

netfilter archeology: 18 years from 2.3 to 4.x

Harald Welte <laforge@gnumonks.org>

What is this about

- netfilter history
- netfilter who-is-who
- netfilter anecdotes
- netfilter folklore
- netfilter world domination

Context

- late 1990ies
- Internet was still new to many people
- Internet Security was still rather new
 - think e.g. of the "ping of death" problems.
- no git, not even subversion, but: CVS(!)
 - even pre-bitkeeper, so no kernel global revision control
 - Linus was applying patches and making "pre" releases as tar-ball every so often
- no authorship annotation / commit history
- no virtual machines, testing on physical boxes, long boot cycles

pre-netfilter

The pre-netfilter days

- Linux 1.2, 1.3 and 2.0 had `ipfwadm` (Jos Vos et al)
 - who in the audience has still used that? Raise your hand!
- Linux 2.2 had `ipchains` (Rusty Russell)
 - who in the audience has still used that? Raise your hand!
 - Rusty was doing some sysadmin work at an ISP and was doing his job so well that he had plenty of spare time
 - He was *immensely* inspired by a talk by DaveM on beating the hell out of Solaris on SPARC
 - wanted to do more Linux stuff, met WatchGuard
 - proposed to do a proper redesign of the Linux firewall if they pay him for 6-12 months
 - ... which they did, so mid-1998 to mid-1999, he hacked away on it.

Creation Timeline

Who the hell are you, and why are you playing with my kernel? I want to clear up some people's misconceptions: I am *no kernel guru*. I know this, because my kernel work has brought me into contact with some of them: David S. Miller, Alexey Kuznetsov, Andi Kleen, Alan Cox. However, they're all busy doing the deep magic, leaving me to wade in the shallow end where it's safe.

— Rusty Russell

- July 20, 1998: Rusty posts initial netfilter design to netdev list
- January 29, 1999: netfilter v0.1 released
- August 26, 1999: netfilter included in kernel 2.3.15
- November 1999: *core team* established with Marc Boucher + Rusty Russell

netfilter v0.1

```
Date: 1999-01-29 10:36:34
From: Paul Rusty Russell <Paul.Russell@rustcorp.com.au>
Subject: [ipchains-dev] netfilter v0.1 released.
```

Hi all,

Just in case people are sick of all this "stable kernel" crap, and want their interesting behaviour back, here's the first alpha-quality cut of my new firewall/NAT/masquerading/transproxy/redirect/portforward framework, from [...]

This work is a result of the vast amount of user feedback I've had on things such as transparent proxying, masquerading, ipfwadm, ipchains, etc. Includes a 400k patch against 2.2.0. (84 files changed, 2509 insertions, 12097 deletions).

It's all set in slush at this stage, so if you have any comments, please feel free to leap forward and abuse me...

In other news of January 1999:

- Yahoo bought Geocities
- Clinton faced impeachemnt trial

netfilter merged in kernel 2.3.15

```
From: Linus Torvalds <torvalds@transmeta.com>  
Subject: Linux-2.3.15..  
Date: Wed, 25 Aug 1999 16:36:10 -0700 (PDT)
```

There's a rather huge patch-set out there now, taking the 2.3.x series to 2.3.15. [...]

Other features that don't impact everybody, but are rather major:

```
* firewalling is gone (again), replaced by an even more generic  
netfilter facility.  
[...]
```

```
Have fun,  
Linus
```

In other news:

- East Timor becomes independent of Indonesia
- Vladimir Putin becomes Prime Minister of Russia for the first time

Rusty (at Linux Beer Hike 2000)



Linux Beer Hike 2000



Linux Beer Hike 2000



Marc Boucher (at OLS 2000)



Core Team Timeline

- November 1999: *core team* established with *Marc Boucher* + *Rusty Russell*
- Sydney Linux Expo: *James Morris* joins core team
- September 2000: *Harald Welte* joins core team
- November 2001: *Jozsef Kadlecsik* joins core team
- August 2003: *Martin Josefsson* joins core team
 - Rusty, Marc and James become *emeritus* members
- January 2004: *Patrick McHardy* joins core team
- October 2005: *Yasuyuki Kozakai* joins core team
- February 2007: *Pablo Neira* joins core team
- October 2012: *Eric Leblond* and *Florian Westphal* join core team
 - Harald, Martin and Yasuyuki enter *emeritus* status

James Morris (in 2008)



(sorry, I have no earlier picture of him)

(Rustys) Humor

From <http://www.netfilter.org/about.html#history>

Following James' assimilation into the collective, our efforts were mainly directed towards preparations for the release of Netfilter as part of the upcoming 2.4 kernel.

It was the dawn of the third age of Linux firewalling; a time of great struggle and heroic deeds. It was our last, best hope for peace. Great communities were founded, old civilizations were lost, and new alliances were formed.

James' missions during this period included the *continued perversion of the networking code*, such that it was now possible to load an ASN.1 parser into the kernel and *inflict grave terror upon unsuspecting SNMP packets*; and to extend the IP stack into userspace with Perl.

Now peering squarely into the abyss, we noticed the good deeds of a young kernel warrior named Harald Welte, who seemed to actually understand the NAT code. Accordingly, his distinctiveness was added to the collective. With balance restored, the netfilter juggernaut was now free to accelerate into the brave new world of Linux 2.4 and face it's greatest challenge: users.

(Rustys) Humor

```
Date: Fri, 13 Oct 2000 16:26:06 +1100
From: Rusty Russell <rusty@linuxcare.com.au>
To: netfilter@lists.samba.org, netfilter-devel@lists.samba.org
Subject: [CORE TEAM] New Member Announce
```

The Netfilter Core Team is proud to welcome Harald Welte into its hallowed botherhood.

Harald Welte has frequently answered user questions on the mailing list, and authored the IRC connection tracking and NAT modules. He even documented what he'd done! And then fixed some of the bugs!

This shocking and revolutionary approach to software development will fill a much-needed void in the Netfilter Team. Assuming he survives the inauguration ceremony.

Meanwhile, in other news:

- Bill Gates steps down as CEO of Microsoft

2000: Harald Welte



- active in German BBS community and pre-internet offline e-mail networking
- sysadmin work at first German *online bistro* later turning into first *internet cafe*
- volunteer sysadmin at volunteer-based non-profit ISP from 1994 onwards
- interest: packet filtering and IT security in general

2001: Jozsef Kadlecsik

Date: Fri, 7 Dec 2001 21:19:57 +1100 (EST)
From: James Morris <jmorris@intercode.com.au>
Subject: [netfilter-announce] [ANNOUNCE] New Core Team Member - Jozsef Kadlecsik

The Netfilter Core Team is proud to announce the addition Jozsef Kadlecsik as a new member.

Jozsef joins us as a dedicated and talented member of the Netfilter development community.

His demonstrated insight and high coding standards will be highly valuable assets to the project as development focus shifts to the 2.5 kernel series.

Welcome Jozsef!

- James, on behalf of the Netfilter Core Team.

--

James Morris <jmorris@intercode.com.au>

2001: Jozsef Kadlecsik



Jozsef joins netfilter core team in December 2001

- Physicist at Hungarian Physics Research Institute KFKI
 - does lots of sysadmin work there, including firewalling
 - btw: what's it with physicists and Linux networking, just like Alexey Kuznetsov?
- focus on connection tracking (he added TCP window tracking)
- still active in the project ever since (longest standing core team member)

2001: Jozsef Kadlecsik

Prior to Jozsef joining, but note-worthy:

- Kernel 2.4.0 is released in January 2001 (with netfilter/iptables)

Meanwhile in December 2001:

- Enron files for Chapter 11 bankruptcy
- UN authorizes ISAF in Afghanistan (post 9'11 attacks)
- President Karzai is selected to lead Afghan Interim Administration

Documentation

One key aspect was lots of good, easy to read documentation

- netfilter hacking HOWTO
- netfilter extensions HOWTO
- Linux 2.4 Packet Filtering HOWTO
- Linux Networking-concepts HOWTO
- NAT HOWTO

Getting into the project as both a user or developer was helped enormously by the HOWTOs.

The original versions of those documents were all created in early 2000.

The netfilter scoreboard

- a *scoreboard* was established
 - high-score for number of patches/contributions
 - counts not only code but also documentation updates
 - manually maintained by scoreboard
 - bonus points for patches that apply to correct version
- motivation for developers, particularly junior ones!

Remember, this was the pre-git and even pre-bitkeeper days!

Guess these days, people would count this as *gamification*?

The netfilter scoreboard (April 2002)

Contributors

Top 20 Non-Core-Team Netfilter Hackers			
Kis-Szabo Andras	88	Fabrice Marie	63
Henrik Nordstrom	22	Martin Josefsson	15
Jan Rekorajski	11	Andreas Herrmann	11
Bob Hockney	10	Gianni Tedesco	9
Brad Chapman	9	Andries van Schie	9
David Miller	7	Emmanuel Roger	7
Willy Tarreau	6	Patrick Schaaf	6
Philip Blundell	6	Guillaume Morin	6
Imran Patel	6	Magnus Boden	6
Marco Masetti	6	Fernando Anton	6

The netfilter scoreboard (April 2002)

Contributions of Kis-Szabo Andras

Mon Jul 08 00:00:00 2002

- 10 points for 2 x libip6t_tcp.c bugfixes
- 1 points for patch in correct form

Fri Oct 12 13:22:00 2001

- 5 points for Security fix for IPv6 mac matching.
- 1 points for patch in correct form

Tue Dec 18 00:00:00 2001

- 15 points for IPv6 tools updates (x3)
- 1 points for Patch in correct form
- 0 points for

Mon Mar 25 00:00:00 2002

- 5 points for tcpreplay6 tool for testsuite
- 1 points for patch correct form

modularity / extensibility

- netfilter is just a set of hooks for call-back functions
- iptables matches and targets are just plug-ins for both kernel and userspace
- good documentation on the APIs and how to write one
- get people involved, implement their favorite feature

Problem: How to distribute / maintain them?

patch-o-matic

1) The netfilter core team is maintaining a set of extensions / new features which are not yet committed to the mainstream kernel tree.

They are a collection of maybe-broken maybe-cool third-party extensions.

Please note that you cannot apply any combination of any of those patches. Some of them are incompatible....

If you want to try some extensions, and be sure that they don't break each other, you can do the following:

```
% ./runme base KERNEL_DIR=<<where-you-built-your-kernel>>
```

It will modify your kernel source (so back it up first!). You will have to recompile / rebuild your kernel and modules.

Alternatively, if you really know what you are doing, you can use the following command in order to offer you the full list of choices. Be aware that we don't prevent you from shooting yourself in the foot.

```
% ./runme extra KERNEL_DIR=<<where-you-built-your-kernel>>
```

patch-o-matic

Date: Mon, 30 Oct 2000 14:28:30 +1100
From: Rusty Russell <rusty@linuxcare.com.au>
To: Netfilter Development Mailinglist <netfilter-devel@us4.samba.org>

On Mon, Oct 23, 2000 at 09:15:23PM -1100, Daniel Stone wrote:
> This, to me, reflects a problem. Basically, I can only see two
> things causing this:
> a) no testing at all, or
> b) a mis-paste. Please tell me it was the latter.

Completely untested. I looked at the patch as I threw it into patch-o-matic.

That's what patch-o-matic is for: to get stuff out there without waiting for the Rusty Linus planet alignment thing...

Rusty.

--

Hacking time.

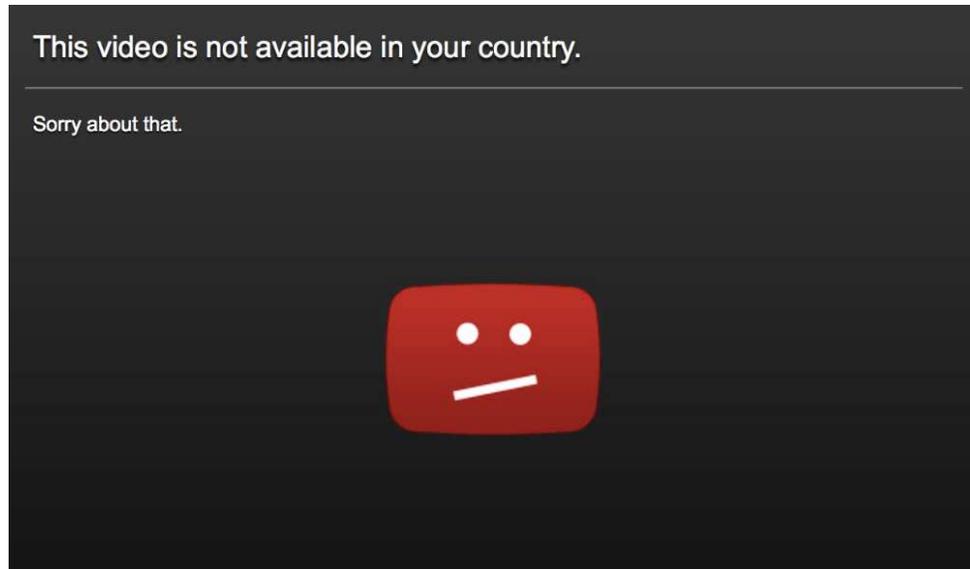
Early Success

What contributed to the early success with lots of developers writing netfilter/iptables code:

- loads of good documentation
- modular framework with
 - netfilter hooks
 - pluggable iptables targets + matches
- system for maintaining non-mainline code + merging it

⇒ Everyone could easily write his favorite match/target/plugin

2003: Martin Josefsson



Martin Josefsson joins core team in August 2003

- mainly optimizations e.g. on connection tracking hash tables
- he worked at large ISP where that performance actually mattered

Historical Context:

- kernel 2.6.x released in December 2003
- DPRK withdraws from nuclear non-proliferation treaty
- The US space shuttle Columbia crashes
- US launches war on Iraq; Saddam Hussein is captured

2003: Pablos first messages

Date: Wed, 12 Nov 2003 00:06:11 +0100
From: pablo neira <pablo@eurodev.net>
To: netfilter-devel@lists.netfilter.org
Subject: ip_contrack_get

Hi everyone,

I've been for almost two weeks trying to understand netfilter code, at this point I'm trying to understand contrack table code.

I have a problem with how contrack manages the ip_contrack_info stuff.

[...]

Don't blame me if it's obvious for you, I'm just a guy trying to understand a **really really nice piece of code**. Thanks!

cheers,
Pablo

Pablos first messages

Date: Wed, 03 Dec 2003 15:23:39 +0100
From: pablo neira <pablo@eurodev.net>
To: netfilter-devel@lists.netfilter.org
Subject: sending event to user space

Hi list!

I programmed a dummy module for netfilter which tries to match a packet and if it does, it will do nothing (NF_ACCEPT) but I would want it to send an event to a program in user space to do something, how can I do that?

So something like:

- a) packet gets my hook and do NF_ACCEPT.
- b) modules sends an event to user space.
- c) program in user space does something.

Thanks!
Pablo

P.S: thanks for this *great piece of code*!!

2004: Patrick McHardy



Patrick joins core team in January 2004

- lots of good work in many areas; moved beyond netfilter and even entered the iproute2/tc lands ;)
- most recently suspended from core team due to questionable practises in copyright/license enforcement

Historical Context:

- Facebook was born
- Bluetooth 2.0 EDR spec released
- Skype becomes really popular

core team emeritus members

Date: Fri, 09 Jan 2004 15:17:19 +1100
From: Rusty Russell <rusty@rustcorp.com.au>
Subject: [ANNOUNCE] Core Team Announces Emeritus Members

The Netfilter Core Team has long discussed the issue of Core Team members who are no longer active. Dismissing them from the Core Team would deny them the benefits of such a prestigious title, should any become apparent.

Hence the conclusion is that Marc Boucher, James Morris and Rusty Russell are now "emeritus"[1] members of the Netfilter Core Team.

[...]

[1] Latin for "burnt-out freeriding slacker", I believe.

the story behind Rustys departure

- Until recently, I thought
 - Rusty simply had too many other tempting distracting projects (kernel module loader, qemu, paravirtualization, ...)
- Recently, Rusty told me
 - it was a deliberate decision to leave netfilter
 - the new core team and maintainers should run the project without interference from the project father
 - kids have to stand on their own feet

2005: Yasuyuki Kozakai joins core team

Yasuyuki Kozakai joins netfilter core team

- member of Japanese USAGI project for Linux IPv6
- Main contribution: IPv6 connection tracking, including
 - nf_conntrack generalization
 - conntrack extensions
 - nf_nat for IPv6

2007: Pablo Neira Ayuso joins core team

Date: Thu, 15 Feb 2007 14:02:03 +0900 (JST)
From: Yasuyuki KOZAKAI <yasuyuki.kozakai@toshiba.co.jp>
Subject: [ANNOUNCE]: New Coreteam Member Pablo Neira Ayuso

The Netfilter Core Team is proud to announce the addition of Pablo Neira Ayuso as a new member.

He has repeatedly demonstrated high insight and coding standards, and has already been responsible for several parts of the codebase, especially ctnetlink, conntrack and conntrackd.

By joining the Core Team, Pablo will definitely help advance the development of the Netfilter project to a higher level.

Welcome Pablo!

Yasuyuki,
on behalf of the Netfilter Core Team.

Pablo (2009)



- initially known for work on ctnetlink and contrackd
- later known for a jack of all [netfilter] trades
- official head of core team since 2013, already more or less de-facto before

Harald (2009)

As I'm showing various old pictures of other people, for fairness' sake...

Historical Context:

- President Obama is inaugurated
- Conficker virus infects 9.5 million PCs
- Michael Jackson died
- Google starts ChromeOS

nftables (2009)

Date: Wed, 18 Mar 2009 05:29:42 +0100
From: Patrick McHardy <kaber@trash.net>
To: Netfilter Development Mailinglist <netfilter-devel@vger.kernel.org>
CC: Linux Netdev List <netdev@vger.kernel.org>
Subject: [ANNOUNCE]: First release of nftables

Finally, with a lot of delay, I've just released the first full public version of my nftables code (including userspace), which is intended to become a successor to iptables. Its written from scratch and there are numerous differences to iptables in both features and design, so I'll start with a brief overview.

There are three main components:

- the kernel implementation
- libnl netlink communication
- nftables userspace frontend

2012: Eric Leblond and Florian Westphal

In October 2012, Eric Leblond and Florian Westphal join core team

- Eric: nf_nat port randomization, lots of nfnetlink* fixes, later also nftables
- Florian: NFQUEUE load balancing, NFQUEUE fixes and improvements, later pretty much every area

Also in October 2012: Harald, Martin and Yasuyuki finally enter emeritus state

Historical Context:

- US begins retaliation action against embassy attack in Libya
- Turkey retaliates against Syria
- Windows 8 makes its debut
- Great Patent war Apple vs. Samsung
- Megaupload gets shut down

nftables (2013)

Pablo picked up a lot of the loose ends left by Patrick after some time and in 2013, nftables finally goes mainline!

```
commit 96518518cc417bb0a8c80b9fb736202e28acdf96
Author: Patrick McHardy <kaber@trash.net>
Date:   Mon Oct 14 11:00:02 2013 +0200
```



bugs

Date: Mon, 14 Jan 2002 22:33:16 +1100
From: Rusty Russell <rusty@rustcorp.com.au>
To: "David S. Miller" <davem@redhat.com>
Subject: Re: Kernel 2.4.16: NetFilter Bug Report

In message <20020113.231425.73653921.davem@redhat.com> you write:
> Any ideas on that NAT timer one from cat@zip.com.au? I've for now
asked Marcelo to revert that 2.4.17 change until a different fix is
obtained.

I am a fucking retard.

I was looking at what was wrong with the code, and came up with **five**
separate problems. I know the compat layer was a hack, but what the
fuck was I doing?

Read and weep,
Rusty.

--

Anyone who quotes me in their sig is an idiot. -- Rusty Russell.

nfsim / testsuite

Big problem with lots of code, including netfilter: Lack of automatic testing.

Rusty returns to netfilter with *nfsim*, a netfilter simulator, co-authored with Jeremy Kerr.

- nfsim runs netfilter kernel code in userspace against test suite
- emulates kernel environment in userspace
- imports netfilter kernel code + builds it in userspace
- {get,set}sockopt() wrapper for userspace tools
- can simulate allocation failures
- manual control over time (important for conntrack state tables)

⇒ Great Idea, and lots of useful work

nfsim / testsuite

Reality sucks:

- very few contributions
- very few users beyond Rusty + Jeremy
- very limited adoption/use by netfilter developers

nfsim / testsuite

Date: Tue, 31 May 2005 23:48:24 +1000
From: Rusty Russell <rusty@rustcorp.com.au>
To: Patrick McHardy <kaber@trash.net>

On Tue, 2005-05-31 at 15:02 +0200, Patrick McHardy wrote:
> Second of all, I spent like 10 hours to verify the proposed fixes,
and I am still convinced that it is correct.

Which shows exactly **why** we have a testsuite. Dammit, I didn't spend all those hours on it for fun.

You spent **10** hours, and the testsuite runs in 5 seconds (60 seconds counting build time the first time).

<sigh>

nfsim / testsuite

Reality sucks:

- patches get validated only in test suite, not real kernel

nfsim / testsuite

Date: Sun, 23 Jan 2005 20:15:17 -0800
From: "David S. Miller" <davem@davemloft.net>
Subject: Re: [PATCH 3/2] Fix compile with NAT but without modules

On Mon, 24 Jan 2005 01:33:47 +0100 Patrick McHardy <kaber@trash.net>
wrote:

> I'll apply your patches and push them to Dave tomorrow, bkbkits.com
> is unreachable currently so I can't resync my tree.

To be frank, this is one of several severe fallouts from Rusty's patches. I really think they were not ready for submission when he sent them to me. It even broke the build if you had modules enabled in any way.

I'm only mentioning this because it appears that nfsim is becoming partially a crutch, because I know this is what Rusty and others use heavily for testing. Which is fine, but if your patches break the build in many ways in the real kernel tree you're relying too heavily on the userland simulator IMHO.

nfsim / testsuite

Reality sucks:

- very few contributions
- very limited adoption/use by netfilter developers
- bit-rot of kernel environment simulation
- constant lag in terms of completeness
 - no netlink simulation, i.e. no nfnetlink/ctnetlink/nf_queue/nf_log

⇒ no replacement / successor, till today :(

nfsim / testuite

So all we can do is join DaveM and pray for code correctness



humor

Date: Wed, 26 Feb 2003 11:54:59 +1100
From: Rusty Russell <rusty@rustcorp.com.au>
To: Jozsef Kadlecsik <kadlec@blackhole.kfki.hu>
Subject: [netfilter-core] Re: conntrack patches

> Hi Rusty,

>

> Last year I started to go through all of your unpublished conntrack related patches. [...]

> Are you working on the patches or plan to finish them? Or can I go back and complete the half-done job on the patches?

Jozsef,

Let me put it this way: take over those patches and I will name my first child after you.

How many beers did I owe you now?

Rusty.

Harald and IPv6 NAT

Date: Thu, 20 Nov 2003 14:40:42 +0100
From: Harald Welte <laforge@netfilter.org>
Cc: netfilter-devel@lists.netfilter.org
Subject: Re: NAT for IPv6

On Wed, Nov 19, 2003 at 01:38:47PM +0100, Maciej Soltysiak wrote:
> out of curiosity - are there plans to incorporate NAT into
ip6tables or future pkttables ?

over my dead body. NAT is what broke ipv4 end-to-end. Let's not do
the same with ipv6.

The only reasonable application is ipv4-to-ipv6 transition-nat.

--

- Harald Welte <laforge@netfilter.org>
<http://www.netfilter.org/>

=====
"Fragmentation is like classful addressing -- an interesting early
architectural error that shows how much experimentation was going
on while IP was being designed." -- Paul Vixie

netfilter Workshops

- 1998/1999/2000: Informal meetings of some of the people involved
 - like James + Marc + Rusty at Sydney Linux Expo
 - like Harald + Rusty at Linux Beer Hike
- *workshop* established from 2001 onwards to get developers meet up
- not every year, but almost: 13 workshops in 18 years
- invitation-only
- organization done by community for community
- sponsors typically among commercial netfilter users

netfilter Workshops

- 2001: Enschede, Netherlands
- 2003: Budapest, Hungary
- 2004: Erlangen, Germany
- 2005: Seville, Spain
- 2007: Karlsruhe, Germany
- 2008: Paris, France
- 2010: Seville, Spain
- 2011: Freiburg im Breisgau, Germany
- 2013: Copenhagen, Denmark
- 2014: Montpellier, France
- 2015: Budapest, Hungary
- 2016: Amsterdam, Netherlands
- 2017: Faro, Portugal

Workshop 2003: Group Picture



Workshop 2005: Fun fact



Workshop 2013



- 2011: Kernel 3.0 is released, with netfilter/iptables

Workshop 2014



Workshop 2015



- Kernel 4.0 is released, with netfilter/iptables and nftables

Workshop 2016



Interesting Challenges

- iptables kernel code used to never verify ruleset integrity
 - you could crash kernel using malicious ruleset
 - believed to be non-issue due to NET_CAP_ADMIN requirement
 - assumption broke horribly when unprivileged containers appeared



netfilter.org infrastructure

- self-hosted physical servers for web/svn/bugzilla/git (and even lists) for long time
 - lists moved to vger.kernel.org eventually
- firewall machine in front of netfilter.org for many years: iptables on UltraSPARC
 - because we can, and because script kiddies don't do SPARC assembly
- netfilter.org servers for many years Linux on PPC (G5 Clusternode)
 - because we can, and because script kiddies don't do PPC assembly

Summary: Why successful?

- smart people got funded to implement things the way they want
- extensible architecture from day one
- good documentation for developers and users from day one
- passionate developers who picked netfilter as their own topic of interest

Regrets?

- not having time for netfilter work anymore :/
- not officially stepping down sooner, giving Pablo + Patrick more credit
- conntrack/nat helpers are still in kernel space
- people think they need dynamic IPv6-to-IPv6 NA(P)T
- nftsim without replacement; netfilter kernel code remains largely without tests
- with the size and relevance of the Linux industry in 2017, why don't people invest in automatic test suites for netfilter (and other kernel networking code)?
- not having pushed for more ulogd adoption. Lots of people still use LOG, 17 years after ULOG and ulogd

Thanks

- to the audience, for bearing with me
- to the netdev 2.2 committee, for inviting me
- to Rusty, for being my hero
- to Pablo, for picking up the pieces when I left
- to Dave, for being everyone's hero
- to Jesper, for group (and other) pictures
- to every single netfilter contributor out there

EOF

End of File.

No packets were harmed in the making of this presentation.