Qcow2 snapshot chain length impact on performance and memory consumption

pierre.olivier@manchester.ac.uk

In this experiment we investigate the impact of a Qcow2 snapshot chain size of I/O performance and RAM consumption.

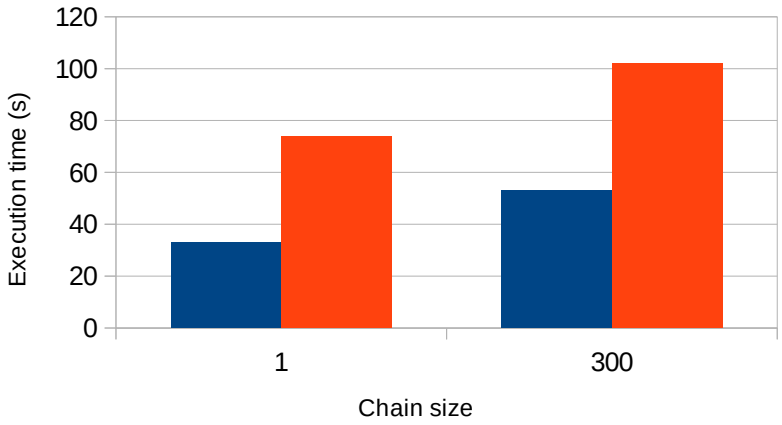
A Look at Performance

A long snapshot chain is first created as follows: starting from a base 20GB disk image of ubuntu 18.04, the VM is snapshotted every 10 seconds. During the experiment the VM runs a light write workload in the form of a fio job writing at 53KB/s in a 10GB file. In the end, we have a chain size of 300 snapshots. The size of each snapshot is about 4.9 MB. On the host the disk image and snapshots are on a standard SSD on my ex4 filesystem.

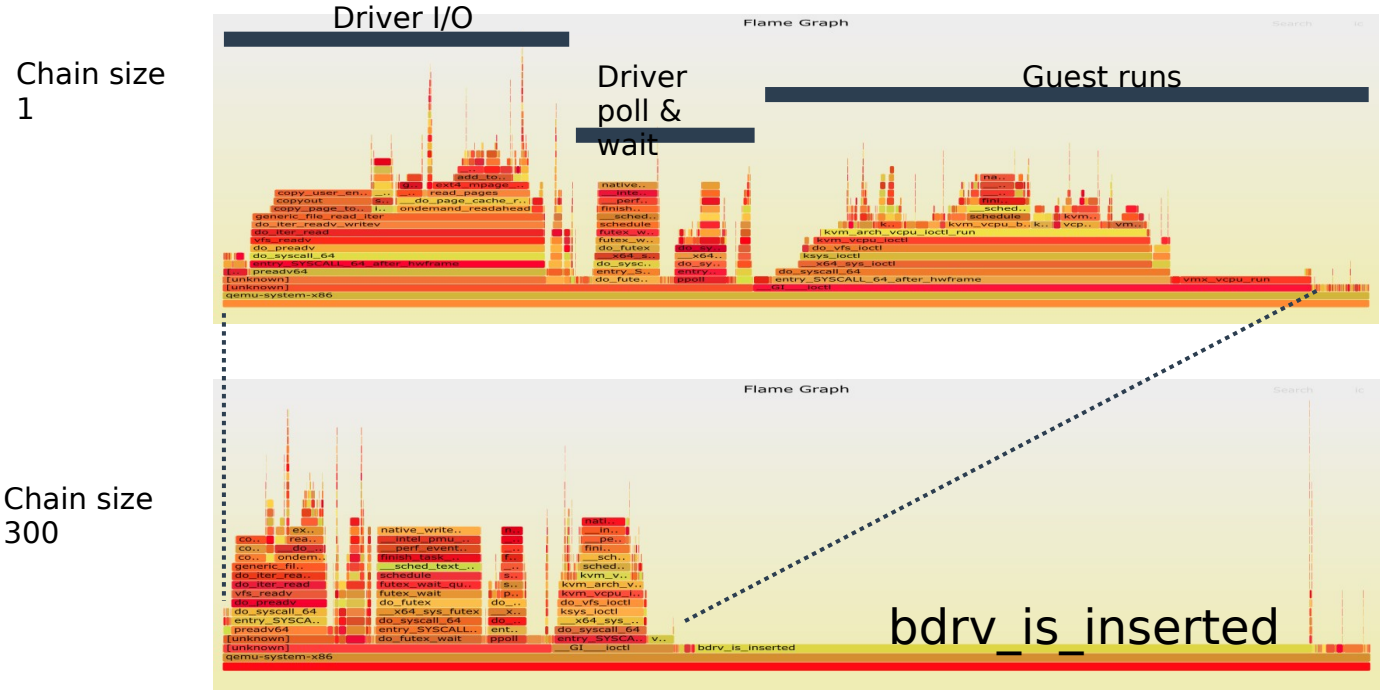
After that we boot the VM on a given snapshot to select a given snapshot chain length and run two tests, measuring their execution time. Another run is done where qemu is profiled with perf to further understand the performance.

1. Reading the entire disk from the guest: `dd if=/dev/sda of=/dev/null bs=4M`. This command is first launched without measurement to populate Qcow's indexing caches, then the guest page cache is dropped and the command is launched another time, this time with measurements.
2. Streaming all the layers into the current one.

Results are presented for A) a very small chain (1 snapshot) and B) a very long one (300 snapshots). Execution times are below.



As one can see the execution time for reading the disk as well as streaming are seriously impacted by the chain length. Now let's look at the flame graphs obtained with perf profiling qemu during the disk read experiment.



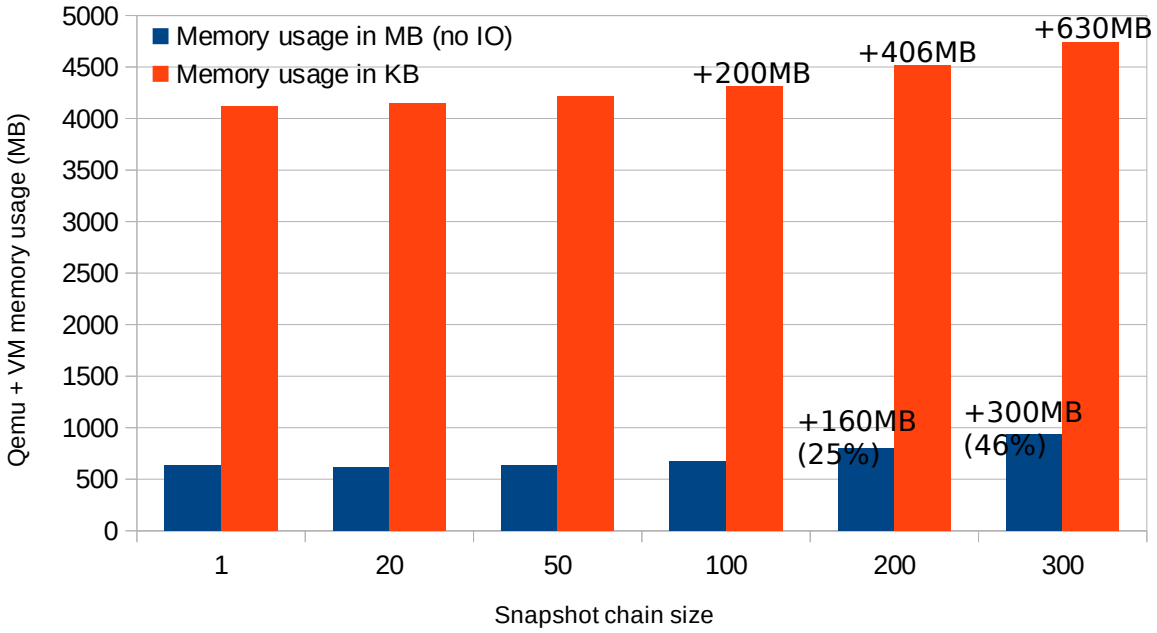
The interactive flame graphs can be found here: size 1, size 300.

With a long chain a very large portion of the time is spent calling `bdrv_is_inserted`. The function can be found here. It is calling itself recursively in a loop which seems to iterate over the snapshot chain. It seems that the complexity of all calls to this function is factorial n.

We see a similar behaviour when profiling the streaming operation. The flame graphs are available here: size 1, size 300.

A Look at Memory Consumption

In this experiment the RAM footprint of Qemu (peak Resident Set Size – RSS) is measured when varying the chain length. 2 runs are done, one in which the VM is simply booted then halted, the other where the VM is booted, the entire disk is read with the `dd` command previously presented, then the VM is halted. As one can see the chain length has a significant impact on the memory consumption.



When streaming the memory consumption is also much higher on long chains: 1.2 GB for 300 vs. 650 MB for 1.