

# Welcome & Project Update

Ian Pratt

Keir Fraser



# Attendee Profile

- 100+ Attendees
- 8+ Countries
- 36+ Companies
- 5+ Universities



# Program Committee

- Dan Magenheimer – Oracle
- Jose Renato Santos – HP
- Jun Nakajima - Intel
- John Janakiraman – Skytap
- Alex Vasilevsky – Virtual Computer
- Victor Hugo dos Santos
- Sang-bum Suh - Samsung
- Hitoshi Matsumoto - Fujitsu

**Thanks for putting together a great event!**



# Event Information

- Breakout Room Available During Event
  - Room 105
- See handout for Wireless Setup
- Lunch – Provided for Registered Attendees at Conference Center
- Abstracts & Speaker Profiles at



<http://www.xen.org/community/xensummit.html>



# Evening Event – Tuesday



6:30 pm – 9:30 pm

<http://www.computerhistory.org/>

- Dinner, Wine, Beer, and Sodas
- 2 Guides for Private Museum Tours
- Directions at Registration Table
- Sign-up Sheet for Carpooling



# Xen Summit Gear

- Xen Summit Jackets
- Carabiners for Event Pass



# Agenda Overview - Tuesday

9:30 – 10:00 am	Welcome and Project Status
10:00 – 10:30 am	Roadmap & Releases
10:50 – 11:10 am	Difference Engine
11:10 – 11:40 am	PVOPS Status
11:40 – 12:15 pm	Transcendent Memory on Xen
1:15 - 1:45 pm	Satori: Enlightened Page-Sharing
1:45 – 2:05 pm	Paravirtualized USB Support
2:05 – 2:35 pm	PCI-Pass Through Techniques
2:35 – 3:05 pm	Status of SR-IOV & VT-D
3:30 – 4:00 pm	Cross Vendor Migration
4:00 – 4:20 pm	Power Management in Xen
4:20 – 4:40 pm	Detecting and Correcting Transient Errors via Xen

# Agenda Overview - Wed

9:00 – 9:30 am	Open Nebula VM Manager
9:30 – 10:00 am	Client Virtualization Framework
10:00 – 10:30 am	Tralfamadore
10:45 – 11:00 am	VM Snapshots
11:00 – 11:30 am	Real-time and VMM
11:30 – 12:00	Nested Page Tables
1:00 – 1:30 pm	Project Zentific
1:30 – 2:00 pm	Just Run It
2:00 – 2:30 pm	3D Rendering on Xen
2:30 – 3:00 pm	REFLINK Operation in ocfs2
3:30 – 4:00 pm	Achieving 10GB/s Paravirt Drivers
4:00 – 4:30 pm	Xen Scalability



# Xen Summit Europe at LinuxTAG



# Overview

- XenEmbedded
- XenClient
- HostedXen
- Xen Introspection API



# XenEmbedded

- Small footprint Xen environment
- For embedding into Servers, Desktops,

Laptops, Routers, Storage Arrays

- buildroot/busybox/uclibc based
- 8MB compressed image size
- xenvm/xenops embedded toolstack
  - Simple configuration file
  - Process per VM with control socket
- Fast booting
- See [xenbits.xen.org/xenclient](http://xenbits.xen.org/xenclient)



# XenClient

- Repository for bleeding-edge client related work, feeding in to xen-unstable
- Enhanced device pass-through
  - Intel, ATI, nVidia graphics
- Power conservation and suspend
- ACPI/SMBIOS virtualization
  - function keys, battery state, etc
- Secure mouse/keyboard routing
- Graphics virtualization
- Fast boot and measured launch
- Enhanced USB emulation
- “In-place P2V”, easy to install



# Hosted Xen

- Xen engine run as a kernel module to enable hosted (type-2) VMM
  - Runs on Windows and OSX today
- Enables Xen to also compete with other

type-2 VMMs

- KVM, VirtualBox, Parallels, VirtualPC, VMWare Workstation/ACE/Fusion
- Loses the security and isolation benefits of a true type-1 hypervisor, but

still useful in some scenarios



# Hosted Xen

- Xen loaded as an ELF module, linked via dispatch table into host kernel
- Less than 1000 LOC to plumb Xen module into Windows and OSX
- Leverages all of the great feature development, optimization, broad testing that's done on Xen
  - Latest hardware support, SMP guests, PV drivers etc



# Xen Introspection API

- API to enable monitoring and control of VMs by a suitably privileged entity
  - CPU, memory, disk, network, etc
  - Enables Security, Forensics, Debugging, System Management
- Georgia Tech Xen Access library
  - Accessing memory, pagetable walking
- Shadow/HAP enhancements for trap-on access/write/execute
- UofAlaska/UC Davis VIX



# Observations

- Security is becoming increasingly important in virtualization
  - Xen as a true type-1 is well placed
  - Must continue disaggregation and de-privileging campaign
- Must continue to foster academic research on and using Xen
  - Community must help turn prototypes into production code



# Xen Releases and Roadmap

Keir Fraser



02/25/09

Xen Summit at Oracle Feb 24-25, 2009

# Release plans

Current stable releases: 3.2.3 and 3.3.1

Both released in early January

Next releases: 3.3.2 and 3.4.0

Both anticipated around Easter time

Ongoing strategy:

Maintain two stable branches until the later one has matured enough for switchover

Quarterly releases from stable branches

Six to nine months between major releases



# Memory management

More efficient heap allocators

- No separate 'xenheap'

- Better 'malloc'

Populate-on-demand HVM guest memory

- Boot an HVM guest with a big memory map

- But no need to allocate it all up front

- OS won't use much memory during boot

- And then balloon driver can claim large swathes



# Page sharing

Potential for reducing memory pressure by sharing identical pages across VMs

Significant savings in 'ideal' cases

Rather smaller gains in typical heterogeneous scenarios (10-20%)

How to find identical pages?

Memory scanning vs. disc block tracking

Allows memory overcommit

Hence requires demand paging

Or don't give spare pages directly back to guests



# Virtual block devices: blktap2

Support VHD storage format

Snapshot virtual disks

- Checkpoints, backups, gold images, etc

- Live coalescing of snapshots

Simplified kernel support

- Leverage blkback

Simpler invocation model

- More generic, easier test and debug



# Cross-hypervisor compatibility

Viridian interface

CPUID, hypercalls

Actually turn on just a few optimisations

And turn off some annoying checks

VHD format support



# High availability

Detect hardware failures

Hardware-based: Machine Check Architecture

Software-based: e.g., compare replicas

React appropriately...

CPU/memory offlining

Disable the offending hardware

Switch to a 'hot spare'

UBC's ongoing Remus project

Kemari developed at NTT Japan



# Network performance

Network virtualisation is particularly hard

- High packet rates; latency sensitive

Existing netfront/back drivers have limitations

- High cost for packet receive

- Not designed for next-generation NICs

Ongoing work on netchannel2 to address this

Lazy copy in the guest (reduces dom0 load)

- Provide guest a copy-only, sub-page, revocable grant

Support multi-queue NICs

- DMA directly to guest buffers

Reusable extensible ring architecture



# Security

## Deprivileged service domains

Qemu-dm, pvgrub, ...

## Secure boot

Measurement and containment

## Xen Introspection Project

Allow guest state to be monitored and dissected

Read memory, registers, etc

Callbacks when critical state is modified

Virus scanners, test/debug, ...



# Power management

## Range timers

fuzzy deadlines, allowing batched firing and fewer wakeups

## Selectable PM policy

Admin can choose governor to trade off power vs performance

## Smarter scheduling

Further work is ongoing: George Dunlap, Intel, etc



# Managing development

Use the developer mailing list

[xen-devel@lists.xensource.com](mailto:xen-devel@lists.xensource.com)

Requests for comments

Announcements of dev plans

Patches posted for comment, review and  
checkin

Use the wiki

<http://wiki.xensource.com/xenwiki/XenRoadMap>



# Welcome & Project Update

Ian Pratt

Keir Fraser



Xen Summit at Oracle Feb 24-25, 2009

# Attendee Profile

- 100+ Attendees
- 8+ Countries
- 36+ Companies
- 5+ Universities



Xen Summit at Oracle Feb 24-25, 2009



## Program Committee

- Dan Magenheimer – Oracle
- Jose Renato Santos – HP
- Jun Nakajima - Intel
- John Janakiraman – Skytap
- Alex Vasilevsky – Virtual Computer
- Victor Hugo dos Santos
- Sang-bum Suh - Samsung
- Hitoshi Matsumoto - Fujitsu

**Thanks for putting together a great event!**



Xen Summit at Oracle Feb 24-25, 2009

## Event Information

- Breakout Room Available During Event
  - Room 105
- See handout for Wireless Setup
- Lunch – Provided for Registered Attendees at Conference Center
- Abstracts & Speaker Profiles at



<http://www.xen.org/community/xensummit.html>



## Evening Event – Tuesday



6:30 pm – 9:30 pm

<http://www.computerhistory.org/>

- Dinner, Wine, Beer, and Sodas
- 2 Guides for Private Museum Tours
- Directions at Registration Table
- Sign-up Sheet for Carpooling



Xen Summit at Oracle Feb 24-25, 2009

# Xen Summit Gear

- Xen Summit Jackets
- Carabiners for Event Pass



Xen Summit at Oracle Feb 24-25, 2009



# Agenda Overview - Tuesday

9:30 – 10:00 am	Welcome and Project Status
10:00 – 10:30 am	Roadmap & Releases
10:50 – 11:10 am	Difference Engine
11:10 – 11:40 am	PVOPS Status
11:40 – 12:15 pm	Transcendent Memory on Xen
1:15 - 1:45 pm	Satori: Enlightened Page-Sharing
1:45 – 2:05 pm	Paravirtualized USB Support
2:05 – 2:35 pm	PCI-Pass Through Techniques
2:35 – 3:05 pm	Status of SR-IOV & VT-D
3:30 – 4:00 pm	Cross Vendor Migration
4:00 – 4:20 pm	Power Management in Xen
4:20 – 4:40 pm	Detecting and Correcting Transient Errors via Xen

Xen Summit at Oracle I/O 24-25, 2009

TM

## Agenda Overview - Wed

9:00 – 9:30 am	Open Nebula VM Manager
9:30 – 10:00 am	Client Virtualization Framework
10:00 – 10:30 am	Traifamadore
10:45 – 11:00 am	VM Snapshots
11:00 – 11:30 am	Real-time and VMM
11:30 – 12:00	Nested Page Tables
1:00 – 1:30 pm	Project Zentific
1:30 – 2:00 pm	Just Run It
2:00 – 2:30 pm	3D Rendering on Xen
2:30 – 3:00 pm	REFLINK Operation in ocfs2
3:30 – 4:00 pm	Achieving 10GB/s Paravirt Drivers
4:00 – 4:30 pm	Xen Scalability

Xen Summit at Oracle Feb 24-25, 2009



# Xen Summit Europe at LinuxTAG



Xen Summit at Oracle Feb 24-25, 2009



# Overview

- XenEmbedded
- XenClient
- HostedXen
- Xen Introspection API

Xen Summit at Oracle Feb 24-25, 2009



# XenEmbedded

- Small footprint Xen environment
- For embedding into Servers, Desktops,  
Laptops, Routers, Storage Arrays
  - buildroot/busybox/uclibc based
  - 8MB compressed image size
- xenvm/xenops embedded toolstack
  - Simple configuration file
  - Process per VM with control socket
- Fast booting
- See [xenbits.xen.org/xenclient](http://xenbits.xen.org/xenclient)



# XenClient

- Repository for bleeding-edge client related work, feeding in to xen-unstable
- Enhanced device pass-through
  - Intel, ATI, nVidia graphics
- Power conservation and suspend
- ACPI/SMBIOS virtualization
  - function keys, battery state, etc
- Secure mouse/keyboard routing
- Graphics virtualization
- Fast boot and measured launch
- Enhanced USB emulation
- “In-place P2V”, easy to install



## Hosted Xen

- Xen engine run as a kernel module to enable hosted (type-2) VMM
  - Runs on Windows and OSX today
- Enables Xen to also compete with other

type-2 VMMs

- KVM, VirtualBox, Parallels, VirtualPC, VMWare Workstation/ACE/Fusion
- Loses the security and isolation benefits of a true type-1 hypervisor, but

still useful in some scenarios



## Hosted Xen

- Xen loaded as an ELF module, linked via dispatch table into host kernel
- Less than 1000 LOC to plumb Xen module into Windows and OSX
- Leverages all of the great feature development, optimization, broad testing that's done on Xen
  - Latest hardware support, SMP guests, PV drivers etc



## Xen Introspection API

- API to enable monitoring and control of VMs by a suitably privileged entity
  - CPU, memory, disk, network, etc
  - Enables Security, Forensics, Debugging, System Management
- Georgia Tech Xen Access library
  - Accessing memory, page table walking
- Shadow/HAP enhancements for trap-on access/write/execute
- UofAlaska/UC Davis VIX



## Observations

- Security is becoming increasingly important in virtualization
  - Xen as a true type-1 is well placed
  - Must continue disaggregation and de-privileging campaign
- Must continue to foster academic research on and using Xen
  - Community must help turn prototypes into production code



# Xen Releases and Roadmap

Keir Fraser



02/25/09  
Xen Summit at Oracle Feb 24-25, 2009

## Release plans

Current stable releases: 3.2.3 and 3.3.1

Both released in early January

Next releases: 3.3.2 and 3.4.0

Both anticipated around Easter time

Ongoing strategy:

Maintain two stable branches until the later one has matured enough for switchover

Quarterly releases from stable branches

Six to nine months between major releases

# Memory management

More efficient heap allocators

No separate 'xenheap'

Better 'malloc'

Populate-on-demand HVM guest memory

Boot an HVM guest with a big memory map

But no need to allocate it all up front

OS won't use much memory during boot

And then balloon driver can claim large swathes

02/25/09  
Xen Summit at Oracle Feb 24-25, 2009



Amazon

- \* Mar 1999 HotOS paper on XenoServers
- \* XXX grant application
  - [lessons from the Nemesis experience - the importance of compatibility]
  - [accounting and billing at core (unlike PlanetLab)]
  - [speed of light, something fundamental]
  - [struggle for funding]
- \* Aug 2002 Xen development started
  - [pub bet]
- \* [2002 XenoServers project funded]
- \* Oct 2003 SOSP paper "Xen and the Art of Virtualization"
  - [could have stopped, interesting platform, keep going to build something real]
  - [Enterprise software is hard, respect for folk that make stuff work]
  - [resource revocation, 24x7 operation, benchmarks]
- \* Apr 2004 Xen 1.0 released
- \* Jun 2004 First Xen developer's meeting at OLS
- \* Nov 2004 Xen 2.0 released
- \* Dec 2004 XenSource formed
- \* working with CPU and IO hardware vendors
- \* RedHat, Novell, Sun adopt Xen in their OSes
- \* Microsoft and VMware adopt paravirtualization
- \* Sep 2006 XenEnterprise released
- [Oct 2007 XenSource acquired by Citrix Systems Inc]
- \* May 2008 XenServer 3.2 embedded in flash memory on Dell and HP servers
  - [proud, part of the platform, ubiquitous, back to the 1970's]

## Page sharing

Potential for reducing memory pressure by sharing identical pages across VMs

- Significant savings in 'ideal' cases

- Rather smaller gains in typical heterogeneous scenarios (10-20%)

How to find identical pages?

- Memory scanning vs. disc block tracking

Allows memory overcommit

- Hence requires demand paging

- Or don't give spare pages directly back to guests



02/25/09  
Xen Summit at Oracle Feb 24-25, 2009

Satori – Enlightened guests detect sharing opportunities and give up memory when sharing breaks

Difference Engine – all automatic

Tmem – a different plan for spare memory

## Virtual block devices: blktap2

- Support VHD storage format

- Snapshot virtual disks

  - Checkpoints, backups, gold images, etc

  - Live coalescing of snapshots

- Simplified kernel support

  - Leverage blkback

- Simpler invocation model

  - More generic, easier test and debug



# Cross-hypervisor compatibility

Viridian interface

CPUID, hypercalls

Actually turn on just a few optimisations

And turn off some annoying checks

VHD format support



## High availability

Detect hardware failures

Hardware-based: Machine Check Architecture

Software-based: e.g., compare replicas

React appropriately...

CPU/memory offlining

Disable the offending hardware

Switch to a 'hot spare'

UBC's ongoing Remus project

Kemari developed at NTT Japan



Xen Summit at Oracle Feb 24-25, 2009

MCA – AMD, Intel, Sun

Replication – John Byrne

# Network performance

Network virtualisation is particularly hard

- High packet rates; latency sensitive

Existing netfront/back drivers have limitations

- High cost for packet receive

- Not designed for next-generation NICs

Ongoing work on netchannel2 to address this

Lazy copy in the guest (reduces dom0 load)

- Provide guest a copy-only, sub-page, revocable grant

Support multi-queue NICs

- DMA directly to guest buffers

Reusable extensible ring architecture



# Security

## Deprivileged service domains

Qemu-dm, pvgrub, ...

## Secure boot

Measurement and containment

## Xen Introspection Project

Allow guest state to be monitored and dissected

Read memory, registers, etc

Callbacks when critical state is modified

Virus scanners, test/debug, ...



# Power management

## Range timers

fuzzy deadlines, allowing batched firing and fewer wakeups

## Selectable PM policy

Admin can choose governor to trade off power vs performance

## Smarter scheduling

Further work is ongoing: George Dunlap, Intel, etc



# Managing development

Use the developer mailing list

[xen-devel@lists.xenproject.org](mailto:xen-devel@lists.xenproject.org)

Requests for comments

Announcements of dev plans

Patches posted for comment, review and  
checkin

Use the wiki

<http://wiki.xenproject.org/wiki/XenRoadMap>

