Subject: Limits for low memory
Posted by Wolfgang Schnerring on Wed, 25 Apr 2007 13:43:40 GMT
View Forum Message <> Reply to Message

Hello!

I've been doing some experiments with OpenVZ, and one of them was, how many VE's can I run on a single machine? The guinea pig machine was a Pentium4/3GHz with 2GB RAM. 230 VE's run just fine, but upon starting the 231st, I reproducably get a kernel panic, "out of memory" (VE configuration is appended below).

What I'd really like to know is, why does that happen?
So I collected resource consumption numbers from /proc/bc for all VE's.
Summed up for 220 VE's (I didn't dare 230 ;-), the maxheld values are
  kmemsize 160957978
  socket buffers 167838032 (tcpsndbuf+tcprcvbuf+dgramrcvbuf+othersockbuf)
  physpages 143708

The physpages amount to 562MB; that should fit into 2GB nicely in my book.

If I understand <http://wiki.openvz.org/UBC_systemwide_configuration> correctly, those two things need to go into low memory, and the total amount for that on x86 is 832MB.

-- Question: Where do the 832 MB come from?

The two numbers above are roughly 314 MB, that's nowhere near 832.
But the wiki-page has the formula kmemsize+sockbuf / 0.4*832MB should be smaller than 1 to be safe. Interestingly enough, that quotient is 0.94, so there might be something up here.

-- Question: Where does the 0.4 come from?

Can anybody help me figure out what's going on here?

Thank you very much for your help,
Wolfgang

```
# VE configuration file
NUMFLOCK="100:110"
VMGUARPAGES="6144:2147483647"
OTHERSOCKBUF="132096:459776"
DISTRIBUTION="debian"
NUMPTY="16:16"
OOMGUARPAGES="6144:2147483647"
DGRAMRCVBUF="132096:132096"
```

CPUUNITS="1000"
NUMIPTENT="128:128"
NUMTCPSOCK="768:768"
PHYSPAGES="0:2147483647"
AVNUMPROC="50:50"
NUMPROC="100:100"
LOCKEDPAGES="32:32"
TCPSNDBUF="10485760:12451840"
VE_ROOT="/var/tmp/vte-wosc/vm33745/root"
NUMSIGINFO="256:256"
KMEMSIZE="5324800:5857280"
SHMPAGES="8192:8192"
TCPRCVBUF="10485760:12451840"
VE_PRIVATE="/var/tmp/vte-wosc/vm33745/private"
NUMFILE="3200:3200"
PRIVVMPAGES="32768:36044"
DCACHESIZE="1228800:1351680"
NUMOTHERSOCK="128:128"
OSTEMPLATE="debian-3.1"
HOSTNAME="vm0.local"
NETIF=" ifname=eth0,mac=02:7B:6A:56:00:02,host_ifname=vif33745.0,hos
t_mac=02:7B:6A:56:00:02 "

Subject: Re: Limits for low memory
Posted by dev on Wed, 25 Apr 2007 14:19:38 GMT
View Forum Message <> Reply to Message

Wolfgang Schnerring wrote:
> Hello!
>
> I've been doing some experiments with OpenVZ, and one of them was, how
> many VE's can I run on a single machine? The guinea pig machine was a
> Pentium4/3GHz with 2GB RAM. 230 VE's run just fine, but upon starting
> the 231st, I reproducably get a kernel panic, "out of memory" (VE
> configuration is appended below).

1. 230 VEs correspond to kir@ results when he was running fine ~120 VEs on 1Gb notebook.
2. can you please attach your kernel panic
(if it was a really panic and not an OOM kill message)?
3. what kernel version do you use?

> What I'd really like to know is, why does that happen?

panic or ?

> So I collected resource consumption numbers from /proc/bc for all VE's.
> Summed up for 220 VE's (I didn't dare 230 ;-), the maxheld values are

> kmemsize 160957978
> socket buffers 167838032 (tcpsndbuf+tcprcvbuf+dgramrcvbuf+othersockbuf)
> physpages 143708
>
> The physpages amount to 562MB; that should fit into 2GB nicely in my book.

is it i686 or x86-64 kernel?

actually it is not all.

1. You also need to take into account dcachesize
(which is not immeadeately accounted by default for optimization purposes,
  but can be enabled with dentry_watermark sysctl and
  usually dcachesize is around ~2Mb per VE, so 460Mb of low memory in your case)
2. plus there are some in-kernel structures which are not-accounted in kmemsize,
   since can't be used for DoS. i.e. kmemsize accounts only those resources
   which can be used for abusing. But real usage of kernel memory can be slightly higher.

> If I understand <http://wiki.openvz.org/UBC_systemwide_configuration>
> correctly, those two things need to go into low memory, and the
> total amount for that on x86 is 832MB.

No. lowmemory is required for: kmemsize, tcp/socket/udp buffers and dcachesize

physpages can go both to low memory and highmemory and you don't know how
it was spread from UBC counters. (*)

> -- Question: Where do the 832 MB come from?

32bit systems can address up to 4GB of virtual RAM (2^32 bytes).
this 4GB address space is split by default as 3 GB for user space
and 1GB for kernel space.
1GB of kernel has kernel itself, different reserved areas (e.g. for vmalloc by default 128Mb),
So around ~850Mb of address space is left for free memory.

The number depends on kernel and configuration:

[dev@dev ~]$ cat /proc/meminfo | grep Low
LowTotal: 903348 kB

the most impact on the size of low memory have vmalloc reserved area.
We used to reserve more when different kernel structures were using it, so low memory was smaller.
Nowdays only 128Mb are reserved by default and about ~890Mb of Low memory should be available.

> The two numbers above are roughly 314 MB, that's nowhere near 832.
> But the wiki-page has the formula kmemsize+sockbuf / 0.4*832MB should

> be smaller than 1 to be safe. Interestingly enough, that quotient is
> 0.94, so there might be something up here.
>
> -- Question: Where does the 0.4 come from?

in your case:
kmemsize ~160Mb
buffers ~160Mb
dcachesize ~460Mb

i.e. ~780Mb in total (see also (*)).

plus you need to take into account fragmentation issues.
e.g. if some memory page was allocated for objects of size 128 bytes,
and has only one such object allocated on it, the page won't be used for objects of another size
and thus some of the memory will by unused and can't be used even under memory pressure.

i.e. it's not that easy like a sum of counters == total memory.

> Can anybody help me figure out what's going on here?

you are welcome!

> Thank you very much for your help,
> Wolfgang
>
> # VE configuration file
> NUMFLOCK="100:110"
> VMGUARPAGES="6144:2147483647"
> OTHERSOCKBUF="132096:459776"
> DISTRIBUTION="debian"
> NUMPTY="16:16"
> OOMGUARPAGES="6144:2147483647"
> DGRAMRCVBUF="132096:132096"
> CPUUNITS="1000"
> NUMIPTENT="128:128"
> NUMTCPSOCK="768:768"
> PHYSPAGES="0:2147483647"
> AVNUMPROC="50:50"
> NUMPROC="100:100"
> LOCKEDPAGES="32:32"
> TCPSNDBUF="10485760:12451840"
> VE_ROOT="/var/tmp/vte-wosc/vm33745/root"
> NUMSIGINFO="256:256"
> KMEMSIZE="5324800:5857280"
> SHMPAGES="8192:8192"
> TCPRCVBUF="10485760:12451840"
> VE_PRIVATE="/var/tmp/vte-wosc/vm33745/private"

> NUMFILE="3200:3200"
> PRIVVMPAGES="32768:36044"
> DCACHESIZE="1228800:1351680"
> NUMOTHERSOCK="128:128"
> OSTEMPLATE="debian-3.1"
> HOSTNAME="vm0.local"
> NETIF=" ifname=eth0,mac=02:7B:6A:56:00:02,host_ifname=vif33745.0,hos
t_mac=02:7B:6A:56:00:02 "
Kirill

Subject: Re: Limits for low memory
Posted by Wolfgang Schnerring on Wed, 25 Apr 2007 16:14:57 GMT
View Forum Message <> Reply to Message

Hello Kirill,

thank you very much for your insights!

* Kirill Korotaev <dev@sw.ru> [2007-04-25 18:19]:
> 1. 230 VEs correspond to kir@ results when he was running fine ~120
> VEs on 1Gb notebook.

That's good to hear, so the number seems to be about right.

> 2. can you please attach your kernel panic
> (if it was a really panic and not an OOM kill message)?

Yes, sorry, it was an OOM kill message. The machine was dead either
way, though.;-) But since that's my fault, I don't mind -- after all,
I pretty much told it to kill itself by asking for more memory than
there is physically available.

The traceback goes over several screens, I've uploaded them to
<http://wosc.de/tmp/oom/> if anybody is interested (please pardon my
bad photography skills).

> 3. what kernel version do you use?
> is it i686 or x86-64 kernel?

It's 2.6.18-028test19 on i686.

> 1. You also need to take into account dcachesize
> (which is not immeadeately accounted by default for optimization purposes,
> but can be enabled with dentry_watermark sysctl and
> usually dcachesize is around ~2Mb per VE, so 460Mb of low memory in your case)

Ah I see, so dcachesize seems to be the missing piece.

I tried to enable dcachesize accounting by "echo 0 0 >
/proc/sys/ubc/dentry_watermark", is that correct?

I reran the 220 VE experiment, and the numbers it yielded were
  kmemsize   127537452
  sockbuf   156184224
  dcachesize 16297263
which again is only 290MB... strange.

> physpages can go both to low memory and highmemory and you don't
> know how it was spread from UBC counters.

I understand. But I would have thought that the kernel tries to put
everything that's possible to high memory first, especially before
starting OOM-killing.

> > -- Question: Where do the 832 MB come from?
> The number depends on kernel and configuration:
> [dev@dev ~]$ cat /proc/meminfo | grep Low
> LowTotal: 903348 kB
> the most impact on the size of low memory have vmalloc reserved area.

You see, this is the kind of insight you'll rarely find in any books or
something. Either, you read the source (which is a little daunting
given the size of the kernel) or you are lucky -- like I am right now
-- and are able to ask an expert. Thanks very much for your help!

> > -- Question: Where does the 0.4 come from?

I'm afraid I still haven't understood what the meaning of this factor
is in the formula on the wiki-page.

> plus you need to take into account fragmentation issues.
> i.e. it's not that easy like a sum of counters == total memory.

That makes sense, but I'd still expect that the sum of the counters to
be at least in the same genereal area -- or is that a wrong expectation?

Wolfgang

Subject: Debian Kernel
Posted by Sidnei Rodrigo Basei on Wed, 25 Apr 2007 21:52:03 GMT
View Forum Message <> Reply to Message

Hi, I installed Open VZ in my Ubuntu with a deb package.

How can I recompile the kernel? I must set my sound, becouse it doesn't work with pre-compiled kernel.

Thanks.

Subject: RE: Debian Kernel
Posted by Michael Flaig on Fri, 27 Apr 2007 10:30:34 GMT
View Forum Message <> Reply to Message

Hi,

If you are using Ubuntu there should be a kernel-source-<version> package

Install that package, go to /usr/src, patch the kernel with openvz und configure it to your needs.

Compile it afterwards with make-kpkg (package: kernel-package) to get a kernel-image package and install it then.

Regards,

Michael

> -----Original Message-----

> From: users-bounces@openvz.org

> [mailto:users-bounces@openvz.org] On Behalf Of "BASEI, Sidnei

> Rodrigo" <sidnei.basei@gmail.com>

> Sent: Wednesday, April 25, 2007 6:52 PM

> To: users@openvz.org

> Subject: [Users] Debian Kernel

>

> Hi, I installed Open VZ in my Ubuntu with a deb package.

>

> How can I recompile the kernel? I must set my sound, becouse

> it doesn't

> work with pre-compiled kernel.

>

>

> Thanks.

> _____

> Users mailing list

> Users@openvz.org

> https://openvz.org/mailman/listinfo/users

>

>