Subject: Weird memory numbers
Posted by dagr on Thu, 13 Sep 2007 08:03:55 GMT
View Forum Message <> Reply to Message

I made expirement to understand whether i can see how much memory exactly my vps consumes

I installed apache and run it with 100 start processes (prefork mpm model)

[[email]root@vps-10811[/email] conf]# ps -eo "sz,size,rss,vsize,pid,cmd" | grep httpd| wc -l
102

[[email]root@vps-10811[/email] conf]# ps -eo "sz,size,rss,vsize,pid,cmd" | grep httpd
 1451 1244 2520 5804 30193 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2136 5936 32109 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2136 5936 32110 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2136 5936 32111 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32112 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32113 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32114 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32115 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32116 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32117 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2144 5936 32118 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32119 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32120 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32121 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32122 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32123 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32124 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32125 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32126 /vps_core/apache-2.2.3/bin/httpd -k start
 1484 1376 2140 5936 32127 /vps_core/apache-2.2.3/bin/httpd -k start
.....
 1451 1244 1720 5804 32217 /vps_core/apache-2.2.3/bin/httpd -k start
 1451 1244 1720 5804 32219 /vps_core/apache-2.2.3/bin/httpd -k start
 395 160 520 1580 5144 grep httpd

ps shows memory usage in tricky way , so i checked it by pmap

[[email]root@vps-10811[/email] conf]# pmap -d 32217
.....

mapped: 5808K   writeable/private: 1428K   shared: 48K

writeable/private: 1428K - is the minimum memory which process uses and doesnt share.
Its close to size 1451K from ps in our case . Ie 100x1458 = 150 Mb

Now lets see what beancounters give us

[[email]root@vps-10811[/email] conf]# cat /proc/user_beancounters
Version: 2.5

uid	resource	held	maxheld	barrier	limit	failcnt
10811:	kmemsize	11638464	13987632	20752512	29036012	0
	lockedpages	0	0	32	32	0
	privvmpages	24687	27632	600000	700000	0
	shmpages	652	661	249152	249152	0
	dummy	0	0	0	0	0
	numproc	110	119	200	200	0
	physpages	7099	7854	0	2147483647	0
	vmguarpages	0	0	60000	2147483647	0
	oomguarpages	7099	7854	60000	2147483647	0
	numtcpsock	6	174	300	300	0
	numflock	3	4	100	110	0
	numpty	1	1	16	16	0
	numsiginfo	0	101	256	256	0
	tcpsndbuf	51428	206412	1319488	2524288	0
	tcprcvbuf	84888	942160	1319488	2524288	0
	othersockbuf	2236	9800	5036896	5536896	0
	dgramrcvbuf	0	13980	232096	500000	0
	numothersock	4	9	100	100	0
	dcachesize	0	0	1800000	2000000	0
	numfile	1087	1398	4048	4048	0
	dummy	0	0	0	0	0

Both privvmpages (~100Mb) and oomguarpages(~28Mb) show me numbers far from expected.
Im pretty sure about pmap method , so whats the trick ?

(2.6.9-023stab044.4-enterprise , vzctl-3.0.13-1)

Subject: Re: Weird memory numbers
Posted by dev on Thu, 13 Sep 2007 08:17:55 GMT

View Forum Message <> Reply to Message

taking a look at pmap: AFAICS it scans /proc/<pid>/maps file, which is not enough in reallity to tell how much RAM is used.

Look, read-only mappings are consuming memory as well.
glibc, executable etc. all these goes to UBC::oomguarpages and in pmap::Mapped. So taking into account only "Writeable/Private" is wrong.

Subject: Re: Weird memory numbers
Posted by dagr on Thu, 13 Sep 2007 08:21:48 GMT
View Forum Message <> Reply to Message

I mot taking ONLY. I said that at least it must be private/writable size. it means at least 150Mb . Not ?

Subject: Re: Weird memory numbers
Posted by dev on Thu, 13 Sep 2007 08:51:02 GMT
View Forum Message <> Reply to Message

1. I don't remember whether mpm model implies all threads or all processes or mixed. It depends very much on this.
so please add to your ps output pid and tgid plz.

2. plus plz add full output of pmap -d, since I can't see what regions are used.

3. oomguarpages are much lower then "writeable/private"*100, since many of these writeable areas are not fully touched/used and all the writeable areas from libraries can be shared between the processes until modified. So *100 is a very high estimation. oomguarpages shows real page usages.

4. So the only thing to understand left: why privvmpages << "writeable/private" * 100, correct?

Subject: Re: Weird memory numbers
Posted by dagr on Thu, 13 Sep 2007 09:21:57 GMT
View Forum Message <> Reply to Message

4. So the only thing to understand left: why privvmpages << "writeable/private" * 100, correct?

Yes

1. MPM prefork implies simple forking just like in Apache 1.3 - no threads at all.
tgid - you mean gid ?

[root@vps-10811 extra]# ps -eo "sz,size,rss,vsize,pid,gid,ppid,cmd" | grep httpd

```
 1452 1248 2488 5808 25791   0     1 /vps_core/apache-2.2.3/bin/httpd -k start
 1485 1380 1704 5940 30233   2 25791 /vps_core/apache-2.2.3/bin/httpd -k start
 1485 1380 1700 5940 30234   2 25791 /vps_core/apache-2.2.3/bin/httpd -k start
```

group 2 - daemon

2.

[root@vps-10811 extra]# pmap -d 30233

```
30233:   /vps_core/apache-2.2.3/bin/httpd -k start
Address   Kbytes Mode  Offset           Device    Mapping
00111000      56 r-x-- 0000000000000000 007:0000f libpthread-2.3.4.so
0011f000       4 r---- 000000000000d000 007:0000f libpthread-2.3.4.so
00120000       4 rw--- 000000000000e000 007:0000f libpthread-2.3.4.so
00121000       8 rw--- 0000000000121000 000:00000   [ anon ]
00123000      76 r-x-- 0000000000000000 007:0000f libgssapi_krb5.so.2.2
00136000       4 rw--- 0000000000013000 007:0000f libgssapi_krb5.so.2.2
00137000      60 r-x-- 0000000000000000 007:0000f libresolv-2.3.4.so
00146000       4 r---- 000000000000f000 007:0000f libresolv-2.3.4.so
00147000       4 rw--- 0000000000010000 007:0000f libresolv-2.3.4.so
00148000       8 rw--- 0000000000148000 000:00000   [ anon ]
001fc000      36 r-x-- 0000000000000000 007:0000f libnss_files-2.3.4.so
00205000       4 r---- 0000000000008000 007:0000f libnss_files-2.3.4.so
00206000       4 rw--- 0000000000009000 007:0000f libnss_files-2.3.4.so
00229000       8 r-x-- 0000000000000000 007:0000f libuuid.so.1.2
0022b000       4 rw--- 0000000000001000 007:0000f libuuid.so.1.2
002c5000      60 r-x-- 0000000000000000 007:0000f libz.so.1.2.1.2
002d4000       4 rw--- 000000000000e000 007:0000f libz.so.1.2.1.2
002f5000      76 r-x-- 0000000000000000 0fd:00000 libaprutil-1.so.0.2.7
00308000       4 rw--- 0000000000012000 0fd:00000 libaprutil-1.so.0.2.7
00319000      84 r-x-- 0000000000000000 007:0000f ld-2.3.4.so
0032e000       4 r---- 0000000000015000 007:0000f ld-2.3.4.so
0032f000       4 rw--- 0000000000016000 007:0000f ld-2.3.4.so
00349000     100 r-x-- 0000000000000000 0fd:00000 libexpat.so.0.1.0
00362000       8 rw--- 0000000000019000 0fd:00000 libexpat.so.0.1.0
00385000     124 r-x-- 0000000000000000 0fd:00000 libapr-1.so.0.2.7
003a4000       4 rw--- 000000000001f000 0fd:00000 libapr-1.so.0.2.7
003a9000     132 r-x-- 0000000000000000 007:0000f libm-2.3.4.so
003ca000       4 r---- 0000000000020000 007:0000f libm-2.3.4.so
003cb000       4 rw--- 0000000000021000 007:0000f libm-2.3.4.so
003e9000      32 r-x-- 0000000000000000 007:0000f librt-2.3.4.so
```

```
003f1000       4 r---- 0000000000007000 007:0000f librt-2.3.4.so
003f2000       4 rw--- 0000000000008000 007:0000f librt-2.3.4.so
003f3000      40 rw--- 00000000003f3000 000:00000   [ anon ]
00475000     196 r-x-- 0000000000000000 007:0000f libssl.so.0.9.7a
004a6000      12 rw--- 0000000000031000 007:0000f libssl.so.0.9.7a
004a9000    1172 r-x-- 0000000000000000 007:0000f libc-2.3.4.so
005ce000       8 r---- 0000000000124000 007:0000f libc-2.3.4.so
005d0000       8 rw--- 0000000000126000 007:0000f libc-2.3.4.so
005d2000       8 rw--- 00000000005d2000 000:00000   [ anon ]
0097f000     128 r-x-- 0000000000000000 007:0000f libk5crypto.so.3.0
0099f000       4 rw--- 0000000000020000 007:0000f libk5crypto.so.3.0
009a3000     396 r-x-- 0000000000000000 007:0000f libkrb5.so.3.2
00a06000       8 rw--- 0000000000063000 007:0000f libkrb5.so.3.2
00b87000       8 r-x-- 0000000000000000 007:0000f libcom_err.so.2.1
00b89000       4 rw--- 0000000000001000 007:0000f libcom_err.so.2.1
00bf4000       8 r-x-- 0000000000000000 007:0000f libdl-2.3.4.so
00bf6000       4 r---- 0000000000001000 007:0000f libdl-2.3.4.so
00bf7000       4 rw--- 0000000000002000 007:0000f libdl-2.3.4.so
00cf1000     848 r-x-- 0000000000000000 007:0000f libcrypto.so.0.9.7a
00dc5000      72 rw--- 00000000000d4000 007:0000f libcrypto.so.0.9.7a
00dd7000      12 rw--- 0000000000dd7000 000:00000   [ anon ]
00f15000      20 r-x-- 0000000000000000 007:0000f libcrypt-2.3.4.so
00f1a000       4 r---- 0000000000004000 007:0000f libcrypt-2.3.4.so
00f1b000       4 rw--- 0000000000005000 007:0000f libcrypt-2.3.4.so
00f1c000     156 rw--- 0000000000f1c000 000:00000   [ anon ]
08048000     668 r-x-- 0000000000000000 0fd:00000 httpd
080ef000      16 rw--- 00000000000a6000 0fd:00000 httpd
080f3000      20 rw--- 0000000080f3000 000:00000   [ anon ]
09520000    1080 rw--- 0000000009520000 000:00000   [ anon ]
b7fc2000      48 rw-s- 0000000000000000 000:00033   [ anon ]
b7fce000      20 rw--- 00000000b7fce000 000:00000   [ anon ]
bfff7000      28 rw--- 00000000bfff7000 000:00000   [ stack ]
ffffe000       4 ----- 0000000000000000 000:00000   [ anon ]
mapped: 5944K    writeable/private: 1564K    shared: 48K
```

3. So you say that real consumption is oomguarpages. May be oomguarpages << privmpages when processes idle or processing same document - in other words, in docs you call privmpages some kind of memusage prediction. So in what case apache will consume privmpages which i see now?

Subject: Re: Weird memory numbers
Posted by dev on Thu, 13 Sep 2007 09:32:03 GMT

View Forum Message <> Reply to Message

1. let's honestly sum up all pmap info from all the processes. multiplication by 100 can be not honest.

2.
> So you say that real consumption is oomguarpages. May be
> oomguarpages << privmpages when processes idle or processing same
> document - in other words, in docs you call privmpages some kind
> of memusage prediction. So in what case apache will consume
> privmpages which i see now?

for any normal application privvmapes is around 20-50% of oomguarpages. it is normal, since apps never touch all the memory they map.
however, in special case or in maliscuios apps all the memory counted to privvmpages can be really allocated and thus oomguarpages can be equal to privvmpages.

yes, oomguarpages is current real usage of physical pages + swap pages.

privvmpages is a high estimation of what can be allocated by application w/o denials.

Subject: Re: Weird memory numbers
Posted by dagr on Thu, 13 Sep 2007 11:07:34 GMT
View Forum Message <> Reply to Message

1.

```
[root@vps-10811 extra]# s=0;n=0;for pid in ` ps -ef | grep httpd  | grep -v "gre
p"| awk '{print $2}'`;do
>  a=`pmap -d $pid | grep private | awk '{print $4}'`;
>  l=`expr length $a - 1`;
>  b=`expr substr $a 1 $l`
>  s=`expr $s + $b`
>  n=`expr $n + 1 `
> done
[root@vps-10811 extra]# echo "s=$s"
s=157832
[root@vps-10811 extra]# echo "num=$n"
num=101
```

2. I cant understand this.
"it is normal, since apps never touch all the memory they map"
But OS cant know for sure that app wont touch it later , so it must ne isolated ie - busy, and other app cant use it same time . In this terms i dont understand sense of 2 parameters like oom and priv, should be only 1. Example above is simple. I need to estimate how much processes my apache can run and nothing more, ie i should take overall memory and divide by lets say 1.5 Mb.

Subject: Re: Weird memory numbers
Posted by dev on Thu, 13 Sep 2007 11:24:15 GMT
View Forum Message <> Reply to Message

1. hmm... strange
I'll check

2. No. OS doesn't know in reallity how much RAM you will consume.
Look, you can mmap 1GB private mappings on machine with 100Mb of RAM. You will be allowed to do so on std kernel (without disabled overcimmitment). i.e. this 1GB will be charged to privvmpages.
The real memory pages are allocated later when application really wants to touch the memory. This is what charged to oomguarpages.

Applications very often map files privately with write access, so potentially they can have a full private copy of file in memory. But in reallity they touch only small pieces of it, so little memory is required. And that's why privvmpages >> oomguarpages.

Subject: Re: Weird memory numbers
Posted by xemul on Wed, 19 Sep 2007 10:03:26 GMT
View Forum Message <> Reply to Message

Please, check with "ps -o pid,tid,comm" that all the http processes are not threads (I didn't notice whether you did this).

If they are threads, then the answer is: pmap can't handle this at all.

Otherwise, could you please create a tarball with this VE and let us download it?

Thanks,
Pavel

Subject: Re: Weird memory numbers
Posted by dagr on Wed, 19 Sep 2007 12:39:51 GMT
View Forum Message <> Reply to Message

Its not threads

```
[root@vps-10811 conf]# ps -eo "pid,tid,comm"
  PID   TID COMMAND
    1     1 init
31992 31992 syslogd
32002 32002 sshd
32013 32013 httpd
```

32028 32028 crond
32035 32035 httpd
32036 32036 httpd
32037 32037 httpd
32038 32038 httpd
32039 32039 httpd
32040 32040 httpd
32041 32041 httpd
32042 32042 httpd
32043 32043 httpd
32044 32044 httpd
32045 32045 httpd
32046 32046 httpd
32047 32047 httpd
32048 32048 httpd
32049 32049 httpd
32050 32050 httpd
32051 32051 httpd
32052 32052 httpd
.....

[root@vps-10811 conf]# httpd -l | grep prefork
  prefork.c

I checked situation on another HN - its the same

VE can be downloaded here

http://209.3.12.11/d/10811.img.tgz

(confs,private dir and loopback_ro_image)

(apache starts automatically after VE start)

Subject: Re: Weird memory numbers
Posted by xemul on Fri, 21 Sep 2007 11:43:49 GMT
View Forum Message <> Reply to Message

Well, I've got it

The answer is - this is normal. Point is that httpd uses private readonly mappings. Reading from

these regions results in so called ZERO_PAGE appears in the address space. ZERO_PAGE is a reserved 4k chunk of memory which is used only for this or similar purposes. We do not account for this page, since it doesn't increase the RSS at all, but its presence is "accounted" with the pmap utility.

Subject: Re: Weird memory numbers
Posted by dagr on Sat, 22 Sep 2007 15:54:56 GMT
View Forum Message <> Reply to Message

Cool , and the last question .

When do i really need to take into account privvmpages ?
So far it looks more like inner technical parameter.

Subject: Re: Weird memory numbers
Posted by dev on Mon, 24 Sep 2007 07:03:10 GMT
View Forum Message <> Reply to Message

you need to take it into account when limiting your VE.
plz note that physpages limit doesn't affect VE.
and oomguarpages is a gurantee, not a limit.
So privvmpages is the only handle to control VE memory.

P.S. In future we'll add physpages limiting leading to VE swap-out, but this is not efficient way of control memory, since swap bandwidth is scarce resource.