

## Re: simplify disksort, please review.

**Source:** <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/arch/2005-06/0055.html>

---

**From:** Poul-Henning Kamp (*phk\_at\_phk.freebsd.dk*)

**Date:** 06/10/05

To: Jeff Roberson <jroberson@chesapeake.net>

Date: Fri, 10 Jun 2005 10:10:47 +0200

In message <20050610040302.S16943@mail.chesapeake.net>, Jeff Roberson writes:

>On Thu, 9 Jun 2005, Poul-Henning Kamp wrote:

>

>> In message <20050609193008.GB837@darkness.comp.waw.pl>, Pawel Jakub Dawidek writes:

>>

>> >The one example of how the order can be broken (write(offset, size)):

>> >

>> > write(1024, 512)

>> > write(0, 2048)

>>

>> If you issue these two requests just like that, you get no guarantee

>> which order they get written in.

>>

>> It's not just disksort which might surprise you, tagged queuing and

>> write caches may mess up your day as well.

>

>Anyway, it's really not important to solve at the disk queue layer,

>because it's only an issue with raw devices, and it's very irregular to

>have multiple processes opening the same raw device to issue a lot of io

>of different sizes.

It will not be solved at any level, because it would require the GEOM layer to either make all writes strictly synchronous, a total no-go, or keep very expensive housekeeping of which requests are outstanding so we can block conflicting writes etc.

This is something that belongs 100% in the code issuing the I/O requests and if some of that code does something stupid, it looses and that is how it should be.

Jeff is correct that the buffer-cache mostly will prevent people from shooting their feet off, but nothing prevents a GEOM class from making a fool of itself.

--

Poul-Henning Kamp	UNIX since Zilog Zeus 3.20
phk@FreeBSD.ORG	TCP/IP since RFC 956
FreeBSD committer	BSD since 4.3-tahoe

freebsd-arch: Re: simplify disksort, please review.

Never attribute to malice what can adequately be explained by incompetence.

---

freebsd-arch@freebsd.org mailing list

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

To unsubscribe, send any mail to "freebsd-arch-unsubscribe@freebsd.org"