



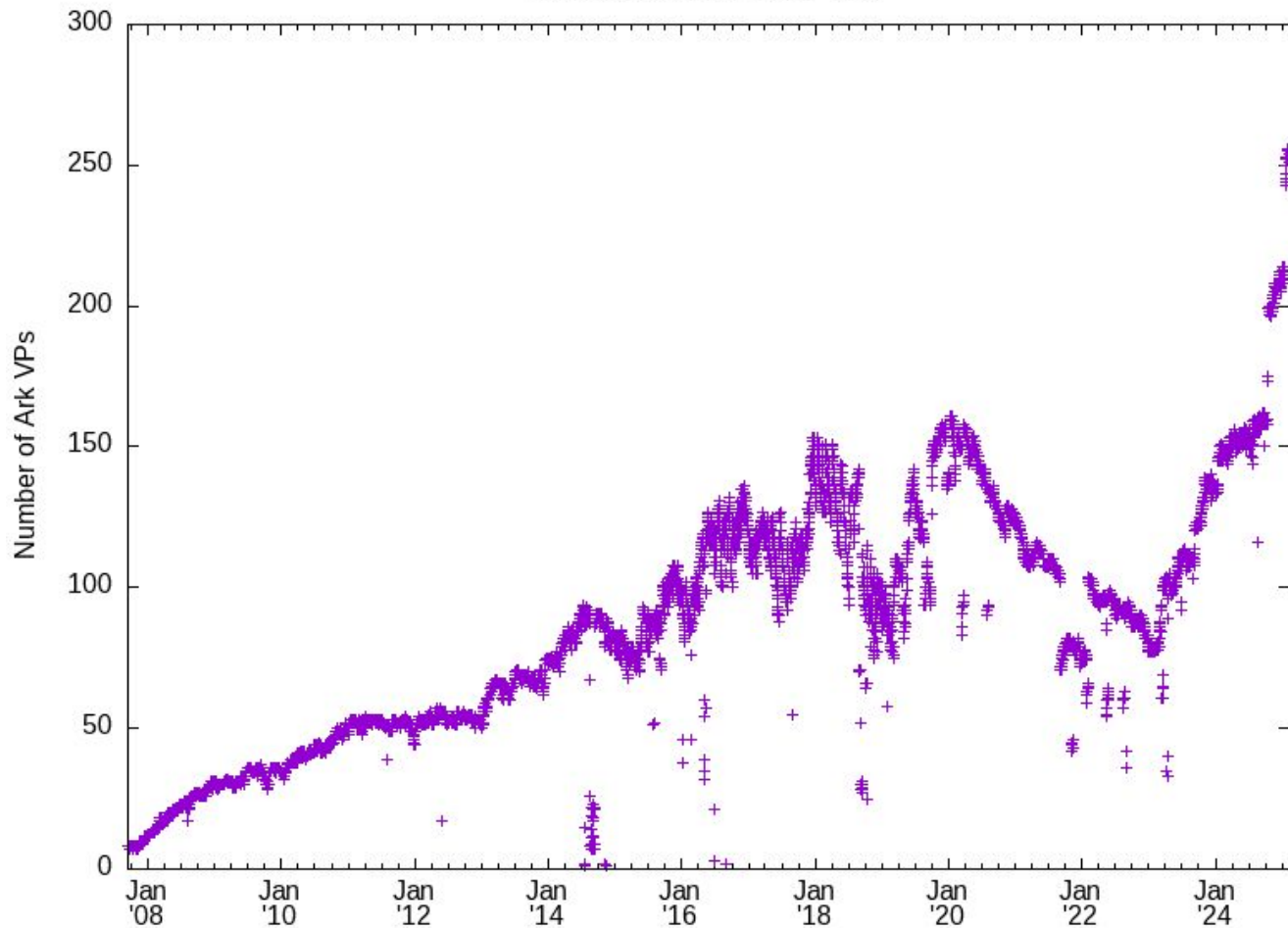
# Globally Distributed Active Measurement Platform

GMI-AIMS-5 February 2025

# In The Beginning...

- Ark began almost two decades ago (2007)
- Pre-dates many of the tools we now take for granted
- Portable across a variety of operating systems (FreeBSD, Linux, etc)
- Small enough to manage by hand
- `ssh`, `rsync`, `cp`, `tmux`, `openssl`

Number of Ark VPs over time



# Scaling Ark For The Future

- Lower the software maintenance burden
- Fewer manual processes
- Simplify creating new nodes

# Use Existing Software

- A lot of what we do isn't unique
- Better use of our time to let other people maintain software
- Avoid incidental technical debt and complexity
- Replace custom software where it makes sense
  - Marinda custom tuple-space messaging vs kafka
  - Tmux "services" vs systemd
  - Natportal custom reverse ssh tunnel vs ssh
- Spend our energy on writing quality measurement tools

# Use Packaged Software

- Dependency management
- Clear versioning
- Reproducible
- Consistent

# Use Packaged Software

- All node management software is installed from Debian packages
  - ssh tunnel
  - ramdisk
  - user credentials
- All ark activities are installed from Debian packages
  - team probing
  - topo-v6
  - scamper
  - spoofer
- We build packages if upstream doesn't
- Nothing gets installed without a package!

# Use Packaged Software

- New code releases automatically build packages using `gitlab-ci`
- Automatically published to our repository using `aptly`



# Automate Certificate Management

- Short-lived certificates
  - Shrinks attack window
  - Ensures renewal workflows actually get tested
- Generate credentials on the node
  - Avoid copying private keys around
- Use X509 certificates
  - Ark components can prove who they are
- Use SSH certificates
  - Less node-specific configuration
  - Avoid needing to update `authorized_keys`

# Automate Certificate Management with Smallstep

- Certificate authority and cli “swiss army knife”
- Single-use tokens to issue initial certificates
- X509 Certificates
- SSH Certificates
- Certificate issuance policies
- Services / timers to renew certificates
- Certificate revocation

# Update Ark Activities To Be Client Driven

- Pull, rather than push, target lists
- Push, rather than pull, results
- Avoids issues with NAT
- Avoids maintaining lists of clients and activities on the server
- Use certificates to prove data provenance

# Help Users To Self-Manage

- Give users tools to manage their Ark nodes so we don't have to
- Bootstrapping a new node
  - Need a token to get the initial certificate
- Node status
  - “Is my node working?”
  - “Why isn't my node showing on the website?”
- Node diagnostics

# Embrace Containerization

- Easier to scale than physical nodes
- Low touch setup and configuration makes containerization possible
  - Simple generic base image
  - Sensible default configuration
- Provide a magic token once on first start
  - Save generated certificates to a persistent volume
- Minimal ongoing maintenance
  - Stripped down image containing the bare minimum
  - Upgrade entire image in one go

# Embrace Containerization

- Run containers on Ark nodes too?
- Minimal, immutable filesystem images
- A/B system updates
- Reduce configuration drift
- Still exploring this space...

# Ark on Internet2

- > 40 container-based Ark nodes
- Running on Nautilus/NRP
- Our first large-scale container deployment

# Ark on Vultr

- 32 container-based Ark nodes
- Running on UTwente Vultr hosts
- Co-existing with their own experiments, and spoofer receivers



# Ark on Measurement Lab

- Running 3 container-based nodes on the sandbox / testing cluster
- Added very basic Prometheus monitoring, liveness probes
- Need to finish testing, and automate more onboarding processes
- Could then have access to ~115 sites

# Scaling Ark For The Future

- Lower the software maintenance burden
  - Use existing software
  - Use packaged software
- Fewer manual processes
  - Automate certificate management
  - Update activities to be client driven
- Simplify creating new nodes
  - Embrace containerization
  - Help users to self-manage
- Reduce the effort per node



<https://www.caida.org/projects/ark/>

[brendonj@caida.org](mailto:brendonj@caida.org)

[ark-info@caida.org](mailto:ark-info@caida.org)