Virtio and kvm networking
status update and plans

2013

Michael S. Tsirkin
Red Hat

VIRTIO / VHOST
KVM NETWORKING
OASIS Virtio TC

Virtio 1.0

- PCI
- MMIO (ARM)
- CCW (PPC)
Virtio 1.0

- Virtio PCI:
  - Replace Port IO with Memory mapped IO
  - PCI Express (hotplug, AER, multi-root, SRIOV)
  - Infinite features

- Reduced memory requirements

- Fixed endianness

- Compatibility
## Port vs Memory mapped IO

<table>
<thead>
<tr>
<th></th>
<th>Port IO</th>
<th>MM IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandated for PCI Express</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Portable</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>HW Virtualization</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Fast on x86</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
Port IO versus memory mapped IO on KVM x86: cycles per access (lower is better)
Port IO: outl

VM Exit

EF
OUT

notify
(%DX)

VQ#
%EAX

REASON
QUALIFICATION
STATE
Memory mapped IO: writel

89 MOV (%EDI) 3E %RSI

PTE VALID?

VM Exit

REASON GUEST ADDRESS RIP
Fast MMIO

VM Exit

MOV

notify

VQ#

(%EDI)

%RSI

PTE

VALID?

REASON

GUEST ADDRESS
Access times on KVM x86: Cycles per access (lower is better)
Multiple interfaces

- CAPABILITY LIST
- IO BAR
- MEMORY BAR
- VIRTIO CAPABILITY #1
- VIRTIO CAPABILITY #2
Memory requirements

0.9

VQ → desc | avail | used

1.0

VQ → desc | avail | used
features

DEVICE FEATURES

0
1
1
-

DRIVER FEATURES

0
1
-

DRIVER

SEL
1

2
3
4
.....

0...

....

....

....

....

STATUS = FEATURES_OK
Endianness

Virtio 0.9

Virtio LE

Device LE

intel

Virtio BE

Device BE

PPC

Virtio 1.0

Virtio LE

Device

Device

Device

Device
compatibility

0.9

Driver

Device

1.0

Driver

Device

compatibility

✓

✓
Packet layout

Virtio 0.9

INDIRECT → next → header

Virtio 1.0

header
Packet layout: transactions per sec (higher is better)
More: virtio 1.0 versus 0.9.5

- Virtio 9p
- Virtio blk: WCE
- Virtio-net Multiqueue
- Virtio-net dynamic offloads
- Already upstream (based on spec draft)
vhost updates

- Vhost scsi
- Vhost-net zero copy transmit
- No need for driver changes
Kvm networking

- Openvswitch – if time allows
- Ethernet bridge
Bridge FDB

<table>
<thead>
<tr>
<th>London</th>
<th>Heathrow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>CDG</td>
</tr>
</tbody>
</table>

uplink

London

Heathrow

Paris

CDG
Flood: DOS potential

<table>
<thead>
<tr>
<th>London</th>
<th>Heathrow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>CDG</td>
</tr>
</tbody>
</table>

Diagram showing connections between London, Heathrow, and Paris with an uplink to Bangkok.
Disable flood

<table>
<thead>
<tr>
<th>London</th>
<th>Heathrow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>CDG</td>
</tr>
</tbody>
</table>

uplink

London

Bangkok

Paris

Heathrow

CDG

London

Paris

uplink
softmac

- `ifconfig eth0 hw ether 00:12:23:45:67:89`
Using softmac/non promiscuous

<table>
<thead>
<tr>
<th>London</th>
<th>Heathrow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>CDG</td>
</tr>
</tbody>
</table>

HEATHROW

CDG

NEW

uplink
Work in progress

- ELVIS (vhost blk/vhost net)
- Virgl
- Vhost-net performance
RX latency

NIC

HOST

VHOST

VM
Fast rx

NIC

HOST

VHOST

VM

Current?

RAM?
Fast rx: transactions per sec (higher is better)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>331668</td>
</tr>
<tr>
<td>Miss</td>
<td>79</td>
</tr>
</tbody>
</table>
Vhost-net threading

<table>
<thead>
<tr>
<th>tap</th>
<th>RX</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHOST</td>
<td></td>
</tr>
<tr>
<td>VM</td>
<td>TX</td>
</tr>
</tbody>
</table>

NIC
VHOST
VM
Vhost-net thread pool

tap
VM

NIC
VM

WQ
VHOST
VHOST
threading: UDP RR
transactions/sec (higher is better)
threading: TCP STREAM transactions/sec (higher is better)
summary

- Performance
- Manageability
- Security
Questions?
OVS: flow match

 PACKET FLOW

192.68.0.1 22

192.68.0.1 12865

22 12865

VM

kernel

userspace

OVS-VSWITCHD
OVS: wildcard match

22
12865

PACKET FLOW
192.68.0.1 *

VM

kernel

userspace

OVS-VSWITCHD
Wilcard: netperf CRR (higher is better)