pfSense - Bug #958

reply-to for 1:1 from other directly connected subnets not functioning correctly

10/19/2010 12:46 AM - Chris Buechler

Status:	Resolved	**Start date:**	10/19/2010
Priority:	High	**Due date:**	
Assignee:		**% Done:**	0%
Category:	Operating System	**Estimated time:**	0.00 hour
Target version:	2.0		
Affected Version:	2.0	**Affected Architecture:**	

Description

Where you have a system with two WANs, such as WAN1 and WAN2, when sourcing traffic from a host on the WAN1's IP subnet to a 1:1 NAT on WAN2, connectivity does not work. The SYN is passed in correctly, the internal host responds with its SYN ACK, but the firewall sends that SYN ACK out of WAN1, with the private IP as the source IP.

Ermal committed a potential fix for this today, just opening this for tracking.

History

#1 - 10/29/2010 09:48 PM - Chris Buechler

- Status changed from Feedback to New

this reportedly not fixed

#2 - 10/30/2010 09:23 PM - Adam Thompson

Here's some hard data describing the problem. The host "lisa.muug.mb.ca" is connected nearby via MRNet, but still needs to send email via the Terago (public) interface. The most obvious symptom is that email stops working for all MRNet-connected hosts (which includes all of CA*Net, Internet2, etc. - all research & education sites, basically), but this does affect all NAT'd connections as far as I can tell. I ran "nc remote.c3a.ca 25" from lisa.muug.mb.ca to generate these traces - "telnet remote.c3a.ca 25" or just sending an email produces the same behaviour.
-Adam

```
MRNET (wan)               -> em2        -> 192.139.69.168
      C3A_LAN (lan)             -> em0        -> 192.168.232.1
      MRPS_LAN (opt1)           -> em1_vlan233 -> 192.168.233.1
      TERAGO (opt2)             -> em3        -> 67.226.137.177
```

First, with BGP FIB coupling turned **OFF**, the path is out em3:

```
[2.0-BETA4][root@pfsense.c3a.ca]/root(8): traceroute lisa.muug.mb.ca
    traceroute to lisa.muug.mb.ca (130.179.31.46), 64 hops max, 40 byte packets
     1  172.16.30.30 (172.16.30.30)  3.206 ms  3.634 ms  4.214 ms
     2  10.64.24.9 (10.64.24.9)  7.587 ms  12.950 ms  7.557 ms
     3  10.64.24.58 (10.64.24.58)  4.972 ms  5.165 ms  5.103 ms
     4  h64-141-108-9.bigpipeinc.com (64.141.108.9)  7.094 ms  8.673 ms  6.847 ms
     5  * * h64-141-108-18.bigpipeinc.com (64.141.108.18)  8.440 ms
     6  cwrouter-ci.cc.umanitoba.ca (192.139.114.22)  7.467 ms  7.274 ms  7.842 ms
     7  enrouter-cw.cc.umanitoba.ca (130.179.42.202)  8.598 ms  7.804 ms  8.729 ms
     8  e2b1-enrouter.cc.umanitoba.ca (130.179.41.42)  9.470 ms  10.094 ms  10.064 ms
     9  * * *
    10  * * *
    11  *^C
```

```
[2.0-BETA4][root@pfsense.c3a.ca]/root(1): tcpdump -i em3 host lisa.muug.mb.ca
    tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
    listening on em3, link-type EN10MB (Ethernet), capture size 96 bytes
    20:10:51.463801 IP lisa.muug.mb.ca.54542 > static-67-226-137-178.ptr.terago.net.smtp: Flags [S], seq 12917
05018, win 5840, options [mss 1460,sackOK,TS val 379025534 ecr 0,nop,wscale 7], length 0
    20:10:51.464846 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.54542: Flags [S.], seq 1974
```

```
267749, ack 1291705019, win 8192, options [mss 1460,nop,wscale 8,sackOK,TS val 30140665 ecr 379025534], length
 0
    20:10:51.471046 IP lisa.muug.mb.ca.54542 > static-67-226-137-178.ptr.terago.net.smtp: Flags [.], ack 1, wi
n 46, options [nop,nop,TS val 379025542 ecr 30140665], length 0
    20:10:51.476244 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.54542: Flags [P.], ack 1, w
in 514, options [nop,nop,TS val 30140666 ecr 379025542], length 89
    20:10:51.484662 IP lisa.muug.mb.ca.54542 > static-67-226-137-178.ptr.terago.net.smtp: Flags [.], ack 90, w
in 46, options [nop,nop,TS val 379025555 ecr 30140666], length 0
    20:10:54.721215 IP lisa.muug.mb.ca.54542 > static-67-226-137-178.ptr.terago.net.smtp: Flags [F.], seq 1, a
ck 90, win 46, options [nop,nop,TS val 379028792 ecr 30140666], length 0
    20:10:54.721670 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.54542: Flags [.], ack 2, wi
n 514, options [nop,nop,TS val 30140990 ecr 379028792], length 0
    20:10:54.722275 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.54542: Flags [F.], seq 90,
 ack 2, win 514, options [nop,nop,TS val 30140991 ecr 379028792], length 0
    20:10:54.727965 IP lisa.muug.mb.ca.54542 > static-67-226-137-178.ptr.terago.net.smtp: Flags [.], ack 91, w
in 46, options [nop,nop,TS val 379028800 ecr 30140991], length 0
```

```
[2.0-BETA4][root@pfsense.c3a.ca]/root(1): tcpdump -i em0 host lisa.muug.mb.ca
    tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
    listening on em0, link-type EN10MB (Ethernet), capture size 96 bytes
    20:10:51.464003 IP lisa.muug.mb.ca.54542 > remote.c3a.ca.smtp: Flags [S], seq 1291705018, win 5840, option
s [mss 1460,sackOK,TS val 379025534 ecr 0,nop,wscale 7], length 0
    20:10:51.464746 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.54542: Flags [S.], seq 1974267749, ack 1291705019,
 win 8192, options [mss 1460,nop,wscale 8,sackOK,TS val 30140665 ecr 379025534], length 0
    20:10:51.471082 IP lisa.muug.mb.ca.54542 > remote.c3a.ca.smtp: Flags [.], ack 1, win 46, options [nop,nop,
TS val 379025542 ecr 30140665], length 0
    20:10:51.476209 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.54542: Flags [P.], ack 1, win 514, options [nop,no
p,TS val 30140666 ecr 379025542], length 89
    20:10:51.484692 IP lisa.muug.mb.ca.54542 > remote.c3a.ca.smtp: Flags [.], ack 90, win 46, options [nop,nop
,TS val 379025555 ecr 30140666], length 0
    20:10:54.721293 IP lisa.muug.mb.ca.54542 > remote.c3a.ca.smtp: Flags [F.], seq 1, ack 90, win 46, options
 [nop,nop,TS val 379028792 ecr 30140666], length 0
    20:10:54.721566 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.54542: Flags [.], ack 2, win 514, options [nop,nop
,TS val 30140990 ecr 379028792], length 0
    20:10:54.722249 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.54542: Flags [F.], seq 90, ack 2, win 514, options
 [nop,nop,TS val 30140991 ecr 379028792], length 0
    20:10:54.728006 IP lisa.muug.mb.ca.54542 > remote.c3a.ca.smtp: Flags [.], ack 91, win 46, options [nop,nop
,TS val 379028800 ecr 30140991], length 0
```

(no matching traffic recorded on either em1 or em2)

Then, with BGP FIB coupling turned **ON**, the path is out em2 instead:

```
[2.0-BETA4][root@pfsense.c3a.ca]/root(20): traceroute lisa.muug.mb.ca
    traceroute to lisa.muug.mb.ca (130.179.31.46), 64 hops max, 40 byte packets
     1  192.139.69.161 (192.139.69.161)  0.765 ms  0.395 ms  0.780 ms
     2  * cirouter-mr.cc.umanitoba.ca (192.139.69.18)  0.865 ms  0.804 ms
     3  cwrouter-ci.cc.umanitoba.ca (192.139.114.22)  1.334 ms  1.594 ms  1.416 ms
     4  alrouter-cw.cc.umanitoba.ca (130.179.42.206)  1.370 ms  1.416 ms  1.305 ms
     5  e2b1-alrouter.cc.umanitoba.ca (130.179.41.46)  2.351 ms  2.978 ms  2.202 ms
     6  lisa.muug.mb.ca (130.179.31.46)  1.496 ms  1.385 ms  1.478 ms
```

```
[2.0-BETA4][root@pfsense.c3a.ca]/root(3): tcpdump -i em0 host lisa.muug.mb.ca
    tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
    listening on em0, link-type EN10MB (Ethernet), capture size 96 bytes
    20:16:27.241040 IP lisa.muug.mb.ca.43897 > remote.c3a.ca.smtp: Flags [S], seq 1645264572, win 5840, option
s [mss 1460,sackOK,TS val 379361355 ecr 0,nop,wscale 7], length 0
    20:16:27.241308 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.43897: Flags [S.], seq 2209064005, ack 1645264573,
 win 8192, options [mss 1460,nop,wscale 8,sackOK,TS val 30174242 ecr 379361355], length 0
    20:16:30.231888 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.43897: Flags [S.], seq 2209064005, ack 1645264573,
 win 8192, options [mss 1460,nop,wscale 8,sackOK,TS val 30174541 ecr 379361355], length 0
    20:16:36.227075 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.43897: Flags [S.], seq 2209064005, ack 1645264573,
 win 65535, options [mss 1460,sackOK,TS val 30175141 ecr 379361355], length 0
    20:16:48.226350 IP remote.c3a.ca.smtp > lisa.muug.mb.ca.43897: Flags [R], seq 2209064006, win 0, length 0
```

(no matching traffic on em1)

```
[2.0-BETA4][root@pfsense.c3a.ca]/root(2): tcpdump -i em2 host lisa.muug.mb.ca
    tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
    listening on em2, link-type EN10MB (Ethernet), capture size 96 bytes
    20:16:17.058992 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.43896: Flags [R], seq 41750
60663, win 0, length 0
    20:16:27.241380 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.43897: Flags [S.], seq 2209
064005, ack 1645264573, win 8192, options [mss 1460,nop,wscale 8,sackOK,TS val 30174242 ecr 379361355], length
 0
    20:16:30.231960 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.43897: Flags [S.], seq 2209
064005, ack 1645264573, win 8192, options [mss 1460,nop,wscale 8,sackOK,TS val 30174541 ecr 379361355], length
 0
    20:16:36.227147 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.43897: Flags [S.], seq 2209
064005, ack 1645264573, win 65535, options [mss 1460,sackOK,TS val 30175141 ecr 379361355], length 0
    20:16:48.226430 IP static-67-226-137-178.ptr.terago.net.smtp > lisa.muug.mb.ca.43897: Flags [R], seq 22090
64006, win 0, length 0
```

```
[2.0-BETA4][root@pfsense.c3a.ca]/root(5): tcpdump -i em3 host lisa.muug.mb.ca
    tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
    listening on em3, link-type EN10MB (Ethernet), capture size 96 bytes
    20:16:27.240856 IP lisa.muug.mb.ca.43897 > static-67-226-137-178.ptr.terago.net.smtp: Flags [S], seq 16452
64572, win 5840, options [mss 1460,sackOK,TS val 379361355 ecr 0,nop,wscale 7], length 0
```

Note that the reply's source address is correct, it's just leaving out the wrong interface. So the NAT is half-working... this looks like something that I ought to be able to work around using reply-to rules, no?

#3 - 10/30/2010 09:24 PM - Adam Thompson

I'm open to workarounds as well as "fixing" the problem - this is suddenly getting a lot more important for us.

#4 - 10/31/2010 10:50 PM - Chris Buechler

- Priority changed from Normal to High

This attempted fix has caused issues with reply-to across the board.

#5 - 11/02/2010 06:20 PM - Ermal Luçi

- Status changed from New to Feedback

Fix committed test new snapshots.
For reference:

https://rcs.pfsense.org/projects/pfsense-tools/repos/mainline/commits/a0c58058079fbd633a912006a5b2bfcb2103cc65
https://rcs.pfsense.org/projects/pfsense-tools/repos/mainline/commits/d12a156d81269d32ad9bdffdf3ce3b07c6c36161

#6 - 11/05/2010 01:36 PM - Jim Pingle

- Status changed from Feedback to Resolved

#7 - 11/05/2010 08:56 PM - Adam Thompson

I finally had a chance to upgrade, and I'm sorry, but this **still** doesn't work (for me).
Now testing with 2.0-BETA4 (i386) built on Fri Nov 5 01:04:55 EDT 2010 FreeBSD 8.1-RELEASE-p1.

Using the exact same scenario as above, I still see the reply traffic trying to leave on em2 instead of em3.
In fact, it looks like I see **exactly** the same behaviour as previously documented.

1) do I need to be running a different binary snapshot to test this?
2) are you **sure** I don't need to specify a gateway or anything special in the firewall rules? Right now, the inbound NAT rule is a floating rule...

#8 - 11/07/2010 09:29 PM - Chris Buechler

- Status changed from Resolved to Feedback

#9 - 11/08/2010 01:20 PM - Ermal Luçi

Can you show some debugging info's?!
Tracing of traffic etc to actually see what the issue is for you?

#10 - 11/11/2010 11:37 AM - Adam Thompson

As I mentioned in #7, the tcpdump output didn't change at all after updating to the snapshot mentioned there.

So the debug output is still identical what what I posted in #2.

I'll update after business hours today and re-test.

Please let me know what other tracing/debug information might help fix this problem, I'll be happy to provide that as soon as possible.

#11 - 11/11/2010 11:44 AM - Adam Thompson

I know you said it wasn't relevant, but I tried setting the Gateway parameter on the firewall rule: no effect whatsoever, reply packets still aren't going the right way.

#12 - 11/11/2010 02:13 PM - Ermal Luçi

Can you post your config here?
Please sanitize you ip's but leave the subnets real per se.
I still do not understand why you say it does not work! Please post even /tmp/rules.debug to check.

#13 - 11/11/2010 08:09 PM - Adam Thompson

I've just upgraded and confirmed that it does work perfectly for port-forward NAT, but it does still break for 1:1 NAT.

I'll email you an unencrypted config.xml (as-is, not anonymized) separately for now, to your @pfsense.org address - I'll be late for an appointment if I

stop to sanitize the data right now!

#14 - 11/12/2010 05:45 AM - Ermal Luçi

Can you please try adding rules to your opt2 interface to allow the traffic for all your 1:1 ips.
This will make sure that reply-to is added to rules for those addresses and that will make sure it will work.

Since you are doing this from floating rules there is no reply-to added to the resulting rules and that is your problem.

#15 - 11/12/2010 01:12 PM - Adam Thompson

Well, I did mention it was a floating rule, back in comment #7...

Adding the rules to the specific interface works perfectly.

Thank you for making this work!!!

#16 - 11/12/2010 09:56 PM - Chris Buechler

- Status changed from Feedback to Resolved