Subject: Re: privvmpages problem in 2.6.15beta
Posted by dev on Fri, 10 Mar 2006 08:45:37 GMT
View Forum Message <> Reply to Message

Miroslav,

Can you please check if this patch helps you?

Thanks,
Kirill

> Hello,
>
> I'm testing 2.6.15-025stab014-default on AMD64 machine and I have some
> problems with privvmpages. It seems to me that value of privvmpages is
> nevere decreased.
> This is current situation in one VPS:
>
> root# cat /proc/user_beancounters
> Version: 2.5
> uid ref resource held maxheld
> barrier limit failcnt
> 102 2481: kmemsize 1072 2147
> 3578265 3816816 0
> lockedpages 0 0
> 32 32 0
> privvmpages 1909456 1915750
> 13107200 13107200 36
> shmpages 0 0
> 8192 8192 0
> dummy 0 0
> 9223372036854775807 9223372036854775807 0
> numproc 17 24
> 65 65 0
> physpages 1310 21194
> 0 9223372036854775807 0
> vmguarpages 0 0
> 6144 9223372036854775807 0
> oomguarpages 1310 21194
> 6144 9223372036854775807 0
> numtcpsock 15 19
> 80 80 0
> numflock 1 10
> 100 110 0
> numpty 1 1
> 16 16 0
> numsiginfo 0 6
> 256 256 0

> tcpsndbuf 2304 44032
> 319488 524288 0
> tcprcvbuf 0 8704
> 319488 524288 0
> othersockbuf 25344 33024
> 132096 336896 0
> dgramrcvbuf 0 4352
> 132096 132096 0
> numothersock 14 19
> 80 80 0
> dcachesize 115497 202857
> 1048576 1097728 0
> numfile 165 1280
> 1280 1280 2995
> dummy 0 0
> 9223372036854775807 9223372036854775807 0
> dummy 0 0
> 9223372036854775807 9223372036854775807 0
> dummy 0 0
> 9223372036854775807 9223372036854775807 0
> numiptent 10 10
> 128 128 0
>
> root# ps axu
> USER PID %CPU %MEM VSZ RSS TTY STAT START TIME COMMAND
> root 1 0.0 0.0 1516 524 ? Ss 13:13 0:00 ini
> daemon 16306 0.0 0.0 1632 352 ? Ss 13:13 0:00
> /sbin/portmap
> root 16368 0.0 0.0 2268 724 ? Ss 13:13 0:00
> /sbin/syslogd
> root 16371 0.0 0.0 1516 364 ? Ss 13:13 0:00 /sbin/klogd
> root 17430 0.0 0.0 2248 676 ? Ss 13:13 0:00
> /usr/sbin/inetd
> root 17533 0.0 0.1 4784 1044 ? Ss 13:13 0:00
> /usr/sbin/sshd
> root 17539 0.0 0.0 2396 876 ? Ss 13:13 0:00
> /sbin/rpc.statd
> root 17546 0.0 0.0 1772 748 ? Ss 13:13 0:00
> /usr/sbin/cron
> root 17591 0.0 0.2 5940 2636 ? Rs 13:14 0:00 sshd:
> root@pts/0
> root 17599 0.0 0.1 2604 1512 pts/0 Rs 13:14 0:00 -bash
> root 22137 0.0 0.0 2508 856 pts/0 R+ 14:02 0:00 ps axu
>
> Avery time I start some program, held value for privvmpages is increased
> and never changed back when program finish. Is it know bug or my
> misunderstanding
> of privvmpages?

>
> Best regards
>

```
diff -urp --new-file ../git/t/arch/x86_64/ia32/ia32_binfmt.c
linux-2.6.15/arch/x86_64/ia32/ia32_binfmt.c
--- ../git/t/arch/x86_64/ia32/ia32_binfmt.c 2006-02-01 13:51:42.000000000 +0300
+++ linux-2.6.15/arch/x86_64/ia32/ia32_binfmt.c 2006-02-01 12:56:12.000000000 +0300
@@ -34,7 +34,7 @@
 #define AT_SYSINFO 32
 #define AT_SYSINFO_EHDR  33

-int sysctl_vsyscall32 = 1;
+int sysctl_vsyscall32 = 0;

 #define ARCH_DLINFO do {  \
  if (sysctl_vsyscall32) { \
diff -urp --new-file ../git/t/arch/x86_64/ia32/syscall32.c linux-2.6.15/arch/x86_64/ia32/syscall32.c
--- ../git/t/arch/x86_64/ia32/syscall32.c 2006-02-01 13:51:42.000000000 +0300
+++ linux-2.6.15/arch/x86_64/ia32/syscall32.c 2006-02-01 12:56:12.000000000 +0300
@@ -10,6 +10,7 @@
 #include <linux/init.h>
 #include <linux/stringify.h>
 #include <linux/security.h>
+#include <linux/elfcore.h>
 #include <asm/proto.h>
 #include <asm/tlbflush.h>
 #include <asm/ia32_unistd.h>
@@ -60,6 +61,10 @@ int syscall32_setup_pages(struct linux_b
    flags, NULL, UB_SOFT))
   goto err_charge;

+ if (sysctl_at_vsyscall == 0)
+  return 0;
+
+ printk(KERN_WARNING "WARN! vsyscalls are broken on x86-64");
  vma = kmem_cache_alloc(vm_area_cachep, SLAB_KERNEL);
  if (!vma)
   goto err_alloc;
diff -urp --new-file ../git/t/fs/exec.c linux-2.6.15/fs/exec.c
--- ../git/t/fs/exec.c 2006-02-01 13:51:59.000000000 +0300
+++ linux-2.6.15/fs/exec.c 2006-02-01 12:56:12.000000000 +0300
@@ -66,6 +66,8 @@ int suid_dumpable = 0;
 EXPORT_SYMBOL(suid_dumpable);
 /* The maximal length of core_pattern is also specified in sysctl.c */

+int sysctl_at_vsyscall;
```

```
+
 static struct linux_binfmt *formats;
 static DEFINE_RWLOCK(binfmt_lock);

diff -urp --new-file ../git/t/include/asm-i386/elf.h linux-2.6.15/include/asm-i386/elf.h
--- ../git/t/include/asm-i386/elf.h 2006-02-01 13:52:03.000000000 +0300
+++ linux-2.6.15/include/asm-i386/elf.h 2006-02-01 12:56:12.000000000 +0300
@@ -136,8 +136,10 @@ extern void __kernel_vsyscall;

 #define ARCH_DLINFO      \
 do {       \
+ if (sysctl_at_vsyscall) {    \
   NEW_AUX_ENT(AT_SYSINFO, VSYSCALL_ENTRY); \
   NEW_AUX_ENT(AT_SYSINFO_EHDR, VSYSCALL_BASE); \
+ }       \
 } while (0)

 /*
diff -urp --new-file ../git/t/include/linux/elfcore.h linux-2.6.15/include/linux/elfcore.h
--- ../git/t/include/linux/elfcore.h 2006-02-01 13:52:08.000000000 +0300
+++ linux-2.6.15/include/linux/elfcore.h 2006-02-01 12:56:12.000000000 +0300
@@ -6,6 +6,8 @@
 #include <linux/time.h>
 #include <linux/user.h>

+extern int sysctl_at_vsyscall;
+
 struct elf_siginfo
 {
  int si_signo;   /* signal number */
diff -urp --new-file ../git/t/include/linux/sysctl.h linux-2.6.15/include/linux/sysctl.h
--- ../git/t/include/linux/sysctl.h 2006-02-01 13:52:08.000000000 +0300
+++ linux-2.6.15/include/linux/sysctl.h 2006-02-01 12:56:12.000000000 +0300
@@ -755,6 +755,7 @@ enum
  FS_AIO_NR=18, /* current system-wide number of aio requests */
  FS_AIO_MAX_NR=19, /* system-wide maximum number of aio requests */
  FS_INOTIFY=20, /* inotify submenu */
+  FS_AT_VSYSCALL=21, /* int: to announce vsyscall data */
 };

 /* /proc/sys/fs/quota/ */
diff -urp --new-file ../git/t/kernel/sysctl.c linux-2.6.15/kernel/sysctl.c
--- ../git/t/kernel/sysctl.c 2006-02-01 13:52:09.000000000 +0300
+++ linux-2.6.15/kernel/sysctl.c 2006-02-01 12:56:12.000000000 +0300
@@ -62,6 +62,7 @@ extern int max_threads;
 extern int sysrq_enabled;
 extern int core_uses_pid;
 extern int suid_dumpable;
```

```
+extern int sysctl_at_vsyscall;
 extern char core_pattern[];
 extern int cad_pid;
 extern int pid_max;
@@ -995,6 +996,14 @@ static ctl_table fs_table[] = {
   .mode  = 0644,
   .proc_handler = &proc_dointvec,
  },
+ {
+  .ctl_name = FS_AT_VSYSCALL,
+  .procname = "vsyscall",
+  .data  = &sysctl_at_vsyscall,
+  .maxlen  = sizeof(int),
+  .mode  = 0644,
+  .proc_handler = &proc_dointvec
+ },
  { .ctl_name = 0 }
 };
```