# Caveat

These slides were used to run a 20m session as a warm-up for a 30m open Q&A. They are not very content dense :-)

# Why Johnny can't scan at hyperscale

Tales and adventures building security scanning at Google

Claudio Criscione (@paradoxengine)
**Sebastian Lekies (@slekies)**

# About us

## Security Scanning & Vulnerability Management @ Google

### Claudio Criscione (@paradoxengine)

- Robot Overlord (yes, that’s my official title)
- *“Automated tools are only good for monkeys” me, ca 2010*

### Sebastian Lekies (@slekies)

- Staff Software Engineer
- Manager of the Security Scanner Engineering Team

# Hyper*mega*scale Scanning at Alphabet

"Hyperscale computing refers to the facilities and provisioning required in distributed computing environments to efficiently scale from a few servers to ~~thousands of~~ *millions of* servers"

## Our team is responsible for scanning & vuln management at Alphabet

- Includes 50+ organizations
- Several million assets in various environments scanned daily
- Hundreds of different asset types: Corp, Prod, Cloud and Other assets
- Various scanning capabilities: Network-, Code/App-level-, Dependency- & Configuration-Scanning

## At hyperscale all scanning-related problems get much worse

- If an edge case exists, you will hit it
- “Mildly annoying” becomes “Absolutely disastrous”

# The world is not ready for scanning*

*some or all details of this story have been altered for educational purposes

# Scanning challenges at scale

What if you had to scan a high availability network (99.999 % uptime)

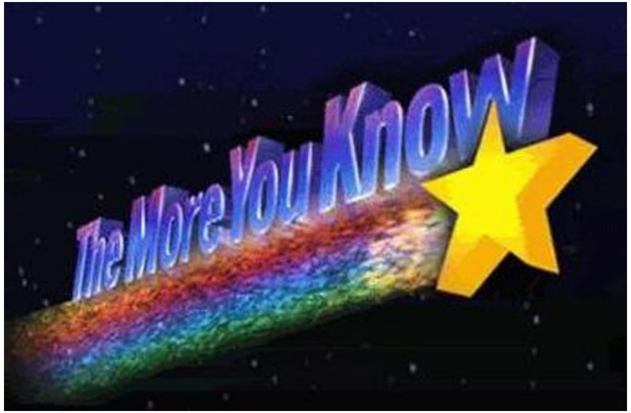
...or if the scanned host is called: acid-pump.corp.google.com

*At hyperscale, you just can't even guess what's going to be out there.*

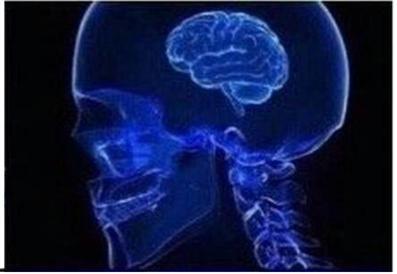
Scanning IP ranges at hyperscale → sadness :-(

**Solution**: Intelligent Scheduling

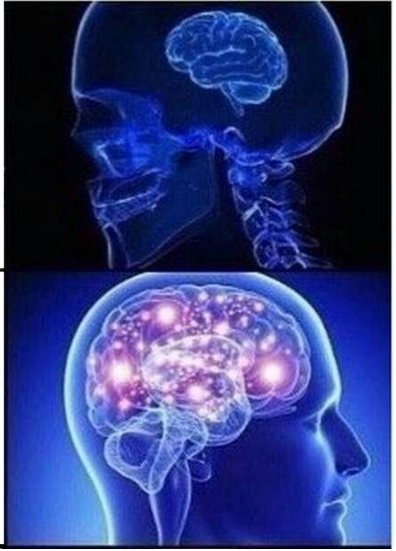
- Know what you are scanning
  - (Type of device, Role in the network)
- Know when it is ok to scan
- Have a good process for exclusion and recovery



IP sweeps
scans



- Ephemeral devices come and go. “Managed” assets with agent are not problematic!
- Without clear attribution, who’s going to fix those vulnerabilities?



IP sweeps
scans

- Ephemeral devices come and go. “Managed” assets with agent are not problematic!
- Without clear attribution, who’s going to fix those vulnerabilities?

“Inventoried IPs”
scans

- OK, now you can cover the assets you know about, and track quality metrics. Hello IPv6!
- What are you going to do about higher level abstractions? Logical routing? Assets with no IPs?

IP sweeps
scans

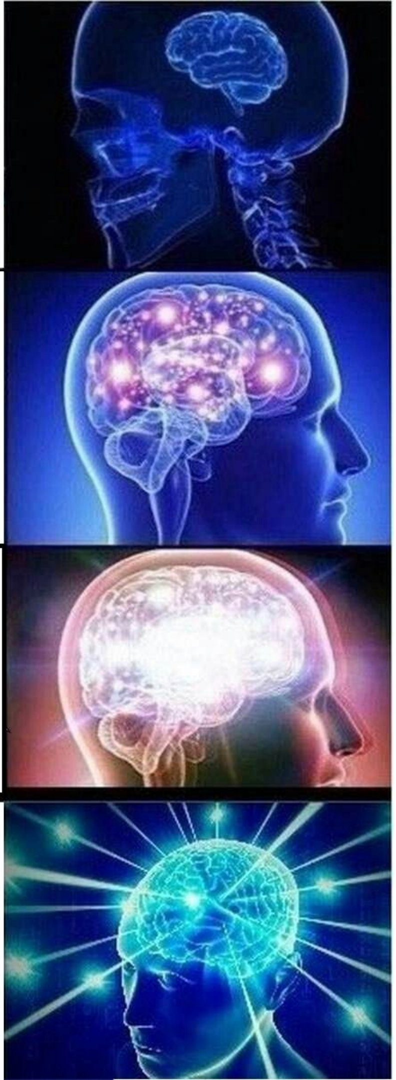
- Ephemeral devices come and go. “Managed” assets with agent are not problematic!
- Without clear attribution, who’s going to fix those vulnerabilities?

“Inventoried IPs”
scans

- OK, now you can cover the assets you know about, and track quality metrics. Hello IPv6!
- What are you going to do about higher level abstractions? Logical routing? Assets with no IPs?

Asset driven
scans

- IPs are “targeting data” only, and not the only type
- Your management processes and tools now scale to different challenges (projects, buckets, VMs..)



IP sweeps
scans

- Ephemeral devices come and go. “Managed” assets with agent are not problematic!
- Without clear attribution, who’s going to fix those vulnerabilities?

“Inventoried IPs”
scans

- OK, now you can cover the assets you know about, and track quality metrics. Hello IPv6!
- What are you going to do about higher level abstractions? Logical routing? Assets with no IPs?

Asset driven
scans

- IPs are “targeting data” only, and not the only type
- Your management processes and tools now scale to different challenges (projects, buckets, VMs..)

Move back to paper
and typewriters

- When is the last time you had to patch your typewriter?

# More fragile than a butterfly

*SANS' survival time*

300
Unix
0
Jul
Aug
Sep

- Attackers don't have license limits on their daily scanned assets
- "Survival time" is a worst case, but real-life experience likely measured in hours
- Scanning once a week/day won't cut it
- **Automation is key**

# Quality matters

1% false positive/irrelevant is **hundreds of findings** at hyperscale.

Hundreds of findings that you can’t manually review quickly enough.

Or hundreds of false reports that undermine your engineers trust.

**Treat false positives as nasty bugs**

Don't cry wolf

# Summary

We are all doomed. Hyperscale is hard!

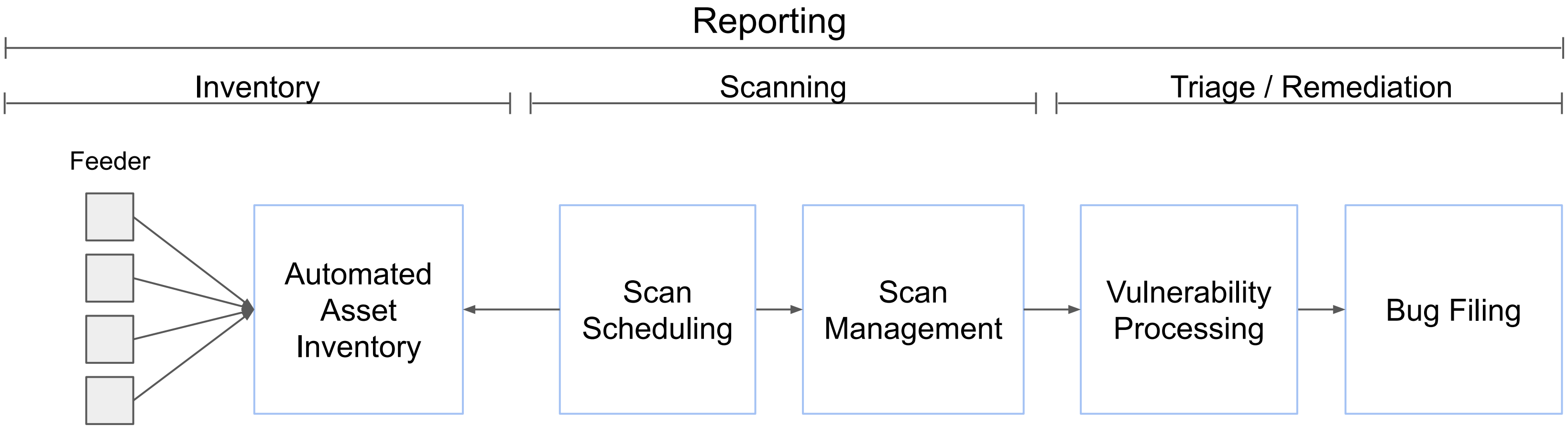
- Automate & manage your inventory
- Leave IPs behind
- Event-based, smart scheduling
- False positive is no laughing matter

End of talk
motivational quote

# Questions?

@paradoxengine, @slekies

# A structure



Reporting

Inventory

Scanning

Triage / Remediation

Feeder

Automated Asset Inventory

Scan Scheduling

Scan Management

Vulnerability Processing

Bug Filing