WLCG Cloud Traceability Working Group progress

Ian Collier

Pre-GDB Amsterdam

10th March 2015

Working Group History

- October 2014 call for volunteers to do practical work
  - investigating cloud traceability gaps
  - Testing possible solutions
- Face to face meeting at CERN February 10th
  - 18 participants - Many sites and all main LHC VOs represented
  - Gathered ideas/concerns
  - 5 initial areas of interest identified

The problem

Security incidents in our infrastructure are an operational reality

- 70+ in WLCG since 2006

When incidents occur, we need to contain them and limit their impact.

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

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

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

Well established model, backed up by 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.

Many ideas

Condensed to 4 areas for ‘immediate’ investigation

Areas for further investigation

- Just use syslog
  - Logging from inside VMs
  - Exploit the high degree of trust between sites and Vos within WLCG
  - Evaluate merits of using job features vs. contextualisation to pass syslog configuration info
- Quarantining VMs
  - Ability to keep VM images for forensics
- Hypervisor & network flow logging
  - Can depend on the externally observable behaviour
  - Have until now neglected netflow
- VO Logging
  - Gap analysis

In addition

- Giving sites root access
  - Exploit VO trust in sites to support incident response

Agreed this can just happen, does not need further work

- Policy evolution
  - There will be policy changes
  - Should track work and evolve policy as appropriate
  - No immediate action

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
- Ability to cross checking/aggregate from multiple sources increasingly important
- Tools for storing, aggregating & searching increasingly important
- *Within WLCG* images are well controlled by VOs
  - This allows a greater degree of trust
  - User and superuser 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
  - 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
- **Manchester: Initial work on using machinefeatures to pass local syslog server information**
- **Prototype cernvm image supporting contextualisation of syslog**
- **At RAL, getting some logs, ironing out problems.**

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

- **Initial contacts made**

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
  - RAL, IN2P3 & Glasgow to investigate

- **At RAL virtualised WNs already quarantined (managed by Condor). Investigating building this in to storage service.**

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

- **Awaiting progress in other areas**

Summary

- A productive meeting
- A set of clear actions
- Some progress in last month

- 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

(need a CERN account to join on your own)