



Xen Development Roadmap

Keir Fraser / Ian Pratt

- Development model
 - Removing Linux sparse tree
 - Target systems
 - ABI Stability
- Previous roadmap scorecard
- New Roadmap

- Aim: stabilize “unstable” tree every ~10 wks
 - Sweep unstable into 3.0.x-testing
 - Release as 3.0.1, 3.0.2 etc
- But, often a couple of key features
 - (3.0.5 XenAPI and HVM live relocation)
- Bug fixes cherry picked into 3.0.x-testing at least until next release
 - After being in unstable for a few days, requests to push into 3.0.x-testing
 - Release as 3.0.x-y etc.

- Linux 'sparse tree' removed from xen-unstable post 3.0.5
- Makefile support for building external Linux trees
 - Linux 2.6.18
 - Linux last release
 - Linux tip
- Natural progression to paravirt_ops
 - Dom0 support

- 1-4 socket systems are the priority
- Performance and scalability work
 - Optimizations for bigger systems must not hurt smaller ones (they often help)
 - Onus is on submitter to demonstrate
 - (Patches that clearly hurt larger systems should be rejected too)
- Good performance tools now available
 - s/w perf counters, xen oprofile, xentrace etc

- Guest ABIs are stable (hypercall, IO)
 - Backward guaranteed by Xen:
 - Old 3.0 guests must run on new xen
 - Forward compatibility provided by guest
 - E.g. Linux build option (default on)
- Privileged domain hypercall API
 - Dom0 kernel ABI stable
 - Tools hypercall API still evolving
- Xen API control protocol and CLI syntax
 - Version 1.0 in 3.0.5.
 - XML-RPC with language bindings for C and Python

- Last roadmap published July 2006
- 4 priorities listed:
 - 1. end Q3 2006
 - 2. end Q4 2006
 - 3. end Q1 2007
 - 4. beyond...

Roadmap Scorecard #1



Prii	Area	Description
1	tools	xend VM life-cycle management
2	tools	XML cong file and conversion tools
2	tools	standardized xen control API: xml-rpc over https/unixdomain sockets
2	tools	C++/perl/python bindings for control API
2	tools	simple storage management in xend
3	tools	revive guest coredump support
3	tools	split VM relocation operation into two parts and authenticate
3	tools	DMTF CIM providers
4	tools	Web GUI for Xen
1	storage	blktap (or other) support for file-based virtual disk storage.
2	storage	blktap plugins for common formats
2	storage	optimized qcow implementation
3	storage	consider adding write accounting/throttling on current loop driver
3	storage	support for block IO QoS. Use CFQ and ionice, or implement in blkback
4	storage	'ioctl' support between blkfront/back
4	storage	media change, size change event propagation to guest userspace
4	storage	consider SCSI level storage virtualization option

Roadmap Scorecard #2



Pri	Area	Description
2	network	TCP Segmentation Offload support in device channel
2	network	checksum offload cleanup
2	network	hypervisor chooses to copy vs. page flip
2	network	dynamic allocation of grant table entries; grant table resize
2	network	investigate whether bridge code needs to be 'streamlined'
2	network	jumbo frames support in dom0 and device channel
4	network	investigate static shared buffer approach
4	network	TCP Offload Engine support in device channel
4	network	investigate high-performance point-to-point link support
4	network	RDMA support in device channel

Roadmap Scorecard #3



Pri	Area	Description
1	xen	extensive benchmarking and perf tuning
1	xen	CPU scheduler that balances VCPUs, implements weight & caps
2	xen	initial NUMA mechanism checkin
2	xen	live relocation tuning, robustication, tools safety interlock
3	xen	support for running 32b PAE guests on a 64b hypervisor
3	xen	improved NUMA policy code
3	xen	add order>0 guest memory allocation accounting
3	xen	extend x86 64 heap size; merge xen and domain pools
3	xen	investigate bad pre-emption avoidance/mitigation strategies
3	xen	add superpage support for PV guests
3	xen	IOMMU support: isolation of devices to domains; grant table integration
4	xen	lazy memory allocation for live relocation of ballooned guests
4	xen	fine-grained delegation for dom0ops; hierarchical resource model
4	xen	power management enhancements: CPU sleep, freq scaling
4	xen	power management enhancements: suspend/hibernate
4	xen	accounting and billing time IO domains spend on behalf of guests

Roadmap Scorecard #4



Pri	Area	Description
1	hvm	fix current shadow pagetable code, add PAE-on-PAE mode, SMP
1	hvm	upgrade QEMU version, maintain as a patch queue
2	hvm	rewrite shadow pagetable code to optimize, simplify
2	hvm	finalize interface for making hypercalls from VT guests
2	hvm	HVM save/restore support; qemu, xen, and tools changes
2	hvm	basic SMP HVM guest support; ACPI tables, locking safety
3	hvm	change QEMU-xen interface to use the 'v2e' approach
3	hvm	SMP HVM guest performance and scalability
3	hvm	support real superpage mappings for HVM guests
3	hvm	implement high-performance SCSI HBA emulation
3	hvm	implement high-performance Ethernet emulation
3	hvm	live relocation. Add log-dirty support
3	hvm	move QEMU into a 'stub domain' linked against a linux kernel
4	hvm	move QEMU into a 'stub domain' linked against a minios
4	hvm	HVM hotplug CPU emulation

Roadmap Scorecard #5



Pri	Area	Description
1	linux	extensive benchmarking and perf tuning
2	linux	SMP scalability improvements
2	linux	work to get xen port in to kernel.org linux
2	linux	code review of x86 64 port
2	linux	investigate proposed x86 64 optimizations
2	linux	improve interaction between balloon driver and page allocator
3	linux	support for multiple virtual serial consoles
3	linux	consider hybrid round-robin/priority scheme to service events
2	client	basic kernel fbdev paravirtual framebuffer implementation
3	client	USB virtualization; investigate USB-over-IP code
3	client	Xserver support for 'h/w cursor', copy rect and fill rect
4	client	OpenGL/Direct3D virtualization

Roadmap Scorecard #6



Pri	Area	Description
3	misc	support for dom0 kexec/kdump to get a machine core
3	misc	infiniband direct guest IO support
3	misc	support for hiding CPU feature flags from guests (PV and HVM)
4	misc	smart NIC direct guest IO support
4	misc	submit xen for scanning by Coverity tool; investigate warnin
4	misc	tools support for doing auto CPU/memory resource allocation
4	misc	support to checkpoint/rollback guests
4	misc	port Grub2 bootloader to net/blockfront devices
4	misc	investigate 'pluggable driver architecture'
4	misc	xenfs filesystem-level virtualization; shared buffer cache
4	misc	do we need support for ISA/PCMCIA DMA (below 16MB)?

- Performance and scalability enhancements
- Client features (laptop/desktop)
- I/O improvements
- HVM enhancements
- Control APIs
- Security

- Ongoing profiling and optimization
- MSI Support
- Scheduler measurement and tuning
 - Wakeup latencies
 - Load balancing
- Locking granularity
 - Remove domain lock from hot paths
- Super page support
 - Need per-guest accounting

- Power management
 - CPU Frequency and voltage scaling
 - Xen hibernate and suspend to RAM
- IOMMUs
 - Device passthrough for HVM guests
- Graphics virtualization
 - Paravirtual framebuffer
 - Direct3D, OpenGL virtualization

- Virtual ACPI S3 support
 - suspend to RAM
- Stub domains
 - Better scalability for I/O emulation
 - Minimal Linux integrated tightly with qemu-dm
- Extend and use `x86_emulate`
 - For real-mode emulation on VT-x
 - To replace the ad-hoc mmio decoder
- SCSI boot BIOS support

- PV removable media support
- NetChannel2 protocol
- Support for Smart NICs
- SCSI front/back
 - CD, tape etc.
- XenSocket stream transport
- USB front/back

- XenAPI extensions
 - ACM/XSM management
 - Enhanced storage management
 - Live relocation interlocks
 - Improved log management
- Continue CIM provider implementation
 - Track DMTF standard

- Guest interface fuzzing tools
- Dom0 disaggregation
 - Reduce size of TCB by moving privileged operations into isolated VMs
 - Driver domains
 - Bulk of dom0 becomes deprivileged
- Mandatory access control
 - ACM and XSM
- Fine-grained hierarchical delegation

- Need to incorporate information from summit and prepare a new Roadmap document to go on the wiki
- Live document
 - Link through to project update pages on wiki

Thanks to IBM!



- Many thanks to IBM for a great venue, and for LTC/Corporate sponsorship
 - Jennifer Hall - local arrangements
 - Mike Day
 - Warren Grunbok II
 - Ray Valdez
 - Reiner Sailer
 - Ron Perez - WiFi
 - Hollace Goodman - badges
 - Terri Reinhart and Aimee Francioni from XenSource

- Social event at Ottawa Linux Symposium in July
- Proposal for XenSummit in Cambridge in September, or possibly West Coast US later in October

- IA64
- Source tree organization