Subject: *SOLVED* optimization on privvmpages failcnt
Posted by victorskl on Thu, 09 Nov 2006 09:08:33 GMT
View Forum Message <> Reply to Message

```
-bash-3.00# cat /proc/user_beancounters
Version: 2.5
      uid  resource          held    maxheld    barrier      limit    failcnt
      102: kmemsize       12114638   12647900   34333450   37766795          0
           lockedpages          0          8       1676       1676          0
           privvmpages     620219     667841     620896     682985        886
           shmpages           573       2189      62089      62089          0
           dummy                0          0          0          0          0
           numproc             66         82        838        838          0
           physspages       401426     422359          0 2147483647          0
           vmguarpages          0          0     620896 2147483647          0
           oomguarpages     401426     422359     620896 2147483647          0
           numtcpsock           15         32        838        838          0
           numflock             17         26       1000       1100          0
           numpty                1          1         83         83          0
           numsiginfo            0          4       1024       1024          0
           tcpsndbuf        134160     261612    8012035   11444483          0
           tcprcvbuf        245760     376912    8012035   11444483          0
           othersockbuf     155632     435680    4006017    7438465          0
           dgramrcvbuf           0       8380    4006017    4006017          0
           numothersock        131        161        838        838          0
           dcachesize            0          0    7498066    7723008          0
           numfile            2019       2651      13408      13408          0
           dummy                0          0          0          0          0
           dummy                0          0          0          0          0
           dummy                0          0          0          0          0
           numiptent           39         39        100        100          0
-bash-3.00#
```

Hi i was trying and i rise it up all the way 2G+.. but failcnt very oftenly hit on privvmpages.. I knew it related to applications. most probably, Apache.. Can guide me how could i trace effectively..

Thanks a lot..

Subject: Re: optimization on privvmpages failcnt
Posted by dim on Thu, 09 Nov 2006 09:27:52 GMT
View Forum Message <> Reply to Message

It is definitely bug. Please, fill it at the http://bugzilla.openvz.org in order to force us to don't miss it. Please, post as much info as possible there, in particular used kernel version, it's config.

Concerning how to debug. If you have assumption, that this could be because of the apache, you could just start/stop it and check, does privvmpages held value is changed.

Subject: Re: optimization on privvmpages failcnt
Posted by xemul on Thu, 09 Nov 2006 10:41:58 GMT
View Forum Message <> Reply to Message

This doesn't look like a BUG for me. This looks like some task eats more memory than it should. Could you please give us access to this VE so we could track what has happened?

Subject: Re: optimization on privvmpages failcnt
Posted by victorskl on Thu, 09 Nov 2006 14:56:15 GMT
View Forum Message <> Reply to Message

Thanks thanks..

I used latest RHEL Kernel as stated at my signature... The VE use CentOS template.

[root@localhost ~]# uname -a
Linux localhost.localdomain 2.6.9-023stab032.1-enterprise #1 SMP Fri Oct 20 03:13:34 MSD 2006 i686 i686 i386 GNU/Linux
[root@localhost ~]#

These are for example, after that respective VE getting mad...

[root@localhost ~]# /usr/sbin/vzctl exec 102 service --status-all
bash: error while loading shared libraries: libtermcap.so.2: failed to map segment from shared object: Cannot allocate memory
[root@localhost ~]# /usr/sbin/vzctl enter 102
entered into VE 102
Unable to set raw mode: Interrupted system call
-bash: error while loading shared libraries: libtermcap.so.2: failed to map segment from shared object: Cannot allocate memory
exited from VE 102
[root@localhost ~]#

This is the last portion of dmesg..

Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.

Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
VZDQ: detached inode not in creation, orig 5, dev dm-0, inode 73828039, fs ext3
current 18761 (httpd), VE 102, time 77463.605075
 [<ed79c296>] vzquota_det_qmblk_recalc+0x256/0x270 [vzdquota]
 [<ed79c302>] vzquota_inode_qmblk_recalc+0x52/0x70 [vzdquota]
 [<ed79c573>] vzquota_inode_data+0xb3/0xf0 [vzdquota]
 [<ed79c449>] vzquota_inode_init_call+0x19/0x80 [vzdquota]
 [<021fb940>] ext3_delete_inode+0x0/0x120
 [<ed79e47f>] vzquota_initialize+0xf/0x20 [vzdquota]
 [<0219d983>] generic_delete_inode+0x173/0x190
 [<021996f6>] dput_recursive+0x56/0x230
 [<0217f333>] __fput+0x123/0x1b0
 [<0217d4c2>] filp_close+0x52/0xa0
 [<0217d57a>] sys_close+0x6a/0xa0
VZDQ: detached inode not in creation, orig 5, dev dm-0, inode 73827282, fs ext3
current 8753 (mysqld), VE 102, time 77463.644383
 [<ed79c296>] vzquota_det_qmblk_recalc+0x256/0x270 [vzdquota]
 [<ed79c302>] vzquota_inode_qmblk_recalc+0x52/0x70 [vzdquota]
 [<ed79c573>] vzquota_inode_data+0xb3/0xf0 [vzdquota]
 [<ed79c449>] vzquota_inode_init_call+0x19/0x80 [vzdquota]
 [<021fb940>] ext3_delete_inode+0x0/0x120
 [<ed79e47f>] vzquota_initialize+0xf/0x20 [vzdquota]
 [<0219d983>] generic_delete_inode+0x173/0x190
 [<021996f6>] dput_recursive+0x56/0x230
 [<0217f333>] __fput+0x123/0x1b0
 [<0217d4c2>] filp_close+0x52/0xa0
 [<0212e17a>] put_files_struct+0x6a/0xf0
 [<0212f6d0>] do_exit+0x1b0/0x560
 [<02138665>] __dequeue_signal+0x115/0x210
 [<0212fb60>] do_group_exit+0x40/0xb0
 [<0213ab3b>] get_signal_to_deliver+0x2ab/0x410
 [<0210877b>] do_signal+0x9b/0x180
 [<02194f04>] sys_select+0x444/0x530
 [<0210ab6d>] handle_IRQ_event+0x5d/0xb0
 [<02108897>] do_notify_resume+0x37/0x40
VZDQ: detached inode not in creation, orig 5, dev dm-0, inode 73828028, fs ext3
current 8753 (mysqld), VE 102, time 77463.652955
 [<ed79c296>] vzquota_det_qmblk_recalc+0x256/0x270 [vzdquota]
 [<ed79c302>] vzquota_inode_qmblk_recalc+0x52/0x70 [vzdquota]
 [<ed79c573>] vzquota_inode_data+0xb3/0xf0 [vzdquota]
 [<ed79c449>] vzquota_inode_init_call+0x19/0x80 [vzdquota]
 [<021fb940>] ext3_delete_inode+0x0/0x120
 [<ed79e47f>] vzquota_initialize+0xf/0x20 [vzdquota]
 [<0219d983>] generic_delete_inode+0x173/0x190
 [<021996f6>] dput_recursive+0x56/0x230

[<0217f333>] __fput+0x123/0x1b0
[<0217d4c2>] filp_close+0x52/0xa0
[<0212e17a>] put_files_struct+0x6a/0xf0
[<0212f6d0>] do_exit+0x1b0/0x560
[<02138665>] __dequeue_signal+0x115/0x210
[<0212fb60>] do_group_exit+0x40/0xb0
[<0213ab3b>] get_signal_to_deliver+0x2ab/0x410
[<0210877b>] do_signal+0x9b/0x180
[<02194f04>] sys_select+0x444/0x530
[<0210ab6d>] handle_IRQ_event+0x5d/0xb0
[<02108897>] do_notify_resume+0x37/0x40
VZDQ: detached inode not in creation, orig 5, dev dm-0, inode 73828029, fs ext3
current 8753 (mysqld), VE 102, time 77463.661461
[<ed79c296>] vzquota_det_qmblk_recalc+0x256/0x270 [vzdquota]
[<ed79c302>] vzquota_inode_qmblk_recalc+0x52/0x70 [vzdquota]
[<ed79c573>] vzquota_inode_data+0xb3/0xf0 [vzdquota]
[<ed79c449>] vzquota_inode_init_call+0x19/0x80 [vzdquota]
[<021fb940>] ext3_delete_inode+0x0/0x120
[<ed79e47f>] vzquota_initialize+0xf/0x20 [vzdquota]
[<0219d983>] generic_delete_inode+0x173/0x190
[<021996f6>] dput_recursive+0x56/0x230
[<0217f333>] __fput+0x123/0x1b0
[<0217d4c2>] filp_close+0x52/0xa0
[<0212e17a>] put_files_struct+0x6a/0xf0
[<0212f6d0>] do_exit+0x1b0/0x560
[<02138665>] __dequeue_signal+0x115/0x210
[<0212fb60>] do_group_exit+0x40/0xb0
[<0213ab3b>] get_signal_to_deliver+0x2ab/0x410
[<0210877b>] do_signal+0x9b/0x180
[<02194f04>] sys_select+0x444/0x530
[<0210ab6d>] handle_IRQ_event+0x5d/0xb0
[<02108897>] do_notify_resume+0x37/0x40
VZDQ: detached inode not in creation, orig 5, dev dm-0, inode 73828038, fs ext3
current 8753 (mysqld), VE 102, time 77463.671746
[<ed79c296>] vzquota_det_qmblk_recalc+0x256/0x270 [vzdquota]
[<ed79c302>] vzquota_inode_qmblk_recalc+0x52/0x70 [vzdquota]
[<ed79c573>] vzquota_inode_data+0xb3/0xf0 [vzdquota]
[<ed79c449>] vzquota_inode_init_call+0x19/0x80 [vzdquota]
[<021fb940>] ext3_delete_inode+0x0/0x120
[<ed79e47f>] vzquota_initialize+0xf/0x20 [vzdquota]
[<0219d983>] generic_delete_inode+0x173/0x190
[<021996f6>] dput_recursive+0x56/0x230
[<0217f333>] __fput+0x123/0x1b0
[<0217d4c2>] filp_close+0x52/0xa0
[<0212e17a>] put_files_struct+0x6a/0xf0
[<0212f6d0>] do_exit+0x1b0/0x560
[<02138665>] __dequeue_signal+0x115/0x210
[<0212fb60>] do_group_exit+0x40/0xb0

[<0213ab3b>] get_signal_to_deliver+0x2ab/0x410
[<0210877b>] do_signal+0x9b/0x180
[<02194f04>] sys_select+0x444/0x530
[<0210ab6d>] handle_IRQ_event+0x5d/0xb0
[<02108897>] do_notify_resume+0x37/0x40
VPS: 102: stopped
VPS: 102: started
Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
Fatal resource shortage: privvmpages, UB 102.
VPS: 102: stopped
VPS: 102: started
VPS: 102: stopped
VPS: 102: started
VPS: 102: stopped
VPS: 102: started

Indeed, this VE loaded Bind, Apache (mod_security and other modules..), MySQL, PostgreSQL, Postfix, Cyrus IMAP, vsftp, denyhosts, ..

The occurrence of problem is really unpredictable. What i m guilty of is that it could be due to my poor configuration over those daemons.. Sometime during ftp, i get this problem.. Right now i was thinking to turn off mod_security which i suspect the most and according to apache error log.

I dont mind to pm this VE login details. Please let me know if you might need sudoer. So that i will create temp account and add in to wheel group for temporary..

thanks....

Subject: Re: optimization on privvmpages failcnt
Posted by dev on Fri, 10 Nov 2006 16:12:56 GMT
View Forum Message <> Reply to Message

When it happens again, please do the following:
in VE0 (host system) run:
top
then press 'M'
After that 'top' will show you most memory consuming proccesses first. The column 'RSS' shows you how much memory proccess has.
Please post it here and try to kill the offending processes. It should help.

BTW, how much RAM do you have on the node?

Subject: Re: optimization on privvmpages failcnt
Posted by victorskl on Mon, 13 Nov 2006 01:58:48 GMT
View Forum Message <> Reply to Message

Thanks again.

I will debug and post it asap. For now, i turn off XSS setting in mod_security and this VE seem quiet.

HN have Ram 4G and Swap 8G.. but by following vzsplit, it end up with

vmguarpages 0 0 620896 2147483647

That's another question i'd like to ask also. It seem from this point onwards, if i want to increase guarantee ram i have to manually increase pages? bec vzsplit end at 620896 pages which splitting around 16 VEs. If i split 10 VEs only, vmguarpages won't increase anymore than 620896 pages but other parameters such as cpu power increase accordingly.

Subject: Re: optimization on privvmpages failcnt
Posted by dev on Mon, 13 Nov 2006 09:48:59 GMT
View Forum Message <> Reply to Message

ok. it could be a memory leak in mod_security... Report it if it happens again.

how did you run vzsplit (i.e. command line options)?

Subject: Re: optimization on privvmpages failcnt
Posted by victorskl on Mon, 13 Nov 2006 15:31:25 GMT
View Forum Message <> Reply to Message

Sure.. I will follow up with Apache and mod_security community.

For vzsplit, yes i'm using command line vzsplit to perform the initial equal-shared VE splitting. After that i tune(increase/decrease) according to HN's capability/handling and how much VEs this HN can offer. So that i can get some initial point for packaging respect to quota and traffic accounting.

Subject: Re: optimization on privvmpages failcnt
Posted by dev on Mon, 13 Nov 2006 16:02:40 GMT

Sorry, looks like we don't get each other
can you rephrase your question about vzsplit please?

Subject: Re: optimization on privvmpages failcnt
Posted by victorskl on Mon, 13 Nov 2006 17:06:46 GMT

Yap.. sorry might be off topic..

- Yes, i split VE using command line vzsplit. So, I split 20, 40, 80 etc and i have those config files.

- The resource parameters are changing(decrease/increase) according to the amount of VE that i splitting..

- But start from splitting around 17 onward lower value, i.e.. 16, 15, 14,..., 10

- vmguarpages parameters wont increase from barrier 620736 pages. It still showing barrier 620736 pages on 17, and 16, ..., 10 config files.

- But cpuunits/other parameters increase/decrease accordingly on this situation.

Please have a look these splitted config files.

```
# Configuration file generated by vzsplit for 40 VPS
# on HN with total amount of physical mem 4042 Mb
# low memory 3274 Mb, swap size 8063 Mb, Max treads 8000
# Resourse commit level 0:
# Free resource distribution. Any parameters may be increased
# Primary parameters
NUMPROC="418:418"
AVNUMPROC="209:209"
NUMTCPSOCK="418:418"
NUMOTHERSOCK="418:418"
VMGUARPAGES="68777:2147483647"

# Secondary parameters
KMEMSIZE="17166725:18883397"
TCPSNDBUF="4010113:5722241"
TCPRCVBUF="4010113:5722241"
OTHERSOCKBUF="2005056:3717184"
DGRAMRCVBUF="2005056:2005056"
```

OOMGUARPAGES="68777:2147483647"
PRIVVMPAGES="412662:453928"

# Auxiliary parameters
LOCKEDPAGES="838:838"
SHMPAGES="41266:41266"
PHYSPAGES="0:2147483647"
NUMFILE="6688:6688"
NUMFLOCK="668:734"
NUMPTY="41:41"
NUMSIGINFO="1024:1024"
DCACHESIZE="3740085:3852288"
NUMIPTENT="50:50"
DISKSPACE="1666190:1832810"
DISKINODES="860512:946564"
CPUUNITS="7808"

# Configuration file generated by vzsplit for 20 VEs
# on HN with total amount of physical mem 4041 Mb
# low memory 3273 Mb, swap size 8063 Mb, Max treads 8000
# Resourse commit level 0:
# Free resource distribution. Any parameters may be increased
# Primary parameters
NUMPROC="836:836"
AVNUMPROC="418:418"
NUMTCPSOCK="836:836"
NUMOTHERSOCK="836:836"
VMGUARPAGES="620736:2147483647"

# Secondary parameters
KMEMSIZE="34322554:37754809"
TCPSNDBUF="8016595:11440851"
TCPRCVBUF="8016595:11440851"
OTHERSOCKBUF="4008297:7432553"
DGRAMRCVBUF="4008297:4008297"
OOMGUARPAGES="620736:2147483647"
PRIVVMPAGES="620736:682809"

# Auxiliary parameters
LOCKEDPAGES="1675:1675"
SHMPAGES="62073:62073"
PHYSPAGES="0:2147483647"
NUMFILE="13376:13376"
NUMFLOCK="1000:1100"
NUMPTY="83:83"
NUMSIGINFO="1024:1024"

DCACHESIZE="7480170:7704576"
NUMIPTENT="100:100"
DISKSPACE="3332381:3665620"
DISKINODES="1721025:1893128"
CPUUNITS="15242"

# Configuration file generated by vzsplit for 10 VEs
# on HN with total amount of physical mem 4041 Mb
# low memory 3273 Mb, swap size 8063 Mb, Max treads 8000
# Resourse commit level 0:
# Free resource distribution. Any parameters may be increased
# Primary parameters
NUMPROC="1674:1674"
AVNUMPROC="837:837"
NUMTCPSOCK="1674:1674"
NUMOTHERSOCK="1674:1674"
VMGUARPAGES="620736:2147483647"

# Secondary parameters
KMEMSIZE="68645109:75509619"
TCPSNDBUF="16024999:22881703"
TCPRCVBUF="16024999:22881703"
OTHERSOCKBUF="8012499:14869203"
DGRAMRCVBUF="8012499:8012499"
OOMGUARPAGES="620736:2147483647"
PRIVVMPAGES="620736:682809"

# Auxiliary parameters
LOCKEDPAGES="3351:3351"
SHMPAGES="62073:62073"
PHYSPAGES="0:2147483647"
NUMFILE="26784:26784"
NUMFLOCK="1000:1100"
NUMPTY="167:167"
NUMSIGINFO="1024:1024"
DCACHESIZE="14978236:15427584"
NUMIPTENT="200:200"
DISKSPACE="6664763:7331240"
DISKINODES="3442051:3786257"
CPUUNITS="29100"

So compare the last two config, 20 and 10, all other parameters increase/decrease accordingly but VMGUARPAGES, PRIVVMPAGES, OOMGUARPAGES barrier parameters are not. So, my question is
- i have to manually set the amount OR

- is this normal and i'm missing something OR
- vzsplit can not split more than 2G+ of RAM OR
- the rest of RAM(around 1.5G+) reserved for HN?

But of course, i wont offer or package more than 2G+ for VPS hosting.. I just curious to know at least.

Thanks

Subject: Re: optimization on privvmpages failcnt
Posted by dev on Mon, 13 Nov 2006 17:41:17 GMT
View Forum Message <> Reply to Message

AFAICS, it it as follows:

vmguarpages = 620736 = 620736*4 Mb = 2.482 GB
so, for 10 VEs it have allocated 24.82GBs which is overcommitment equal to x2 (you have 4+8 = 12GB total). AFAICS, vzsplit tries to keep overcommitment bound.

also you can always check your configs with vzcfgvalidate after any manual changes.

Subject: Re: optimization on privvmpages failcnt
Posted by victorskl on Mon, 13 Nov 2006 18:57:46 GMT
View Forum Message <> Reply to Message

Wow..
Thanks.. This is clear all my doubts and lack of leveling.. Now i know how to ride the horse!!