
Subject: OOM didn't save the machine

Posted by [lazy](#) on Sun, 29 Mar 2009 17:33:53 GMT

[View Forum Message](#) <> [Reply to Message](#)

64 bit rhel5 60.2, 2 32 bit VE's 8GB of ram without swap

what is filp cache ?

Am i right that it was 4,9GB big ?

Mar 29 15:49:11 hd1 kernel: filp : size 4979023872 objsize 256

it looked like some memory leak (for 6 hours memory usage was raising) 2 VE's have total 6GB
privvmpages + 4,9GB for flip cache makes it possible.

Mar 29 15:49:11 hd1 kernel: apache2 invoked oom-killer: gfp_mask=0x201d2, order=0, oomkilladj=0

Mar 29 15:49:11 hd1 kernel:

Mar 29 15:49:11 hd1 kernel: Call Trace:

Mar 29 15:49:11 hd1 kernel: [<ffffffff800b5373>] out_of_memory+0x9a/0x16a

Mar 29 15:49:11 hd1 kernel: [<ffffffff8000f00d>] __alloc_pages+0x236/0x323

Mar 29 15:49:11 hd1 kernel: [<ffffffff80012535>] __do_page_cache_readahead+0xcd/0x24d

Mar 29 15:49:11 hd1 kernel: [<ffffffff8003086a>] release_sock+0x13/0xaa

Mar 29 15:49:11 hd1 kernel: [<ffffffff8001ce4d>] tcp_recvmmsg+0x803/0x916

Mar 29 15:49:11 hd1 kernel: [<ffffffff80012fac>] filemap_nopage+0x147/0x313

Mar 29 15:49:11 hd1 kernel: [<ffffffff80007ff2>] __handle_mm_fault+0x26d/0xe76

Mar 29 15:49:11 hd1 kernel: [<ffffffff8000c801>] do_sync_read+0xc7/0x104

Mar 29 15:49:11 hd1 kernel: [<ffffffff8000a849>] do_page_fault+0x3aa/0x6e3

Mar 29 15:49:11 hd1 kernel: [<ffffffff8003086a>] release_sock+0x13/0xaa

Mar 29 15:49:11 hd1 kernel: [<ffffffff8000b1b0>] vfs_read+0x11b/0x150

Mar 29 15:49:11 hd1 kernel: [<ffffffff8005fe39>] error_exit+0x0/0x84

Mar 29 15:49:11 hd1 kernel:

Mar 29 15:49:11 hd1 kernel: Mem-info:

Mar 29 15:49:11 hd1 kernel: DMA per-cpu:

Mar 29 15:49:11 hd1 kernel: cpu 0 hot: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: cpu 0 cold: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: cpu 1 hot: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: cpu 1 cold: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: cpu 2 hot: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: cpu 2 cold: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: cpu 3 hot: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: cpu 3 cold: high 0, batch 1 used:0

Mar 29 15:49:11 hd1 kernel: DMA32 per-cpu:

Mar 29 15:49:11 hd1 kernel: cpu 0 hot: high 186, batch 31 used:182

```

Mar 29 15:49:11 hd1 kernel: cpu 0 cold: high 62, batch 15 used:56
Mar 29 15:49:11 hd1 kernel: cpu 1 hot: high 186, batch 31 used:158
Mar 29 15:49:11 hd1 kernel: cpu 1 cold: high 62, batch 15 used:23
Mar 29 15:49:11 hd1 kernel: cpu 2 hot: high 186, batch 31 used:154
Mar 29 15:49:11 hd1 kernel: cpu 2 cold: high 62, batch 15 used:2
Mar 29 15:49:11 hd1 kernel: cpu 3 hot: high 186, batch 31 used:178
Mar 29 15:49:11 hd1 kernel: cpu 1 hot: high 186, batch 31 used:183
Mar 29 15:49:11 hd1 kernel: cpu 1 cold: high 62, batch 15 used:14
Mar 29 15:49:11 hd1 kernel: cpu 2 hot: high 186, batch 31 used:172
Mar 29 15:49:11 hd1 kernel: cpu 2 cold: high 62, batch 15 used:5
Mar 29 15:49:11 hd1 kernel: cpu 3 hot: high 186, batch 31 used:172
Mar 29 15:49:11 hd1 kernel: cpu 3 cold: high 62, batch 15 used:14
Mar 29 15:49:11 hd1 kernel: HighMem per-cpu: empty
Mar 29 15:49:11 hd1 kernel: Free pages:      42888kB (0kB HighMem)
Mar 29 15:49:11 hd1 kernel: Active:636733 inactive:1008 dirty:0 writeback:59 unstable:0
free:10722 slab:1341299 mapped-file:27 mapped-anon:637545 pagetables
:8471
Mar 29 15:49:11 hd1 kernel: DMA free:12304kB min:16kB low:20kB high:24kB active:0kB
inactive:0kB present:11868kB pages_scanned:0 all_unreclaimable? yes
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 3246 8034 8034
Mar 29 15:49:11 hd1 kernel: DMA32 free:23768kB min:4632kB low:5788kB high:6948kB
active:1100012kB inactive:336kB present:3324872kB pages_scanned:2028965 all
_unreclaimable? yes
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 0 4788 4788
Mar 29 15:49:11 hd1 kernel: Normal free:6816kB min:6828kB low:8532kB high:10240kB
active:1446920kB inactive:3696kB present:4902912kB pages_scanned:6614355 a
ll_unreclaimable? yes
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 0 0 0
Mar 29 15:49:11 hd1 kernel: HighMem free:0kB min:128kB low:128kB high:128kB active:0kB
inactive:0kB present:0kB pages_scanned:0 all_unreclaimable? no
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 0 0 0
Mar 29 15:49:11 hd1 kernel: DMA: 2*4kB 5*8kB 4*16kB 3*32kB 3*64kB 3*128kB 1*256kB
0*512kB 1*1024kB 1*2048kB 2*4096kB = 12304kB
Mar 29 15:49:11 hd1 kernel: DMA32: 8*4kB 13*8kB 5*16kB 0*32kB 0*64kB 0*128kB 0*256kB
0*512kB 1*1024kB 1*2048kB 5*4096kB = 23768kB
Mar 29 15:49:11 hd1 kernel: Normal: 120*4kB 4*8kB 0*16kB 1*32kB 32*64kB 3*128kB 1*256kB
1*512kB 1*1024kB 1*2048kB 0*4096kB = 6816kB
Mar 29 15:49:11 hd1 kernel: HighMem: empty
Mar 29 15:49:11 hd1 kernel: 1208 pagecache pages
Mar 29 15:49:11 hd1 kernel: Swap cache: add 0, delete 0, find 0/0, race 0+0+0
Mar 29 15:49:11 hd1 kernel: Free swap = 0kB
Mar 29 15:49:11 hd1 kernel: Total swap = 0kB
Mar 29 15:49:11 hd1 kernel: Free swap:      0kB
Mar 29 15:49:11 hd1 kernel: 2293760 pages of RAM
Mar 29 15:49:11 hd1 kernel: 253465 reserved pages
Mar 29 15:49:11 hd1 kernel: 173691 pages shared
Mar 29 15:49:11 hd1 kernel: 0 pages swap cached
Mar 29 15:49:11 hd1 kernel: Top 10 caches:

```

```

Mar 29 15:49:11 hd1 kernel: ip_contrack      : size  4235264 objsize    312
Mar 29 15:49:11 hd1 kernel: dentry_cache    : size 335552512 objsize    240
Mar 29 15:49:11 hd1 kernel: size-2048(UBC)   : size  7345536 objsize   2048
Mar 29 15:49:11 hd1 kernel: files_cache     : size 10207232 objsize    768
Mar 29 15:49:11 hd1 kernel: size-128       : size  8962048 objsize   128
Mar 29 15:49:11 hd1 kernel: ext3_inode_cache : size 18158976 objsize    808
Mar 29 15:49:11 hd1 kernel: vm_area_struct  : size  7004160 objsize   168
Mar 29 15:49:11 hd1 kernel: filp           : size 4979023872 objsize   256
Mar 29 15:49:11 hd1 kernel: task_struct     : size  43917312 objsize  2160
Mar 29 15:49:11 hd1 kernel: page_beancounter : size  53186560 objsize    64
Mar 29 15:49:11 hd1 kernel: Mem-info:
Mar 29 15:49:11 hd1 kernel: DMA per-cpu:
Mar 29 15:49:11 hd1 kernel: cpu 0 hot: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: cpu 0 cold: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: cpu 1 hot: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: cpu 1 cold: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: cpu 2 hot: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: cpu 2 cold: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: cpu 3 hot: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: cpu 3 cold: high 0, batch 1 used:0
Mar 29 15:49:11 hd1 kernel: DMA32 per-cpu:
Mar 29 15:49:11 hd1 kernel: cpu 0 hot: high 186, batch 31 used:182
Mar 29 15:49:11 hd1 kernel: cpu 0 cold: high 62, batch 15 used:56
Mar 29 15:49:11 hd1 kernel: cpu 1 hot: high 186, batch 31 used:158
Mar 29 15:49:11 hd1 kernel: cpu 1 cold: high 62, batch 15 used:23
Mar 29 15:49:11 hd1 kernel: cpu 2 hot: high 186, batch 31 used:154
Mar 29 15:49:11 hd1 kernel: cpu 2 cold: high 62, batch 15 used:2
Mar 29 15:49:11 hd1 kernel: cpu 3 hot: high 186, batch 31 used:178
Mar 29 15:49:11 hd1 kernel: cpu 3 cold: high 62, batch 15 used:49
Mar 29 15:49:11 hd1 kernel: Normal per-cpu:
Mar 29 15:49:11 hd1 kernel: cpu 0 hot: high 186, batch 31 used:159
Mar 29 15:49:11 hd1 kernel: cpu 0 cold: high 62, batch 15 used:24
Mar 29 15:49:11 hd1 kernel: cpu 1 hot: high 186, batch 31 used:183
Mar 29 15:49:11 hd1 kernel: Normal per-cpu:
Mar 29 15:49:11 hd1 kernel: cpu 0 hot: high 186, batch 31 used:159
Mar 29 15:49:11 hd1 kernel: cpu 0 cold: high 62, batch 15 used:24
Mar 29 15:49:11 hd1 kernel: cpu 1 hot: high 186, batch 31 used:183
Mar 29 15:49:11 hd1 kernel: cpu 1 cold: high 62, batch 15 used:14
Mar 29 15:49:11 hd1 kernel: cpu 2 hot: high 186, batch 31 used:172
Mar 29 15:49:11 hd1 kernel: cpu 2 cold: high 62, batch 15 used:59
Mar 29 15:49:11 hd1 kernel: cpu 3 hot: high 186, batch 31 used:172
Mar 29 15:49:11 hd1 kernel: cpu 3 cold: high 62, batch 15 used:14
Mar 29 15:49:11 hd1 kernel: HighMem per-cpu: empty
Mar 29 15:49:11 hd1 kernel: Free pages:      43852kB (0kB HighMem)
Mar 29 15:49:11 hd1 kernel: Active:637098 inactive:750 dirty:0 writeback:59 unstable:0
free:10963 slab:1341299 mapped-file:27 mapped-anon:637545 pagetables:
8471
Mar 29 15:49:11 hd1 kernel: DMA free:12304kB min:16kB low:20kB high:24kB active:0kB

```

inactive:0kB present:11868kB pages_scanned:0 all_unreclaimable? yes
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 3246 8034 8034
Mar 29 15:49:11 hd1 kernel: DMA32 free:23768kB min:4632kB low:5788kB high:6948kB
active:1101240kB inactive:496kB present:3324872kB pages_scanned:2590683 all
_unreclaimable? yes
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 0 4788 4788
Mar 29 15:49:11 hd1 kernel: Normal free:7780kB min:6828kB low:8532kB high:10240kB
active:1447152kB inactive:2504kB present:4902912kB pages_scanned:15392 all
_unreclaimable? no
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 0 0 0
Mar 29 15:49:11 hd1 kernel: HighMem free:0kB min:128kB low:128kB high:128kB active:0kB
inactive:0kB present:0kB pages_scanned:0 all_unreclaimable? no
Mar 29 15:49:11 hd1 kernel: lowmem_reserve[]: 0 0 0 0
Mar 29 15:49:11 hd1 kernel: DMA: 2*4kB 5*8kB 4*16kB 3*32kB 3*64kB 3*128kB 1*256kB
0*512kB 1*1024kB 1*2048kB 2*4096kB = 12304kB
Mar 29 15:49:11 hd1 kernel: DMA32: 8*4kB 13*8kB 5*16kB 0*32kB 0*64kB 0*128kB 0*256kB
0*512kB 1*1024kB 1*2048kB 5*4096kB = 23768kB
Mar 29 15:49:11 hd1 kernel: Normal: 361*4kB 4*8kB 0*16kB 1*32kB 32*64kB 3*128kB 1*256kB
1*512kB 1*1024kB 1*2048kB 0*4096kB = 7780kB
Mar 29 15:49:11 hd1 kernel: HighMem: empty
Mar 29 15:49:11 hd1 kernel: 952 pagecache pages
Mar 29 15:49:11 hd1 kernel: Swap cache: add 0, delete 0, find 0/0, race 0+0+0
Mar 29 15:49:11 hd1 kernel: Free swap = 0kB
Mar 29 15:49:11 hd1 kernel: Total swap = 0kB

Subject: Re: OOM didn't save the machine
Posted by [maratrus](#) on Mon, 30 Mar 2009 13:42:15 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

Quote:

it looked like some memory leak (for 6 hours memory usage was raising) 2 VE's have total 6GB privvmpages + 4,9GB for filp cache makes it possible.

Whenever you open a file the kernel uses a piece of memory that belongs to filp cache. So, a numerous open files might cause this situation. Are you sure that user_beancounters are configured properly, i.e. they restrict the VEs consumption?
Could you please show /proc/user_beancounters from HN?

Subject: Re: OOM didn't save the machine
Posted by [lazy](#) on Mon, 30 Mar 2009 19:24:58 GMT

Thank's for Your answer.

It's possible the "leak" started after the machine was rsynced (f

Now it's happening again, apache process is using 100% cpu, I can't enter the vps beancounters bellow

3000: kmemsize	68901194	68920466	668435456	836870912	
0					
lockedpages	0	0	2562	2562	0
privvmpages	157683	171574	512000	537600	
0					
shmpages	284	284	15374	15374	0
dummy	0	0	0	0	0
numproc	98	127	2000	2000	0
physpages	50345	59833	0	9223372036854775807	
0					
vmguarpages	0	0	35236	9223372036854775807	
0					
oomguarpages	50345	59833	35236	9223372036854775807	
0					
numtcpsock	15	95	2000	2000	0
numflock	1	13	1000	1100	0
numpty	0	1	200	200	0
numsiginfo	1	30	1024	1024	0
tcpsndbuf	178640	1460304	9300923	17492923	
0					
tcprcvbuf	1296	286512	638976	1048576	
0					
othersockbuf	11600	1270944	4650461	12842461	
0					
dgramrcvbuf	0	4368	4650461	4650461	
0					
numothersock	9	33	8000	9000	0
dcachesize	0	0	11452893	11796480	0
numfile	3101	4801	20480	20480	0
dummy	0	0	0	0	0
dummy	0	0	0	0	0
dummy	0	0	0	0	0
numiptent	10	10	200	200	0
0: kmemsize	49891537	49960180	9223372036854775807		
9223372036854775807	0				
lockedpages	0	0	9223372036854775807		
9223372036854775807	0				
privvmpages	6277	18544	9223372036854775807		
9223372036854775807	0				

shmpages	647	663	9223372036854775807
9223372036854775807	0		
dummy	0	0	9223372036854775807 9223372036854775807
0			
numproc	90	98	9223372036854775807
9223372036854775807	0		
physpages	3690	15435	9223372036854775807
9223372036854775807	0		
vmguarpages	0	0	9223372036854775807
9223372036854775807	0		
oomguarpages	3697	15435	9223372036854775807
9223372036854775807	0		
numtcpsock	5	6	9223372036854775807
9223372036854775807	0		
numflock	1	7	9223372036854775807 9223372036854775807
0			
numpty	4	4	9223372036854775807 9223372036854775807
0			
numsiginfo	1	3	9223372036854775807 9223372036854775807
0			
tcpsndbuf	85216	698000	9223372036854775807
9223372036854775807	0		
tcprecvbuf	81920	1312608	9223372036854775807
9223372036854775807	0		
othersockbuf	9280	24832	9223372036854775807
9223372036854775807	0		
dgramrcvbuf	0	8464	9223372036854775807
9223372036854775807	0		
numothersock	22	27	9223372036854775807
9223372036854775807	0		
dcachesize	0	0	9223372036854775807 9223372036854775807
0			
numfile	1680	1825	9223372036854775807
9223372036854775807	0		
dummy	0	0	9223372036854775807 9223372036854775807
0			
dummy	0	0	9223372036854775807 9223372036854775807
0			
dummy	0	0	9223372036854775807 9223372036854775807
0			
numiptent	10	10	9223372036854775807
9223372036854775807	0		

this apache is heavily modified and it can be stuck in some recvmsg, i cant kill -9, memory is starting to be eaten


```

meminfo from hn
MemTotal:    8161180 kB
MemFree:     55608 kB
Buffers:     114284 kB
Cached:      2716572 kB
SwapCached:    0 kB
Active:      3450172 kB
Inactive:    2065236 kB
HighTotal:    0 kB
HighFree:     0 kB
LowTotal:    8161180 kB
LowFree:     55608 kB
SwapTotal:   1023992 kB
SwapFree:    1023964 kB
Dirty:       9504 kB
Writeback:    188 kB
AnonPages:   2683688 kB
Mapped:      28620 kB
Slab:        2479572 kB
PageTables:   32560 kB
NFS_Unstable: 0 kB
Bounce:      0 kB
CommitLimit: 5104580 kB
Committed_AS: 4361756 kB
VmallocTotal: 34359738364 kB
VmallocUsed:  273424 kB
VmallocChunk: 34359464744 kB

```

vps are eating total 2,6G physp + 2,6G cache and 140M free

any pointers how to kill that vps, i'm thinging about taking away its privvmpages and thus forcing oom

this process is untracable

its wchan is init_level4_pg, and is eating 100% cpu

```

vzctl enter 3000 ends
brk(0xa41e000) = 0xa41e000
rt_sigaction(SIGPIPE, {SIG_IGN}, NULL, 0) = 0
open("/etc/vz/vz.conf", O_RDONLY) = 3
stat("/etc/vz/vz.conf", {st_mode=S_IFREG|0644, st_size=1103, ...}) = 0
fstat(3, {st_mode=S_IFREG|0644, st_size=1103, ...}) = 0
mmap(NULL, 4096, PROT_READ|PROT_WRITE, MAP_PRIVATE|MAP_ANONYMOUS, -1, 0) =
0x2b3071a0b000
read(3, "## Global parameters\nVIRTUOZZO=y"..., 4096) = 1103
read(3, "", 4096) = 0

```

```

close(3) = 0
munmap(0x2b3071a0b000, 4096) = 0
open("/var/log/vzctl.log", O_WRONLY|O_APPEND|O_CREAT, 0666) = 3
fstat(3, {st_mode=S_IFREG|0644, st_size=0, ...}) = 0
mmap(NULL, 4096, PROT_READ|PROT_WRITE, MAP_PRIVATE|MAP_ANONYMOUS, -1, 0) =
0x2b3071a0b000
fstat(3, {st_mode=S_IFREG|0644, st_size=0, ...}) = 0
lseek(3, 0, SEEK_SET) = 0
stat("/etc/vz/conf/3000.conf", {st_mode=S_IFREG|0644, st_size=1194, ...}) = 0
open("/etc/vz/conf/3000.conf", O_RDONLY) = 4
stat("/etc/vz/conf/3000.conf", {st_mode=S_IFREG|0644, st_size=1194, ...}) = 0
fstat(4, {st_mode=S_IFREG|0644, st_size=1194, ...}) = 0
mmap(NULL, 4096, PROT_READ|PROT_WRITE, MAP_PRIVATE|MAP_ANONYMOUS, -1, 0) =
0x2b3071a0c000
read(4, "# Configuration file generated b"... , 4096) = 1194
read(4, "", 4096) = 0
close(4) = 0
munmap(0x2b3071a0c000, 4096) = 0
fcntl(0, F_GETFL) = 0x8002 (flags O_RDWR|O_LARGEFILE)
fcntl(1, F_GETFL) = 0x8002 (flags O_RDWR|O_LARGEFILE)
fcntl(2, F_GETFL) = 0x8002 (flags O_RDWR|O_LARGEFILE)
open("/dev/vzctl", O_RDWR) = 4
ioctl(4, 0x400c2e05, 0x7fff392db2f0) = 0
ioctl(4, 0x400c2e05, 0x7fff392db150) = 0
clone(child_stack=0, flags=CLONE_CHILD_CLEARTID|CLONE_CHILD_SETTID|SIGCHLD,
child_tidptr=0x2b3071e52b70) = 4189
wait4(4189,

root  2283 0.0 0.0 10228 872 pts/2  S+  21:01  0:00 vzctl enter 3000
root  2284 0.0 0.0 10228 376 ?    Ds  21:01  0:00 vzctl enter 3000

```

Subject: Re: OOM didn't save the machine
 Posted by [lazy](#) on Mon, 30 Mar 2009 21:41:06 GMT
[View Forum Message](#) <> [Reply to Message](#)

thanks to sysreq i managed to reboot the machine without a crash, I didnt find anything interesting in proc beside filp beaing eaten in slab and this

CPU 1, VCPU 3000:0

Modules linked in: vzethdev(U) vznetdev(U) simfs(U) vzrst(U) ip_nat(U) vzcpt(U) ip_conntrack(U) nfnetlink(U) ipip(U) tunnel4(U) tun(U) vxdquota(U) vzmon(U) vzdev(U) xt_tcpudp(U) xt_length(U) ipt_ttl(U) xt_tcpmss(U) ipt_TCPMSS(U) iptable_mangle(U) iptable_filter(U) xt_multiport(U) xt_limit(U) ipt_tos(U) ipt_REJE

CT(U) ip_tables(U) x_tables(U) button(U) dm_snapshot(U) dm_mirror(U) dm_mod(U) mptctl(U)
loop(U) sg(U) sr_mod(U) cdrom(U) sd_mod(U) ehci_hcd(U) ata_piix(U)
libata(U) tg3(U) uhci_hcd(U) mptsas(U) mptscsih(U) mptbase(U) scsi_transport_sas(U)
scsi_mod(U)

Pid: 10722, comm: httpd Tainted: P

~~~~~ this is the unkillable 100% cpu eating monster

2.6.18-92.1.18.el5.028stab060.2 #1 028stab060

RIP: 0060:[<ffffff8004a9b7>] [<ffffff8004a9b7>] unix\_stream\_sendmsg+0x194/0x3d7

RSP: 0000:ffff8101f250fb88 EFLAGS: 00000203

RAX: 0000000000000000 RBX: ffff8101f250fee8 RCX: 000000000000003b8

RDX: ffffffffef0 RSI: ffff8101e35b4c0 RDI: ffff81012fbae000

RBP: 0000000000000227 R08: ffff8101c476cb80 R09: 0000000000000286

R10: 000053fe5fafdbc6 R11: ffff8101f250fb38 R12: 0000000000000000

R13: ffff8101f34b50c0 R14: 0000000000000000 R15: ffff8101b17cbc80

FS: 0000000000000000(0000) GS:ffff81022f494b40(0033) knlGS:00000000b7bd06c0

CS: 0060 DS: 007b ES: 007b CR0: 000000008005003b

CR2: 00000000b6d19030 CR3: 00000001f2634000 CR4: 00000000000006e0

Call Trace:

<NMI> <<EOE>> [<ffffff8001df4c>] \_\_pollwait+0x0/0xe1

[<ffffff80055a09>] sock\_sendmsg+0xd4/0xec

[<ffffff8003f92b>] memcpy\_toiovec+0x36/0x66

[<ffffff80064edf>] \_spin\_lock\_bh+0x9/0x14

[<ffffff800960cc>] autoremove\_wake\_function+0x0/0x2e

[<ffffff801de53f>] cmsghdr\_from\_user\_compat\_to\_kern+0x180/0x20b

[<ffffff801ca06a>] sys\_sendmsg+0x217/0x28a

[<ffffff8000c801>] do\_sync\_read+0xc7/0x104

[<ffffff80064edf>] \_spin\_lock\_bh+0x9/0x14

[<ffffff800960cc>] autoremove\_wake\_function+0x0/0x2e

[<ffffff801ddd6c>] compat\_sys\_socketcall+0x159/0x172

[<ffffff800615fa>] ia32\_sysret+0x0/0xa

this was the unkillable process, this apache process recives http requests and sends sockets to other http processes using sendmsg, there can be some error in this msghdr can be wrong or there might be some loop but it shouldn't kill the machine.

FD\_SETSIZE is raised to 2048 if it makes any diference.

I think i might be able to reproduce some of the flow from this trace, to bad I did sysreq p only once

Basicly it should be sthing like poll(), accept a socket, sendmsg the socket to another process threw one of sockets created by socketpair(PF\_UNIX, SOCK\_STREAM, 0, socks)

Same code runs without problems on non openvz kernels 2.6.22 25 27 for months

Any pointers how to debug it ?

---

---

Subject: Re: OOM didn't save the machine  
Posted by [maratrus](#) on Tue, 31 Mar 2009 11:36:36 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello,

Quote:  
It's possible the "leak"

Yes, it might be a leak. Moreover, you said that  
Quote:  
Same code runs without problems on non openvz kernels 2.6.22 25 27 for months

So, let's try to find out if it's indeed a memory leak.

Quote:  
Now it's happening again, apache process is using 100% cpu, I can't enter the vps beancounters bellow

Looks like you have to adjust CPUUNITS/CPULIMITS on the HN. You may read about them in OpenVZ user's guide  
<http://download.openvz.org/doc/OpenVZ-Users-Guide.pdf>

Quote:

oomguarpages 50345 59833 35236 9223372036854775807 0

Oomguarpages exceeded barrier value, so this is likely the reason why a process inside that VE was killed.  
<http://wiki.openvz.org/Oomguarpages#oomguarpages>

Do you have a single VE on the HN? If no, please show the full user\_beancounters output.

Next time this issue will occur try to stop the problem VE (hope CPULIMIT/CPUUNITS adjustment will allow to do that) and look at the user\_beancounters again. There should be no usage when VE is stopped. Please, also look at the slab state. If the consumption value is decreased. And show please /proc/slabinfo output after and before stopping the problem VE.

---

Subject: Re: OOM didn't save the machine  
Posted by [lazy](#) on Tue, 31 Mar 2009 12:34:21 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

maratrus wrote on Tue, 31 March 2009 06:36Hello,

Quote:

It's possible the "leak"

Yes, it might be a leak. Moreover, you said that

Quote:

Same code runs without problems on non openvz kernels 2.6.22 25 27 for months

So, let's try to find out if it's indeed a memory leak.

Quote:

Now it's happening again, apache process is using 100% cpu, I can't enter the vps beancounters bellow

Looks like you have to adjust CPUUNITS/CPULIMITS on the HN. You may read about them in OpenVZ user's guide

<http://download.openvz.org/doc/OpenVZ-Users-Guide.pdf>

this VE has 300% cpulimit, and themachine has 4 cores

Quote:

oomguarpages 50345 59833 35236 9223372036854775807 0

Oomguarpages exceeded barrier value, so this is likely the reason why a process inside that VE was killed.

<http://wiki.openvz.org/Oomguarpages#oomguarpages>

yes, first time when memory run out oom tried to kill some processes but I don't know if the one eating all the cpu was killed

Do you have a single VE on the HN? If no, please show the full user\_beancounters output.

Version: 2.5

| uid   | resource     | held      | maxheld   | barrier   | limit               | failcnt   |
|-------|--------------|-----------|-----------|-----------|---------------------|-----------|
| 3001: | kmemsize     | 164187239 | 168681365 | 268435456 |                     | 536870912 |
| 0     |              |           |           |           |                     |           |
|       | lockedpages  | 0         | 0         | 2562      | 2562                | 0         |
|       | privvmpages  | 750001    | 863788    | 1048576   | 1048576             |           |
| 0     |              |           |           |           |                     |           |
|       | shmpages     | 28        | 48        | 15374     | 15374               | 0         |
|       | dummy        | 0         | 0         | 0         | 0                   | 0         |
|       | numproc      | 202       | 247       | 2000      | 2000                | 0         |
|       | physpages    | 667127    | 774657    | 0         | 9223372036854775807 |           |
| 0     |              |           |           |           |                     |           |
|       | vmguarpages  | 0         | 0         | 35236     | 9223372036854775807 |           |
| 0     |              |           |           |           |                     |           |
|       | oomguarpages | 667127    | 774657    | 35236     | 9223372036854775807 |           |
|       |              |           |           |           |                     |           |

|       |              |          |          |           |                     |   |
|-------|--------------|----------|----------|-----------|---------------------|---|
|       | numtcpsock   | 134      | 309      | 2000      | 2000                | 0 |
|       | numflock     | 10       | 32       | 1000      | 1100                | 0 |
|       | numpty       | 0        | 1        | 200       | 200                 | 0 |
|       | numsiginfo   | 0        | 101      | 1024      | 1024                | 0 |
| 0     | tcpsndbuf    | 1125360  | 4048272  | 8388608   | 10485760            |   |
| 0     | tcprcvbuf    | 1046080  | 2213216  | 8388608   | 10485760            |   |
| 0     | othersockbuf | 171680   | 1371728  | 4650461   | 12842461            |   |
| 0     | dgramrcvbuf  | 0        | 8736     | 4650461   | 4650461             |   |
| 0     | numothersock | 101      | 140      | 2048      | 3000                | 0 |
|       | dcachesize   | 0        | 0        | 11452893  | 11796480            | 0 |
|       | numfile      | 3896     | 4609     | 20480     | 20480               | 0 |
|       | dummy        | 0        | 0        | 0         | 0                   | 0 |
|       | dummy        | 0        | 0        | 0         | 0                   | 0 |
|       | dummy        | 0        | 0        | 0         | 0                   | 0 |
|       | numiptent    | 10       | 10       | 200       | 200                 | 0 |
| 3000: | kmemsize     | 80854359 | 83363201 | 125829120 | 131072000           |   |
| 0     | lockedpages  | 0        | 0        | 2562      | 2562                | 0 |
| 0     | privvmpages  | 135920   | 171574   | 512000    | 537600              |   |
| 0     | shmpages     | 284      | 284      | 15374     | 15374               | 0 |
|       | dummy        | 0        | 0        | 0         | 0                   | 0 |
|       | numproc      | 65       | 127      | 2000      | 2000                | 0 |
| 0     | physpages    | 37459    | 59833    | 0         | 9223372036854775807 |   |
| 0     | vmguarpages  | 0        | 0        | 35236     | 9223372036854775807 |   |
| 0     | oomguarpages | 37503    | 59833    | 30000     | 30000               |   |
| 0     | numtcpsock   | 15       | 95       | 2000      | 2000                | 0 |
|       | numflock     | 2        | 13       | 1000      | 1100                | 0 |
|       | numpty       | 0        | 1        | 200       | 200                 | 0 |
|       | numsiginfo   | 8        | 30       | 1024      | 1024                | 0 |
| 0     | tcpsndbuf    | 266800   | 1460304  | 9300923   | 17492923            |   |
| 0     | tcprcvbuf    | 1296     | 286512   | 638976    | 1048576             |   |
| 0     | othersockbuf | 13136    | 1270944  | 4650461   | 12842461            |   |
| 0     | dgramrcvbuf  | 0        | 4368     | 4650461   | 4650461             |   |
| 0     | numothersock | 11       | 33       | 8000      | 9000                | 0 |
|       | dcachesize   | 0        | 0        | 11452893  | 11796480            | 0 |
|       | numfile      | 2126     | 4801     | 20480     | 20480               | 0 |

|                     |          |          |                     |                     |   |
|---------------------|----------|----------|---------------------|---------------------|---|
| dummy               | 0        | 0        | 0                   | 0                   | 0 |
| dummy               | 0        | 0        | 0                   | 0                   | 0 |
| dummy               | 0        | 0        | 0                   | 0                   | 0 |
| numiptent           | 10       | 10       | 200                 | 200                 | 0 |
| 0: kmemsize         | 74806381 | 74857340 | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| lockedpages         | 0        | 0        | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| privvmpages         | 8703     | 20362    | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| shmpages            | 1287     | 1303     | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| dummy               | 0        | 0        | 9223372036854775807 | 9223372036854775807 |   |
| 0                   |          |          |                     |                     |   |
| numproc             | 99       | 109      | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| physpages           | 4995     | 15435    | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| vmguarpages         | 0        | 0        | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| oomguarpages        | 5002     | 15435    | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| numtcpsock          | 6        | 7        | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| numflock            | 1        | 7        | 9223372036854775807 | 9223372036854775807 |   |
| 0                   |          |          |                     |                     |   |
| numpty              | 8        | 8        | 9223372036854775807 | 9223372036854775807 |   |
| 0                   |          |          |                     |                     |   |
| numsiginfo          | 2        | 6        | 9223372036854775807 | 9223372036854775807 |   |
| 0                   |          |          |                     |                     |   |
| tcpndbuf            | 125696   | 698000   | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| tcpvcbuf            | 98304    | 1312608  | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| othersockbuf        | 9280     | 24832    | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| dgramrcvbuf         | 0        | 8464     | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| numothersock        | 25       | 30       | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| dcachesize          | 0        | 0        | 9223372036854775807 | 9223372036854775807 |   |
| 0                   |          |          |                     |                     |   |
| numfile             | 1935     | 2135     | 9223372036854775807 |                     |   |
| 9223372036854775807 | 0        |          |                     |                     |   |
| dummy               | 0        | 0        | 9223372036854775807 | 9223372036854775807 |   |
| 0                   |          |          |                     |                     |   |
| dummy               | 0        | 0        | 9223372036854775807 | 9223372036854775807 |   |
| 0                   |          |          |                     |                     |   |

```

dummy          0          0 9223372036854775807 9223372036854775807
0
numiptent      10          10 9223372036854775807
9223372036854775807      0

```

Quote:

Next time this issue will occur try to stop the problem VE (hope CPULIMIT/CPUUNITS adjustment will allow to do that) and look at the user\_beancounters again. There should be no usage when VE is stopped. Please, also look at the slab state. If the consumption value is decreased. And show please /proc/slabinfo output after and before stopping the problem VE.

after i tried to set vzctl set 3000 --cpus 1 all other vzctl's got stuck

there was monstrous filp cache usage in slab, (taken at second event at Tue)

```

mnt_cache      44   75  256  15   1 : tunables 120  60   8 : slabdata   5   5   0
inode_cache    1665 1740  608   6   1 : tunables  54  27   8 : slabdata  290  290   0
dentry_cache   896624 896624  240  16   1 : tunables 120  60   8 : slabdata 56039
56039   0
filp           9343650 9343650  256  15   1 : tunables 120  60   8 : slabdata 622910 622910
0
names_cache    33   33  4096   1   1 : tunables  24  12   8 : slabdata   33   33   0
idr_layer_cache 102  105  528   7   1 :

```

Please look at the attached images, there is a lot of kernel activity about 100% i guess this is runaway unkillable httpd which was in some kind of loop which was also eating filp cache thus crashing the machine,

this process had under 20 open files and was sending msgs threw af\_unix socket as I mentioned bellow

memory consumption by userspace processes were normal under 4 GB all the time, it looks like filp cache was eating the rest

Today we roll out plain apache on that vps, but later I'll try to trigger this event on a test machine.

## File Attachments

- 1) [ram.png](#), downloaded 231 times
- 2) [cpu.png](#), downloaded 208 times

---

Subject: Re: OOM didn't save the machine

Posted by [maratrus](#) on Wed, 01 Apr 2009 15:59:11 GMT

[View Forum Message](#) <> [Reply to Message](#)

---



I have to admit that I'm a little bit confused.  
The only assumption for now is the following.  
The memory freeing process takes place in two steps:  
- uncharging beancounters  
- freeing memory with kfree via RCU.

If some process is stuck inside kernel space, RCU will never be scheduled and thus the second step will never be done. Hence, user\_beancounters output shows nothing interesting (everything is ok) but slab cache consumes a huge amount of memory.

So, we have to single out the process that is stuck inside kernel space. You mentioned that you'd look at the wchan output but it is not the reliable way of debugging.

So, the ideal situation would be a serial console  
[http://wiki.openvz.org/Remote\\_console\\_setup](http://wiki.openvz.org/Remote_console_setup)  
and alt-sysrq-  
- p (twice the number of CPUs)  
- w (several times)  
- t (for all processes calltrace. this is a time consuming operation)

---

---

Subject: Re: OOM didn't save the machine  
Posted by [lazy](#) on Wed, 01 Apr 2009 16:36:26 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

maratrus wrote on Wed, 01 April 2009 10:59I have  
So, the ideal situation would be a serial console  
[http://wiki.openvz.org/Remote\\_console\\_setup](http://wiki.openvz.org/Remote_console_setup)  
and alt-sysrq-  
- p (twice the number of CPUs)  
- w (several times)  
- t (for all processes calltrace. this is a time consuming operation)

ok, thank's for the tips

now we moved to plain apache on this machine(so far it's stable)

but I'll try to run it on test machine and try to trigger this issue again and log to netconsole

---