Towards Truly Burst-Aware Evaluation of Data Center Congestion Control

Pragna Mamidipaka, Srikanth Sundaresan, Theophilus A. Benson

The 10th Asia-Pacific Workshop on Networking, 2026

Poor CCA behavior impacts application performance

Production rollback to legacy baseline

[1] DCTCP disabled at Meta

10% lower throughput
4.5× retransmissions

[1] Data-intensive read service

65→40 ms p90
130→80 ms p99

Modern CCAs

[2] Swift vs. Poseidon for message transfer workloads

14.5× higher
p99.9 message time

[3] AI training on 400G RoCE

9.6% training throughput
left unrealized

[1] A large-scale deployment of DCTCP, NSDI 2024. [2] Poseidon: Efficient, Robust, and Practical Datacenter CC via Deployable INT, NSDI 2023. [3] Barre: Empowering Simplified and Versatile Programmable Congestion Control in High-Speed AI Clusters, ATC 2025.

Today’s evaluation workload choices are CCA-agnostic

General workflow

- Pick some workloads (background, incast flows)
- Run simulations / testbed eval
- Report metrics (FCT, throughput)

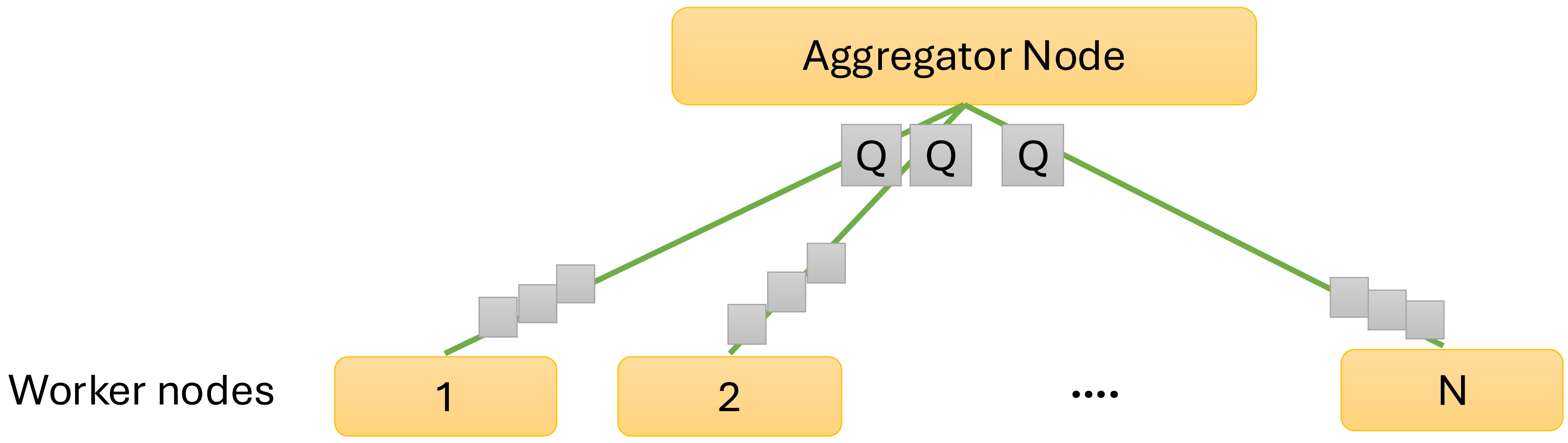
What this misses

- ❑ Workloads are usually CCA-agnostic
- ❑ Results do not transfer to new workloads without re-running
- ❑ Operator’s worry - unrepresentative evaluation.

Research question: How can we systematically choose workloads that cover distinct behaviors of the CCA?

Incast burst pattern in datacenters

Common traffic pattern arising from partition aggregate workflows.



CCAs handle incast bursts, and this is crucial for good application performance.

Data Center Networking Papers Survey

Survey scope We survey ~30 datacenter networking papers from last six years.

How evaluation workloads are constructed

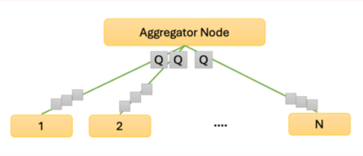
Background traffic

- Sample flow sizes from a published CDF
- Poisson arrivals
- Random source and destination

+

Optional incast overlay

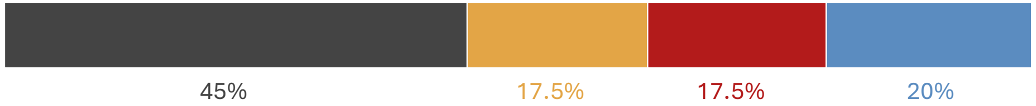
- N simultaneous senders
- X KB from each sender
- repeats at QPS



Evaluation workload = CDF-based background + optional incast overlay

"Burst-aware workloads"

Survey results



Canonical background only	13 papers · 45%	Bursts occur only by chance
Background + fixed incast	5 papers · 17.5%	Only one operating point
Background + varied incast	5 papers · 17.5%	More expressive, but still ad hoc
Fully synthetic	6 papers · 20%	Custom microbenchmarks / stress scenarios

62.5% papers either omit incast or evaluate only one fixed incast point

Burst-aware workloads still miss production temporal clustering

What is being compared?

Production ground truth Millisampler

Burst-aware workloads from papers

Analysis

1 ms utilization time series

Markov r-value
= temporal clustering of high-utilization intervals

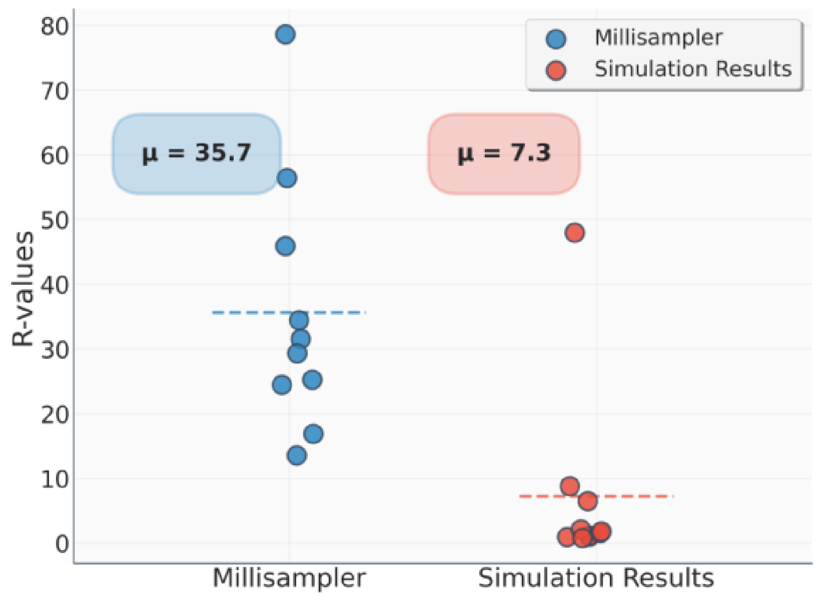


Figure 1: Millisampler hosts exhibit higher r values.

Interpretation: Production bursts are rare but persistent; simulated burst-aware workloads create bursts, but with weaker temporal clustering.

Vision: CCA evaluations should explore all CCA regimes

Arbitrary workloads

Can the results transfer cleanly to a different workload?

Which CCA behaviors were covered and missed?

Regime Centric CCA eval

Analyze the CCA to identify behavioral regimes

Generate bursts that place workloads in regimes

Compare CCAs and workloads at regime level

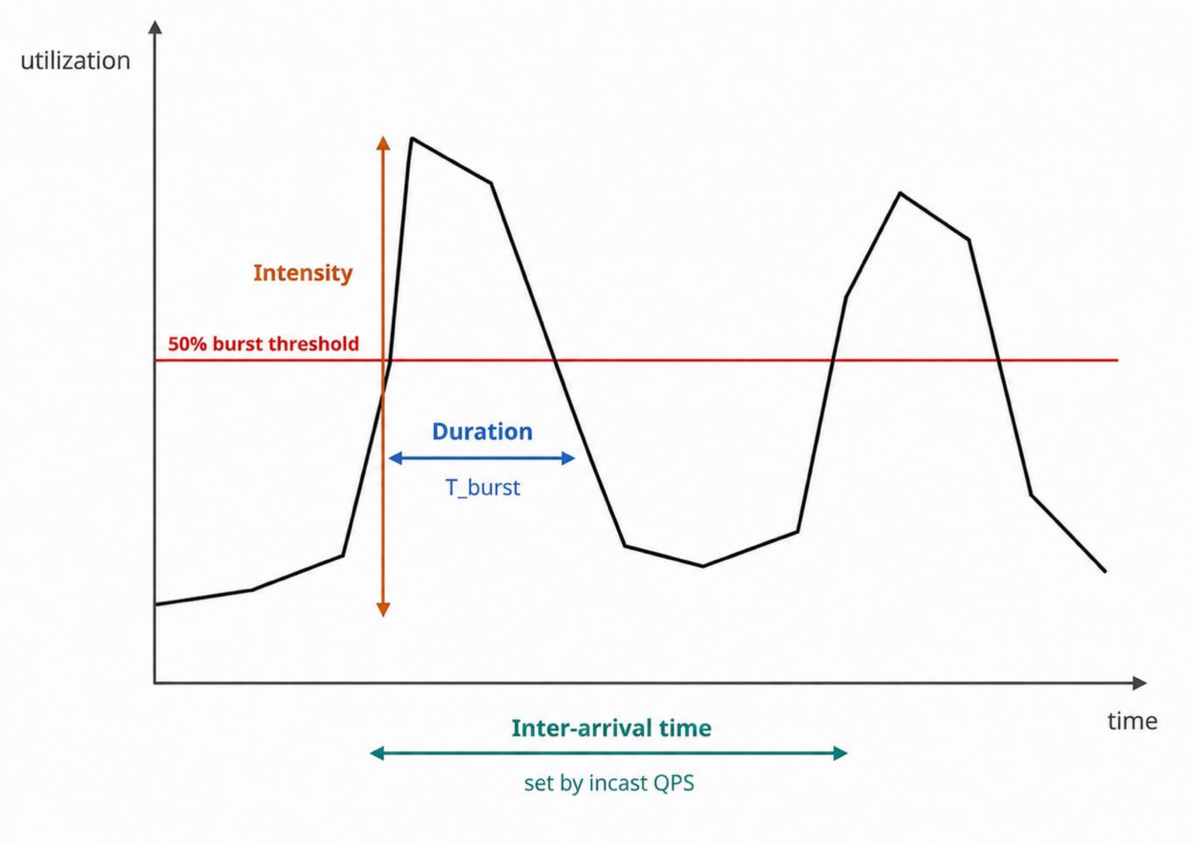
Regime-centric evaluation enables -

Transfer to new workloads

Operator representativeness

Fair CCA comparison

A burst has four dimensions that stress parts of a CCA



Four axes derived from the burst episode

Intensity
Data injected during one RTT at burst onset

Duration
How long the burst episode lasts
T_burst

Synchronicity
Whether senders are phase-aligned

Inter-arrival time
Time between successive burst episodes

These dimensions provide the axes for identifying CCA-specific behavioral regimes.

DCTCP regimes in Intensity dimension

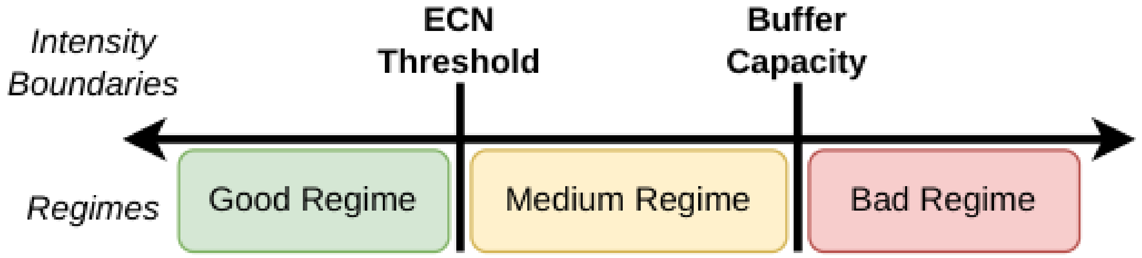
Intensity: how much data is injected in a single RTT?

$$I = N \times cwnd$$

N = number of incast senders/flows

cwnd_init = initial congestion window per flow

cwnd_min = minimum congestion window per flow



Intensity asks how many concurrent DCTCP senders are enough to cross ECN or buffer boundaries

Related work

Workload realism

Encore: flow-size CDFs lose temporal and spatial structure. Does not talk about bursts.

Millisampler / incast measurements: bursts and temporal clustering matter in production.

Automated evaluation

Active learning like **Mahak** approaches search parameter spaces, but do not explain failure boundaries.

CCA modeling

Contracts framework, **DCTCP** fluid models analyze CCA properties.

We use modeling to find boundaries in workload space.

Our position

Use burst dimensions to derive CCA-specific regime boundaries, then generate workloads that cover those regimes.

Encore (HotNets 2023), Millisampler (IMC 2022)., Mahak (HotNets 2023), Contracts (NiNeS 2026), DCTCP (sigmetrics 2011)

Regime centric CCA evaluation = burst-aware + CCA-aware

Takeaways

- ➢ Current DCN workloads are mostly CCA-agnostic.
- ➢ Burst-aware incast workloads still miss production temporal clustering.
- ➢ Burst dimensions let us define CCA-specific regimes.
- ➢ Holistic workload generation should target regime coverage, and easy comparisons.

Future directions

- ❑ Extend analysis to INT-based, delay-based, and receiver-driven CCAs.
- ❑ Build a burst generator with coverage guarantees.
- ❑ Develop a regime-centric comparator for CCAs and workloads.