UEC 1.0 – Technical Deep Dive

Ultra Ethernet Consortium

OCP APAC SUMMIT | 5-6 AUGUST, 2025
TAIPEI, TAIWAN

Technical Deep Dive – UEC 1.0

NETWORKING

J Metz, Ph.D

Chair / Ultra Ethernet

THE NEW ERA NEEDS A NEW NETWORK

Ultra Ethernet

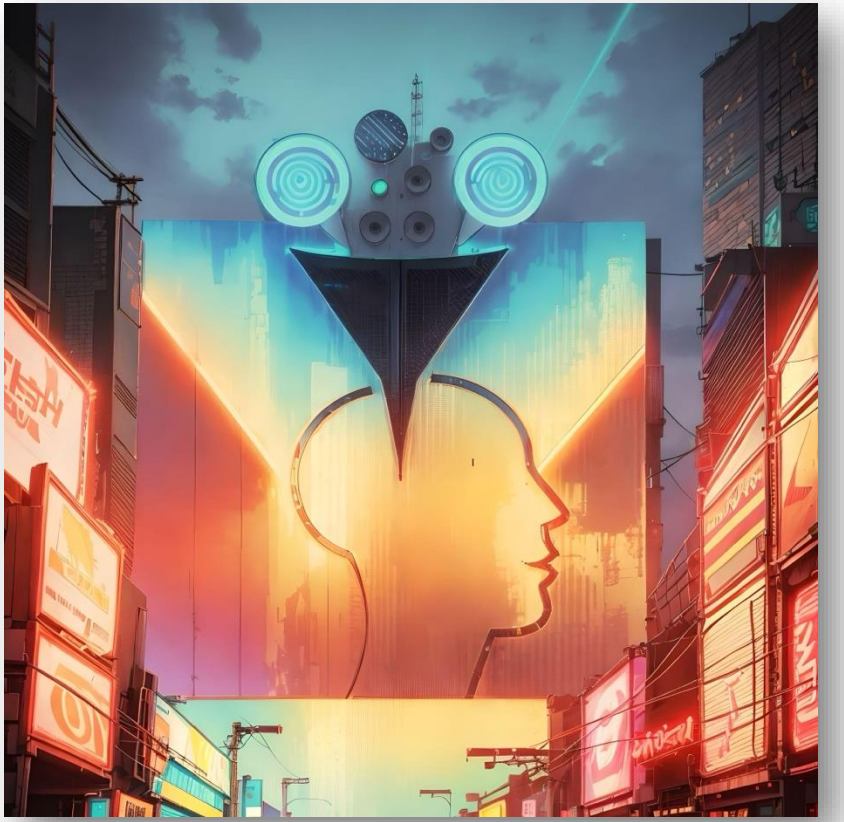
As **performant** as a supercomputing interconnect

As **ubiquitous** and **cost-effective** as Ethernet

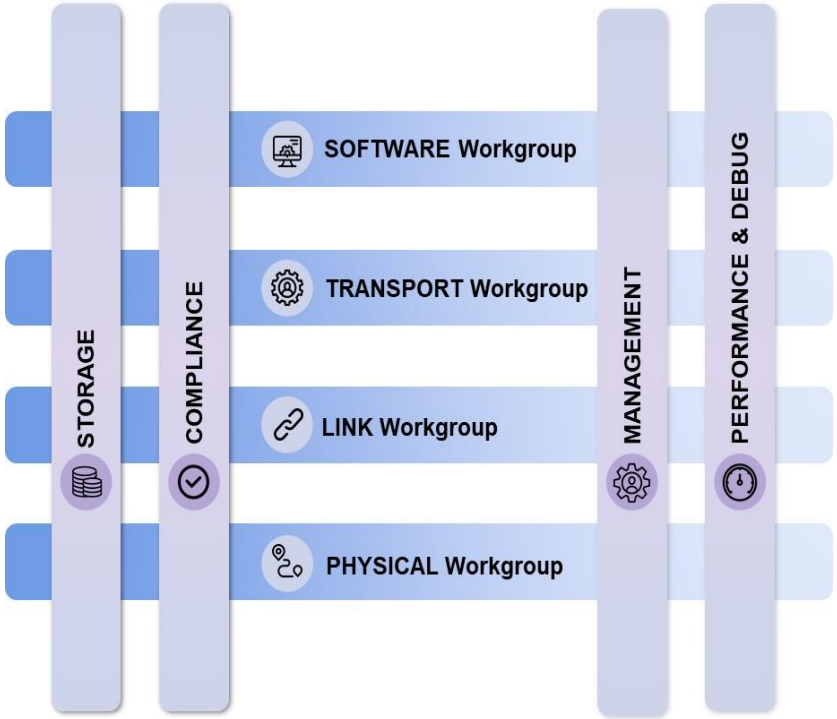
As **scalable** as a cloud data center

Outline

1. Who is UEC, and Why?
2. The UEC Philosophy/Solution
3. The UEC Theory of Operation
4. Summary

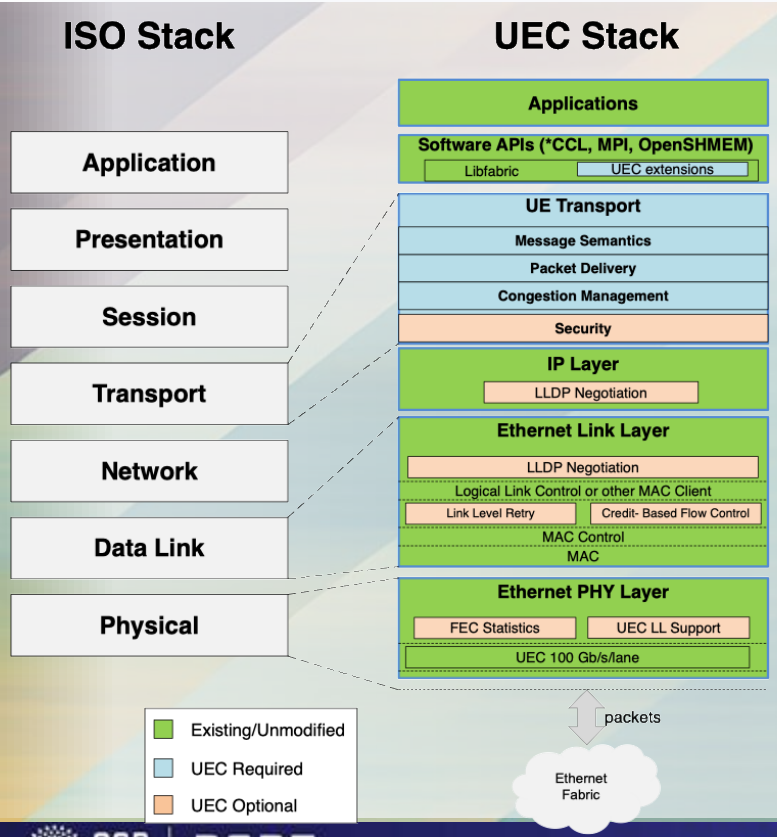


Who is UEC? And Why?



- Full Standards Development Organization
- (One of the?) Fastest growing projects in Linux Foundation
- 130+ Companies
- 1500+ individual active contributor volunteers
- 8 Workgroups
  - Physical
  - Link Layer
  - Transport
  - Software
  - Storage
  - Management
  - Compliance & Test
  - Performance & Debug

UEC – Standing on the Shoulders of Giants



- Backwards-compatible
  - Uses libfabric as its north-bound API
  - Designed to integrate into existing frameworks where libfabric is commonly utilized
- Key driving force is in the Ultra Ethernet Transport (UET)

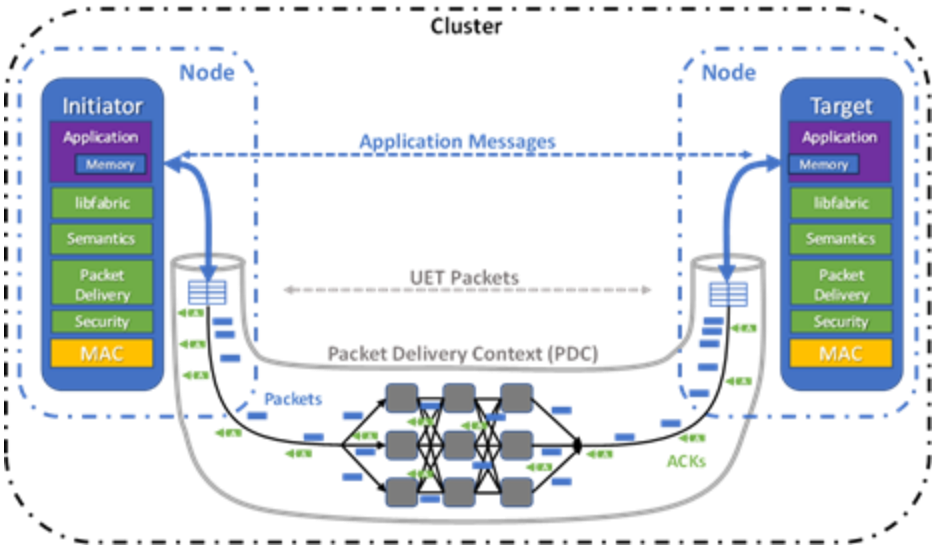
- Software API
  - Libfabrics 2.0 with extensions
  - New Transport Layer
  - Multi-pathing Packet spraying
  - Ordered (ROD) and un-ordered (RUD)
  - Lossy (no PFC) or Lossless
  - Congestion Control: Enhanced Tx and new Rx
  - Trimming
  - In Network Collective

- Network Layer
  - IP v4/v6
  - ECN
- Data Link Layer
  - Negotiation – LLDP
  - Link Level Retry - LLR
  - Header Efficiency Improvements
- Physical Layer
  - IEEE Compliant 100G Signaling
- AI and HPC Profiles

More System View and Nomenclature

More important concepts

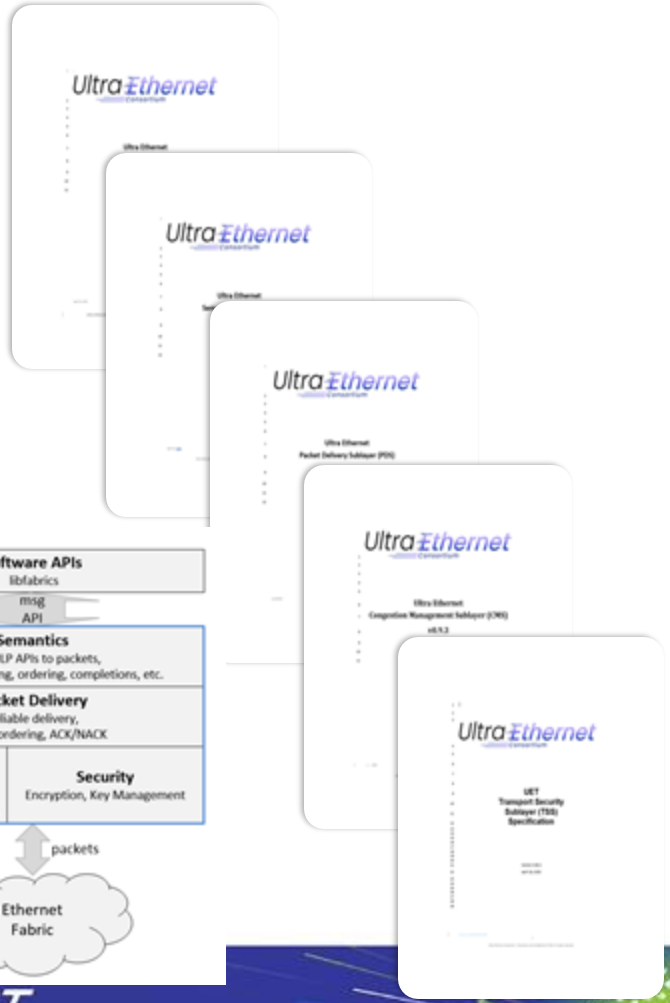
- SES exchanges **messages**
- PDS exchanges **packets** over Packet Delivery Contexts (**PDCs**)
- PDCs are ephemeral connections and have multiple reliability modes
  - **ROD, RUD, RUDI, UUD**
- CMS allows packets spraying across the fabric
- The fabric can be **lossless** or **best-effort**
- Link and PHY reliability features aim to eliminate retransmissions



Transport

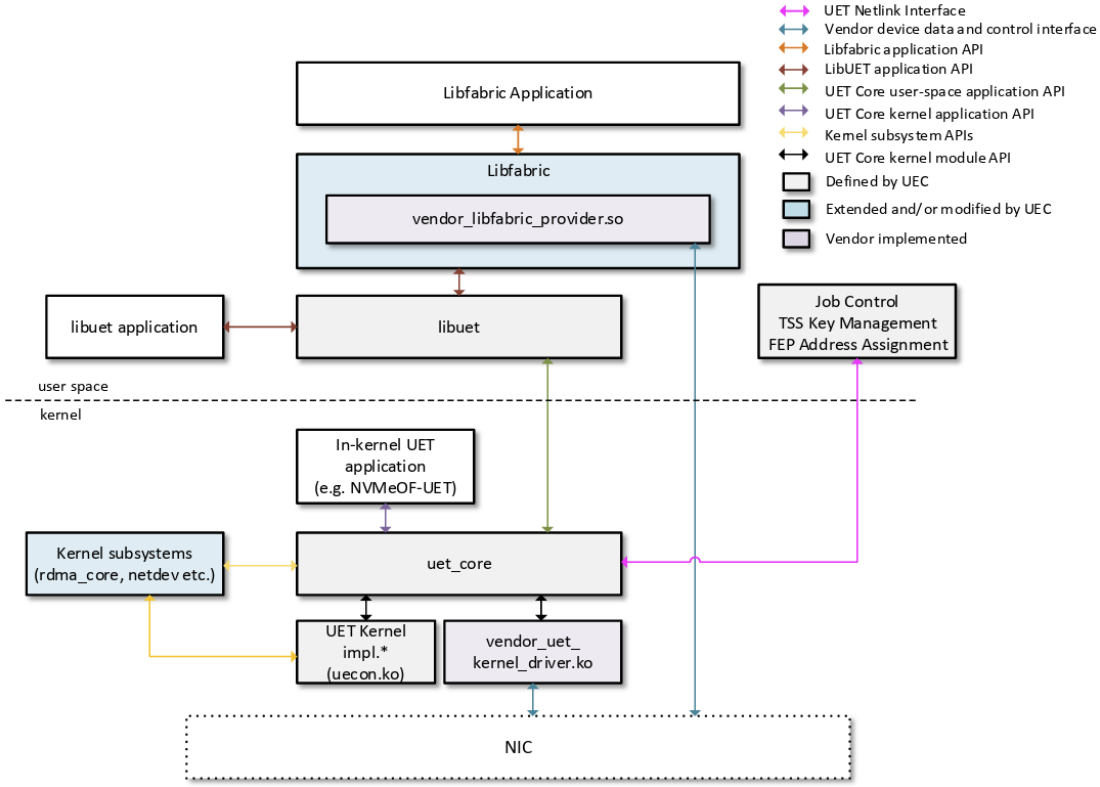
- Transport Overview
  - Semantics Sublayer (SES)
  - Packet Delivery Sublayer (PDS)
  - Congestion Management Sublayer (CMS)
  - Transport Security Sublayer (TSS)

- Includes the core sub-layered architecture and protocol specifications while providing implementation flexibility
- Tightly coupled with software



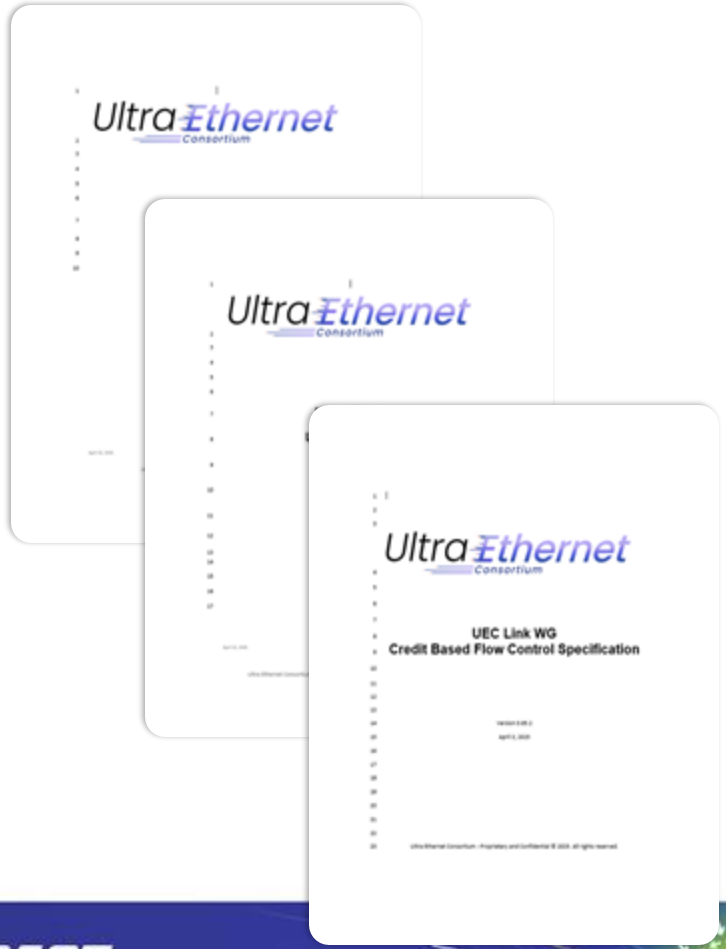
Software

- SW Overview
  - Introduces SW components and interfaces
  - Provides pointers to UE OpenSource reference software
- Libfabric Mapping
  - Defines the mapping and requirements of libfabric v2.0 APIs on a UET provider
  - Describes the intricacies of libfabric binding to UET addressing schemes and packet delivery modes



Link Layer

- Link Layer Retry (LLR)
  - Defines a sublayer and preamble for standard Ethernet interfaces that provides lossless operation at the link level by supporting retransmission of lost frames.
  - Works with existing PFC and new CBFC flow-control mechanisms.
- Link Negotiations
  - Defines a negotiation protocol and YANG management objects on top of LLDP for UE link level features.
- Credit-based Flow Control (CBFC)
  - Specifies an alternative to PFC (but can co-exist)
  - Allows up to 32 virtual channels for individual lossless operation.



Physical Layer

- Specifies UE 100G per lane signaling
- Specifies FEC statistics for prediction of UE link quality
- Specifies UE changes to the PHY layer to support LLR and CBFC

Modern Transport and RDMA Services for AI and HPC

Requirement	UEC Transport	Legacy RDMA	UEC Advantage
Multi-Pathing	Packet spraying	Flow-level multi-pathing	Higher network utilization
Flexible Ordering	Out-of-order packet delivery with in-order message delivery	N/A	Matches application requirements, lower tail latency
AI and HPC Congestion Control	Workload-optimized, configuration free, lower latency, programmable	DCQCN: configuration required, brittle, signaling requires additional round trip	Incast reduction, faster response, future-proofing
In Network Collective	Built-In	NONE	Faster Collective operation, lower latency
Simplified RDMA	Streamlined API, native workload interaction, minimal endpoint state	Based on IBTA Verbs	App-level performance, lower cost implementation
Security	Scalable, 1st class citizen	Not addressed, external to spec	High scale, modern security
Large Scale with Stability and Reliability	Targeting 1M endpoints	Typically, a few thousand simultaneous end points	Current and future-proof scale

Call to Action

- Where to download the specification:
  - https://www.ultraethernet.org

- How to join UltraEthernet:
  - https://www.ultraethernet.org/membership

Open Discussion