Virtualisation Working Group Report

Tony Cass

HEPiX Fall 2010

November 1st 2010

Organisation

- Slow start, mostly due to re-org @ CERN.
- 66 people on mailing list; core of ~10 regular participants, but many more join meetings on occasion.
  - Many thanks to Ian Gable for taking notes during meetings.
- Group identified 5 work areas
  - Image Generation Policy
    - » Dave Kelsey & Keith Chadwick
  - Image Exchange
    - » Owen Synge
  - Image Expiry/Revocation
    - » Later agreed to be part of policy & exchange area
  - Image Contextualisation
    - » Sebastien Goasguen
  - Multiple Hypervisor Support
    - » Andrea Chierici

Policy for Trusted Image Generation

- You recognise that VM base images, VO environments and VM complete images, must be generated according to current best practice, the details of which may be documented elsewhere by the Grid. These include but are not limited to:
  - any image generation tool used must be fully patched and up to date;
  - all operating system security patches must be applied to all images and be up to date;
  - images are assumed to be world-readable and as such must not contain any confidential information;
  - there should be no installed accounts, host/service certificates, ssh keys or user credentials of any form in an image;
  - images must be configured such that they do not prevent Sites from meeting the fine-grained monitoring and control requirements defined in the Grid Security Traceability and Logging policy to allow for security incident response;
  - the image must not prevent Sites from implementing local authorisation and/or policy decisions, e.g. blocking the running of Grid work for a particular user.
- http://www.jspg.org/wiki/Policy_Trusted_Virtual_Machines

Image Cataloguing and Exchange

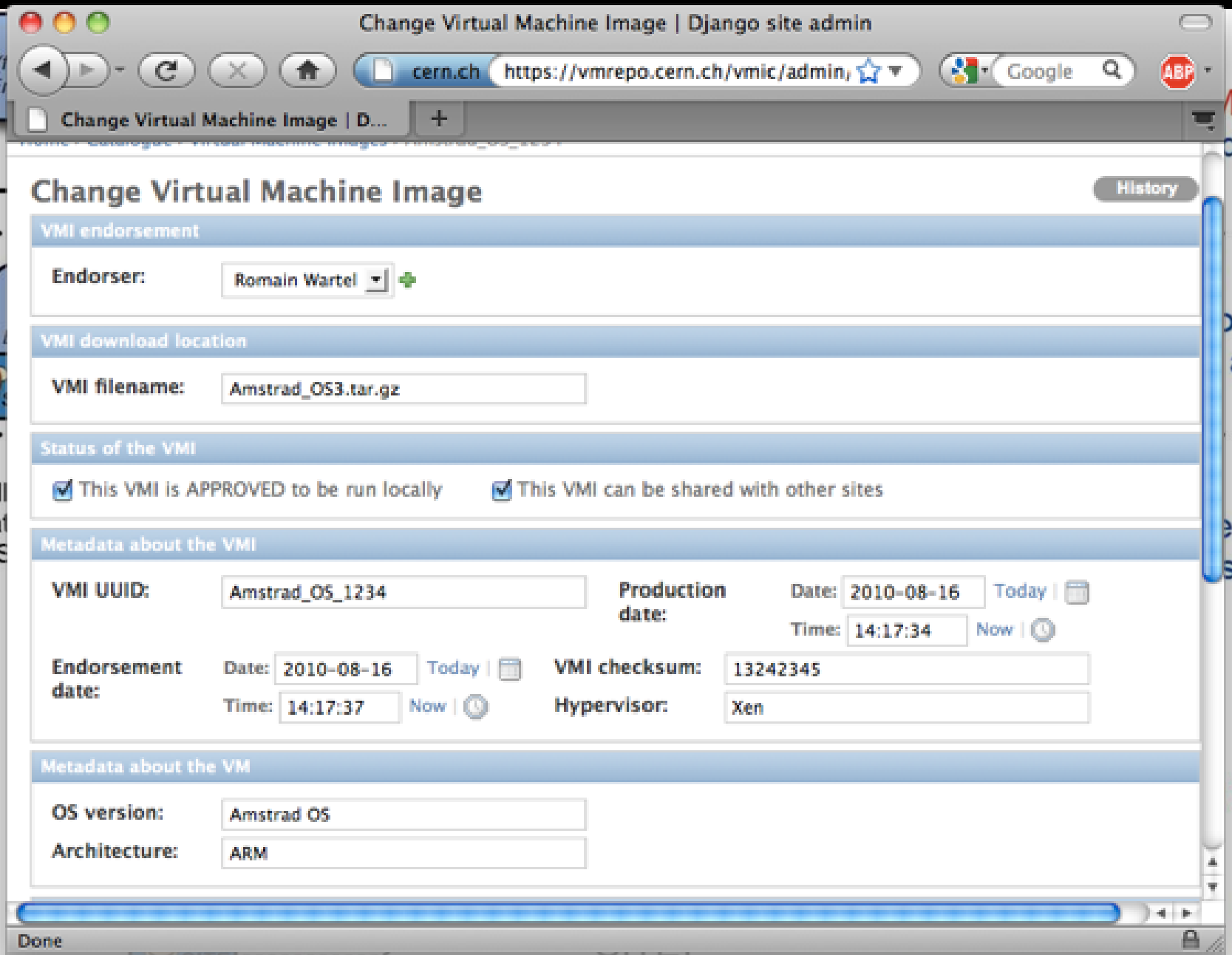

Image Contextualisation

- Contextualisation is needed so that sites can configure images to interface to local infrastructure
  - e.g. for syslog, monitoring & batch scheduler.
- Contextualisation is limited to these needs! Sites may not alter the image contents in any way.
  - Any site are concerned about security aspects of an image should refuse to instantiate it and notify the endorser.
- Contextualisation mechanism
  - Images should attempt to mount a CDROM image provided by the sites and, if successful, invoke two scripts from the CDROM image:
    - prolog.sh before network initialisation
    - epilog.sh after network initialisation

Multiple Hypervisor Support

- Andrea
  - Surveyed sites; results show that kvm and Xen dominate as hypervisors, especially in batch virtualisation area.
  - Documented method to produce VM image that can be used with both kvm and Xen
    - Method tested by Sebastien Goasguen and Abdeslem Djaoui (RAL)

Current Status

- Generation policy
  - Clear, but probably needs to be formally approved by JSPG?
- Contextualisation & kvm/Xen support
  - Also clear.
- Image Cataloguing & Exchange
  - Ideas sound, and working internal catalogue @ CERN, but we need functioning inter-site exchange!
  - Key issue is lack (to date) of working group member(s) with management support to deliver (and support!) a solution.
  - Stratus Lab, as Michel will report later this week, has developed similar ideas.
    - » Joint intention to explore collaboration, but no opportunity to do so before late November.

Other thoughts

- The CernVM filesystem offers an attractive way to ensure sites have the correct VO software, reducing the need for VM images as a mechanism for this.
  - See Ian Collier's presentation shortly.
  - CVMFS team have asked Romain Wartel to lead security audit
    - This should allay any fears from sites about using CVMFS.
- Virtualisation is an area with much scope for communication failures!
  - We must be clear that "image endorsement" is a very rapid process
    - The person creating endorses an image which is then immediately available for instantiation by all sites who trust the endorser; there is no need for a lengthy process of verification at sites.
  - Some sites talk about restricting instantiated VM images but the actual impact for end-users is likely small
    - e.g. VM images would have no need to connect to a NFS-based shared storage area, so it would not matter if "isolated from the rest of the network" just means "no access to our NFS servers".
      - This appears to be the likely situation at NIKHEF, one of the sites the most reluctant to enable instantiation of remotely generated images.

Summary

- The working group has made good progress in establishing policies to allow the exchange of VM images…
- … but not such good progress in delivering a distributed catalogue of endorsed images.
- CVMFS is probably the neatest solution to the problem of VO software distribution…
- … but VM exchange remains interesting
  - as an option for sites to run hypervisors not OSes and automatically migrate to latest patched system as images instantiate, and
  - if the VM images can contact pilot job frameworks directly, simplifying the scheduling problems at sites.