

## Re: freebsd-net Digest, Vol 207, Issue 2

---

*Source:* <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/net/2007-03/msg00305.html>

---

- *From:* rhinux <[linuxinfoplus@xxxxxxxxxx](mailto:linuxinfoplus@xxxxxxxxxx)>
  - *Date:* Wed, 21 Mar 2007 10:33:59 +0800
- 

( 2007-3-20 H8:00 freebsd-net-request@xxxxxxxxxx TMS

Send freebsd-net mailing list submissions to  
[freebsd-net@xxxxxxxxxx](mailto:freebsd-net@xxxxxxxxxx)

To subscribe or unsubscribe via the World Wide Web, visit  
<http://lists.freebsd.org/mailman/listinfo/freebsd-net>  
or, via email, send a message with subject or body 'help' to  
[freebsd-net-request@xxxxxxxxxx](mailto:freebsd-net-request@xxxxxxxxxx)

You can reach the person managing the list at  
[freebsd-net-owner@xxxxxxxxxx](mailto:freebsd-net-owner@xxxxxxxxxx)

When replying, please edit your Subject line so it is more specific  
than "Re: Contents of freebsd-net digest..."

Today's Topics:

1. Re: Wireshark (Shteryana Shopova)
2. Interface index hack in IP\_ADD\_MEMBERSHIP (Bruce M Simpson)
3. Re: Interface index hack in IP\_ADD\_MEMBERSHIP (Eugene Grosbein)
4. Re: [PATCH] bge(4) patch for -STABLE (Pete French)
5. Re: Interface index hack in IP\_ADD\_MEMBERSHIP (Bruce M Simpson)
6. Re: rc.order wrong (ipfw) (Doug Barton)
7. Re: Wireshark (Randall Stewart)
8. Re: [PATCH] bge(4) patch for -STABLE (Mike Tanca)
9. Re: [PATCH] bge(4) patch for -STABLE (Pete French)
10. [PATCH] Multicast reccounting in network stack (Bruce M Simpson)
11. Re: [PATCH] Multicast reccounting in network stack  
(Andre Oppermann)
12. Re: netisr\_direct (Keith Arner)
13. networking code and splx() (Ignacio Rey)
14. Re: rc.order wrong (ipfw) (David Gilbert)
15. Re: networking code and splx() (Julian Elischer)
16. Re: networking code and splx() (Bruce M. Simpson)
17. Re: PMTU Discovery support (Kevin Lahey)

Re: freebsd-net Digest, Vol 207, Issue 2

18. Re: PMTU Discovery support (Kevin Lahey)
19. Re: PMTU Discovery support (Bruce M. Simpson)
20. Re: [PATCH] Multicast reccounting in network stack (Bruce M Simpson)
21. ifstated behavior (Alexandre Biancalana)
22. Re: networking code and splx() (John Hay)
23. Re: networking code and splx() (Eygene Ryabinkin)
24. "em" driver shuts down interface when "ifconfig media 100baseTX" is invoked (Infraservice hostmaster)

---

Message: 1

Date: Mon, 19 Mar 2007 14:25:49 +0200

From: "Shteryana Shopova" <shteryana@xxxxxxxx>

Subject: Re: Wireshark

To: "manuel.ochoa@xxxxxxxx" <manuel.ochoa@xxxxxxxx>

Cc: Max Laier <max@xxxxxxxxxxxx>, freebsd-net@xxxxxxxx

Message-ID:

<61b573980703190525s30f22648od0ecdec01879d0c@xxxxxxxxxxxxxxxx>

Content-Type: text/plain; charset=UTF-8; format=flowed

On 3/19/07, manuel.ochoa@xxxxxxxx <manuel.ochoa@xxxxxxxx> wrote:

Max, correct me if I'm wrong but tcpdump will only give you the headers, is that correct? This is fine most of the time but sometimes I need to capture full frames.

Nope – that's not correct –

#tcpdump -s 0

will capture full frames.

Shteryana

Thanks

Manuel Ochoa CCNP MCSA MCSE MCDBA

----- Original Message -----

From: Max Laier <max@xxxxxxxxxxxx>

To: freebsd-net@xxxxxxxx

Cc: manuel.ochoa@xxxxxxxx

Sent: Saturday, March 17, 2007 2:05:06 PM

Re: freebsd-net Digest, Vol 207, Issue 2

Subject: Re: Wireshark

On Saturday 17 March 2007 19:16, manuel.choa@xxxxxxxx wrote:

Can someone please explain the difference between  
Wireshark and  
Wireshark-lite. I would like to install a packet sniffer on my  
FreeBSD  
box for CLI only. Thanks,

What's wrong with tcpdump(8)? Other than that building either the real or  
the -lite version with "WITHOUT\_X11" defined will get you the  
cli-version. "-lite" seems to just disable a couple of dissectors that  
have a lot of external dependencies.

--

/"\ Best regards, | mlaier@xxxxxxxxxxxxx  
\/ Max Laier | ICQ #67774661  
X <http://pf4freebsd.love2party.net/> | mlaier@EFnet  
/\ ASCII Ribbon Campaign | Against HTML Mail and News

---

Expecting? Get great news right away with email Auto-Check.  
Try the Yahoo! Mail Beta.  
[http://advision.webevents.yahoo.com/mailbeta/newmail\\_tools.html](http://advision.webevents.yahoo.com/mailbeta/newmail_tools.html)

---

freebsd-net@xxxxxxxxxxxxx mailing list  
<http://lists.freebsd.org/mailman/listinfo/freebsd-net>  
To unsubscribe, send any mail to "freebsd-net- unsubscribe@xxxxxxxxxxxxx"

---

Message: 2  
Date: Mon, 19 Mar 2007 14:28:52 +0000  
From: Bruce M Simpson <bms@xxxxxxxxxxxxxxxxxxxx>  
Subject: Interface index hack in IP\_ADD\_MEMBERSHIP  
To: net@xxxxxxxxxxxxx  
Message-ID: <45FE9E24.8010201@xxxxxxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Hi,

I plan to get rid of the ugly little ip\_multicast\_if() hack in the IP

Re: freebsd-net Digest, Vol 207, Issue 2

stack.=

Before I do, is anyone actually using this?

RFC 3678 specifies a protocol independent API for socket group memberships which allow joins on interfaces referenced by index. This is intended to support IGMPv3 and MLDv2.

Regards,  
BMS

---

Message: 3

Date: Mon, 19 Mar 2007 22:28:37 +0700

From: Eugene Grosbein <eugen@xxxxxxxxxxxx>

Subject: Re: Interface index hack in IP\_ADD\_MEMBERSHIP

To: Bruce M Simpson <bms@xxxxxxxxxxxxxxxx>

Cc: net@xxxxxxxxxxxx

Message-ID: <20070319152837.GA3984@xxxxxxxxxxxxxxxxxxxx>

Content-Type: text/plain; charset=us-ascii

On Mon, Mar 19, 2007 at 02:28:52PM +0000, Bruce M Simpson wrote:

I plan to get rid of the ugly little ip\_multicast\_if() hack in the IP stack.=

Before I do, is anyone actually using this?

RFC 3678 specifies a protocol independent API for socket group memberships which allow joins on interfaces referenced by index. This is intended to support IGMPv3 and MLDv2.

I recall that routed and ripd used to utilize something similar long time ago. I'm not sure if they have switched to another API.

Eugene

---

Message: 4

Date: Mon, 19 Mar 2007 16:28:58 +0000

From: Pete French <petefrench@xxxxxxxxxxxxxxxx>

Subject: Re: [PATCH] bge(4) patch for -STABLE

To: freebsd-net@xxxxxxxxxxxx, jkim@xxxxxxxxxxxx

Cc: freebsd-stable@xxxxxxxxxxxx

Message-ID: <E1HTKjC-000GQc-No@xxxxxxxxxxxxxxxxxxxx>

I have made bge(4) patch for -STABLE (sorry, not suitable for RELENG\_6\_2):

What dates stable is this relative to ? I am trying to apply your patch to a cvsup of stable pulled on the day/time you sent your email, but parts of it are failing for me unfortunately. I would like to test this as I have a number of bge interfaces running on some systems.

-pete.

---

Message: 5  
Date: Mon, 19 Mar 2007 16:51:29 +0000  
From: Bruce M Simpson <bms@xxxxxxxxxxxxxxxx>  
Subject: Re: Interface index hack in IP\_ADD\_MEMBERSHIP  
To: Eugene Grosbein <eugen@xxxxxxxx>  
Cc: net@xxxxxxxx  
Message-ID: <45FEBF91.2000709@xxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Eugene Grosbein wrote:

I recall that routed and ripd used to utilize something similar long time ago. I'm not sure if they have switched to another API.

You're right -- this would break routed on point-to-point interfaces.

They didn't, unless it was updated at the upstream, i.e. rhyolite.com.

This means that the RFC1724 hack can't be safely deprecated without breaking this use case, until routed is updated to use the RFC 3678 protocol-independent ASM API.

Linux uses a slightly different technique to work-around this; ip\_mreq is expanded to ip\_mreqn internally, and the interface index is explicitly passed around in the kernel.

The blocker in the FreeBSD case which prevents us simply adopting this is the source interface selection logic in ip\_output().

Regards,  
BMS

Message: 6

Date: Mon, 19 Mar 2007 10:03:42 -0700

From: Doug Barton <doug@xxxxxxxxxxxx>

Subject: Re: rc.order wrong (ipfw)

To: Kian Mohageri <kian.mohageri@xxxxxxxx>

Cc: freebsd-net@xxxxxxxxxxxx, Mark Andrews <Mark\_Andrews@xxxxxx>, freebsd-rc@xxxxxxxxxxxx

Message-ID: <45FEC26E.40504@xxxxxxxxxxxx>

Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Kian Mohageri wrote:

After re-reading your original idea, I think I understand a little better what you mean to do. For clarification, are you proposing that the [early] firewall scripts do nothing if firewall\_late\_enable=YES, and then have all firewalling taken care of later in the boot process (i.e. post-networking) by firewall\_late?

I think I might have misunderstood your original proposal:)

I think so too. :) To be clear, what I'm suggesting is that we move ipfw and pf to a spot in the rcorder that is ahead of netif, along with ipfilter which is already there. I am not suggesting that we change their functionality, just the ordering. As a completely separate thing (although they could be done at the same time) I am suggesting \_adding\_ a new script for "late" firewall rules (where "late" is defined as after netif) so that people who want to do firewall-related things that require netif (like cloned interfaces, FQDN rules, etc.) will have a standard way to accomplish that.

Thanks for the opportunity to clarify,

Doug

--

This .signature sanitized for your protection

---

Message: 7

Date: Mon, 19 Mar 2007 13:41:21 -0400

From: Randall Stewart <rrs@xxxxxxxx>

Subject: Re: Wireshark

To: Shteryana Shopova <shteryana@xxxxxxxx>

Re: freebsd-net Digest, Vol 207, Issue 2

Cc: Max Laier <max@xxxxxxxxxxxxxxxx>, "manuel.ochoa@xxxxxxxx" <manuel.ochoa@xxxxxxxx>, freebsd-net@xxxxxxxx  
Message-ID: <45FECB41.3070601@xxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Shteryana Shopova wrote:

On 3/19/07, manuel.ochoa@xxxxxxxx <manuel.ochoa@xxxxxxxx> wrote:

Max, correct me if I'm wrong but tcpdump will only give you the headers, is that correct? This is fine most of the time but sometimes I need to capture full frames.

Nope - that's not correct -

```
#tcpdump -s 0
```

will capture full frames.

But nothing IMO beats wireshark for being able to go in and analyze a dump .. searching on various condition's fields etc..

It does not matter to me generally how its collected wireshark/tcpdump -s 0..

But to analyze it.. give me wireshark any day :-D

R

--

Randall Stewart  
NSSTG - Cisco Systems Inc.  
803-345-0369 <or> 803-317-4952 (cell)

---

Message: 8  
Date: Mon, 19 Mar 2007 13:51:48 -0400  
From: Mike Tancsa <mike@xxxxxxxx>  
Subject: Re: [PATCH] bge(4) patch for -STABLE  
To: Pete French <petefrench@xxxxxxxxxxxxxxxx>, freebsd-net@xxxxxxxx, jkim@xxxxxxxx  
Cc: freebsd-stable@xxxxxxxx

Re: freebsd-net Digest, Vol 207, Issue 2

Message-ID: <200703191753.12JHreQ3061174@xxxxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset="us-ascii"; format=flowed

At 12:28 PM 3/19/2007, Pete French wrote:

I have made bge(4) patch for -STABLE (sorry, not suitable  
for  
RELENG\_6\_2):

What dates stable is this relative to ? I am trying to apply your  
patch to a cvsup of stable pulled on the day/time you sent your email,  
but parts of it are failing for me unfortunately. I would like to test this  
as I have a number of bge interfaces running on some systems.

Mine applied cleanly to sources from last Friday. So far so good for  
me in that I have not yet seen the watchdog timeout (previously once  
every 4 days or so) but its too early to tell. Still, I have not  
seen any regressions with it yet since installing it last  
Saturday. This is a fairly busy recursive DNS server

---Mike

---

Message: 9  
Date: Mon, 19 Mar 2007 18:08:23 +0000  
From: Pete French <petefrench@xxxxxxxxxxxxxxxxxx>  
Subject: Re: [PATCH] bge(4) patch for -STABLE  
To: freebsd-net@xxxxxxxxxxxx, jkim@xxxxxxxxxxxx, mike@xxxxxxxxxxxx  
Cc: freebsd-stable@xxxxxxxxxxxx  
Message-ID: <E1HTMHP-000HZU-H9@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>

Mine applied cleanly to sources from last Friday.

O.K., that works (now I have the correct date in my supfile). Will  
give it a shot...

-pete.

---

Message: 10  
Date: Mon, 19 Mar 2007 18:52:32 +0000  
From: Bruce M Simpson <bms@xxxxxxxxxxxxxxxxxx>

Re: freebsd-net Digest, Vol 207, Issue 2

Subject: [PATCH] Multicast reccounting in network stack  
To: freebsd-net@xxxxxxxxxxxxx  
Message-ID: <45FEDBF0.4050105@xxxxxxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Hi,

A patch against -CURRENT is now available:  
[http://people.freebsd.org/~bms/dump/multi\\_reccounting.diff](http://people.freebsd.org/~bms/dump/multi_reccounting.diff)

This is a fairly sweeping architectural change which should resolve memory leaks and potential panics with the network stack as a whole, to better support interface detach at runtime.

I'd like to check it in as soon as possible as it fixes the root cause of the problems we have had with carp and pfsync in our stack. NetBSD has implemented reccounting like this for some time now, so it does not suffer from the same problems.

Regards,  
BMS

---

Message: 11  
Date: Mon, 19 Mar 2007 20:11:56 +0100  
From: Andre Oppermann <andre@xxxxxxxxxxxxx>  
Subject: Re: [PATCH] Multicast reccounting in network stack  
To: Bruce M Simpson <bms@xxxxxxxxxxxxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxxxxx  
Message-ID: <45FEE07C.4060501@xxxxxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Bruce M Simpson wrote:

Hi,

A patch against -CURRENT is now available:  
[http://people.freebsd.org/~bms/dump/multi\\_reccounting.diff](http://people.freebsd.org/~bms/dump/multi_reccounting.diff)

This is a fairly sweeping architectural change which should resolve memory leaks and potential panics with the network stack as a whole, to better support interface detach at runtime.

I'd like to check it in as soon as possible as it fixes the root cause of the problems we have had with carp and pfsync in our stack. NetBSD has implemented reccounting like this for some time now, so it does not suffer from the same problems.

Patch looks good. :-)

--

Andre

-----

Message: 12

Date: Mon, 19 Mar 2007 15:54:55 -0400

From: "Keith Arner" <vornum@xxxxxxxx>

Subject: Re: netisr\_direct

To: "Robert Watson" <rwatson@xxxxxxxx>

Cc: net@xxxxxxxx

Message-ID:

<8e552a500703191254qba4e194y810eb99a9f07bed8@xxxxxxxxxxxxxxxx>

Content-Type: text/plain; charset=ISO-8859-1; format=flowed

On 3/11/07, Robert Watson <rwatson@xxxxxxxx> wrote:

There are several ways we could start to reduce contention on that lock:

(3) Move towards greater granularity of locking for the tcbinfo: instead of a single mutex, move to more than one locks, so that different connections processed simultaneously are likely to involve different locks. For listen sockets, we would have to have a special case, such as a single lock to serialize simultaneous lock acquisitions of multiple chain locks (for example).

I've been thinking about this approach, and it does sound like the simplest to implement of the three proposed. However, the special case of the listen socket seems like it would complicate matters.

It seems to me, however, that the complication of the listen socket could be simplified if the listen sockets were maintained in a separate pcb list from the rest of the TCP sockets (the fully connected sockets). If the two types of sockets were thus separated,

Re: freebsd-net Digest, Vol 207, Issue 2

the code would acquire the lock on the bucket in the connect hash, and search the connect hash; if there was a miss, acquire the lock on the listen list and then search the listen list.

The lock on the listen list should follow the locks on the connect buckets in the locking order. The connection bucket should be locked first, as delivery of data would be the common case. To create a new connection in the tcp\_input path, both the connect bucket and the listen list would need to be locked (connect bucket as a new entry would be added to the list, and listen list as the accept socket would be protected from going away).

Keith

---

Message: 13  
Date: Mon, 19 Mar 2007 20:37:09 +0100  
From: Ignacio Rey <unixero@xxxxxxxx>  
Subject: networking code and splx()  
To: freebsd-net@xxxxxxxxxxxx  
Message-ID: <20070319203709.1272a470@debian>  
Content-Type: text/plain; charset=US-ASCII

Hello everyone,

I'm studying a bit the FreeBSD networking code.

I've read "TCP/IP illustrated vol 2" by G. R. Wright and W. R. Stevens, which describes code in 4.4BSD-lite.

Now I'm taking a look at FreeBSD 6.2 release. Some things are different, many others kept the same. What I'm confused about is that in 4.4BSD there are many calls to the splx() family of functions, and I didn't see any of them in the networking code in FreeBSD. However the 'grep' program showed me that they are used in other parts of the kernel.

The question is: Have calls to these functions been wrapped? or are they simply not used in this context?

Thanks in advance,

Ignacio

---

Message: 14  
Date: Mon, 19 Mar 2007 15:12:52 -0500  
From: David Gilbert <dgilbert@xxxxxxx>  
Subject: Re: rc.order wrong (ipfw)  
To: Doug Barton <doughb@xxxxxxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxxxx, Mark Andrews <Mark\_Andrews@xxxxxxx>, Kian Mohageri <kian.mohageri@xxxxxxxx>, freebsd-rc@xxxxxxxxxxxx  
Message-ID: <17918.61124.353668.804988@xxxxxxxxxxxx>  
Content-Type: text/plain; charset=us-ascii

"Doug" ==  
Doug  
Barton  
<doughb@xxxxxxxxxxxx>  
writes:

Doug> Kian Mohageri wrote:

I agree VERY MUCH with this sort of approach. It would be a much cleaner solution than completely separate handling of all of these different problems. I'm trying to get an idea of what all of the major problems with the current order are, and these are the ones I'm aware of:

– ipfw blocks by default (names unresolvable, rtsol breaks) –  
ipf/pf pass by default (services are unprotected)

I think a firewall\_boot script (similar to what you've proposed) could potentially solve all of these problems.

Doug> exception, not the rule. Furthermore (and I'm betraying a Doug> prejudice here) I think that firewall rules that rely on name Doug> resolution are absolutely nuts, and I say that with many years Doug> of experience as a professional DNS and system administrator.

I think you're misreading the above. The poster is saying that because ipfw's default behaviour is block, loading it at the wrong time can break other startup items because they require name resolution or the sending of packets (rtsol).

Dave.

--

=====  
=====  
David Gilbert, Independent Contractor.	Two things can be
Mail: dave@xxxxxxx	equal if and only if they
<http://daveg.ca>	are precisely opposite.

=====  
=====  
-----GLO-----  
=====

-----  
Message: 15  
Date: Mon, 19 Mar 2007 13:23:17 -0700  
From: Julian Elischer <julian@xxxxxxxxxxxx>  
Subject: Re: networking code and splx()  
To: Ignacio Rey <unixero@xxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxxxx  
Message-ID: <45FEF135.2050203@xxxxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Ignacio Rey wrote:

Hello everyone,

I'm studying a bit the FreeBSD networking code.

I've read "TCP/IP illustrated vol 2" by G. R. Wright and W. R. Stevens, which describes code in 4.4BSD-lite.

Now I'm taking a look at FreeBSD 6.2 release. Some things are different, many others kept the same. What I'm confused about is that in 4.4BSD there are many calls to the splx() family of functions, and I didn't see any of them in the networking code in FreeBSD. However the 'grep' program showed me that they are used in other parts of the kernel.

There was a major change in the way that synchronization was done between FreeBSD 4.x and FreeBSD 5.X.

The changes were to support real multiprocessor operation and were extensive. The 'spl' method of operation was only able to protect operation on a single CPU.

You should probably get a copy of "The design of the FreeBSD (um 5.2 I think) Operating System by Kirk McKusick and George Neville-Neil. It goes into some of the changes. (many changes have happened since then too). the New locking largely depends on Mutexes.

Re: freebsd-net Digest, Vol 207, Issue 2

The question is: Have calls to these functions been wrapped? or are they simply not used in this context?

Thanks in advance,

Ignacio

---

freebsd-net@xxxxxxxxxxxxx mailing list  
<http://lists.freebsd.org/mailman/listinfo/freebsd-net>  
To unsubscribe, send any mail to "freebsd-net-unsubscribe@xxxxxxxxxxxxx"

---

Message: 16

Date: Mon, 19 Mar 2007 22:14:33 +0000  
From: "Bruce M. Simpson" <bms@xxxxxxxxxxxxx>  
Subject: Re: networking code and splx()  
To: Ignacio Rey <unixero@xxxxxxxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxxxxx  
Message-ID: <45FF0B49.9060008@xxxxxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Ignacio Rey wrote:

...  
The question is: Have calls to these functions been wrapped? or are they simply not used in this context?

splx() and friends have been no-ops since FreeBSD 5.x was branched. Synchronization is now done using other mechanisms such as mutexes and spin locks. See the new man page locking(9) in -CURRENT.

Regards,  
BMS

---

Message: 17

Date: Mon, 19 Mar 2007 14:54:22 -0700  
From: Kevin Lahey <kml@xxxxxxxxxxxxxxxxxxxxx>  
Subject: Re: PMTU Discovery support  
To: freebsd-net@xxxxxxxxxxxxx  
Message-ID: <20070319145422.39bfdcd@xxxxxxxxxxxxxxxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset=US-ASCII

Re: freebsd-net Digest, Vol 207, Issue 2

On Tue, 6 Mar 2007 10:35:42 +0530  
"aditya kiran" <adityaa.kiran@xxxxxxxxxx> wrote:

RFC 1191 says to increase the PMTU at some interval (15 minutes default)

10 minutes.

next time a packet is sent, this will be used... and if PMTU is really increased, no ICMP error will be received. that shows an increase in the PMTU. I'm trying to understand if this mechanism is there in freebsd. any on this is appreciated

It looks to me as though FreeBSD stores per-host MTU data in the hostcache, which gets purged after five minutes of inactivity. If that's actually how it works, then, yes, FreeBSD should indeed periodically probe for larger PMTUs.

Of course, the real test is to set up a few hosts and see what happens, rather than speculating based on a quick perusal of the code. :-)

Kevin  
kml@xxxxxxxxxxxxxxxxxxxx

---

Message: 18  
Date: Mon, 19 Mar 2007 17:21:56 -0700  
From: Kevin Lahey <kml@xxxxxxxxxxxxxxxxxxxx>  
Subject: Re: PMTU Discovery support  
To: freebsd-net@xxxxxxxxxxxx  
Message-ID: <20070319172156.68cba0a9@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset=US-ASCII

On Mon, 19 Mar 2007 14:54:22 -0700  
Kevin Lahey <kml@xxxxxxxxxxxxxxxxxxxx> wrote:

Of course, the real test is to set up a few hosts and see what happens, rather than speculating based on a quick perusal of the code. :-)

After my slap-dash read of the current FreeBSD code, I was a little concerned that I'd missed something. As penance, I set up a quick

Re: freebsd-net Digest, Vol 207, Issue 2

experiment with four hosts connected in a line, A <-> B <-> C <-> D, set the MTU on the links from B to C to 512, and ran ttcp from A to D. PMTUD worked correctly. Then I suspended the ttcp process, went away for an hour, and resumed it. Watching tcpdump, it appears that 512 octet packets continued to be sent, with no attempt at probing.

That would seem to be a bug.

The boxes were running FreeBSD-6.1, but I can't really vouch for the particular kernel configuration. It could well be that the problem is with the loose nut behind the wheel, rather than with FreeBSD. :-)

Kevin  
kml@xxxxxxxxxxxxxxxxxx

---

Message: 19  
Date: Tue, 20 Mar 2007 01:47:27 +0000  
From: "Bruce M. Simpson" <bms@xxxxxxxxxx>  
Subject: Re: PMTU Discovery support  
To: Kevin Lahey <kml@xxxxxxxxxxxxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxx  
Message-ID: <45FF3D2F.3040000@xxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Kevin Lahey wrote:

The boxes were running FreeBSD-6.1, but I can't really vouch for the particular kernel configuration. It could well be that the problem is with the loose nut behind the wheel, rather than with FreeBSD. :-)

I believe PMTU measurements may only be relied upon for active TCP connections, but it's been a while since I read this code.

It would be useful if non-TCP drivers such as gre(4) could be extended to perform PMTU discovery and auto-tune their MTU based on this, as manually setting the MTU is a bit random and can result in horrible fragmentation when going across the big-I Internet.

I imagine doing this would require changes to the icmp input path and a bit of abstraction.

Regards,  
BMS

---

Message: 20  
Date: Tue, 20 Mar 2007 01:48:00 +0000  
From: Bruce M Simpson <bms@xxxxxxxxxxxxxxx>  
Subject: Re: [PATCH] Multicast refcounting in network stack  
To: Andre Oppermann <andre@xxxxxxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxxxx  
Message-ID: <45FF3D50.3050604@xxxxxxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Andre Oppermann wrote:

[http://people.freebsd.org/~bms/dump/multi\\_refcounting.diff](http://people.freebsd.org/~bms/dump/multi_refcounting.diff)

Patch looks good. :-)

Committed, with some changes.

Regards,  
BMS

---

Message: 21  
Date: Tue, 20 Mar 2007 00:22:55 -0300  
From: Alexandre Biancalana <ale@xxxxxxxx>  
Subject: ifstated behavior  
To: freebsd-net@xxxxxxxxxxxx  
Message-ID: <45FF538F.1050405@xxxxxxxx>  
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

Hi list,

First, excuse-me by the off-topic message, I asked this on - questions but I don't have any answer.

I'm trying to setup ifstated to check two links and if some go down, do some actions like change pf rules and machine's route.

My doubt is about the execution order/repetition of the states body of ifstated.conf, in all configs that I tried just the last check is executed always, follow and example:

ifstated.conf:

```
=====
loglevel debug
```

```
ping1 = '( "ping -q -c 1 -t 3 www.site1.com <http://www.site1.com> > /dev/null" every 10 ) '
```

Re: freebsd-net Digest, Vol 207, Issue 2

```
ping2 = '( "ping -q -c 1 -t 3 www.site2.com <http://www.site2.com> >
/dev/null" every 10 ) '
```

```
state one {
if ! ( $ping1 && $ping2 ) {
set-state two
}
}
```

```
state two {
```

```
init {
run "logger -p console.notice -t ifstated 'Restarting
network !'"
}
```

```
if ( $ping && $ping2 ) {
set-state one
}
}
```

```
=====
```

```
# ifstated -dv
ping1 = "( "ping -q -c 1 -t 3 www.site1.com <http://www.site1.com> >
/dev/null" every 10 ) "
ping2 = "( "ping -q -c 1 -t 3 www.site2.com <http://www.site2.com> >
/dev/null" every 10 ) "
ifstated: initial state: one
ifstated: changing state to one
ifstated: running ping -q -c 1 -t 3 www.site1.com <http:// www.site1.com>
/dev/null
ifstated: running ping -q -c 1 -t 3 www.site2.com <http:// www.site2.com>
/dev/null
ifstated: started
ifstated: changing state to two
ifstated: running ping -q -c 1 -t 3 www.site1.com <http:// www.site1.com>
/dev/null
ifstated: running ping -q -c 1 -t 3 www.site2.com <http:// www.site2.com>
/dev/null
ifstated: running ping -q -c 1 -t 3 www.site2.com <http:// www.site2.com>
/dev/null
```

Re: freebsd-net Digest, Vol 207, Issue 2

ifstated: running ping -q -c 1 -t 3 www.site2.com <http:// www.site2.com>

/dev/null

As you can see, after change state ifstated execute only the \*last\* check command of the statement (ping2) forever....

This is the expected behavior ?

I'm running 6-STABLE + ifstated-20050505 (instaled via /usr/ports/net/ifstated)

Thanks for any help.

Alexandre

---

Message: 22  
Date: Tue, 20 Mar 2007 09:31:50 +0200  
From: John Hay <jhay@xxxxxxxxxxxxxx>  
Subject: Re: networking code and splx()  
To: "Bruce M. Simpson" <bms@xxxxxxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxxxx  
Message-ID: <20070320073150.GA19859@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset=us-ascii

On Mon, Mar 19, 2007 at 10:14:33PM +0000, Bruce M. Simpson wrote:

Ignacio Rey wrote:

...  
The question is: Have calls to these functions been wrapped?  
or are they  
simply not used in this context?

splx() and friends have been no-ops since FreeBSD 5.x was branched.  
Synchronization is now done using other mechanisms such as mutexes and  
spin locks. See the new man page locking(9) in -CURRENT.

It does not seem to get installed:

Doing a grep for locking in /usr/src/share/man/man9/Makefile produce  
nothing.

John

John Hay --- John.Hay@xxxxxxxxxxxxxxxxxxxxx / jhay@xxxxxxxxxxxxx

---

Message: 23  
Date: Tue, 20 Mar 2007 10:40:51 +0300  
From: Eygene Ryabinkin <rea-fbsd@xxxxxxxxxxxxx>  
Subject: Re: networking code and splx()  
To: John Hay <jhay@xxxxxxxxxxxxxxxxxx>  
Cc: freebsd-net@xxxxxxxxxxxxx, "Bruce M. Simpson" <bms@xxxxxxxxxxxxx>  
Message-ID: <20070320074051.GH96806@xxxxxxxxxxxxx>  
Content-Type: text/plain; charset=koi8-r

John, good day.

Tue, Mar 20, 2007 at 09:31:50AM +0200, John Hay wrote:

splx() and friends have been no-ops since FreeBSD 5.x was branched. Synchronization is now done using other mechanisms such as mutexes and spin locks. See the new man page locking(9) in -CURRENT.

It does not seem to get installed:

The locking.9 is not the part of the current build as the comment of the initial commit of that file says.

Doing a grep for locking in /usr/src/share/man/man9/Makefile produce nothing.

But if you're running -CURRENT, then you can view the page from the sources (assuming that you have the system sources in /usr/src/):  
\$ groff -Tascii -mandoc /usr/src/share/man/man9/locking.9 | less

Our you can download the locking.9 from  
<http://www.freebsd.org/cgi/cvsweb.cgi/src/share/man/man9/locking.9>

---  
Eygene

Re: freebsd-net Digest, Vol 207, Issue 2

Message: 24  
Date: Tue, 20 Mar 2007 04:19:55 -0400 (EDT)  
From: Infraservice hostmaster <hostmaster@xxxxxxxxxxxxxxxxxxxx>  
Subject: "em" driver shuts down interface when "ifconfig media 100baseTX" is invoked  
To: freebsd-net@xxxxxxxxxxxxx  
Message-ID: <m1HTZZT-000t9XC@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>  
Content-Type: text/plain; charset=us-ascii

FreeBSD cg-gw 6.2-PRERELEASE FreeBSD 6.2-PRERELEASE #10: Tue Jan 16 01:40:27 EST 2007 root@taint:/usr/obj/usr/src/sys/FW-YQ i386

I did:  
"ifconfig em2 media 100baseTX mediaopt full-duplex"

ifconfig em2 now reports:  
...  
media: Ethernet 100baseTX <full-duplex> (autoselect)  
status: no carrier

"ifconfig em2 media 100baseTX" also does this

We tried it on 2 different 100baseTX interfaces with the same result

"ifconfig em2 media autonegotiate" brings the interface back up

It doesn't seem to happen on another interface which autonegotiates 10baseT half-duplex, however

This seems to indicate there might be a driver problem since the 2 interfaces connect to very different kinds of equipment

Any suggestions?

-----

---

freebsd-net@xxxxxxxxxxxxx mailing list  
<http://lists.freebsd.org/mailman/listinfo/freebsd-net>  
To unsubscribe, send any mail to "freebsd-net-unsubscribe@xxxxxxxxxxxxx"

End of freebsd-net Digest, Vol 207, Issue 2

Re: freebsd-net Digest, Vol 207, Issue 2

\*\*\*\*\*

---

freebsd-net@xxxxxxxxxxx mailing list

<http://lists.freebsd.org/mailman/listinfo/freebsd-net>

To unsubscribe, send any mail to "freebsd-net-unsubscribe@xxxxxxxxxxx"