OPEN POSSIBILITIES.

NICs for Hyperscalers

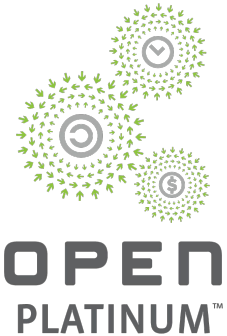
OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

NICs for Hyperscalers

Parveen Patel,
Network Infrastructure @ Google

Jakub Kicinski
Network Infrastructure @ Facebook

OPEN POSSIBILITIES.



Why care about the NIC?

NETWORKING

- NICs are critical to data center innovation
  - Core data center and edge applications
  - Accelerators (e.g., crypto, compression) critical to DC efficiency
  - Enable composable infrastructure

Foundational vs Smart NIC

NETWORKING

- Foundational NIC
  - Traditional NICs with modern acceleration capabilities
  - The entire control plane runs on the host CPU
  - TCO-optimal for trusted workloads, dedicated appliances
  - Primary focus of this proposal

- Smart NIC
  - Foundational NIC plus compute cores
  - Enables offload of the control plane
  - Unified infrastructure for Baremetal and VMs
  - Outside the scope of this proposal

OPEN POSSIBILITIES.

Challenge - enable more innovation

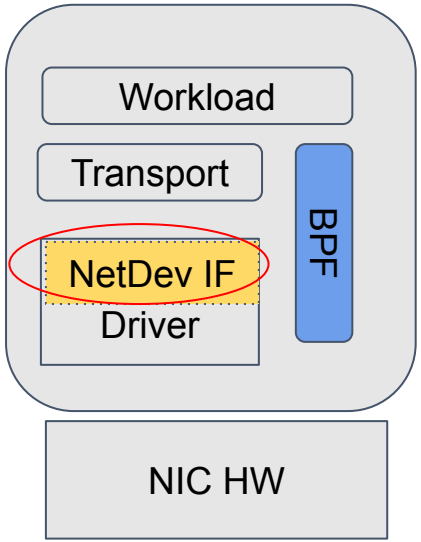
NETWORKING

- Difficult for NIC vendors to understand the needs of the operators
  - Unclear requirements
    - Examples: scale, security, packet formats

- Difficult for operators to bring in a new NIC
  - Proprietary interfaces
    - Examples: telemetry, packet steering APIs
  - Complex set of features
    - Slowing down deployment
    - New features affect basic functionality

High-level proposal

- Standardize APIs for NIC hardware features
- Consistent APIs to make porting easier
- Initial focus on the NetDev Interface
- Establish meaningful benchmarks



NETWORKING

Example hyperscale features

NETWORKING

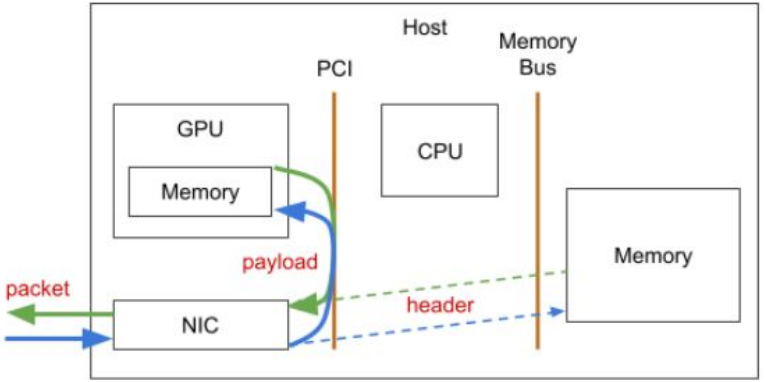
- Flow steering
- Traffic engineering
- Inline crypto
- Transport
- Operations at scale

OPEN POSSIBILITIES.

Flow steering

- Uniform API for “Application Queues”
  - Enable alternative transport stacks: QUIC, SNAP
  - Userspace (e.g., DPDK/AF_XDP) workloads
- Enable CPU memory bypass
  - Direct data to/from GPU/TPU/SSD memory
  - Dealing with TCP ordering constraints
- Multiqueue
  - #queues < #cores
  - Traffic prioritization across queues

NETWORKING



TCP Direct Data Placement

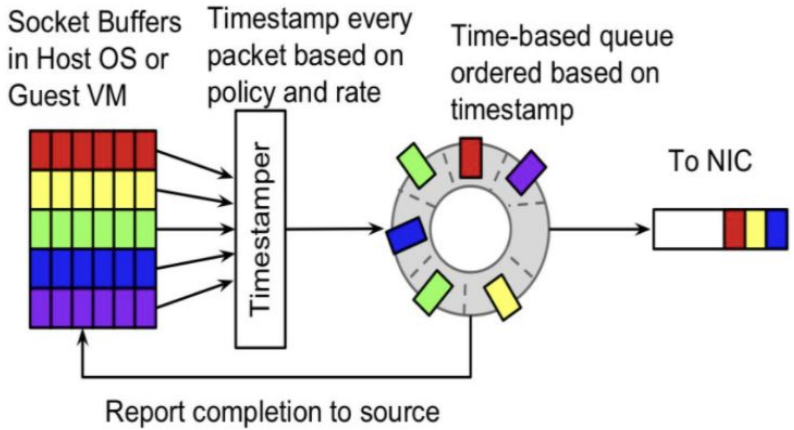
OPEN POSSIBILITIES.

QUIC: https://www.chromium.org/quic
SNAP: https://research.google/pubs/pub48630/

Traffic engineering

NETWORKING

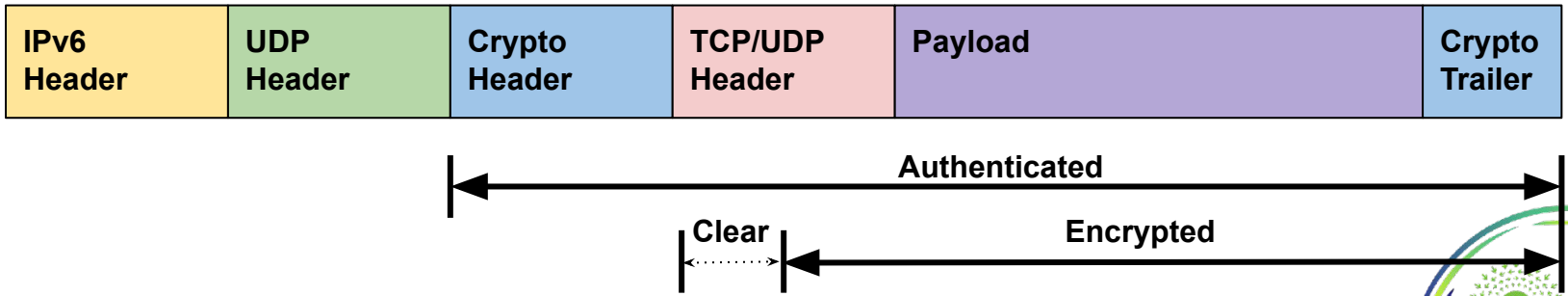
- Flexible tunnel *stack* support
  - Multiple encapsulations: UDP-over-MPLS-over-GRE
  - Idempotent offloads: replicate, do not parse headers
  - 9K client MTU *plus* maximal TE envelope
- Rate limiting via Earliest Departure Time (EDT)
  - https://www.files.netdevconf.info/d/4ee0a09788fe49709855/



Inline crypto

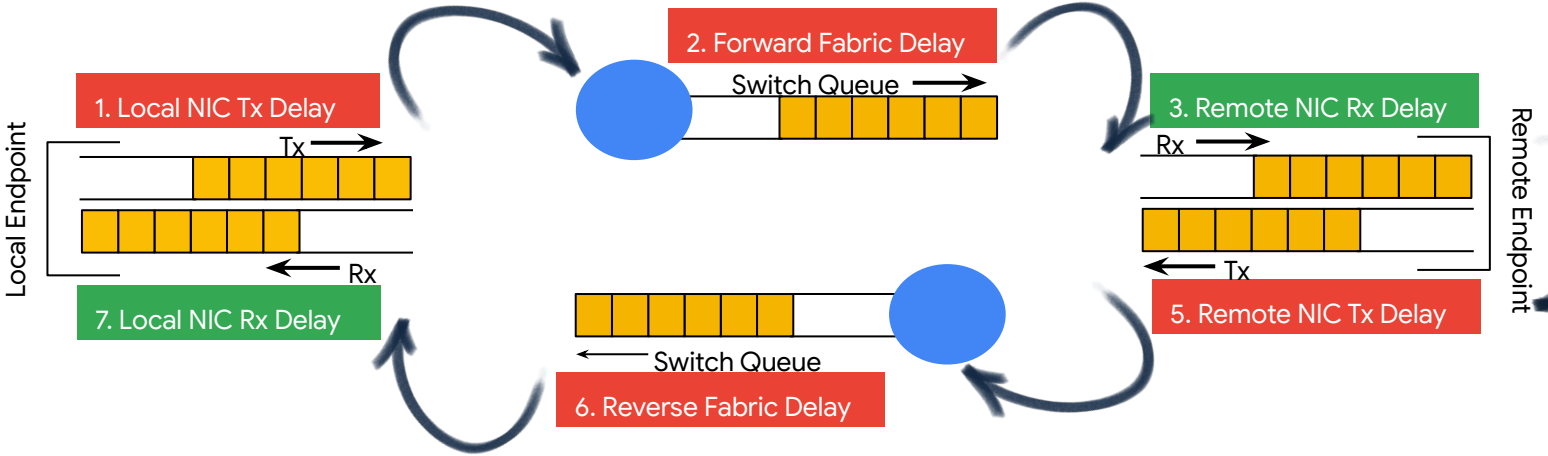
- Scaling state to millions of connections
  - Alternative to TLS offload
  - Hybrid: offload top-N flows capturing X ~= 98% of traffic
  - Stateless: O(1) scaling
- Flexible encapsulation format
  - Carry inner entropy, metadata
  - Partial confidentiality to enable network telemetry
- Accelerate QUIC with crypto offload

NETWORKING



Transport

- Improvements to congestion control algorithms requiring HW assist
- Access to high precision timing events
  - https://dl.acm.org/doi/pdf/10.1145/3387514.3406591



Operations at scale

NETWORKING

- Uniform visibility
  - Flow counters
  - Packet sampling
  - Drop counters
  - Identify NIC bottlenecks
  - Metrics for identifying misbehaving workloads
- Fleet health management
  - Self-tests
  - Device error reporting
- Acceptance criteria
  - Testing frameworks
  - Scale and stress tests

OPEN POSSIBILITIES.

Join the conversation

- Operator and vendors need to come together
- Is OCP the right forum for this?

Open Discussion