cumulus networks™

Breaking Open Linux Switching Drivers

Andy Gospodarek

Cumulus Networks

16 Feb 2015

netdev01

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Agenda

- Why
- History
- Design proposal
- Future possibilities

Why am I here?

Linux kernel should enable others to create the next generation of forwarding devices

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Integrate support for offload hardware directly into the the Linux kernel

Hardware Platform History

Market dominated by switch and router vendors providing expensive proprietary solutions

Proprietary software running on switches and routers was not open for developers and users to enhance

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Today's hardware platforms are significantly higher-performance and more generally available

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Spare CPU cycles are available for applications to run directly on the switch

Bare-metal platforms are now appealing and available to commercial Linux vendors, developers, and users

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Software History

15+ years with Linux as a viable OS for host processor on switches/routers

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

10+ years Linux “support” by ASIC vendors

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

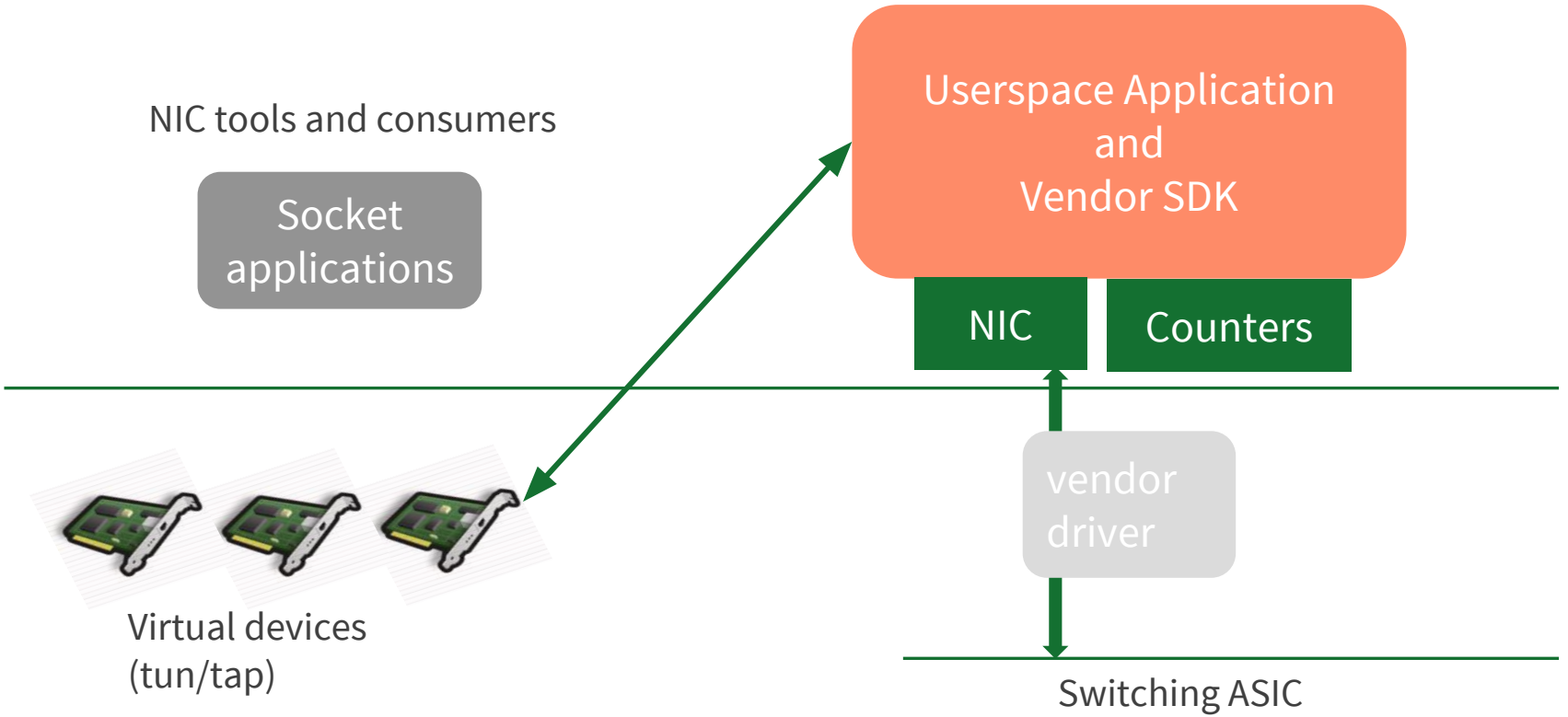
Basic in-kernel switching/offload layer support in v3.19

Software architecture to control ASICs has not fundamentally changed in the last decade

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

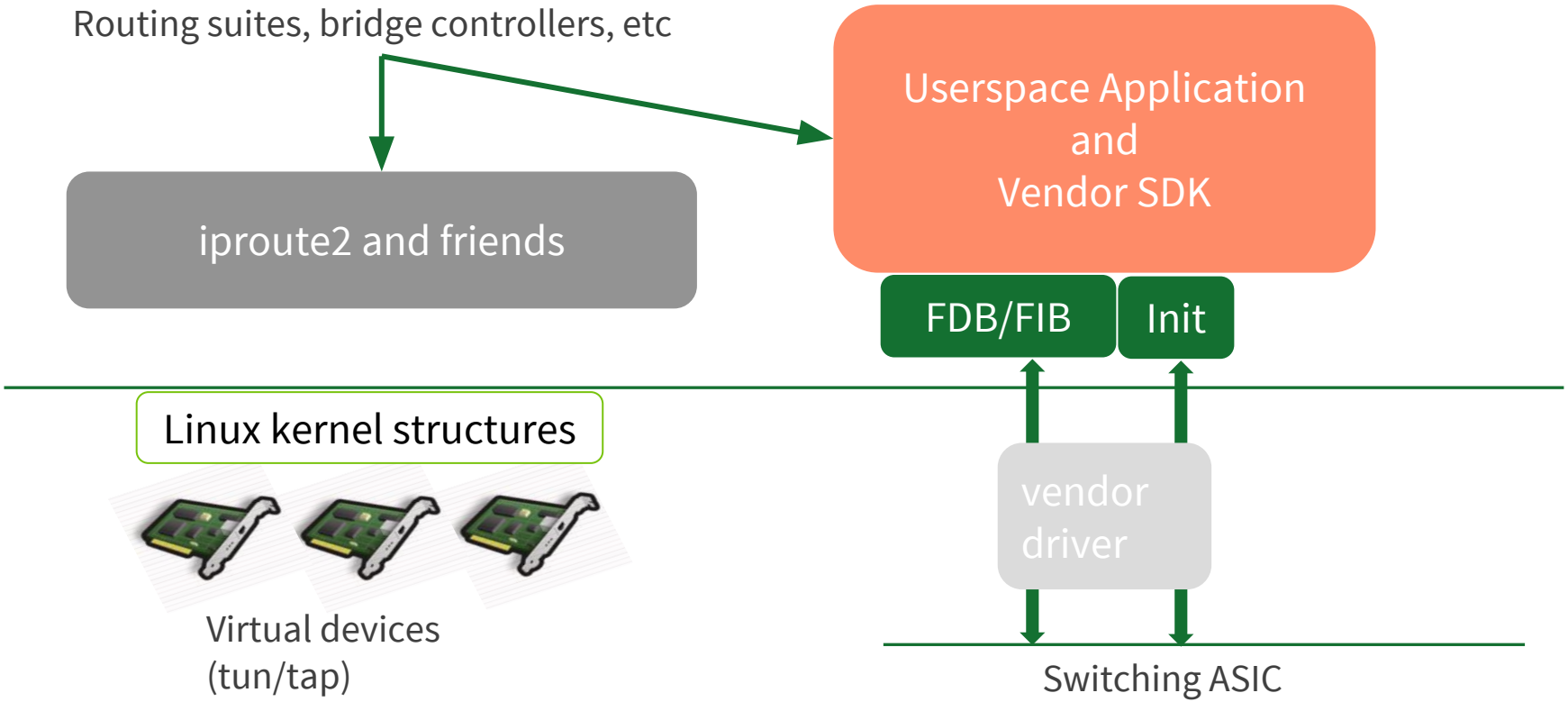
What exactly does that look like?

Typical packet path



Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Control Plane Programming

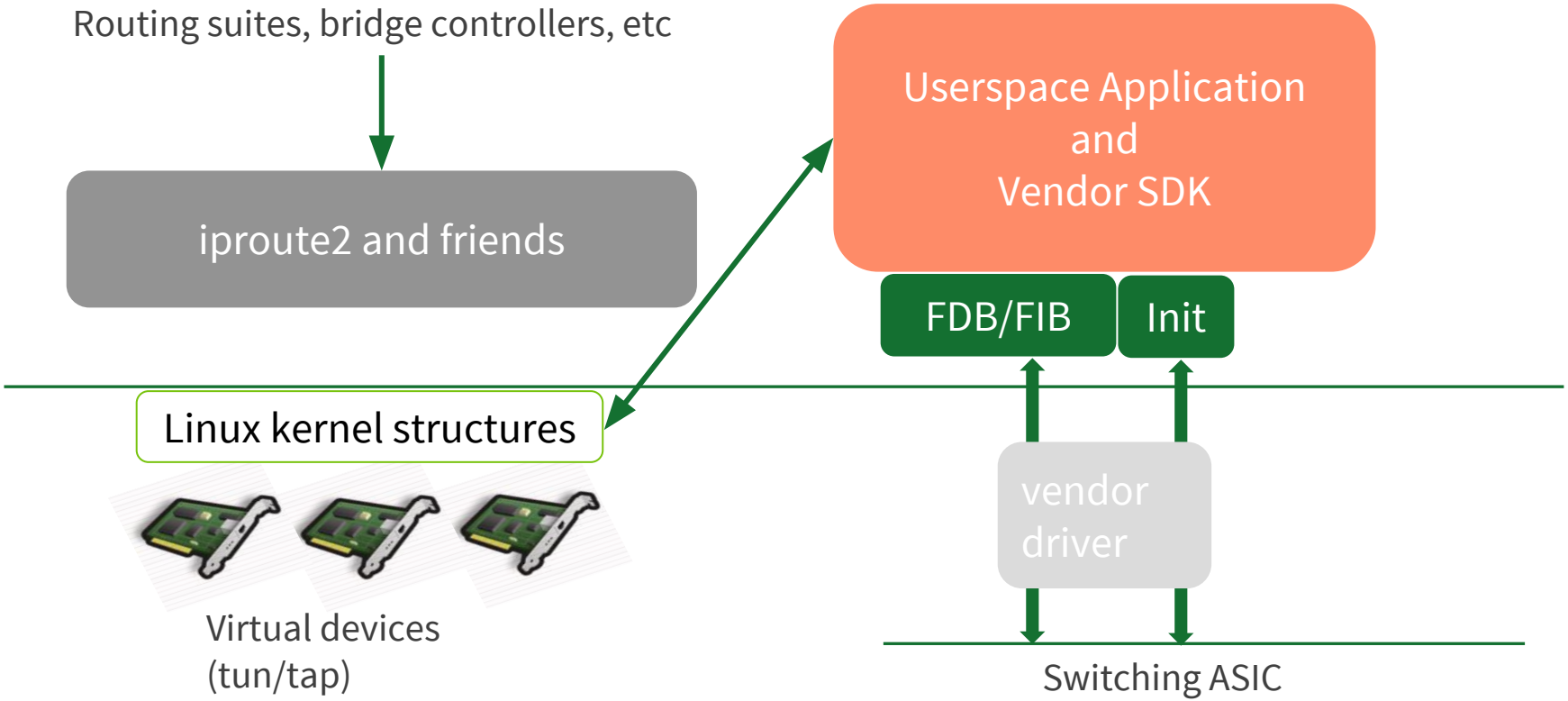


Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Painful for those developing switches as management applications need to talk to kernel/netlink and SDKs

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Netlink Control Path



Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Much better design, but each SDK supported needs a new translation between netlink and SDK

Kernel hackers and distribution vendors see a simple solution

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Get rid of all closed-source SDKs

Great idea!

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Vendors do not want to open source their SDKs

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Can we use a userspace SDK and a kernel driver at the same time?

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Not if you want it upstream!

OK...how do I get started?

Phased Approach

- Participate!
- Pick a hardware platform
- Write and post a switchdev-compatible network driver
- Enhance that driver to add ndo/offload_ops to driver

Attend conferences, participate on mailing-lists, and post patches

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Write and post a switchdev-compatible network driver

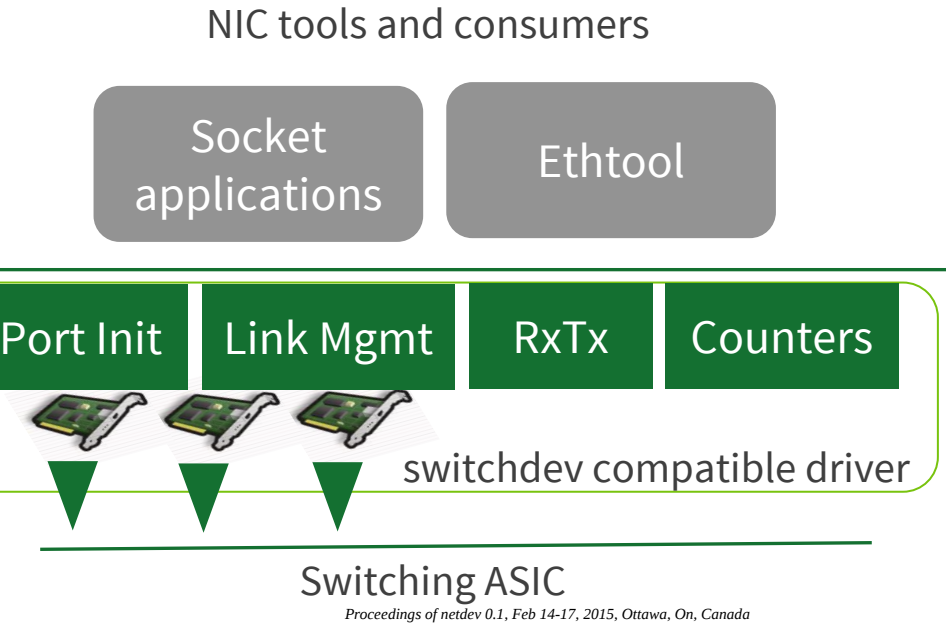
Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Advantages

- Provide network access via front panel ports
- Phased approach to working upstream
- Applications can developed without need for hardware offload

What might that look like?

Base Switchdev Driver



Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Great, we are upstream!

Are we done?

No!

Add offload support to driver as upstream infrastructure is developed

What might *that* look like?

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

Switchdev Offload Driver

NIC tools, Routing suites, bridge controllers, etc

Socket applications

Ethtool

iproute2 and friends

Port Init | Link Mgmt | Counters | RxTx | Offload Ops

switchdev compatible driver

Switching ASIC

Proceedings of netdev 0.1, Feb 14-17, 2015, Ottawa, On, Canada

“If you are the first you will be so cool.”

-DaveM

Get coding

“..and we’ll help you maintain it”

-DaveM

cumulus networks

Thank You!

© 2013 Cumulus Networks. Cumulus Networks, the Cumulus Networks Logo, and Cumulus Linux are trademarks or registered trademarks of Cumulus Networks, Inc. or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners. The registered trademark Linux® is used pursuant to a sublicense from LMI, the exclusive licensee of Linus Torvalds, owner of the mark on a world-wide basis.