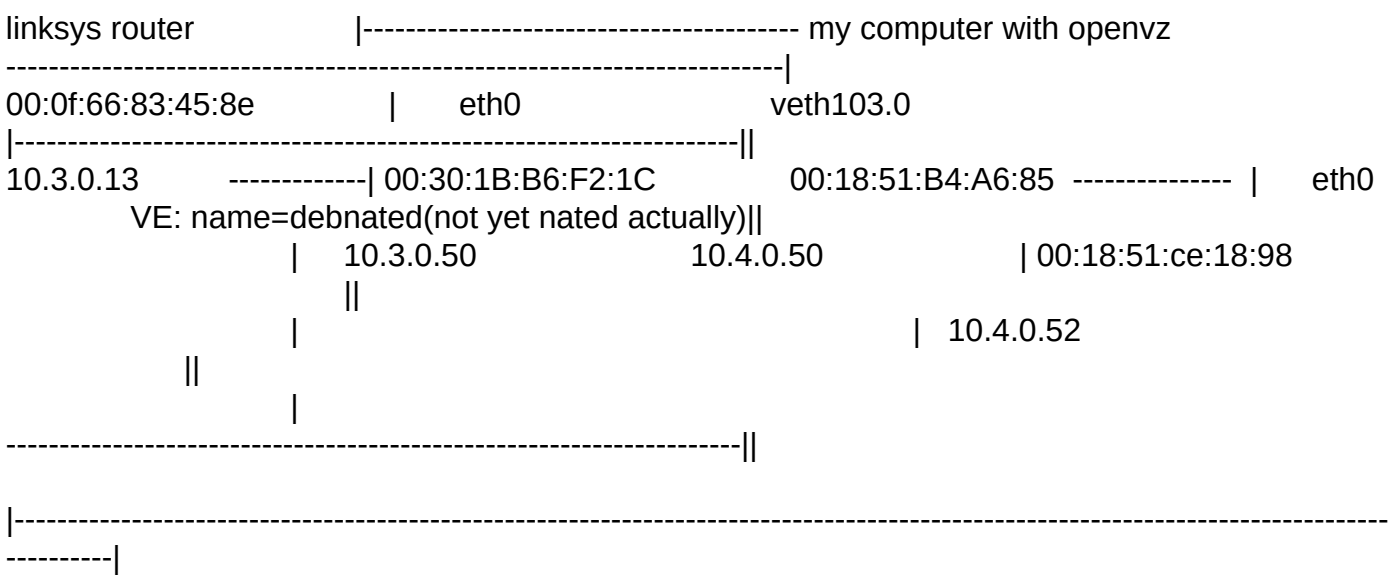
Subject: *SOLVED* VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Sun, 04 Feb 2007 10:02:47 GMT
View Forum Message <> Reply to Message

hello,

Here is a picture, of what I have: (I don't know how it will look for you, so you can also find it here (EDIT: or as an attachement)



on the VE "debnated" (ip: 10.4.0.52):
ping 10.3.0.13

one tcpdump process running on the eth0 and the other on veth103.0 : (after sorting them (and added the "== interface" at the end of the line), here is what I get)

1) 10:02:24.541939 00:18:51:ce:18:98 (oui Unknown) > Broadcast, ethertype ARP (0x0806), length 42: arp who-has 10.4.0.50 tell debnated == on veth103.0
2) 10:02:24.541968 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype ARP (0x0806), length 42: arp reply 10.4.0.50 is-at 00:18:51:b4:a6:85 (oui Unknown) == on veth103.0
3) 10:02:24.541973 00:30:1b:b6:f2:1c (oui Unknown) > 00:0f:66:83:45:8e (oui Unknown), ethertype IPv4 (0x0800), length 98: debnated > linksys: ICMP echo request, id 56597, seq 1, length 64 == on veth103.0
4) 10:02:24.542005 00:30:1b:b6:f2:1c (oui Unknown) > 00:0f:66:83:45:8e (oui Unknown), ethertype IPv4 (0x0800), length 98: debnated > linksys: ICMP echo request, id 56597, seq 1, length 64 == on eth0
5) 10:02:24.542615 00:0f:66:83:45:8e (oui Unknown) > 00:30:1b:b6:f2:1c (oui Unknown), ethertype IPv4 (0x0800), length 98: linksys > debnated: ICMP echo reply, id 56597, seq 1, length 64 == on eth0
6) 10:02:24.542633 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype IPv4 (0x0800), length 98: linksys > debnated: ICMP echo reply, id 56597, seq 1, length

64 == on veth103.0
7) 10:02:29.540780 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype ARP (0x0806), length 42: arp who-has debnated tell 10.4.0.50 == on veth103.0
8) 10:02:29.540809 00:18:51:ce:18:98 (oui Unknown) > 00:18:51:b4:a6:85 (oui Unknown), ethertype ARP (0x0806), length 42: arp reply debnated is-at 00:18:51:ce:18:98 (oui Unknown) == on veth103.0

take the 3) line for example, look at the MAC address, shouldn't it be 00:18:51:ce:18:98 > 00:18:51:B4:A6:85 instead ? the VE "debnated" should not even know thess MAC address? (unless I misunderstood something)

here is the arp cache on 'debnated", just after the ping:

```
debnated:/ #  arp
 Address                  HWtype  HWaddress           Flags Mask            Iface
 10.4.0.50                ether   00:18:51:B4:A6:85   C                     eth0
```

I don't know if it's because of how openvz works, or if I made something wrong?

thanks in advance

File Attachments

1) temp.txt, downloaded 469 times

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by Andrey Mirkin on Mon, 05 Feb 2007 11:08:47 GMT
View Forum Message <> Reply to Message

Can you please post here output of commands "ip a ls" and "ip ro ls" in VE and VE0.

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Mon, 05 Feb 2007 16:02:38 GMT
View Forum Message <> Reply to Message

oh yes I didn't give their routing tables :/

in the VE0

```
ip addr ls
2: lo: <LOOPBACK,UP,10000> mtu 16436 qdisc noqueue
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
4: eth0: <BROADCAST,MULTICAST,UP,10000> mtu 1500 qdisc pfifo_fast qlen 1000
    link/ether 00:30:1b:b6:f2:1c brd ff:ff:ff:ff:ff:ff
    inet 10.3.0.50/24 scope global eth0
```

1: venet0: <BROADCAST,POINTOPOINT,NOARP,UP,10000> mtu 1500 qdisc noqueue
    link/void
3: veth103.0: <BROADCAST,MULTICAST,UP,10000> mtu 1500 qdisc noqueue
    link/ether 00:18:51:b4:a6:85 brd ff:ff:ff:ff:ff:ff
    inet 10.4.0.50/24 scope global veth103.0
and
ip route ls
10.4.0.0/24 dev veth103.0 proto kernel scope link src 10.4.0.50
10.3.0.0/24 dev eth0 proto kernel scope link src 10.3.0.50
127.0.0.0/8 dev lo scope link
default via 10.3.0.13 dev eth0

In the VE
 ip addr l
1: lo: <LOOPBACK,UP,10000> mtu 16436 qdisc noqueue
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
3: venet0: <BROADCAST,POINTOPOINT,NOARP> mtu 1500 qdisc noqueue
    link/void
    inet 127.0.0.1/32 scope host venet0
5: eth0: <BROADCAST,MULTICAST,UP,10000> mtu 1500 qdisc noqueue
    link/ether 00:18:51:ce:18:98 brd ff:ff:ff:ff:ff:ff
    inet 10.4.0.52/24 brd 10.4.0.255 scope global eth0

and
ip r l10.4.0.0/24 dev eth0 proto kernel scope link src 10.4.0.52
default via 10.4.0.50 dev eth0

There is nothing special. Let me know if you need anything else.

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by Andrey Mirkin on Mon, 05 Feb 2007 16:18:32 GMT
View Forum Message <> Reply to Message

Can you please post here output of "ip neigh ls" in VE and VE0.
Also please post here tcpdump output without sorting for veth103.0 and eth0 in VE0.

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Mon, 05 Feb 2007 17:43:41 GMT
View Forum Message <> Reply to Message

sure
in the VE
ping 10.3.0.13
PING 10.3.0.13 (10.3.0.13) 56(84) bytes of data.
64 bytes from 10.3.0.13: icmp_seq=1 ttl=63 time=1.43 ms

--- 10.3.0.13 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 1.430/1.430/1.430/0.000 ms

then ip neigh ls10.4.0.50 dev eth0 lladdr 00:18:51:b4:a6:85 nud reachable

in the VE0
on eth0tcpdump -i 1 src net 10 and dst net 10 -e
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 68 bytes
18:30:12.066702 00:30:1b:b6:f2:1c (oui Unknown) > 00:0f:66:83:45:8e (oui Unknown), ethertype IPv4 (0x0800), length 98: debnated > linksys: ICMP echo request, id 26134, seq 1, length 64
18:30:12.067175 00:0f:66:83:45:8e (oui Unknown) > 00:30:1b:b6:f2:1c (oui Unknown), ethertype IPv4 (0x0800), length 98: linksys > debnated: ICMP echo reply, id 26134, seq 1, length 64
18:30:17.065558 00:30:1b:b6:f2:1c (oui Unknown) > 00:0f:66:83:45:8e (oui Unknown), ethertype ARP (0x0806), length 42: arp who-has linksys tell 10.3.0.50
18:30:17.065866 00:0f:66:83:45:8e (oui Unknown) > 00:30:1b:b6:f2:1c (oui Unknown), ethertype ARP (0x0806), length 60: arp reply linksys is-at 00:0f:66:83:45:8e (oui Unknown)
on veth103.0tcpdump -i 3 src net 10 and dst net 10 -e
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on veth103.0, link-type EN10MB (Ethernet), capture size 68 bytes
18:30:12.066645 00:18:51:ce:18:98 (oui Unknown) > Broadcast, ethertype ARP (0x0806), length 42: arp who-has 10.4.0.50 tell debnated
18:30:12.066671 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype ARP (0x0806), length 42: arp reply 10.4.0.50 is-at 00:18:51:b4:a6:85 (oui Unknown)
18:30:12.066677 00:30:1b:b6:f2:1c (oui Unknown) > 00:0f:66:83:45:8e (oui Unknown), ethertype IPv4 (0x0800), length 98: debnated > linksys: ICMP echo request, id 26134, seq 1, length 64
18:30:12.067191 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype IPv4 (0x0800), length 98: linksys > debnated: ICMP echo reply, id 26134, seq 1, length 64
18:30:17.065571 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype ARP (0x0806), length 42: arp who-has debnated tell 10.4.0.50
18:30:17.065599 00:18:51:ce:18:98 (oui Unknown) > 00:18:51:b4:a6:85 (oui Unknown), ethertype ARP (0x0806), length 42: arp reply debnated is-at 00:18:51:ce:18:98 (oui Unknown)

ip neigh ls
10.4.0.52 dev veth103.0 lladdr 00:18:51:ce:18:98 REACHABLE
10.3.0.13 dev eth0 lladdr 00:0f:66:83:45:8e REACHABLE

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by Andrey Mirkin on Tue, 06 Feb 2007 11:03:16 GMT
View Forum Message <> Reply to Message

Do you use NAT in VE0? Or you have route to 10.4.0.x network on linksys?

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Tue, 06 Feb 2007 11:44:36 GMT
View Forum Message <> Reply to Message

No NAT in VE0, but I did set a route to 10.4.0/24 in the linksys.(going through 10.3.0.50)
I fail to see why it would be important (except that I would have any "echo reply") The VE is using (at least that's what I understand looking at tcpdump output) the VE0 eth0 MAC address, and the linksys MAC address.

I know the VE is "sharing/using" (?) the same kernel as VE0, so this is probably normal, but if it is, why, in the VE, does ip n l only give veth103.0 MAC address? In short what would be the use of this "neighbour table" if it isn't used

The fact is, that, on the other way (line 6 in the sorted tcpdump output, first post), the mac addresses are "correct" :
6) 10:02:24.542633 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype IPv4 (0x0800), length 98: linksys > debnated: ICMP echo reply, id 56597, seq 1, length 64 == on veth103.0where 00:18:51:b4:a6:85 is veth103.0 MAC address, and 00:18:51:ce:18:98 is the VE eth0 MAC address.

I going to test the last openvz kernel asap, tonight if I can.

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Thu, 08 Feb 2007 18:33:04 GMT
View Forum Message <> Reply to Message

well I did try with the latest openvz available : 2.6.18-openvz-028.015
(For what it's worth I'm using gentoo, and I'm using the kernel provided in the gentoo repository)

but I have an other problem with this kernel, I don't know if I did something wrong with the .config ? I can't enter any VE with this kernel:
vzctl enter 103
entered into VE 103
exited from VE 103

I can still exec in the VE though

I don't see anything in /var/log/vzctl.log, vzctl --verbose didn't help. The following is what I've been asked on #openvz:
vzctl exec 103 mount

simfs on / type simfs (rw)
proc on /proc type proc (rw)
sysfs on /sys type sysfs (rw)
devpts on /dev/pts type devpts (rw)
tmpfs on /dev/shm type tmpfs (rw)
and vzctl exec 103 ls -lh /dev/ptmx
crw-rw-rw- 1 root tty 5, 2 Apr 29 2006 /dev/ptmx
 and when connecting through ssh:
ssh 10.4.0.52
Password:
Last login: Thu Feb 8 19:59:08 2007 from 10.4.0.50
Connection to 10.4.0.52 closed.(as you can see, the session is closed as soon as I log in :/

So for now I'm back with 2.6.18-openvz-028.010

For the initial problem, is this normal then? or can this be a bug? Should I bug report?

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by Andrey Mirkin on Fri, 09 Feb 2007 12:07:53 GMT
View Forum Message <> Reply to Message

I'm sorry for not answering so long.
Right now I have only one explanation of this bug - the problem actually in tcpdump:
1. tcpdump gets information that veth received packet.
2. packet passes through route table and source MAC address in this packet is changed to eth0 MAC address.
3. tcpdump shows packet with changed MAC address.

tcpdump doesn't copy packet when receive it, that is why it shows changed packet.

Usually packets are copied before routing, but this doesn't performed for veth and venet devices due to optimization (reduce overhead).

Actually I was not able to reproduce this bug on the same configuration. What version of tcpdump do you use?

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Fri, 09 Feb 2007 16:20:19 GMT
View Forum Message <> Reply to Message

I'm using tcdump-3.9.5

Quote:Usually packets are copied before routing, but this doesn't performed for veth and venet devices due to optimization (reduce overhead).
I'm not sure I've understood, "usually packets are copied", does this mean, usually, tcpdump display the copied packet?

This is wierd you cannot reproduce it:/ I have an other VE (precreated slackware template if that matters..) with the same configuration (with veth) and I can reproduce it.

Thanks you for your help

(given that I had no response before, I did a bug report: #466 :/ , the bug is probably invalid then)

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by Andrey Mirkin on Fri, 09 Feb 2007 17:17:36 GMT
View Forum Message <> Reply to Message

I'm using tcpdump-3.8.2-10.RHEL4.
Maybe I just can't reproduce this bug because of race:
Packet should be changed before printing.
Quote:I'm not sure I've understood, "usually packets are copied", does this mean, usually, tcpdump display the copied packet?
For not virtualized interfaces (e.g. when packets go from eth0 to eth1) packets are copied before changing MAC address. And tcpdump shows a copy of packet, which was not changed.

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Fri, 09 Feb 2007 18:41:47 GMT
View Forum Message <> Reply to Message

And now the million dollar (or euro..) question, why does it work if I make a bridge with only this interface (well this also work if I bridge several interfaces..) ?
brctl show
bridge name bridge id STP enabled interfaces
br0 8000.001851b4a685 no veth103.0and ip a l
4: eth0: <BROADCAST,MULTICAST,UP,10000> mtu 1500 qdisc pfifo_fast qlen 1000
link/ether 00:30:1b:b6:f2:1c brd ff:ff:ff:ff:ff:ff
inet 10.3.0.50/24 scope global eth0
...
9: veth103.0: <BROADCAST,MULTICAST,UP,10000> mtu 1500 qdisc noqueue
link/ether 00:18:51:29:5c:12 brd ff:ff:ff:ff:ff:ff
6: br0: <BROADCAST,MULTICAST,UP,10000> mtu 1500 qdisc noqueue
link/ether 00:18:51:b4:a6:85 brd ff:ff:ff:ff:ff:ff
inet 10.4.0.50/24 scope global br0on the VE0 tcpdump -i 3 -e
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode

listening on br0, link-type EN10MB (Ethernet), capture size 68 bytes
(...)
19:33:34.851958 00:18:51:ce:18:98 (oui Unknown) > 00:18:51:b4:a6:85 (oui Unknown), ethertype IPv4 (0x0800), length 98: debnated > linksys: ICMP echo request, id 15397, seq 1, length 64
19:33:34.852498 00:18:51:b4:a6:85 (oui Unknown) > 00:18:51:ce:18:98 (oui Unknown), ethertype IPv4 (0x0800), length 98: linksys > debnated: ICMP echo reply, id 15397, seq 1, length 64
This means the packet are copied when the interface is a bridge(or at least when it's not a veth nor a venet interface)?

By the way, when you said, the packet are copied, it's the kernel which makes the copy, right?

Thank you

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by Andrey Mirkin on Fri, 09 Feb 2007 23:00:28 GMT
View Forum Message <> Reply to Message

samlt wrote on Fri, 09 February 2007 13:41And now the million dollar (or euro..) question, why does it work if I make a bridge with only this interface (well this also work if I bridge several interfaces..) ?
This means the packet are copied when the interface is a bridge(or at least when it's not a veth nor a venet interface)?

If you will create a bridge with veth device you will have following situation: all outgoing packets from veth will be delivered to VE0 from bridge interface and these packets will be copied before routing as bridge is not virtual interface. Thus tcpdump will show not changed packets.

Quote:
By the way, when you said, the packet are copied, it's the kernel which makes the copy, right?

Right.

Subject: Re: VE with veth, using MAC address it shouldn't be aware of
Posted by samlt on Sat, 10 Feb 2007 01:31:03 GMT
View Forum Message <> Reply to Message

well, so, answer to my question is, yes this is normal, but what you see is not what's really happenning!

Thank you Andrey Mirkin, I tag this thread as solved