From: kerneljasonxing@gmail.com
To: edumazet@google.com, davem@davemloft.net, kuznet@ms2.inr.ac.ru,
        yoshfuji@linux-ipv6.org
Cc: netdev@vger.kernel.org, kerneljasonxing@gmail.com,
        linux-kernel@vger.kernel.org, liweishi@kuaishou.com,
        lishujin@kuaishou.com
Subject: [PATCH] tcp: fix TCP socks unreleased in BBR mode
Date: Tue, 2 Jun 2020 16:04:25 +0800   [thread overview]
Message-ID: <20200602080425.93712-1-kerneljasonxing@gmail.com> (raw)

From: Jason Xing <kerneljasonxing@gmail.com>

TCP socks cannot be released because of the sock_hold() increasing the
sk_refcnt in the manner of tcp_internal_pacing() when RTO happens.
Therefore, this situation could increase the slab memory and then trigger
the OOM if the machine has beening running for a long time. This issue,
however, can happen on some machine only running a few days.

We add one exception case to avoid unneeded use of sock_hold if the
pacing_timer is enqueued.

Reproduce procedure:
0) cat /proc/slabinfo | grep TCP
1) switch net.ipv4.tcp_congestion_control to bbr
2) using wrk tool something like that to send packages
3) using tc to increase the delay in the dev to simulate the busy case.
4) cat /proc/slabinfo | grep TCP
5) kill the wrk command and observe the number of objects and slabs in TCP.
6) at last, you could notice that the number would not decrease.

Signed-off-by: Jason Xing <kerneljasonxing@gmail.com>
Signed-off-by: liweishi <liweishi@kuaishou.com>
Signed-off-by: Shujin Li <lishujin@kuaishou.com>
---
 net/ipv4/tcp_output.c | 3 ++-
 1 file changed, 2 insertions(+), 1 deletion(-)

diff --git a/net/ipv4/tcp_output.c b/net/ipv4/tcp_output.c
index cc4ba42..5cf63d9 100644
--- a/net/ipv4/tcp_output.c
+++ b/net/ipv4/tcp_output.c
@@ -969,7 +969,8 @@ static void tcp_internal_pacing(struct sock *sk, const struct sk_buff
*skb)
        u64 len_ns;
        u32 rate;

-       if (!tcp_needs_internal_pacing(sk))
+   if (!tcp_needs_internal_pacing(sk) ||
+   hrtimer_is_queued(&tcp_sk(sk)->pacing_timer))
                return;
        rate = sk->sk_pacing_rate;
        if (!rate || rate == ~0U)
--
1.8.3.1

---

next   reply   other threads:[~2020-06-02 8:05 UTC|newest]

Thread overview: 22+ messages / expand[flat|nested]   mbox.gz Atom feed top
2020-06-02 8:04 kerneljasonxing [this message]

```
2020-06-02 13:05 ` [PATCH] tcp: fix TCP socks unreleased in BBR mode Eric Dumazet
2020-06-03  1:53   ` Jason Xing
2020-06-03  2:14     ` David Miller
2020-06-03  2:29     ` Eric Dumazet
2020-06-03  2:41       ` Jason Xing
2020-06-03  2:43         ` Eric Dumazet
2020-06-03  2:48           ` Jason Xing
2020-06-03  5:05           ` Jason Xing
2020-06-03  5:44             ` Eric Dumazet
2020-06-03  6:32               ` Jason Xing
2020-06-03 12:02               ` Neal Cardwell
2020-06-03 13:51                 ` Jason Xing
2020-06-03 13:54                 ` Eric Dumazet
2020-06-03 14:07                   ` Neal Cardwell
2020-06-04  8:40                     ` Jason Xing
2020-06-04  9:00 ` [PATCH v2 4.19] " kerneljasonxing
2020-06-04 13:10   ` Eric Dumazet
2020-06-04 13:47     ` Jason Xing
2020-08-11 10:37       ` Jason Xing
2020-08-11 15:32         ` Eric Dumazet
2021-04-30 10:51           ` Jason Xing
```

find likely ancestor, descendant, or conflicting patches for this message:

```
dfblob:cc4ba42 dfblob:5cf63d9
```

search   (help)

Reply instructions:

You may reply publicly to this message via plain-text email
using any one of the following methods:

* Save the following mbox file, import it into your mail client,
  and reply-to-all from there: mbox

  Avoid top-posting and favor interleaved quoting:
  https://en.wikipedia.org/wiki/Posting_style#Interleaved_style

* Reply using the **--to, --cc,** and **--in-reply-to**
  switches of git-send-email(1):

  git send-email \
    --in-reply-to=20200602080425.93712-1-kerneljasonxing@gmail.com \
    --to=kerneljasonxing@gmail.com \
    --cc=davem@davemloft.net \
    --cc=edumazet@google.com \
    --cc=kuznet@ms2.inr.ac.ru \
    --cc=linux-kernel@vger.kernel.org \
    --cc=lishujin@kuaishou.com \
    --cc=liweishi@kuaishou.com \
    --cc=netdev@vger.kernel.org \
    --cc=yoshfuji@linux-ipv6.org \
    /path/to/YOUR_REPLY

  https://kernel.org/pub/software/scm/git/docs/git-send-email.html

* If your mail client supports setting the **In-Reply-To** header
  via mailto: links, try the mailto: link

Be sure your reply has a **Subject:** header at the top and a blank line before the message body.