
Subject: [PATCH 1/4] netns: Tag the network flow with the network namespace it is in (v2)

Posted by [den](#) on Tue, 04 Dec 2007 09:52:45 GMT

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

As well as marking flows this indirectly marks the ipv4 routing cache as every routing entry contains a flow.

It is useful to add the network namespace into flows as frequently the routing information for ingoing and outgoing network packets is collected into a flow structure which is then used for several functions as it sorts out what is going on.

Changes from v1:

- remove flow.h dependency from net_namespace.h

Signed-off-by: Denis V. Lunev <den@openvz.org>

Signed-off-by: Eric W. Biederman <ebiederm@xmission.com>

include/net/flow.h | 2 ++

1 files changed, 2 insertions(+), 0 deletions(-)

diff --git a/include/net/flow.h b/include/net/flow.h

index af59fa5..9590bbe 100644

--- a/include/net/flow.h

+++ b/include/net/flow.h

@ @ -10,7 +10,9 @ @

#include <linux/in6.h>

#include <asm/atomic.h>

+struct net;

struct flowi {

+ struct net *fl_net;

int oif;

int iif;

__u32 mark;

--

1.5.3.rc5

Subject: Re: [PATCH 1/4] netns: Tag the network flow with the network namespace it is in (v2)

Posted by [Stephen Hemminger](#) on Tue, 04 Dec 2007 14:26:05 GMT

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

On Tue, 4 Dec 2007 12:53:33 +0300

"Denis V. Lunev" <den@openvz.org> wrote:

> As well as marking flows this indirectly marks the ipv4 routing cache
> as every routing entry contains a flow.
>
> It is useful to add the network namespace into flows as frequently
> the routing information for ingoing and outgoing network packets is
> collected into a flow structure which is then used for several functions
> as it sorts out what is going on.
>
> Changes from v1:
> - remove flow.h dependency from net_namespace.h
>
> Signed-off-by: Denis V. Lunev <den@openvz.org>
> Signed-off-by: Eric W. Biederman <ebiederm@xmission.com>
> ---
> include/net/flow.h | 2 ++
> 1 files changed, 2 insertions(+), 0 deletions(-)
>
> diff --git a/include/net/flow.h b/include/net/flow.h
> index af59fa5..9590bbe 100644
> --- a/include/net/flow.h
> +++ b/include/net/flow.h
> @@ -10,7 +10,9 @@
> #include <linux/in6.h>
> #include <asm/atomic.h>
>
> +struct net;
> struct flowi {
> + struct net *fl_net;
> int oif;
> int iif;
> __u32 mark;
> --

Can this be made conditional on network namespaces being configured on?
That way the flow structure won't have to grow taking more space.
It matters in DoS attacks where flow cache becomes a critical resource.

Subject: Re: [PATCH 1/4] netns: Tag the network flow with the network namespace it is in (v2)

Posted by [den](#) on Tue, 04 Dec 2007 18:42:11 GMT

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

Stephen Hemminger wrote:

> Can this be made conditional on network namespaces being configured on?
> That way the flow structure won't have to grow taking more space.
> It matters in DoS attacks where flow cache becomes a critical resource.

could you exactly point me out the flow cache your are talking about.
Is this dst entry cache or struct flow_cache described in the
net/core/flow.c

For the latter case, there is completely no difference in the size on my
x86_64 host with SLAB allocator, i.e. there are 30 objects per slab
with/without fl_net (objsize = 128).

Regards,
Den

Subject: Re: [PATCH 1/4] netns: Tag the network flow with the network namespace
it is in (v2)

Posted by [Herbert Xu](#) on Tue, 04 Dec 2007 22:40:33 GMT

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

Denis V. Lunev <den@sw.ru> wrote:

>

> could you exactly point me out the flow cache your are talking about.

> Is this dst entry cache or struct flow_cache described in the

> net/core/flow.c

The flow object is embedded in struct rtable so does its size change?

Cheers,

--

Visit Openswan at <http://www.openswan.org/>

Email: Herbert Xu ~{PmV>Hl~} <herbert@gondor.apana.org.au>

Home Page: <http://gondor.apana.org.au/~herbert/>

PGP Key: <http://gondor.apana.org.au/~herbert/pubkey.txt>

Subject: Re: [PATCH 1/4] netns: Tag the network flow with the network namespace
it is in (v2)

Posted by [den](#) on Wed, 05 Dec 2007 06:49:15 GMT

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

Herbert Xu wrote:

> Denis V. Lunev <den@sw.ru> wrote:

>> could you exactly point me out the flow cache your are talking about.

>> Is this dst entry cache or struct flow_cache described in the

>> net/core/flow.c

>

> The flow object is embedded in struct rtable so does its size change?

>

> Cheers,

SLAB allocator, x86_64 host

Before the patch:

```
ip6_dst_cache      384  10
xfrm_dst_cache     384  10
ip_dst_cache       384  10
```

After the patch:

```
ip6_dst_cache      384  10
xfrm_dst_cache     384  10
ip_dst_cache       384  10
```

Regards,
Den

Subject: Re: [PATCH 1/4] netns: Tag the network flow with the network namespace it is in (v2)

Posted by [davem](#) on Wed, 05 Dec 2007 10:10:36 GMT

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

From: "Denis V. Lunev" <den@sw.ru>

Date: Tue, 04 Dec 2007 21:42:49 +0300

> Stephen Hemminger wrote:

> > Can this be made conditional on network namespaces being configured on?

> > That way the flow structure won't have to grow taking more space.

> > It matters in DoS attacks where flow cache becomes a critical resource.

>

> could you exactly point me out the flow cache your are talking about.

> Is this dst entry cache or struct flow_cache described in the

> net/core/flow.c

>

> For the latter case, there is completely no difference in the size on my

> x86_64 host with SLAB allocator, i.e. there are 30 objects per slab

> with/without fl_net (objsize = 128).

This may be true, but another thing to consider is that flow objects sit on the stack in many call sites.

I won't let this block your patch, but I want you to be cognizant of this issue in the future, it's not all about SLAB.

You should also BTW consider how this change will effect D-cache access patterns and L2 cache utilization. Some object access patterns may not fit in the cache, which did beforehand, which can kill performance. We're talking about something which gets

touched multiple times per packet at routing rates in the million packet per second range.

Subject: Re: [PATCH 1/4] netns: Tag the network flow with the network namespace it is in (v2)

Posted by [davem](#) on Wed, 05 Dec 2007 10:13:18 GMT

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

From: "Denis V. Lunev" <den@openvz.org>

Date: Tue, 4 Dec 2007 12:53:33 +0300

> As well as marking flows this indirectly marks the ipv4 routing cache
> as every routing entry contains a flow.
>
> It is useful to add the network namespace into flows as frequently
> the routing information for ingoing and outgoing network packets is
> collected into a flow structure which is then used for several functions
> as it sorts out what is going on.
>
> Changes from v1:
> - remove flow.h dependency from net_namespace.h
>
> Signed-off-by: Denis V. Lunev <den@openvz.org>
> Signed-off-by: Eric W. Biederman <ebiederm@xmission.com>

Hmmm, actually I change my mind.

> @@ -10,7 +10,9 @@
> #include <linux/in6.h>
> #include <asm/atomic.h>
>
> +struct net;
> struct flowi {
> + struct net *fl_net;
> int oif;
> int iif;
> __u32 mark;

I'm not applying this, it's going to have a negative impact on routing performance.

It also changes the semantics of the flowi object in a way I very much dislike, in that there is now non-clobberable state in there.

Previously only addressing identifying objects were present in the flow, you could use it any context, and there were no pointer dereferencing or object references from this thing. It was very

simple.

That is no longer the case after your patch and I don't want us to go down this path.

Please find another way to implement this.
