Traceability in the face of Clouds

EGI-GEANT Symposium – cloud security track

With grateful thanks for the input from
Romain Wartel, CERN and wLCG
Sven Gabriel, Nikhef and EGI
Ian Collier, STFC/RAL

David Groep
Nikhef
Amsterdam
PDP & Grid

wLCG experience

- Incidents happen on a regular basis, 10-12 per year
- Attacks continue to improve over the years
  - More and more sophisticated
    - For example, Zeus (Windows botnet) used to steal HEP accounts
    - No easy or public mean to detect modern malware
  - No longer a side-effect of being connected to the Internet
    - State-of-the-art malware used against WLCG
    - Attackers being arrested for attacking WLCG resources
  - No reduction of the severity or # of incidents in the recent years
    - Yet most of them follow the same pattern
    - We have now built the necessary expertise and have experience

David Groep
Nikhef
Amsterdam
PDP & Grid

The Traceability Premise

"Be able to answer the basic questions who, what, where, and when concerning any incident."

- Prevent re-occurances of the incident
- Prevent a 'waterbed effect' in our federated infrastructure

'in building our infrastructure to federate we also help miscreants spread through federated access – so we now also need rapid, coordinated, and federated response'

- Larger federation ⇨ larger risk of (apparent) 'insider actions'

David Groep
Nikhef
Amsterdam
PDP & Grid

Traceability for the HTC platform

- **Record** ('who, what, when, where')
  - at minimum be able to identify the source of all actions (executables, file transfers, portal jobs) and the individual who initiated them
  - traceability commensurate with scope of action
- and **React**
  - sufficiently fine-grained controls, such as blocking the originating user and monitoring
  - communicate controls information rapidly throughout the federation (resource centres, users, communities)
- and only then **Recover**
  - understand the cause and to fix any problems before re-enabling access for the user

David Groep
Nikhef
Amsterdam
PDP & Grid

A policy framework

- Number of security policies apply to participants:
  - http://wlcg.web.cern.ch/security/computer-security
- Important operational security:
  - Security Incident Response Policy
    - https://edms.cern.ch/document/428035
    - "A security incident is the act of violating an explicit or implied security policy "
    - Report suspected incidents locally and to the infrastructure
    - "Perform appropriate investigations and forensics and share the results with the incident coordinator"
    - "Aim at preserving the privacy of involved participants and identities"
  - Traceability and Logging Policy
    - https://edms.cern.ch/document/428037
      https://documents.egi.eu/document/81

David Groep
Nikhef
Amsterdam
PDP & Grid

Current EGI/wLCG Security Traceability and Logging Policy

- Idea: understand and prevent incidents*
- Requirements:
  - Software MUST produce application logs:
    - Source of any action
    - Initiator of any action
  - Logs MUST be collected centrally [resource centre]
  - Logs MUST be kept 180 days
- Sites currently know what to do in order to be able to answer who, what, where & when

David Groep
Nikhef
Amsterdam
PDP & Grid

Capabilities

- Forensics & trace analysis capabilities scarce
  - Mostly at the larger resource centres and with a few specialised institutes and individuals
  - Logs and audit records needed for experts to work on
  - Collaborate widely with the trusted community to maintain integrity of our ecosystem at large

David Groep
Nikhef
Amsterdam
PDP & Grid

Beyond the HTC platform offering

With new service classes (like IaaS clouds) our 'attack surface' increases

- **Record?**
  - we now need traceability capabilities for **all access methods**
  - with **expertise for forensics** and analysis
- **React?**
  - controlling access for suspected miscreants
  - both to the innards of the VM as well as to the 'external controls' (management interfaces, KVM console, networks, …)
- **Recover?**
  - **different entities** now responsible for the resolution
  - But re-enabling any service should wait for *full* resolution!

David Groep
Nikhef
Amsterdam
PDP & Grid

‘Sites’ can’t do it all

- We cannot implement traceability in exactly the same way
- Sites can log observable behaviour
  - VM launched at such and such a time
  - Network connection to such and such an address at a certain time
  - Etc.
- Sites can no longer see
  - Credential used to run workload(s) inside VMs
  - Detailed application logs from within past VMs
- **CAN** isolate running VMs for analysis

David Groep
Nikhef
Amsterdam
PDP & Grid

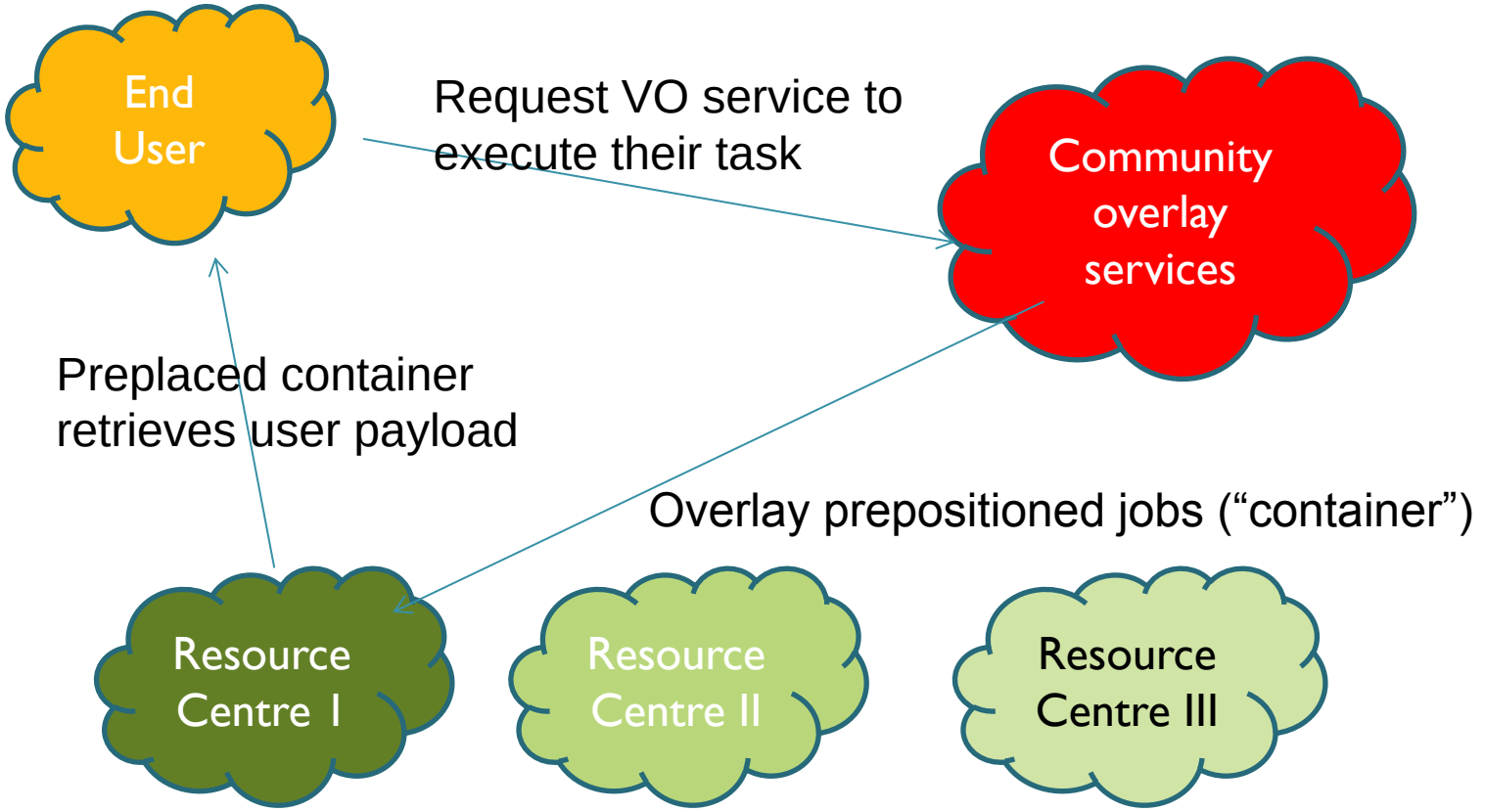
New territory

- VOs (research communities) will have to participate in incident response to provide the missing information.
- Are VOs going to maintain detailed central application logs and retain them?
- Could sites provide a central syslog service for VMs run at their site?
  - But that would not help for public cloud work
  - Perhaps just for some nodes
- **Many more issues and questions**

David Groep
Nikhef
Amsterdam
PDP & Grid

Distributed traceability in practice?

- wLCG already faced distributed responsibility



David Groep
Nikhef
Amsterdam
PDP & Grid

Exercising traceability

For a test, a fake-malicious user payload was submitted through community container portal …

- Common & multiple-use containers made tracing impossible for the VO – and the VO-CSIRT existed (unique!) and was involved
- After a week (!) the intruder was not yet found
- Remarkable resources would have be needed for a proper response
- Retention times for needed logs are to short (<30 days).
- "*It would have taken O(1 week) to scan all input sources for the offending code*"

David Groep
Nikhef
Amsterdam
PDP & Grid

Next steps

- We need to address this *before* workflows become too firmly established.
  - Easier to build in early than to add on afterwards
  - Traceability requires specific design at every level

- Working group (sites, communities, and users)
  - Test different approaches to filling traceability gaps
  - Update guidelines
  - Disseminate
  - **Exercise the system** –
    with planned and unscheduled challenges

David Groep
Nikhef
Amsterdam
PDP & Grid