

## Re: MFS ROOT over 100MB

---

*Source:* <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/current/2008-03/msg00605.html>

---

- *From:* "Peter Wemm" <[peter@xxxxxxxx](mailto:peter@xxxxxxxx)>
  - *Date:* Fri, 28 Mar 2008 17:53:32 -0700
- 

On Fri, Mar 28, 2008 at 1:12 PM, Julian Elischer <[julian@xxxxxxxxxxxxxx](mailto:julian@xxxxxxxxxxxxxx)> wrote:

Miroslav Chlastak wrote:

- > Hi,
- >
- >
- > it's possible create MFS ROOT over 100MB?
- > If I create MFS ROOT less then 90MB, system boot normally.
- > With size 90MB–100MB of MFS ROOT kernel panic with boot and
- > over 100MB system is rebooted in loop.
- >

I suspect that the MFS root is loaded into kernelVM which is a limited resource. in fact it must be I think because it is linked with the kernel by the loader.

There isn't a lot of spare kernel virtual memory by default, but there are other things that are likely to be being hit first.

First, have a look at the NKPT kernel compile option. This is probably what is being run into.

By default, for an i386, non-PAE kernel, the value is 30. This is the initial number of page table pages that are statically allocated for early kernel bootup. Once the VM system is fully up and running, we can allocate more. But during early bootup, this is what we have.

In this case, 30 page table pages, and there are 1024 entries per page table page, with 4K per entry. ie:  $30 \times 1024 \times 4k = 122MB$ . If the VM system isn't fully up and running by the time this 122MB of kernel VM is consumed, bad things will happen.

This is where the 90–100MB threshold in MFS root image fits in. While there is 1GB of KVM available in total, the MFS image has to fit inside this 122MB chunk, **\*\*and\*\*** the kernel has to complete bootstrap with what's left.

On an i386 kernel without PAE, the KVA\_PAGES option is 256. It is

Re: MFS ROOT over 100MB

also means  $256 \times 1024 \times 4k = 1GB$ . In theory, you could increase the NKPT option to close to this amount. I don't know what the exact threshold is, but it'll be pretty close.

ie: you should be able to add 'options NKPT=220' to the kernel config. This would let you load a mfsroot in the order of about 900MB, **\*\*IN THEORY\*\***. The problem is that you'll then run into other kernel VM exhaustion. It still has a total of 1GB available by default. This is changeable too.

If you're trying to do a massive mfsroot, you'll probably have to increase KVA\_PAGES too. For example, 'options KVA\_PAGES=512' gives you 2GB of kernel VM, and 2GB left for user processes. You could then set NKPT to 300+ and easily get a 1GB mfsroot ramdisk loaded.

For PAE, page table pages have different numbers of entries. You have to double the numbers because they're in units of  $512 \times 4k$ , or 2MB per page table page. So with PAE, you'd need 'options KVA\_PAGES=1024' and 'options NKPT=600' to get the same 2G/2G user/kernel split, and have the 1gb mfsroot image.

I hope this is helpful. You'll need to experiment to make this work and not waste too much memory. There is no way to avoid compiling custom kernels to do this.

- > MFS ROOT image is ok (i can mount this image via mdconfig and mount).
- > I'am testing this on RELENG\_6\_2 and RELENG\_7\_0.
- >
- > The same problem like this:
- > <http://lists.freebsd.org/pipermail/freebsd-stable/2008-February/040626.html>
- >
- > Thanks,
- > Mira

—  
Peter Wemm – peter@xxxxxxxx; peter@xxxxxxxxxxxx; peter@xxxxxxxxxxxxxxxx  
"All of this is for nothing if we don't go to the stars" – JMS/B5  
"If Java had true garbage collection, most programs would delete themselves upon execution." — Robert Sewell  
**\*\*WANTED TO BUY/TRADE: Garmin Streetpilot 2650 or 2660. Not later model! \*\***

---

freebsd-current@xxxxxxxxxxxxx mailing list  
<http://lists.freebsd.org/mailman/listinfo/freebsd-current>  
To unsubscribe, send any mail to "freebsd-current-unsubscribe@xxxxxxxxxxxxx"