pfSense - Bug #4523

master.passwd/group file corruption may occur after kernel panic or unclean shut down

03/14/2015 11:27 PM - Chris Buechler

Status:	Resolved	**Start date:**	03/14/2015
Priority:	Very High	**Due date:**	
Assignee:	Chris Buechler	**% Done:**	0%
Category:	Operating System	**Estimated time:**	0.00 hour
Target version:	2.2.4		
Affected Version:	2.2.x	**Affected Architecture:**	

Description

After a kernel panic, the passwd and/or group files may be corrupt. This seems to be a problem common to FreeBSD 10.1, and potentially fsck-related.

Associated revisions

Revision decb0b11 - 04/14/2015 09:52 PM - Chris Buechler

set forcesync to 1 by default for now, testing potential impact for Ticket #4523.

Revision 5d5fff67 - 04/14/2015 09:55 PM - Chris Buechler

set forcesync to 1 by default for now, testing potential impact for Ticket #4523.

Revision ab37f56f - 05/25/2015 07:25 AM - Ermal Luçi

Disable this tunable for now. Ticket #4523

Revision 85a37985 - 05/25/2015 07:25 AM - Ermal Luçi

Disable this tunable for now. Ticket #4523

Revision f2e36920 - 05/27/2015 03:20 AM - Ermal Luçi

Ticket #4523 Run fsck with -C flag and alway in foreground during bootup to prevent any issues that might schedule background mode.

Revision 7fd93993 - 05/27/2015 03:49 AM - Ermal Luçi

Ticket #4523 Major changes to how fsck is done.

Follow best practice of using fsck from FreeBSD rc.d/fsck script.
This means run preen mode first and later on try forcefully to fix issues.

Try to give as much information during boot on the status of the filesystem.

Revision fc123231 - 05/27/2015 03:50 AM - Ermal Luçi

Ticket #4523 Run fsck with -C flag and alway in foreground during bootup to prevent any issues that might schedule background mode.

Revision 427d36b4 - 05/27/2015 03:51 AM - Ermal Luçi

Ticket #4523 Major changes to how fsck is done.

```
Follow best practice of using fsck from FreeBSD rc.d/fsck script.
    This means run preen mode first and later on try forcefully to fix issues.
```

```
Try to give as much information during boot on the status of the filesystem.
```

Follow best practice of using fsck from FreeBSD rc.d/fsck script.
This means run preen mode first and later on try forcefully to fix issues.

Try to give as much information during boot on the status of the filesystem.

Conflicts:
etc/inc/services.inc

Revision 36314cba - 05/27/2015 09:13 AM - Ermal Luçi

Revert "Disable this tunable for now. Ticket #4523"

This reverts commit ab37f56f404a41dc5c5c26a83d594f0f883bd88d.

Revision face47a5 - 05/27/2015 09:14 AM - Ermal Luçi

Revert "Disable this tunable for now. Ticket #4523"

This reverts commit 85a37985b15c7a7c935d0028aa7a520110c2e649.

Revision ed97bf78 - 06/10/2015 11:49 AM - Jim Pingle

Activate sync for the root slice in fstab during upgrade. Ticket #4523

Revision b8947f8f - 06/10/2015 11:54 AM - Jim Pingle

Activate sync for the root slice in fstab during upgrade. Ticket #4523

Revision 2300307e - 07/03/2015 11:11 PM - Chris Buechler

de-activate sync on upgrade where it's enabled now that the root passwd/group problem is fixed. Ticket #4523

Revision aaf07882 - 07/03/2015 11:11 PM - Chris Buechler

de-activate sync on upgrade where it's enabled now that the root passwd/group problem is fixed. Ticket #4523

History

#1 - 03/14/2015 11:44 PM - Phillip Davis

Jeremy already put a similar bug report #4519 some hours ago.

#2 - 03/16/2015 06:23 PM - Chris Buechler

we'll keep this one, it's more specific to the root problem at hand, closed other as duplicate

#3 - 04/02/2015 04:37 PM - Chris Buechler

- Target version changed from 2.2.2 to 2.2.3

#4 - 05/22/2015 11:28 AM - Jim Pingle

I thought I added this here a while back but apparently not.

I have tried combinations of:

- Soft updates
- SU+J
- Sync vs Async
- Disabling atime (mostly to see if less writes helped)

Each time it only took a handful of power pulls (usually 1-3) or manually initiated panics (sysctl debug.kdb.panic=1) before /etc was corrupted.

Sometimes whole files are swapped, other times portions of them are overlapping.

#5 - 05/22/2015 09:49 PM - Chris Buechler

- Subject changed from /etc file corruption may occur after kernel panic to /etc file corruption may occur after kernel panic or unclean shut down

this is replicable with just an unclean shut down

#6 - 05/23/2015 02:59 AM - Kill Bill

Reading this like this:
- https://forums.freebsd.org/threads/freebsd-on-ufs-preventing-data-loss-on-crash.30683/
- https://bugs.freebsd.org/bugzilla/show_bug.cgi?id=183042

I am rather surprised this has not been observed earlier and wondering which was the last fsck version that was not completely braindead. Now, pulling the plug is one thing, however screwing the FS on kernel panic or unclean shutdown is really just WTF.

#7 - 05/25/2015 10:09 AM - Ermal Luçi

The installer and nano has been switched to SU+J same as default FreeBSD.

#8 - 05/27/2015 03:54 AM - Ermal Luçi

- Status changed from Confirmed to Feedback

Improvements on how filesystem check/correction is being done have been merged which should help with corruption to not be as easily reproducible.

#9 - 05/28/2015 12:07 AM - Chris Buechler

- Status changed from Feedback to Confirmed

still an issue

#10 - 05/29/2015 11:21 PM - ky41083 -

Nano using SU+J = bad. Either go back to plain sync or just SU. All journaling does is **double all meta-data writes** with the goal of making fsck faster.

Doubling meta-data writes is very bad for the flash memory nano is almost always installed on. We know this. The entire goal of SU+J vs SU is essentially a faster fsck run, and is designed with spinny drives and minimizing head thrashing in mind. This should all be gladly sacrificed for extended flash storage life.

Going out on a limb, has anyone tried running fsck multiple (2+) times back to back? I've dealt with this issue on Linux running from flash, fsck had to be called at LEAST twice in a row to get a properly clean filesystem. Some parts had to be fixed to further fix additional parts, etc. Eventually my fix

was to write a script that ran it 3 times in a row, haven't had an issue for years since.

Seems to be an issue on FreeBSD as well:
https://forums.freebsd.org/threads/softupdate-with-journaling-decrease-reliability.39125/

Hope this helps.

#11 - 05/30/2015 01:33 AM - Chris Buechler

- Subject changed from /etc file corruption may occur after kernel panic or unclean shut down to master.passwd/group file corruption may occur after kernel panic or unclean shut down

- Priority changed from High to Very High

updated subject to narrowed down problem.

With SU, with or without J, you end up with 0 byte master.passwd, passwd, group, pwd.db, spwd.db. Or some subset of those. Without SU, you end up with master.passwd and/or group corrupted, containing parts of other files in /etc/.

It's replicable on stock FreeBSD 9.0 through 11-CURRENT by running the following:

```
#!/bin/sh

/usr/sbin/pw userdel -n 'admin'
/usr/sbin/pw groupadd all -g 1998
/usr/sbin/pw groupadd admins -g 1999
/usr/sbin/pw groupmod all -g 1998 -M ''
echo \$6\$O/T6GYkcgYvOBTGm\$KvOh3zhFKiA6HMEPktuImAI8/cetwEFsgj7obXdeTcQvn6mhs50HgkWt6nfnxNhTIb2w4Je6dqdKtARavx
ThP1 | /usr/sbin/pw usermod -q -n root -s /bin/tcsh -H 0
echo \$6\$O/T6GYkcgYvOBTGm\$KvOh3zhFKiA6HMEPktuImAI8/cetwEFsgj7obXdeTcQvn6mhs50HgkWt6nfnxNhTIb2w4Je6dqdKtARavx
ThP1 | /usr/sbin/pw useradd -m -k /etc/skel -o -q -u 0 -n admin -g wheel -s /bin/sh -d /root -c 'System Admini
strator' -H 0
/usr/sbin/pw unlock admin -q
/usr/sbin/pw groupmod all -g 1998 -M '0'
/usr/sbin/pw groupmod admins -g 1999 -M '0'
```

then power cycling the system. If using SU, you'll end up with 0 byte files. Without SU, you'll have corrupted files containing parts of some other file(s) in /etc.

Still investigating, we'll be reporting specifics upstream soon.

#12 - 05/30/2015 01:49 AM - Kill Bill

Chris Buechler wrote:

> If using SU, you'll end up with 0 byte files. Without SU, you'll have corrupted files containing parts of some other file(s) in /etc.

Is this **before** or **after** running fsck? IOW, is UFS just unusable, or is fsck being full retard?

Meanwhile - is this ZFS howto still valid for 2.2.x? Is this howto still valid for 2.2.x? https://forum.pfsense.org/index.php?topic=71953.0

Since, frankly I've had enough of this. I haven't created a user/group on any of the boxes that get randomly screwed for **ages**. WTH these files are being constantly damaged?

#13 - 05/30/2015 01:57 AM - Chris Buechler

That's after fsck (including after multiple runs). They aren't "constantly damaged", only after unclean shut downs, and only a minority of the time at that. The above shell script replicates 100% reliably on stock FreeBSD, but in real world usage it's less likely you'll hit it nearly that reliably. It seems to affect slower drives more often than faster ones. Maybe 1 in a handful of times pulling the plug up to one in 2-3 dozen times.

no idea if those ZFS instructions still work. They might. We haven't done anything to intentionally disable that, but we don't test it either.

#14 - 05/30/2015 02:12 AM - Kill Bill

Chris Buechler wrote:

> That's after fsck (including after multiple runs).

Well what I meant is actually whether it's fsck screwing those files or whether they were truncated/mangled already before running fsck. (IOW, power off the system and stick the drive into some other box and mount it, instead of trying to boot from it and letting fsck do its sloppy job.)

#15 - 06/01/2015 02:25 AM - Jim Thompson

It's not fsck.

it's likely a bug in SU (with or without journaling.)

the fix (for now) is to mount / "sync" on all pfSense installs (nano or full).

> WTH these files are being constantly damaged?

because of the internals of what the "pw" command does, crossed by some bug in UFS that has yet to be found.

#16 - 06/01/2015 06:39 AM - Phillip Davis

"sync" seems like "a good thing" on root file system "/" for pfSense use cases anyway. pfSense uses would not modify stuff in "/" very often at run time, and thus having that root file-system activity be synchronous would have almost imperceptible performance impact. Might as well use "sync" whether the underlying bug here is fixed or not.
Real-time file system activity on pfSense is mostly to /var and /tmp for updating DHCP lease files, writing log entries and such like, plus packages that cache stuff (Squid...).

#17 - 06/01/2015 11:15 AM - Kill Bill

Updated ZFS howto for people who are on full install and are simply tired of this... https://forum.pfsense.org/index.php?topic=94656

(One would assume UFS to be somewhat mature after all those years... ugh.)

#18 - 06/01/2015 11:58 AM - Denny Page

Does sync actually avoid the issue? Update 4 suggested that this was not the case...

Sync for root fs generally seems like a good idea, but only if is not updated infrequently. Given that the default install has var in the root fs, this would not be a good choice if there are packages that update /var frequently (ntopng, squid, etc.).

#19 - 06/01/2015 12:06 PM - Jim Pingle

It was apparently an error in my notes... I looked back at a forum post I made when I first tested that mid-April and and I had noted that although fsck still ran and found issues with sync, the files remained intact: https://forum.pfsense.org/index.php?topic=88439.msg511477#msg511477

#20 - 06/01/2015 10:59 PM - Chris Buechler

sync definitely avoids the root issue. I have a system that's now upwards of 1000 power cycles with 0 issues with sync.

The root problem seems to be within pw rather than anything to do with UFS. We'll pursue a proper fix there. In the mean time, setting sync does fix the problem and shouldn't have a negative impact for our use cases.

I updated the installer to set sync. We'll need to add code to add that to fstab on upgraded systems.

#21 - 06/01/2015 11:58 PM - Denny Page

Wow, there's a name I haven't heard in 20+ years.

#22 - 06/02/2015 11:38 AM - Jim Thompson

Kill Bill wrote:

> Updated ZFS howto for people who are on full install and are simply tired of this... https://forum.pfsense.org/index.php?topic=94656
>
> (One would assume UFS to be somewhat mature after all those years... ugh.)

Do let me know when you have sufficient experience with filesystems to decide if something is "mature" or not.

#23 - 06/02/2015 11:41 AM - Jim Thompson

Denny Page wrote:

> Wow, there's a name I haven't heard in 20+ years.

Yes, and cmb shouldn't have quoted a private communication without permission. I've edited his post.

#24 - 06/07/2015 06:55 PM - Ermal Luçi

- Status changed from Confirmed to Feedback

#25 - 06/08/2015 01:28 AM - Kill Bill

There's something badly broken on nanobsd with this...

https://forum.pfsense.org/index.php?topic=94900.0

#26 - 06/09/2015 03:00 PM - Jim Pingle

- Status changed from Feedback to Confirmed

Moving this back to Confirmed since the upgrade code is still missing for existing installations, and it appears as though on the 2.2.3 snapshots the sync flag is not being added to the root slice during install. I see the code in the bsdinstaller repo but it's not in the snapshots.

#27 - 06/10/2015 07:04 PM - Ermal Luçi

- Status changed from Confirmed to Feedback

Installer has been updated for new snaps and upgrade code been put in place.

#28 - 06/17/2015 12:21 AM - Chris Buechler

- Status changed from Feedback to Resolved

fixed. We'll again verify as part of the release test matrix on each install type.

#29 - 07/03/2015 11:11 PM - Chris Buechler

- Status changed from Resolved to Feedback

- Target version changed from 2.2.3 to 2.2.4

- Affected Version changed from 2.2 to 2.2.x

this is adequately worked around in 2.2.3 with the usage of sync. Now that we have a proper fix for pw in 2.2.4, and sync has been removed from the installer, and upgrade code changed to remove sync where it's enabled, moving this back to feedback to confirm those sync changes.

#30 - 07/05/2015 09:10 AM - Thomas X

- File 2015-07-05_pfsense_2.2.3_corrupted_seriallog.txt added

Today I had a power loss with pfSense 2.2.3 AMD64 NanoBSD, which seems to have corrupted the installation. The system was upgraded from

pfSense 2.2.1 AMD64 NanoBSD 7 days ago.

Afterwards, when power was available again, the system didn't came up correctly, the web frontend showed an internal server error, login was not possible even with serial console.

See the attached log which was recorded when doing another hard reset. Switching the bootup slice made my day, now running 2.2.1 just fine.

I'm not sure if this corruption is releated to this issue, please ignore if it's not. I was just wondering why this could happen although sync was added in 2.2.3.

Best regards
Thomas

#31 - 07/05/2015 09:16 AM - Thomas X

One addition: Filesystem has been in standard NanoBSD mode (ReadOnly) when the loss of power appeared.

#32 - 07/05/2015 11:47 AM - Kill Bill

Thomas X wrote:

> I was just wondering why this could happen although sync was added in 2.2.3.

Probably because the sync mount option was never a proper fix in the first place; plus performs absolutely horribly even on full installs with fast SATA HDDs. Try with latest 2.2.4 snapshots.

#33 - 07/05/2015 09:36 PM - Jim Thompson

The sync option was not an **optimal** fix, but it was a proper fix, as it does fix the corruption issue, and was what we could get done (with testing) prior to the correct fix (which is in 2.2.4 **and** in FreeBSD.)

#34 - 07/14/2015 11:00 AM - Jim Thompson

- Assignee set to Chris Buechler

#35 - 07/14/2015 02:02 PM - Chris Buechler

- Status changed from Feedback to Resolved

sync no longer added to new installs, and confirmed the upgrade code removes it where it's set and doesn't change anything where it isn't.

Files

2015-07-05_pfsense_2.2.3_corrupted_seriallog.txt	15.1 KB	07/05/2015	Thomas X