

## Re: Packet loss every 30.999 seconds

---

*Source:* <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/stable/2007-12/msg00394.html>

---

- *From:* Bruce Evans <[brde@xxxxxxxxxxxxxxxxxx](mailto:brde@xxxxxxxxxxxxxxxxxx)>
  - *Date:* Tue, 18 Dec 2007 23:36:50 +1100 (EST)
- 

On Mon, 17 Dec 2007, Scott Long wrote:

Bruce Evans wrote:

On Mon, 17 Dec 2007, David G Lawrence wrote:

One more comment on my last email... The patch that I included is not meant as a real fix – it is just a bandaid. The real problem appears to be that a very large number of vnodes (all of them?) are getting synced (i.e. calling `ffs_syncvnode()`) every time. This should normally only happen for dirty vnodes. I suspect that something is broken with this check:

```
if (vp->v_type == VNON || ((ip->i_flag &
(IN_ACCESS | IN_CHANGE | IN_MODIFIED |
IN_UPDATE)) == 0 &&
vp->v_bufobj.bo_dirty.bv_cnt == 0)) {
VI_UNLOCK(vp);
continue;
}
```

Isn't it just the  $O(N)$  algorithm with  $N$  quite large? Under ~5.2, on

Right, it's a non-optimal loop when  $N$  is very large, and that's a fairly well understood problem. I think what DG was getting at, though, is that this massive flush happens every time the syncer runs, which doesn't seem correct. Sure, maybe you just rsynced 100,000 files 20 seconds ago, so the upcoming flush is going to be expensive. But the next flush 30 seconds after that shouldn't be just as expensive, yet it

Re: Packet loss every 30.999 seconds

appears to be so.

I'm sure it doesn't cause many bogus flushes. iostat shows zero writes caused by calling this incessantly using "while ;; do sync; done".

This is further supported by the original poster's claim that it takes many hours of uptime before the problem becomes noticeable. If vnodes are never truly getting cleaned, or never getting their flags cleared so that this loop knows that they are clean, then it's feasible that they'll accumulate over time, keep on getting flushed every 30 seconds, keep on bogging down the loop, and so on.

Using "find / >/dev/null" to grow the problem and make it bad after a few seconds of uptime, and profiling of a single sync(2) call to show that nothing much is done except the loop containing the above:

under ~5.2, on a 2.2GHz A64 UP ini386 mode:

after booting, with about 700 vnodes:

```
% % cumulative self self total % time seconds seconds calls ns/call ns/call name % 30.8 0.000 0.000 0
100.00% mcount [4]
% 14.9 0.001 0.000 0 100.00% mexitcount [5]
% 5.5 0.001 0.000 0 100.00% cputime [16]
% 5.0 0.001 0.000 6 13312 13312 vfs_msync [18]
% 4.3 0.001 0.000 0 100.00% user [21]
% 3.5 0.001 0.000 5 11321 11993 ffs_sync [23]
```

after "find / >/dev/null" was stopped after saturating at 64000 vnodes (desiredvodes is 70240):

```
% % cumulative self self total % time seconds seconds calls ns/call ns/call name % 50.7 0.008 0.008 5
1666427 1667246 ffs_sync [5]
% 38.0 0.015 0.006 6 1041217 1041217 vfs_msync [6]
% 3.1 0.015 0.001 0 100.00% mcount [7]
% 1.5 0.015 0.000 0 100.00% mexitcount [8]
% 0.6 0.015 0.000 0 100.00% cputime [22]
% 0.6 0.016 0.000 34 2660 2660 generic_bcopy [24]
% 0.5 0.016 0.000 0 100.00% user [26]
```

vfs\_msync() is a problem too. It uses an almost identical loop for the case where the vnode is not dirty (but has a different condition for being dirty). ffs\_sync() is called 5 times because there are 5 ffs file systems mounted r/w. There is another ffs file system mounted r/o and that combined with a missing r/o optimization might give the extra call to vfs\_msync(). With 64000 vnodes, the calls take 1–2 ms each. That is already quite a lot, and there are many calls. Each call only looks at vnodes under the mount point so the number of mounted

Re: Packet loss every 30.999 seconds

file systems doesn't affect the total time much.

ffs\_sync() is taking 125 ns per vnode. That is more than I would have expected.

Bruce

---

freebsd-stable@xxxxxxxxxxx mailing list

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

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