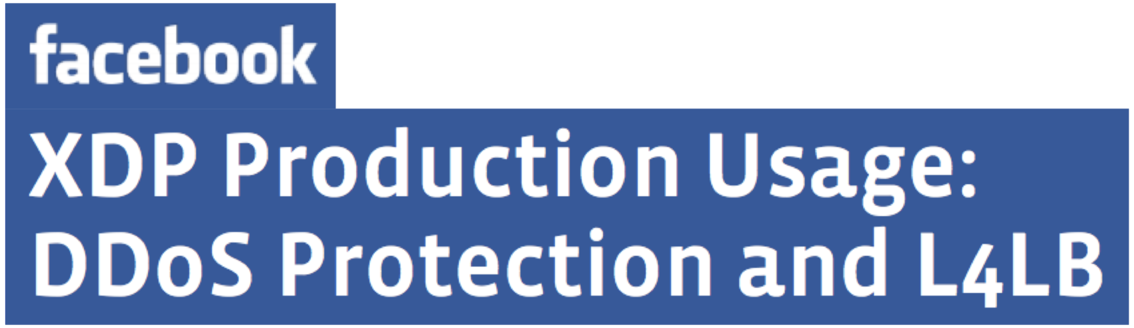
Cilium

Accelerating Envoy and Istio with Cilium and the Linux Kernel

Thomas Graf, Covalent

BPF - The *Superpowers* inside Linux

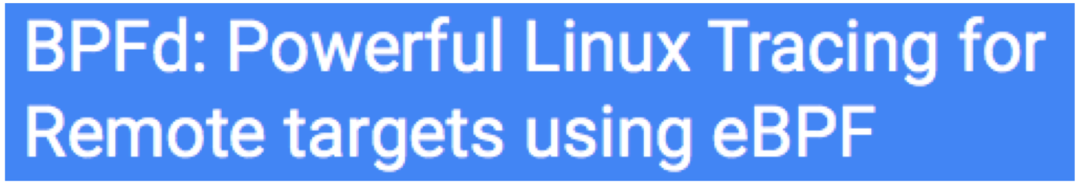
The Rise of BPF [and XDP]



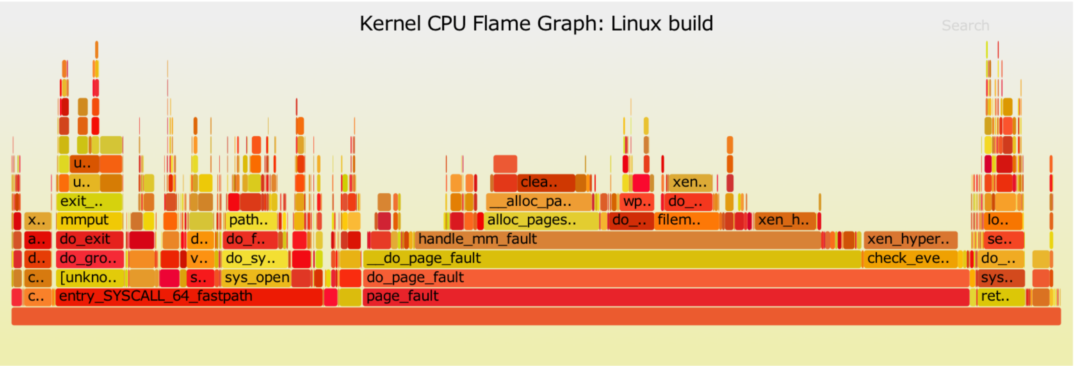
facebook

XDP Production Usage: DDoS Protection and L4LB

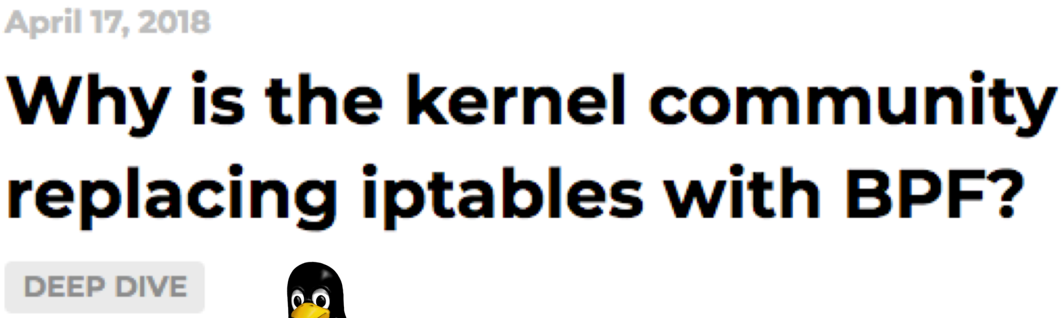
Google



BPFd: Powerful Linux Tracing for Remote targets using eBPF


Kernel CPU Flame Graph: Linux build

NETFLIX

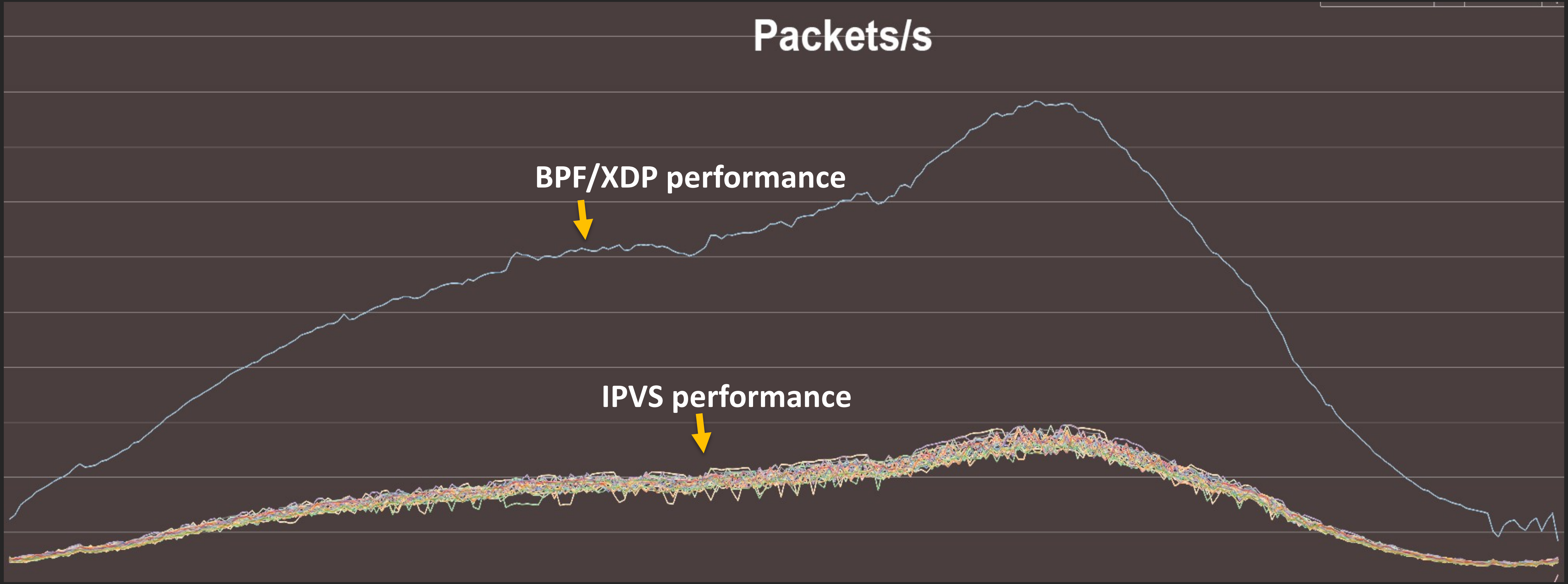


April 17, 2018

Why is the kernel community replacing iptables with BPF?

DEEP DIVE

Facebook published BPF numbers for L3/L4 LB at NetDev 2.1



Source: https://www.netdevconf.org/2.1/slides/apr6/zhou-netdev-xdp-2017.pdf

BPF/XDP: DDoS mitigation

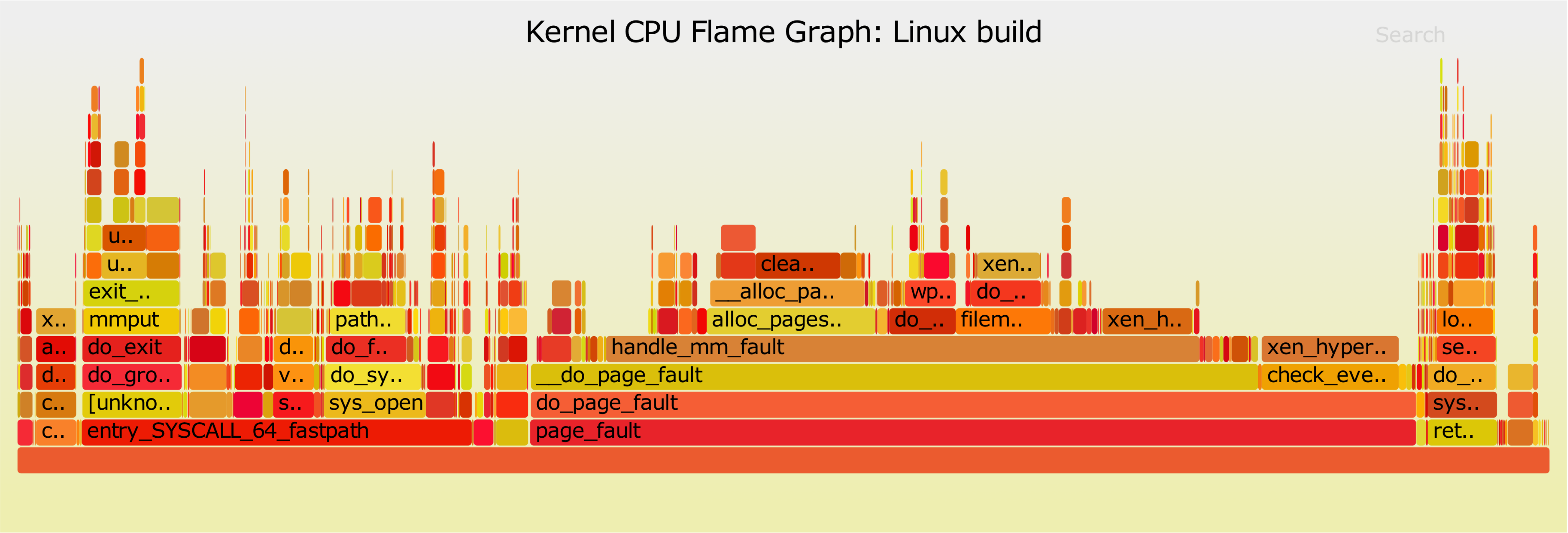
Metric	iptables / ipset	BPF / XDP
DDoS rate [packets/s]	11.6M	11.6M
Drop rate [packets/s]	7.1M	11.6M
Latency under load [ms]	2.3ms	0.1ms
Requests/s under DDoS [Requests/s]	280	82’800

Sender: Send 64B packets as fast as possible
Receiver: Drop as fast as possible

Source: http://schd.ws/hosted_files/ossna2017/da/BPFandXDP.pdf

BPF: “dtrace for Linux”

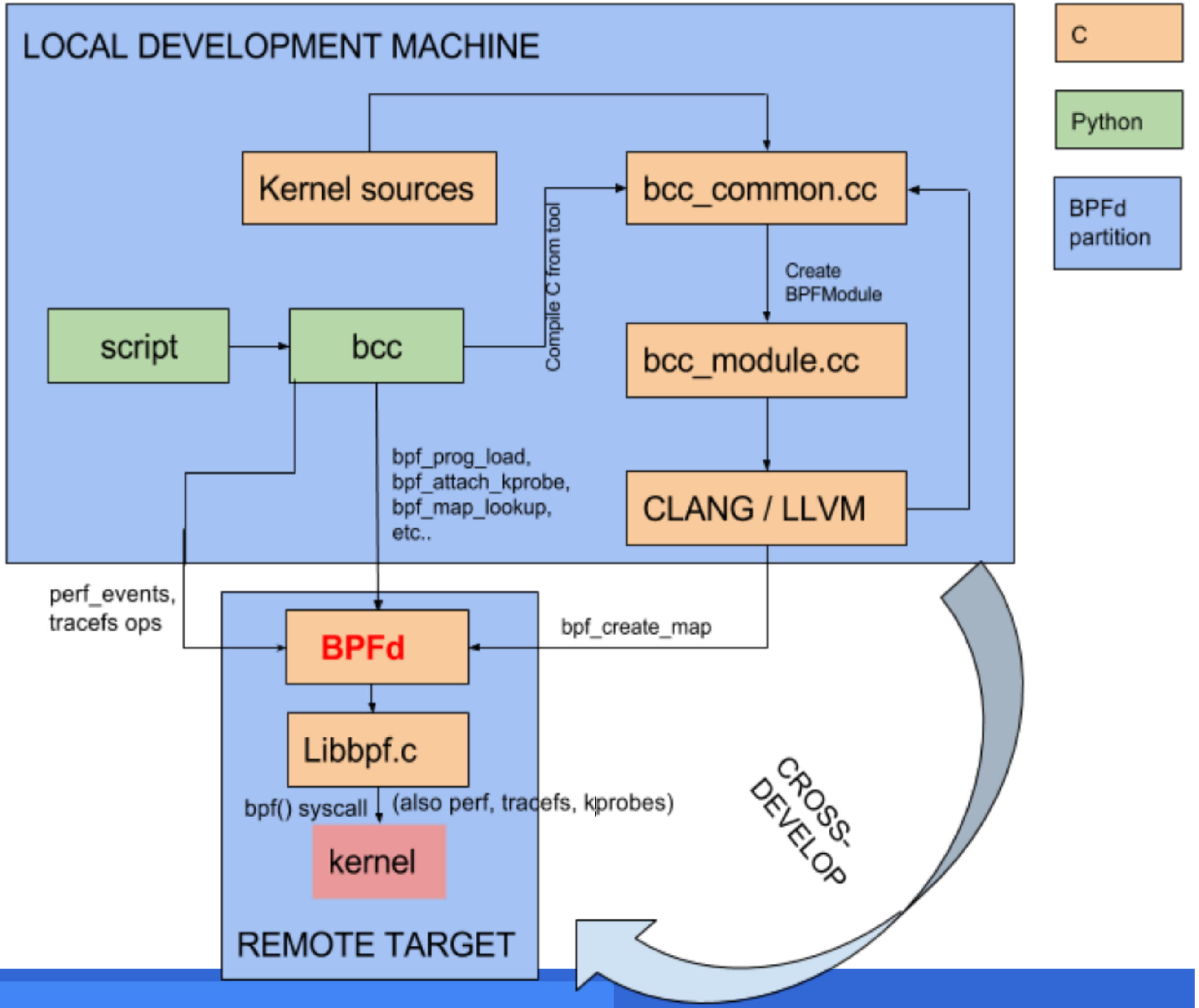
NETFLIX



Source: http://www.brendangregg.com/blog/

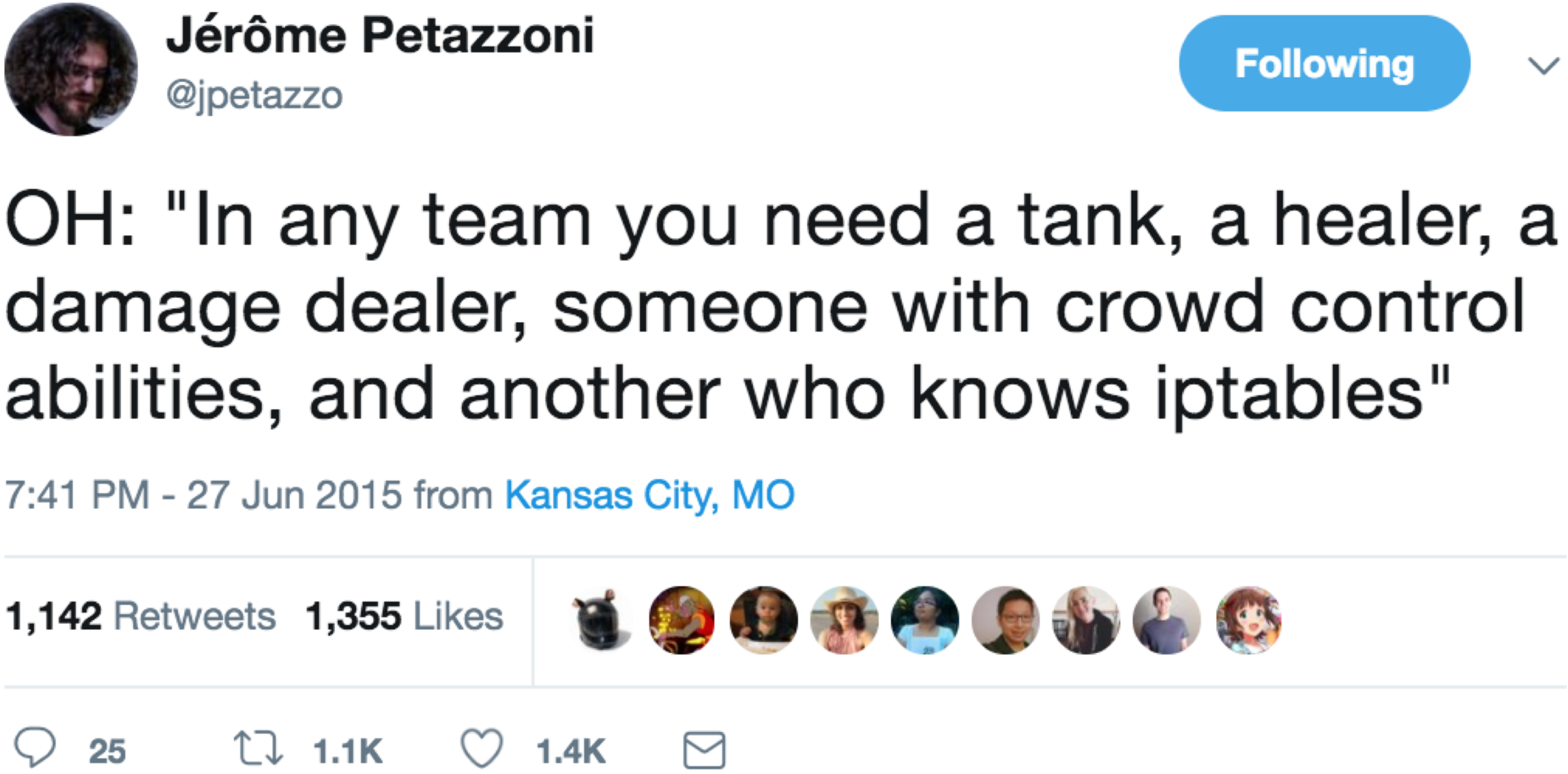
BPFd: Powerful Linux Tracing for Remote targets using eBPF

BCC for remote:



What is your favorite iptables memory?

What is your favorite iptables memory?



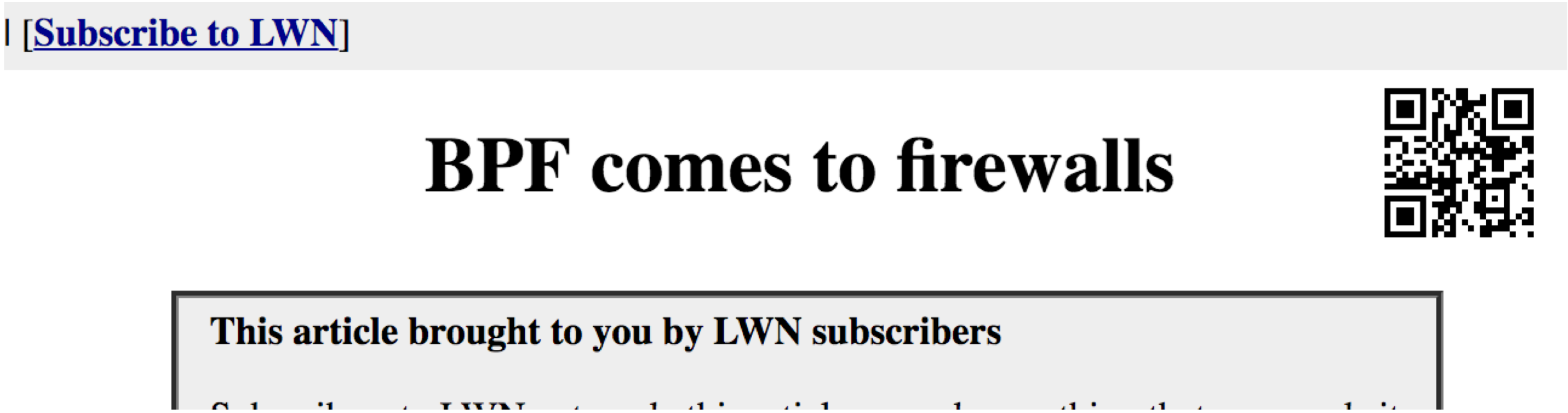
Jérôme Petazzoni
@jpetazzo

OH: "In any team you need a tank, a healer, a damage dealer, someone with crowd control abilities, and another who knows iptables"

7:41 PM - 27 Jun 2015 from Kansas City, MO

1,142 Retweets 1,355 Likes

Kernel developers are saying goodbye to iptables



| [Subscribe to LWN]

BPF comes to firewalls

This article brought to you by LWN subscribers

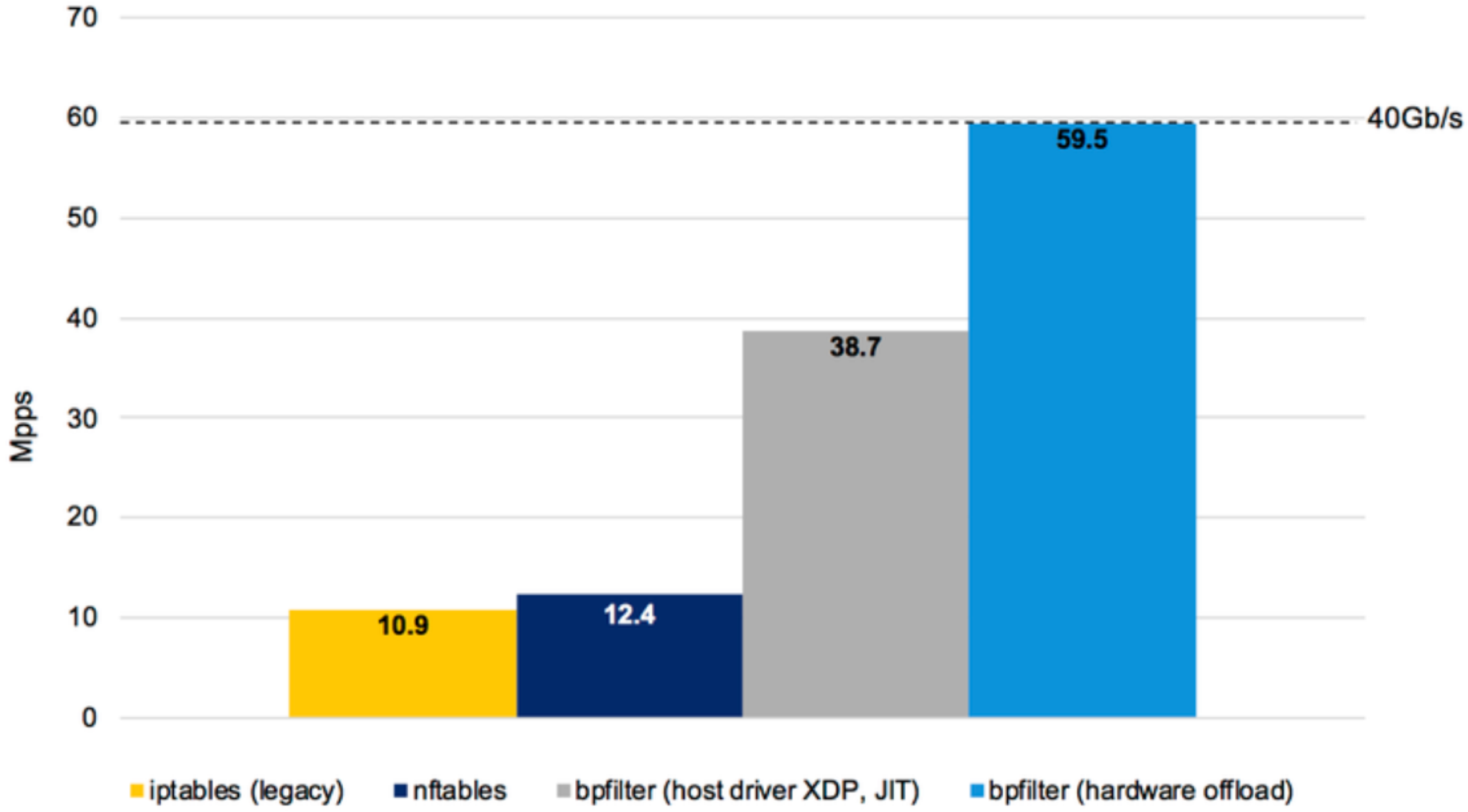
https://lwn.net/Articles/747504/

April 17, 2018

Why is the kernel community replacing iptables with BPF?

DEEP DIVE

Early performance benchmark

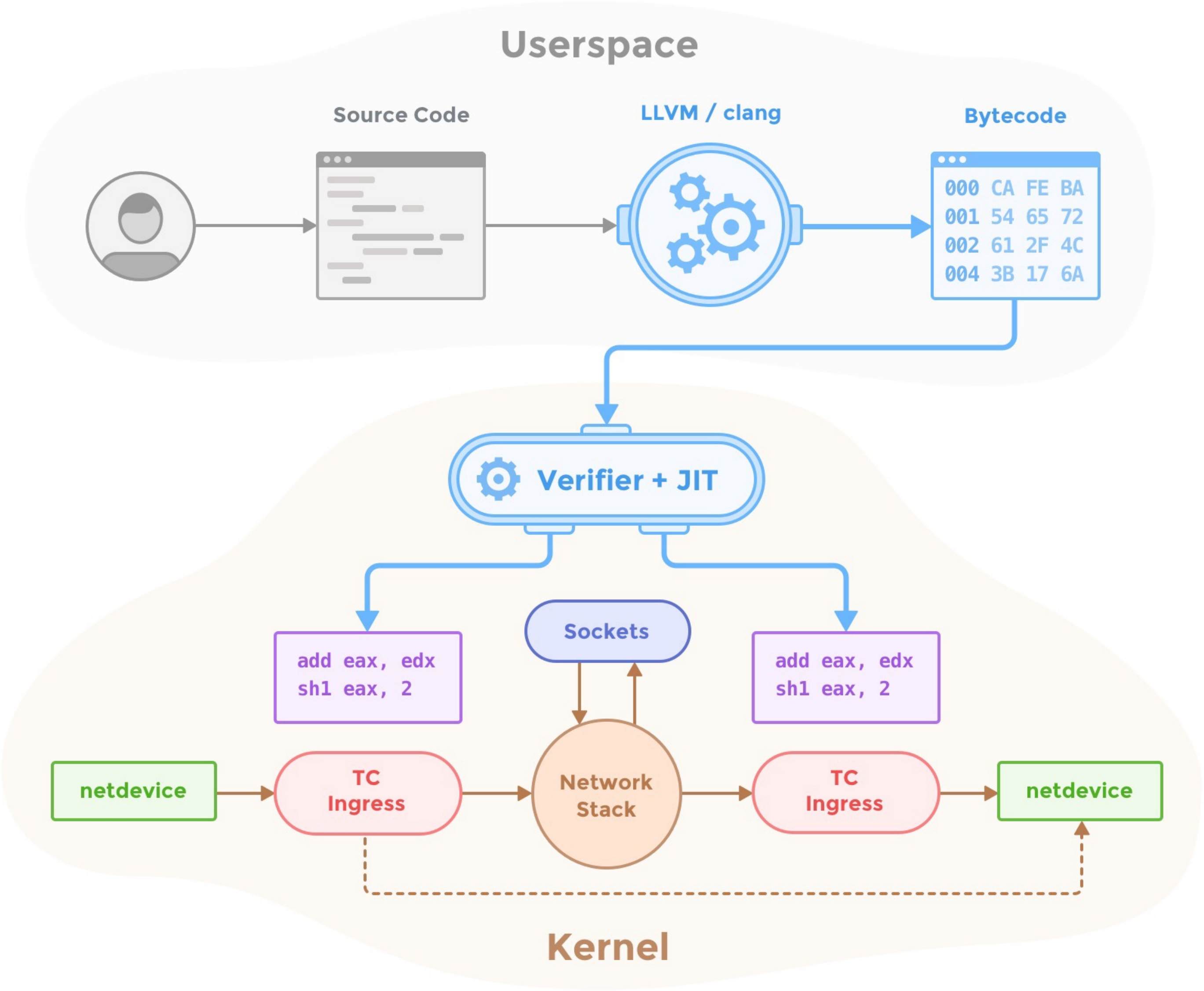


Source: https://bit.ly/2Ks9gJH

So many more examples….

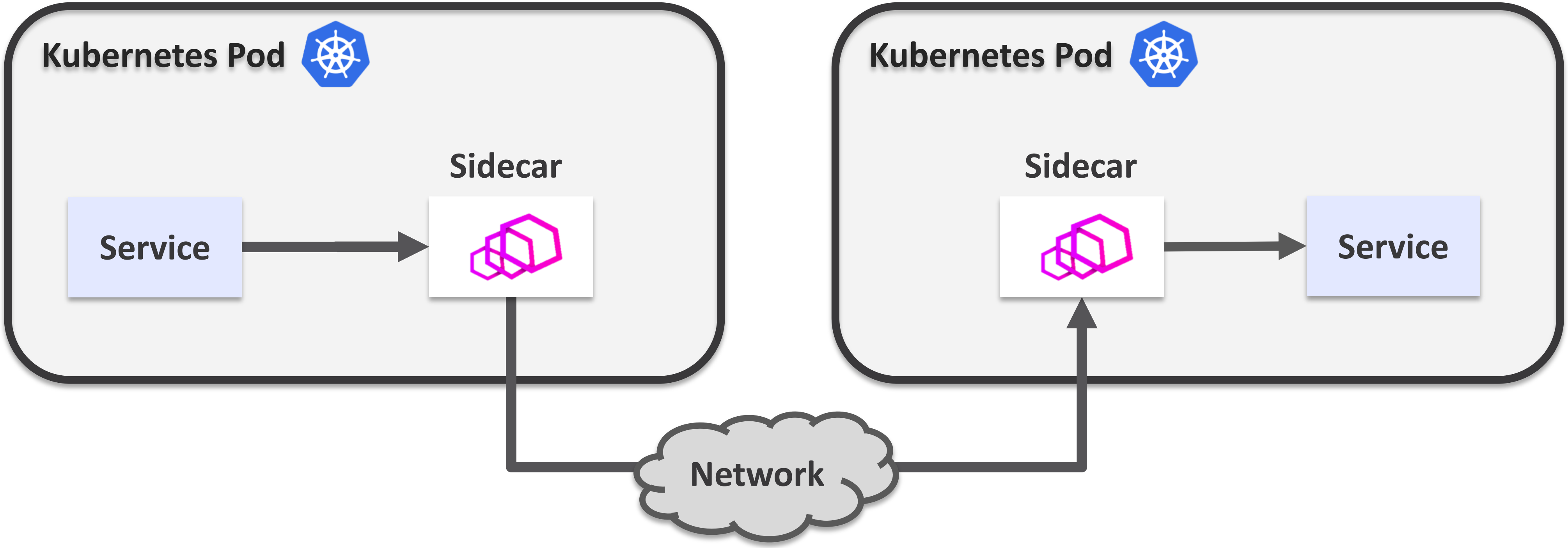
- **Cloudflare DDoS mitigation** [https://www.netdevconf.org/2.1/slides/apr6/bertin_Netdev-XDP.pdf]
- **BCC** [https://github.com/iovisor/bcc]
- **Systemtap** [https://sourceware.org/git/gitweb.cgi?p=systemtap.git;a=summary]
- **Weave Scope** [https://github.com/weaveworks/scope]
- **Suricata** [http://suricata.readthedocs.io/en/latest/capture-hardware/ebpf-xdp.html]
- **systemd** [http://0pointer.net/blog/ip-accounting-and-access-lists-with-systemd.html]
- **gobpf** [https://github.com/iovisor/gobpf]
- **ply** [https://github.com/wkz/ply]
- **bpfps** [https://github.com/genuinetools/bpfps]
- **Perf** [https://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux.git/tree/tools/perf]
- **bpftrace** [https://github.com/ajor/bpftrace]
- **Open vSwitch** [http://www.openvswitch.org//support/ovscon2016/7/1120-tu.pdf]
- **PCP** [https://github.com/performancecopilot/pcp]
- …

BPF toolchain

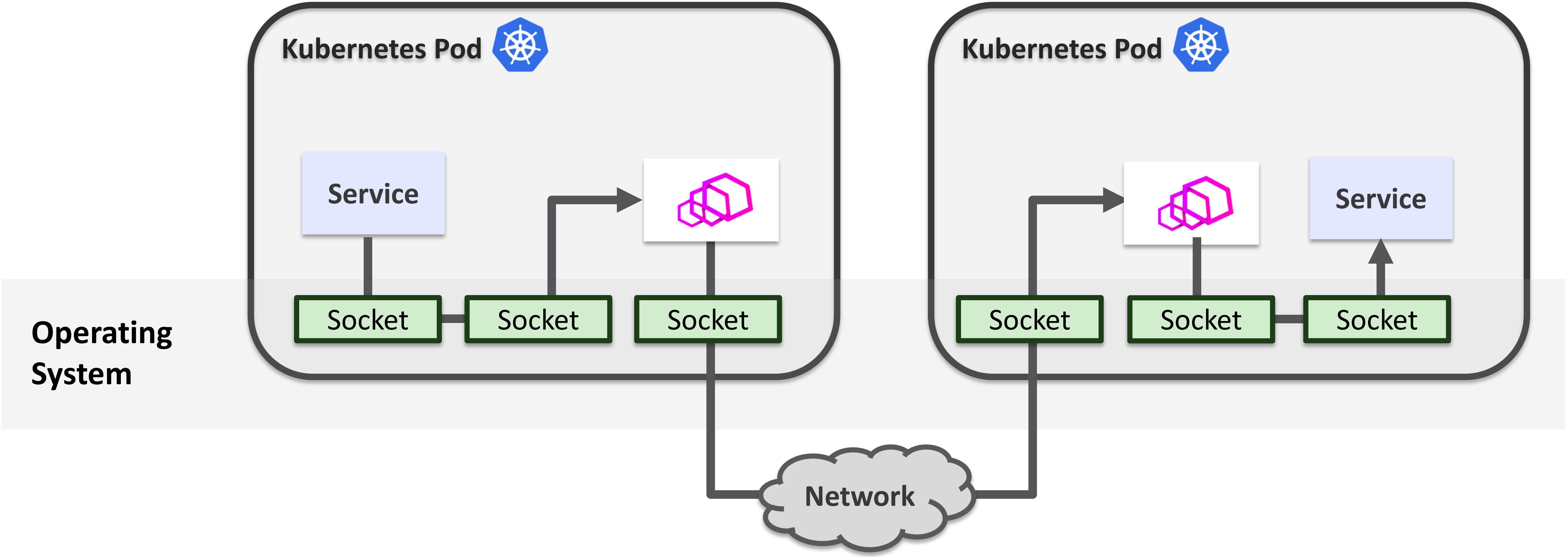


How does BPF apply to Envoy and Service Mesh?

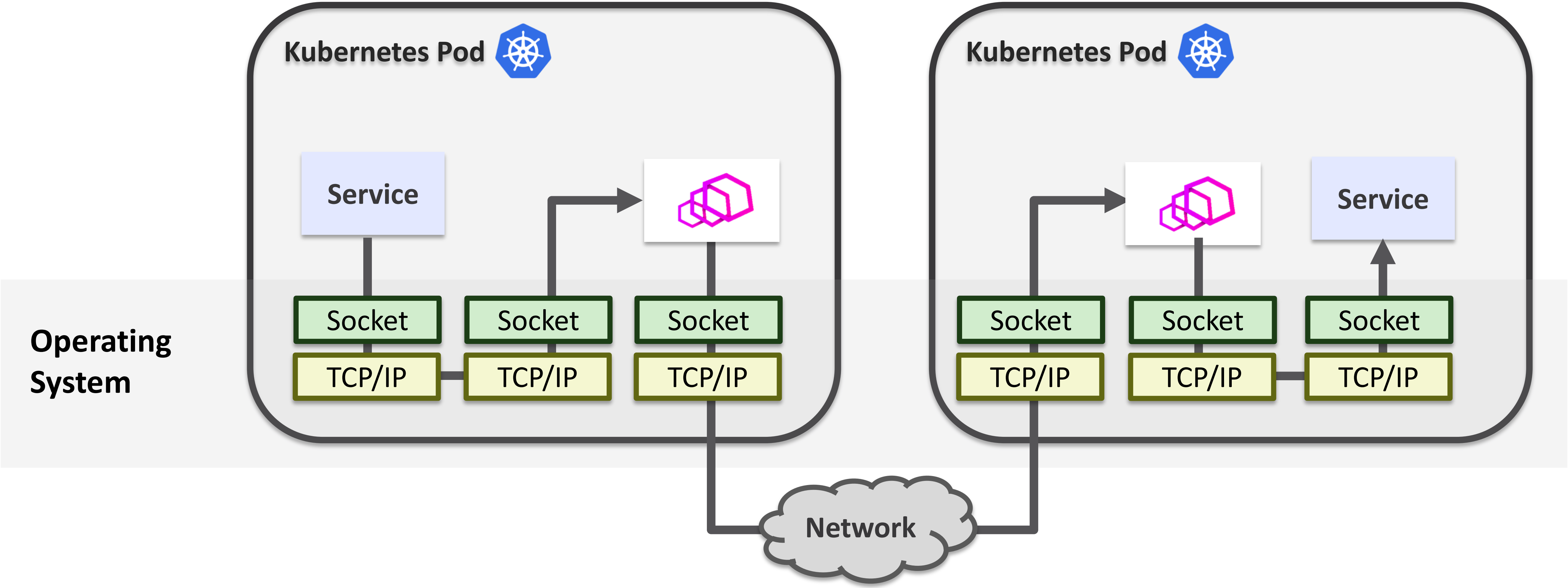
Service Mesh / Sidecar Architecture



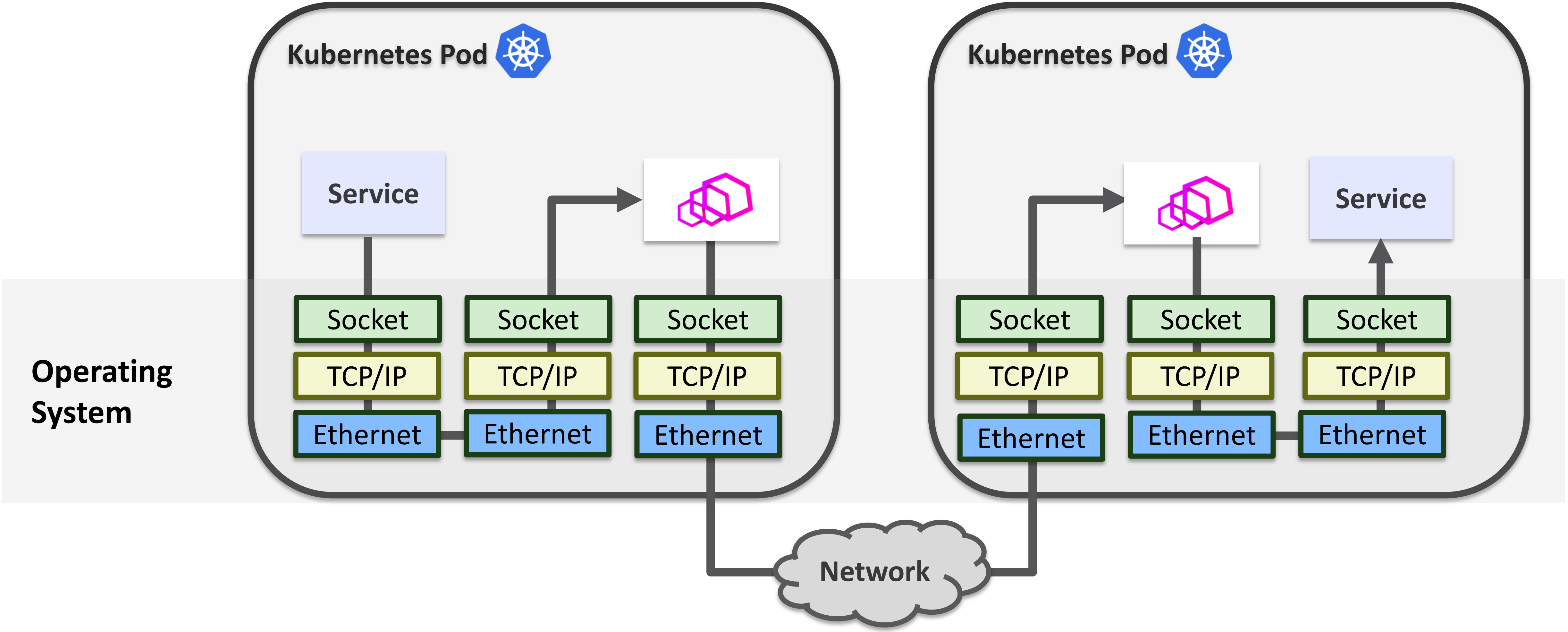
Sidecar Injection



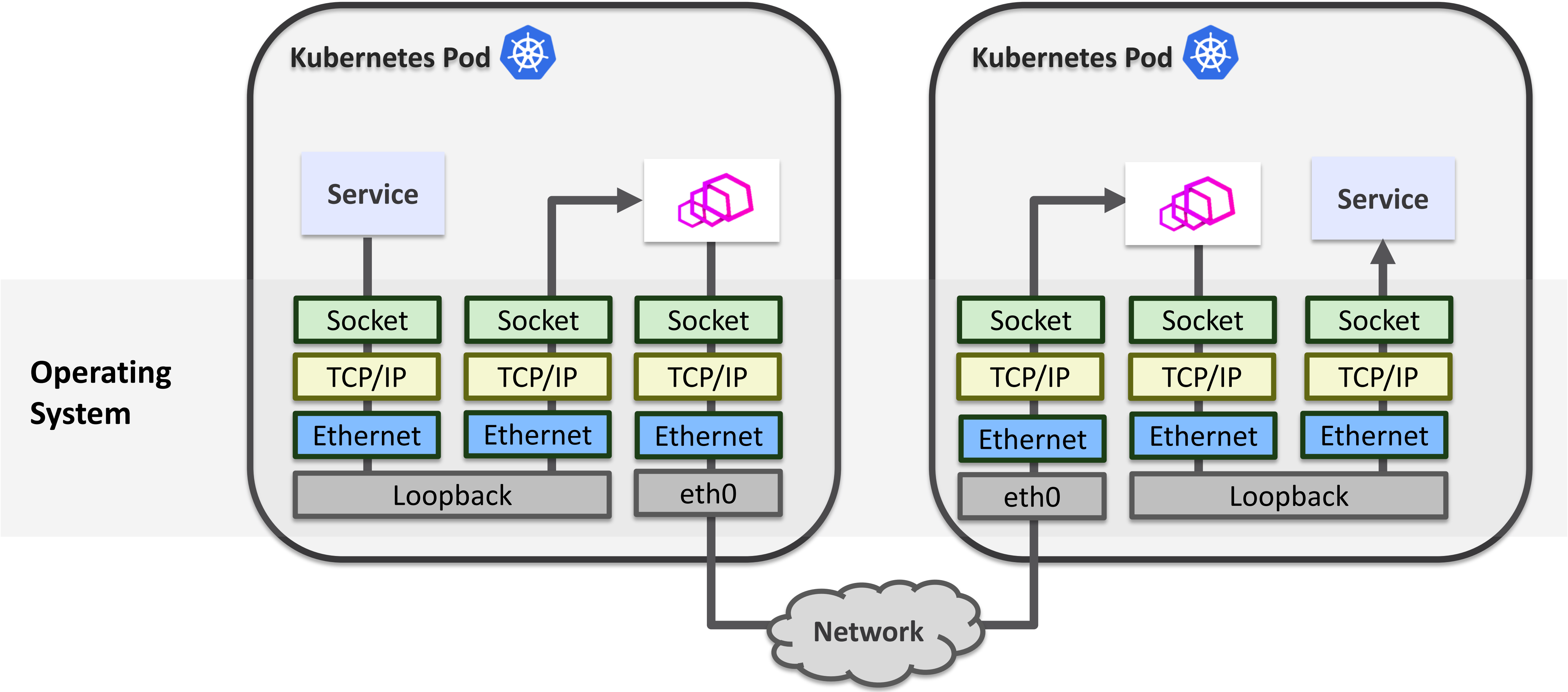
Sidecar Injection



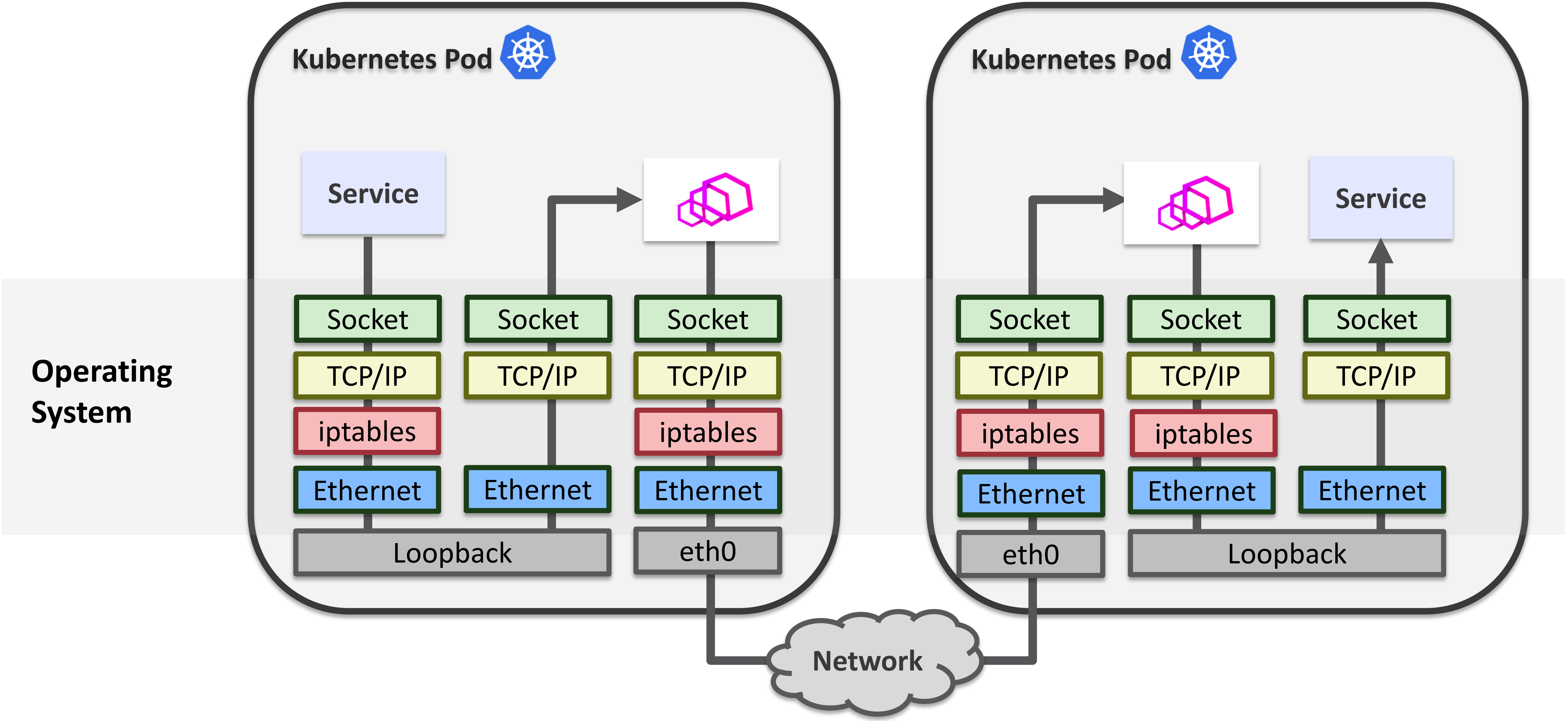
Sidecar Injection



Sidecar Injection (Non-transparent)

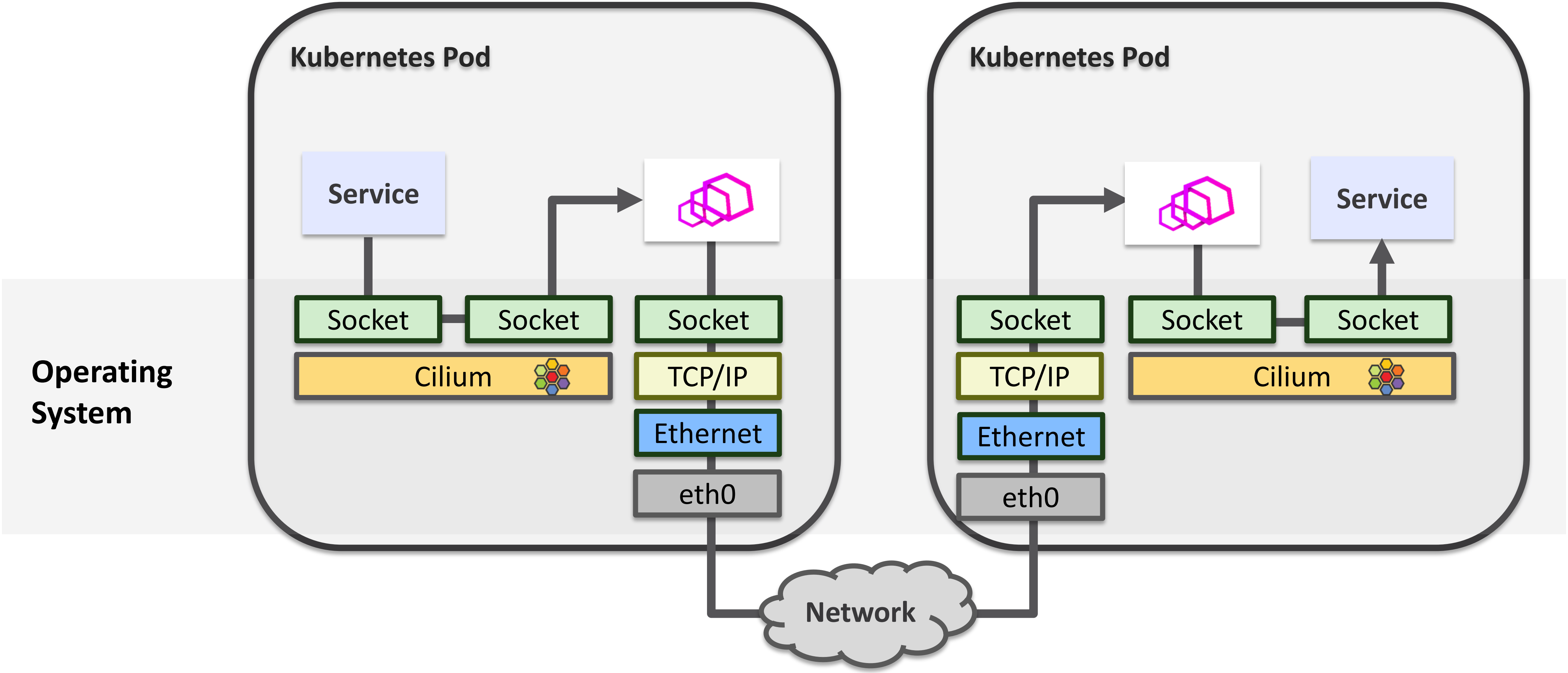


Sidecar Injection (Transparent)

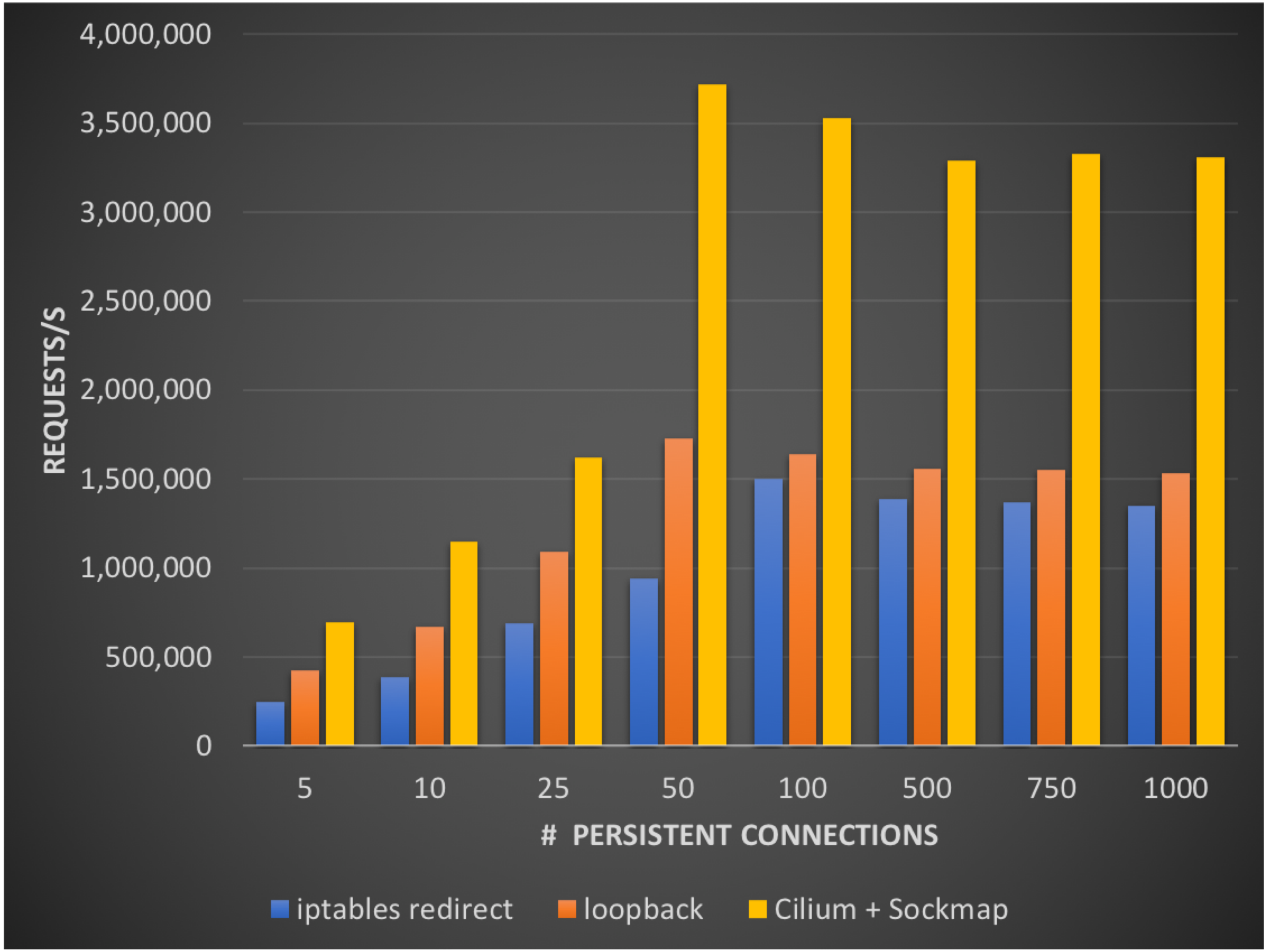


Why use TCP and Ethernet in a single-node, lossless environment?

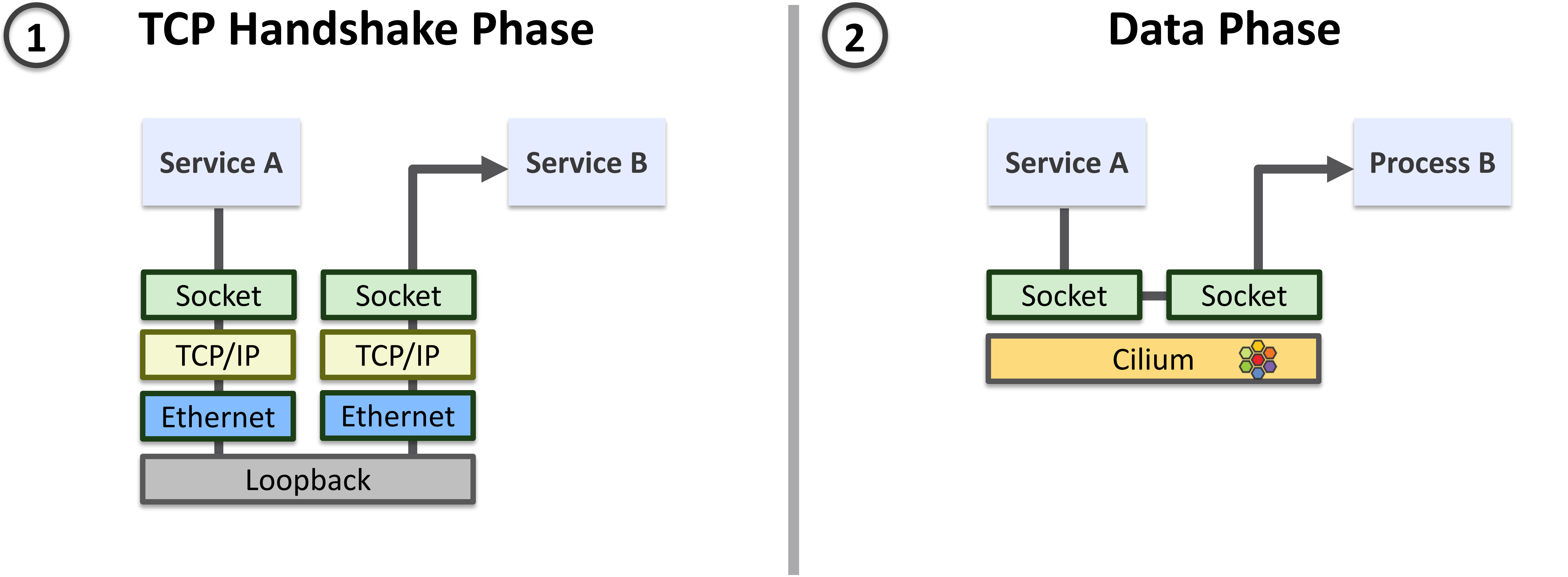
Transparent Sidecar Injection with Cilium



Sidecar Injection Performance



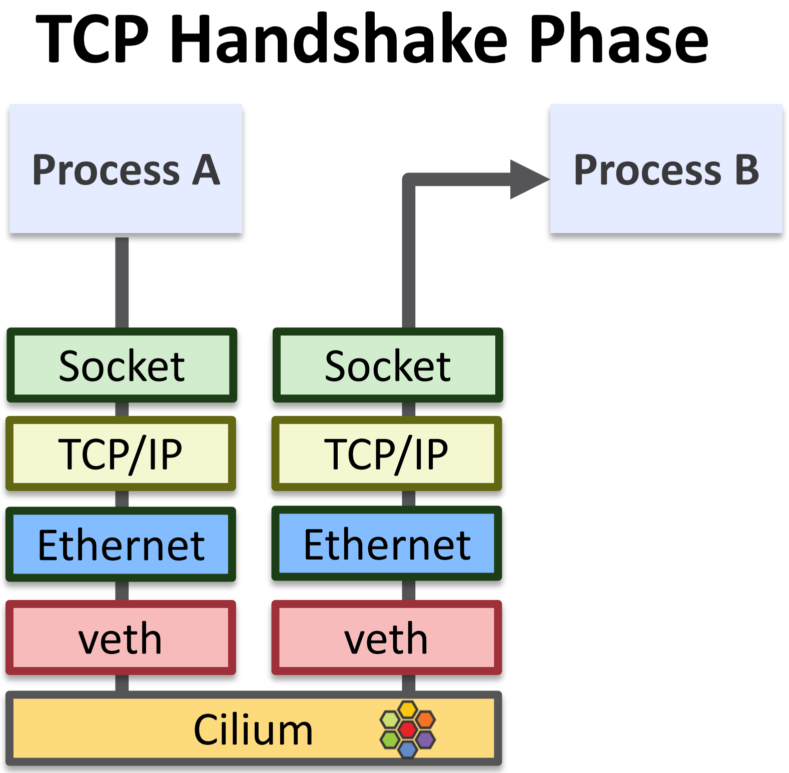
Transparent acceleration during data phase

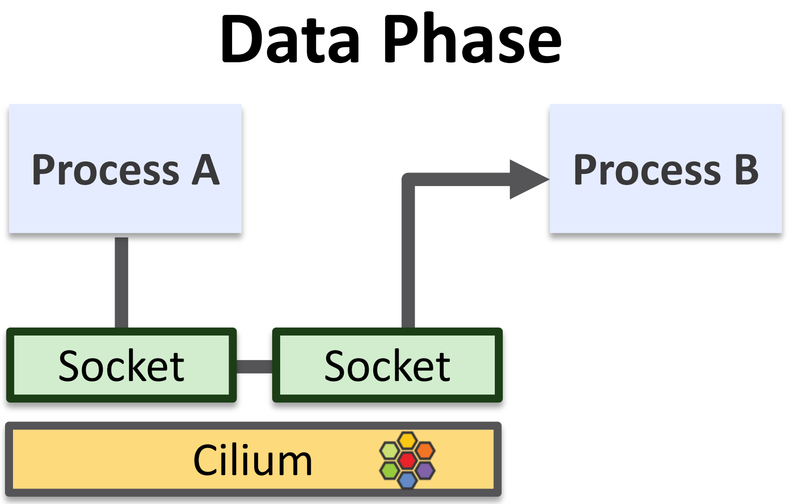


- Transparent acceleration, behavior stays the same
- No changes to application or Envoy needed

But wait...

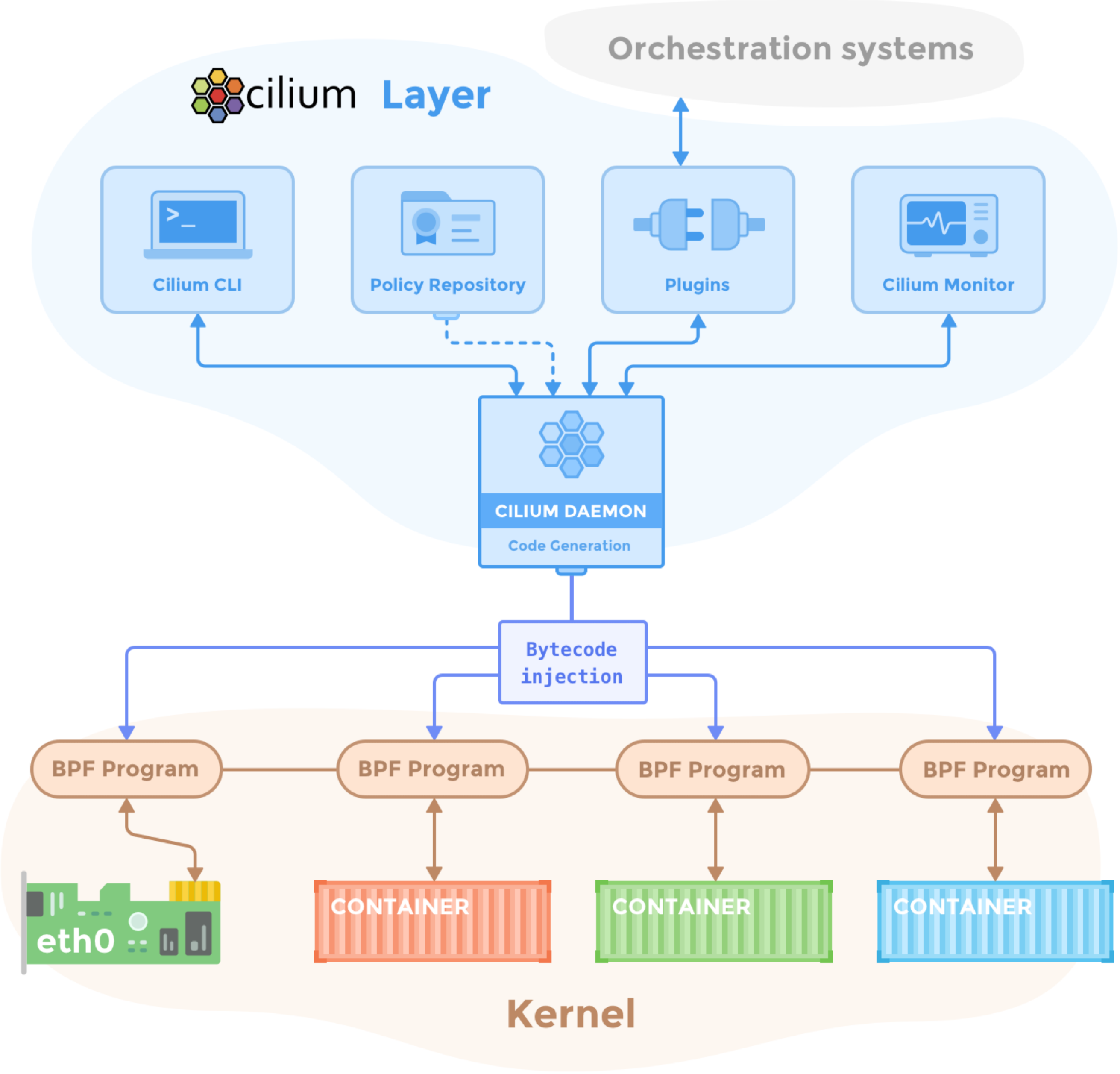
Works with any local socket communication





How do I get this?

Cilium



Cilium:

- Kernel: 4.9

Sidecar Accel:

- Kernel: 4.16
- Cilium: 1.1/1.2

http://github.com/cilium/cilium

Cilium in a Nutshell

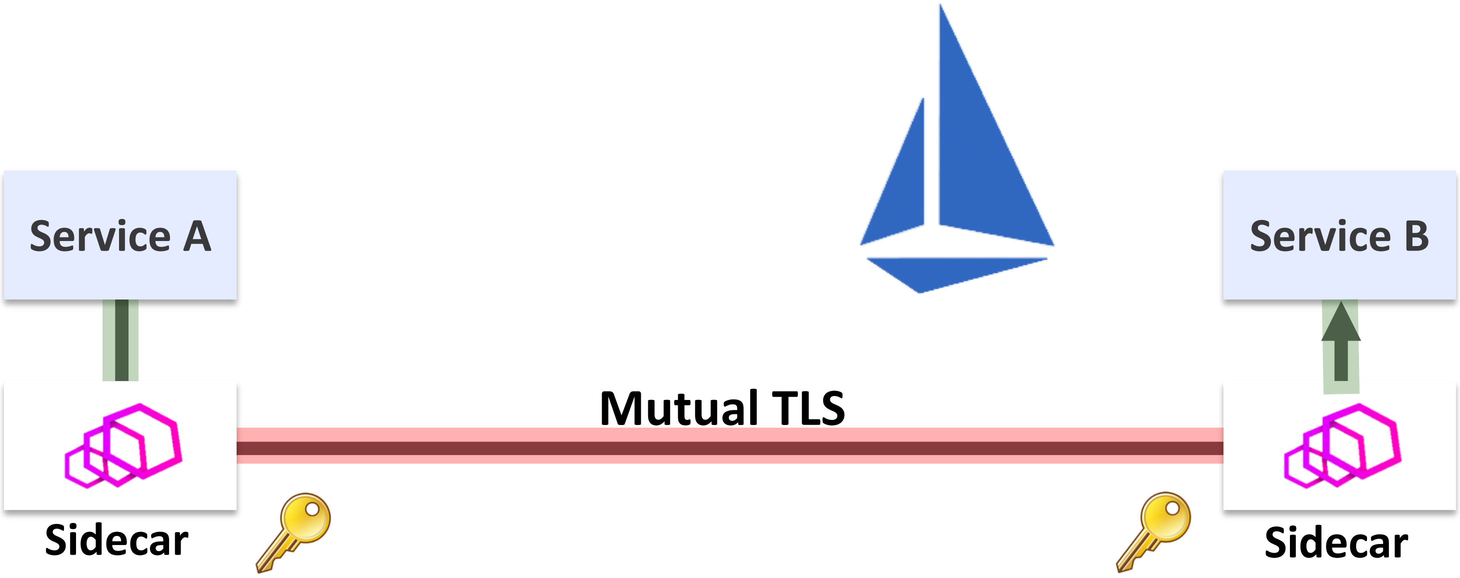
- **Current Release: 1.0.1**
- **Highly efficient BPF datapath**
  - Fully Distributed
  - Service Mesh datapath
- **CNI** and **CMM** plugin
- **Network Security on both Packet and API level**
  - Identity Based
  - IP/CIDR as fallback
  - API Aware (HTTP, gRPC, Kafka, [more coming soon])
- **Distributed and Scalable Load Balancing**
- **Simplified Networking Model**
  - Overlay
  - Direct Routing
  - Delegation to other plugin (1.1)
- **Visibility / Tracing**

It gets better…

It gets better…

… and a little bit scary

TLS in Service Mesh



Service A

Service B

Mutual TLS

Sidecar

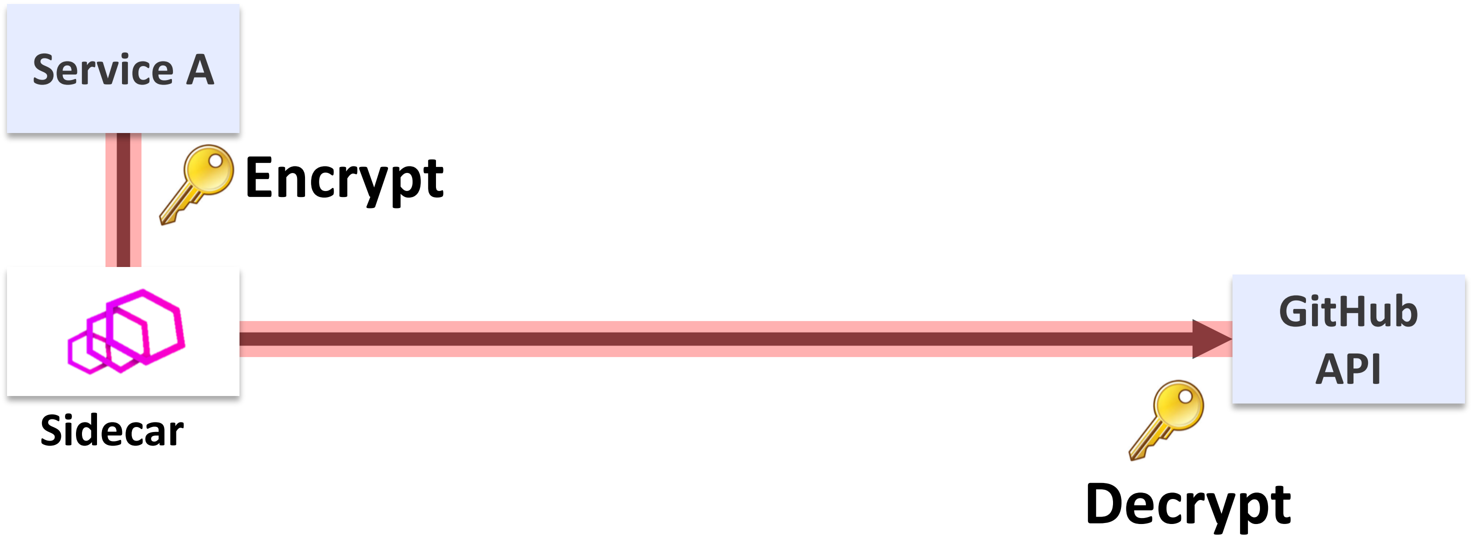
Sidecar

🟩 Clear Text

🟥 Encrypted

🔑 Key

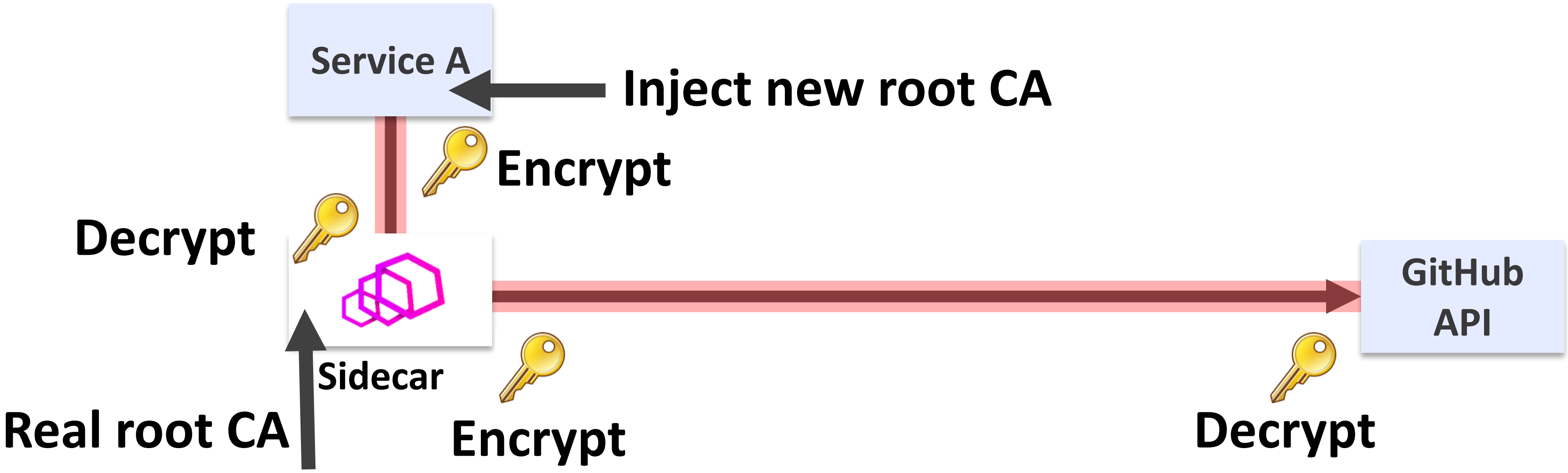
Accessing external services via sidecar using TLS



🟩 Clear Text

🟥 Encrypted

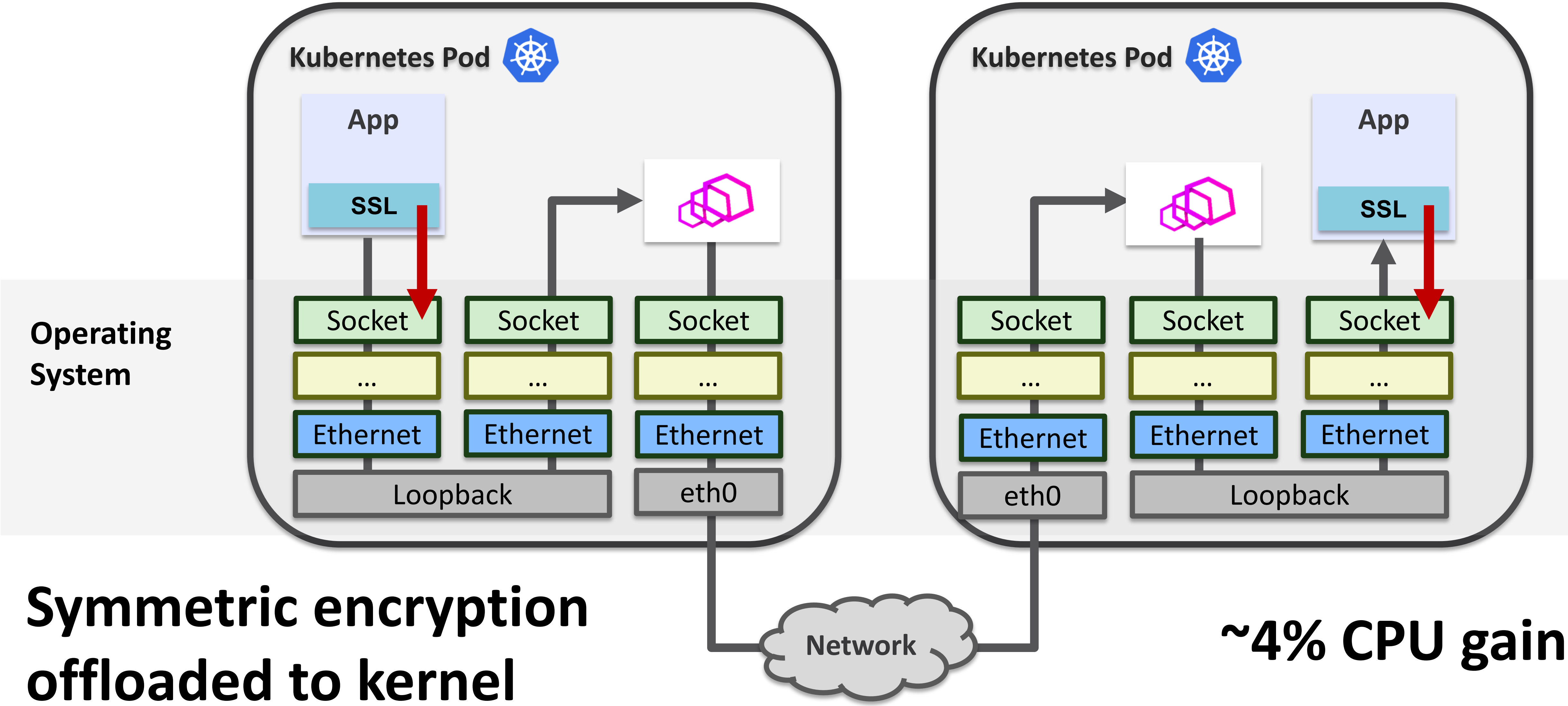
Option #1: Inject Root CA



🟩 Clear Text

🟥 Encrypted

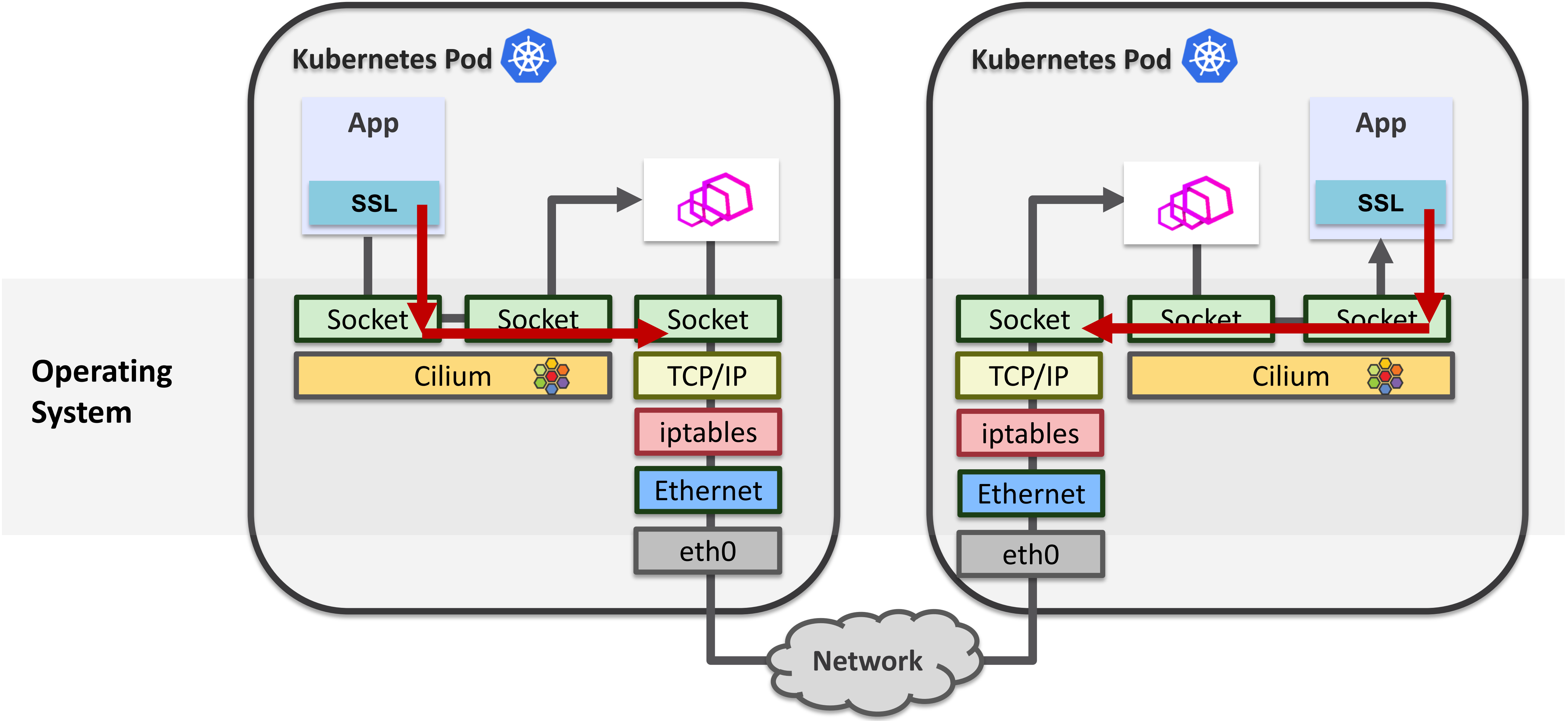
kTLS



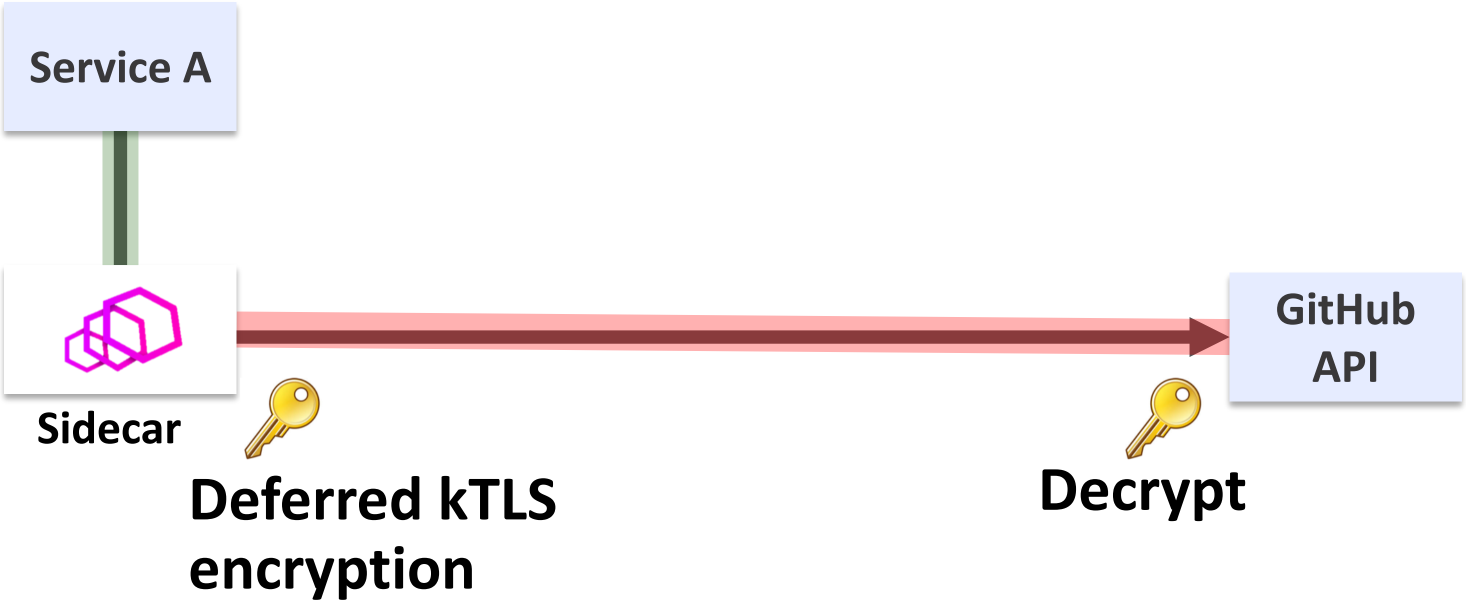
Symmetric encryption offloaded to kernel

~4% CPU gain

kTLS + Cilium + Envoy



Option #2: kTLS



Service A

Sidecar

GitHub API

Deferred kTLS encryption

Decrypt

🟩 Clear Text

🟥 Encrypted

Sidecar for TLS encrypted connections without CA injection & decryption

What date is today?

Thank You!
Questions?

Getting Started:
http://cilium.io/

@ciliumproject

http://github.com/cilium/cilium