Subject: [PATCH] Don't create tunnels with '%' in name.
Posted by Pavel Emelianov on Thu, 21 Feb 2008 12:05:27 GMT
View Forum Message <> Reply to Message

Four tunnel drivers (ip_gre, ipip, ip6_tunnel and sit) can
receive a pre-defined name for a device from the userspace.
Since these drivers call the register_netdevice() after this
(rtnl_lock is held), the device's name may contain a '%'
character.

Not sure how bad is this to have a device with a '%' in its
name, but all the other places either use the register_netdev(),
or explicitly call dev_alloc_name() before registering, i.e.
do not allow for such names.

Signed-off-by: Pavel Emelyanov <xemul@openvz.org>

```
diff --git a/net/ipv4/ip_gre.c b/net/ipv4/ip_gre.c
index 63f6917..6b9744f 100644
--- a/net/ipv4/ip_gre.c
+++ b/net/ipv4/ip_gre.c
@@ -274,19 +274,24 @@ static struct ip_tunnel * ipgre_tunnel_locate(struct ip_tunnel_parm
*parms, int
  if (!dev)
    return NULL;

+ if (strchr(name, '%')) {
+  if (dev_alloc_name(dev, name) < 0)
+   goto failed_free;
+ }
+
  dev->init = ipgre_tunnel_init;
  nt = netdev_priv(dev);
  nt->parms = *parms;

- if (register_netdevice(dev) < 0) {
-  free_netdev(dev);
-  goto failed;
- }
+ if (register_netdevice(dev) < 0)
+  goto failed_free;

  dev_hold(dev);
  ipgre_tunnel_link(nt);
  return nt;
```

```
+failed_free:
+ free_netdev(dev);
 failed:
  return NULL;
 }
diff --git a/net/ipv4/ipip.c b/net/ipv4/ipip.c
index da28158..118e7d9 100644
--- a/net/ipv4/ipip.c
+++ b/net/ipv4/ipip.c
@@ -236,19 +236,24 @@ static struct ip_tunnel * ipip_tunnel_locate(struct ip_tunnel_parm
*parms, int c
  if (dev == NULL)
   return NULL;

+ if (strchr(name, '%')) {
+  if (dev_alloc_name(dev, name) < 0)
+   goto failed_free;
+ }
+
  nt = netdev_priv(dev);
  dev->init = ipip_tunnel_init;
  nt->parms = *parms;

- if (register_netdevice(dev) < 0) {
-  free_netdev(dev);
-  goto failed;
- }
+ if (register_netdevice(dev) < 0)
+  goto failed_free;

  dev_hold(dev);
  ipip_tunnel_link(nt);
  return nt;

+failed_free:
+ free_netdev(dev);
 failed:
  return NULL;
 }
diff --git a/net/ipv6/ip6_tunnel.c b/net/ipv6/ip6_tunnel.c
index cd94064..fa83d70 100644
--- a/net/ipv6/ip6_tunnel.c
+++ b/net/ipv6/ip6_tunnel.c
@@ -245,17 +245,24 @@ static struct ip6_tnl *ip6_tnl_create(struct ip6_tnl_parm *p)
  if (dev == NULL)
   goto failed;

+ if (strchr(name, '%')) {
```

```
+  if (dev_alloc_name(dev, name) < 0)
+   goto failed_free;
+ }
+
  t = netdev_priv(dev);
  dev->init = ip6_tnl_dev_init;
  t->parms = *p;

- if ((err = register_netdevice(dev)) < 0) {
-  free_netdev(dev);
-  goto failed;
- }
+ if ((err = register_netdevice(dev)) < 0)
+  goto failed_free;
+
  dev_hold(dev);
  ip6_tnl_link(t);
  return t;
+
+failed_free:
+ free_netdev(dev);
 failed:
  return NULL;
 }
diff --git a/net/ipv6/sit.c b/net/ipv6/sit.c
index e77239d..a09a6b0 100644
--- a/net/ipv6/sit.c
+++ b/net/ipv6/sit.c
@@ -179,6 +179,11 @@ static struct ip_tunnel * ipip6_tunnel_locate(struct ip_tunnel_parm
*parms, int
  if (dev == NULL)
   return NULL;

+ if (strchr(name, '%')) {
+  if (dev_alloc_name(dev, name) < 0)
+   goto failed_free;
+ }
+
  nt = netdev_priv(dev);
  dev->init = ipip6_tunnel_init;
  nt->parms = *parms;
@@ -186,16 +191,16 @@ static struct ip_tunnel * ipip6_tunnel_locate(struct ip_tunnel_parm
*parms, int
  if (parms->i_flags & SIT_ISATAP)
   dev->priv_flags |= IFF_ISATAP;

- if (register_netdevice(dev) < 0) {
-  free_netdev(dev);
```

```
-  goto failed;
- }
+ if (register_netdevice(dev) < 0)
+  goto failed_free;

  dev_hold(dev);

  ipip6_tunnel_link(nt);
  return nt;

+failed_free:
+ free_netdev(dev);
 failed:
  return NULL;
 }
```

Subject: Re: [PATCH] Don't create tunnels with '%' in name.
Posted by Patrick McHardy on Thu, 21 Feb 2008 12:10:16 GMT
View Forum Message <> Reply to Message

Pavel Emelyanov wrote:
> Four tunnel drivers (ip_gre, ipip, ip6_tunnel and sit) can
> receive a pre-defined name for a device from the userspace.
> Since these drivers call the register_netdevice() after this
> (rtnl_lock is held), the device's name may contain a '%'
> character.
>
> Not sure how bad is this to have a device with a '%' in its
> name, but all the other places either use the register_netdev(),
> or explicitly call dev_alloc_name() before registering, i.e.
> do not allow for such names.
>
> Signed-off-by: Pavel Emelyanov <xemul@openvz.org>
>
> ---
>
> diff --git a/net/ipv4/ip_gre.c b/net/ipv4/ip_gre.c
> index 63f6917..6b9744f 100644
> --- a/net/ipv4/ip_gre.c
> +++ b/net/ipv4/ip_gre.c
> @@ -274,19 +274,24 @@ static struct ip_tunnel * ipgre_tunnel_locate(struct ip_tunnel_parm
*parms, int
>   if (!dev)
>   return NULL;
>
> + if (strchr(name, '%')) {
> + if (dev_alloc_name(dev, name) < 0)

> +   goto failed_free;
> + }
> +

It would be nicer to replace the entire hand-made name
allocation to remove the 100 device limit.

Subject: Re: [PATCH] Don't create tunnels with '%' in name.
Posted by Pavel Emelianov on Thu, 21 Feb 2008 12:17:27 GMT
View Forum Message <> Reply to Message

Patrick McHardy wrote:
> Pavel Emelyanov wrote:
>> Four tunnel drivers (ip_gre, ipip, ip6_tunnel and sit) can
>> receive a pre-defined name for a device from the userspace.
>> Since these drivers call the register_netdevice() after this
>> (rtnl_lock is held), the device's name may contain a '%'
>> character.
>>
>> Not sure how bad is this to have a device with a '%' in its
>> name, but all the other places either use the register_netdev(),
>> or explicitly call dev_alloc_name() before registering, i.e.
>> do not allow for such names.
>>
>> Signed-off-by: Pavel Emelyanov <xemul@openvz.org>
>>
>> ---
>>
>> diff --git a/net/ipv4/ip_gre.c b/net/ipv4/ip_gre.c
>> index 63f6917..6b9744f 100644
>> --- a/net/ipv4/ip_gre.c
>> +++ b/net/ipv4/ip_gre.c
>> @@ -274,19 +274,24 @@ static struct ip_tunnel * ipgre_tunnel_locate(struct ip_tunnel_parm
*parms, int
>>   if (!dev)
>>   return NULL;
>>
>> + if (strchr(name, '%')) {
>> + if (dev_alloc_name(dev, name) < 0)
>> +   goto failed_free;
>> + }
>> +
>
>
> It would be nicer to replace the entire hand-made name
> allocation to remove the 100 device limit.

>

Actually, I thought the same, but fixing % in names looks like a
BUG-fix for 2.6.25, while removing the hand-made name allocation
looks like an enhancement for 2.6.26. No?

Thanks,
Pavel

Subject: Re: [PATCH] Don't create tunnels with '%' in name.
Posted by Patrick McHardy on Thu, 21 Feb 2008 12:22:01 GMT
View Forum Message <> Reply to Message

Pavel Emelyanov wrote:
> Patrick McHardy wrote:
>
>> It would be nicer to replace the entire hand-made name
>> allocation to remove the 100 device limit.
>>
>
> Actually, I thought the same, but fixing % in names looks like a
> BUG-fix for 2.6.25, while removing the hand-made name allocation
> looks like an enhancement for 2.6.26. No?

Well, its so closely related that I guess it would still look
like a bugfix :) But changing this in 2.6.26 is also fine of
course, your patch just reminded me since I wanted to change
this for a long time and repeatedly forgot about it again.

Subject: [PATCH] Don't limit the number of tunnels with generic name explicitly.
Posted by Pavel Emelianov on Thu, 21 Feb 2008 12:38:16 GMT
View Forum Message <> Reply to Message

Patrick McHardy wrote:
> Pavel Emelyanov wrote:
>> Patrick McHardy wrote:
>>
>>> It would be nicer to replace the entire hand-made name
>>> allocation to remove the 100 device limit.
>>>
>> Actually, I thought the same, but fixing % in names looks like a
>> BUG-fix for 2.6.25, while removing the hand-made name allocation
>> looks like an enhancement for 2.6.26. No?
>

>
> Well, its so closely related that I guess it would still look
> like a bugfix :) But changing this in 2.6.26 is also fine of
> course, your patch just reminded me since I wanted to change
> this for a long time and repeatedly forgot about it again.

Ok, point taken ;) Here's the 2nd patch that does so. If David
decides it can go to 2.6.25, that would be good, otherwise this
patch will fit the 2.6.26 as well.

Changelog:

Use the added dev_alloc_name() call to create tunnel device name,
rather than iterate in a hand-made loop with an artificial limit.

Thanks Patrick for noticing this.

Signed-off-by: Pavel Emelyanov <xemul@openvz.org>

```
diff --git a/include/net/ip6_tunnel.h b/include/net/ip6_tunnel.h
index c17fa1f..6512d85 100644
--- a/include/net/ip6_tunnel.h
+++ b/include/net/ip6_tunnel.h
@@ -14,8 +14,6 @@
 /* capable of receiving packets */
 #define IP6_TNL_F_CAP_RCV 0x20000

-#define IP6_TNL_MAX 128
-
 /* IPv6 tunnel */

 struct ip6_tnl {
diff --git a/net/ipv4/ip_gre.c b/net/ipv4/ip_gre.c
index 6b9744f..e7821ba 100644
--- a/net/ipv4/ip_gre.c
+++ b/net/ipv4/ip_gre.c
@@ -259,16 +259,8 @@ static struct ip_tunnel * ipgre_tunnel_locate(struct ip_tunnel_parm
*parms, int

  if (parms->name[0])
   strlcpy(name, parms->name, IFNAMSIZ);
- else {
-  int i;
-  for (i=1; i<100; i++) {
-   sprintf(name, "gre%d", i);
-   if (__dev_get_by_name(&init_net, name) == NULL)
```

```
-    break;
-  }
-  if (i==100)
-    goto failed;
- }
+ else
+  sprintf(name, "gre%%d");

  dev = alloc_netdev(sizeof(*t), name, ipgre_tunnel_setup);
  if (!dev)
@@ -292,7 +284,6 @@ static struct ip_tunnel * ipgre_tunnel_locate(struct ip_tunnel_parm
*parms, int

 failed_free:
  free_netdev(dev);
-failed:
  return NULL;
 }

diff --git a/net/ipv4/ipip.c b/net/ipv4/ipip.c
index 118e7d9..dbaed69 100644
--- a/net/ipv4/ipip.c
+++ b/net/ipv4/ipip.c
@@ -221,16 +221,8 @@ static struct ip_tunnel * ipip_tunnel_locate(struct ip_tunnel_parm
*parms, int c

  if (parms->name[0])
   strlcpy(name, parms->name, IFNAMSIZ);
- else {
-  int i;
-  for (i=1; i<100; i++) {
-   sprintf(name, "tunl%d", i);
-   if (__dev_get_by_name(&init_net, name) == NULL)
-    break;
-  }
-  if (i==100)
-   goto failed;
- }
+ else
+  sprintf(name, "tunl%%d");

  dev = alloc_netdev(sizeof(*t), name, ipip_tunnel_setup);
  if (dev == NULL)
@@ -254,7 +246,6 @@ static struct ip_tunnel * ipip_tunnel_locate(struct ip_tunnel_parm *parms,
int c

 failed_free:
  free_netdev(dev);
```

```
-failed:
  return NULL;
 }

diff --git a/net/ipv6/ip6_tunnel.c b/net/ipv6/ip6_tunnel.c
index fa83d70..78f4388 100644
--- a/net/ipv6/ip6_tunnel.c
+++ b/net/ipv6/ip6_tunnel.c
@@ -229,18 +229,11 @@ static struct ip6_tnl *ip6_tnl_create(struct ip6_tnl_parm *p)
  char name[IFNAMSIZ];
  int err;

- if (p->name[0]) {
+ if (p->name[0])
   strlcpy(name, p->name, IFNAMSIZ);
- } else {
-  int i;
-  for (i = 1; i < IP6_TNL_MAX; i++) {
-   sprintf(name, "ip6tnl%d", i);
-   if (__dev_get_by_name(&init_net, name) == NULL)
-    break;
-  }
-  if (i == IP6_TNL_MAX)
-   goto failed;
- }
+ else
+  sprintf(name, "ip6tnl%%d");
+
  dev = alloc_netdev(sizeof (*t), name, ip6_tnl_dev_setup);
  if (dev == NULL)
   goto failed;
diff --git a/net/ipv6/sit.c b/net/ipv6/sit.c
index a09a6b0..1656c00 100644
--- a/net/ipv6/sit.c
+++ b/net/ipv6/sit.c
@@ -164,16 +164,8 @@ static struct ip_tunnel * ipip6_tunnel_locate(struct ip_tunnel_parm
*parms, int

  if (parms->name[0])
   strlcpy(name, parms->name, IFNAMSIZ);
- else {
-  int i;
-  for (i=1; i<100; i++) {
-   sprintf(name, "sit%d", i);
-   if (__dev_get_by_name(&init_net, name) == NULL)
-    break;
-  }
-  if (i==100)
```

```
-   goto failed;
- }
+ else
+   sprintf(name, "sit%%d");

  dev = alloc_netdev(sizeof(*t), name, ipip6_tunnel_setup);
  if (dev == NULL)
```

Subject: Re: [PATCH] Don't limit the number of tunnels with generic name explicitly.
Posted by Patrick McHardy on Thu, 21 Feb 2008 12:45:15 GMT
View Forum Message <> Reply to Message

Pavel Emelyanov wrote:
> Patrick McHardy wrote:
>> Pavel Emelyanov wrote:
>>> Patrick McHardy wrote:
>>>
>>>> It would be nicer to replace the entire hand-made name
>>>> allocation to remove the 100 device limit.
>>>>
>>> Actually, I thought the same, but fixing % in names looks like a
>>> BUG-fix for 2.6.25, while removing the hand-made name allocation
>>> looks like an enhancement for 2.6.26. No?
>>
>> Well, its so closely related that I guess it would still look
>> like a bugfix :) But changing this in 2.6.26 is also fine of
>> course, your patch just reminded me since I wanted to change
>> this for a long time and repeatedly forgot about it again.
>
> Ok, point taken ;) Here's the 2nd patch that does so. If David
> decides it can go to 2.6.25, that would be good, otherwise this
> patch will fit the 2.6.26 as well.
>
> Changelog:
>
> Use the added dev_alloc_name() call to create tunnel device name,
> rather than iterate in a hand-made loop with an artificial limit.
>
> Thanks Patrick for noticing this.
>
> Signed-off-by: Pavel Emelyanov <xemul@openvz.org>

Looks good to me, thanks.

Subject: Re: [PATCH] Don't limit the number of tunnels with generic name explicitly.
Posted by davem on Sun, 24 Feb 2008 04:19:52 GMT
View Forum Message <> Reply to Message

From: Pavel Emelyanov <xemul@openvz.org>
Date: Thu, 21 Feb 2008 15:38:16 +0300

> Changelog:
>
> Use the added dev_alloc_name() call to create tunnel device name,
> rather than iterate in a hand-made loop with an artificial limit.
>
> Thanks Patrick for noticing this.
>
> Signed-off-by: Pavel Emelyanov <xemul@openvz.org>

Applied, but I had to rework this in two places that didn't
apply cleanly.

The ip_gre.c and ipip.c changes remove a "failed" label but
that can't be done in the current tree as there are other
existing references.

Subject: Re: [PATCH] Don't limit the number of tunnels with generic name explicitly.
Posted by Pavel Emelianov on Tue, 26 Feb 2008 07:47:44 GMT
View Forum Message <> Reply to Message

David Miller wrote:
> From: Pavel Emelyanov <xemul@openvz.org>
> Date: Thu, 21 Feb 2008 15:38:16 +0300
>
>> Changelog:
>>
>> Use the added dev_alloc_name() call to create tunnel device name,
>> rather than iterate in a hand-made loop with an artificial limit.
>>
>> Thanks Patrick for noticing this.
>>
>> Signed-off-by: Pavel Emelyanov <xemul@openvz.org>
>
> Applied, but I had to rework this in two places that didn't
> apply cleanly.

That's because you skipped the first patch titled "Don't create
tunnels with '%' in name.", which adds the dev_alloc_name() call
and tosses the error paths a bit. Without this first patch, these
four drivers become broken :(When user doesn't specify the name,

the device's name will be e.g. "tunl%d", but not "tunl0" like
he expects.

> The ip_gre.c and ipip.c changes remove a "failed" label but
> that can't be done in the current tree as there are other
> existing references.
>

Yup :(this code was removed in that first patch...

Subject: Re: [PATCH] Don't limit the number of tunnels with generic name explicitly.
Posted by davem on Tue, 26 Feb 2008 21:30:54 GMT
View Forum Message <> Reply to Message

From: Pavel Emelyanov <xemul@openvz.org>
Date: Tue, 26 Feb 2008 10:47:44 +0300

> That's because you skipped the first patch titled "Don't create
> tunnels with '%' in name.", which adds the dev_alloc_name() call
> and tosses the error paths a bit. Without this first patch, these
> four drivers become broken :(When user doesn't specify the name,
> the device's name will be e.g. "tunl%d", but not "tunl0" like
> he expects.

Please respin and post the first patch, I had no idea there was a
dependency.
