

Re: Interesting TCP issue

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/hackers/2007-01/msg00017.html>

- *From:* Julian Elischer <julian@xxxxxxxxxxxxx>
 - *Date:* Tue, 02 Jan 2007 13:40:17 -0800
-

Steve Watt wrote:

```
On Jan 2, 0:06, Steve Watt wrote:
} Subject: Re: Interesting TCP issue
} On Jan 1, 23:56, Julian Elischer wrote:
} } Subject: Re: Interesting TCP issue
} } Steve Watt wrote:
} } > One of my users is having trouble receiving mail from Skype. So,
} } > after some sniffing, I discovered this:
} } > } } > # tcpdump -vv -s 1500 -i dc0 -X net 213.244.128.0/18
} } > tcpdump: listening on dc0, link-type EN10MB (Ethernet), capture size 1500 bytes
} } > 13:18:13.607493 IP (tos 0x20, ttl 58, id 12896, offset 0, flags [DF], proto: TCP (6),
length: 74) share.skype.net.50406 > wattres.watt.com.smtp: P, cksum 0x9297 (correct),
4072464914:4072464936(22) ack 1248591103 win 46 <nop,nop,timestamp 2511885672
520058954>
} [ sneck ]
} } > } } > And no responses from my system.
} } > } } > Interesting. I presume it has something to do with the
} } > idiotically small window the remote server is advertising. So I
} } > set net.inet.tcp.minmss down to 46, and that resulted in a RST
} } > being spit back to skype's server when its retransmit happened.
} } > [...]
} } } } turn off window scaling (I forget the sysctl) and see if that helps
} } It's broken in some versions of freeBSD at least.
} } Duh, should've mentioned the version:
} } FreeBSD wattres.Watt.COM 6.2-PRERELEASE FreeBSD 6.2-PRERELEASE #6: Tue
Dec 26 11:46:36 PST 2006 root@xxxxxxxxxxxxxxxxx:/usr/obj/usr/src/sys/WATTRES i386
} } I did the cvsup just before the build time above.
} } I just turned off net.inet.tcp.rfc1323; we'll see if that helps on the
} next polling attempt by skype's server.
```

We have a winner — setting `net.inet.tcp.rfc1323=0` let the mail message come in on the next try.

p0f's guess at the remote machine is that it's a newer Linux 2.6 box; that doesn't seem like an interoperability problem that should've slipped through -BETA, given how common those are...

The exchange with `rfc1323` looks completely normal, with the remote end

Re: Interesting TCP issue

giving windows of 5840. But it ended the conversation in a very Windows way by sending a couple of RSTs after the FIN exchange. Or has that brokenness now extended itself to Linux as well?

we have seen this since 4.x

I think a fix may be in 7.0 but I'm not sure..

I think there is a problem when the far end sets the window down to 1 but scales it by a factor of $2^{\{\text{big number}\}}$.

the FreeBSD tcp code apparently scales the wrong variable leading it to have a 1 byte window.

Andre, can you check out this problem and MFC the correct fix if it is indeed the same problem in 6.2?

I enclose part of our trouble ticket info. on the topic.

----- Problem Description -----

The TCP window scaling (rfc 1323) implementation of FreeBSD is broken such that we initially under-estimate the TCP receive window size of the remote server.

Depending on the scaling options requested by the remote server, this can cause us to send our SMTP banner spanning more than a single TCP packet.

In at least one customer case, the remote MTA doesn't like what appears to be a truncated SMTP banner and immediately RST's the connection (instead of ACK'ing the first packet and receiving the rest of the banner in the next packet).

The fact that the remote server won't accept our SMTP banner in more than one packet is a problem on their end. But the fact that we break up the banner in the first place is due to a bug in FreeBSD related to TCP window scaling.

See AE ticket # 78590.

here's what happens:

- In the originating TCP SYN packet, the remote side says "we want to use a window scaling factor of 9" (ie: multiply their TCP window size values by $2^{*9}=512$ to calculate the real window size they are able to accept.)
- We ACK this SYN, saying that we support their scaling request, but we don't need any window scaling done. (We send TCP option window scale = 0.

Re: Interesting TCP issue

Re: Interesting TCP issue

nothing wrong here yet)

- They send us an ACK with a window size of 12 (by which they really mean $12 * 512 = 6144$)
- We seem to forget about the scaling and think they can only handle a 12 byte TCP window, so we send only the first 12 bytes of our SMTP banner in the first packet.
- We intend to send the rest of the banner when they ACK the previous packet, but they are too impatient and hang up the phone without ACK'ing our partial banner so we can send the rest.

This problem likely affects all cases in which window scaling is requested in the TCP options, but is usually not noticed because it just causes us to send smaller packets than we could. In most instances, this only affects theoretical peak throughput (since we could conceivably be sending more data before requiring an ACK)

a possible workaround is to use 'sysctl' to completely disable our support for rfc1323. for example,

```
snooper:service 37] sysctl net.inet.tcp.rfc1323=0
net.inet.tcp.rfc1323: 1 -> 0
```

This SHOULD cause us to not repond to the TCP window scaling option at all, thus telling the remote server that we are not supporting it (which is better than saying we do support it, then getting it wrong).

Pending customer feedback to see if this workaround addresses the issue, then we might make the change permanent in /etc/sysctl.conf. Also pending information about what MTA software and OS is running on the remote server.

Evan did a little digging and noticed that TCP window scaling implementation is just being addressed in FreeBSD HEAD.

see:

http://cvs.ironport.com/cgi-bin/viewcvs.cgi/freebsd/src/sys/netinet/tcp_syncache.c.diff?r1=1.84&r2=1.85

It could well be that window scaling works fine AFTER the initial connection is made, but in at least this one case, it is affecting customer's ability to receive email from certain domains...

We have info on reproducing the problem by crafting TCP SYN packets with window scaling options.

freebsd-hackers@xxxxxxxxxxx mailing list

<http://lists.freebsd.org/mailman/listinfo/freebsd-hackers>

To unsubscribe, send any mail to "freebsd-hackers-unsubscribe@xxxxxxxxxxx"