

# Next Generation of Programmable Datapath



**OVS Conference, San Jose, 2016**  
**Thomas Graf**

**What happened in  
2016 so far?**

# OVS 2.6 was released



**Justin Pettit**

@Justin\_D\_Pettit



Following

We just released OVS 2.6 with non-experimental support for OVN and a number of other nice additions! [#openvswitch](#)  
[openvswitch.org/pipermail/anno ...](http://openvswitch.org/pipermail/anno...)

RETWEETS

21

LIKES

17



9:03 AM - 28 Sep 2016



21



17



# First ever OVN release



**Ben Pfaff**

@Ben\_Pfaff



Following

In honor of the first release of OVN, the speaker gifts for #ovscon are actual ovens. Speakers, please email preference for gas or electric.

RETWEETS

9

LIKES

24



8:20 PM - 16 Oct 2016



# OVS Orbit podcast (May 16)



**Ben Pfaff**

@Ben\_Pfaff



Following

The Open vSwitch podcast that I've been working on is in "beta". Please report bugs. 2 more episodes to edit. [ovsorbit.benpfaff.org](http://ovsorbit.benpfaff.org)

RETWEETS

4

LIKES

11



8:36 AM - 1 May 2016



***Not subscribed yet? [ovsorbit.org](http://ovsorbit.org)***

# Docker followed (Aug 16)



**Docker** ✓  
@docker



Following

Announcing the Official [@docker #Podcast](#) -  
[#Dockercast!](#) Hosted by [@botchagalupe!](#)



## Introducing Dockercast – the Docker Podcast

Today, we're thrilled to introduce Dockercast the official Docker Podcast. The Docker and container ecosystem is moving fast and it can be hard to catch up with the latest projects or features...

[blog.docker.com](http://blog.docker.com)

RETWEETS

35

LIKES

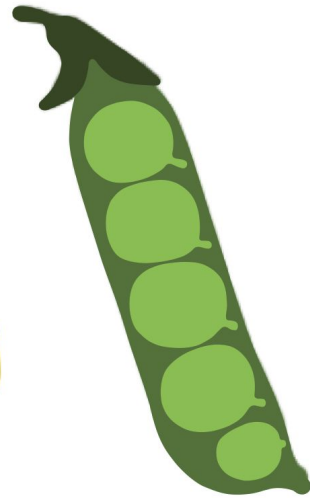
39



11:00 PM - 2 Aug 2016

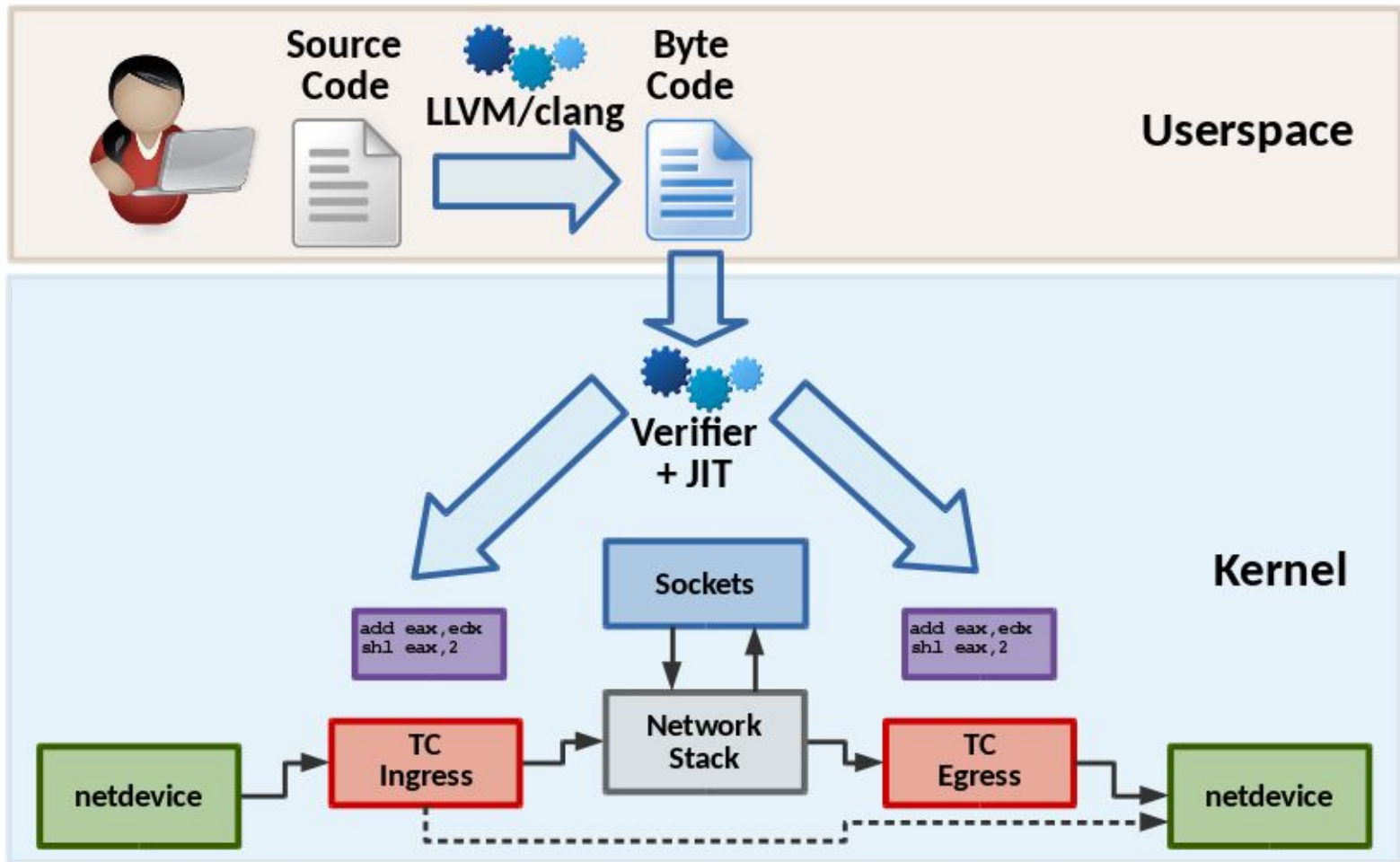


**What is up next?**

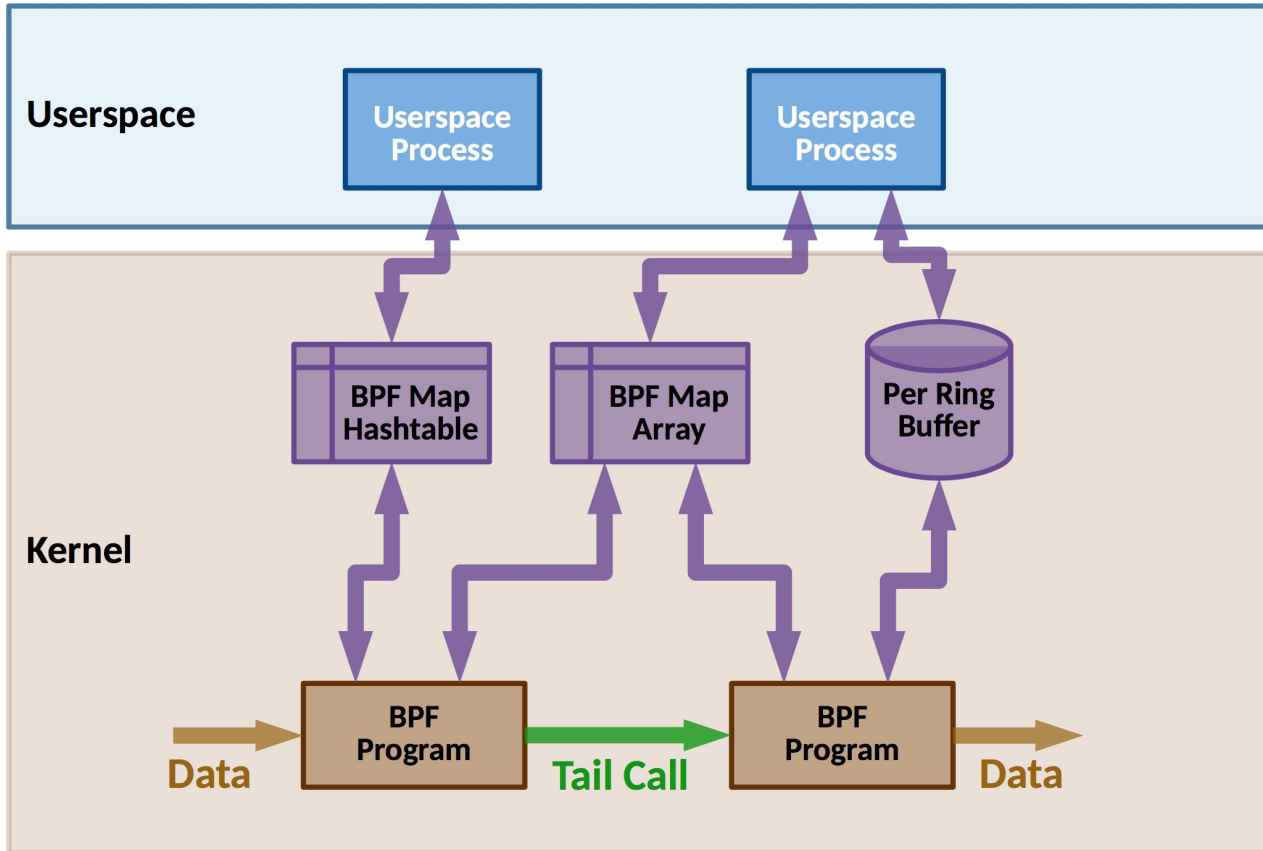


**Stickers  
Available!**





# Programs, Maps, Tail Calls



# BPF Helpers: Interacting with the outside world

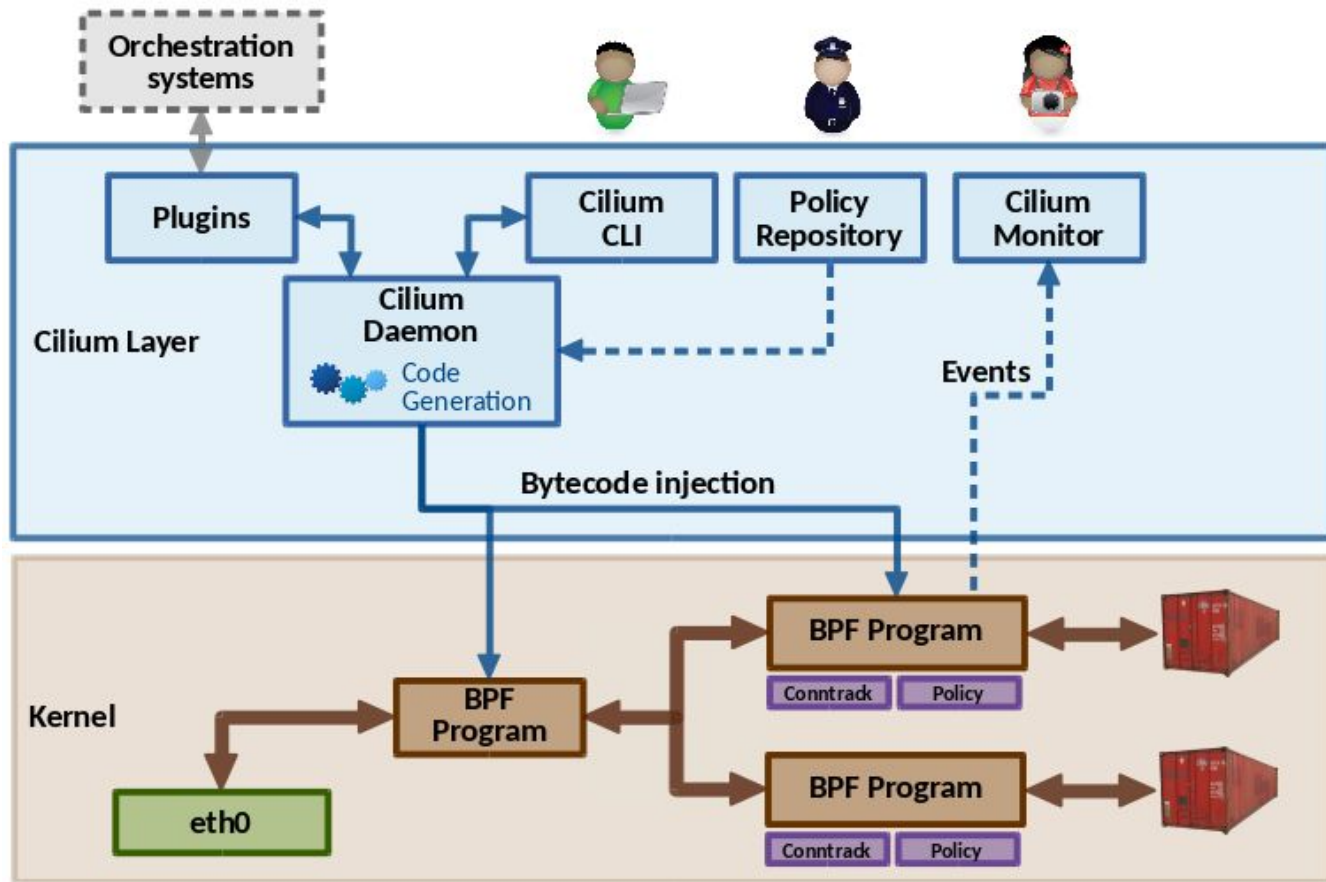
- Map Lookup/Update/Delete
- Get ktime
- printk to trace buffer
- Get random number
- Get SMP processor ID
- Load/store n bytes in skb data
- Replace L3/L4 checksum of skb
- Name/UID/GID of current process
- Push/pop VLAN header
- Set/get tunnel key and options
- Tail call
- Read/write perf event ring buffer
- Redirect/clone to other net\_device
- Get routing realm
- Calculate checksum diff over memory buffer
- Change protocol of skb
- Change type of skb (local/broadcast/...)
- Check for cgroup membership
- Access skb->hash or mark invalid
- Trim tail of skb
- Make skb linear



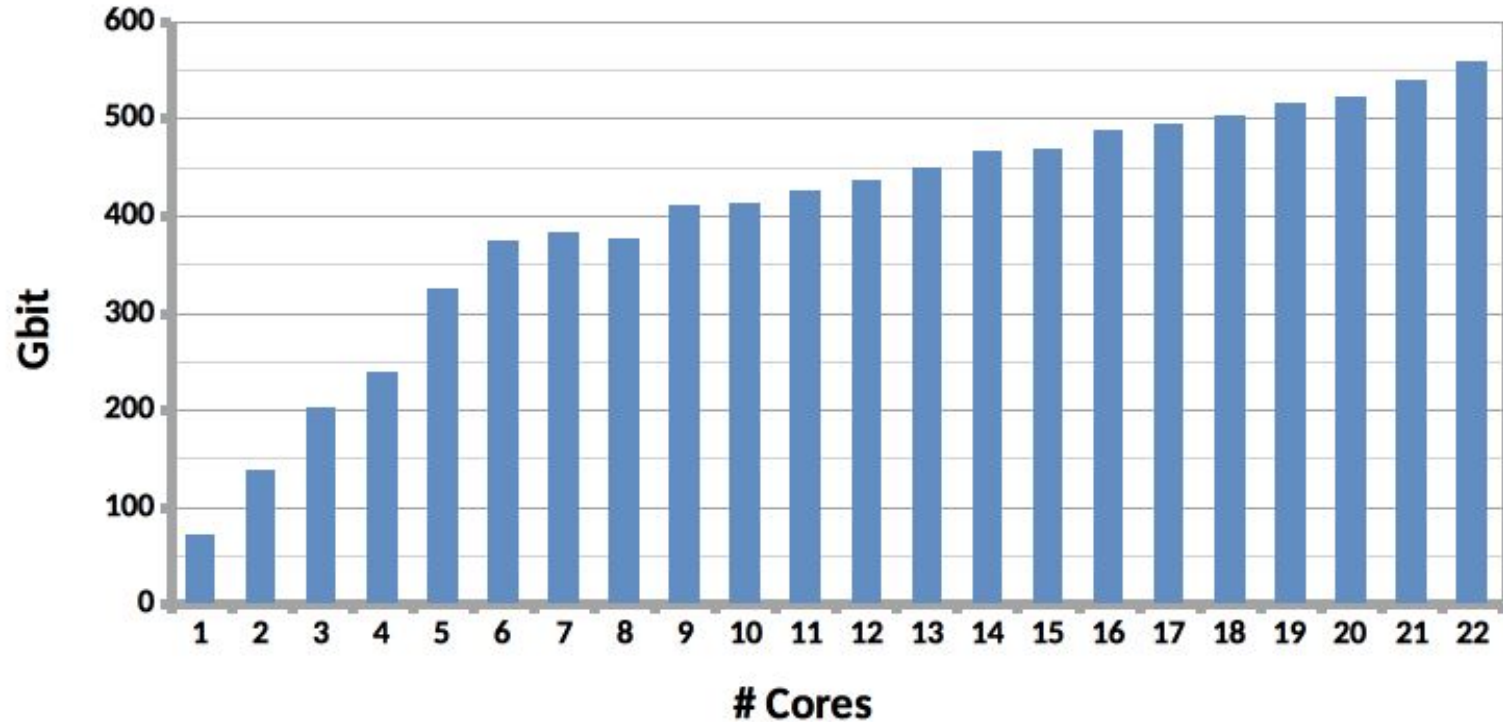
# Experimenting with BPF

- **Apply BPF to container networking & security**
- **Generate individual bytecode for each container at startup**
  - **Incredible flexibility and versatility**
  - **Majority of configuration becomes constant**
- **Decouple code (bytecode) and state (maps)**
  - **Allow for regeneration at any time without breaking connections**

# Cilium Architecture



## Container to container on local node



**Intel Xeon 3.5Ghz Sandy Bridge, 24 cores,  
1 TCP flow per core, netperf -t TCP\_SENDFILE, 10'000 policies**

# Learned Lessons

- **Datapath development on steroids.**
- **Verifier complexity limit requires to split programs into multiple blocks that inherit state.**
- **We have probably not added the last helper yet.**
- **Helper requirements define minimal kernel version, so it's pretty much kernel 4.8+ for anything non trivial.**

# Q&A

**Remember to get your BPF stickers!**

**E-Mail:**

**tgraf@tgraf.ch**

**Twitter**

**@tgraf\_\_**