Subject: [PATCH] ext3: wrong error behavior
Posted by vaverin on Sat, 02 Sep 2006 11:41:04 GMT
View Forum Message <> Reply to Message

SWsoft Virtuozzo/OpenVZ Linux kernel team has discovered that ext3 error
behavior was broken in linux kernels since 2.5.x versions by the following patch:

2002/10/31 02:15:26-05:00 tytso@snap.thunk.org
Default mount options from superblock for ext2/3 filesystems
 http://linux.bkbits.net:8080/linux-2.6/gnupatch@3dc0d88eKbV9 ivV4ptRNM8fBuA3JBQ

In case ext3 file system is mounted with errors=continue (EXT3_ERRORS_CONTINUE)
errors should be ignored when possible. However at present in case of any error
kernel aborts journal and remounts filesystem to read-only. Such behavior was
hit number of times and noted to differ from that of 2.4.x kernels.

This patch fixes this:
- do nothing in case of EXT3_ERRORS_CONTINUE,
- set EXT3_MOUNT_ABORT and call journal_abort() in all other cases
- panic() should be called after ext3_commit_super() to save
 sb marked as EXT3_ERROR_FS

Signed-off-by: Vasily Averin <vvs@sw.ru>
Ack-by: Kirill Korotaev <dev@sw.ru>

Thank you,
 Vasily Averin

SWsoft Virtuozzo/OpenVZ Linux kernel team

```
--- linux-2.6.18-rc5/fs/ext3/super.c.orig 2006-09-02 12:54:01.000000000 +0400
+++ linux-2.6.18-rc5/fs/ext3/super.c 2006-09-02 13:10:02.000000000 +0400
@@ -159,20 +159,21 @@ static void ext3_handle_error(struct sup
  if (sb->s_flags & MS_RDONLY)
   return;

- if (test_opt (sb, ERRORS_RO)) {
-  printk (KERN_CRIT "Remounting filesystem read-only\n");
-  sb->s_flags |= MS_RDONLY;
- } else {
+ if (!test_opt (sb, ERRORS_CONT)) {
   journal_t *journal = EXT3_SB(sb)->s_journal;

   EXT3_SB(sb)->s_mount_opt |= EXT3_MOUNT_ABORT;
   if (journal)
    journal_abort(journal, -EIO);
  }
+ if (test_opt (sb, ERRORS_RO)) {
```

```
+  printk (KERN_CRIT "Remounting filesystem read-only\n");
+  sb->s_flags |= MS_RDONLY;
+ }
+ ext3_commit_super(sb, es, 1);
  if (test_opt(sb, ERRORS_PANIC))
   panic("EXT3-fs (device %s): panic forced after error\n",
    sb->s_id);
- ext3_commit_super(sb, es, 1);
 }

 void ext3_error (struct super_block * sb, const char * function,
```

Subject: Re: [PATCH] ext3: wrong error behavior
Posted by Eric Sandeen on Tue, 05 Sep 2006 22:20:42 GMT
View Forum Message <> Reply to Message

Vasily Averin wrote:
> SWsoft Virtuozzo/OpenVZ Linux kernel team has discovered that ext3 error
> behavior was broken in linux kernels since 2.5.x versions by the following patch:
>
> 2002/10/31 02:15:26-05:00 tytso@snap.thunk.org
> Default mount options from superblock for ext2/3 filesystems
> http://linux.bkbits.net:8080/linux-2.6/gnupatch@3dc0d88eKbV9 ivV4ptRNM8fBuA3JBQ
>
> In case ext3 file system is mounted with errors=continue (EXT3_ERRORS_CONTINUE)
> errors should be ignored when possible. However at present in case of any error
> kernel aborts journal and remounts filesystem to read-only. Such behavior was
> hit number of times and noted to differ from that of 2.4.x kernels.

I've also noticed this differing behavior,

 http://marc.theaimsgroup.com/?l=linux-ext4&m=11537696682 1953&w=2

It passed w/o comment. :)

Unless Ted had a specific reason for changing the behavior, 2.4 and 2.6
should probably be brought into line. Calling ext3_commit_super()
before the panic may defeat (some of) the purpose of the panic option,
though, which is to preserve as much state as possible at the time of
the error for later analysis...

-Eric

Subject: Re: [PATCH] ext3: wrong error behavior
Posted by vaverin on Wed, 06 Sep 2006 06:53:02 GMT

Eric Sandeen wrote:
> Vasily Averin wrote:
>> In case ext3 file system is mounted with errors=continue
>> (EXT3_ERRORS_CONTINUE)
>> errors should be ignored when possible. However at present in case of
>> any error
>> kernel aborts journal and remounts filesystem to read-only. Such
>> behavior was
>> hit number of times and noted to differ from that of 2.4.x kernels.
>
> I've also noticed this differing behavior,
>
> http://marc.theaimsgroup.com/?l=linux-ext4&m=11537696682 1953&w=2
>
> It passed w/o comment. :)

I would note that one of our developers has noticed it 3 year ago.
 http://marc.theaimsgroup.com/?l=linux-kernel&m=104824948 712104&w=2
it was w/o comments too.

> Unless Ted had a specific reason for changing the behavior, 2.4 and 2.6
> should probably be brought into line.

I did not found any arguments explaining Ted's reasons, I believe he did not
noticed that he had inversed this condition. If otherwise he had some important
reasons, there is another bug: he must change the documentation at least.

I would like to add that our customers are really discontented by this change.
Now they pays a high price for any minor ext3 errors: they should stop any
processes that uses this partition, check the partition, restart the processes
again... Downtime is too long and there is not any workarounds.

> Calling ext3_commit_super()
> before the panic may defeat (some of) the purpose of the panic option,
> though, which is to preserve as much state as possible at the time of
> the error for later analysis...

On the other hand if error is not saved on the disc, fsck will not check this
filesystem automatically on the next reboot and therefore node will mount
corrupted filesystem, it's dangerous too. In the worst scenario the node find
the same error on the disc and rebooted again and it is real nightmare for
remote admins.

I would note that Eric Lammerts has tried to fix this issue, however I'm not
sure that he has achieved a success:
 http://marc.theaimsgroup.com/?l=ext3-users&m=11065177390 7882&w=2

Also I would like to add that Andrew Morton has signed off the patch and added it to the -mm tree.

Thank you,
Vasily Averin

SWsoft Virtuozzo/OpenVZ Linux kernel team
