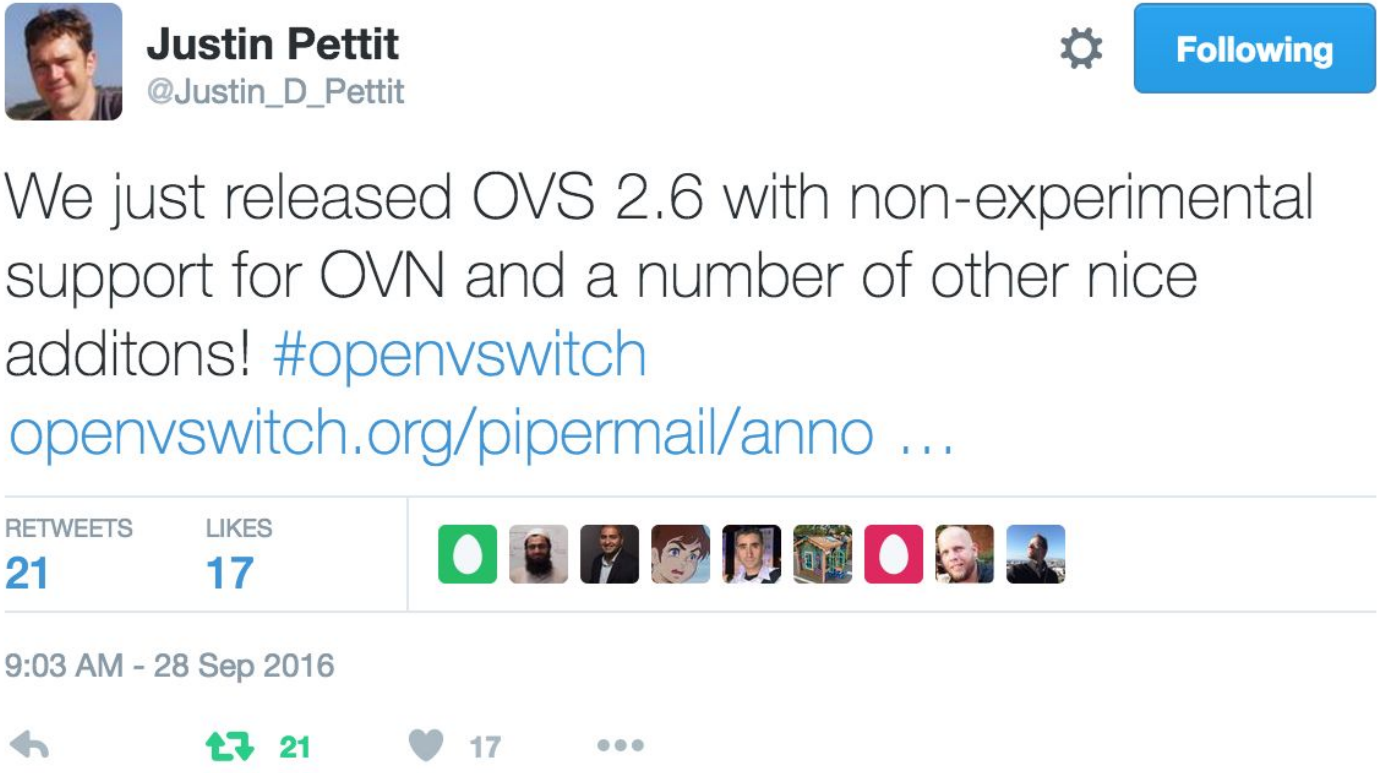
Next Generation of Programmable Datapath

OVS Conference, San Jose, 2016

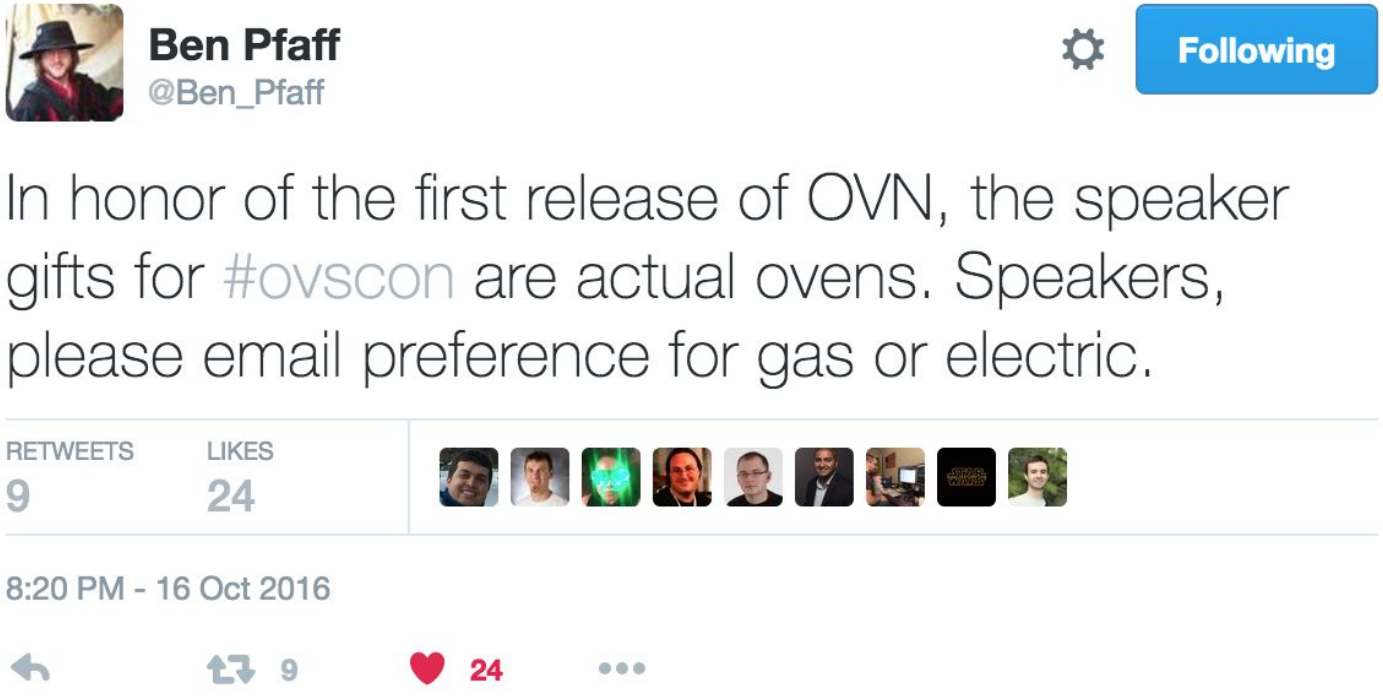
Thomas Graf

What happened in 2016 so far?

OVS 2.6 was released



First ever OVN release

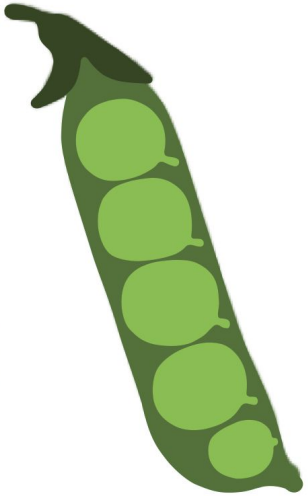


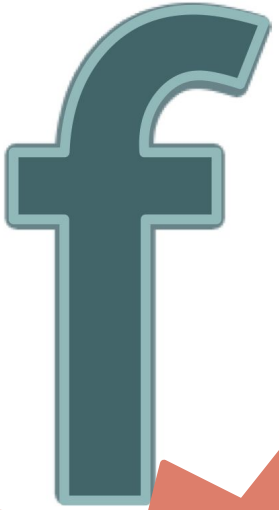
OVS Orbit podcast (May 16)

Not subscribed yet? ovsorbit.org

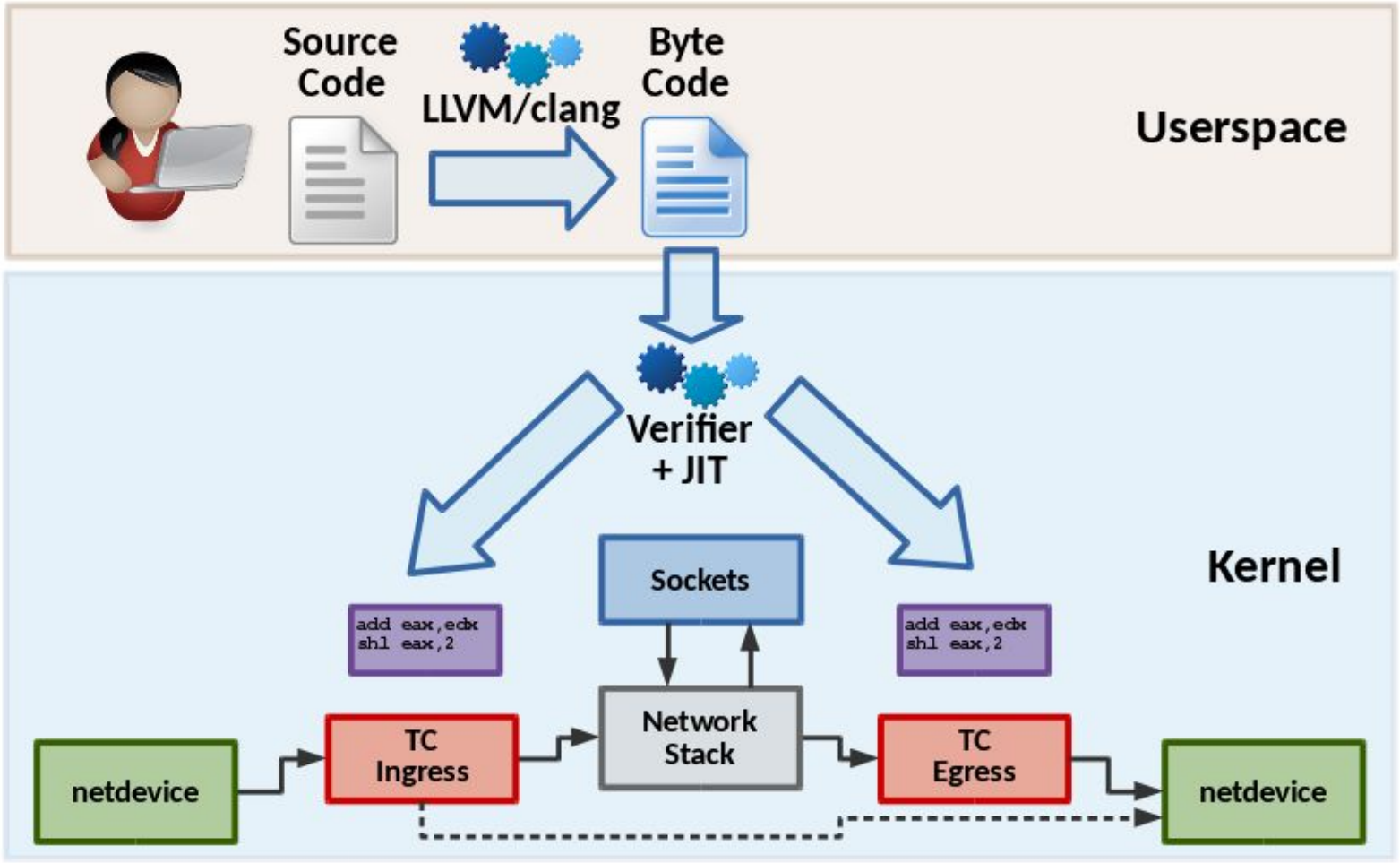
Docker followed (Aug 16)

What is up next?

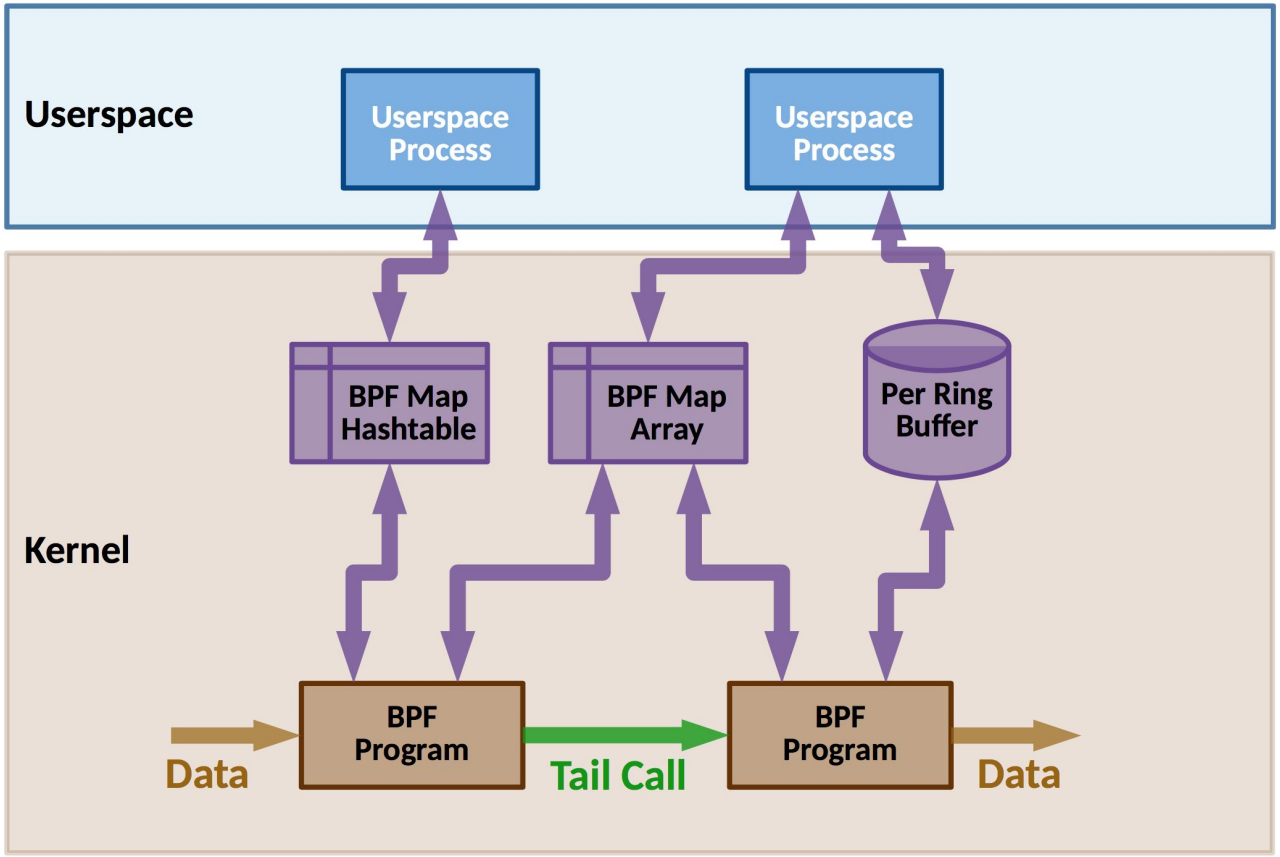




Stickers Available!


Source Code
LLVM/clang
Byte Code
Userspace
Verifier + JIT
Kernel
Sockets
add eax,ecx
shl eax,2
add eax,ecx
shl eax,2
netdevice
TC Ingress
Network Stack
TC Egress
netdevice

Programs, Maps, Tail Calls



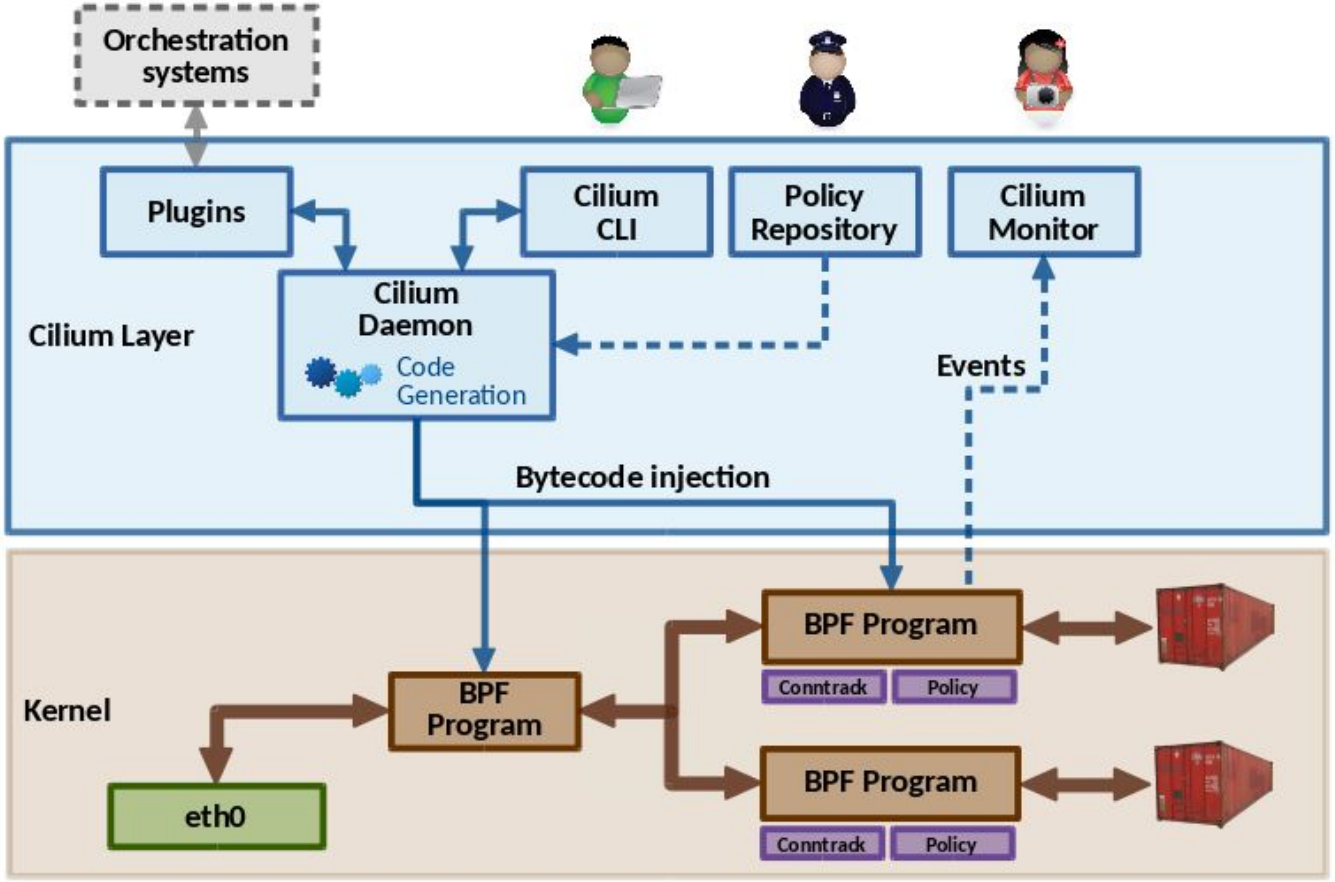
BPF Helpers: Interacting with the outside world

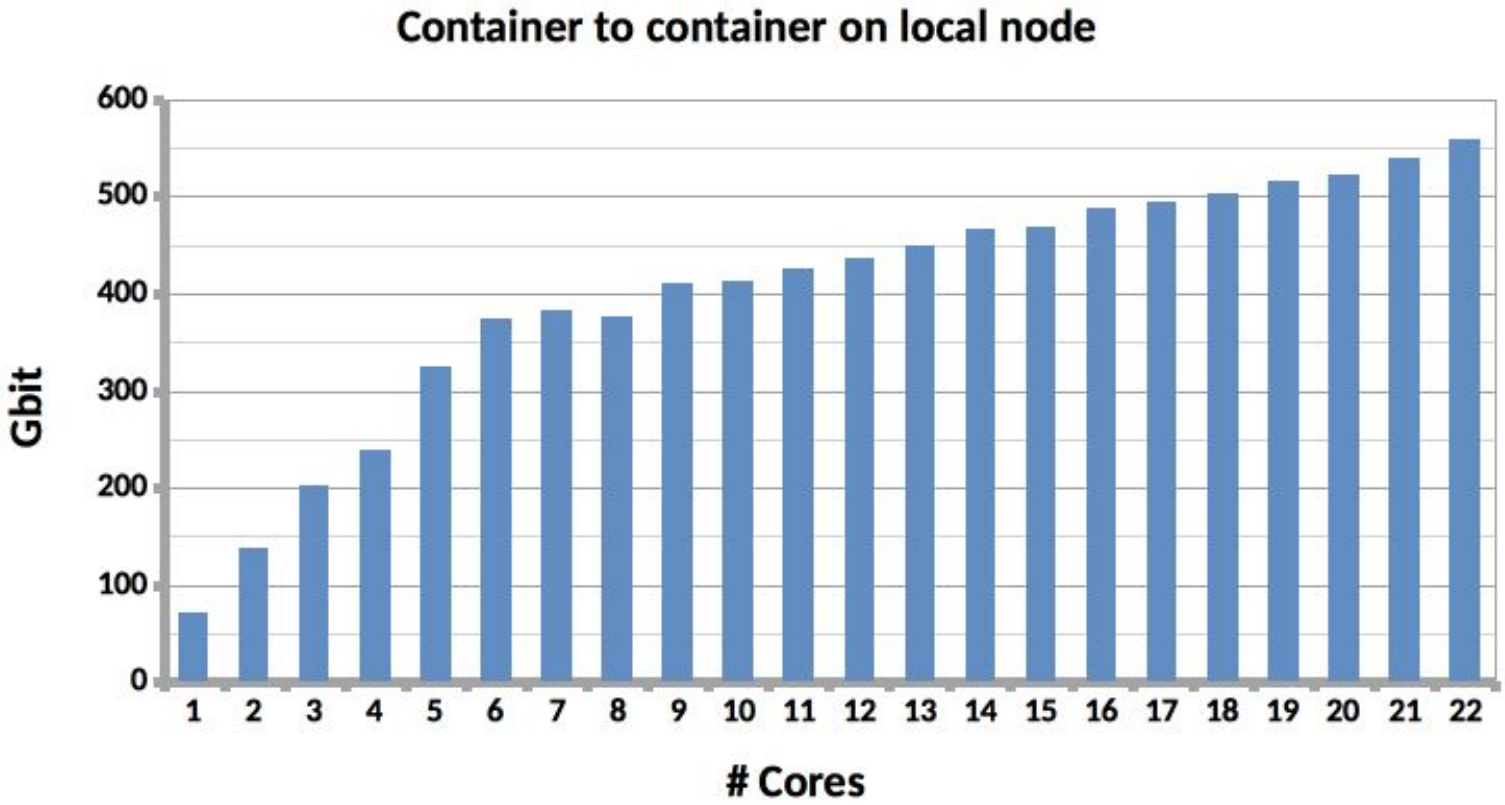
- Map Lookup/Update/Delete
- Get ktime
- printk to trace buffer
- Get random number
- Get SMP processor ID
- Load/store n bytes in skb data
- Replace L3/L4 checksum of skb
- Name/UID/GID of current process
- Push/pop VLAN header
- Set/get tunnel key and options
- Tail call
- Read/write perf event ring 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





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__