

Re: [2.6.24.3][net] bug: TCP 3rd handshake abnormal timeouts

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2008-03/msg06079.html>

- *From:* Willy Tarreau <w@xxxxxx>
 - *Date:* Sat, 15 Mar 2008 07:58:49 +0100
-

[forgot to CC netdev]

Hi Gabriel,

On Sat, Mar 15, 2008 at 02:39:53AM +0100, Gabriel Barazer wrote:

Hi,

[cc'd netdev]

I am experiencing a very annoying bug which I think is a kernel bug, related to how a client establishes a TCP connection to a server (both linux, same vanilla 2.6.24.3 kernels but the problem happened also in the previous 2.6.23.14 and kernel installed).

The case is about a bunch of web servers accessing a MySQL database server via TCP and non-persistent connections and all application level errors have been excluded.

"Sometimes" when establishing a TCP connection to the server, we are seeing a 3000ms delay before the connection is effectively made. This has been confirmed first by some `strace-d` telnet tests, then by `tcpdump` captures on both side at the same time.

Here is a simplified version of what `_both_` the server and the client see (it's important because this means the packets are not lost and the bug not caused by a lossy network). I can send the original pcap file if necessary.

- The client send a TCP packet with the SYN bit
- The server replies by a TCP packet with the SYN,ACK bits (connection appears in `netstat -a` as in the `SYN_RECV` state)
- Here the client should have replied with a TCP ACK packet, but for some reason wait for exactly 3 seconds instead (the delay for retransmitting a SYN packet).
- The client re-send a TCP SYN packet

Re: [2.6.24.3][net] bug: TCP 3rd handshake abnormal timeouts

- The server replies by its TCP SYN,ACK packet
- And finally the client send its TCP ACK packet, completing the 3-way handshake and beginning communication with the MySQL server.

During the time when the bug happens (~1 minute), I can log into the server, and confirm this 3 second delay when strace-ing a telnet session to the mysql port.

I believe the problem is on the client side because when this abnormal delay happens, all the other mysql connections from the other web servers are doing fine (the MySQL server serves other host's queries without any connection delay), and I think I can rule out the userland processes as well because they are not aware of this TCP handshake failing and retrying, only the connect() syscall does.

I have reports of other people having this problem on other networks with different hardware specs, and on different networks (I run this on a local network).

I have no syncookies enabled, and have the netfilter connection tracking enabled but the tracked connections count is very stable, not overflowing and not even increasing when the problem happens.

Obviously, no message in the syslog or the kernel buffer ring.

Has anyone any idea on where to start to find and fix this bug? the situation is becoming critical as this affects all of our installations and I can't find a way to solve or even workaround this.

Any comment or idea would be very appreciated.

You should carefully check the the SYN-ACK received by the client has a correct checksum ("cksum OK" in tcpdump output). It would be possible that for some reason, something on the network randomly corrupts it.

Also, you say you have netfilter with conntrack. Is this on the client ? If so, you should try disabling it to rule out any possible bug in the connection tracking.

Regards,
Willy

--

To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@xxxxxxxxxxxxxxxxx
More majordomo info at <http://vger.kernel.org/majordomo-info.html>
Please read the FAQ at <http://www.tux.org/lkml/>

Re: [2.6.24.3][net] bug: TCP 3rd handshake abnormal timeouts