Relevance of eBPF* to Telcos (NFV)

Sridhar

LF NETWORKING

THE LINUX FOUNDATION

eBPF provides immense scope for innovation

eBPF & Anuket: Summary

Currently actively involved in this.

Performance Benchmarking [ViNePerf]

Specification R*-2

eBPF++

RI-2

What exactly?

Driven by Usecases of eBPF

THE **LINUX** FOUNDATION

LF NETWORKING

Basics
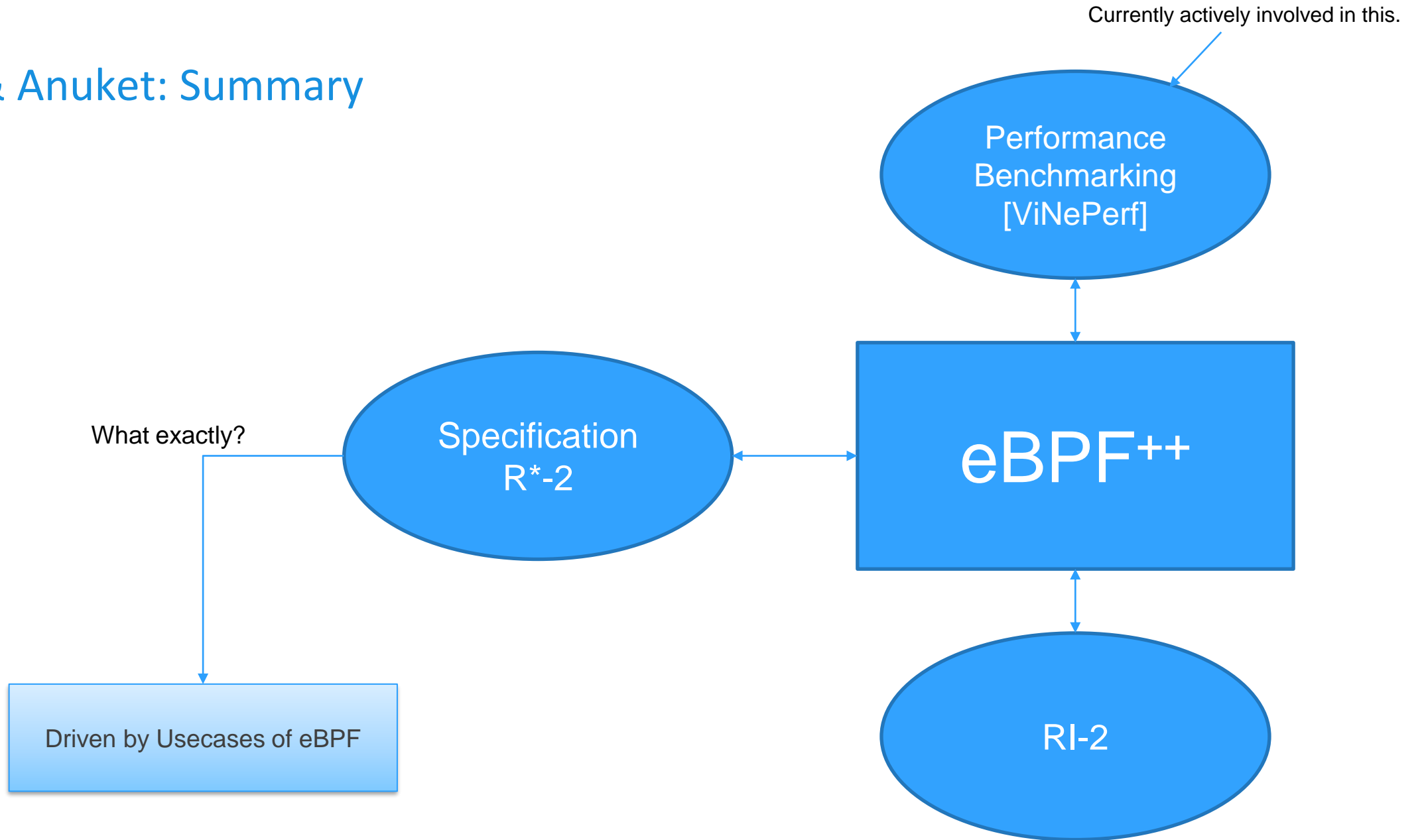
Extended Berkeley Packet Filter (eBPF)

› Run sandboxed programs in a privileged context

    › operating system kernel.

    › safely and efficiently extend the capabilities of the kernel without requiring to change kernel source code or load kernel modules

        › run eBPF programs

        › OS guarantees safety and execution efficiency as if natively compiled with the aid of a Just-In-Time (JIT) compiler and verification engine.

        › Wave of eBPF-based projects

› *Linux networking paradigm that exposes programmable hooks to the network stack inside the Linux kernel.*
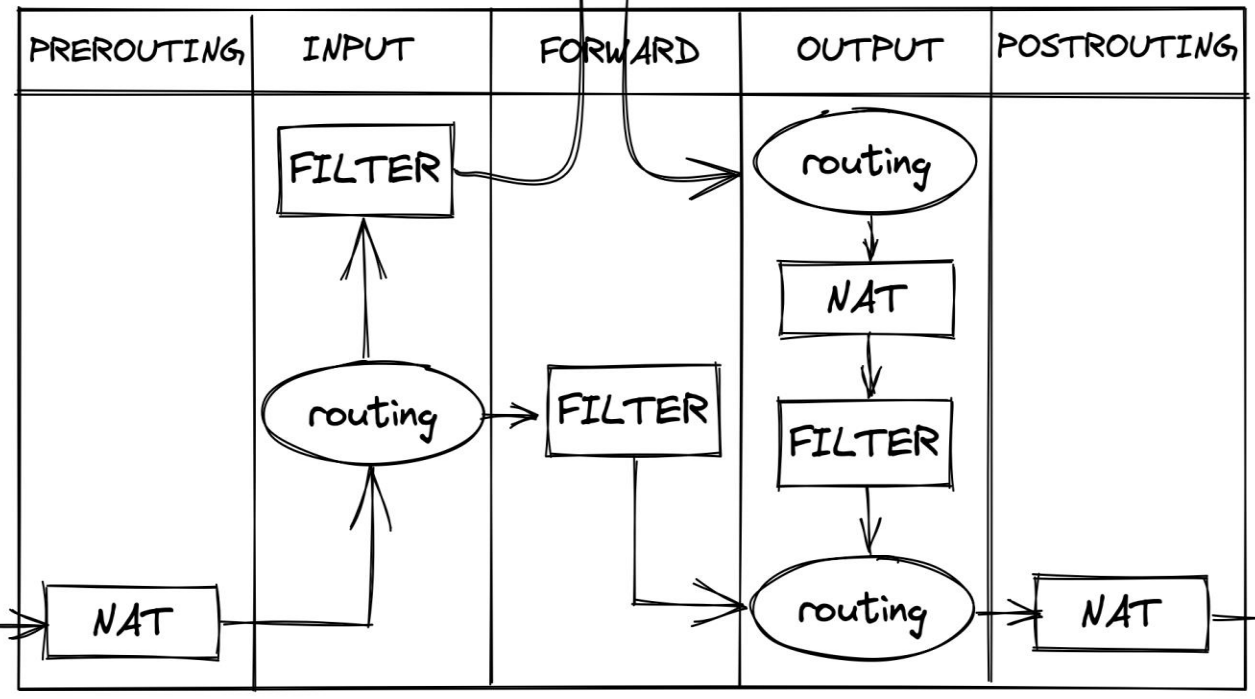
THE **LINUX** FOUNDATION

LF NETWORKING

eXpress Data Path: XDP

› Allows developers to attach eBPF programs to low-level hooks

  › implemented by network device drivers in the Linux kernel*

› runs a BPF program as early* as possible

  › used to achieve high-performance packet processing in an eBPF architecture, primarily using kernel bypass.

    › Control of the network interface card (NIC) is transferred to an eBPF program

› XDP program can perform 4-5 operations with the packets it receives

User space

Kernel space

Local Processes

PREROUTING | INPUT | FORWARD | OUTPUT | POSTROUTING

FILTER

routing

NAT

FILTER

routing

FILTER

routing

skb alloc

TC (ingress)

NAT

NAT

TC (egress)

XDP

Netfilter

Hardware

netdev (eth0)

netdev (eth0)

Source: John Lin

Cilium

› Open source project that has been designed on top of eBPF to address the networking, security, and visibility requirements of container workloads

› *High-level abstraction on top of eBPF.*

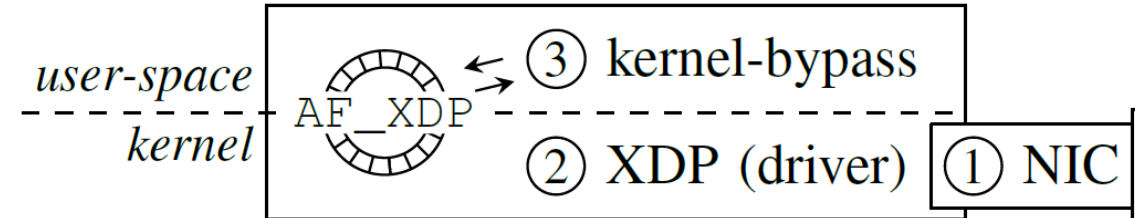   › Cilium is to eBPF what Kubernetes and container runtimes are to Linux kernel namespaces, cgroups, and seccomp

eBPF and NFV

eBPF Landscape

Source: ebpf.io
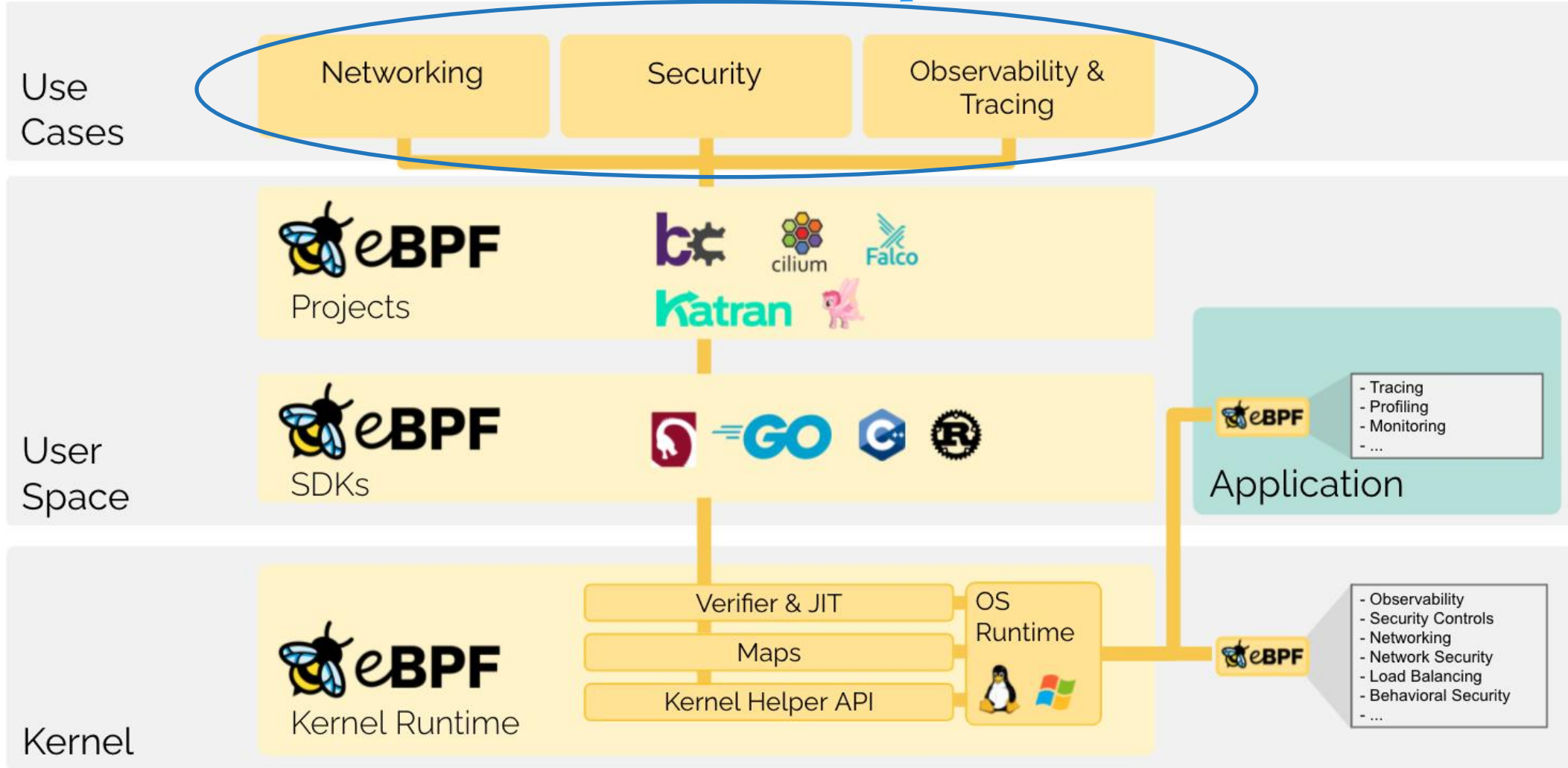
eBPF Usecases for NFV: Networking

› eBPF Based Network Functions
  › Lot of works since last 5 years.
    › Xhonneux et. al., (CoNext 2018) – END.EBPF function
    › Mobile Gateway (Parola et. Al., IEEE NetSoft 2021)
    › Polycube (Miano et. al., IEEE Trans on NSM 2021)
    › Rethinking NFs – 5G UPF using BPF (IoT usecase) - navarrothiago/upf-bpf

› Improving linux dataplane performance (Ben Plaff et. al. Sigcomm 2021, Hohlfeld et. Al. 2019)
  › OVS + xDP. 2 Options for OVS with eBPF : datapath in userspace (AF_XDP) and xdp based datapath in kernel

› CNIs
  › Cilium CNI
    › Why Cilium?
      › Kube-Proxy Replacement, Multi-Cluster, NetworkPolicy Logging, DNS-Aware Network Policy
  › Calico eBPF (native support for Kubernetes services)
  › af_xdp plugins (device and CNI)

› N/w Infrastructure Components (middlebox)
  › LoadBalancer (L4)

› Cilium based Service-Mesh
  › Service-Mesh and CNFs

Latest Cilium Release

eBPF Usecases for NFV: Observability

› End-2-End In-Band Telemetry

  › intel/host-int

› Analyse TCP retransmission

› Container-Level Network Observability

› Userspace tracing to solve DNS problems.

› Bandwidth Management with eBPF

› AI/ML problems based on eBPF Tracing Data

  › Network Anomaly Event Prediction and Optimal Resource Control in Cloud Native Network Functions

eBPF Usecases for NFV: Security

› Cilium

    › Container Security

        › Network Policy, Policy-driven SSL termination & injection, etc.

› DDoS Attacks detection and Mitigation (as early as 2016)

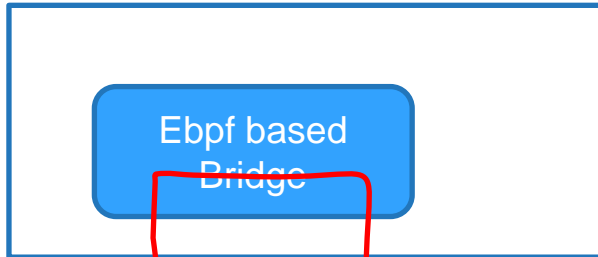   › Rispetto et. al., and Ognibene et. al.,

Slow Adoption

› Not many published works – eBPF & NFV – **YET!**

› Cannot use Cilium CNI with DPDK-based VNC/CNF.

    › Incompatible virtual interface.

› Concerns about operations support.

› Performance Concerns.

› eBPF Community's Opinion: Not many takers in Telco Community for eBPF/Cilium
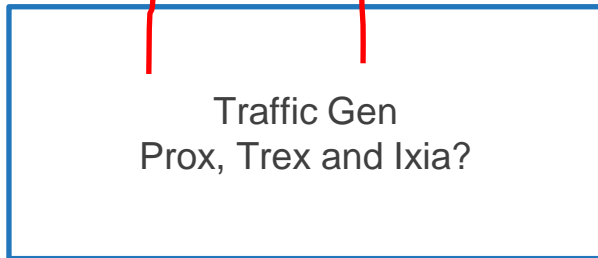
eBPF Performance Analysis (Networking)

› Kernel Networking with eBPF.

  › XDP performance analysis works (ex: Zachary et. al., 2021).

  › VM and Container workloads (Hohlfeld and Ben Plaff et. al)

› CNIs: Cilium and Calico

  › https://docs.cilium.io/en/stable/operations/performance/benchmark/

    › TCP Throughput, %CPU consumption, TCP Req/Response Rate.

    › Netperf (kubenetbench). Pod2Pod. 100Gig I/F.

  › https://itnext.io/benchmark-results-of-kubernetes-network-plugins-cni-over-10gbit-s-network-updated-august-2020-6e1b757b9e49

    › Similar metrics, UDP included, Cilium performance is not there yet!

› Gaps:

  › Standards (ETSI/IETF/..) based testing.

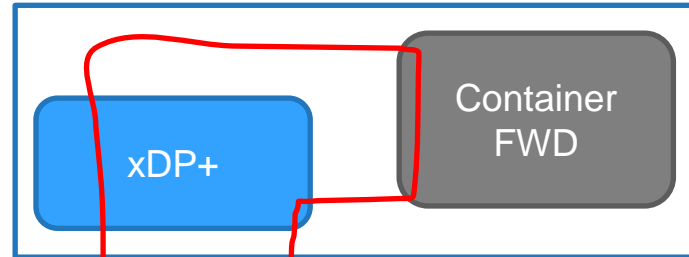 › Studies with DPDK-based (accelerated) workloads – Telco Use cases.
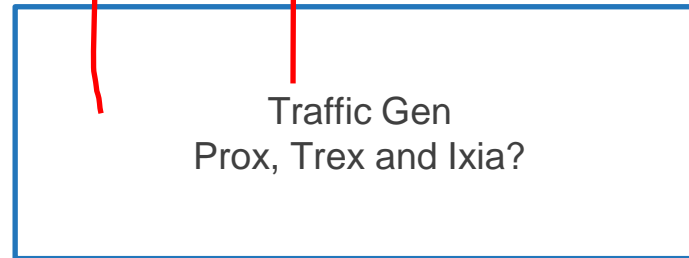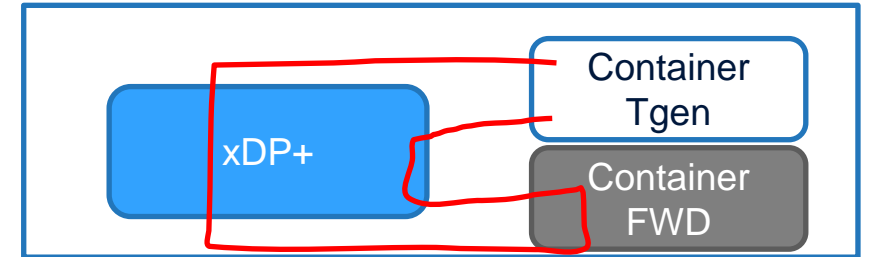
ViNePerf

Topologies

1

Ebpf based
Bridge

P2P

Traffic Gen
Prox, Trex and Ixia?

2

xDP+

Container
FWD

PCP

North-South

Traffic Gen
Prox, Trex and Ixia?

3

xDP+

Container
Tgen

Container
FWD

East-West

ViNePerf Goals

› Questions to Answer:

  › Can this be an alternative to DPDK based solutions? (ex: userspace virtual switches)

    › Reference: Ben Plaff et. al.,

  › What are the parameters that affect the performance - Golden Configuration

  › What are the right test methodologies?

  › What are the right test-tools - Traffic Generators with test traffic measurements, Packet Forwarders, etc.

Thank You!

srao@linuxfoundation.org