Real-time Virtual NIC on KVM for Real-Time Network with OpenFlow

Kenichi Suzuki
Department of Computer and Information Sciences
Tokyo University of Agriculture and Technology

LinuxCon Japan 2013
Background

- Increasing use of OpenFlow network
  - Virtual Machines (VMs) are often used for network nodes
- Virtual Network Interface Controllers (vNIC) are used for network communications
  - Virtual Machine Monitor (VMM) virtualizes a physical NIC and provides vNIC for Guest OS

Real-time communication in OpenFlow and VMM environment e.g. voice, audio, video and control traffic...
Background

• Issues
  ▫ Real-time and non-real-time traffic are treated in a same way
  ◦ Special control is required for real-time communication

OpenFlow and VMM should support priority based traffic-control for real-time ability
Goal

Proposing the “RTvNIC System” for real-time communication between VMs by collaboration of KVM and OpenFlow

- Provides Real-Time Virtual NIC (RTvNIC) on KVM
  - KVM provides a special vNIC for real-time communication between VMs without modification of Guest OS
- QoS control for OpenFlow network with KVM
  - OpenFlow is a programmable network
  - KVM controls the bandwidths of the real-time communication path on OpenFlow network dynamically
Basic Concept

“Real-time Communication Controller (RCC)” on KVM

- **RTvNIC control**
  - Packets from RTvNIC are processed preferentially for QoS.
  - packet scheduling, deadline detection, etc..

- **OpenFlow control**
  - The bandwidths of OpenFlow network are adjusted dynamically for RTvNIC packets

![Diagram showing the interaction between RTvNIC, OpenFlow network, and Guest OS on KVM]
Basic Concept

• **What is RTvNIC?**
  - Special vNIC supporting real-time communication between VMs

• Special packet-processing for RTvNIC
  - *Real-time information* in each packet
    - DSCP value for DiffServ
    - Time information for detecting deadline
  - Packet scheduling (EDF scheduling)
  - Detecting a packet deadline miss
Basic Concept

- Special Rate-control for OpenFlow network
  - OpenFlow prepares the bandwidth for RTvNIC
  - KVM requests OpenFlow network to re-optimize the rate when deadline miss occurs.

Enhance real-time communication in OpenFlow Network
Design of RTvNIC System

I. RTvNIC Control
   ▫ vNIC Handling
     • Attach real-time information
     • Apply EDF Scheduling
   ▫ Deadline Miss Handling
     • Detect deadline miss

II. OpenFlow Control
   ▫ Management of network QoS
     • Rate control in OpenFlow network
vNIC Handling

- EDF Scheduling
  - Take a packet out on EDF priority based form RTvNIC queues to keep the deadline

- Adding real-time information
  - Add DSCP value to control packets from RTvNIC in OpenFlow Network
  - Add the time of packet sent for Deadline Miss Handling
vNIC Handling

- Removing real-time information
  - Remove real-time information from a real-time packet for Guest OS to receive the packet correctly

Deadline Miss Handling

- Deadline miss detection
  - Detect by comparing real-time information with the time of packet received

This result is used by OpenFlow control
Deadline Miss Handling

• Detects deadline miss by using real-time information of a packet
  ▫ If deadline miss is detected, OpenFlow Control notify deadline miss to OpenFlow Controller in order to avoid next deadline miss

• Information used deadline miss handling
  ▫ The time of packet sent : Ts
  ▫ The time of packet received : Tr
  ▫ The deadline registered in the RTvNIC : Td
OpenFlow Control on KVM

- **QoS in OpenFlow network**
  - OpenFlow network is controlled by real-time information (DSCP) based flow control
  - Packets from RTvNIC are preferentially transmitted than other packets

- **Management of OpenFlow**
  - When communication between RTvNICs occur deadline miss, to change QoS in OpenFlow network dynamically

- **Main information to communicate to OpenFlow Controller**
  - Information for identifying a path detected deadline miss
    - source and destination IP address of deadline miss packet
  - **Deadline miss information**
    - Exceeded time of deadline (Tr – Ts - Td)

OpenFlow Controller guarantees real-time communication by changing the priority or route of a target path
OpenFlow Control on KVM

- Initialization of real-time communication
  - When KVM starts VM that used RTvNIC, KVM requests assuring real-time path from OpenFlow network

  ➢ Send following information to OpenFlow Controller
    - Destination and source IP address, minimum rates, deadline level
  ➢ If network resources are in short supply, the request is failed

![Diagram of OpenFlow Control on KVM]

- QEMU-KVM
- RTvNIC
- OpenFlow Network
- OpenFlow Controller (Network Management)
- OF SW
- RCC
- Linux/KVM
- Hardware
- NIC
- Guest OS
- Guarantee network QoS between VM
Implementation

- Implementation of adding and removing real-time information, EDF scheduling and deadline miss detection
  - Linux Kernel: 3.4.10 + CONFIG_PREEMPT_RT Patch
  - VMM: kvm-kmod 3.6 + qemu-kvm 1.2.0
Implementation of vNIC handling

- Adding real-time information (DSCP and deadline time)
  - Implemented on vNIC(ne2000_pci) in qemu-kvm
    - In ne2000_ioport_write()
    - Insert deadline time between data field and FCS of Ethernet frame
Implementation of vNIC handling

- EDF packet scheduling
  - Implementing EDF Scheduling in Queuing Disciplines of Traffic Control of the Kernel
    - `linux/net/sched/sch_generic.c`
  - Priority is determined by the following equation:
    \[
    \text{Priority} = \text{Delay time of packet in queue} - \text{Deadline time registered in RTvNIC}
    \]
Evaluation Environment

- **VMM**
  - Host OS: Fedora 17
  - Kernel: 3.4.10 + CONFIG_PREEMPT_RT
  - VMM: kvm-kmod 3.6 + qemu-kvm 1.2.0
  - Guest OS: Fedora 17

- **OpenFlow Controller**
  - Trema 0.2.4 on Ubuntu 12.04

- **OpenFlow Switch**
  - Open vSwitch 1.7.1 with Traffic Control on Fedora 17
Evaluation

• Evaluation of worst-case delay
  ▫ Worst-case delay is one of the important factors to guarantee real-time communication

• Measure delay by ping packet in congested network
  ▫ Round-Trip Time of ping(RTT)
  ▫ Delay of EDF packet scheduling(A)

• Compare RTvNIC System with No RTvNIC System
Evaluation
Results

<table>
<thead>
<tr>
<th></th>
<th>RTT</th>
<th>RTvNIC</th>
<th>No RTvNIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave. (ms)</td>
<td>3.8</td>
<td>20.0</td>
<td></td>
</tr>
<tr>
<td>Max. (ms)</td>
<td>16.0</td>
<td>120.0</td>
<td></td>
</tr>
<tr>
<td>Min. (ms)</td>
<td>2.7</td>
<td>13.0</td>
<td></td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>RTT</th>
<th>RTvNIC (us)</th>
<th>No RTvNIC (us)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave.</td>
<td>53</td>
<td>400</td>
</tr>
<tr>
<td>Max.</td>
<td>590</td>
<td>6800</td>
</tr>
<tr>
<td>Min.</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

- Queuing delay: A: Delay of packet scheduling
- RTT of ping
- Send
- Receive
- Make response packet
- Send

Table:

<table>
<thead>
<tr>
<th>RTT</th>
<th>RTvNIC (ms)</th>
<th>No RTvNIC (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave.</td>
<td>3.8</td>
<td>20.0</td>
</tr>
<tr>
<td>Max.</td>
<td>16.0</td>
<td>120.0</td>
</tr>
<tr>
<td>Min.</td>
<td>2.7</td>
<td>13.0</td>
</tr>
</tbody>
</table>
Discussion

• Reduction of worst-case queuing delay in VMM from 6800us to 590us
  ➢ EDF scheduling can prevent delay in VMM

• Reduction of worst-case delay round-trip time from 120.0ms to 16.0ms
  ➢ RTvNIC system can prevent worst-case end-to-end delay by real-time information (DSCP) in OpenFlow network

RTvNIC System can handle real-time communication
Conclusion

• The RTvNIC System
  ▫ prevented the worst-case queuing delay in VMM by EDF packet scheduling
  ▫ prevented the worst-case round-trip time in OpenFlow network and VMM

Future works

• Implementation of cooperation between KVM and OpenFlow network
  ▫ Management of network QoS when deadline miss occurs
• Evaluation with Actual Applications
  ▫ Video communication etc...
Real-time Virtual NIC on KVM for Real-Time Network with OpenFlow

Kenichi Suzuki
Department of Computer and Information Sciences
Tokyo University of Agriculture and Technology