WLCG Cloud Traceability Working Group face to face report

Ian Collier

11 February 2015

Cloud Traceability – The problem

When things go wrong, we need to contain and limit the impact of security issues.

This preserves our reputation & ensures resources are being used for the purposes intended.

In order to do this we must be able to answer these questions about any problematic activity:

- **Who** did **What, When** did they do it, and **Where**

Well established policy & best practice to allow us (more or less) to know these things in grid context.

Increasing use of cloud platforms & virtualisation mean we must reavaluate, monitor & log some different things, and validate that our measures are effective

Cloud Traceability

- 18 particpants – 12 at CERN, 6 remote
- Range of site (cloud provider and grid), security team & 4 main VOs

- Began by gathering ideas/concerns
- Organised those in to a more complete agenda
- Discussed, prioritised & agreed actions

Ideas/Concerns

Issues

- Just use syslog
  - Logging from inside VMs
  - In WLCG we have some degree of trust
- Using job features vs. contextualisation
  - To pass syslog info
- Hypervisor & netflow logging
  - Externally observable behaviour
- Giving sites root access
- Quarantining VMs
  - Ability to keep VM images for forensics
- Classification of VMs
  - Distinguish between pure workload machines vs. ones running longer term services
- Policy evolution
  - Increase of VO role in maintaining traceability
- VO Logging
  - Gap analysis

Practical steps - Logging

- Should primarily depend on, and increase logging of, externally observable behaviour
  - Hypervisor & Cloud management framework
  - Network activity & flows
    - This may depend upon (expensive) hardware
- Cross checking of multiple sources is vital
- Tools for storing, aggregating & searching increasingly important
- *Within WLCG* images are well controlled by VOs
  - This allows a greater degree of trust
  - User and ‘supervisor’ roles are well separated.
  - Therefore we should also use syslog from within VMs
- Potential for changes to VO workflow logging in order to better support traceability

Action: Syslog

- Investigate providing remote syslog service for running VMs
  - Also frameworks for managing & searching high volumes of logs
  - RAL, Glasgow, CERN, Brunel, INFN (Ian Collier)
  - Manchester will provide a syslog server that the VMs can report to.
  - Investigate creating VM images that can be configured to use site syslog
  - Compare machine/job features and site contextualisations
  - Andy McNab to lead (possibly with Ulrich Schwickerath)

Action: Hypervisor & Netflow logging

- Survey experience of sites that do have hardware supporting flow monitoring
  - Nikhef, CERN, IN2P3
- Investigate network flow monitoring on hypervisors
  - Some possible approaches – need to test especially for any performance impact
- Formalise recommendations for logging instantiation etc. of VMs
- Raul Lopes & David Crooks to lead

Action: Site root access to VMs

- Aid to incident response
  - agreed that this can be set up by site instantiation

Action: Quarantining VMs

- Forensic examination of VM images is one specific benefit of virtualisation
  - Easy in some cloud management frameworks
    - Built in to Stratuslab
  - Has a cost in storage – nice to have but not essential
  - RAL, IN2P3 & Glasgow to investigate (Ian Collier)

Policy Evolution

- Track, follow up depending upon outcomes of other actions (Dave Kelsey)

Action: VO Logging

- In order to ensure traceability & ability to map users to processes on VMs may need enhancement of (already substantial) VO workflow logging
- Will advance this with traceability service challenges
- Nikhef to lead (Sven Gabriel)

Summary

- A productive meeting
- A set of clear actions
- Will report at next GDB/pre-GDB
- Scope for more participants – contact Ian Collier

Questions & Links

Questions?

Links

Meeting notes:
https://twiki.cern.ch/twiki/bin/view/LCG/20150210PreGDB

Egroup:
project-lcg-gdb-traceability-tf@cern.ch