

Oracle Open Source Software

PVH : PV Guest in HVM container

Mukesh Rathor
Oracle Corporation

August 2012

Xen Summit NA 2012

Oracle Open Source Software

HVM : Hardware Virtual Machine

- Provides protected environment
- Guest kernel can run in any ring
- VMEXITs transfers control to xen



Oracle Open Source Software

Why Need for PVH

- 64bit runs in ring 3 because of no segmentation
- System calls are slow as they are bounced from xen to guest and back



Oracle Open Source Software

PVH: Salient Features

- Runs in ring 0
 - Uses the PV entry point, thus skipping BIOS emulations by qemu
 - Boots faster
 - Uses event channel, so no APIC emulation
 - Native Page tables (HAP required)
 - Native IDT
-
-

Oracle Open Source Software

PVH: Salient Features (*contd.*)

- Uses lot of HVM code paths, thus reducing the xen specific PV code in linux
 - Is a PV guest tho, and `xen_hvm_domain()` would be false.
 - PV guest from xen perspective also. Thus, `is_hvm_domain()` will be false.
-
-

Oracle Open Source Software

Design Details

- The guest is populated with pfns instead of mfns
 - HAP is setup during guest creation for pfn -> mfn mappings
 - Kernel CS/DS are setup for ring 0
 - IO space is mapped 1:1 in the HAP.
-
-

Oracle Open Source Software

Performance data (LMBENCH):
(First set on PV dom0, second on PVH dom0)

Processor, Processes - times in microseconds - smaller is better

```
-----  
Host          OS  Mhz null null      open slct sig  sig  fork exec sh  
              call I/O  stat clos TCP  inst hndl proc proc proc  
-----  
PVH           Linux 3.5.0-m 2631 0.13 0.23 1.83 3.84 3.02 0.24 3.29 170. 512. 1680  
PV            Linux 3.5.0-m 2631 1.08 1.20 3.56 7.24 4.10 1.10 3.52 565. 1392 3634  
HVM           Linux 3.5.0-m 2631 0.13 0.22 1.61 2.77 3.09 0.24 3.74 129. 394. 1399  
  
PVH           Linux 3.5.0-m 2631 0.13 0.23 1.87 3.97 3.02 0.24 3.29 158. 492. 1632  
PV            Linux 3.5.0-m 2631 0.71 0.93 3.27 6.15 3.82 0.87 3.00 566. 1322 3369  
HVM           Linux 3.5.0-m 2631 0.13 0.22 1.60 2.78 3.09 0.24 3.77 125. 394. 1395
```

Oracle Open Source Software

Performance data (LMBENCH):
(First set on PV dom0, second on PVH dom0)

Context switching - times in microseconds - smaller is better

Host	OS	2p/0K ctxsw	2p/16K ctxsw	2p/64K ctxsw	8p/16K ctxsw	8p/64K ctxsw	16p/16K ctxsw	16p/64K ctxsw
PVH	Linux 3.5.0-m	4.2700	4.8200	5.2400	5.5700	7.2200	6.13000	7.37000
PV	Linux 3.5.0-m	5.9500	6.0500	5.4900	6.6400	7.6900	7.10000	7.69000
HVM	Linux 3.5.0-m	1.6900	1.7500	2.1300	2.3500	3.8200	2.92000	3.89000
PVH	Linux 3.5.0-m	3.5200	3.7200	4.1400	4.6800	6.2500	5.29000	6.32000
PV	Linux 3.5.0-m	4.9600	5.1400	4.7900	5.7400	7.1500	6.18000	7.16000
HVM	Linux 3.5.0-m	1.8800	1.9200	2.2700	2.4100	3.8500	2.94000	3.89000

Oracle Open Source Software

Performance data (LMBENCH):
(First set on PV dom0, second on PVH dom0)

File & VM system latencies in microseconds - smaller is better

Host	OS	0K File	10K File	Mmap	Prot	Page	100fd
		Create	Delete	Latency	Fault	Fault	selct
PVH	Linux 3.5.0-m			7814.0	0.275	1.05700	1.375
PV	Linux 3.5.0-m			30.5K	1.181	3.72670	2.350
HVM	Linux 3.5.0-m			7277.0	0.173	0.86030	1.454
PVH	Linux 3.5.0-m			7532.0	0.309	1.05870	1.373
PV	Linux 3.5.0-m			27.5K	1.048	3.38740	2.107
HVM	Linux 3.5.0-m			7512.0	0.205	0.88430	1.456

Oracle Open Source Software

Not Done Yet

- Performance fine tuning
 - Goal is to have PVH perform no worse than PV or HVM.
 - Many optimizations, like delivering interrupt directly to dom0, scheduler change for PVH VMCS affinity, etc..
 - FIXME/TBD in the code for certain features (vcpu placement, event channel EOI map, etc..)
-
-

Oracle Open Source Software

Not Done Yet (*Contd.*)

- HVM cacheatter, mtrr, tsc, etc.. not sure if I got them right. Need to study and investigate more. *Phase II.*
- FPU usage might be incomplete.



Oracle Open Source Software

The End

