Subject: Running DHCP on VPS, (on a router..)
Posted by ml-openvz-eyck on Wed, 26 Sep 2007 06:05:12 GMT
View Forum Message <> Reply to Message

Hello,
I'm having problems with setting up a dhcp server on vps, I figured out I need to create veth device, and then bridge it to the eth0.107 interface that I would like to serve dhcp on.
And it sometimes works, but more often then not my dhcp is ignoring requests, not to mention - when I put up a second dhcp on way slower physical machine, it manages to always respond first, even though it's few switch hops further away..

So I figured I'd give this VPS a physical interface to work with, but when I do that, that device disappears from the HN, which is not very good since I'm supposed to be routing packets coming from this interface.

Is there a way to stop such physical IF from disappearing from the HN? I do understand security implications of such setup, I need this to work.

Here's my setup:
on HN, debian's /etc/network/interfaces:

```
auto br107
iface br107 inet static
      bridge_ports eth0.107
      bridge_maxwait 3
     address 192.168.0.0
     netmask 255.255.255.0
     network 192.168.0.0
     broadcast 192.168.0.255
```

the same on VPS:

```
auto eth0
iface eth0 inet static
        address 192.168.0.251
         netmask 255.255.255.0
  network 192.168.0.0
  broadcast 192.168.0.255
  gateway 192.168.0.254
```

vps.conf:

```
NETIF="ifname=eth0,mac=00:18:51:6F:E6:41,host_ifname=veth107.0,host_mac=00:18:51:1C:F7:1D"
```

firewallOne:/home/eyck# less /etc/vz/vznet.conf

#!/bin/sh

EXTERNAL_SCRIPT="/etc/vz/add-bridges"
firewallOne:/home/eyck# less /etc/vz/add-bridges
#!/bin/sh
brctl addif br107 veth107.0
ifconfig veth107.0 up

--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
 Total Existance Failure

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Fri, 28 Sep 2007 11:27:08 GMT
View Forum Message <> Reply to Message

DHCP server should work fine with veth bridged to host eth0 interface.
Can you reproduce your issue when server doesn't reply?

Delegating interface to the VE is always exclusive, though you can add 2nd network
adapter connected to the same network.

Thanks,
Kirill

Dariush Pietrzak wrote:
> Hello,
> I'm having problems with setting up a dhcp server on vps, I figured out
> I need to create veth device, and then bridge it to the eth0.107 interface
> that I would like to serve dhcp on.
> And it sometimes works, but more often then not my dhcp is ignoring
> requests, not to mention - when I put up a second dhcp on way slower
> physical machine, it manages to always respond first, even though it's few
> switch hops further away..
>
> So I figured I'd give this VPS a physical interface to work with, but when
> I do that, that device disappears from the HN, which is not very good since
> I'm supposed to be routing packets coming from this interface.
>
> Is there a way to stop such physical IF from disappearing from the HN? I
> do understand security implications of such setup, I need this to work.
>
> Here's my setup:
> on HN, debian's /etc/network/interfaces:
>
> auto br107

> iface br107 inet static
> bridge_ports eth0.107
> bridge_maxwait 3
> address 192.168.0.0
> netmask 255.255.255.0
> network 192.168.0.0
> broadcast 192.168.0.255
>
> the same on VPS:
> auto eth0
> iface eth0 inet static
> address 192.168.0.251
> netmask 255.255.255.0
> network 192.168.0.0
> broadcast 192.168.0.255
> gateway 192.168.0.254
>
>
> vps.conf:
>
>
NETIF="ifname=eth0,mac=00:18:51:6F:E6:41,host_ifname=veth107.0,host_mac=00:18:51:1C:F7:1D"
>
> firewallOne:/home/eyck# less /etc/vz/vznet.conf
> #!/bin/sh
>
> EXTERNAL_SCRIPT="/etc/vz/add-bridges"
> firewallOne:/home/eyck# less /etc/vz/add-bridges
> #!/bin/sh
> brctl addif br107 veth107.0
> ifconfig veth107.0 up
>

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by ml-openvz-eyck on Fri, 28 Sep 2007 12:14:29 GMT
View Forum Message <> Reply to Message

> DHCP server should work fine with veth bridged to host eth0 interface.
> Can you reproduce your issue when server doesn't reply?
 it looks like this:

HN: tcpdump -n -i eth0.107
08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308

08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97

HN: tcpdump -n -i br107
08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97

and finally, from inside the vps:
VPS: tcpdump -n -i eth0
08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
08:16:21.185110 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
08:16:21.187356 arp who-has 192.168.9.254 tell 192.168.9.97
08:16:21.188600 arp who-has 192.168.8.1 tell 192.168.9.97

as you can see, broadcast request from dhcp client is missing, but strangely enough, the reply is clearly visible.

I also tried doing something like this to verify the software inside VPS: I chroot to ..../vz/private/107/
and then launch the dhcpd against br107, and then it works as expected.

--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by gblond on Fri, 28 Sep 2007 12:44:10 GMT
View Forum Message <> Reply to Message

On the Friday 28 September 2007 16:14 Dariush Pietrzak, wrote:
> it looks like this:
>
> HN: tcpdump -n -i eth0.107
> 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui,
> Flags [Command], length 50 08:16:21.154240 IP 0.0.0.0.68 >
> 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308

Here is request.

> 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP,
> Reply, length 300 08:16:21.187344 arp who-has 192.168.9.254 tell

> 192.168.9.97
>
>
> HN: tcpdump -n -i br107
> 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui,
> Flags [Command], length 50 08:16:21.154240 IP 0.0.0.0.68 >
> 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308

Here is too.

> 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP,
> Reply, length 300 08:16:21.187344 arp who-has 192.168.9.254 tell
> 192.168.9.97
>
> and finally, from inside the vps:
> VPS: tcpdump -n -i eth0
> 08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui,
> Flags [Command], length 50 08:16:21.185110 IP 192.168.8.254.67 >

And here.

> 255.255.255.255.68: BOOTP/DHCP, Reply, length 300 08:16:21.187356 arp
> who-has 192.168.9.254 tell 192.168.9.97
> 08:16:21.188600 arp who-has 192.168.8.1 tell 192.168.9.97
>
> as you can see, broadcast request from dhcp client is missing, but
> strangely enough, the reply is clearly visible.
>
> I also tried doing something like this to verify the software inside VPS:
> I chroot to ..../vz/private/107/ and then launch the dhcpd against br107,
> and then it works as expected.

--
Thanks,
Vitaliy Gusev

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Fri, 28 Sep 2007 12:49:32 GMT
View Forum Message <> Reply to Message

Dariush Pietrzak wrote:
>>DHCP server should work fine with veth bridged to host eth0 interface.
>>Can you reproduce your issue when server doesn't reply?
>
> it looks like this:
>
> HN: tcpdump -n -i eth0.107

is it veth pair interface, right? ok...
and is 192.168.8.254 assigned to veth inside VE?

> 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
> 08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
> 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> 08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>
>
> HN: tcpdump -n -i br107
> 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
> 08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
> 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> 08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>
> and finally, from inside the vps:
> VPS: tcpdump -n -i eth0
> 08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
> 08:16:21.185110 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> 08:16:21.187356 arp who-has 192.168.9.254 tell 192.168.9.97
> 08:16:21.188600 arp who-has 192.168.8.1 tell 192.168.9.97

I wonder why timestamp of Reply here is greater then timestamp in HN...
Maybe... it was reply from someone else? This would make things logical - someone replies to DHCP requests and you see the reply only. Can you please check (if it is true - shutting down dhcp server inside VE won't affect tcpdump output)?

> as you can see, broadcast request from dhcp client is missing, but strangely enough, the reply
> is clearly visible.

Or it works, but only tcpdump is missing the packet?

> I also tried doing something like this to verify the software inside VPS: I chroot to ..../vz/private/107/
> and then launch the dhcpd against br107, and then it works as expected.

Thanks,
Kirill

Subject: Re: Running DHCP on VPS, (on a router..)

Posted by ml-openvz-eyck on Fri, 28 Sep 2007 12:52:22 GMT
View Forum Message <> Reply to Message

On Fri, 28 Sep 2007, Vitaliy Gusev wrote:
> > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
> Here is request.
Yes.

> > HN: tcpdump -n -i br107
> > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
> Here is too.
Yes.

>
> > 08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui,
> > Flags [Command], length 50 08:16:21.185110 IP 192.168.8.254.67 >
>
> And here.
Ooh? where?

> > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300 08:16:21.187356 arp
this is only a reply, this came from the other dhcp server, that I had to
ask to be set up few segments away from here...

--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Fri, 28 Sep 2007 13:07:39 GMT
View Forum Message <> Reply to Message

is it possible to get an access to your node to check?
if so please send me crentials privately.
also, if I will provide you a patch for testing will you be able to rebuild the kernel
and check?

Thanks,
Kirill

Dariush Pietrzak wrote:
> On Fri, 28 Sep 2007, Vitaliy Gusev wrote:
>
>>>255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
>>
>>Here is request.

>
> Yes.
>
>
>>>HN: tcpdump -n -i br107
>>>255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
>>
>>Here is too.
>
> Yes.
>
>
>>>08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui,
>>>Flags [Command], length 50 08:16:21.185110 IP 192.168.8.254.67 >
>>
>>And here.
>
> Ooh? where?
>
>
>>>255.255.255.255.68: BOOTP/DHCP, Reply, length 300 08:16:21.187356 arp
>
> this is only a reply, this came from the other dhcp server, that I had to
> ask to be set up few segments away from here...
>
>

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Fri, 28 Sep 2007 13:08:42 GMT
View Forum Message <> Reply to Message

if possible plz check with the patch attached.

Kirill Korotaev wrote:
> is it possible to get an access to your node to check?
> if so please send me crentials privately.
> also, if I will provide you a patch for testing will you be able to rebuild the kernel
> and check?
>
> Thanks,
> Kirill
>
> Dariush Pietrzak wrote:
>
>>On Fri, 28 Sep 2007, Vitaliy Gusev wrote:
>>

>>
>>>>255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
>>>
>>>Here is request.
>>
>> Yes.
>>
>>
>>
>>>>HN: tcpdump -n -i br107
>>>>255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
>>>
>>>Here is too.
>>
>> Yes.
>>
>>
>>
>>>>08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui,
>>>>Flags [Command], length 50 08:16:21.185110 IP 192.168.8.254.67 >
>>>
>>>And here.
>>
>> Ooh? where?
>>
>>
>>
>>>>255.255.255.255.68: BOOTP/DHCP, Reply, length 300 08:16:21.187356 arp
>>
>> this is only a reply, this came from the other dhcp server, that I had to
>>ask to be set up few segments away from here...
>>
>>
>
>
>

--- ./drivers/net/veth.c.ve8765 2007-09-28 14:21:33.000000000 +0400
+++ ./drivers/net/veth.c 2007-09-28 17:10:47.000000000 +0400
@@ -300,25 +300,6 @@ static int veth_xmit(struct sk_buff *skb

  if (unlikely(rcv->owner_env->disable_net))
   goto outf;
- /* Filtering */
- if (ve_is_super(dev->owner_env) &&
-   !veth_from_netdev(rcv)->allow_mac_change) {
- /* from VE0 to VEX */

```
-  if (ve_is_super(rcv->owner_env))
-   goto out;
-  if (is_multicast_ether_addr(
-      ((struct ethhdr *)skb->data)->h_dest))
-   goto out;
-  if (compare_ether_addr(((struct ethhdr *)skb->data)->h_dest,
-      rcv->dev_addr))
-   goto outf;
- } else if (!ve_is_super(dev->owner_env) &&
-   !entry->allow_mac_change) {
-  /* from VE to VE0 */
-  if (compare_ether_addr(((struct ethhdr *)skb->data)->h_source,
-      dev->dev_addr))
-   goto outf;
- }

 out:
  skb->owner_env = rcv->owner_env;
```

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by ml-openvz-eyck on Fri, 28 Sep 2007 13:26:04 GMT
View Forum Message <> Reply to Message

> > HN: tcpdump -n -i eth0.107
> is it veth pair interface, right? ok...
 I might not exactly follow the term 'veth pair interface', but eth0.107
and veth are joined together by a bridge:

```
br107          8000.0018511cf71d       no              eth0.107
                                                       veth107.0
```

> and is 192.168.8.254 assigned to veth inside VE?
 192.168.8.251 is assigned to veth, .254 is a ucarp-managed IP and it's
assigned to br107.

> > 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags
[Command], length 50
> > 08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from
00:1b:d5:2c:bf:38, length 308
> > 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length
300
> > 08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
> >
> >
> > HN: tcpdump -n -i br107
> > 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags
[Command], length 50

> > 08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
> > 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> > 08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
> >
> > and finally, from inside the vps:
> > VPS: tcpdump -n -i eth0
> > 08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
> > 08:16:21.185110 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> > 08:16:21.187356 arp who-has 192.168.9.254 tell 192.168.9.97
> > 08:16:21.188600 arp who-has 192.168.8.1 tell 192.168.9.97
>
> I wonder why timestamp of Reply here is greater then timestamp in HN...
 yes, it looks strange, especially since the next packet - arp req, has similarly delayed timestamp, like something in veth delayed packets after dhcp req/reply sequence.
 This delay is visible here: .401880 on bridge and eth0.107, but .401886 from veth. But this .000006 grows to .000014 and then shrinks to .000012 after missing the packet.

> to DHCP requests and you see the reply only. Can you please check (if it is true - shutting down
> dhcp server inside VE won't affect tcpdump output)?
 This is true, but why do some packets visible on bridge and physical interface disappear on veth? When I disable the other dhcp server there is no reply at all. dhcpd doesn't see the packet.

> > as you can see, broadcast request from dhcp client is missing, but strangely enough, the reply
> > is clearly visible.
> Or it works, but only tcpdump is missing the packet?
 That would be possible, but still puzzling, and that wouldn't explaing why both tcpdump and dhcpd are missing the packet.
 Besides, this machine is very lightly loaded at the moment, and the traffic at hand is also abysmally small, not to mention that this happens quite regularly, almost all dhcp requests get lost and never arrive to my dhcpd.

thanks, Eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
 Total Existance Failure

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Fri, 28 Sep 2007 13:39:13 GMT

Dariush Pietrzak wrote:
>>>HN: tcpdump -n -i eth0.107
>>
>>is it veth pair interface, right? ok...
>
> I might not exactly follow the term 'veth pair interface', but eth0.107
> and veth are joined together by a bridge:

by pair I mean that veth interface has 2 ends: one inside VE and another one in HN.
then what is eth0.107? vlan?

> br107 8000.0018511cf71d no eth0.107
> veth107.0
>
>
>>and is 192.168.8.254 assigned to veth inside VE?
>
> 192.168.8.251 is assigned to veth, .254 is a ucarp-managed IP and it's
> assigned to br107.

so, this means that your host system replies to DHCP request below.
cause this IP belongs to HN.
are you running DHCP server in HN as well?

>>>08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
>>>08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
>>>08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
>>>08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>>>
>>>
>>>HN: tcpdump -n -i br107
>>>08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
>>>08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
>>>08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
>>>08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>>>
>>>and finally, from inside the vps:
>>>VPS: tcpdump -n -i eth0
>>>08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
>>>08:16:21.185110 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length

300
>>>08:16:21.187356 arp who-has 192.168.9.254 tell 192.168.9.97
>>>08:16:21.188600 arp who-has 192.168.8.1 tell 192.168.9.97
>>
>>I wonder why timestamp of Reply here is greater then timestamp in HN...
>
> yes, it looks strange, especially since the next packet - arp req, has
> similarly delayed timestamp, like something in veth delayed packets after
> dhcp req/reply sequence.
> This delay is visible here: .401880 on bridge and eth0.107, but .401886
> from veth. But this .000006 grows to .000014 and then shrinks to .000012
> after missing the packet.

if my logic above about DHCP server in HN is correct, then no wonder... :)

>>to DHCP requests and you see the reply only. Can you please check (if it is true - shutting down
>>dhcp server inside VE won't affect tcpdump output)?
>
> This is true, but why do some packets visible on bridge and physical
> interface disappear on veth? When I disable the other dhcp server there is
> no reply at all. dhcpd doesn't see the packet.

I guess this can be due to veth filtering which I removed by the patch I send you.
This filtering drops the packets which should not be seen by VE.
If the patch helps - I'm right.

>>> as you can see, broadcast request from dhcp client is missing, but strangely enough, the reply
>>>is clearly visible.
>>
>>Or it works, but only tcpdump is missing the packet?
>
> That would be possible, but still puzzling, and that wouldn't explaing why
> both tcpdump and dhcpd are missing the packet.
> Besides, this machine is very lightly loaded at the moment, and the
> traffic at hand is also abysmally small, not to mention that this happens quite regularly, almost
> all dhcp requests get lost and never arrive to my dhcpd.

Thanks,
Kirill

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Fri, 28 Sep 2007 13:50:22 GMT
View Forum Message <> Reply to Message

Dariush,

BTW, I forgot to ask yo to grab tcpdump output from veth107.0 in HN.
Now from your tcpdumps we see that DHCP request is visible in eth0.107,
bridge br107 interfaces, but we don't know whether bridge has broadcasted
the request to veth107.0

Kirill

Kirill Korotaev wrote:
> Dariush Pietrzak wrote:
>
>>>>HN: tcpdump -n -i eth0.107
>>>
>>>is it veth pair interface, right? ok...
>>
>> I might not exactly follow the term 'veth pair interface', but eth0.107
>>and veth are joined together by a bridge:
>
>
> by pair I mean that veth interface has 2 ends: one inside VE and another one in HN.
> then what is eth0.107? vlan?
>
>
>>br107 8000.0018511cf71d no eth0.107
>> veth107.0
>>
>>
>>
>>>and is 192.168.8.254 assigned to veth inside VE?
>>
>> 192.168.8.251 is assigned to veth, .254 is a ucarp-managed IP and it's
>>assigned to br107.
>
>
> so, this means that your host system replies to DHCP request below.
> cause this IP belongs to HN.
> are you running DHCP server in HN as well?
>
>
>>>>08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags
[Command], length 50
>>>>08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from
00:1b:d5:2c:bf:38, length 308
>>>>08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length
300
>>>>08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>>>>

>>>>
>>>>HN: tcpdump -n -i br107
>>>>08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
>>>>08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from 00:1b:d5:2c:bf:38, length 308
>>>>08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
>>>>08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>>>>
>>>>and finally, from inside the vps:
>>>>VPS: tcpdump -n -i eth0
>>>>08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags [Command], length 50
>>>>08:16:21.185110 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
>>>>08:16:21.187356 arp who-has 192.168.9.254 tell 192.168.9.97
>>>>08:16:21.188600 arp who-has 192.168.8.1 tell 192.168.9.97
>>>
>>>I wonder why timestamp of Reply here is greater then timestamp in HN...
>>
>> yes, it looks strange, especially since the next packet - arp req, has
>>similarly delayed timestamp, like something in veth delayed packets after
>>dhcp req/reply sequence.
>> This delay is visible here: .401880 on bridge and eth0.107, but .401886
>>from veth. But this .000006 grows to .000014 and then shrinks to .000012
>>after missing the packet.
>
>
> if my logic above about DHCP server in HN is correct, then no wonder... :)
>
>
>>>to DHCP requests and you see the reply only. Can you please check (if it is true - shutting down
>>>dhcp server inside VE won't affect tcpdump output)?
>>
>> This is true, but why do some packets visible on bridge and physical
>>interface disappear on veth? When I disable the other dhcp server there is
>>no reply at all. dhcpd doesn't see the packet.
>
>
> I guess this can be due to veth filtering which I removed by the patch I send you.
> This filtering drops the packets which should not be seen by VE.
> If the patch helps - I'm right.
>
>
>>>>as you can see, broadcast request from dhcp client is missing, but strangely enough, the reply

>>>>is clearly visible.
>>>
>>>Or it works, but only tcpdump is missing the packet?
>>
>> That would be possible, but still puzzling, and that wouldn't explaing why
>>both tcpdump and dhcpd are missing the packet.
>> Besides, this machine is very lightly loaded at the moment, and the
>>traffic at hand is also abysmally small, not to mention that this happens quite regularly, almost
>>all dhcp requests get lost and never arrive to my dhcpd.
>
>
>
> Thanks,
> Kirill

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by ml-openvz-eyck on Fri, 28 Sep 2007 17:29:35 GMT
View Forum Message <> Reply to Message

> if possible plz check with the patch attached.
with this patch in, my dhcp server started working...
what should I do next?

--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Sat, 29 Sep 2007 10:20:55 GMT
View Forum Message <> Reply to Message

Can you please revert previous patch and apply the one I attached?
Does it help?

Thanks,
Kirill

Dariush Pietrzak wrote:
>>if possible plz check with the patch attached.
>
> with this patch in, my dhcp server started working...
> what should I do next?
>

```
--- ./drivers/net/veth.c.ve45678 2007-09-29 14:07:21.000000000 +0400
+++ ./drivers/net/veth.c 2007-09-29 14:18:31.000000000 +0400
@@ -306,17 +306,14 @@ static int veth_xmit(struct sk_buff *skb
   /* from VE0 to VEX */
   if (ve_is_super(rcv->owner_env))
    goto out;
-  if (is_multicast_ether_addr(
-      ((struct ethhdr *)skb->data)->h_dest))
+  if (is_multicast_ether_addr(eth_hdr(skb)->h_dest))
    goto out;
-  if (compare_ether_addr(((struct ethhdr *)skb->data)->h_dest,
-      rcv->dev_addr))
+  if (compare_ether_addr(eth_hdr(skb)->h_dest, rcv->dev_addr))
    goto outf;
  } else if (!ve_is_super(dev->owner_env) &&
    !entry->allow_mac_change) {
   /* from VE to VE0 */
-  if (compare_ether_addr(((struct ethhdr *)skb->data)->h_source,
-      dev->dev_addr))
+  if (compare_ether_addr(eth_hdr(skb)->h_source, dev->dev_addr))
    goto outf;
  }
```

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by ml-openvz-eyck on Sat, 29 Sep 2007 17:33:01 GMT
View Forum Message <> Reply to Message

> Can you please revert previous patch and apply the one I attached?
> Does it help?
it crashes on boot on machine with veth-using vps, (clean machine boots ok though).
Screenshot attached

--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

File Attachments

1) veth.xmit.crash.png, downloaded 4405 times

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Sun, 30 Sep 2007 07:30:29 GMT
View Forum Message <> Reply to Message

oh, sorry, wrong patch :/

I've attached new debug patch, please check it.
It should print some information in /var/log/messages about what packets
are dropped and due to which condition.

Thanks a lot for your help!
Kirill
P.S. you can use in production kernel with removed filtering in veth_xmit() until it is resolved.

Dariush Pietrzak wrote:
>>Can you please revert previous patch and apply the one I attached?
>>Does it help?
>
> it crashes on boot on machine with veth-using vps, (clean machine boots
> ok though).
> Screenshot attached
>
>
>
> ---
>

```
--- ./drivers/net/veth.c.ve2346 2007-09-30 11:26:01.000000000 +0400
+++ ./drivers/net/veth.c 2007-09-30 11:30:45.000000000 +0400
@@ -282,22 +282,26 @@ static int veth_xmit(struct sk_buff *skb
  struct net_device *rcv = NULL;
  struct veth_struct *entry;
  int length;
+ int reason = 0;

  stats = veth_stats(dev, smp_processor_id());
  if (unlikely(get_exec_env()->disable_net))
   goto outf;

+ reason = 1;
  entry = veth_from_netdev(dev);
  rcv = entry->pair;
  if (!rcv)
   /* VE going down */
   goto outf;

+ reason = 2;
  if (!(rcv->flags & IFF_UP)) {
   /* Target VE does not want to receive packets */
   goto outf;
  }
```

```
+ reason = 3;
  if (unlikely(rcv->owner_env->disable_net))
   goto outf;
  /* Filtering */
@@ -309,12 +313,14 @@ static int veth_xmit(struct sk_buff *skb
   if (is_multicast_ether_addr(
      ((struct ethhdr *)skb->data)->h_dest))
    goto out;
+  reason = 4;
   if (compare_ether_addr(((struct ethhdr *)skb->data)->h_dest,
      rcv->dev_addr))
    goto outf;
  } else if (!ve_is_super(dev->owner_env) &&
    !entry->allow_mac_change) {
   /* from VE to VE0 */
+  reason = 5;
   if (compare_ether_addr(((struct ethhdr *)skb->data)->h_source,
      dev->dev_addr))
    goto outf;
@@ -361,6 +367,23 @@ out:
  return 0;

 outf:
+ {
+  unsigned char *addr;
+  int i;
+
+  printk("veth_xmit() dropped pkt reason %d:\n", reason);
+
+  addr = ((struct ethhdr *)skb->data)->h_source;
+  printk("  src = ");
+  for (i = 0; i < ETH_ALEN; i++)
+   printk("%02x:", addr[i]);
+
+  addr = ((struct ethhdr *)skb->data)->h_dest;
+  printk(", dst = ");
+  for (i = 0; i < ETH_ALEN; i++)
+   printk("%02x:", addr[i]);
+  printk("\n");
+ }
  kfree_skb(skb);
  stats->tx_dropped++;
  return 0;
```

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by ml-openvz-eyck on Sun, 30 Sep 2007 11:47:51 GMT

> It should print some information in /var/log/messages about what packets
> are dropped and due to which condition.
 All the messages look like this:

Sep 30 13:31:11 dfw1 kernel: veth_xmit() dropped pkt reason 4:
Sep 30 13:31:11 dfw1 kernel: src = 00:1b:d5:84:90:d2:, dst = 00:08:02:ac:36:20:
 src mac seem to belong to the stations that my dhcpd is supposed to be
serving, I don't recognize dst macs though... I'll inwestigate it if that's
important.

 To sum up - all the messages claim that drop is due to reason 4

> + reason = 4;
>   if (compare_ether_addr(((struct ethhdr *)skb->data)->h_dest,
>   rcv->dev_addr))
>   goto outf;

 I don't recall if I've mentioned this, but I'm using both bridge, and
ucarp on top of the bridge.
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
 Total Existance Failure

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Sun, 30 Sep 2007 12:24:49 GMT

Dariush Pietrzak wrote:
>>It should print some information in /var/log/messages about what packets
>>are dropped and due to which condition.
>
> All the messages look like this:
>
> Sep 30 13:31:11 dfw1 kernel: veth_xmit() dropped pkt reason 4:
> Sep 30 13:31:11 dfw1 kernel: src = 00:1b:d5:84:90:d2:, dst = 00:08:02:ac:36:20:
> src mac seem to belong to the stations that my dhcpd is supposed to be
> serving, I don't recognize dst macs though... I'll inwestigate it if that's
> important.

it should be broadcast MAC, not this one.
Can you please check whose MAC is it?
you can you arping for this.

> To sum up - all the messages claim that drop is due to reason 4

>
>
>>+ reason = 4;
>>   if (compare_ether_addr(((struct ethhdr *)skb->data)->h_dest,
>>   rcv->dev_addr))
>>   goto outf;
>
>
> I don't recall if I've mentioned this, but I'm using both bridge, and
> ucarp on top of the bridge.

Yes, most likely ucarp is the reason... I need some reading to understand what it is...
So would be nice if we resolve this and fixed.

Thanks,
Kirill

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Mon, 01 Oct 2007 08:08:05 GMT
View Forum Message <> Reply to Message

Dariush,

can you please add -e option to tcpdump, so that we could check MAC in packets?
and tell me plz MAC addresses of HN, veth inside VE and your DHCP client.

Thanks,
Kirill

Dariush Pietrzak wrote:
>>DHCP server should work fine with veth bridged to host eth0 interface.
>>Can you reproduce your issue when server doesn't reply?
>
> it looks like this:
>
> HN: tcpdump -n -i eth0.107
> 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags
[Command], length 50
> 08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from
00:1b:d5:2c:bf:38, length 308
> 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> 08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>
>
> HN: tcpdump -n -i br107
> 08:16:19.401880 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags

[Command], length 50
> 08:16:21.154240 IP 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request from
00:1b:d5:2c:bf:38, length 308
> 08:16:21.185096 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> 08:16:21.187344 arp who-has 192.168.9.254 tell 192.168.9.97
>
> and finally, from inside the vps:
> VPS: tcpdump -n -i eth0
> 08:16:19.401886 00:1b:d4:7e:76:2a > 01:00:0c:cc:cc:cd SNAP Unnumbered, ui, Flags
[Command], length 50
> 08:16:21.185110 IP 192.168.8.254.67 > 255.255.255.255.68: BOOTP/DHCP, Reply, length 300
> 08:16:21.187356 arp who-has 192.168.9.254 tell 192.168.9.97
> 08:16:21.188600 arp who-has 192.168.8.1 tell 192.168.9.97
>
> as you can see, broadcast request from dhcp client is missing, but strangely enough, the reply
> is clearly visible.
>
> I also tried doing something like this to verify the software inside VPS: I chroot to
..../vz/private/107/
> and then launch the dhcpd against br107, and then it works as expected.
>

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by ml-openvz-eyck on Mon, 01 Oct 2007 09:46:57 GMT
View Forum Message <> Reply to Message

Hello Kirill,

> can you please add -e option to tcpdump, so that we could check MAC in packets?
> and tell me plz MAC addresses of HN, veth inside VE and your DHCP client.
Here's the tcpdump on HN:

11:38:39.651963 00:1b:d5:2c:bf:38 > ff:ff:ff:ff:ff:ff, ethertype IPv4
(0x0800), length 350: 0.0.0.0.68 > 255.255.255.255.67: BOOTP/DHCP, Request
from 00:1b:d5:2c:bf:38, length 308

11:38:39.669960 00:08:02:ac:36:20 > ff:ff:ff:ff:ff:ff, ethertype IPv4
(0x0800), length 342: 172.17.8.254.67 > 255.255.255.255.68: BOOTP/DHCP,
Reply, length 300

and on VPS it's visible as:

11:38:39.669971 00:08:02:ac:36:20 > ff:ff:ff:ff:ff:ff, ethertype IPv4
(0x0800), length 342: 172.17.8.254.67 > 255.255.255.255.68: BOOTP/DHCP,
Reply, length 300

can't find any sign of this packet in /var/log/messages, it looks like
everything in /var/log/messages is completely unrelated to dhcp, some of
those are vrrp heartbeat, others - SNAP,

it looks like this:
grep -i 00:08:02:ac:36:20 /var/log/messages

Oct 1 10:48:09 dfw1 kernel: src = 00:1b:d5:84:90:92:, dst =
00:08:02:ac:36:20:
Oct 1 10:59:42 dfw1 kernel: src = 00:1b:d5:84:7a:1f:, dst =
00:08:02:ac:36:20:

no sign of similiar packet on 11:38:39,

--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
 Total Existance Failure

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Mon, 01 Oct 2007 10:00:09 GMT
View Forum Message <> Reply to Message

Dariush Pietrzak wrote:
>>It should print some information in /var/log/messages about what packets
>>are dropped and due to which condition.
>
> All the messages look like this:
>
> Sep 30 13:31:11 dfw1 kernel: veth_xmit() dropped pkt reason 4:
> Sep 30 13:31:11 dfw1 kernel: src = 00:1b:d5:84:90:d2:, dst = 00:08:02:ac:36:20:
> src mac seem to belong to the stations that my dhcpd is supposed to be
> serving, I don't recognize dst macs though... I'll inwestigate it if that's
> important.

dst MAC is MAC of your host node according to tcpdump you've sent a few minutes ago.
So pkt was dropped correctly.

> To sum up - all the messages claim that drop is due to reason 4
>
>
>>+ reason = 4;
>>   if (compare_ether_addr(((struct ethhdr *)skb->data)->h_dest,
>>   rcv->dev_addr))
>>   goto outf;
>
>
> I don't recall if I've mentioned this, but I'm using both bridge, and

> ucarp on top of the bridge.

can you check w/o ucarp please?
We setup vlan+bridge+veth to check, but it works fine on our side.
Or maybe you can give me an access to check?

Thanks,
Kirill

Subject: Re: Running DHCP on VPS, (on a router..)
Posted by dev on Fri, 19 Oct 2007 11:18:21 GMT
View Forum Message <> Reply to Message

configuration issue. details:
http://wiki.openvz.org/Bridge_doesn%27t_forward_packets

Subject: Re: Running DHCP on VPS, (on a router..) [Summary/Solved]
Posted by ml-openvz-eyck on Fri, 19 Oct 2007 11:58:01 GMT
View Forum Message <> Reply to Message

....summary:
 it looks like the problem was caused by bridge handing over packets to
iptables firewall, so it was enough to add:
 iptables -I FORWARD -i br101 -d 255.255.255.255 -j ACCEPT
(br101 is the bridge on which I joined physical device with veth101.0)
such behaviour is caused by CONFIG_BRIDGE_NETFILTER=y in kernel config.

 Many thanks for much patience and help in resolving this annoying issue.

 regards, eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
 Total Existance Failure
