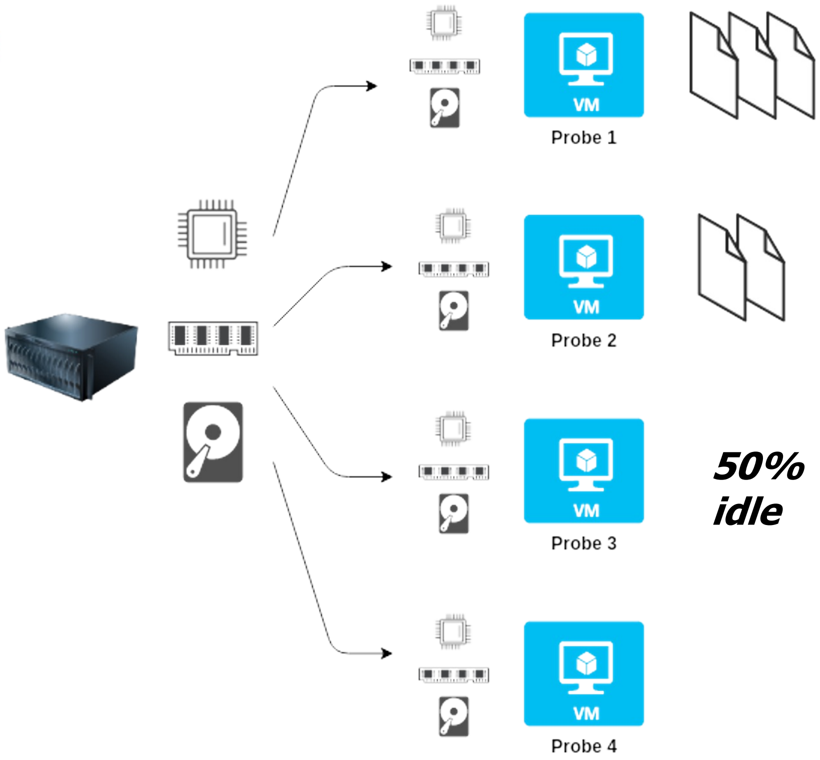
IRMA



SCALING WITH VMS (T=3)

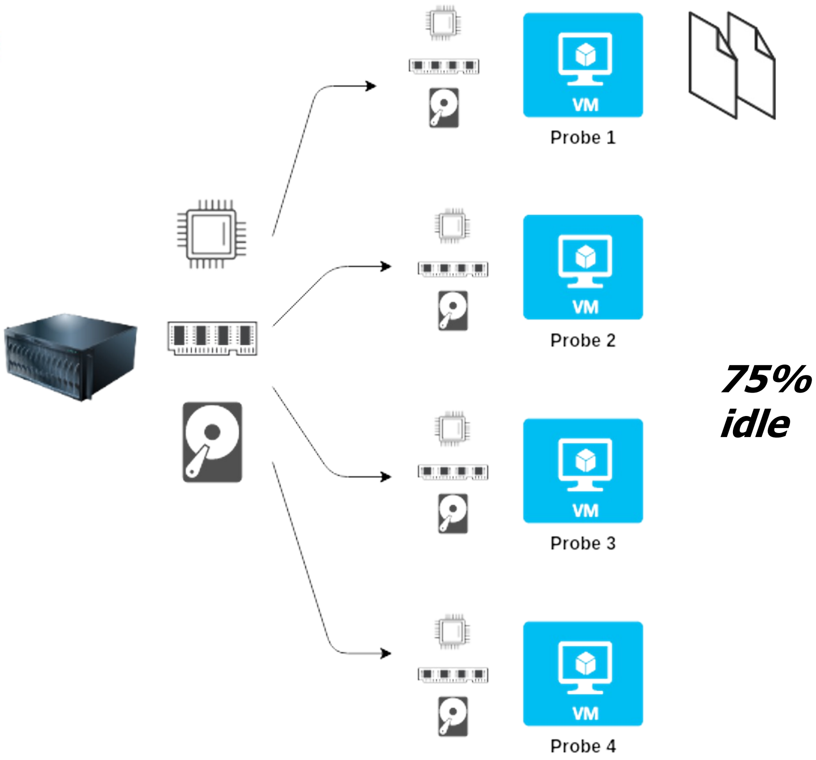


SCALING WITH VMS (T=4)



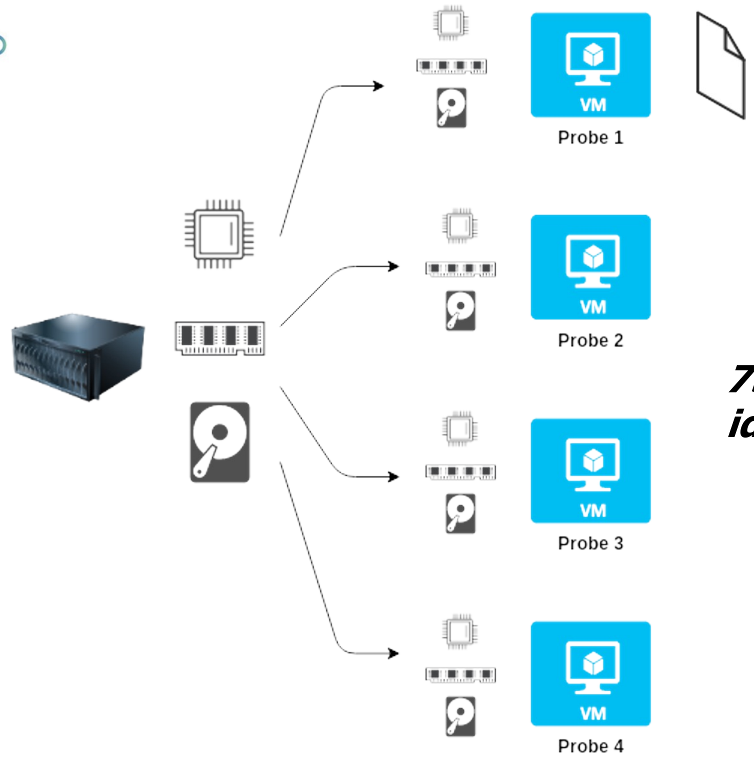
50% idle

SCALING WITH VMS (T=8)



75% idle

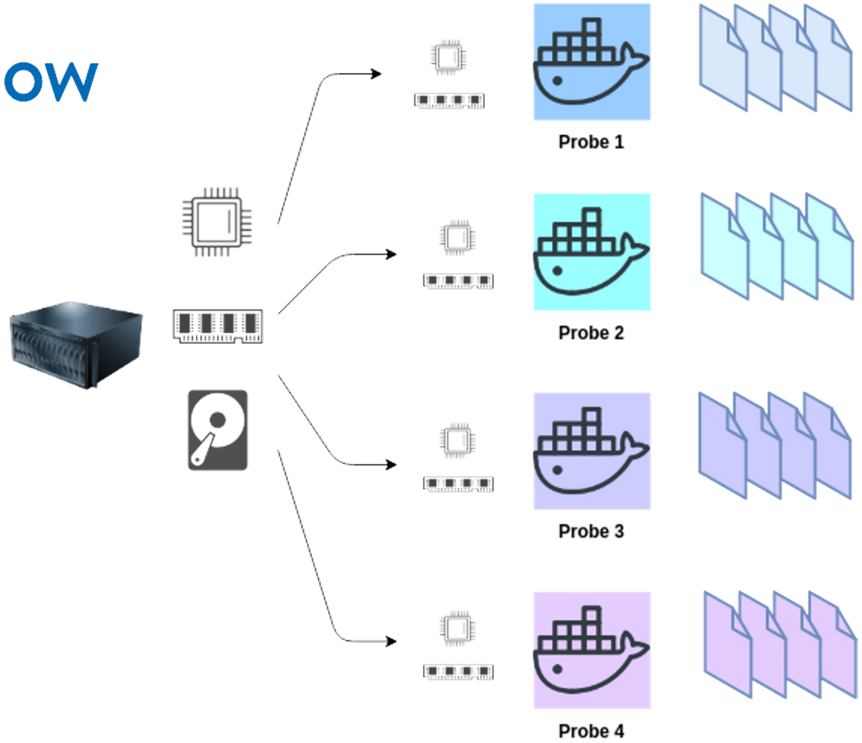
SCALING WITH VMS (T=12)



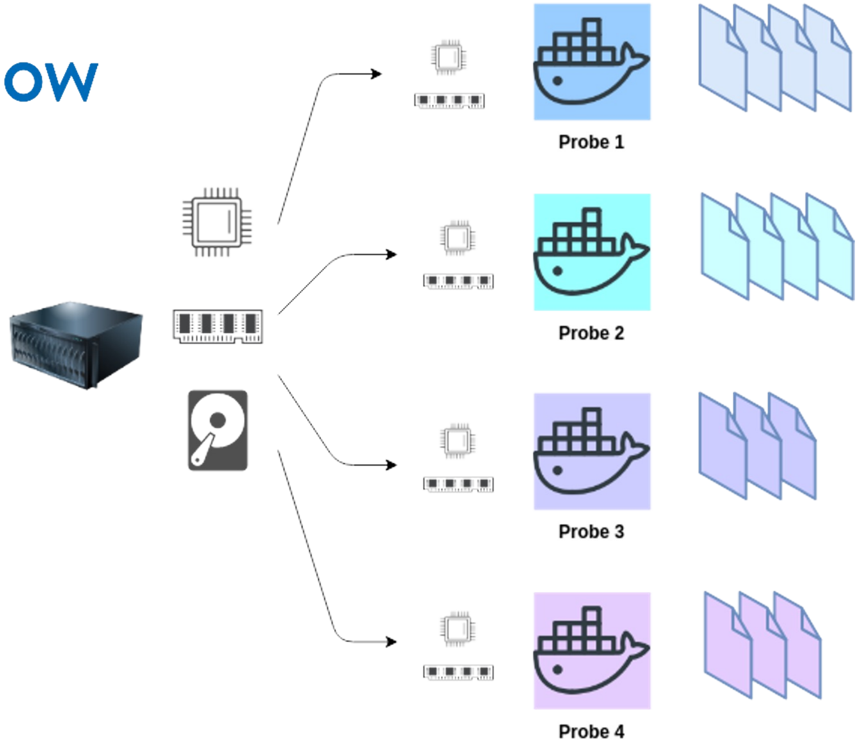
75% idle

SCALING WITH CONTAINERS (T=0)

Quarks Flow

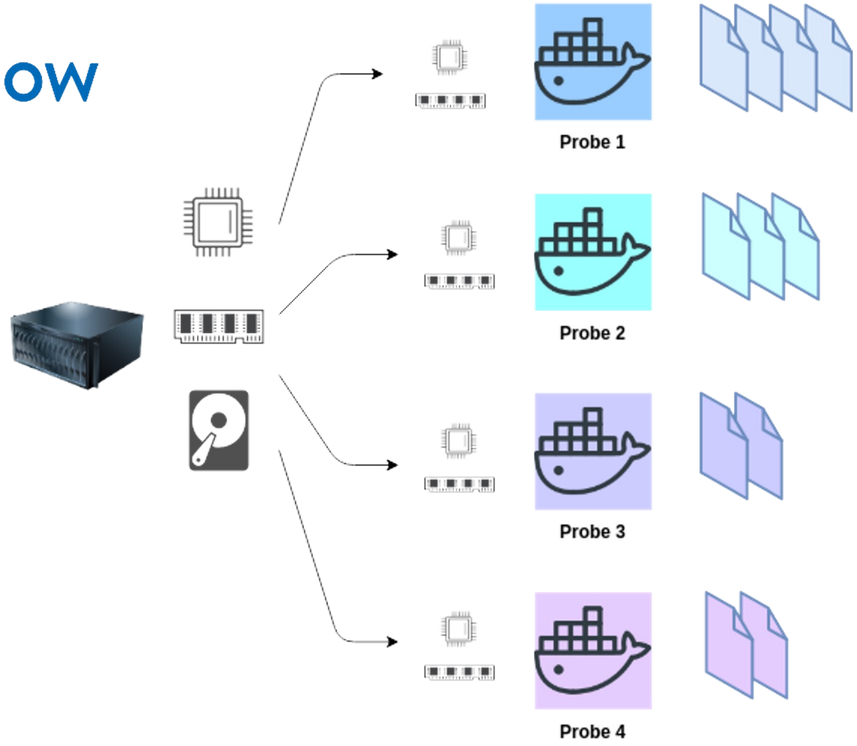


SCALING WITH CONTAINERS (T=1)



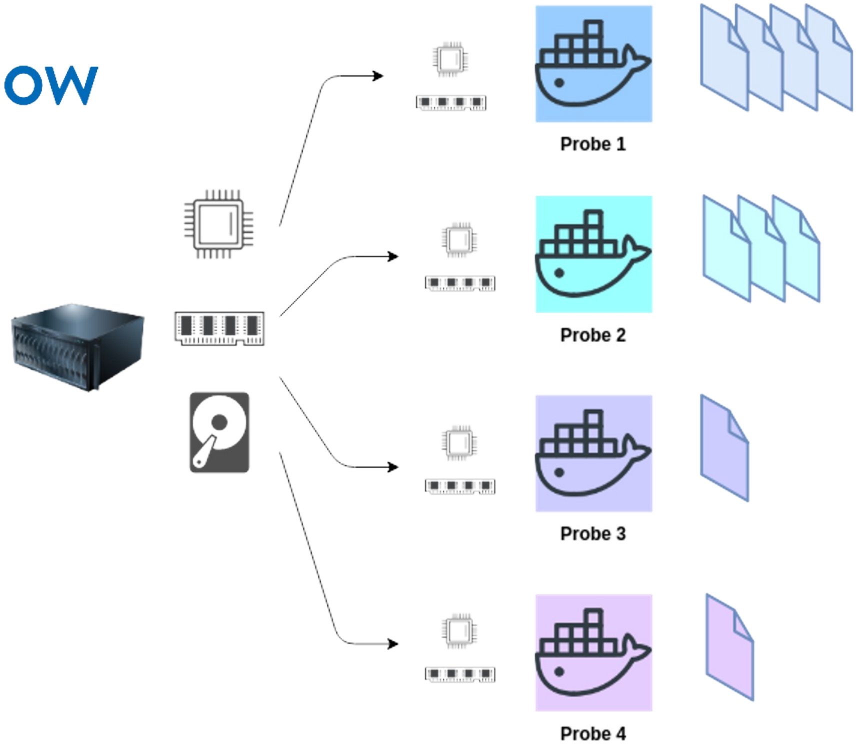
SCALING WITH CONTAINERS (T=2)

Quarks Flow

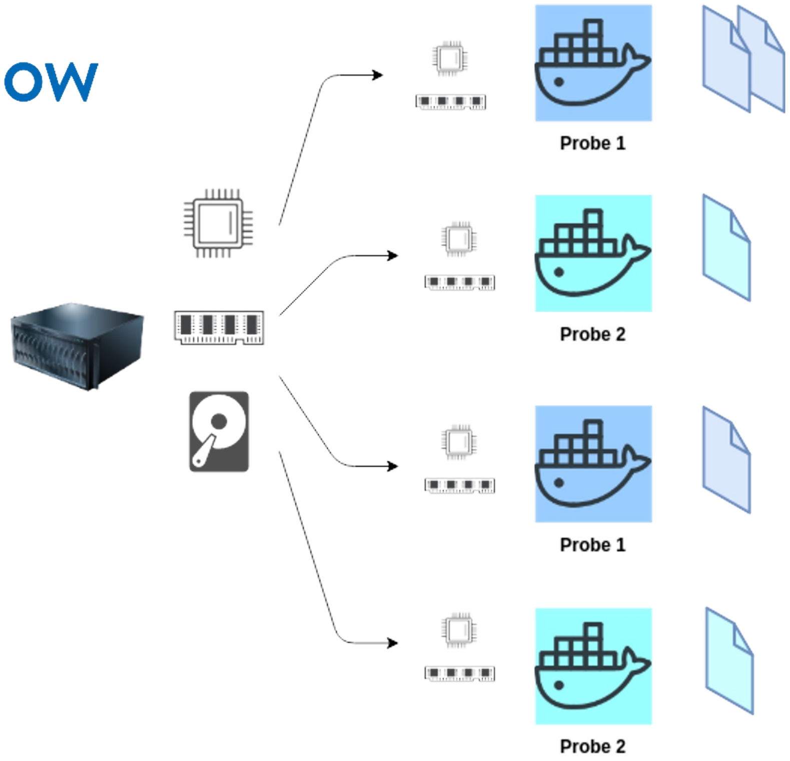


SCALING WITH CONTAINERS (T=3)

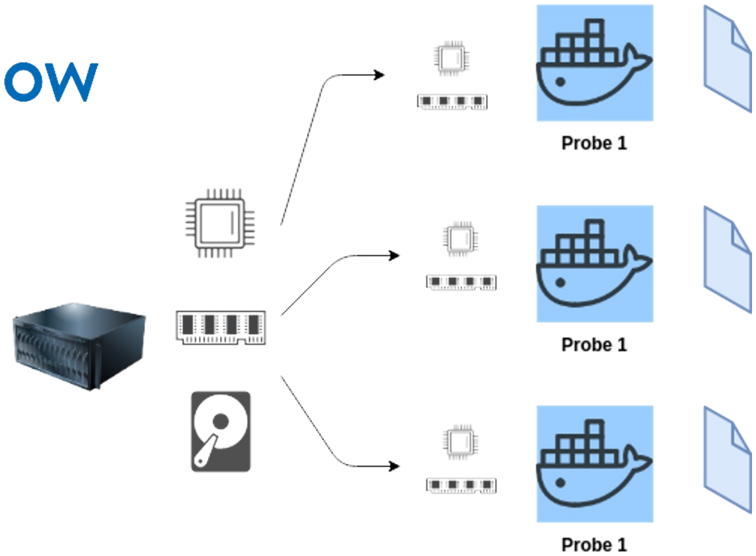
Quarks Flow



SCALING WITH CONTAINERS (T=4)



SCALING WITH CONTAINERS (T=6)



Moving from a single tenant on premise platform to a multitenant SaaS platform

SAAS PLATFORM

Benefits for operating a SaaS platform

- Ability to monitor the usage of customers
- Ability to create, update and manage subscription plans
- And associate them to subscription plans and manage renewals

Cautious: data isolation

- Tenant information added to every data
- API enforcing access control over data

Enough ?

PRIVACY IN SAAS PLATFORM

The problem: uploading files is expensive

- Amplified by the mass scanning (e.g. API in clouds)
- Solution #1: send a hash, upload only if file is unknown yet
- New problem #1: a user can test hash known by the platform

The solution: proof of ownership

- HMAC(tenant_id, hash(file))

Data isolation is not trivial, we play to much with side channels

What is the best AV?
Automation!

DISCLAIMER

Our use case:

- Only static analysis of files
- No runtime (no EDR / “holistic” AV / magic bullet)

EVALUATE AND MUTATE

Question: Is this file a malware? [y/n]

Protocol: mutation based

- **Control** : default group with no mutation
- **Append**: add random bytes to a file
  - ➢ Spots hash based engines
- **Dropper**: stupid exe embedding the file in the data section (no obfuscation)
- **Certifake**: add a spoofed Windows certificate (with a wrong signature)

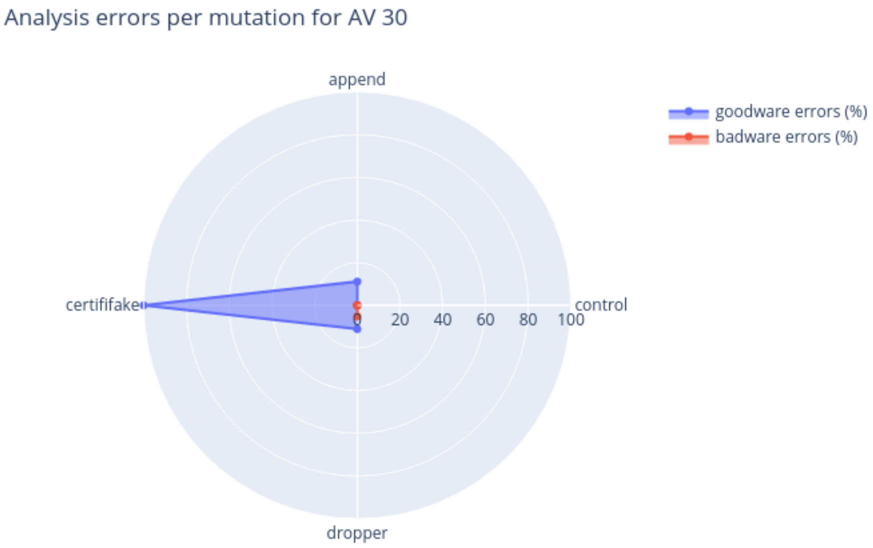
Results:

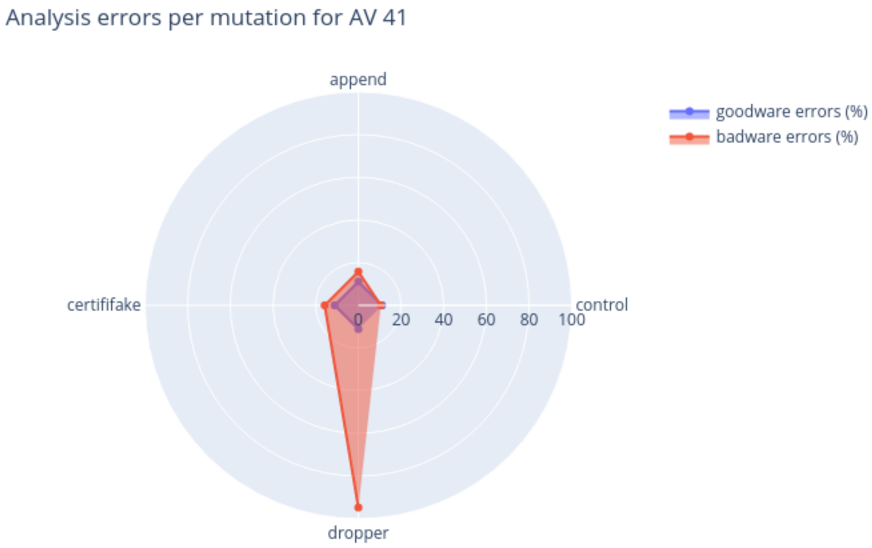
- ◆ False positive (goodware): safe file detected as malware
- ◆ False negative (badware): malware detected as a safe file

Simple tests to understand detection strategies

BLACK BOX BEHAVIOUR ANALYSIS OF AVs

Question: Is this file a malware? [y/n]





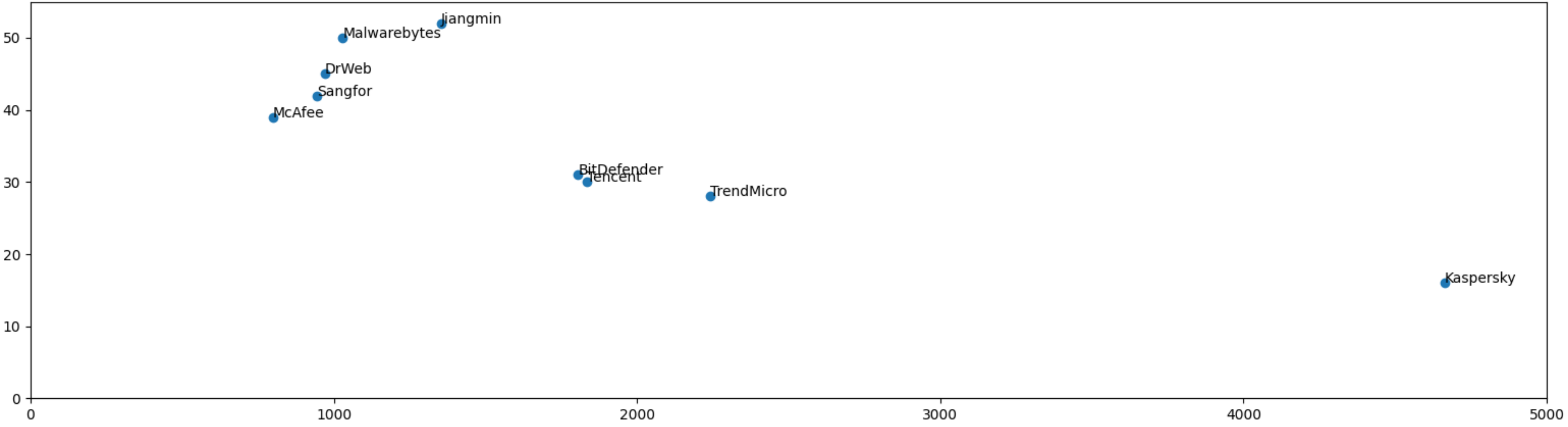
Automation allows to test for 1 engine capabilities…

WEAPONS RACE

- Automation allows to test for 1 engine capabilities… but who cares?
- Does stockpiling engines (AV, EDR, whatever) really improve detection?
  - Without talking about costs, maintenance
- Or we can be smart and choose wisely?

Problem: maximize the detection coverage

COVERAGE IS THE KEY



Legend

- **X-axis**: redundancy (intersection), mutually detected samples
- **Y-axis**: gain, how many *new* samples are detected

Conclusion

WHAT WE HAVE LEARNED

- **Technology changes quickly and opens new possibilities**
  - Adopt too early and you will have something not stable
  - Adopt too late and you will not find resources
- **The challenge was technical, but also cultural**
  - A shift from consulting to building products
  - A shift from geek only to marketing, sales, PM, PMM, …

FUTURE PLANS

- **Advanced Multitenancy**: optimize resource management, application-level tenant configuration and customization
- **White labeling** and **cobranding**
- **Custom workflow** analysis
- **UI/UX Improvements**: i18n, power-user features, improved administration UI

Want to build your future security analysis hub?

- Now: private PoC
- June: private SaaS
- September: open SaaS

Marion Guthmuller, *mguthmuller@quarkslab.com*

Fred Raynal, *fraynal@quarkslab.com*

Quarkslab