



NetReplica: Bottleneck-Centric Data *Generation Substrate*

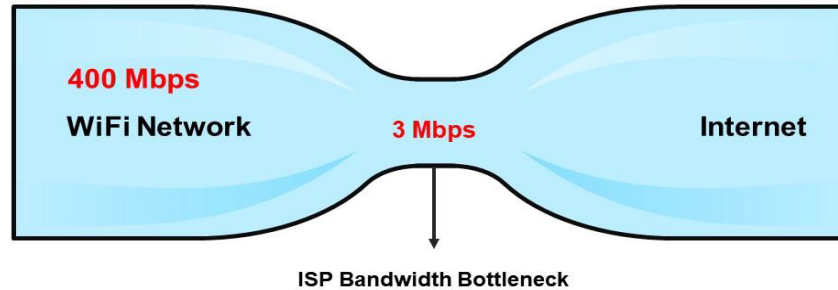
Jaber Daneshamooz

Jaber@ucsb.edu



Why Bottleneck Matters?

- What is network bottleneck?
- What are the impacts of bottleneck link?





Bottleneck-Centric Data Generation

- **The need for "right" data in:**
 - Network measurement
 - ML for networking
 - Education
- **Generating bottleneck-centric data is hard**

What properties must a bottleneck-centric data substrate satisfy?



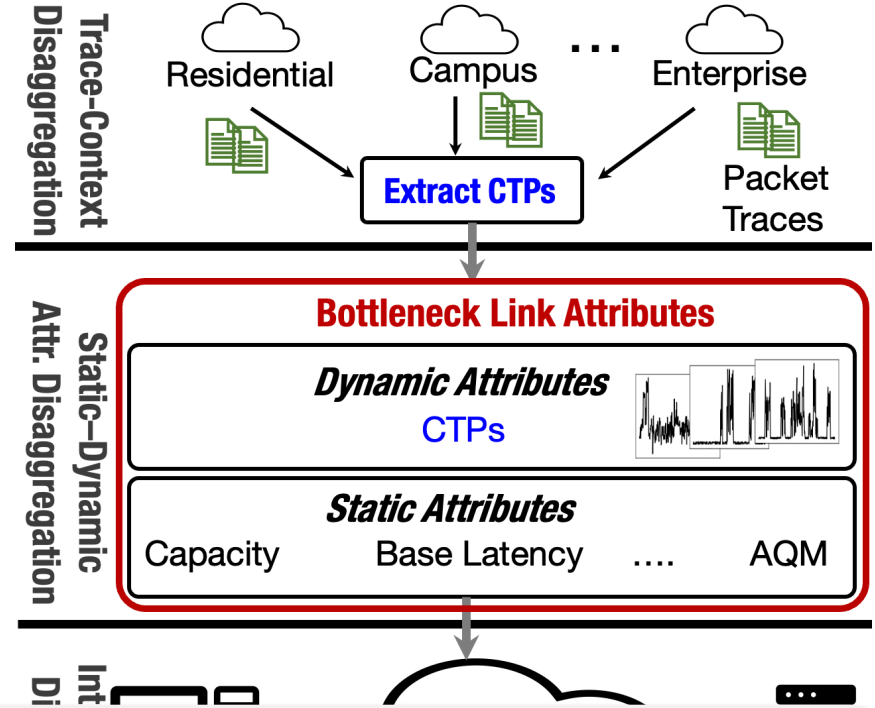
Key Data Generation Requirements

- **Controllability:**
 - Independently set static envelope + dynamic pressure.
- **Composability:**
 - Mix-and-match attributes to build richer regimes.
- **Replicability:**
 - Same spec \Rightarrow statistically equivalent behavior across runs.
- **Fidelity:**
 - Preserve closed-loop interaction between bottleneck and app under test.

Existing tools each satisfy some of these, but not all.

System Design

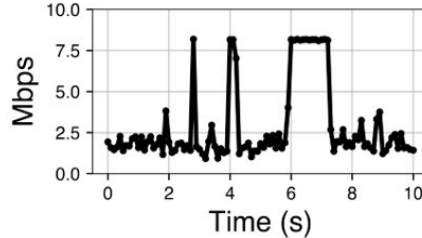
- Reusable pressure
- Enabling composability



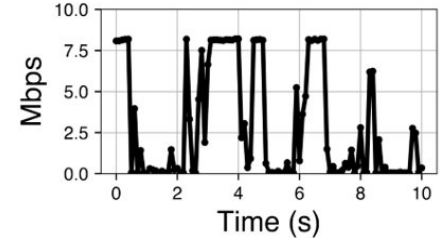
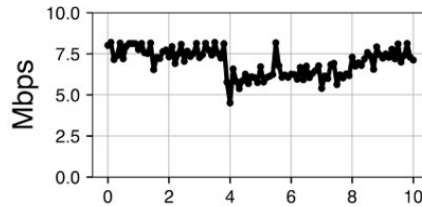
Turning passive traces into reusable pressure points (CTPs).

Understanding CTPs

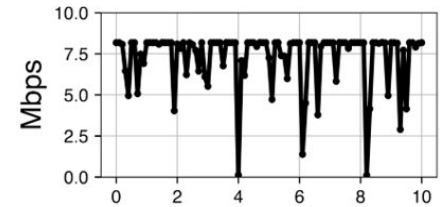
- What are CTPs?
- CTP properties
- How to use CTPs?



(a) CTP1: ↓intensity, ↓bursty

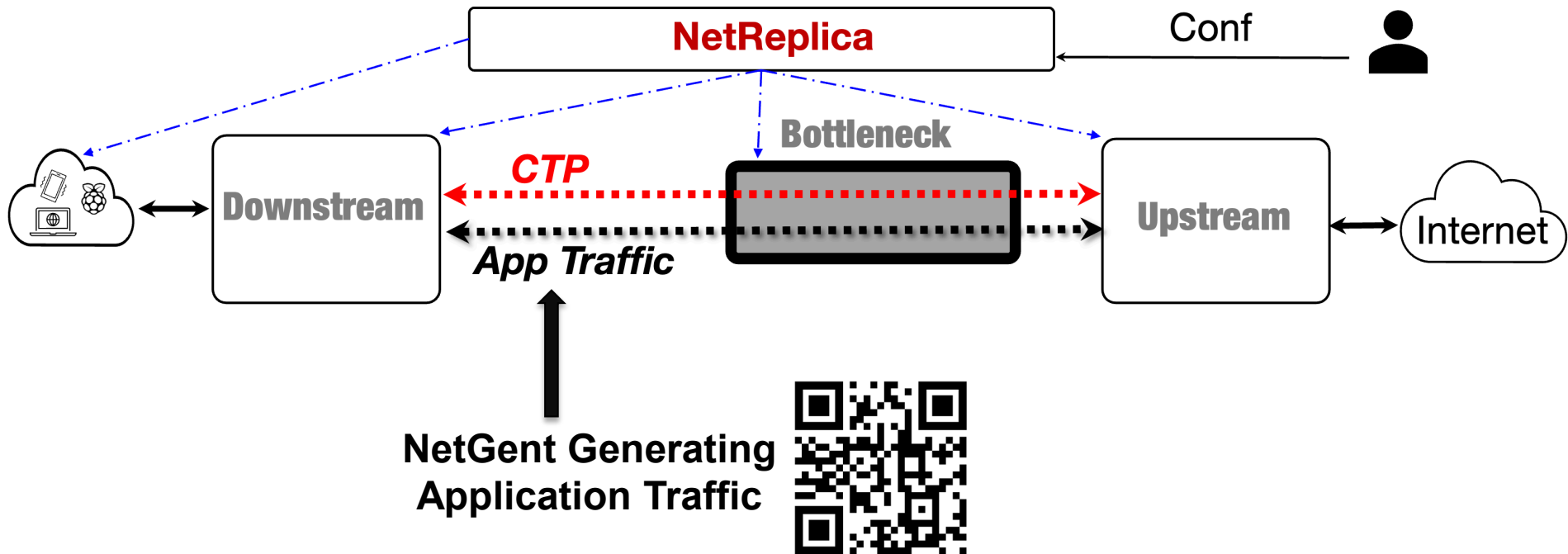


(b) CTP2: ↓intensity, ↑bursty



We can *select/transform/compose* pressure signals.

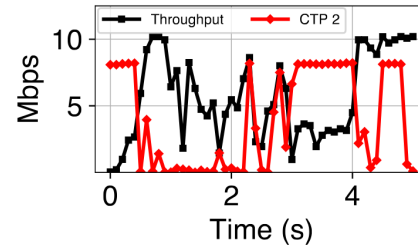
Implementation



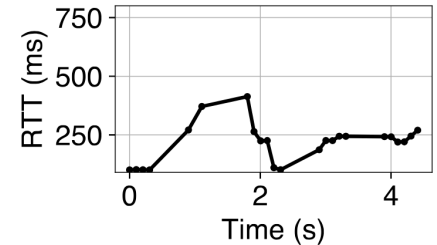
System Evaluation



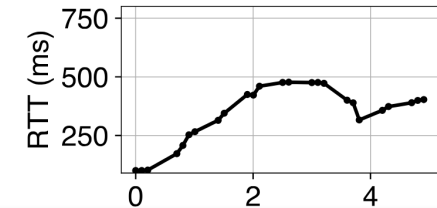
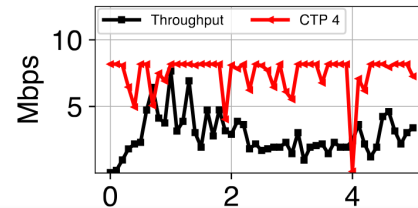
- **Controllability:** independently control
- **Composability:** mix static + dynamic
- **Fidelity:** closed-loop target app
- **Replicability:** same spec = statistically



(a) CTP 2 (Thput)



(b) CTP 2 (RTT)



Does NetReplica produce *useful training data*?



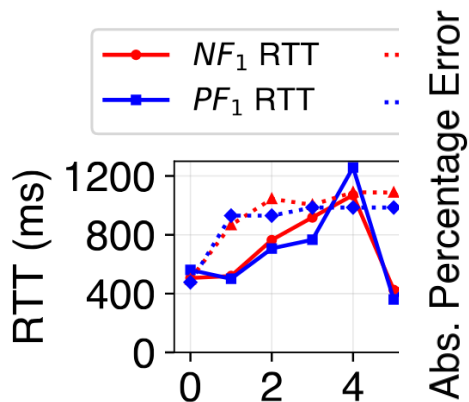
Data Evaluation

- **Using Puffer dataset as case study**
 - Re-broadcasting TV channels on Internet
 - Different ABR (Developed FUGU)
 - Public dataset
- **Feature space for FUGU model:**
 - RTT, throughput, chunk size, ...
 - Past 8 chunks
- **NetReplica generated dataset**
 - RTT > **100ms**, throughput < **6Mbps**
 - Filtering puffer dataset

Data Evaluation

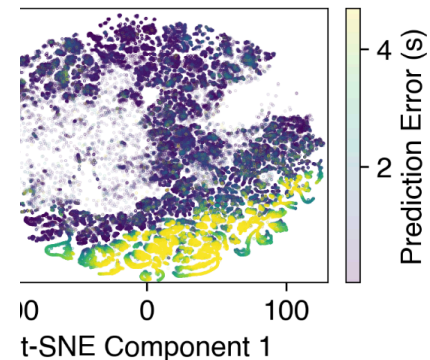
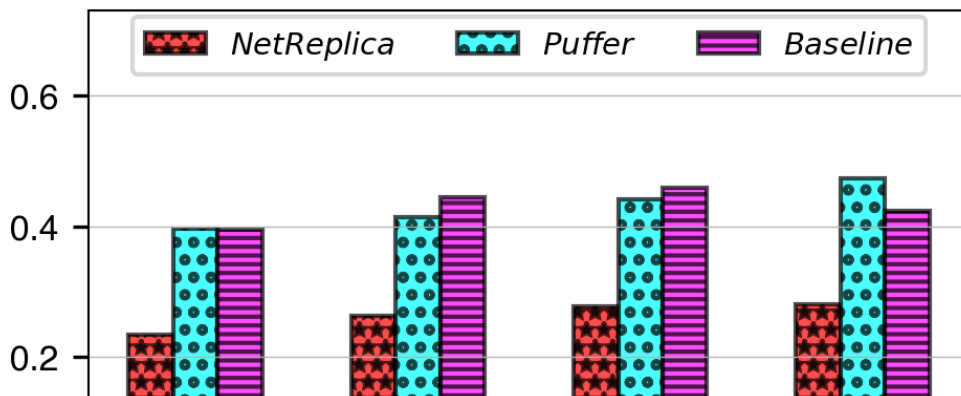


Realism



Model Performance

Coverage



NetReplica: a controllable coverage-expansion tool.

Target Domains

Research Artifact for Teaching



- **Why a substrate matters for teaching?**
 - Composability
 - Replicability
- **Generating application workflow**



NetGent
GitHub



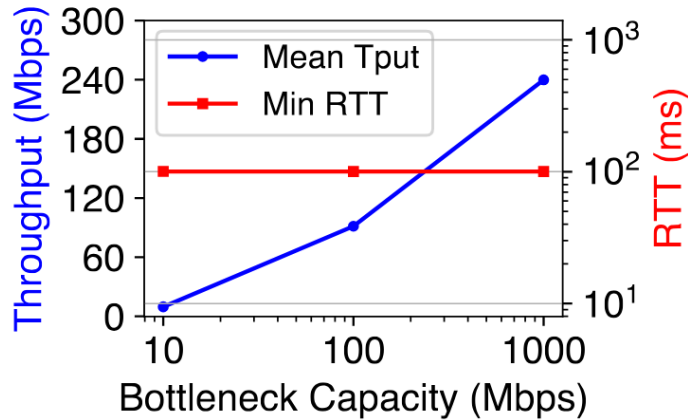
NetReplica
GitHub

NetReplica and NetGent helps teaching, measurement, and developing models for networking and security.

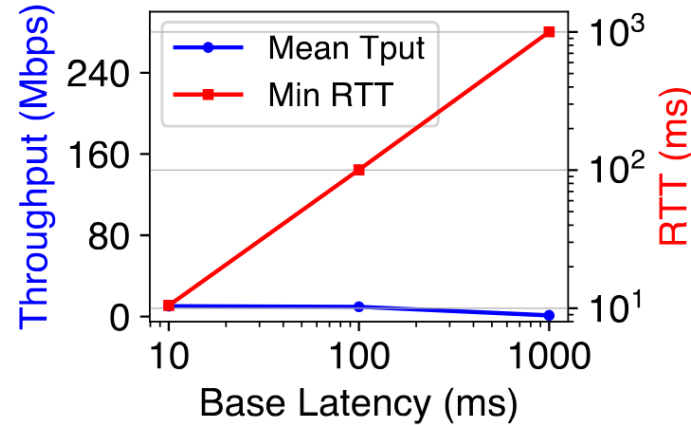
Backup slides



System Evaluation (Controllability)



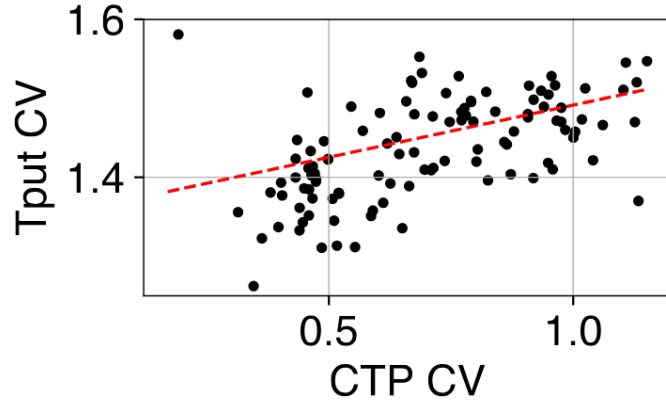
(a) Capacity Sweep



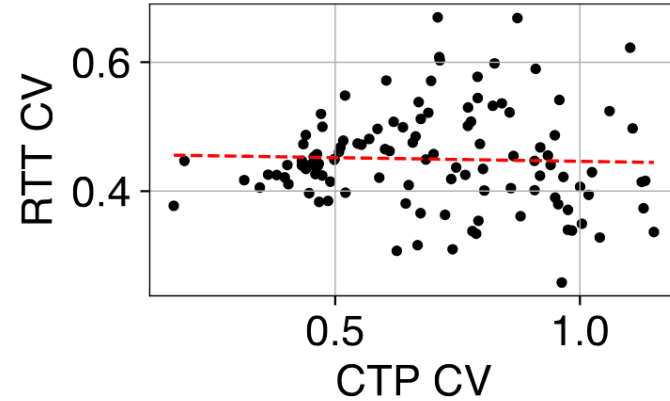
(b) RTT Sweep

Figure 5: Independent control of bottleneck attributes: thrpt scales with capacity, and RTT scales with latency.

System Evaluation (Fidelity)



(a) Throughput



(b) RTT

Figure 9: Fidelity of dynamic congestion pressure.

Implementation

