

Re: [PATCH] sendfile removal

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2007-05/msg13695.html>

- *From:* Jens Axboe <jens.axboe@xxxxxxxxxx>
 - *Date:* Thu, 31 May 2007 13:05:50 +0200
-

On Thu, May 31 2007, Christoph Hellwig wrote:

On Thu, May 31, 2007 at 12:33:16PM +0200, Jens Axboe wrote:

– nfd: The `–>rq_sendfile_ok` optimization is gone for now. I can't determine the value of it, but I'm assuming it's there for a reason. Any chance this can be converted to splice, or use something else than `–>sendfile()`? CC'ed Neil.

sendfile usage in nfsd avoids a data copy and allows to use checksum offloading. it's quite important for nfs server workloads.

OK, I hope Neil can provide some input on how to convert it. Of course I'm just fishing for Neil to actually do that work :-)

Apart from that, it was mostly straight forward. Almost everybody uses `generic_file_sendfile()`, which makes the conversion easy. I changed loop to use `do_generic_file_read()` instead of `sendfile`, it works for me...

```
diff --git a/drivers/block/loop.c b/drivers/block/loop.c
index 5526ead..92bac14 100644
--- a/drivers/block/loop.c
+++ b/drivers/block/loop.c
@@ -435,16 +435,24 @@ do_lo_receive(struct loop_device *lo,
 {
     struct lo_read_data cookie;
     struct file *file;
     – int retval;
     + read_descriptor_t desc;
     +
     + desc.written = 0;
     + desc.count = bvec->bv_len;
     + desc.arg.data = &cookie;
     + desc.error = 0;
```

Re: [PATCH] sendfile removal

```
cookie.lo = lo;
cookie.page = bvec->bv_page;
cookie.offset = bvec->bv_offset;
cookie.bsize = bsize;
file = lo->lo_backing_file;
- retval = file->f_op->sendfile(file, &pos, bvec->bv_len,
- lo_read_actor, &cookie);
- return (retval < 0)? retval: 0;
+
+ do_generic_file_read(file, &pos, &desc, lo_read_actor);
```

This change is wrong. loop or any existing user of ->sendfile absolutely needs to go through a file operations vector so that file-system specific actions such as locking are performed. This is required at least for the clustered filesystems and XFS. The right way to implement this is via do_splice_direct or something similar.

do_generic_file_read is only a library function for filesystem use and should never be called directly.

I'll convert it to do_splice_direct(), thanks.

—
Jens Axboe

—
To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@xxxxxxxxxxxxxxxxx
More majordomo info at <http://vger.kernel.org/majordomo-info.html>
Please read the FAQ at <http://www.tux.org/lkml/>