Packet Pacing, TCP BBRv3 and Jumbo Frames

DC24 workshop, 9-10 November 2023

Scope

Improve performances of data transfer

- being more resilient to packet loss
- better fitting in buffer-constrained networks
- reducing load on sending hosts
- more effectively sharing available bandwidth

Packet pacing with BBRv3

BBR: Bottleneck Bandwidth and Round-trip propagation time

The open-source BBR TCP congestion control algorithm has been developed by Google, and standardised within the IETF. Version1 is widely used within Google.

In traditional loss-based algorithms (RENO, CUBIC), the sending rate is established by the size of the congestion window, and the sender node may send packets in bursts, up to the maximum rate of the sender’s interface. Thus, traditional algorithms rely on routers’ buffers to absorb packet bursts.
BBR uses packet pacing to set the sending rate to the estimated bottleneck bandwidth. The pacing technique spaces out or paces packets at the sender node, spreading them over time.

BBR algorithm has proven superior to simple loss-based congestion control. Google found BBRv3 to have a 12% reduction in the packet re-transmit rate and a slight latency improvement.

V3 vs V1:

- Better coexistence with RENO/CUBIC
- Lower loss rate

BBRv3 is not available yet on any of the WLCG recommended Linux distributions

Packet pacing with Linux TC

TC (Traffic Control) is a Linux kernel module that can shape network traffic and allows for packet pacing

Packet Pacing with TC will also be tested

Jumbo Frames

IPv4 and IPv6 frames have a standard maximum size of 1500 Bytes (IP header + payload)

Jumbo frames: whatever maximum size bigger than 1500B. 9000B is the most used value in the R&D community

Reducing the relative size of the IP header over the payload can reduce the load on the CPU of the sender of large data flows, thus allowing greater throughput for CPU intensive transfers

On the other hand, transfers between hosts using different MTUs can lead to traffic blackholing if the networks in between are not properly configured

BBRv3 testing at ESnet

RTT 150ms:

- Throughput is 2-3 times better with BBR than CUBIC
- Parallel streams step on each other less than expected

Jumbo testing at ESnet

Single stream

- Jumbo frames are 3x faster on 100G hosts
- Jumbo frames are about 15% faster on 10G hosts

8 streams:

- Jumbo frames are about 25% faster on 100G hosts
- Jumbo frames are the same as 1500B 10G hosts

Jumbo testing at JISC

Source	Destination	RTT	9000	1500
SURF (NL)	RNP (Brazil)	100ms	31 Gbit/s	20 Gbit/s
Jisc (London)	BNL (USA)	100ms	14 Gbit/s	6 Gbit/s
SURF (NL)	Jisc (London)	7.2 ms	23 Gbit/s	6 Gbit/s

Plan for DC24

Testing of transfer performances using different Congestion Control protocols and different MTU values, between pairs of sites at different distances, in all possible combinations. Aim to be understood if 9000B is really the best MTU size.

Survey already existing Jumbo frame deployments

Tests and surveys will be carried out mostly before DC24. Available resources and schedule of the tests before DC24 will be documented in this google doc

To be seen if it will be possible to use BBRv3, TC and Jumbo frames on production servers during DC24. It may be complicated to switch production servers to Jumbo and/or BBRv3 during the two weeks of DC24

The most realistic goal of this project is to produce recommendations on packet pacing and Jumbo frames that could be useful for DC26

WLCG DOMA CERNBox Projects covered by this presentation

- Jumbo frames
  https://cernbox.cern.ch/text-editor/public/aCITXJenZxpF5qw/use-of-jumbo-frames.txt?contextRouteName=files-public-link&contextRouteParams.driveAliasAndItem=public/aCITXJenZxpF5qw
- Packet pacing
  https://cernbox.cern.ch/text-editor/public/aCITXJenZxpF5qw/WLCGTrafficPacing.txt?contextRouteName=files-public-link&contextRouteParams.driveAliasAndItem=public/aCITXJenZxpF5qw
- BBR
  https://cernbox.cern.ch/text-editor/public/aCITXJenZxpF5qw/BBR-performances.txt?contextRouteName=files-public-link&contextRouteParams.driveAliasAndItem=public/aCITXJenZxpF5qw

Timeline and Plans Leading to DC24

Spreadsheet tracking hosts and tests
https://docs.google.com/spreadsheets/d/1U0VXIfWHfpK7bX7k2ucFep4Xo_4KQ5x-7rKD7e4az7Y/edit#gid=0