Subject: [PATCH] Fix NULL ->nsproxy dereference in /proc/*/mounts
Posted by adobriyan on Mon, 15 Jan 2007 16:38:22 GMT
View Forum Message <> Reply to Message

/proc/*/mounstats was fixed, all right, but...

To reproduce:

```
 while true; do
  find /proc -type f 2>/dev/null | xargs cat 1>/dev/null 2>/dev/null;
 done
```

```
BUG: unable to handle kernel NULL pointer dereference at virtual address 0000000c
 printing eip:
c01754df
*pde = 00000000
Oops: 0000 [#28]
Modules linked in: af_packet ohci_hcd e1000 ehci_hcd uhci_hcd usbcore xfs
CPU:    0
EIP:    0060:[<c01754df>]    Not tainted VLI
EFLAGS: 00010286   (2.6.20-rc5 #1)
EIP is at mounts_open+0x1c/0xac
eax: 00000000   ebx: d5898ac0   ecx: d1d27b18   edx: d1d27a50
esi: e6083e10   edi: d3c87f38   ebp: d5898ac0   esp: d3c87ef0
ds: 007b   es: 007b   ss: 0068
Process cat (pid: 18071, ti=d3c86000 task=f7d5f070 task.ti=d3c86000)
Stack: d5898ac0 e6083e10 d3c87f38 c01754c3 c0147c91 c18c52c0 d343f314 d5898ac0
       00008000 d3c87f38 ffffff9c c0147e09 d5898ac0 00000000 00000000 c0147e4b
       00000000 d3c87f38 d343f314 c18c52c0 c015e53e 00001000 08051000 00000101
Call Trace:
 [<c01754c3>] mounts_open+0x0/0xac
 [<c0147c91>] __dentry_open+0xa1/0x18c
 [<c0147e09>] nameidata_to_filp+0x31/0x3a
 [<c0147e4b>] do_filp_open+0x39/0x40
 [<c015e53e>] seq_read+0x128/0x2aa
 [<c0147e8c>] do_sys_open+0x3a/0x6d
 [<c0147efa>] sys_open+0x1c/0x20
 [<c0102b76>] sysenter_past_esp+0x5f/0x85
 [<c02a0033>] unix_stream_recvmsg+0x3bf/0x4bf
 =======================
Code: 5d c3 89 d8 e8 06 e0 f9 ff eb bd 0f 0b eb fe 55 57 56 53 89 d5 8b 40 f0 31 d2 e8 02 c1 fa ff
89 c2 85 c0 74 5c 8b 80 48 04 00 00 <8b> 58 0c 85 db 74 02 ff 03 ff 4a 08 0f 94 c0 84 c0 75 74
85 db
EIP: [<c01754df>] mounts_open+0x1c/0xac SS:ESP 0068:d3c87ef0
```

Signed-off-by: Alexey Dobriyan <adobriyan@openvz.org>

```
 fs/proc/base.c |    8 +++++---
 1 file changed, 5 insertions(+), 3 deletions(-)

--- a/fs/proc/base.c
+++ b/fs/proc/base.c
@@ -371,9 +371,11 @@ static int mounts_open(struct inode *ino

  if (task) {
   task_lock(task);
-  ns = task->nsproxy->mnt_ns;
-  if (ns)
-   get_mnt_ns(ns);
+  if (task->nsproxy) {
+   ns = task->nsproxy->mnt_ns;
+   if (ns)
+    get_mnt_ns(ns);
+  }
   task_unlock(task);
   put_task_struct(task);
  }
```

Subject: Re: [PATCH] Fix NULL ->nsproxy dereference in /proc/*/mounts
Posted by serge on Tue, 16 Jan 2007 16:55:47 GMT
View Forum Message <> Reply to Message

Quoting Alexey Dobriyan (adobriyan@openvz.org):
> /proc/*/mounstats was fixed, all right, but...
>
> To reproduce:
>
> while true; do
>   find /proc -type f 2>/dev/null | xargs cat 1>/dev/null 2>/dev/null;
> done
>
> BUG: unable to handle kernel NULL pointer dereference at virtual address 0000000c
> printing eip:
> c01754df
> *pde = 00000000
> Oops: 0000 [#28]
> Modules linked in: af_packet ohci_hcd e1000 ehci_hcd uhci_hcd usbcore xfs
> CPU: 0
> EIP: 0060:[<c01754df>]   Not tainted VLI
> EFLAGS: 00010286   (2.6.20-rc5 #1)
> EIP is at mounts_open+0x1c/0xac
> eax: 00000000   ebx: d5898ac0   ecx: d1d27b18   edx: d1d27a50
> esi: e6083e10   edi: d3c87f38   ebp: d5898ac0   esp: d3c87ef0
> ds: 007b   es: 007b   ss: 0068

> Process cat (pid: 18071, ti=d3c86000 task=f7d5f070 task.ti=d3c86000)
> Stack: d5898ac0 e6083e10 d3c87f38 c01754c3 c0147c91 c18c52c0 d343f314 d5898ac0
>   00008000 d3c87f38 ffffff9c c0147e09 d5898ac0 00000000 00000000 c0147e4b
>   00000000 d3c87f38 d343f314 c18c52c0 c015e53e 00001000 08051000 00000101
> Call Trace:
> [<c01754c3>] mounts_open+0x0/0xac
> [<c0147c91>] __dentry_open+0xa1/0x18c
> [<c0147e09>] nameidata_to_filp+0x31/0x3a
> [<c0147e4b>] do_filp_open+0x39/0x40
> [<c015e53e>] seq_read+0x128/0x2aa
> [<c0147e8c>] do_sys_open+0x3a/0x6d
> [<c0147efa>] sys_open+0x1c/0x20
> [<c0102b76>] sysenter_past_esp+0x5f/0x85
> [<c02a0033>] unix_stream_recvmsg+0x3bf/0x4bf
> =======================
> Code: 5d c3 89 d8 e8 06 e0 f9 ff eb bd 0f 0b eb fe 55 57 56 53 89 d5 8b 40 f0 31 d2 e8 02 c1 fa ff 89 c2 85 c0 74 5c 8b 80 48 04 00 00 <8b> 58 0c 85 db 74 02 ff 03 ff 4a 08 0f 94 c0 84 c0 75 74 85 db
> EIP: [<c01754df>] mounts_open+0x1c/0xac SS:ESP 0068:d3c87ef0
>
> Signed-off-by: Alexey Dobriyan <adobriyan@openvz.org>

Yup, race with do_exit()'s call to exit_namespaces(). Thanks for
the patch.

Acked-by: Serge Hallyn <serue@us.ibm.com>

> ---
>
> fs/proc/base.c | 8 +++++---
> 1 file changed, 5 insertions(+), 3 deletions(-)
>
> --- a/fs/proc/base.c
> +++ b/fs/proc/base.c
> @@ -371,9 +371,11 @@ static int mounts_open(struct inode *ino
>
>   if (task) {
>   task_lock(task);
> - ns = task->nsproxy->mnt_ns;
> - if (ns)
> -   get_mnt_ns(ns);
> + if (task->nsproxy) {
> +   ns = task->nsproxy->mnt_ns;
> +   if (ns)
> +   get_mnt_ns(ns);
> + }
>   task_unlock(task);
>   put_task_struct(task);

> }
>
> -
> To unsubscribe from this list: send the line "unsubscribe linux-kernel" in
> the body of a message to majordomo@vger.kernel.org
> More majordomo info at http://vger.kernel.org/majordomo-info.html
> Please read the FAQ at http://www.tux.org/lkml/