Subject: Re: [RFC PATCH ext3/ext4] orphan list corruption due bad inode
Posted by vaverin on Tue, 05 Jun 2007 06:49:10 GMT
View Forum Message <> Reply to Message

Eric Sandeen wrote:
> Vasily Averin wrote:
>> Bad inode can live some time, ext3_unlink can add it to orphan list, but
>> ext3_delete_inode() do not deleted this inode from orphan list. As result
>> we can have orphan list corruption detected in ext3_destroy_inode().
>
> Ah, I see - so you have confirmed that this inode does have bad ops...? I did
> notice on another orphan inode bug investigation that ext3_inode_delete won't
> clear orphan from a bad_inode...

yes, inode dump shows that i_fop = &bad_file_ops, and IMHO it's possible only
for bad inode.

>> However it is not clear for me how to fix this issue correctly.
>>
>> As far as i see is_bad_inode() is called after iget() in all places
>> excluding ext3_lookup() and ext3_get_parent(). I believe it makes sense to
>> add bad inode check to these functions too and call iput if bad inode
>> detected.
>
> That seems reasonable to me in any case, though as Andrew said, do you know
> for sure how the bad inodes were getting on the orphan list...?
>
> Is it possible that they were recycled after being freed while still on the
> orphan list, as in my previous reply to your previous message?

This incident has been occurred on Virtuozzo kernel based on latest RHEL4
2.6.9-55.el5, and it have your patch applied. btw thank you very much for this fix.

Unfortunately I do not know how this inode become bad, but inode dump in
ext3_destroy_inode shows that it is.

As far as I understand ext3_read_inode has been noticed that raw inode has
i_links_count=0 and therefore inode was marked as bad. However I have no any
ideas is it possible to have an inode on disk with i_links_count=0.

Thank you,
 Vasily Averin