Subject: *SOLVED* networking with veth: TCP inside VE does not work
Posted by mbunkus on Tue, 27 Feb 2007 12:12:33 GMT
View Forum Message <> Reply to Message

Hey,

I'm trying to set up networking with VETH on a Debian box. The HN has one NIC with the address 192.168.1.204. The objective is to have several VEs with services like DHCP and Samba running, so I have to use VETH.

First my problem: ping is partially working, TCP is not working. A running tcpdump inside the VE only shows the first two packets (SYN and SYN+ACK) for a TCP connection initiated from inside the VE, and only one packet (only the SYN) for TCP conenctions initiated from outside the VE to a service running on the VE (e.g. ssh). pinging from the VE to the internal network works fine, pinging to a random web server on the internet works OK but has huge delays between each ping sent out.

Now some basic things about my configuration. OpenVZ 3.0.11 (according to "vzctl --version"); Debian Etch; Kernel is 2.6.18-1-openvz built from Debian's 2.7.18 kernel source with kernel-patch-openvz version 028test007.1d2.

Setting up VETH worked OK, I followed the Wiki. The VE is named 121. Here's the VETH setting:

Quote:VETH="veth121.0,88:00:00:00:79:01,eth0,89:00:00:00:79:01 "

The VE has its interface eth0 configured for the address 192.168.1.223.

After starting 121 a script on the HN runs the following commands:

Quote:ifconfig veth121.0 0
echo '1' > /proc/sys/net/ipv4/conf/veth121.0/forwarding
echo '1' > /proc/sys/net/ipv4/conf/veth121.0/proxy_arp
echo '1' > /proc/sys/net/ipv4/conf/eth1/forwarding
echo '1' > /proc/sys/net/ipv4/conf/eth1/proxy_arp

/sbin/ip route add 192.168.1.223 dev veth121.0

The situation with the pings. Pinging another host on the same network works perfectly, even a flood ping (ping -f ...). No need to show ping's or tcpdump's output here.

Pinging a webserver www.heise.de from the HN works just as well.

Pinging the same webserver www.heise.de from the VE works partially:

Quote:root@basic-net:/# ping www.heise.de
PING www.heise.de (193.99.144.85) 56(84) bytes of data.
64 bytes from www.heise.de (193.99.144.85): icmp_seq=1 ttl=246 time=636 ms
64 bytes from 193.99.144.85: icmp_seq=2 ttl=246 time=11.2 ms

64 bytes from 193.99.144.85: icmp_seq=3 ttl=246 time=11.5 ms
64 bytes from 193.99.144.85: icmp_seq=4 ttl=246 time=11.6 ms
64 bytes from 193.99.144.85: icmp_seq=5 ttl=246 time=11.0 ms

--- www.heise.de ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 31024ms
rtt min/avg/max/mdev = 11.016/136.307/636.040/249.866 ms

The problem is that there's a 10 seconds delay between each packet sent out, and hitting Ctrl-C also only exits after those 10 seconds are done.

This is merely annoying. The real problem are TCP connections which don't work at all. They're simply not established.

I'm trying to telnet www.heise.de on port 80. This is the output of a tcpdump running inside the same VE that the telnet is started in:

Quote:root@basic-net:/# tcpdump -n -i eth0
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 96 bytes
13:14:59.341308 IP 192.168.1.223.32768 > 192.168.1.221.53: 373+ AAAA? www.heise.de. (30)
13:14:59.348365 IP 192.168.1.221.53 > 192.168.1.223.32768: 373 0/1/0 (80)
13:14:59.348476 IP 192.168.1.223.32768 > 192.168.1.221.53: 59679+ AAAA?
www.heise.de.istruhk.de. (41)
13:14:59.348732 IP 192.168.1.221.53 > 192.168.1.223.32768: 59679 NXDomain* 0/1/0 (90)
13:14:59.348772 IP 192.168.1.223.32768 > 192.168.1.221.53: 27466+ A? www.heise.de. (30)
13:14:59.348984 IP 192.168.1.221.53 > 192.168.1.223.32768: 27466 1/0/0 A 193.99.144.85 (46)
13:14:59.349310 IP 192.168.1.223.36470 > 193.99.144.85.80: S 233882351:233882351(0) win
5840 <mss 1460,sackOK,timestamp 90017 0,nop,wscale 4>
13:14:59.349486 IP 193.99.144.85.80 > 192.168.1.223.36470: S 1694916676:1694916676(0) ack
233882352 win 5840 <mss 1460>
13:15:02.341951 IP 193.99.144.85.80 > 192.168.1.223.36470: S 1694916676:1694916676(0) ack
233882352 win 5840 <mss 1460>
13:15:02.346449 IP 192.168.1.223.36470 > 193.99.144.85.80: S 233882351:233882351(0) win
5840 <mss 1460,sackOK,timestamp 90767 0,nop,wscale 4>
13:15:02.346670 IP 193.99.144.85.80 > 192.168.1.223.36470: S 1694916676:1694916676(0) ack
233882352 win 5840 <mss 1460>

As you can see: The first packets are name service resolution, working nicely. Then the TCP connection is about to be established, but it seems like the SYN+ACK is not recognized by the kernel or something.

The next output is tcpdump listening to a SSH from the HN to the VE:

Quote:tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 96 bytes

13:16:41.211266 arp who-has 192.168.1.223 tell 192.168.1.204
13:16:41.211386 arp reply 192.168.1.223 is-at 89:00:00:00:79:01
13:16:41.211408 IP 192.168.1.204.59806 > 192.168.1.223.22: S 338231067:338231067(0) win 5840 <mss 1460,sackOK,timestamp 115482 0,nop,wscale 4>
13:16:44.207232 IP 192.168.1.204.59806 > 192.168.1.223.22: S 338231067:338231067(0) win 5840 <mss 1460,sackOK,timestamp 116232 0,nop,wscale 4>

The same problem, well not the same, here not even the SYN packet is answered.

Does anyone have any idea what's going wrong here?

Subject: Re: networking with veth: TCP inside VE does not work
Posted by Andrey Mirkin on Tue, 27 Feb 2007 13:02:38 GMT
View Forum Message <> Reply to Message

First, please, do not use multicast MAC addresses for your veth devices.

In your example:
Quote:VETH="veth121.0,88:00:00:00:79:01,eth0,89:00:00:00:79:01 "
89:00:00:00:79:01 MAC address is multicast one.

What configuration for veth devices do you have in VE0 (host system)?
Do you use bridge? Or you just have appropriate route to veth devices?

Subject: Re: networking with veth: TCP inside VE does not work
Posted by mbunkus on Tue, 27 Feb 2007 13:24:31 GMT
View Forum Message <> Reply to Message

Hey,

thanks for the hint. I didn't know that the first bit is a multicast indicator... After changing the MAC addresses TCP works nicely. The problem seems to be solved.
