Friday, July 31. 2009

Less known Solaris features: The curious case of the /tmp in Solaris

This article isn't really about a feature, it's about a directory and its misuse. Furthermore it's an article about different default configurations, that lead to misunderstandings. This is a pretty old hat for experienced Solaris admins (many of them learned it the hard tour), but it seems to be totally unknown to many admins new to the business or for people switching from Linux to Solaris, as many distribution are configured in a different way per default. A reader of my blog just found a 2GB .iso in /tmp on a Solaris system and that's not really a good idea. A few days ago, a user in twitter had vast problems with memory usage on a system which boiled down to a crowed /tmp.

tmpfs and its usage
In Linux the /tmp directory is a normal directory. Most often this is just a part of the partition containing root. Many admins consider /tmp as an additional home directory and put data on it they want to use later on. When you really think about the idea behind /tmp is meant as a scratch space, where an application can put some data to process it later. Many applications use this /tmp space to write temporary data to.

Solaris introduced something called tmpfs many years ago, It's a memory based file system. Linux has something with similar functionality and the same name. It´s just used differently in the default configuration of many distributions and there are a vast amount of articles in the web that suggests to configure /tmp in Linux to use the tmpfs filesystem as this can give you advantages for example with a Mysql for example. (when there is a reasonable amout of SELECTs opting for filesort)

The tmpfs is not a ramdisk. A ram disk just resides in the RAM, a tmpfs resides in the virtual memory, thus it uses the swap when the memory is needed for something else. Furthermore it doesn't have a predefined size, albeit you can define a maximum size for the tmpfs to ensure that someone who writes into the /tmp can't eat away all your virtual memory.

It's called tmpfs because everything you write into it is temporary, the next boot or unmount will kill all the files on it. When you look at the mount table of a Solaris System you will recognize, that the usual locations for such temporary files are mounted tmpfs:
jmoekamp@a380:/var$ mount | grep "swap"
/etc/svc/volatile on swap read/write/setuid/devices/xattr/dev=4f00001 on Fri Jul 31 06:33:33 2009
/tmp on swap read/write/setuid/devices/xattr/dev=4f00002 on Fri Jul 31 06:34:14 2009
/var/run on swap read/write/setuid/devices/xattr/dev=4f00003 on Fri Jul 31 06:34:14 2009

Keeping these file systems in virtual memory is a reasonable choice. The stuff in this directory is normally stale after a reboot, most of the time the files are many, but rather small and putting them on disk would just eat away your IOPS budget on your boot disks. As the file system resides in memory, it's much faster and that really helps on jobs with many small files. A good example is compiling software when you use /tmp on as the TMPDIR.

All this advantages come with a big disadvantage, when you are not aware of the nature of the /tmp directory. I assume, you already know why using /tmp for storing ISOs is a bad idea. It eats away your memory and later on your swap. And for all the experienced admins: When someone has memory problems, ask at first about the /tmp directory, we tend to forget about this, as we've learned this lesson a long time ago and thus don't think about this problem. When you really need a temporary place to store data in it for a while you should use /var/tmp. This is a directory on a normal disk based filesystem and thus its content it boot persistent.

Configuring the maximum size of the tmpfs
When you want to enforce a limit on the size of a tmpfs this pretty easy task . You can do this with a mount option. Let's assume you want to provide the /var/application_scratch pad directory for an application, but you want to be sure, that this application can't ruin your day by eating away all your memory:

jmoekamp@a380:/var# mkdir /var/application_scratchpad
jmoekamp@a380:/var# mount -f tmpfs -o size=128m swap /var/application_scratchpad/
jmoekamp@a380:/var# mount | grep "application"
/var/application_scratchpad on swap read/write/setuid/devices/xattr/size=128m/dev=4f00004 on Fri Jul 31 09:32:26 2009

When you want to make a boot persistent change to the maximum size of the /tmp directory, you have to configure this in the /etc/vfstab:
swap   -   /tmp   tmpfs   -   yes   size=512m

Conclusion
I hope it´s now clear, why you shouldn´t use /tmp as a place for storing big files, at least move them directly somewhere else, when you use /tmp as the target directory to ssh a file to a system. At the other side the /tmp in virtual memory gives you some interesting capabilities to speed up your applications. As usual, everything has two sides: Making it default in Solaris gives you speedups per default. But when you are unaware of this default situation your virtual memory may be used for collection of videos Obviously the same is valid, when you configure our Linux system with a tmpfs based /tmp and don't tell it to your fellow admins.

Do you want to learn more?
manpages
man page: tmpfs(7f)
man page: mount_tmpfs(1M)

 Posted by Joerg Moellenkamp in English, Solaris at 11:13

some Linux uses tmpfs for /tmp as well.

On Linux you can use /var/tmp to be on the safe side (or even better use a user specific temp). FHS also requires /var/tmp to persist over boots, whereas /tmp might be cleaned.

Linux has an tmpfs mounted under /dev/shm for shared memory segments, one could link to it for /tmp, but it is better to mount two instances, so you can specify different size restrictions.

Oracle uses /dev/shm on 32bit also for windowing larger SGAs. One has to make sure the window is mlockable for it.

Gruss
Bernd
   Anonymous on Jul 31 2009, 11:56

Then the same things are valid vor Linux as well: Don´t use it as a storage for your isos
   Anonymous on Jul 31 2009, 12:35

Danke!
   Anonymous on Jul 31 2009, 13:30

Bitte Wollte Dir noch eine Mail dazu schreiben, aber war noch nicht dazu gekommen
   Anonymous on Jul 31 2009, 13:42

I wish you would stop bashing on Linux or Linux admins -- I knew the difference between the /temporary/ solutions from year one.

When I saw the headline in the feedreader I was happy (new content from Joerg which hopefully extends my knowledge about Solaris or offers a differnt, interesting view).
After the first paragraph I think: a grumpy old s*ck*r who can't write articles without bashing anything or anyone for a single month.

This is sad. I hope one day you'll find a cure for your cynicism.

But in the mean time you might continue to make jokes about Ubuntu Users in front of your fellow grumpy horde.
I laugh about the Solaris fanboys who feature a ton of features aside from dtrace and ZFS which is based on GPL-Software which most of the time was created on linux desktops -- nothing to be ashamed of, but why the strong need to stress the differences?

I'm interested in both worlds but to be honestly the Solaris folk seems more and more like a strange mixture of marketing and bitterness too me. Not deserving to be called a "sunny" person.

Maybe someday you get inspired by professional cynic Harald Schmidt who stopped making jokes about the pope for only one reason:
it's too easy.

Have a nice day.
   Anonymous on Jul 31 2009, 13:54

Sorry, i didn´t want to bash Linux or Linux Admins. But as it's default on Solaris and it's not on Linux, many Linux admins simply don't think about this little known fact. And even most Solaris Admins, who know this, learned it the hard tour.

Perhaps i should explain how i got the idea to write this article: I've got a mail from a reader which said "Hey, could you please explain

the differences between Linux and Solaris /tmp. I just found an ISO again in a /tmp"

You have to handle the /tmp differently in Linux and Solaris. If you don't do it, there are several pitfalls you can fall into. And many people don't know this difference, otherwise the isos in Solaris /tmp are somewhat unexplainable for me.

On most Linuxes it's the decision of the admin to switch over to a tmpfs based /tmp, the admin is aware of this situation and uses this accordingly. When such an admin switches over from Linux to Solaris, most would't think about the location of /tmp. I have to admit, that i wouldn't think about the /tmp on harddisk when using Linux und would wondering about the fact, that it isn't on a tmpfs disk, just because i'm accustomed to a different situation.

All i wanted to say with this article is: Attention, Solaris /tmp behaves differently than most Linux /tmp. Act accordingly.
    Anonymous on Jul 31 2009, 14:15

There is a reason for making tmpfs a default for /tmp in Solaris and that is that traditionally ufs is very paranoid about making any metadata operations in the filesystem, so file creation and deletion are synchronous - read: slow. On the other hand applications such as a compiler and linker create tons of temporary files in /tmp to pass data between the stages of a compile, and many other programs also use temporary files in $TMPDIR (/tmp if unset).

The immediate solution for this in Solaris was tmpfs, as a kind of stopgap measure, because it was easy to implement - at least easier than optimizing ufs metadata ops. Unfortunately it also was good enough and Sun was busy with ZFS so they never came around to actually fixing ufs and so we still have an 'unusual' /tmp in Solaris.
    Anonymous on Jul 31 2009, 16:21

Yes, Sun engineering has this tendency, i would even say "overparanoid" at some occasions. But i tend to say, that you can't be too paranoid in regard of file systems. I don't know, who coined the phrase, but one of the key differences between Linux and Solaris is : "Solaris is pessimistic in its assumptions, Linux is optimistic in its assumption". Both assumptions are born out of the history of both systems and both have their advantages and disadvantages.
    Anonymous on Jul 31 2009, 16:46

Thank you for explaining the history/motivation of the article.

I totally agree that there are a lot admins who are unaware of this features -- on both sides I guess. (I inherited my Linux setups from a former Solaris Admin and his assistant who made /var/tmp links to /tmp, sadly without documenting the change and the motivation (It's possible he explained the difference to the users back than and together they decided against it!)

I absolutely agree that it is important to explain the differences between Linux and Solaris to interested admins and the reasons (standards being a very good point!) behind it.
Btw. I tried to read this article with relaxed emotions but -- sure, could be I wanted it to read that way, but it sounded a bit to generalized towards Linux (funnily enough "standards" being its weakest spot, generally disable that).
Because I think I'm between the UNIX/BSD and the Linux camps -- dreaming of a system that was created by cherry picking technologies, undetermined: loving & _hating_ GNU tools, prefering Sol. & BSD quality of documentation --
I though it was reasonable to give feedback.

Eventually, I never mistrusted the essence of your article.

Read you soon.
    Anonymous on Jul 31 2009, 16:56

the intent and motivation behind tmpfs(7FS) in Solaris ?

really an oldtimer, easy enough:

http://www.solarisinternals.com/si/reading/tmpfs.pdf
    Anonymous on Aug 1 2009, 15:26

Benjamin,
two things: First I really didn't notice any bashing in the article - in fact, the reasoning was very valid. It's true, that in Linux /tmp is filesystem based and in Solaris virtual memory based, therefore people who have more experience with Linux tend to misunderstand Solaris' /tmp. That's a fact, I don't see why saying that should be considered bashing.
Second if you think Solaris users are grumpy, you probably don't visit Linux sites very often, where people constantly talk about some "world domination" and about how Unix (and all other OSes actually) should already die and "get out of the way". Personally I feel, that Solaris users when compared to Linux guys are actually very nice and friendly bunch of people
    Anonymous on Aug 1 2009, 20:31

This matches my experience: I visited a customer a while ago, who asked me "Why do you Solaris? You fragmenting the Unix space" expressed with strong and open hostility.
    Anonymous on Aug 2 2009, 08:57