

INDIS 2026 · SC26 WORKSHOPS



Who Benefits from Federation?

Heterogeneous Gains in Cross-Operator Network Traffic Representation Learning

Mohammad Arafath Uddin Shariff · **Byrav Ramamurthy**
School of Computing, University of Nebraska-Lincoln





U.S. DEPARTMENT OF ENERGY

What's Ahead

1 **Motivation**
Why operators train alone

2 **Methodology**
A 20-client federation on real telemetry

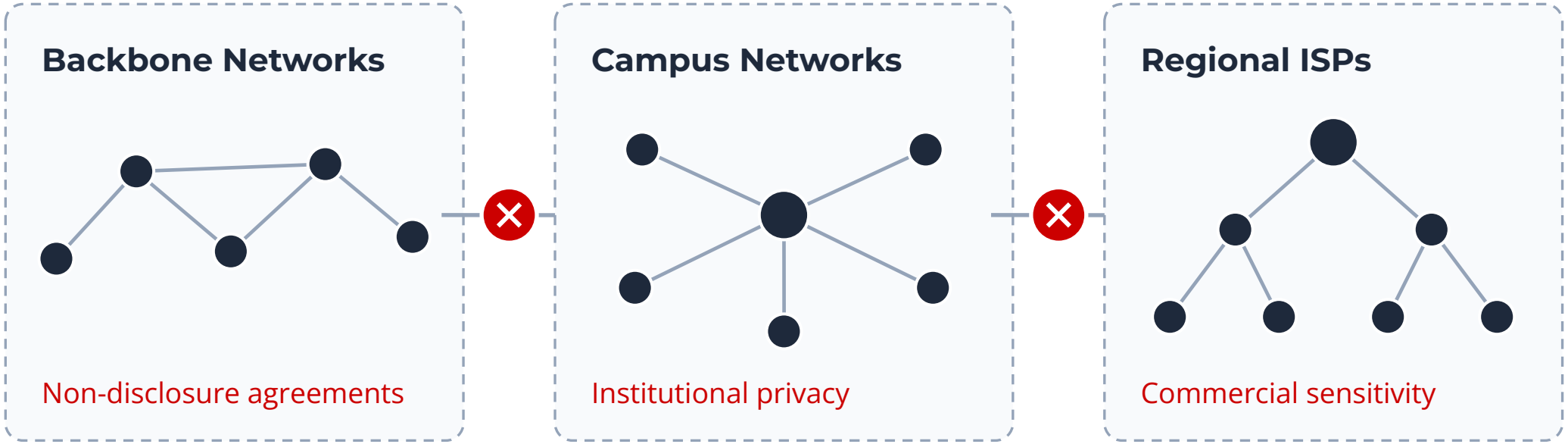
3 **Results**
Cost, who benefits, and why

4 **Discussion**
Implications and limitations

5 **Conclusion**
Takeaways and questions

The Problem

Every operator has the data. No operator can share it.



So each operator trains alone.

Two Open Questions

Federation shares model parameters, never raw data.

1 **What does keeping data local cost?**

Federated vs. centralized pooling

2 **Does everyone benefit equally?**

Who gains from joining, and why

Not asked before, to our knowledge

We answer both on real telemetry from three operators.

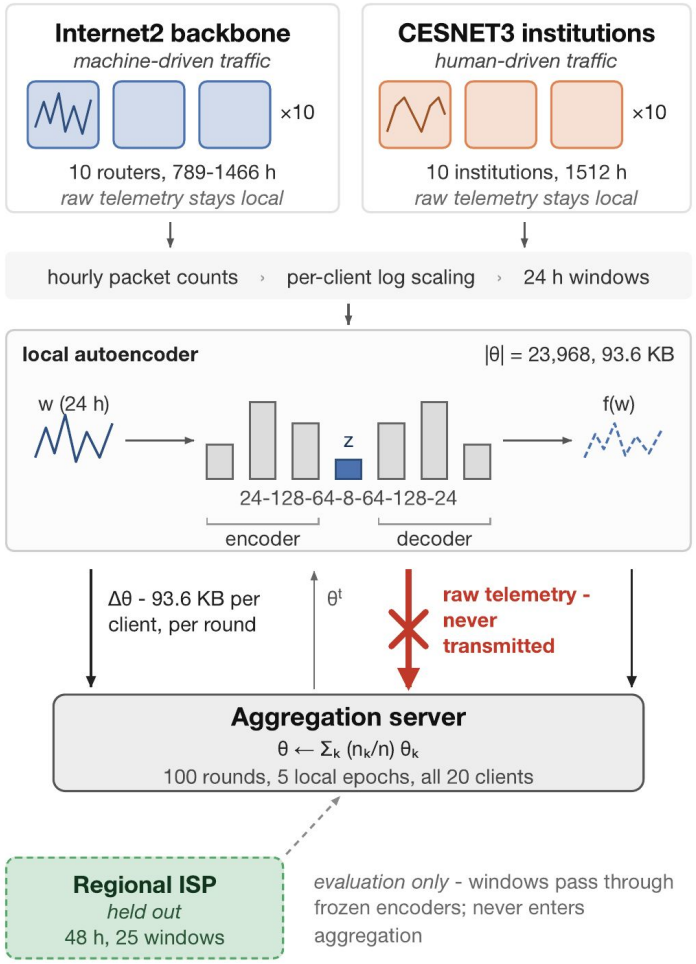
The Federation

10 **Internet2 backbone routers**
Machine-driven · 789–1,466 h each

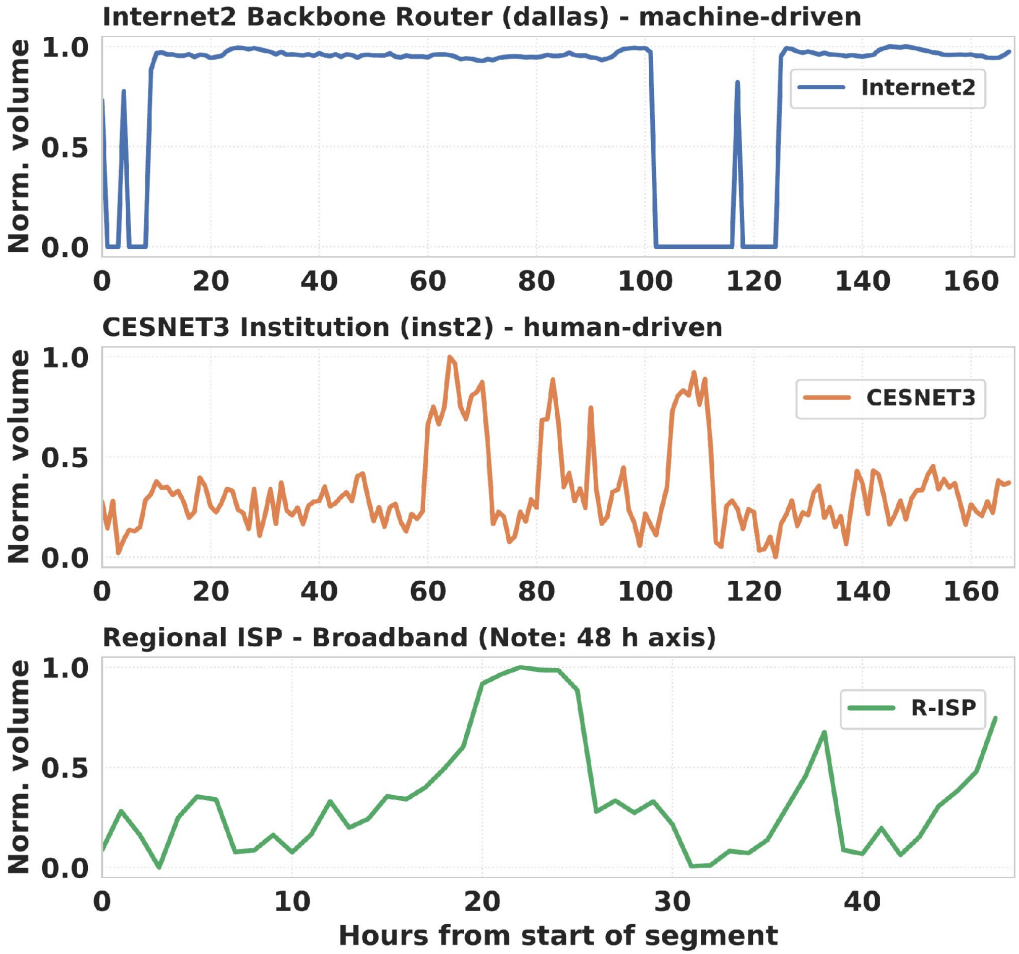
10 **CESNET3 institutions**
Human-driven · 1,512 h each

1 **Regional ISP, held out**
Evaluation only · 48 h, 25 windows

Only model parameters cross administrative boundaries.



Two Traffic Regimes



Machine-driven

Internet2 backbone

Scheduled science transfers: sustained load, abrupt aperiodic shifts, weak daily cycle.

Human-driven

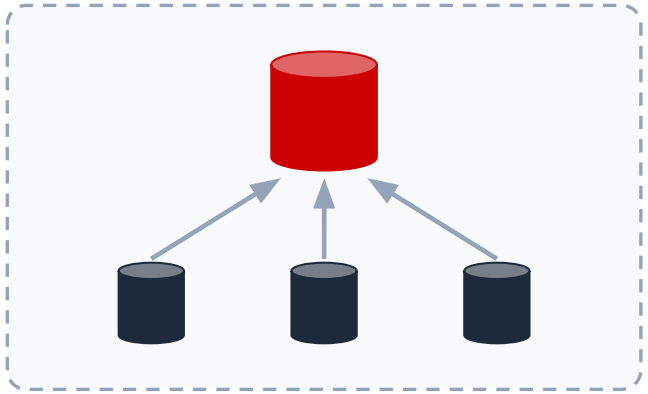
CESNET3 institutions

Follows working hours: strong daily cycle, quieter weekends.

This distinction organizes our main result.

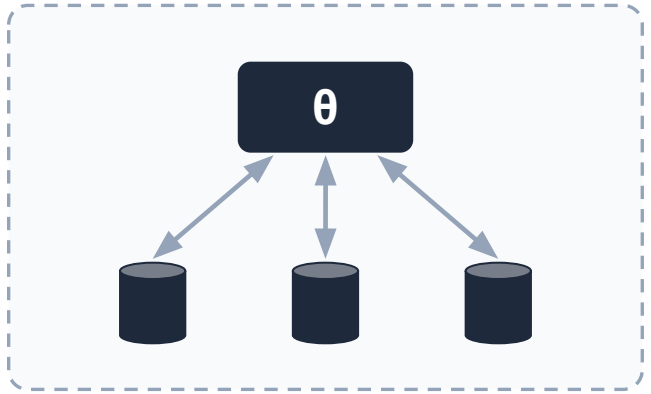
Experimental Setup

Three training regimes, one optimization budget.



Centralized

Pools all data · upper bound
Violates data custody



Federated (FedAvg)

Shares parameters only
100 rounds × 5 local epochs



Local-only

Each operator alone
Today's practice

24→128→64→8→…→24
MLP autoencoder, 24 h windows

500 epochs
equal budget, every regime

5 seeds
mean ± s.d. across seeds

Reconstruction MAE
on each client's future 15%

M. A. U. Shariff and B. Ramamurthy, "Who Benefits from Federation?" INDIS Workshop @ SC26, Chicago, IL, 2026.

Federated ≈ Centralized

Keeping data local costs nothing measurable.

CENTRALIZED		FEDERATED
0.2519	≈	**0.2504**
test MAE · upper bound		test MAE · data stays local

p = 0.020
TOST equivalence at ±5%

0.46%
observed gap, vs. 3.8% seed-to-seed spread

Equivalent, not merely "not significantly different."

M. A. U. Shariff and B. Ramamurthy, "Who Benefits from Federation?" INDIS Workshop @ SC26, Chicago, IL, 2026.

The Mean Misleads

Federation beats training alone (p = 0.0001). But by how much?

MEAN

71.7%

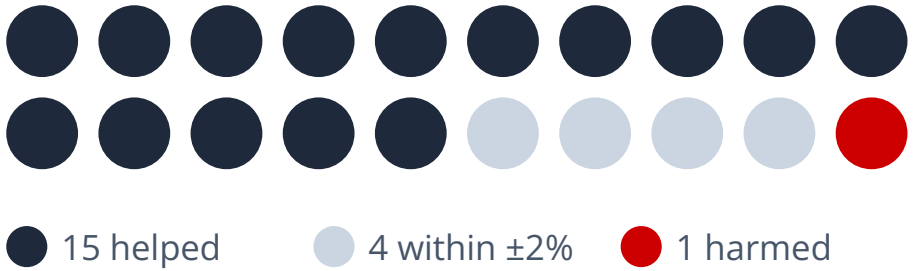
MEDIAN

4.9%

extra error when each operator trains alone

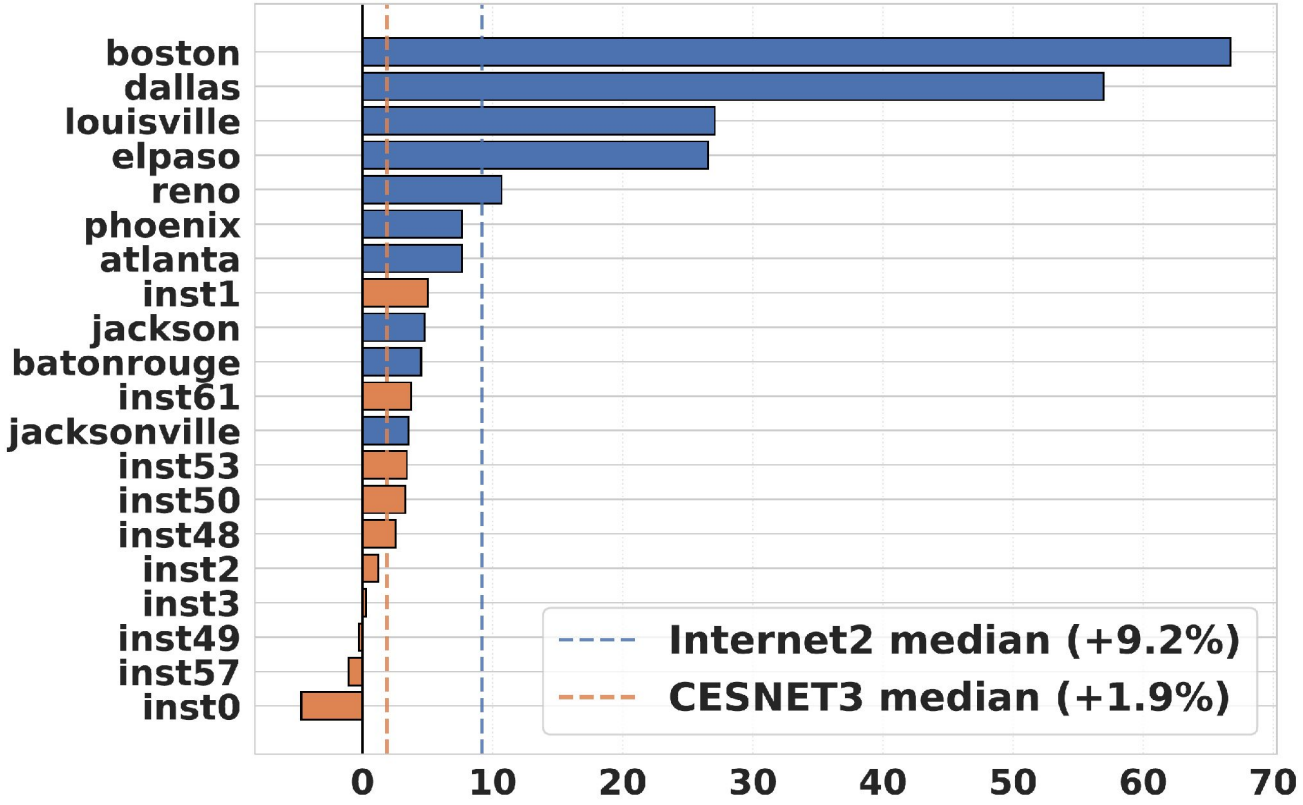
One data-starved router **(529 windows) gains 66.7%** and inflates the mean. Without it: 10.7%.

Per client, federated vs. local-only



From here on, we report medians.

Benefit Is Stratified



+9.2%

Internet2 median benefit

+1.9%

CESNET3 median benefit

p = 0.0003

one-sided Mann-Whitney U

Every backbone router gains; every neutral or harmed client is a campus.

It’s Not Data Quantity

Same data size, opposite outcomes.

Internet2 router

1,003 training windows

+27.1%

vs.

CESNET3 institution

1,035 training windows

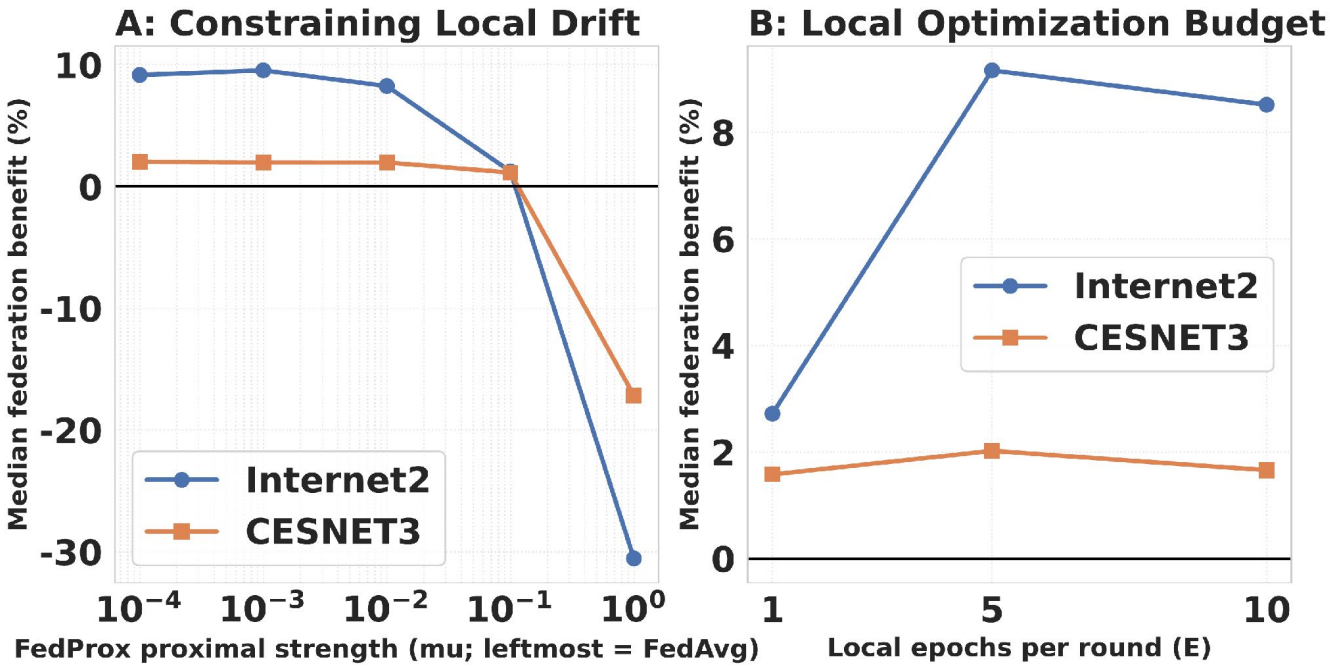
−4.7%

Pooled r = −0.62 is a grouping artifact. **Within Internet2, where size varies: r = −0.43, p = 0.22 (n.s.)**

Network membership orders the benefit; data quantity does not.

The Mechanism

Constrain local adaptation, and the backbone benefit disappears.



A · Penalize drift

FedProx μ = 0.1

Backbone **9.2% → 1.3%**
Campus 1.9% → 1.1%

B · Fewer local epochs

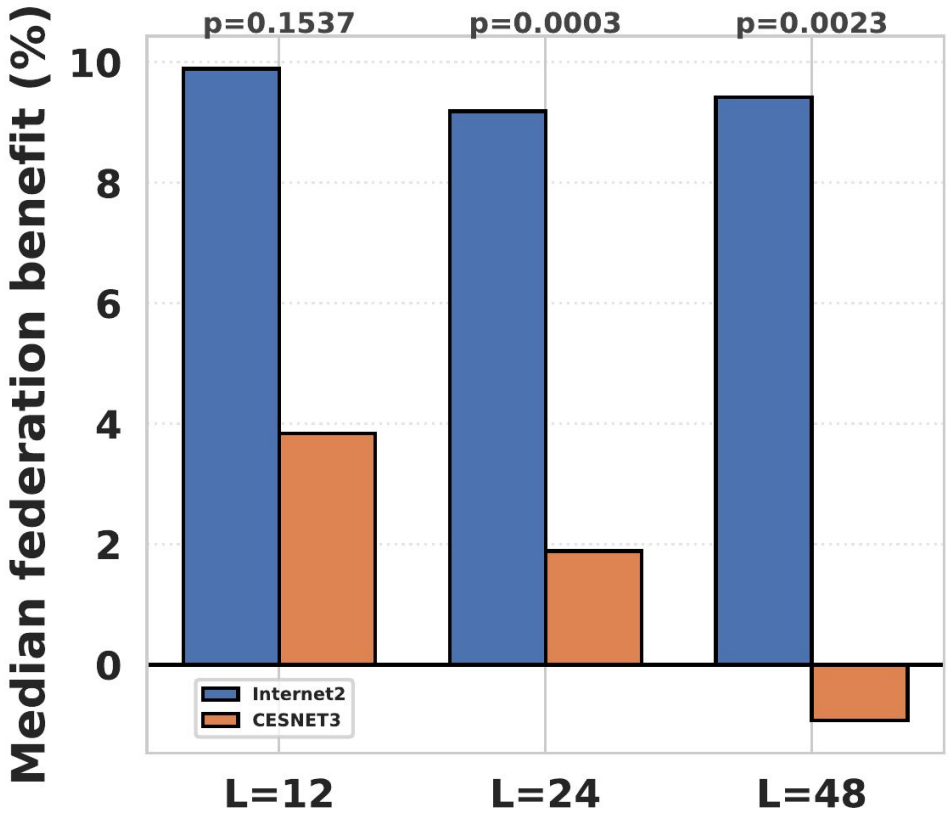
E = 5 → 1, same rounds

Backbone **9.2% → 2.7%**
Campus 1.9% → 1.6%

Client drift is productive for exactly the clients who gain.

Robustness

The effect holds where the regime account predicts.



Window length

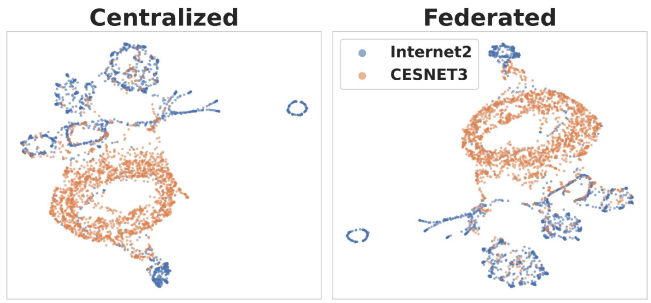
Gap grows 6.1 → 7.3 → 10.3 points at 12, 24, 48 h. Significant at 24 and 48 h; 12 h cannot hold a full daily cycle.

Aggregation method

Best FedProx setting is within 0.3% of FedAvg, inside ~2% seed noise.

Secondary Observations

Shared representation



One encoder across operators; local-only training gives 20 incomparable latent spaces.

Reliable model selection

3.09 vs **1.74**

test / validation error: local-only vs. federated

Local models choose checkpoints on 95–204 windows and overfit; federation selects on pooled validation.

Communication

19.2 MB

per-client crossover where federation becomes cheaper

Today: 383 MB federated vs. 211 KB of raw series. The case is administrative, not bandwidth.

What This Means

Heterogeneity here is signal, not noise.

For operators

Benefit is assessable before joining: traffic regularity and history length.

Irregular or short-history networks gain most, yet contribute least.

For federated optimization

Client drift is not always a pathology; it depends on the deployment.

Personalization should be a per-client quantity, not one global setting.

AT μ = 0.1

0.2504 → 0.2604

mean error: barely moves

p 0.0003 → 0.24

stratified benefit: gone

Aggregate metrics can hide the effect entirely.

Limitations & Next Steps

Limitations

- **Simulated federation**
  One process: no latency, stragglers, or failures
- **Contractual privacy**
  Gradient inversion can still leak from updates
- **Small and univariate**
  20 clients, one metric, CESNET's 10 most active
- **Regime not isolated**
  Groups also differ in operator, vantage, history

Next steps

- **Synthetic regularity sweep**
  Clients at graded periodicity to test the account
- **Personalized federated learning**
  Methods that keep local adaptation by design
- **Real deployment**
  Physically distributed federation on FABRIC
- **Downstream tasks**
  Anomaly detection and forecasting

Takeaways

1 **Federation ≈ centralized pooling**

Keeping data local costs nothing measurable (TOST p = 0.020).

2 **Benefit is stratified**

Backbone +9.2% vs. campus +1.9% median benefit (p = 0.0003).

3 **Local adaptation drives it**

Client drift helps the irregular-traffic clients who gain most.

Look past the aggregate: look at your own traffic.

Questions?

Looking for internships and full-time roles. Let's connect

Thank you.

CONTACT

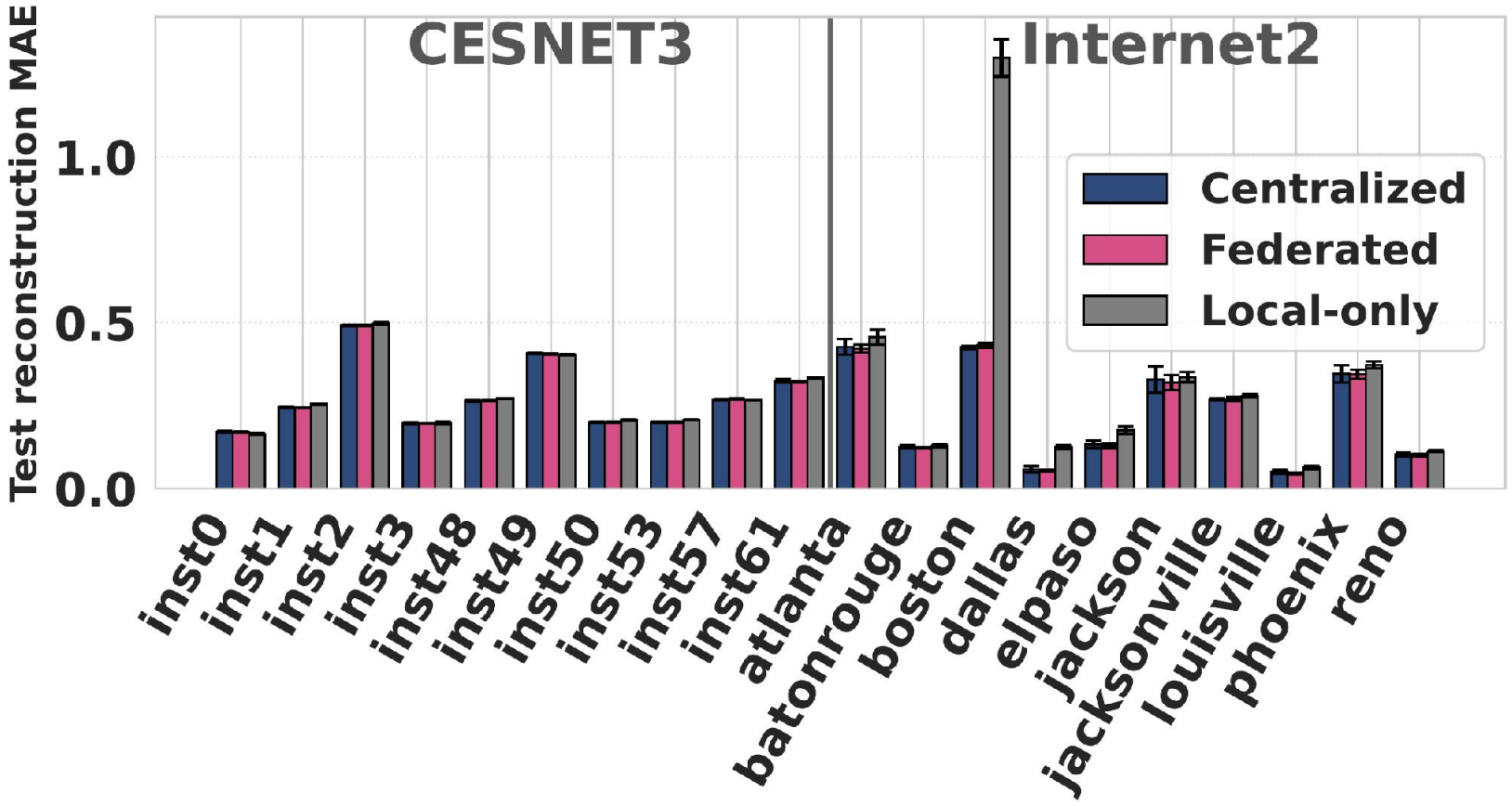
mshariff2@huskers.unl.edu
ramamurthy@unl.edu

CODE & ARTIFACTS

doi.org/10.5281/zenodo.22844433

Supported in part by NSF grants OAC-2322369 and CNS-2413851, and U.S. DOE grant DE-SC0024648.
Data courtesy of Internet2, CESNET, and a regional ISP.

Per-Client Results



Boston: *529 training windows, ~3× worse alone than any other client; +66.7% with federation.*

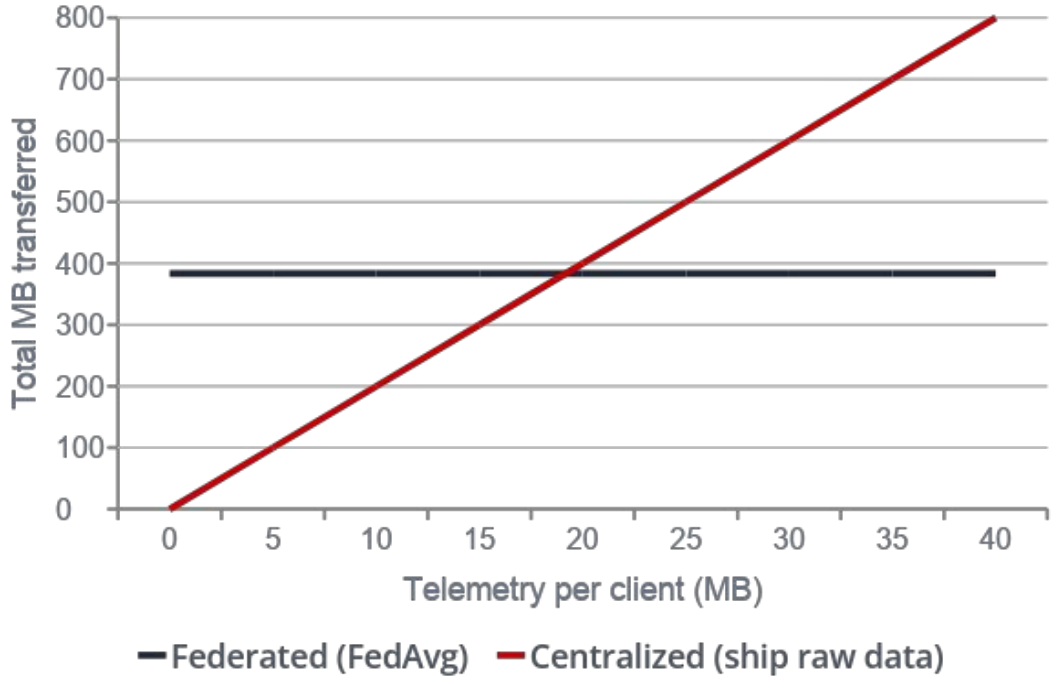
Robustness Table

Check	Configuration	Fed. MAE	B (Internet2)	B (CESNET3)	W/T/L	p
Aggregation	**FedAvg (default)**	**0.2504**	**+9.2%**	**+1.9%**	**15/4/1**	**0.0003**
	FedProx μ = 0.001	0.2496	+9.8%	+1.8%	15/4/1	0.0002
	FedProx μ = 0.01	0.2512	+8.2%	+1.8%	14/5/1	0.0007
	FedProx μ = 0.1	0.2604	+1.3%	+1.1%	7/12/1	0.236
	FedProx μ = 1.0	0.3411	−30.5%	−17.2%	2/0/18	0.977
Window length	L = 12	—	+9.9%	+3.8%	16/3/1	0.154
	L = 24 (default)	**0.2504**	**+9.2%**	**+1.9%**	**15/4/1**	**0.0003**
	L = 48	—	+9.4%	−0.9%	9/7/4	0.002
Local epochs	E = 1	0.2574	+2.7%	+1.6%	—	—
	E = 5 (default)	**0.2504**	**+9.2%**	**+1.9%**	**15/4/1**	**0.0003**
	E = 10	0.2522	+8.5%	+1.7%	—	—

B: median per-client benefit · W/T/L: clients helped / unaffected / harmed at a 2% threshold · p: one-sided Mann-Whitney for stratification

Privacy & Communication

- **Raw telemetry stays local**
  No raw series crosses an administrative boundary.
- **But updates can leak**
  Gradient inversion can reconstruct samples from shared updates.
- **So the claim is contractual**
  An adversarial server would need DP updates; utility cost unmeasured.



Crossover $V^* = 2R|\theta|b \approx 19.2$ MB per client. Today: 383 MB vs. 211 KB.