Subject: dual ethernet woes
Posted by Steve Hodges on Tue, 14 Aug 2007 11:01:09 GMT
View Forum Message <> Reply to Message

The servers I'm trying to put onto a single hardware node each have 2 ethernet interfaces.

each server has 2 addresses, e.g. 192.168.206.11 and 192.168.205.11. (these are the primary and secondary networks respectively)

There is a DNS server on each network. Each DNS provides resolution on that network. For example the DNS at 192.168.205.254 resolves "mail" to 192.168.205.254. The DNS on 192.168.206.254 resolves "mail" to 192.168.206.254.

(actually the name is mail.domain, but you get the idea)

There is some further complication due to the fact that a third network is accessed via a router, and each network has its own gateway to the world.

route -n typically gives something like this:

```
10.60.10.0      192.168.205.1   255.255.255.0   UG    0      0        0 eth1
10.60.10.0      192.168.206.1   255.255.255.0   UG    0      0        0 eth0
192.168.205.0   0.0.0.0         255.255.255.0   U     0      0        0 eth1
192.168.206.0   0.0.0.0         255.255.255.0   U     0      0        0 eth0
0.0.0.0         192.168.206.254 0.0.0.0         UG    0      0        0 eth0
0.0.0.0         192.168.205.254 0.0.0.0         UG    0      0        0 eth1
```

If I set up the hardware node like that (and that is actually from a test hardware node) then I can happily ping all the important places (192.168.205.1, .254, and .11 and .12 (existing servers) as well as 192.168.206.1, .254, .11 and .12 and into the other network (10.60.10.x), and onto the internet).

here's the output of a test script:

```
192.168.205.1 - OK
192.168.205.11 - OK
192.168.205.12 - OK
192.168.205.254 - OK
192.168.205.253 - FAILED
192.168.206.1 - OK
192.168.206.11 - OK
192.168.206.12 - OK
192.168.206.254 - OK
192.168.206.253 - FAILED
10.60.10.131 - OK
```

10.60.10.190 - OK
google.com - OK

(the two that fail are non-existant servers -- that's OK)

However, in a VE things go all pear shaped.

The VE is set up like this:

vzctl set ${VM_NO} --ipadd ${VM_IP_PRI} --save
vzctl set ${VM_NO} --ipadd ${VM_IP_SEC} --save

and they appear on the HN in the route -n like this:

192.168.206.104 0.0.0.0 255.255.255.255 UH 0 0 0
venet0
192.168.205.104 0.0.0.0 255.255.255.255 UH 0 0 0
venet0

however in the VE the test script shows that things are going pear shaped

192.168.205.1 - FAILED
192.168.205.11 - OK
192.168.205.12 - FAILED
192.168.205.254 - OK
192.168.205.253 - FAILED
192.168.206.1 - OK
192.168.206.11 - OK
192.168.206.12 - OK
192.168.206.254 - OK
192.168.206.253 - FAILED
10.60.10.131 - OK
10.60.10.190 - OK
google.com - OK

What has me scratching my head is that it's not really too consistent :-(

I've had a bit of a look here

http://wiki.openvz.org/VEs_in_different_subnets#.282.29_An_OVZ_Hardware_Node_has_two_ethernet_interfaces

but it really doesn't make a lot of sense to me

Is there anything I should look more closely at?

Steve

Subject: Re: dual ethernet woes
Posted by Steve Hodges on Wed, 15 Aug 2007 03:02:27 GMT
View Forum Message <> Reply to Message

On 14/08/2007 7:01 PM, Steve Hodges wrote:
> The servers I'm trying to put onto a single hardware node each have 2
> ethernet interfaces.
>

OK, after extensive testing I have narrowed down what I believe to be the cause of the problem. The VE always seems to chose the first IP address it has been configured with as the source for new outgoing connections (I have no idea if it does this for existing connections, but I would suspect not).

On the HN if I ping 192.168.205.1, my ping appears to come (by default) from my address on that subnet. I can illustrate that by explicitly setting the source address.

ping 192.168.205.1 <-- this works
ping 192.168.205.1 -I 192.168.205.222 <- this works
ping 192.168.205.1 -I 192.168.206.222 <-- this fails

ping 192.168.206.1 <-- this works
ping 192.168.206.1 -I 192.168.205.222 <- this fails
ping 192.168.206.1 -I 192.168.206.222 <-- this works

It seems rather intuitive and reasonable that a machine asked to ping an address that is on the subnet of one of its interfaces will use that interface and set the source address of that ping to the machine's address on the interface it is using.

However, the same does not appear to happen in a VE

ping 192.168.205.1 <-- this fails
ping 192.168.205.1 -I 192.168.205.222 <- this works
ping 192.168.205.1 -I 192.168.206.222 <-- this fails

ping 192.168.206.1 <-- this works
ping 192.168.206.1 -I 192.168.205.222 <- this fails
ping 192.168.206.1 -I 192.168.206.222 <-- this works

It looks like the VE is selecting as its source address, the first IP that is set as the source of packets. (In this case the 192.168.206.111 address is set first).

If I swap those around in my script, setting them in the opposite order (by changing the order of the IP addresses in the /etc/vz/conf/111.conf file) then the behaviour becomes the opposite. That is, the source

address now appears to always be 192.168.205.111, and pings on the 192.168.206.0/24 subnet fail.

I guess that the VE has no idea about what subnet it's on. So I guess it can't know that a ping to 192.168.205.1 should more properly come from one or other of its IP addresses. But for correct routing it does make a difference, so how can I force this to happen?

I happen to have a couple of nodes on my network that are far more aware of the network topology (essentially they're my routers). A ping-R from them may be useful to someone who is a better network person than I am. (and I've done this with the IP addresses on the VE set normally).

This is a ping to the gateway on my primary interface (the primary address is set first on the VE)

```
ping -c 1 -R -n 192.168.206.254
PING 192.168.206.254 (192.168.206.254) 56(124) bytes of data.
64 bytes from 192.168.206.254: icmp_seq=1 ttl=63 time=0.333 ms
RR:     192.168.206.111
        192.168.206.222
        192.168.206.254
        192.168.206.254
        192.168.206.222
        192.168.206.111
```

and here is the same thing, but to the gateway on the secondary network

```
ping -c 1 -R -n 192.168.205.254
PING 192.168.205.254 (192.168.205.254) 56(124) bytes of data.
64 bytes from 192.168.205.254: icmp_seq=1 ttl=62 time=0.513 ms
RR:     192.168.206.111
        192.168.205.222
        192.168.205.254
        192.168.205.254
        192.168.206.1
        192.168.206.222
        192.168.206.111
```

192.168.205.254 and 192.168.206.254 are different machines providing a gateway to the internet
192.168.205.1/192.168.206.1 is the central router that routes between a number of internal networks

If I knew how to use tcpdump I'd probably use that to help diagnose this, but I don't :-(

Steve

Subject: Re: Re: dual ethernet woes
Posted by Steve Hodges on Wed, 15 Aug 2007 23:47:56 GMT
View Forum Message <> Reply to Message

On 15/08/2007 11:02 AM, Steve Hodges wrote:
> On 14/08/2007 7:01 PM, Steve Hodges wrote:
>> The servers I'm trying to put onto a single hardware node each have 2
>> ethernet interfaces.
>>
>
> OK, after extensive testing I have narrowed down what I believe to be
> the cause of the problem. The VE always seems to chose the first IP
> address it has been configured with as the source for new outgoing
> connections (I have no idea if it does this for existing connections,
> but I would suspect not).
>

After sleeping on it, reading some more, and applying my thinking cap, I tried the following:

```
 ifconfig venet0:0 netmask 255.255.255.0
 ifconfig venet0:1 netmask 255.255.255.0
```

these default to 255.255.255.255 and in retrospect the behaviour it was exhibiting was probably predictable.

And that appears to work with all pings to all hosts now working (I haven't checked the routes)

The question is, can I do this automatically, or is it something to put in my rc.local for each VE?

Steve

Subject: Re: Re: dual ethernet woes
Posted by Romeo Theriault on Thu, 16 Aug 2007 01:43:30 GMT
View Forum Message <> Reply to Message

Under Red Hat these routes would be configured for the network interfaces by editing the /etc/sysconfig/network-scripts/route-eth* interface. Here is an example of one such file:

GATEWAY0=192.168.1.254

NETMASK0=255.255.255.0
ADDRESS0=192.168.55.0

GATEWAY1=10.164.234.112
NETMASK1= 255.255.255.240
ADDRESS1=10.164.234.132

Taken from
http://www.cyberciti.biz/tips/configuring-static-routes-in-debian-or-red-hat-linux-systems.html

Hope that helps.

Romeo

On 8/15/07, Steve Hodges <shodges@iinet.net.au> wrote:
>
> On 15/08/2007 11:02 AM, Steve Hodges wrote:
> > On 14/08/2007 7:01 PM, Steve Hodges wrote:
> >> The servers I'm trying to put onto a single hardware node each have 2
> >> ethernet interfaces.
> >>
> >
> > OK, after extensive testing I have narrowed down what I believe to be
> > the cause of the problem. The VE always seems to chose the first IP
> > address it has been configured with as the source for new outgoing
> > connections (I have no idea if it does this for existing connections,
> > but I would suspect not).
> >
>
> After sleeping on it, reading some more, and applying my thinking cap, I
> tried the following:
>
> ifconfig venet0:0 netmask 255.255.255.0
> ifconfig venet0:1 netmask 255.255.255.0
>
> these default to 255.255.255.255 and in retrospect the behaviour it was
> exhibiting was probably predictable.
>
> And that appears to work with all pings to all hosts now working (I
> haven't checked the routes)
>
> The question is, can I do this automatically, or is it something to put
> in my rc.local for each VE?
>
> Steve
--
Romeo Theriault

System Administrator
University of Maine at Fort Kent
Ph#: 207-834-7815
Em@: romeo.theriault@maine.edu