Breaking Up is Hard to Do

Security and Functionality in a Commodity Hypervisor

Presented by Saad Arif
Virtualization in Cloud Computing

• Cloud computing uses virtualization to lease small slices of large scale datacenter facilities to individual paying customers

• Virtualization offers
  – Resource utilization
  – Administrative features
  – Support of existing software

• Results in large scale hosting platforms
Companies in the Cloud
(all these run in EC2 or Rackspace)
Virtualization in Cloud Computing

- Control VM (Dom0)
  - Often a full OS
  - Similar privileges as hypervisor
  - Offers services to guest VMs
Hypervisors are Secure

Small codebase

x86

x86

x86

Hypervisor

Xen: 280 KLOC (based on the current version)
Nova: 9 KLOC (microvisor) + 20 KLOC (VMM) [EuroSys’10]
SecVisor: 2 KLOC [SOSP’07]
Flicker: 250 LOC [EuroSys’08]
Example Attack Vectors

- CERT vulnerability database for type-1 hypervisors
  - 44 attacks
  - 23 originated from within guest VMs
  - Based on attack vector,
    - 14 showed device emulation layer vulnerabilities
    - 2 in virtualized device layer
    - 5 in management component
    - *Only 2 hypervisor exploits*

So 21 out 23 were attacks against service components in the control VM
Hypervisor

Control VM (Dom0)

User A's VM

User B's VM

Platform

IPC

Device Drivers

Device Emulation

Manage devices

Create and destroy VMs

Arbitrarily access memory

“We are the 90%”
Exposure to Risk

- Isolate services into least-privileged service VMs
- Make sharing between components explicit
- Contain scope of exploits in both space and time

Constraint: Don’t reduce functionality, performance, or maintainability of the system
Space

- Control VM
  - IPC
  - Management
  - Device Drivers
  - Platform
- Device Emulation
- User A’s VM
- User B’s VM

Hypervisor
Isolation

- Control VM
- Platform
  - PCI Config
  - System Boot
- IPC
  - XenStore
- Management
  - Builder
  - Tools
- Device Drivers
  - Network
  - Block
- Device Emulation
  - Emulator
Space
Isolation

Platform
- PCI Config

Management
- Builder
- Tools

Device Drivers
- System Boot
- Network
- Block

IPC
- XenStore

Device Emulation
- Emulator

User A’s VM

User B’s VM

Hypervisor
Configurable Sharing

User A’s Tools

User A’s Network

User A’s Block

User A’s VM

User B’s Tools

User B’s Network

User B’s Block

User B’s VM
Configurable Sharing

- Tools
- Network
- Block

- User A’s VM
- User B’s VM
Configurable Sharing

User A’s Tools

User A’s Network

User A’s VM

User A’s Block

User B’s Tools

User B’s Network

User B’s VM

User B’s Block
Space
Isolation
Configurable Sharing

Hypervisor

Platform
  PCI Config

Device Drivers
  System Boot
  Network
  Block

Management
  Builder
  Tools

IPC
  XenStore

Device Emulation
  Emulator

User A’s VM
User B’s VM
Auditing

Which VMs were relying on the Block component while it was compromise?

- VM B and VM C
- User A’s VM
- User B’s VM
- User C’s VM
Space
Isolation
Configurable Sharing
Auditing
TIME
Space

Containment
Configurable Sharing
Auditing

Time

Platform
PCI Config
System Boot

Management
IPC
XenStore
Builder
Tools
Network
Block

Device Drivers

Device Emulation
Emulator

Hypervisor

User A’s VM
User B’s VM
Disposable

PCI Config

System Boot

Services

Hypervisor
Space
Isolation
Configurable Sharing
Auditing

Time
Disposable

Platform
PCI Config
System Boot

Management
Builder
Tools

Device Drivers
Network
Block

IPC
XenStore

Hypervisor

Device Emulation
User A’s VM
User B’s VM
Snapshots

4-25 ms
**Space**

- Isolation
- Configurable Sharing
- Auditing

**Time**

- Disposable
- Timed Restarts

**Components**

- Hypervisor
- System Boot
- PCI Config
- Network
- Block
- Builder
- Tools
- Device Emulation
- User A’s VM
- User B’s VM
- Device Drivers
- XenStore
- IPC

**Features**

- Space Isolation
- Configurable Sharing
- Auditing
- Time Disposable
- Timed Restarts
Stateless VMs

Builder

User B’s VM
Hypervisor

Platform
- PCI Config
- System Boot

Management
- Builder
- Tools

Device Drivers
- Network
- Block

IPC
- XenStore

Space
- Isolation
- Configurable Sharing
- Auditing

Time
- Disposable
- Timed Restarts
- Stateless

Hypervisor

Device Emulation
- Emulator

User A’s VM

User B’s VM
SPACE + TIME
Space
- Isolation
- Configurable Sharing
- Auditing

Space + Time
- IPC
- XenStore

Time
- Disposable
- Timed Restarts
- Stateless

Platform
- PCI Config
- System Boot

Management
- Builder
- Tools

Device Drivers
- Network
- Block

Device Emulation
- Emulator

Hypervisor

User A’s VM

User B’s VM
Composition

B: Network can map 0xDEADBEEF

I’ve enabled 0xPWND

User A’s VM

User B’s VM

OK
Composition

B: Network can map 0xDEADBEEF
A: Please shut me down

A: Please shut me down
I’ve enabled 0xPWND
I've enabled the network driver to map 0xDEADBEEF

OK

B: Network can map 0xDEADBEEF

A: Please shut me down

I’ve enabled 0xPWND
Space
- Isolation
- Configurable Sharing
- Auditing

Space + Time
- Composition

Time
- Disposable
- Timed Restarts
- Stateless

Platform
- PCI Config

Management
- Builder
- Tools

Device Drivers
- Network
- Block

Hypervisor

IPC

XenStore

Device Emulation

Emulator

User A’s VM

User B’s VM

System Boot

Space + Time

Composition
EVALUATION
Evaluation

• What do privileges look like now?
• What is the impact on the security of the system?
• What are the overheads?
• What impact does isolation have on performance?
• What impact do restarts have on performance?
## Privileges

<table>
<thead>
<tr>
<th>Privilege</th>
<th>System Boot</th>
<th>PCI Config</th>
<th>Builder</th>
<th>Tools</th>
<th>Block</th>
<th>Network</th>
<th>XenStore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbitrarily Access Memory</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access and Virtualize PCI devices</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Create VMs</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage VMs</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage Assigned Devices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Security

• Of the 21 vulnerabilities against the control plane, we contain all 21

• TCB is reduced from the control VM’s 7.5 million lines of code (Linux) to Builder’s 13,500 (on top of Xen)
Isolation Performance

Postmark performance

wget performance
Restart Performance

Kernel build performance

- Dom0 (nfs)
- Xoar (nfs)
- Restarts (10s)
- Restarts (5s)
CONCLUSION
Summing it All Up

• Components of control VM a major source of risk

• Xoar isolates components in space and time
  – Contains exploits
  – Provides explicit exposure to risk

• Functionality, performance, and maintainability are not impacted