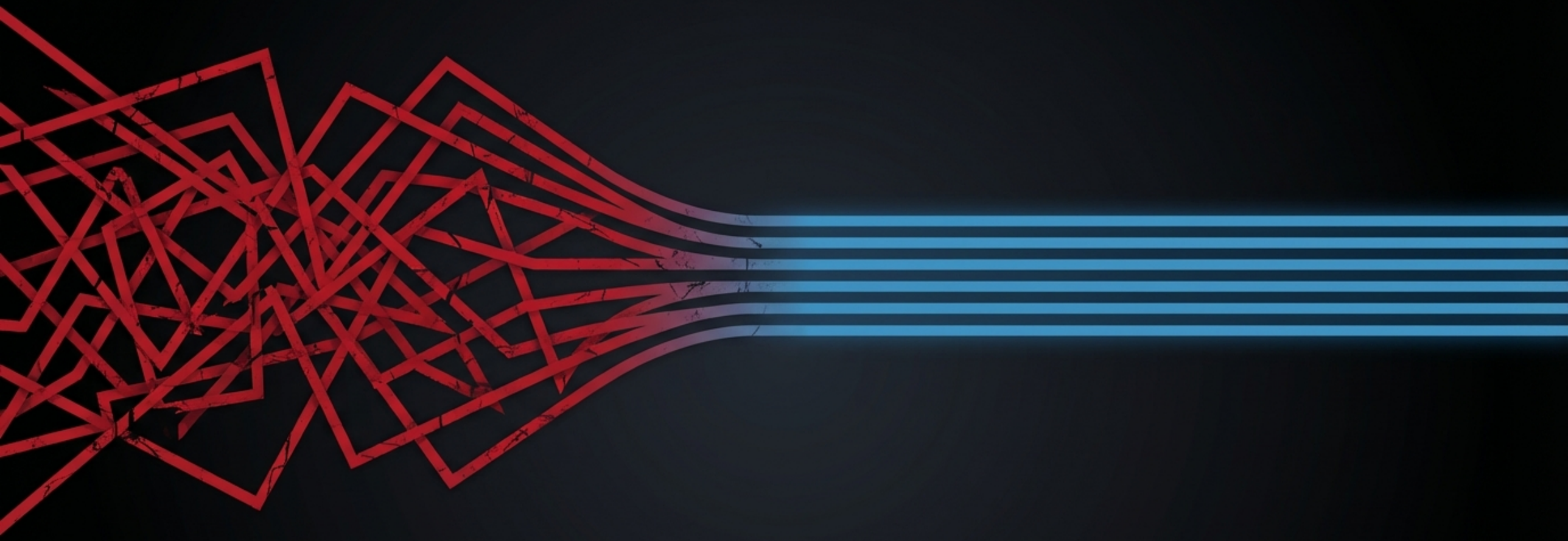


The Deterministic Storage Fabric

ERADICATING NVMe-oF INCAST CONGESTION FOR ENTERPRISE AI | PTCP PILOT AUTHORIZATION READOUT

The Executive Summary

10-15%

GPU Cycle Time Wasted

Multi-million-dollar Exascale compute silicon sits in idle iowait states waiting for dropped storage packets.

50-200ms

Tail Latency Penalties

TCP retransmission timeouts completely stall AI application execution and model training.

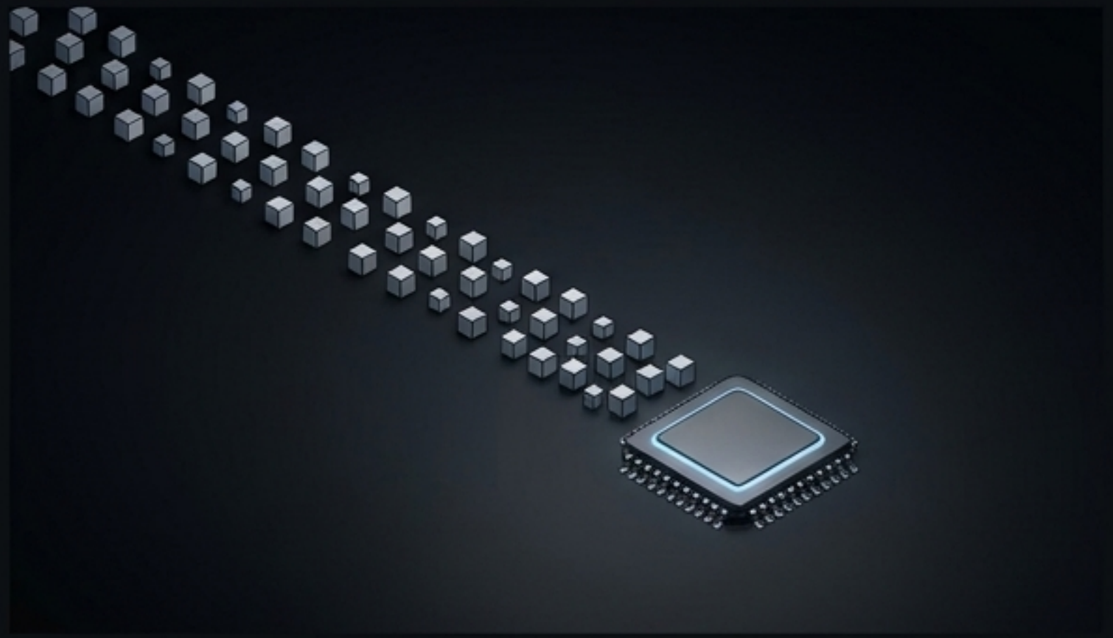
$0

Required Hardware Refresh

Software-defined mathematically prevents packet drops without ripping and replacing existing Ethernet switches.

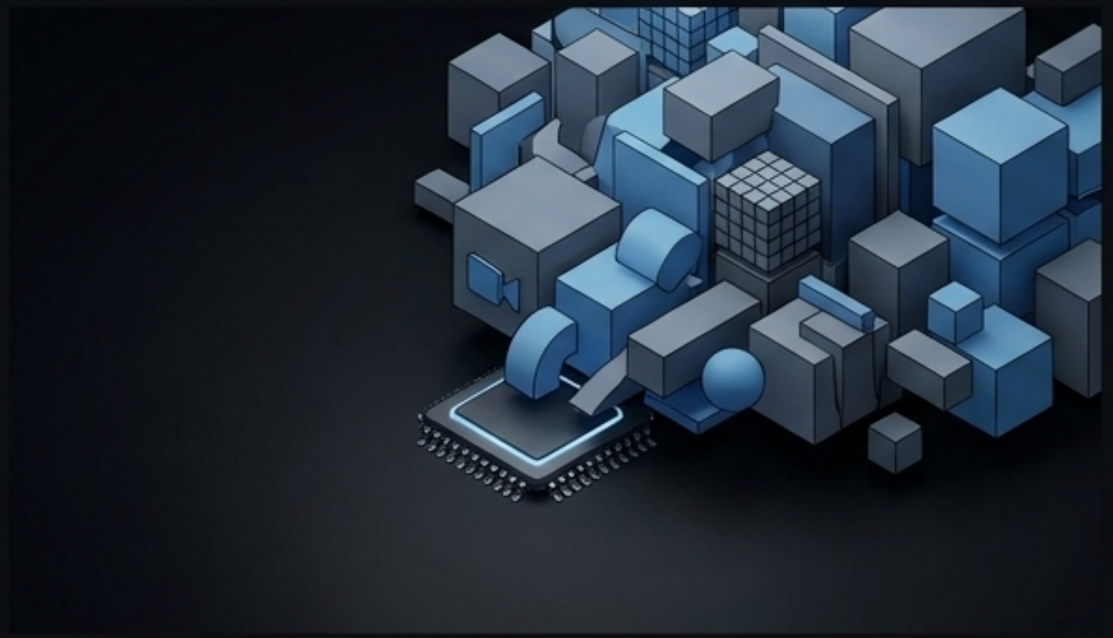
The AI & RAG Data Imperative

Legacy LLMs



Text-based LLMs operate within predictable, lightweight context windows.

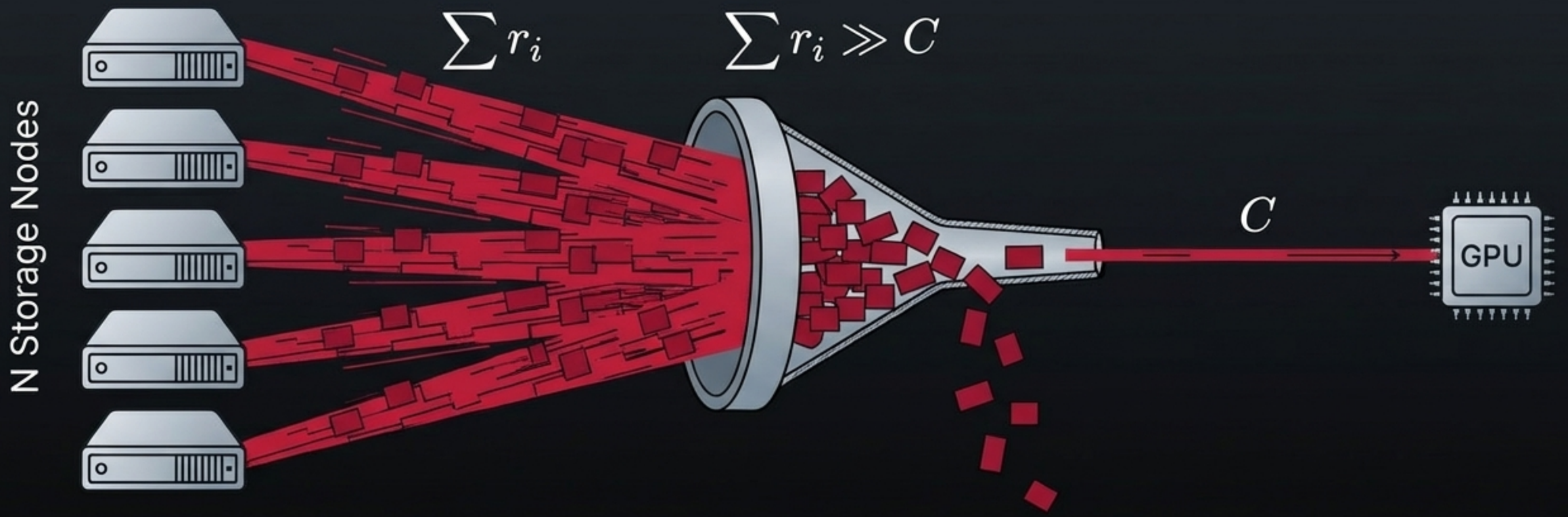
The Multimodal RAG Imperative



Modern Retrieval-Augmented Generation (RAG) requires concurrent semantic search across billions of vectors.

The Time To First Token (TTFT) is fundamentally dictated by storage latency. If even one distributed storage shard is delayed by a 200ms retransmission, the entire inference prompt stalls.

The Physics of Failure



The Incast Microburst

Distributed RAG queries force massive, simultaneous responses from all-flash NVMe arrays.

The Reactive Failure

Standard TCP and RoCEv2 rely on shallow Ethernet switch buffers. When buffers overflow, TCP drops packets, triggering lic AIMD sawtooth collapse.

The HoL Block

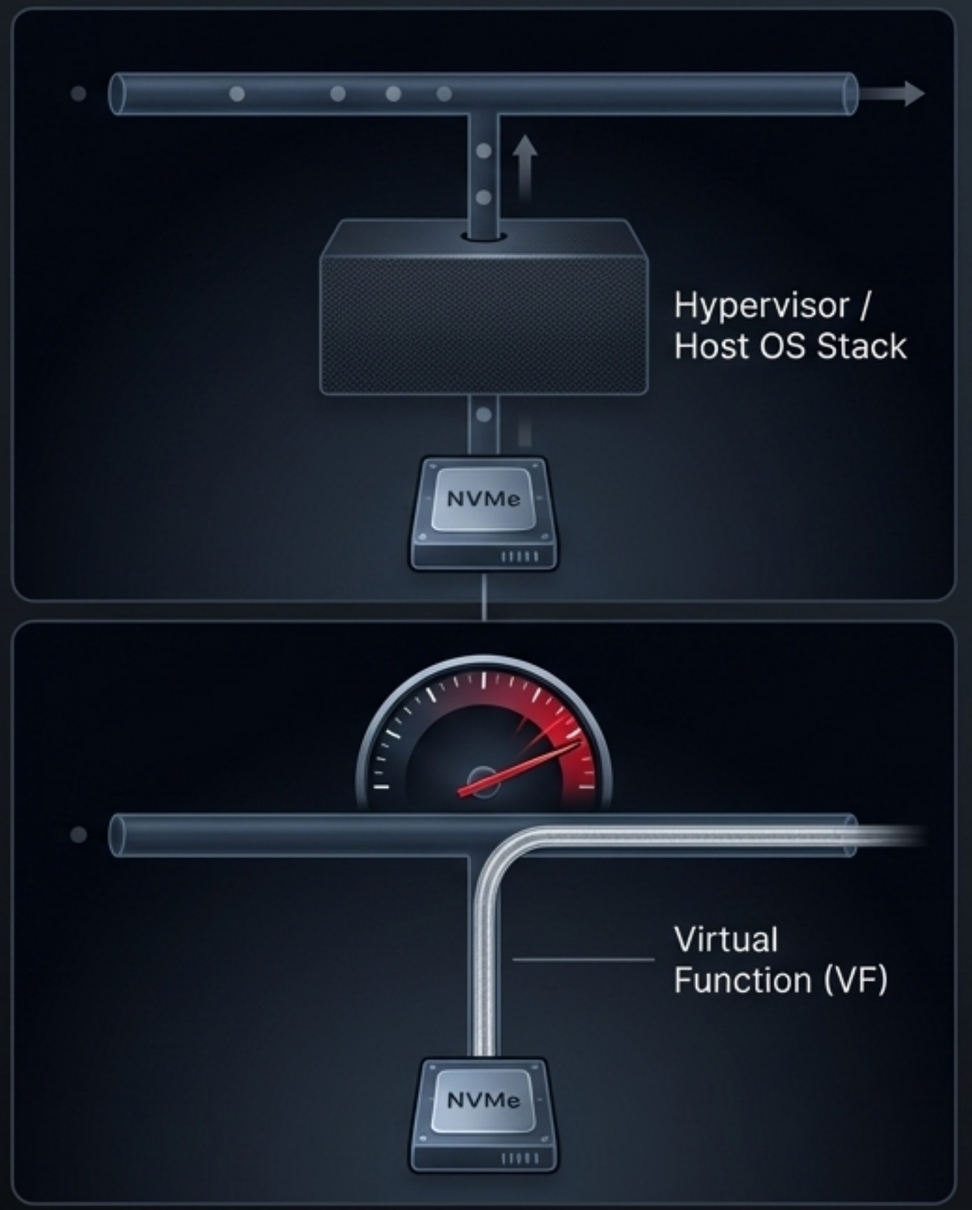
Even lossless networks suffer Priority-based Flow Control (PFC) storms that freeze the entire fabric.

The SR-IOV Paradox

To feed our AI clusters, we deployed Single Root I/O Virtualization (SR-IOV) to bypass host OS bridging and accelerate throughput.

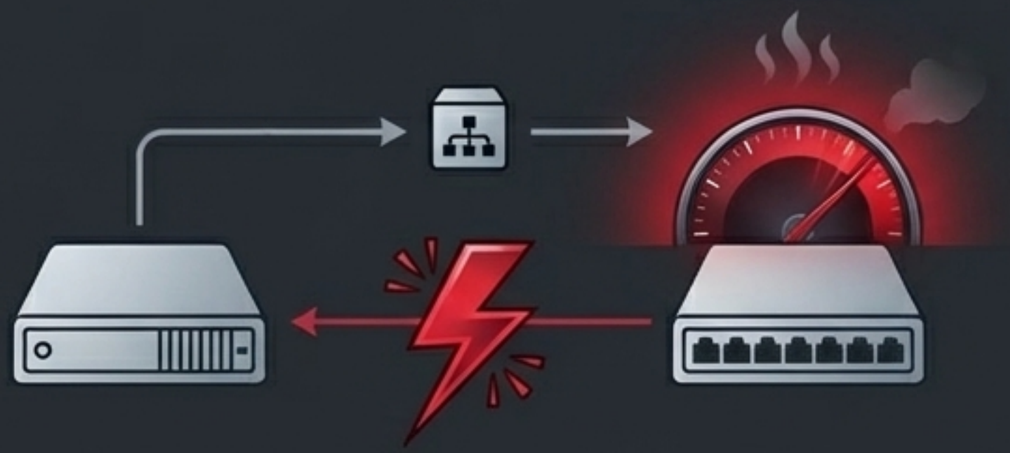
The Paradox:

Because SR-IOV is so brutally efficient, it pushes parallel storage data to the wire instantly. This eliminates local latency but rapidly accelerates the Many-accelerates the Many-to-One microburst, overwhelming the ToR switch buffers even faster.



Paradigm Shift: Reactive vs. Predictive

The Reactive Network (Legacy)



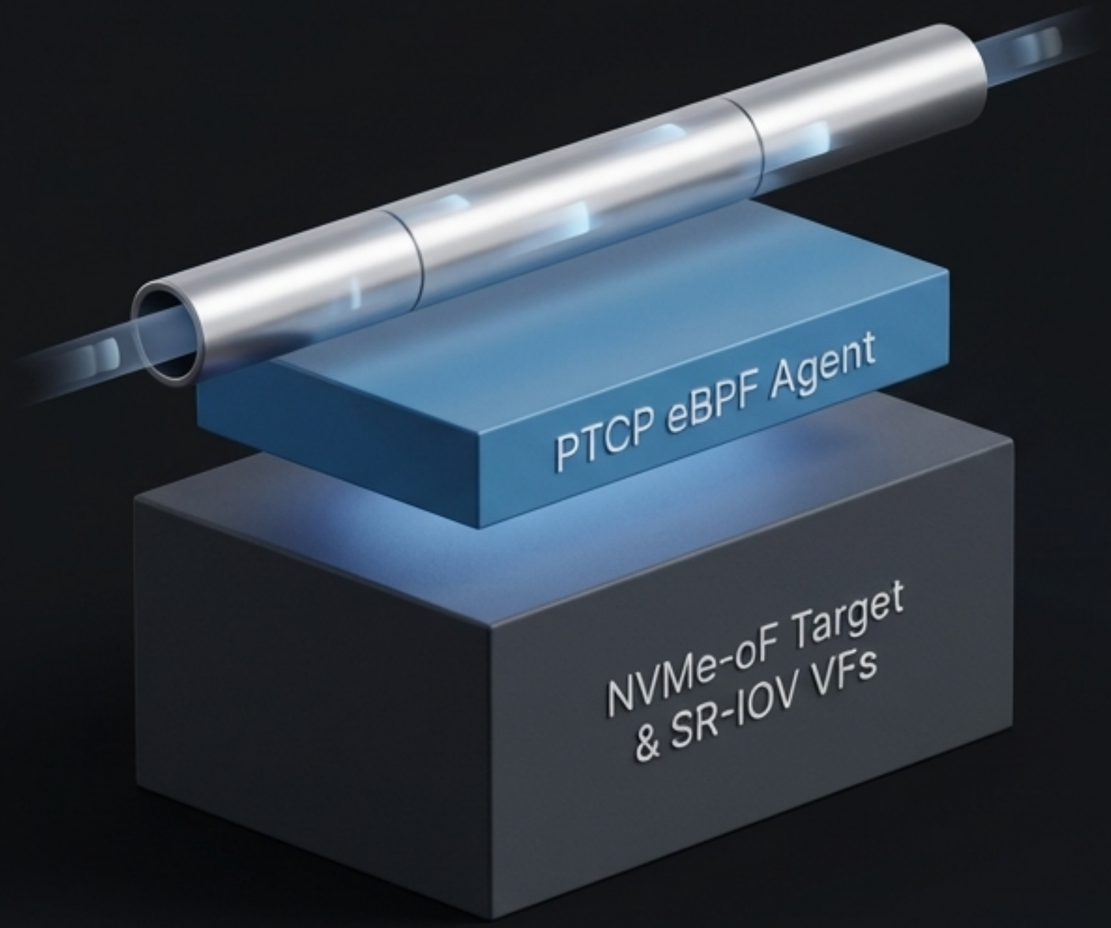
Congestion control is managed after the crash occurs.

The Predictive Edge (PTCP)

Congestion is mathematically prevented before data hits the wire.

Move intelligence away from the physical switches and embed it directly into the Linux kernels of the storage controllers.

Tensor Networks' Predictive Tensor Control Plane (PTCP)



- **Microscopic Footprint**

  We inject compiled eBPF C-code directly into the Traffic Control (TC) layer of the storage array's Virtual Functions.

- **O(1) Offload**

  Operates natively at the Linux netdev layer, maintaining the strict CPU-offload benefits of SR-IOV.

- **Hardware Agnostic**

  Requires zero proprietary switch ASICs; executes entirely on the host edge.

The Mechanics of Earliest c Departure Time (EDT)

1. The PTCP Orchestrator utilizes a Pattern-of-Life Tensor Train (PoL-TT) model to calculate the exact clearing rate of the ToR switch.

2. The eBPF agent intercepts Socket Buffers (`skb`) and manipulates the `tstamp`.

3. An Earliest Departure Time (EDT) schedule is applied, interleaving data precisely.

Autonomous Zero-Trust Storage Security

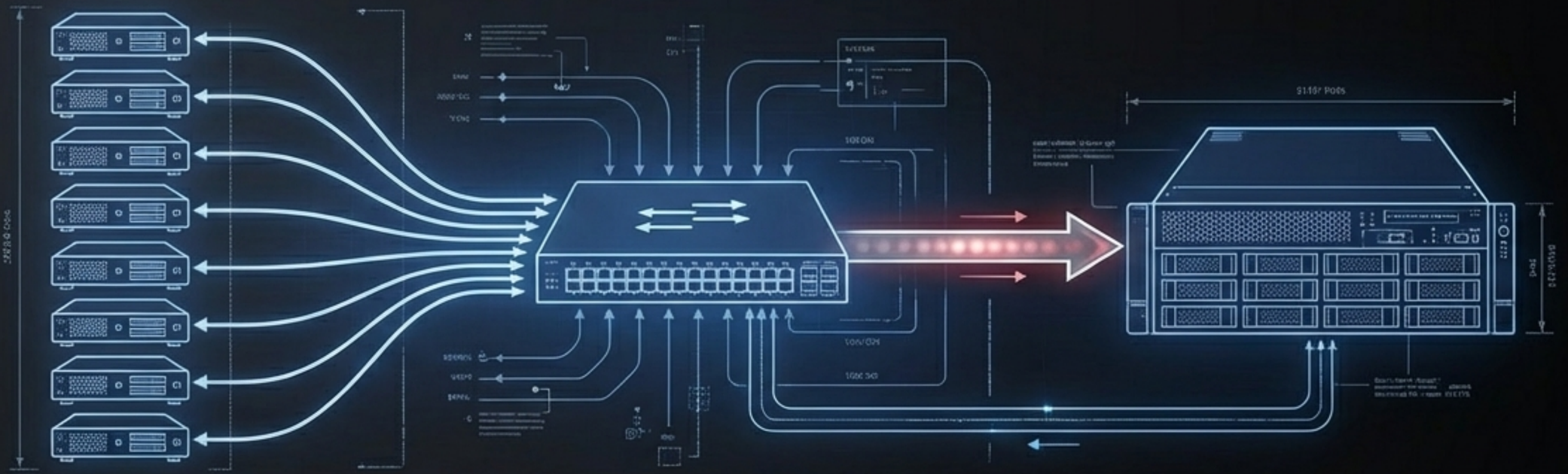
Storage arrays are the ultimate target for ransomware and lateral exfiltration.

Storage arrays are the ultimate target for ransomware and lateral exfiltration. Whenever he reconmented the security condenant can ireent than and stande more arm export.

The Mechanism: Because PTCP's **PoL-TT models the exact flow state of vector queries**, any unauthorized lateral data dump or anomalous mass encryption mathematically violates the baseline.

The Benefit: The eBPF agent autonomously severs the connection at the kernel layer in nanoseconds—containing threats instantly without the latency penalty of inline firewalls.

The Lab Validation Methodology



8 Compute Initiators
running fio + io_uring

100GbE Broadcom ToR Switch
(No PFC/ECN)

Bare-metal NVMe-oF Storage Target with
SR-IOV Virtual Functions (ens1np0v0)

Execution Strategy: We intentionally trigger a devastating Many-to-One Incast event using highly concurrent, random 16KB reads (Queue Depth 128) to force switch buffer overflow. The 45-day PoC captures the raw network destruction of legacy TCP versus the mathematical determinism of PTCP.

Telemetry Synthesis: The Network Weather

baseline_tc_stats.log & Wireshark Pcap

```
qdisc pfifo_fast 0: dev ens1np0 root
   Sent 85938502 bytes 59385 pkt
   dropped 8459 overlimits 1204 requeues 0

TCP Retransmission: True
Duplicate ACK: True
```

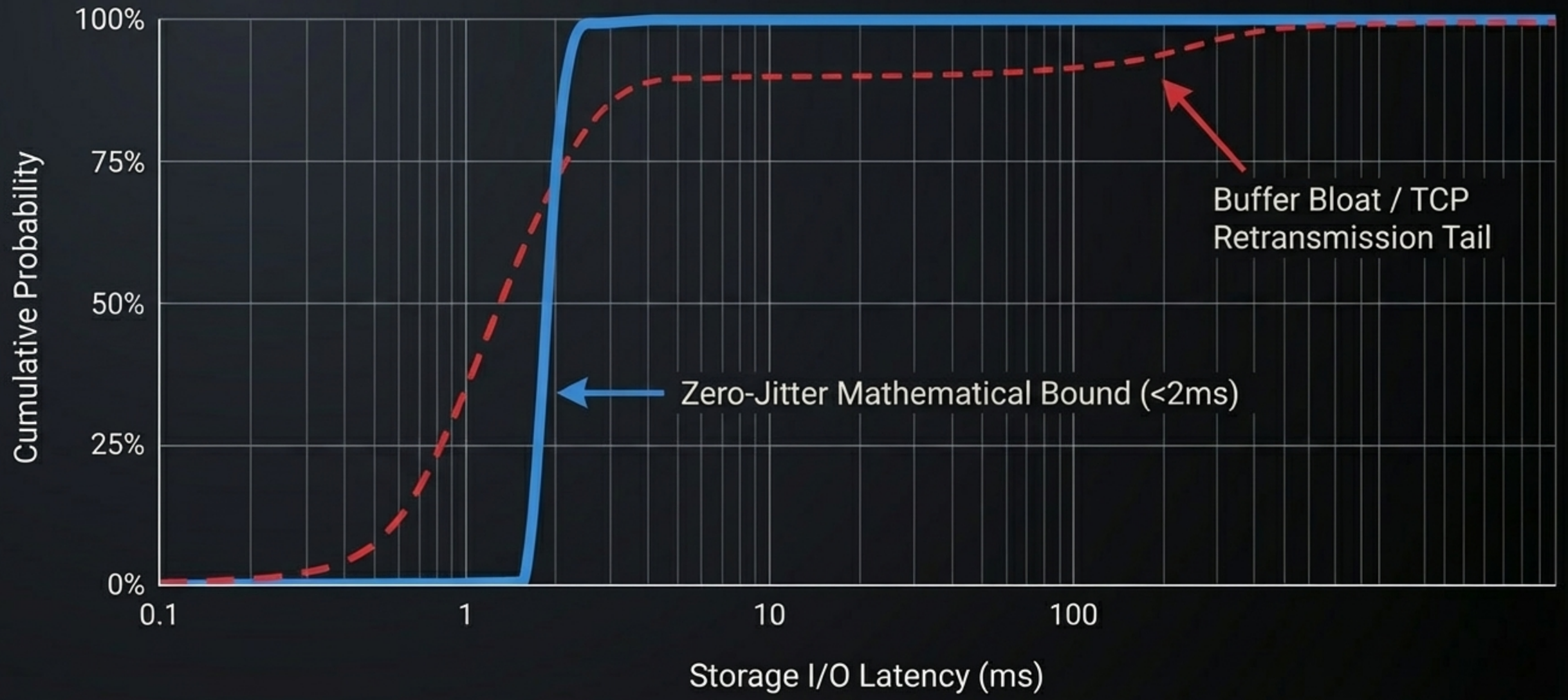
Baseline triggers the AIMD TCP sawtooth **collapse**. Severe packet drops.

ptcp_tc_stats.log & Wireshark Pcap

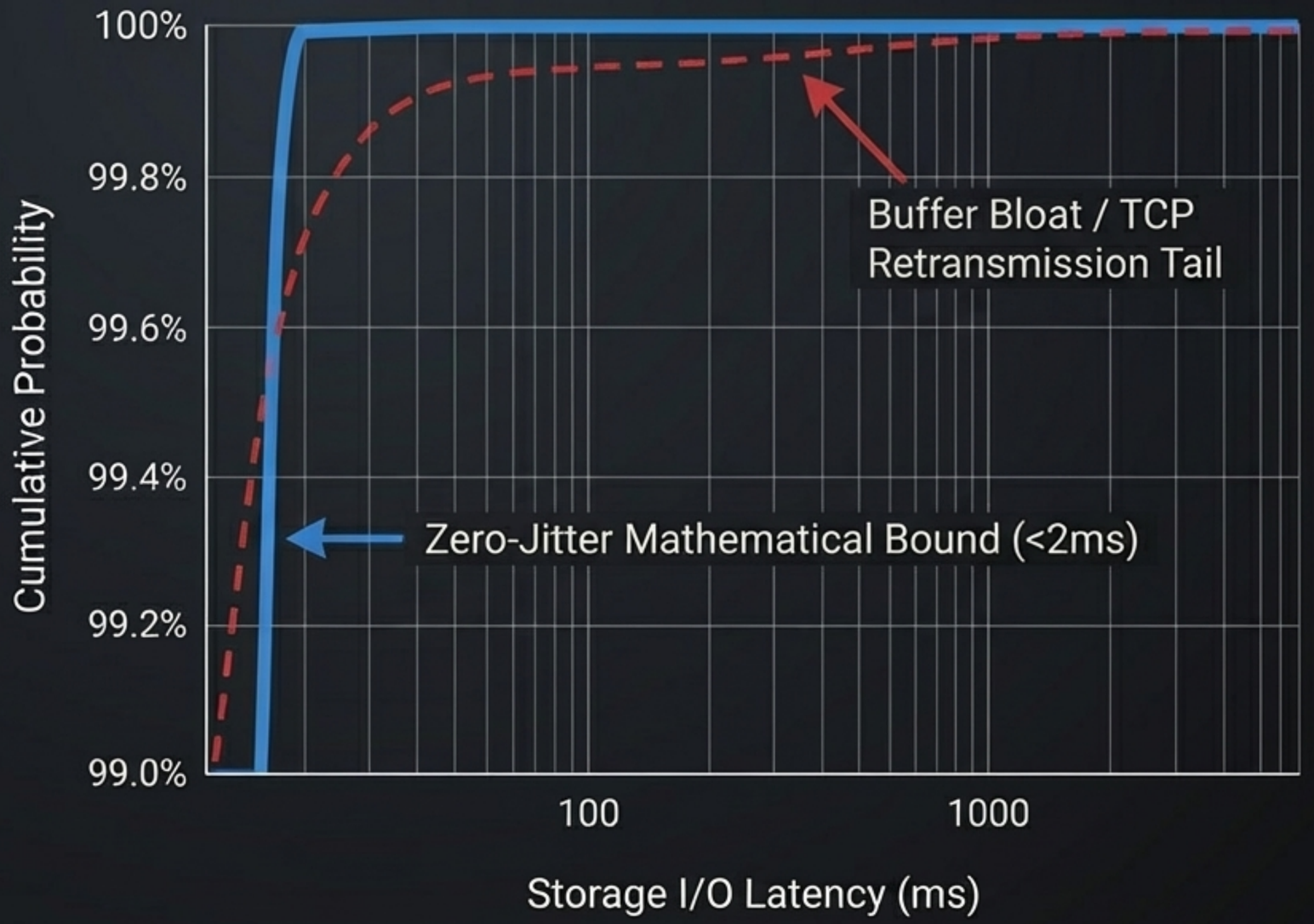
```
qdisc fq 0: dev ens1np0v0 root
   Sent 85938502 bytes 59385 pkt
   dropped 0 overlimits 0 requeues 0

[Filter: tcp.analysis.retransmission] -> 0 Packets
```

PTCP yields zero switch packet drops. Stream behaves like a congestion-free optical pipe.

The Executive “Kill Shot”



Interpreting the Tail: Reclaiming Exascale Yield



The 50% Mark

At the median, standard Ethernet performs adequately. Averages hide the failure.

The Red Tail

That 0.1% of delayed storage requests represents multi-million-dollar AI clusters sitting completely idle, waiting for data.

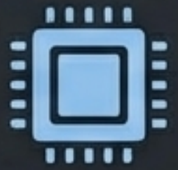
The Blue Wall

PTCP eradicates this weather entirely. 99.99% of packets arrive with identical, bounded latency, guaranteeing that expensive GPUs are continuously fed and never starved for data.

Synthesis Matrix: Legacy NVMe/TCP vs. PTCP

Feature / Metric	Legacy NVMe/TCP (Baseline)	PTCP Determinism
99.9th Tail Latency	50ms to 200ms+ (Variable)	Strictly <2ms (Bounded)
Incast Packet Drops	Severe (TCP Sawtooth Recovery)	Zero (Mathematically Prevented)
ToR Switch Requirement	Deep Buffer ASICs ($$$)	Standard Shallow Ethernet ($)
Target CPU Overhead	High (Due to Retransmissions)	Negligible (O(1) eBPF Offload)

Financial Impact & ROI Vectors



OpEx: GPU Yield Reclamation

Eradicating iowait ensures our Exascale compute silicon is saturated. A 10% reduction in idle time equates to millions reclaimed in compute efficiency and power.

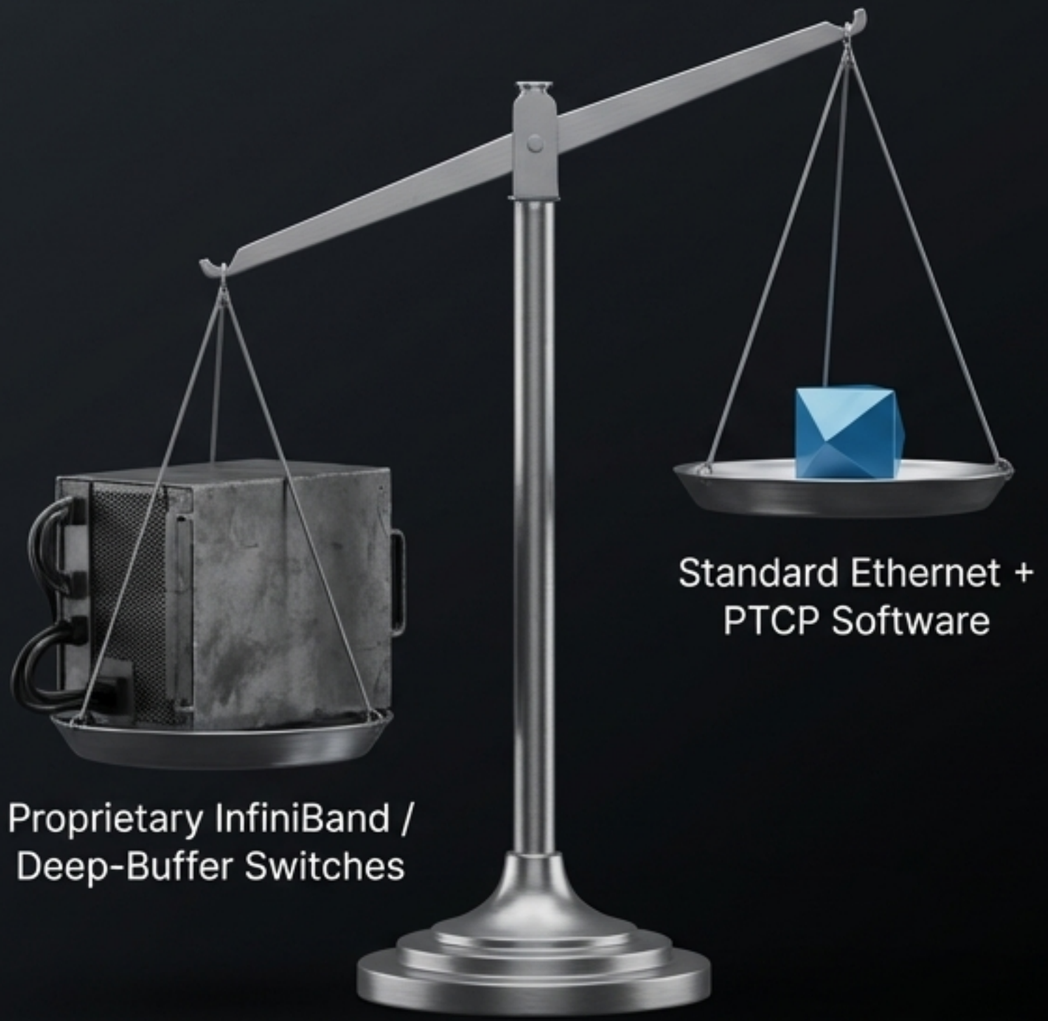
CapEx: Network Hardware Deferral

Achieves lossless, deterministic routing over our existing commoditized commoditized Ethernet footprint, saving up to 40% in planned network upgrades.

Top-Line: Application Performance

Sub-millisecond maximum latencies provide instantaneous, synchronous responses for enterprise Copilots, RAG chatbots, and multimodal video generation.

“The InfiniBand Avoidance”



Historically, solving storage Incast required ripping out standard networks and replacing them with highly proprietary NVIDIA InfiniBand or ultra-deep-buffer ASICs.

The Bottom Line: By pacing data predictively at the source, PTCP delivers InfiniBand-level determinism natively over commodity Broadcom Ethernet. We fix the physics problem with a software upgrade, not a hardware rip-and-replace.

Pilot Scope & Execution

Week 1

Provisioning

Allocate SR-IOV Virtual Functions (`ens1np0v0`) on the Enterprise Storage Lab target.

Bind NVMe-oF target.

Weeks 2-4

Deployment & Incast

Inject `PTCP` Go daemon via `tc qdisc clsact`.

Run massive concurrent `fio` microbursts across 4-8 initiators.

Weeks 5-6

Telemetry Readout

Capture `tcpdump` and `tc` metrics.

Generate internal CDF graphs proving zero switch packet drops and collapsed tail latency.

Pilot Authorization Request

Statement of
Work (SOW)

Phase 1
Licensing Budget

Action Requested:
Authorization of the Phase 1 software licensing budget
to initiate the PTCP NVMe-oF lab validation.

The transition to native multimodal AI requires a fabric that operates at the absolute limits of physical hardware. We **cannot build the next generation** of enterprise AI on a storage network that drops data.