IBM Research, Hawthorne NY

Managing Security of Virtual Machine Images in a Cloud Environment

Jinpeng Wei

Florida International University

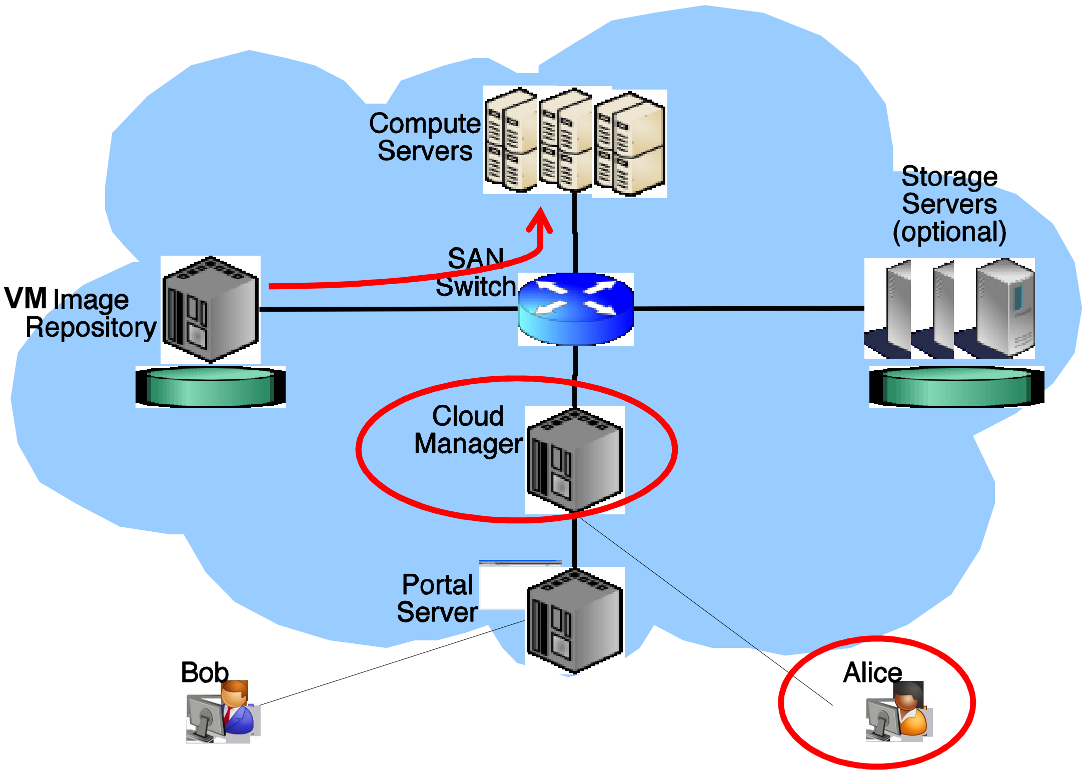
Xiaolan Zhang, Glenn Ammons, Vasanth Bala

IBM T J Watson Research Center

Peng Ning

North Carolina State University

Virtual Machine Images in a Typical Cloud



- Virtual Machine (VM) = logical computer
- Virtual Machine Image (VM Image) = logical computer <u>in a file</u>
- VM image repositories: collections of VM images
  - E.g., VMware virtual appliance market place, Amazon machine images (AMIs) collection in EC2
  - Facilitate deployment of new virtual machines
  - Reduce management/configuration cost of the cloud users

VM image sharing is one of the underpinnings of cloud computing

Security Risks in an Image Repository

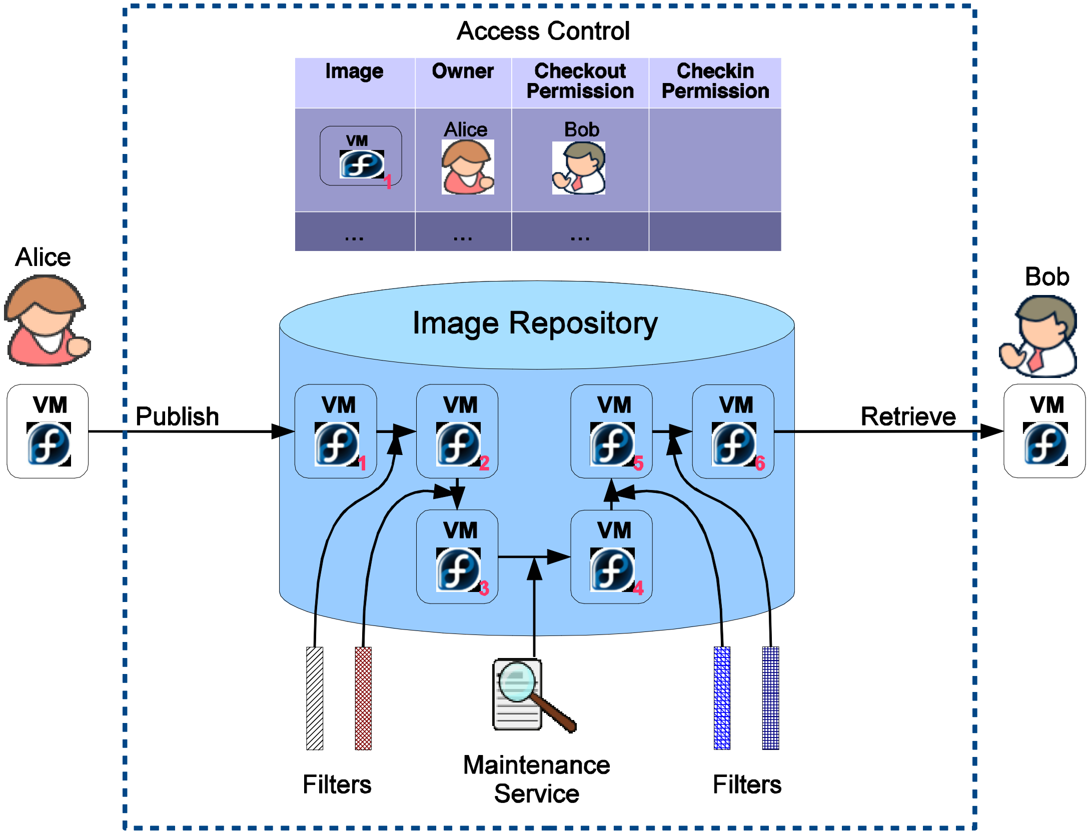
- **The publisher's risk: inadvertent leaking of sensitive information (private data or intellectual properties) and unauthorized access to the image**
  - Sensitive information is often stored without the publisher's awareness. E.g., autocomplete feature of some browsers

- **The retriever's risk: running vulnerable or malicious virtual machine images**
  - A retrieved image may be instantiated into a full-fledged intruder machine inside a corporate network. Easier way to deploy Trojan Horses

- **The repository admin's risk: hosting and distributing images that contain malicious or illegal content**
  - Software patches, software license compliance checks
  - No systematic way to track image ownership, provenance or derivation relationships

Solution Overview: Mirage

- An access control framework: regulates the sharing of VM images
- Image filters: remove unwanted information in the image
- A provenance tracking mechanism: tracks the derivation history of an image and the associated operations performed on the image



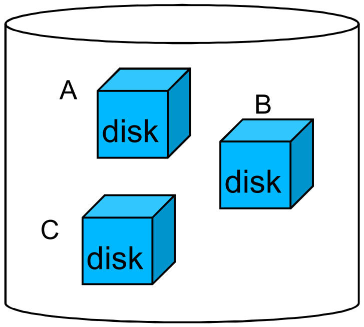
Access Control

Image	Owner	Checkout Permission	Checkin Permission
VM 1	Alice	Bob	
...	...	...	

- A set of repository maintenance services, such as periodic virus scanning of the entire repository, that detect and fix vulnerabilities discovered after images are published

Implementation: the Mirage Image Library

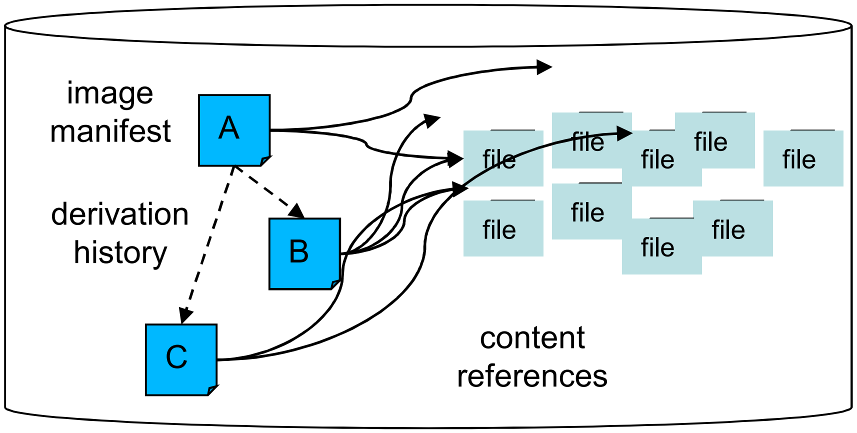
Conventional image library



Disk granularity store

- Disk based representation
- No image relationships
- Hypervisor-dependent
- Merely a storage system for image disks

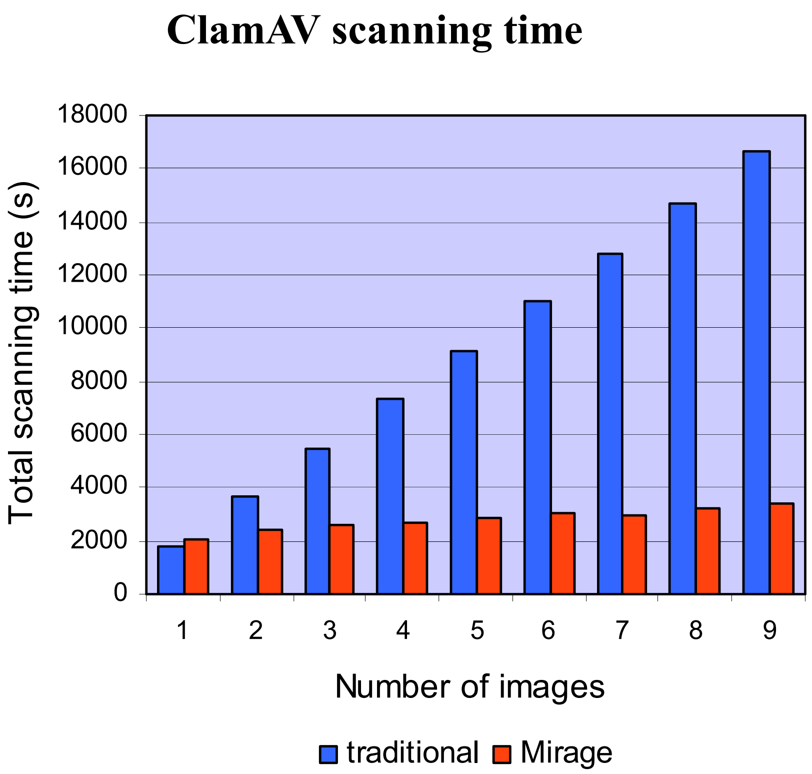
Mirage image library



Content addressable, file granularity store

- File based representation
- Image relationships (think CVS)
- Hypervisor-agnostic
- A sophisticated store with APIs to directly manipulate images without deploying them as instances or fully assembling their disks
- Conventional disk is reconstituted when an image is checked out

Preliminary Experiments



- The VM images are daily snapshots of a large, commercial, Eclipse-based development environment (6GB, ~60,000 files)
- Each unique file is scanned only once, even if shared among many VM images
- Scanning time gains depend on the similarity among VM images

```
Scan the CAS as if it is a single file system;
For each infected file F {
    For all image manifests that contain a reference to F,
        flag the reference as 'infected';
}
```

Uses a reverse index (constant time operation to identify image manifests containing reference to F)

Acknowledgements

- **IBM T. J. Watson Research Center**
  - Bowen Alpern
  - Arun Iyengar
  - Todd Mummert
  - Darrell Reimer
  - Jian Yin

Thank you!