ORACLE®

Zones networking and Crossbow

Phil Kirk
Oracle Solaris Internet RPE

Overview

- Original zones networking
- Enhancments to zones networking
- IP instances
- Crossbow, vnics and resource management

Shared stack model

- Zones were never designed as virtual machines
- When zones shipped this was the only option available
- With the shared stack model the zones share the global zones ip instance
- What this means is the zones, including the global zone, all share the same routing table
- When zones were initially shipped the assumption was people would use them as process containers and the zones would be on the same subnet as the global zone

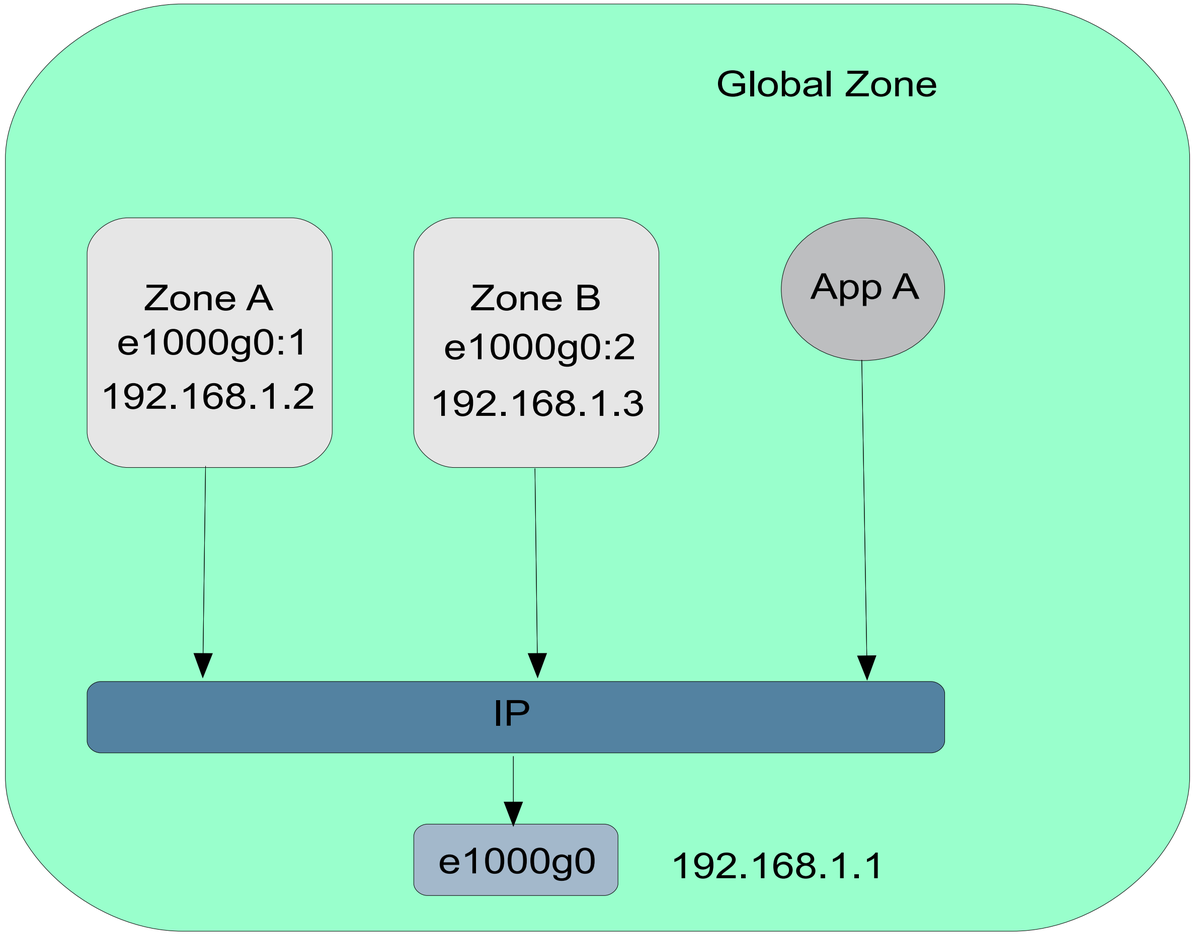
Shared stack model cont.

- When the shared stack model is used a seperate ip alias is created for each zone eg.
  - `qfe0:1: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu 1500 index 721`

    `zone ngz1`

    `inet 192.168.1.2 netmask ffffff00 broadcast 192.168.1.255`
- Seperation of network traffic is done within ip based on ip address
- Within a zone you only see the networking configuration for the zone

Shared stack model cont.



Shared model cont.

- IPMP with shared IP zones works
- Configuration is done from the global zone, ip addresses associated with the zone are placed into the IPMP group
- IP Filter works but you need to turn on loopback filtering. Have to be running OpenSolaris or Solaris 10u4 or later
- Network features like DHCP, IPsec, Raw sockets etc. don't work with shared stacks due to the shared state

Problems with the shared stack model

- What happens when you decide to put the zones on a different subnet to the global zone
- Remember ip information such as the routing table is shared
- Most problems come from using the shared model and putting the zones on a different subnet and the affect this can have on how packets are routed
- Seperating the zones from the global zone by putting them on a different interface won't necessarily fix the problems

Problems with the shared stack model cont.

- With IPMP a common problem with probe based failover is configuring host targets
- The workaround for this problem is to add a null route so for example:
  - `/usr/sbin/route add -host 192.168.0.1 192.168.0.1`
- Why? The route will have the G/W flag set
- This works great for IPMP but what about the impact it can have on zones?
- It's not just host routes, the same applies to network routes

Problems with the shared stack model cont.

- The defrouter option was introduced to so that a default router can be specified when creating and configuring the zone rather than leaving it to the admin to add manually
- In Solaris default routes are selected round robin
- It just adds a default route but uses the -ifp flags, there's no magic here
- Another problem people have run into is inter-zone traffic
- Inter-zone traffic is all sent over the loopback interface
- What if you need all traffic to go via your stateful firewall?

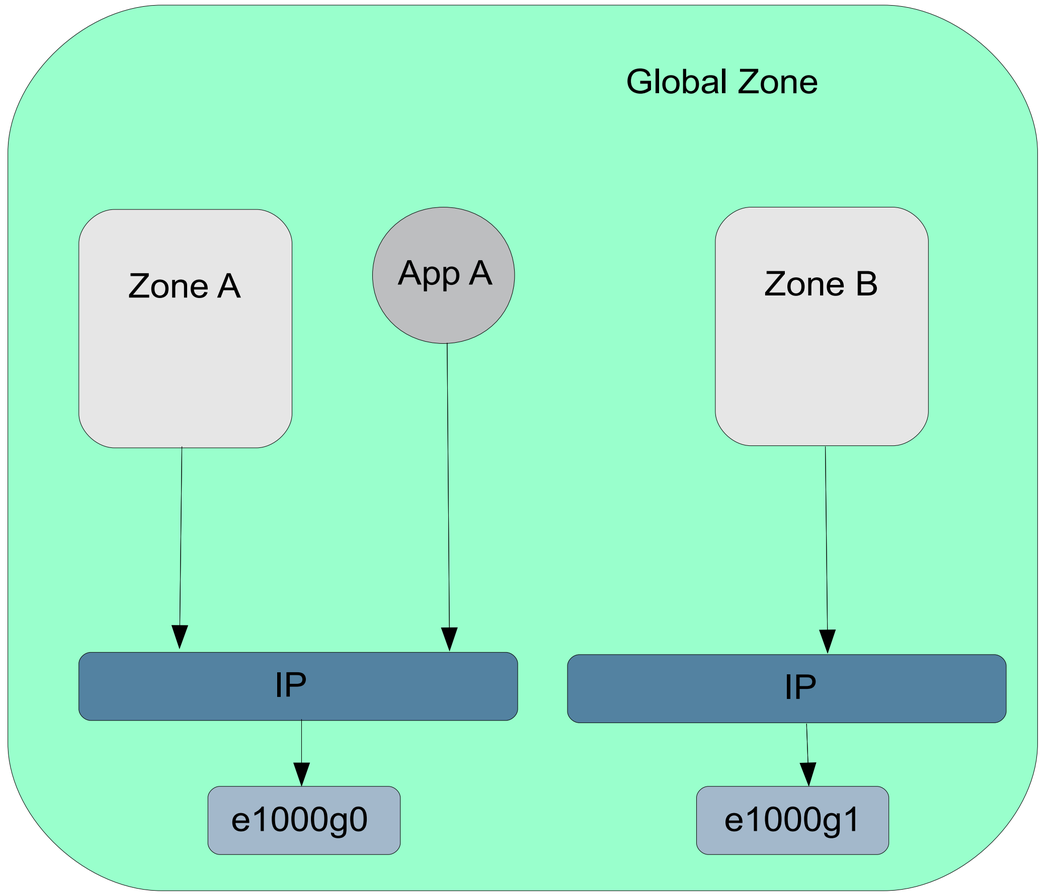
Problems with the shared stack model cont.

- Rather than force major configuration changes on customers a new ndd tubable was introduced:
  - `ip_restrict_interzone_loopback`
- Means that traffic is now forced to go out on the wire rather than been sent over the loopback
- With all the issues around shared stack why should you use it? If your configuration is simple and the zones are going to be on the same subnet(s) as the global zone then this option is fine

Exclusive ip stacks or ip instances

- Virtualization at the ip layer
- Each zone gets a unique ip instance
- Configuration is all done from inside the zone. From the global zone you need to zlogin into the zone to see what's configured
- Networking features like DHCP, IPsec, Raw sockets work
- To specify a zone should use the exclusive stack model you set the zone property ip-type to exclusive

Exclusive ip stacks or ip instances cont.



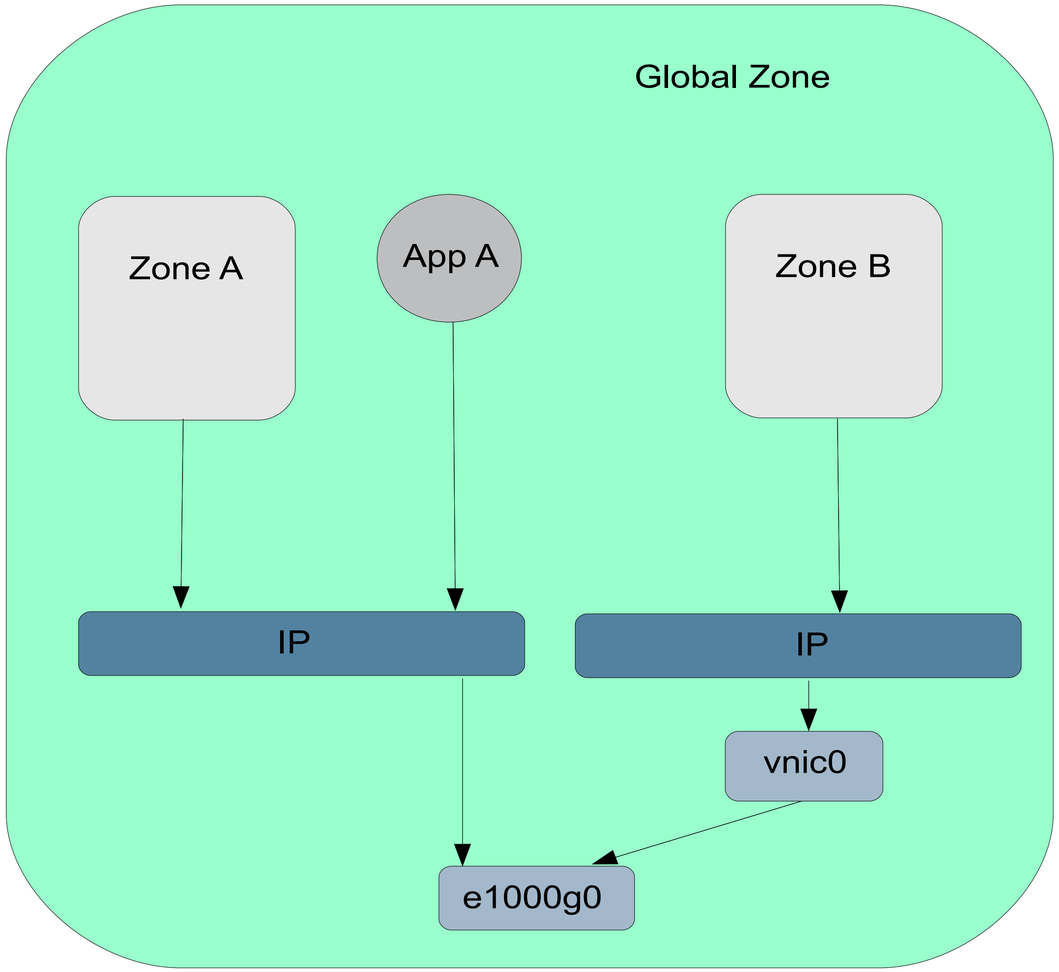
Exclusive ip stacks or ip instances cont.

- Solves all the problems that people have with the share stack model
- However, each zone has to have a nic dedicated to it
- Internally even if you have no zones there'll be one ip instance for the global zone
- Can mix exclusive and shared stack zones
- Problems with the exclusive model? Other than having to have a dedicated nic none

Crossbow and vnics

- Having to have a seperate nic for each zone is clearly a limitation
- With Crossbow vnics have been implemented, as well as much more
- Crossbow vnics allow you to slice a physical nic into virtual nics
- With etherstubs you don't even need a real nic
- Once you've created vnics these can be placed into a zone

Crossbow and vnics cont.



What else Crossbow gives you

- Resource management – bandwidth and flow management
- Bind packet processing for a data link to a given processor or group of processors
- Leverage hardware capabilities provided by the new interfaces
- Real time usage and accounting

Crossbow examples

- Creating vnics is easy, everything is driven through dladm
  - `dladm create-vnic -l <link> <vnic-link>`
  - `#dladm create-vnic -l e1000g0 vnic0`

```
#dladm show-vnic
LINK         OVER         SPEED  MACADDRESS        MACADDRTYPE       VID
vnic0        e1000g0      1000   2:8:20:2c:63:ff   random            0
```

- We can also specify link properties such as b/w restrictions, which cpus we want to process packets received on this interface etc. here

Crossbow examples cont.

- $ dladm set-linkprop -p maxbw=100 vnic0

```
$ dladm show-linkprop vnic0
LINK         PROPERTY        PERM VALUE          DEFAULT        POSSIBLE
vnic0        autopush        -w   --             --             --
vnic0        zone            rw   --             --             --
vnic0        state           r-   unknown        up             up,down
vnic0        mtu             r-   1500           1500           --
vnic0        maxbw           rw     100          --             --
vnic0        cpus            rw   --             --             --
vnic0        priority        rw   high           high           low,medium,high
vnic0        tagmode         rw   vlanonly       vlanonly       normal,vlanonly
$
```

When things go wrong

- When trying to troubleshoot zones networking problems the usual network troubleshooting tools are still the same
- Inter-zone communication is harder if you're on Solaris 10 and using shared stacks as the traffic is looped back internally
- On OpenSolaris you can now snoop the loopback, as well as snoop inside a zone using shared stack

Q&A

SOFTWARE. HARDWARE. COMPLETE.