

CATCHING THE QUIET THIEF

DETECTING DATA EXFILTRATION WITH
GRAYLOG AND PHILEAS

About us

Rob Dickinson

Graylog API Security

rob.dickinson@graylog.com

github.com/robfromboulder

linkedin.com/in/robfromboulder

Jeff Zemerick

Philterd, LLC

jeff.zemerick@philterd.ai

github.com/jzonthemtn

linkedin.com/in/jeffzemerick

A dirty little secret

Your APIs are a great portal
for stealing data



A breach is just one exploit away

- Even if you have a WAF
- Even if you have an API gateway
- Even if you require authentication
- Even if you use secure coding practices
- You have to be **perfect**, attackers just have to be **lucky**
- Vulnerabilities (technical and human) are **unbounded**

Observability is the cure

Because APIs will always be
vulnerable, we should be
monitoring API calls to detect
data exfiltration

Observing exfiltration is hard

- Job falls to security/ops team in many orgs
- Production APIs are big and noisy
- Many data exfiltrations are low and slow
- Detecting different kinds of PII is pretty hard
- Detecting PII in real time is hard too
- Community projects skew towards Python & LLMs

Our requirements for PII tooling

- Suitable for existing Java services & containers
- Configurable matchers for specific PII types
- Minimum performance of 10k detections/sec/core
- A community of fellow PII nerds

Introducing Phileas

- <https://github.com/philterd/phileas>
- Motivation: Gap in Java open-source PII redaction
- Trust: No proprietary components, all open source
- Flexibility: Run anywhere, even air-gapped

Phileas makes PII detection easy

- Designed to be straightforward
- Configure the types of identifiers to mask
- Redact with consistency to maintain meaning
- Iterate spans for the specific PII type found, confidence level, position & original text

Phileas makes PII detection easy

```
FilterService filterService = new PhileasFilterService();  
FilterResponse fr = filterService.filter(text);  
System.out.println(fr.getFilteredText()); // redacted  
for (Span span : fr.getExplanation.appliedSpans()) {  
    System.out.println(span.getFilterType().name());  
    System.out.println(span.getConfidence());  
    System.out.println(span.getCharacterStart());  
    System.out.println(span.getText()); // original  
}
```

Using & contributing to Phileas

- Shared under Apache Software License
- Commercial use is ok
- All contributions welcome
- Help set the project's direction

Questions?

Rob Dickinson

Graylog API Security

rob.dickinson@graylog.com

github.com/robfromboulder

linkedin.com/in/robfromboulder

Jeff Zemerick

Philterd, LLC

jeff.zemerick@philterd.ai

github.com/jzonthemtn

linkedin.com/in/jeffzemerick

THANK YOU.