Virtualizing The Network (Project Crossbow)

Ryan Matteson
matty91@gmail.com
http://prefetch.net

Overview

- Tonight I am going to discuss the OpenSolaris network virtualization project (Crossbow)
- I plan to split my presentation into 3 parts:
  - Part 1 will provide an overview of Crossbow
  - Part 2 will show you how to use the technology
  - Part 3 will be a Q&A period

Virtualization

- Virtualization has reached the mainstream, and most companies are now using one or more virtualization technologies (e.g., Xen, VMWare, Solaris zones, KVM, Linux vservers, OpenVZ, etc.)
- Deploying virtualization allows companies to better utilize server platforms (they keep getting faster and faster don't they?), but ensuring SLAs are met becomes a bit more difficult when multiple guests are running on the same piece of hardware
- To meet SLAs, enforcing QOS at the CPU, memory, I/O and network layers is becoming a core requirement

Network Virtualization

- Most of the major Operating Systems provide a way to partition CPU resources and limit how much memory is available to guest operating systems, but they rarely focus on limiting how guests are using the network
- Wouldn’t it be cool if the Operating System allowed you to segment network resources and allocate them to guests in the same way that you can partition memory and CPU resources?
- Wouldn’t it also be cool if you could add QOS measures to ensure that https traffic is prioritized over quake traffic?

Enter Project Crossbow

- Project Crossbow provides the management tools and kernel plumbing to create virtual networks, and to create and enforce network QOS policies
- Utilizes advancement in network interface technology (e.g., hardware classifiers, RX/TX ring partitioning, multiple DMA channels, etc.) to maximize performance
- Crossbow is comprised of three main technologies:
  - Ethernet stubs (i.e., virtual switches)
  - Virtual NICs (i.e., virtual Ethernet interfaces)
  - QOS policies
- Crossbow was integrated into Nevada build 105 in December of 2008, so you can download and use the technology today!

Ethernet Stubs

- Ethernet stubs are virtual network switches, and act just like a real Ethernet switch (you won’t have to submit a PO to use them!)
- You can create Ethernet stubs with the dladm utilities “create-etherstub” option:

```
$ dladm create-etherstub switch0
```

Ethernet Stubs (cont.)

- The dladm utilities “show-etherstub” option can be used to display the Ethernet stubs that have been created:

```
$ dladm show-etherstub
LINK
switch0
switch1
```

Virtual NICs

- Virtual NICs (VNICs) are virtual network interfaces that are layered on top of a physical interface or Ethernet stub, and act just like a real network interface
- VNICs are managed identically to physical interfaces, and provide all the capabilities (i.e., DHCP, snoopable, traffic is isolated from other NICs, etc.) that a real interface does
- The dladm utilities “create-vnic” subcommand can be used to create virtual NICs on top of a physical interface:

```
$ dladm create-vnic –l e1000g0 vnic1
```

- You can also connect a virtual NIC to an Ethernet stub by specifying the Ethernet stub instead of a physical interface:

```
$ dladm create-vnic –l switch0 vnic1
```

Virtual NICs (cont.)

- The dladm utilities “show-vnic” subcommand can be used to display virtual NIC information:

```
$ dladm show-vnic
 LINK       OVER       SPEED  MACADDRESS        MACADDRTYPE      VID
 vnic0      switch0      0      2:8:20:d3:a8:6a   random            0
 vnic1      e1000g0     0      2:8:20:fb:7a:9a   random            0
```

- In the example above, we can see that two VNICs exist on the system, one VNIC is connected to an Ethernet stub, one VNIC is attached to a physical interface, each interface is configured with a randomly generated MAC address, and both interfaces are in VLAN 0 (the default VLAN)

Network Resource Controls

- Crossbow provides several network resource controls that can be applied to physical and virtual interfaces:
  - Bandwidth limits (maxbw setting)
  - Relative priorities (priority setting)
  - CPU bindings for traffic processing (cpu setting)
- Additionally, Crossbow allows you to enable resource controls for individual traffic flows (e.g., prioritize traffic on port 80 over quake traffic)

Enforcing Bandwidth Controls

- The dladm utilities “set-linkprop” subcommand can be used to configure QOS policies:

  ```
  $ dladm set-linkprop -p maxbw=10m vnic0
  ```

- In the example above, network bandwidth will be capped at 10mb/s for vnic0

Displaying Resource Control Settings

- The dladm utilities “show-linkprop” subcommand can be used to display the QOS settings for virtual and physical interfaces:

```
$ dladm show-linkprop  vnic0
LINK      PROPERTY     PERM VALUE      DEFAULT     POSSIBLE
vnic0     autopush      --    --         --          --
vnic0     zone         rw     --         --          --
vnic0     state        r-     unknown   up          up,down
vnic0     mtu          r-     9000      1500        --
vnic0     maxbw        rw     10         --         --
vnic0     cpus         rw     --         --         --
vnic0     priority     rw     high       high       low,medium,high
```

- In the example above, we can see that bandwidth is capped at 10Mb for vnic0, the NIC priority is set to high, and the virtual NIC hasn’t been bound to a specific set of CPUs

Network Flows

- Crossbow also allows you to assign network resource controls to individual traffic flows
- Traffic flows can consists of:
  - Source and destination addresses
  - TCP and UDP port numbers
  - Header flags
- Network resource control (e.g., maximum amount of bandwidth that can be used, traffic priority, etc.) can be attached to flows, allowing find grained network QOS policies to be created

Creating Network Flows

- Flows are created with the flowadm utilities “add-flow” subcommand:

```
$ flowadm add-flow -l vnic0 transport=tcp,local_port=80 httpflow
```

- Flows can be displayed with the flowadm utilities “show-flow” subcommand:

```
$ flowadm show-flow
FLOW       LINK     IPADDR                   PROTO  PORT    DSFLD
httpflow   vnic0     --                      tcp     80       --
```

Adding Resource Limits to Flows

- Once a flow is created, you can limit the maximum amount of bandwidth available to the flow, adjust the priority of traffic matching the flow, and bind the processing of traffic that matches the flow to one or more CPUs
- To cap the maximum bandwidth for the flow named httpflow, the dladm utility can be run with the “set-flowprop” subcommand, the “-p” option and the maxbw keyword, a maximum bandwidth value (which is expressed in K(bps), M(bps) or G(bps)) and the name of the flow to modify:

```
$ flowadm set-flowprop -p maxbw=5M httpflow

$ flowadm show-flowprop -l vnic0
FLOW       PROPERTY     VALUE     DEFAULT     POSSIBLE
httpflow   maxbw         5        --          5M
httpflow   priority     high      --          high
```

Monitoring Flow Usage

- Once flows are configured, you can use the flowadm utilities “show-flow” subcommand along with the “-s” option to view flow usage statistics:

```
$ flowadm show-flow -s -i 1
FLOW        IPACKETS  RBYTES    IERRORS  OPACKETS  OBYTES    OERRORS
httpflow    278891    19754223 0        232390    29558178  0
httpflow    5551      393179   0        4626      588354    0
httpflow    5616      397800   0        4680      595296    0
httpflow    5664      401200   0        4720      600384    0
httpflow    5532      391850   0        4610      586392    0
httpflow    5532      391850   0        4610      586392    0
```

Flow Usage Accounting

- The OpenSolaris extended accounting facility can be used to capture flow statistics over longer durations, which can be useful for metering and billing customers
- Extended accounting can be enabled with the acctadm command:

```
$ acctadm -e extended -f /var/log/net.log net
```

Flow Usage Accounting (cont.)

- Once extended accounting is enabled, the flowadm “show-usage” subcommand can be used to display flow statistics:

```
$ flowadm show-usage -f /var/log/net.log
FLOW      DURATION  IPACKETS  RBYTES    OPACKETS  OBYTES    BANDWIDTH
httpflow  1620      513064    36337108  427407    54349626  0.447 Mbps
```

Putting It All Together

- Crossbow's utility really shines when it is combined with virtualization technologies such as Solaris zones or Xen
- This combination allows for a number of awesome things:
  - Limiting bandwidth available to virtual machines
  - Adding priorities to specific network protocols
  - Delegating network administration
  - Billing customers for network usage

Configuring Zones To Use VNICs

- The following zonecfg example shows how to configure a zone to use the virtual NIC (vnic1) that was created a few slides back:

```
$ zonecfg -z zone1
zone3: No such zone configured
Use 'create' to begin configuring a new zone.
zonecfg:zone3> create
zonecfg:zone3> set zonepath=/zones/zone1
zonecfg:zone3> set ip-type=exclusive
zonecfg:zone3> add net
zonecfg:zone3:net> set physical=vnic1
zonecfg:zone3:net> end
zonecfg:zone3> verify
zonecfg:zone3> commit
zonecfg:zone3> exit
```

Configuring Zones To Use VNICs (cont.)

- Once a zone is configured to use a virtual NIC, you can login to the zone and configure it using the same steps you would use to configure a physical interface:

```
$ ifconfig vnic1 plumb
$ ifconfig vnic1 inet 192.168.1.2 netmask \
                    255.255.255.0 broadcast +
$ ifconfig vnic1 up
$ ping 192.168.1.1
192.168.1.1 is alive
```

- To make the VNIC settings permanent, you will need to update /etc/hosts, /etc/netmasks and /etc/hostname.vnic[0-9]+ with the VNIC network settings

Conclusion

- Crossbow provides network virtualization in OpenSolaris, which contains the building blocks needed to build virtual networks and enforce network QOS policies
- Everything listed in this presentation is 100% free, and can be downloaded from the OpenSolaris website (http://opensolaris.org)

References

- Ben Rockwood’s blog (everything Solaris): http://cuddletech.com/blog
- Crossbow FAQ: http://opensolaris.org/os/project/crossbow/faq
- Prefetch blog (the slides will be posted here) http://prefetch.net/blog
- Sunay Tripathi’s blog: http://blogs.sun.com/sunay

Questions?