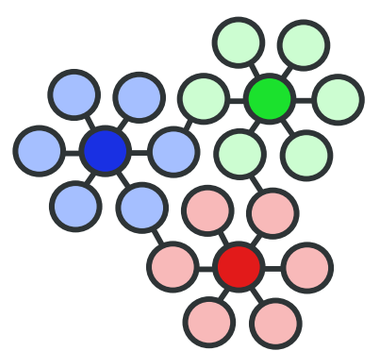
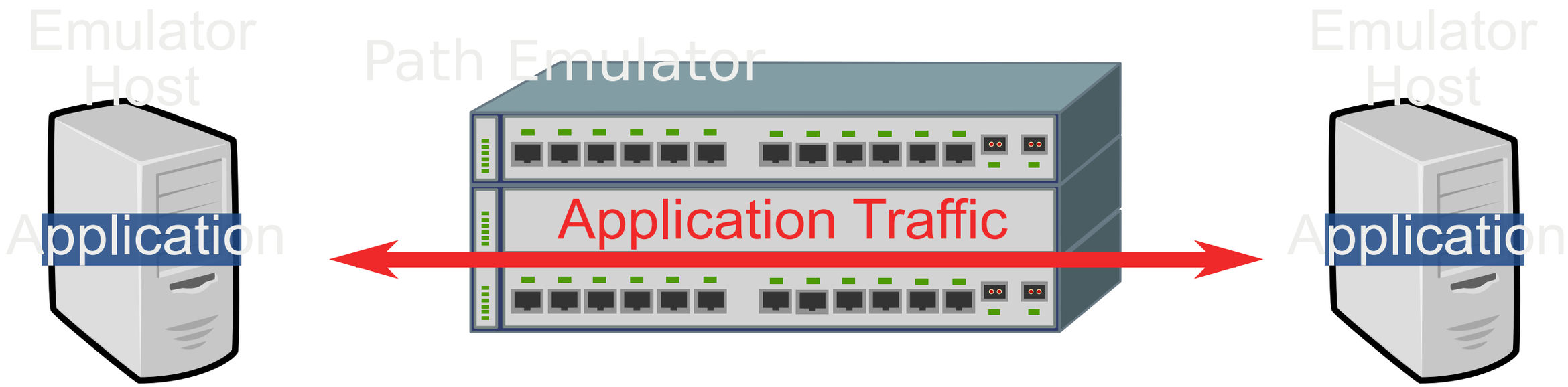
The Flexlab Approach To Realistic Evaluation of Networked Systems

Robert Ricci, Jonathon Duerig,
Pramod Sanaga, Daniel Gebhardt, Mike Hibler,
Kevin Atkinson, Junxing Zhang, Sneha Kasera, and
Jay Lepreau

NSDI 2007
April 12, Cambridge, MA



Emulators



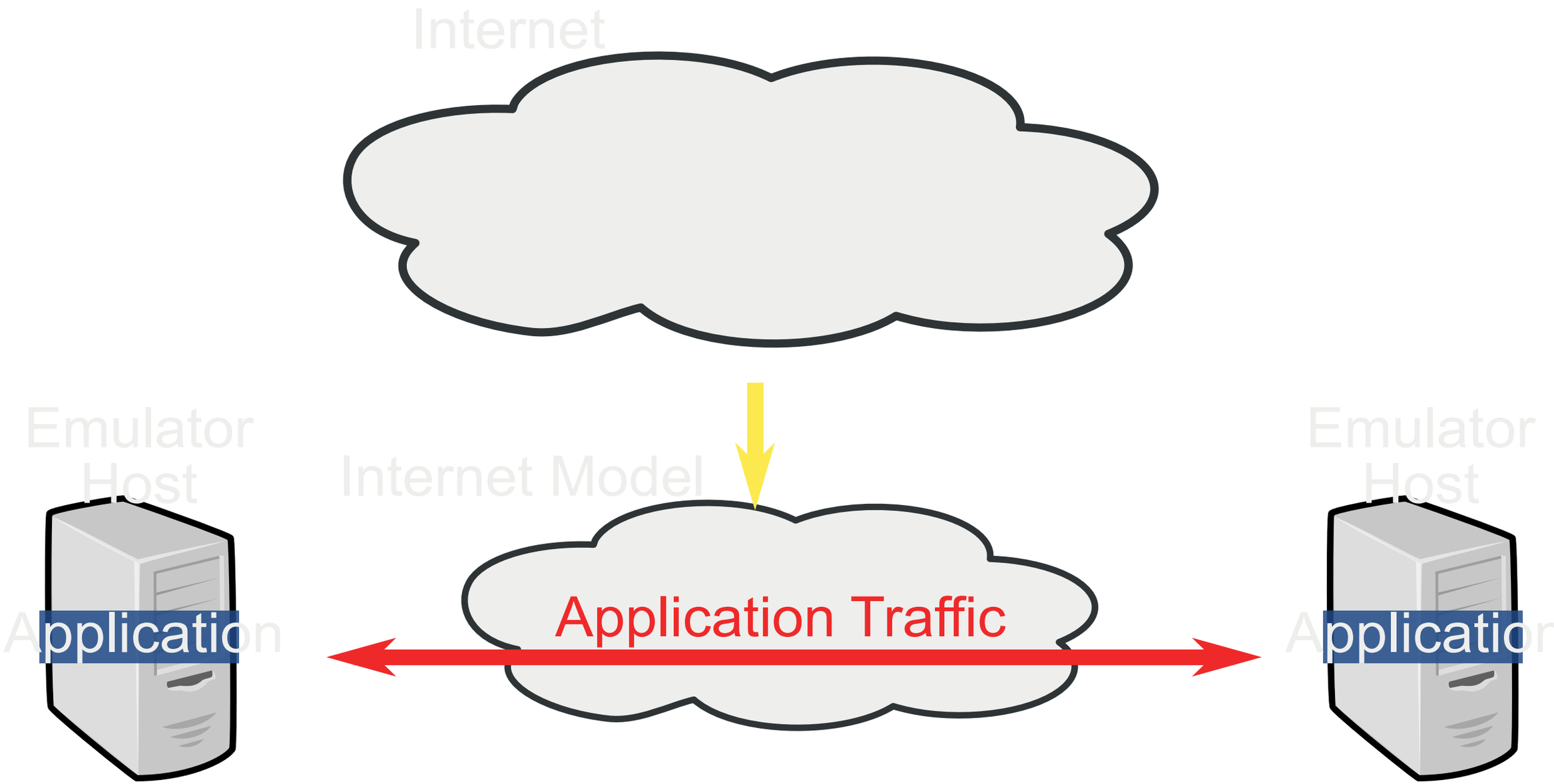
- Examples: ModelNet and Emulab
- The Good: Control, repeatability, wide variety of network conditions
- The Bad: Artificial network conditions

Overlay Testbeds

- Examples: RON and PlanetLab
- The Good: Real network conditions, deployment platform
- The Bad: Overloaded, few privileged operations, poor repeatability, hard to develop/debug on

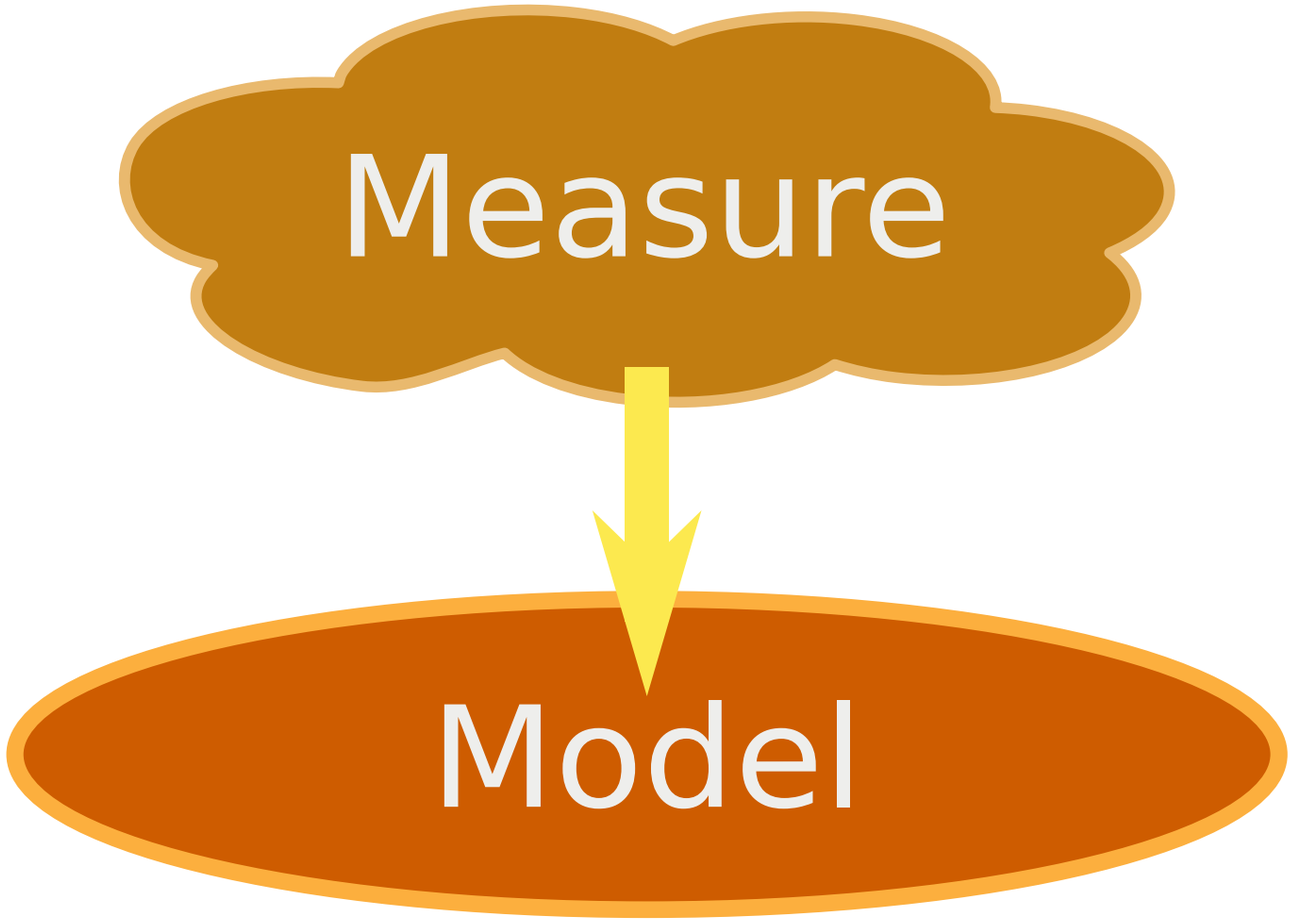
Evaluating Networked Systems: Flexlab

Goal: Real Internet within Emulator

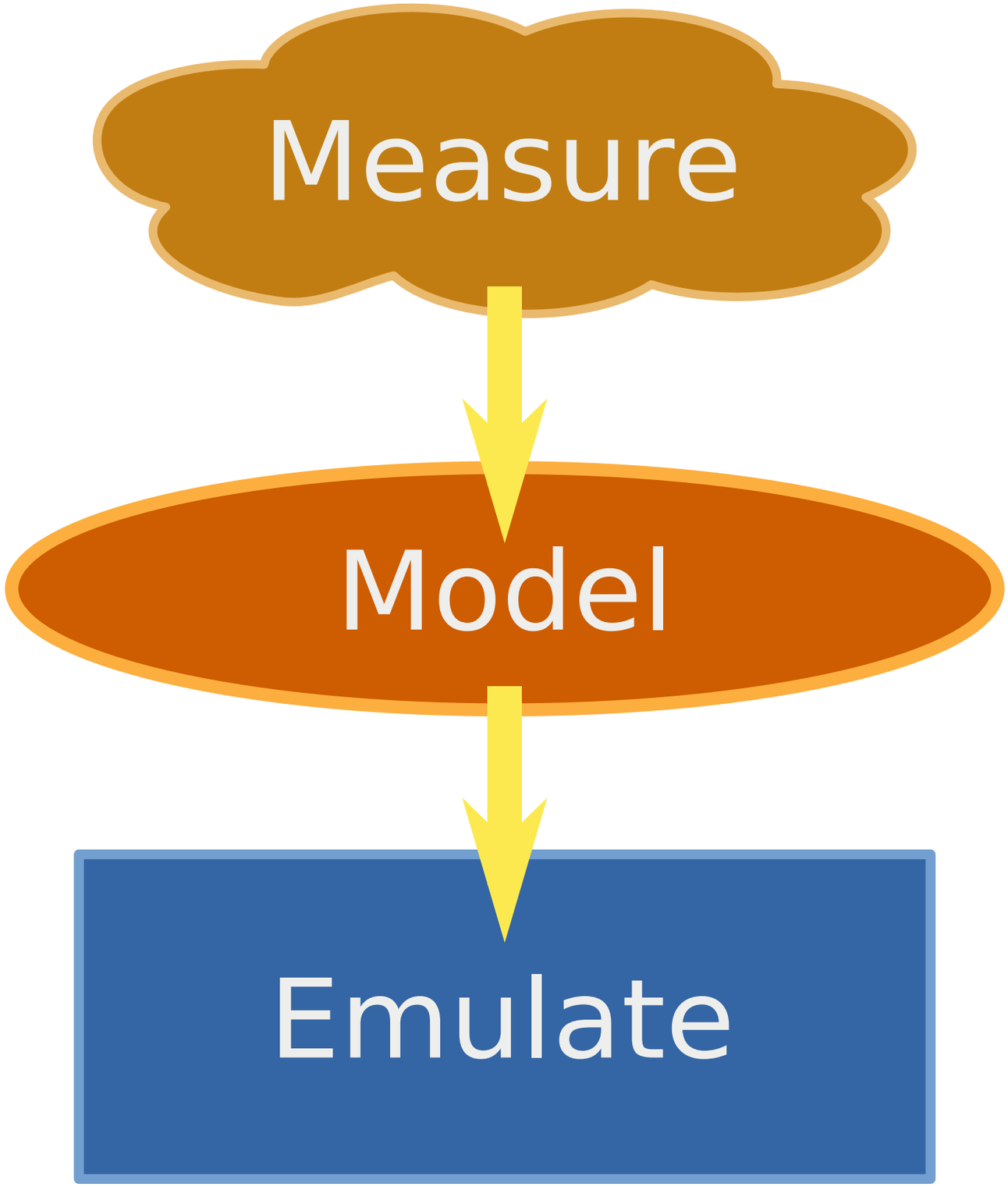


The Flexlab Approach

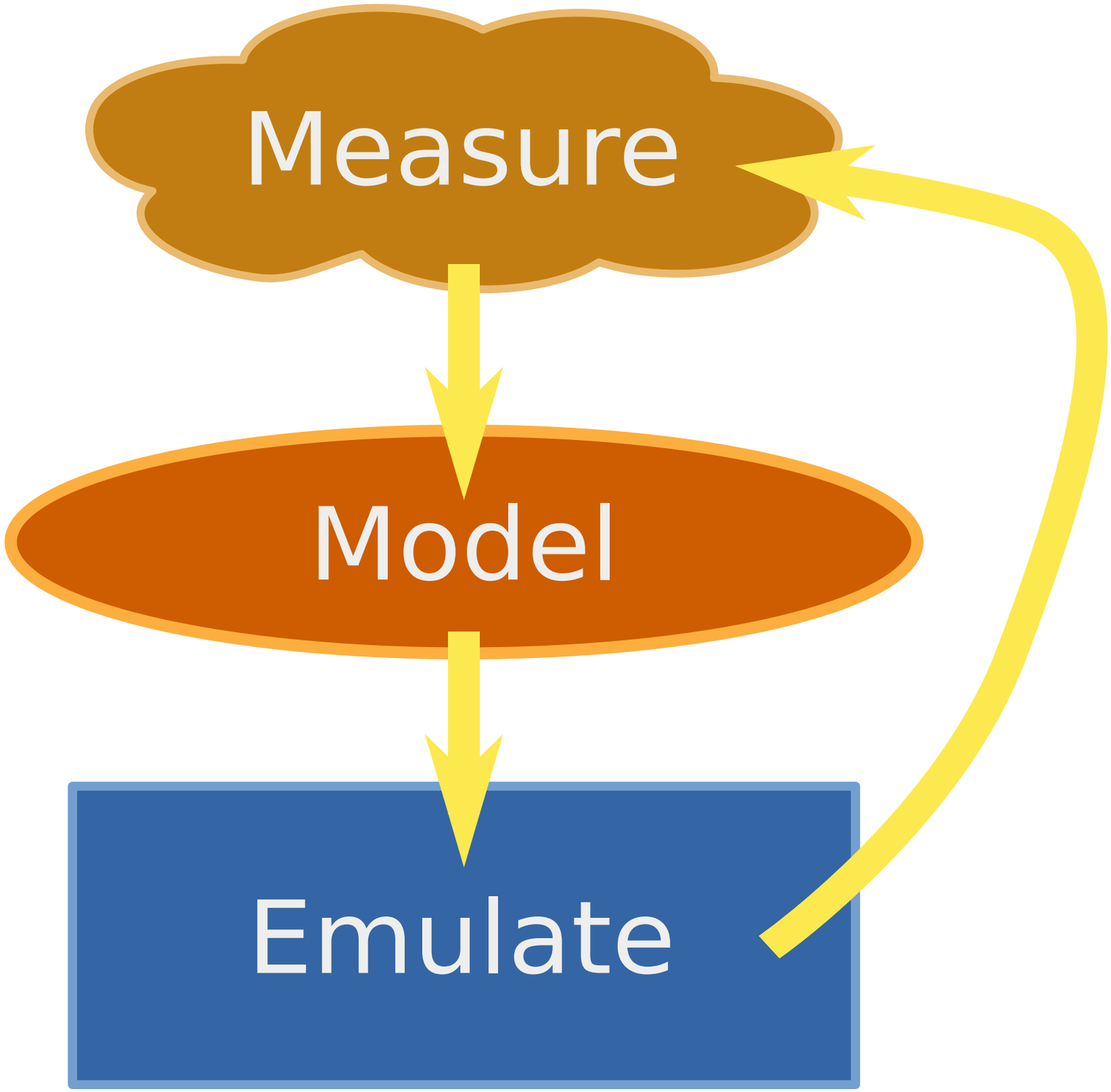
The Flexlab Approach



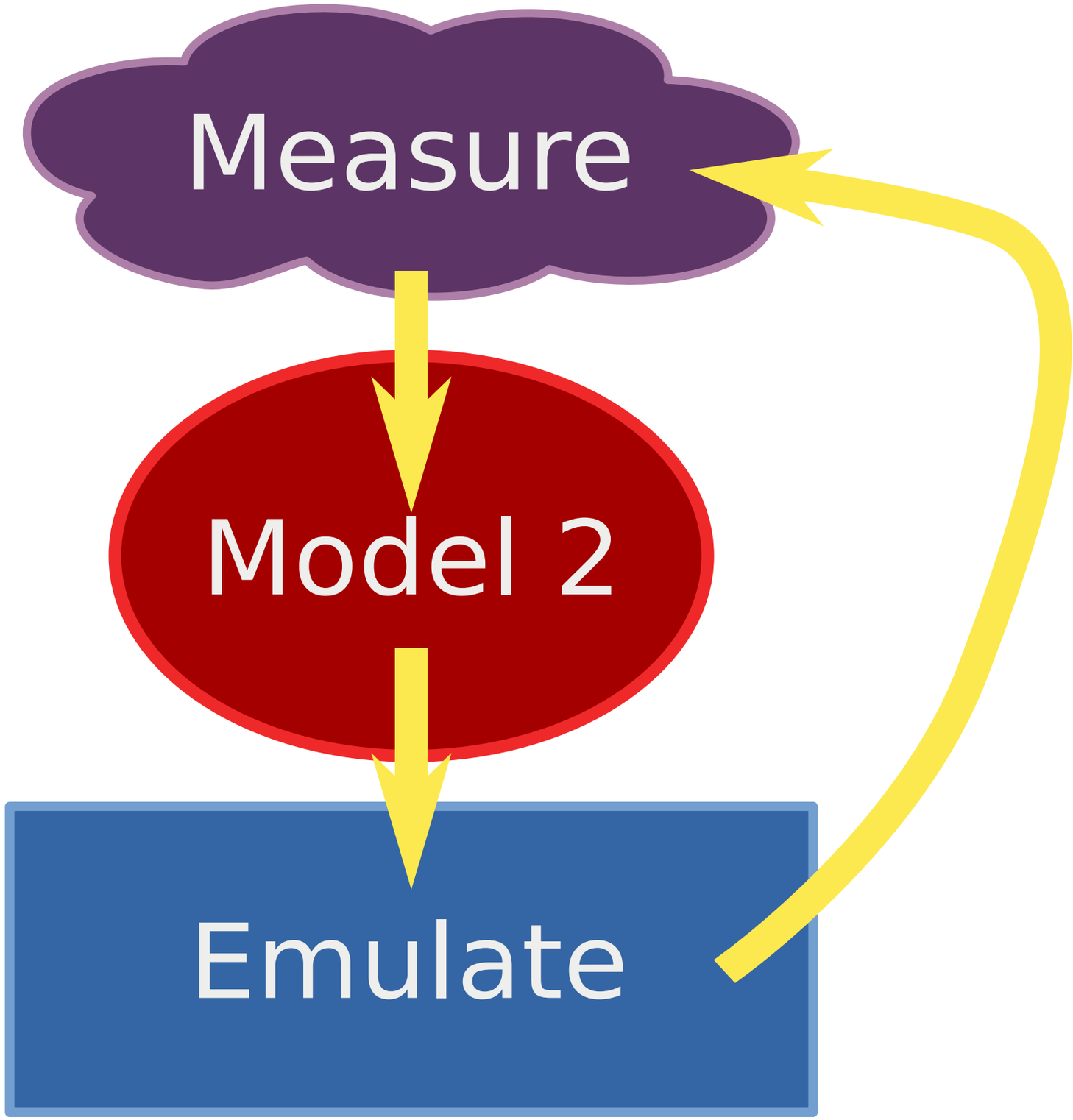
The Flexlab Approach



The Flexlab Approach



The Flexlab Approach



Key Points

- Software framework for pluggable network models
- Application behavior can drive measurements & model in real-time
- Application-Centric Internet Modeling
  - High fidelity measurement/ emulation technique
  - Includes new techniques for ABW measurement

More in the Paper

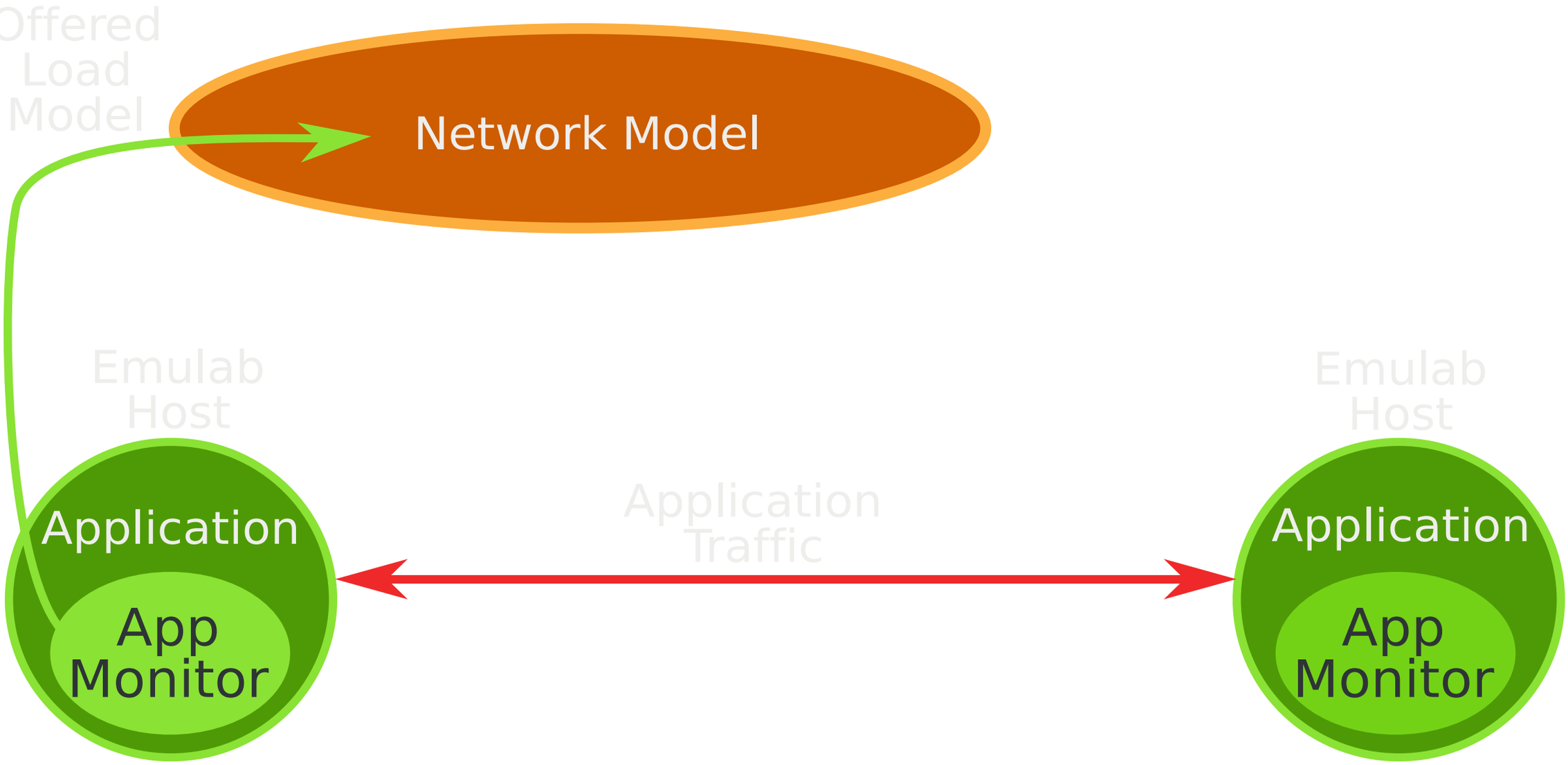
- Flexible network measurement system
- Network stationarity results
- Two straightforward network models
- Shared bottleneck analysis
- PlanetLab scheduling delay measurements

Flexlab Architecture

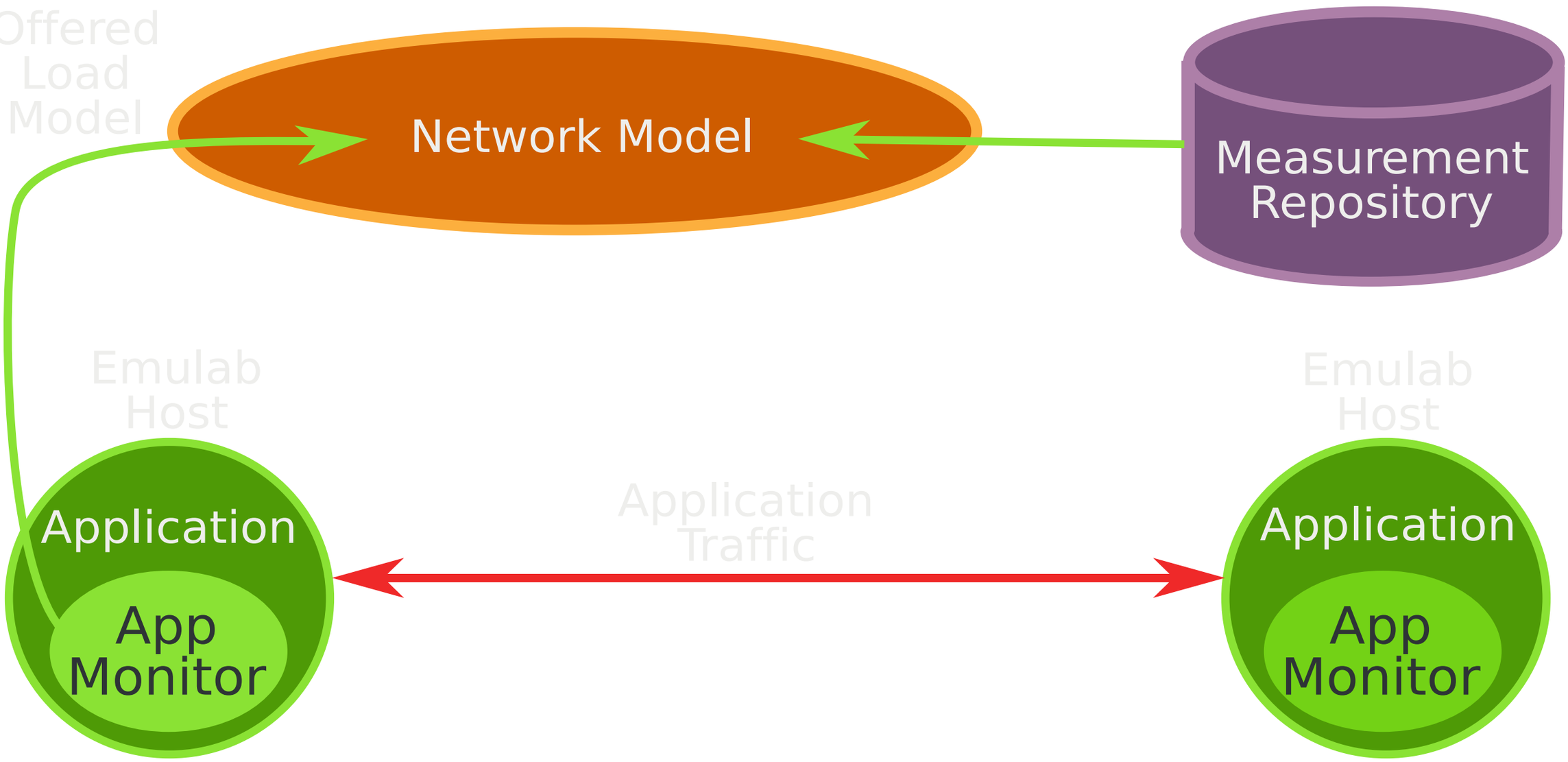
Flexlab: Application

Flexlab: Application Monitor

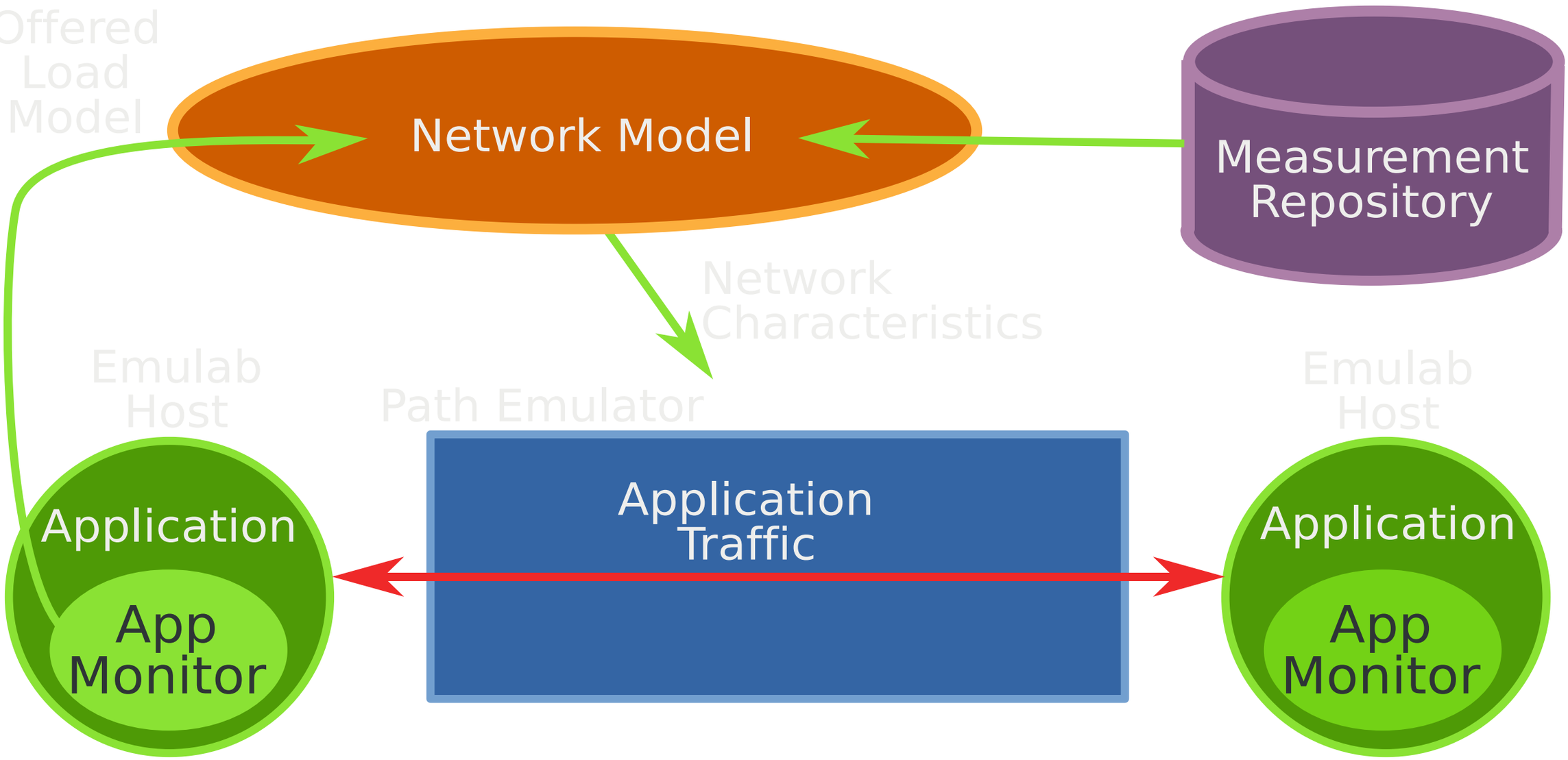
Flexlab: Network Model



Flexlab: Measurement Repo.

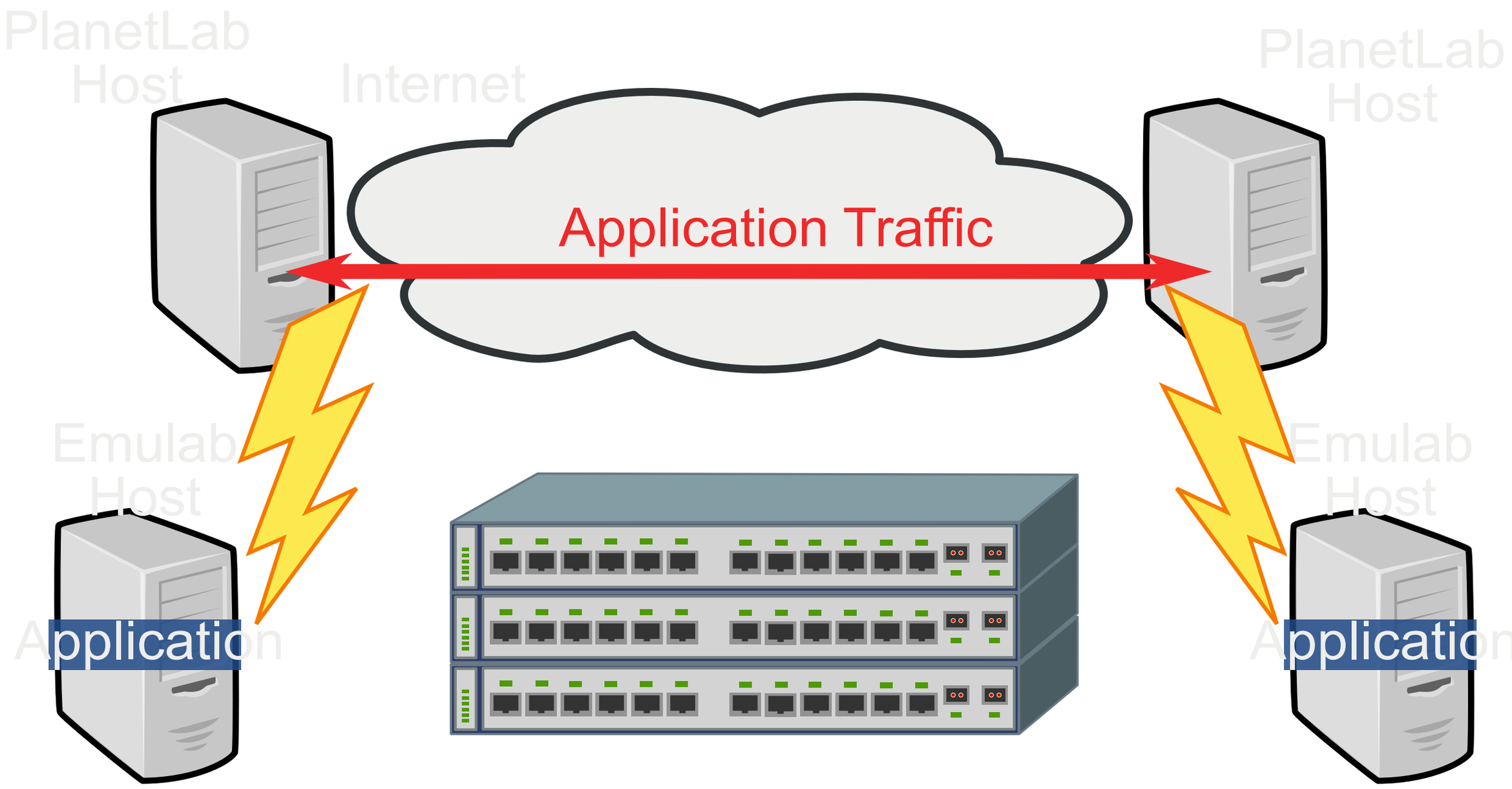


Flexlab: Path Emulator

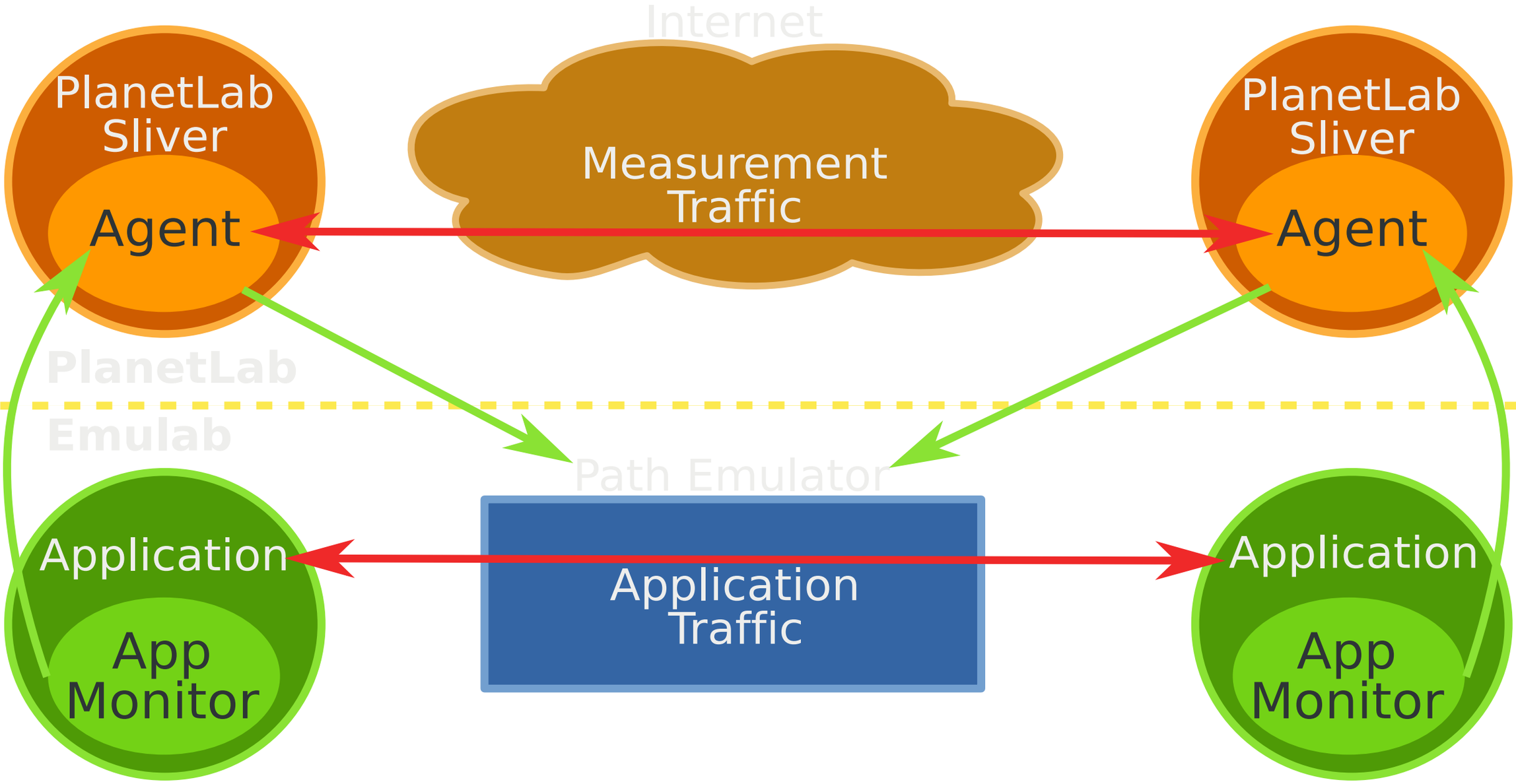


ACIM: Application-Centric Internet Modeling

Imagine Ideal Fidelity



ACIM Architecture

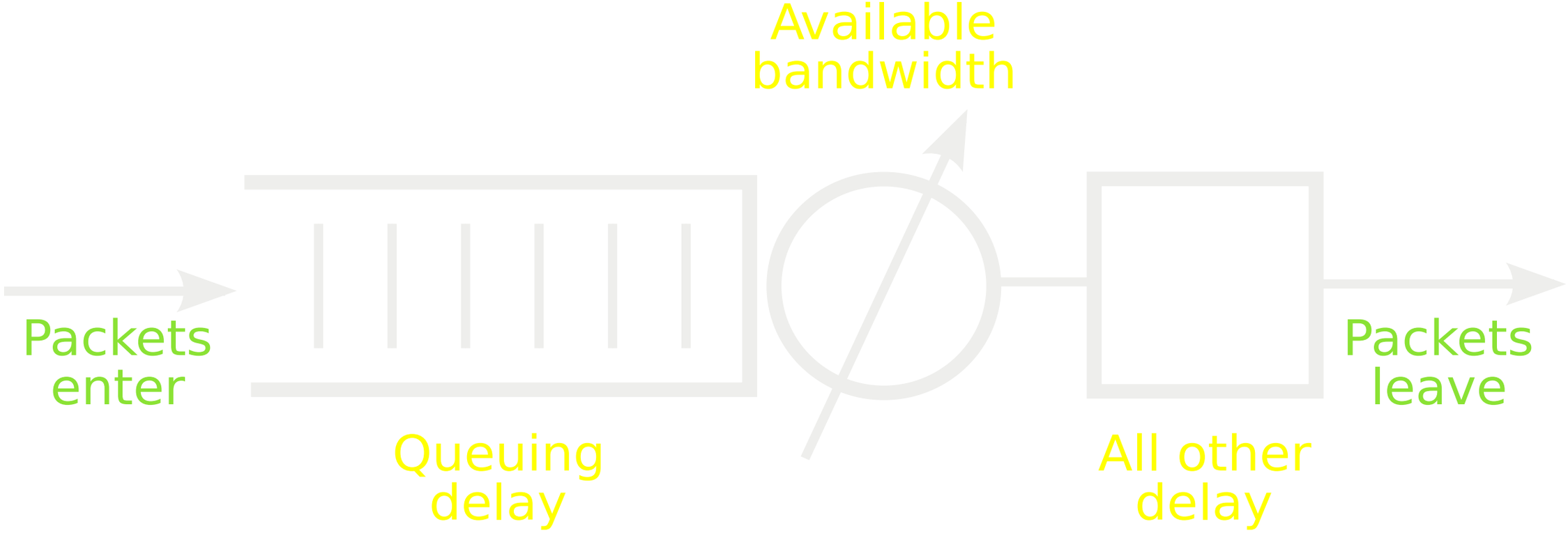


ACIM Design Challenges

- Determining when to drop packets
- Finding relationship between throughput and ABW
- Extension to UDP
- CPU starvation on PlanetLab
  - Host artifacts in throughput
  - Packet loss in libpcap

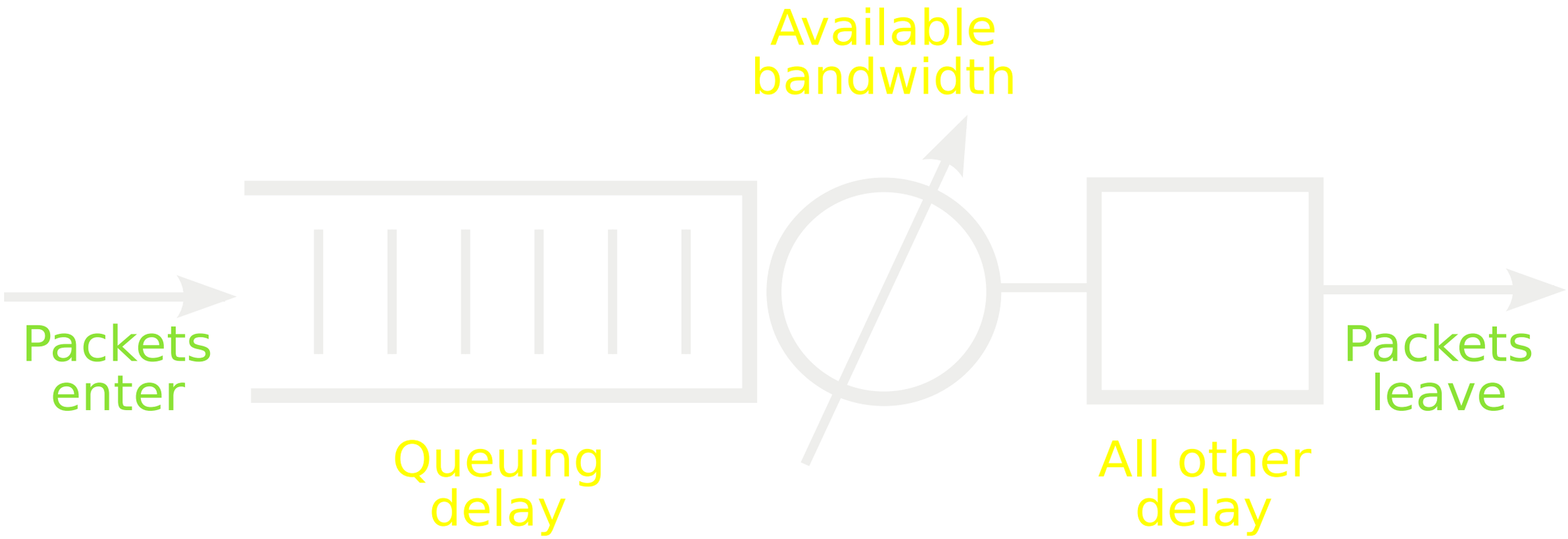
ACIM Path Emulator Parameters

All Other Delay



- Base RTT: Smallest RTT seen recently [Vegas 95]
- Packets saw little or no queueing delay

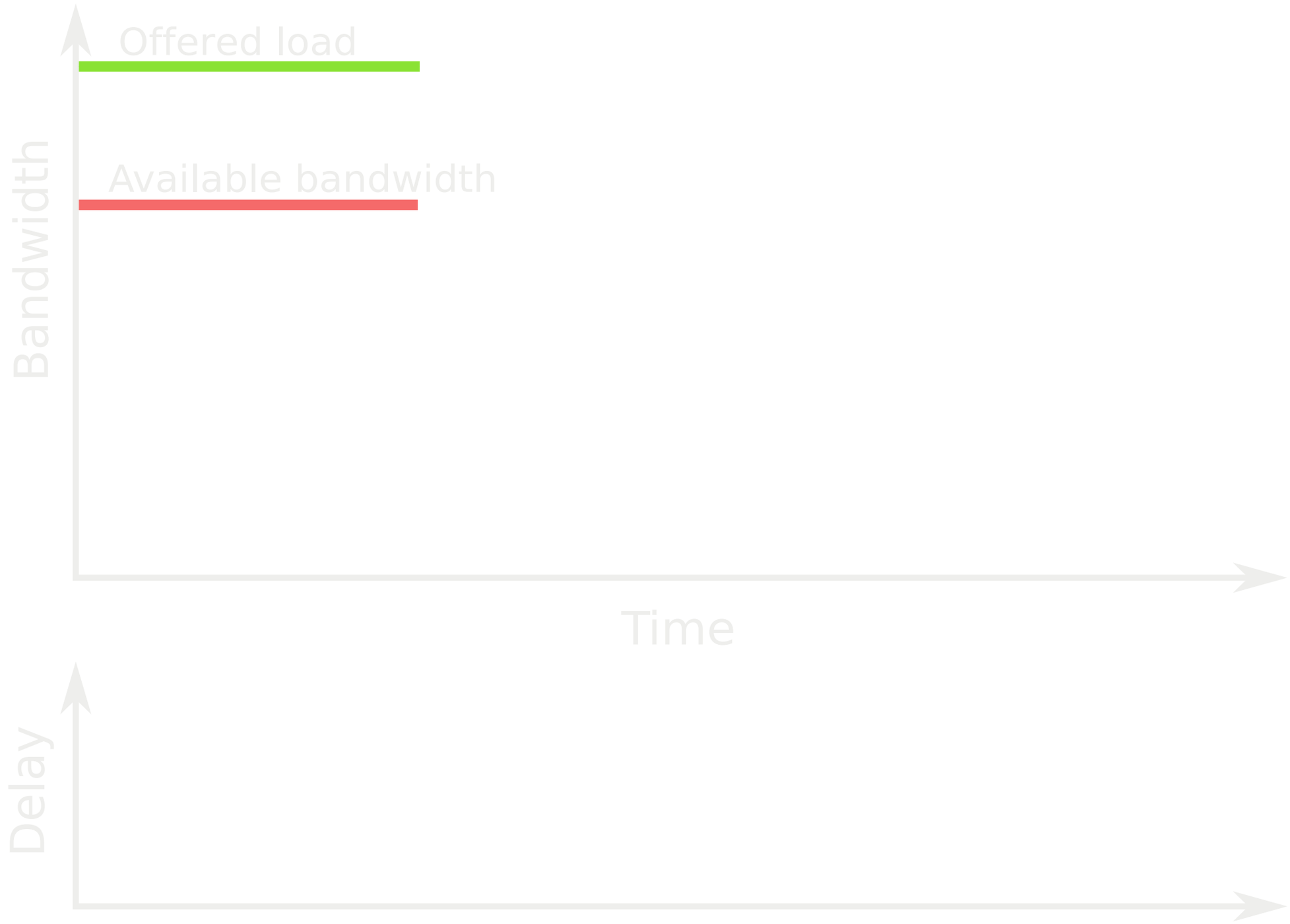
Packet Loss



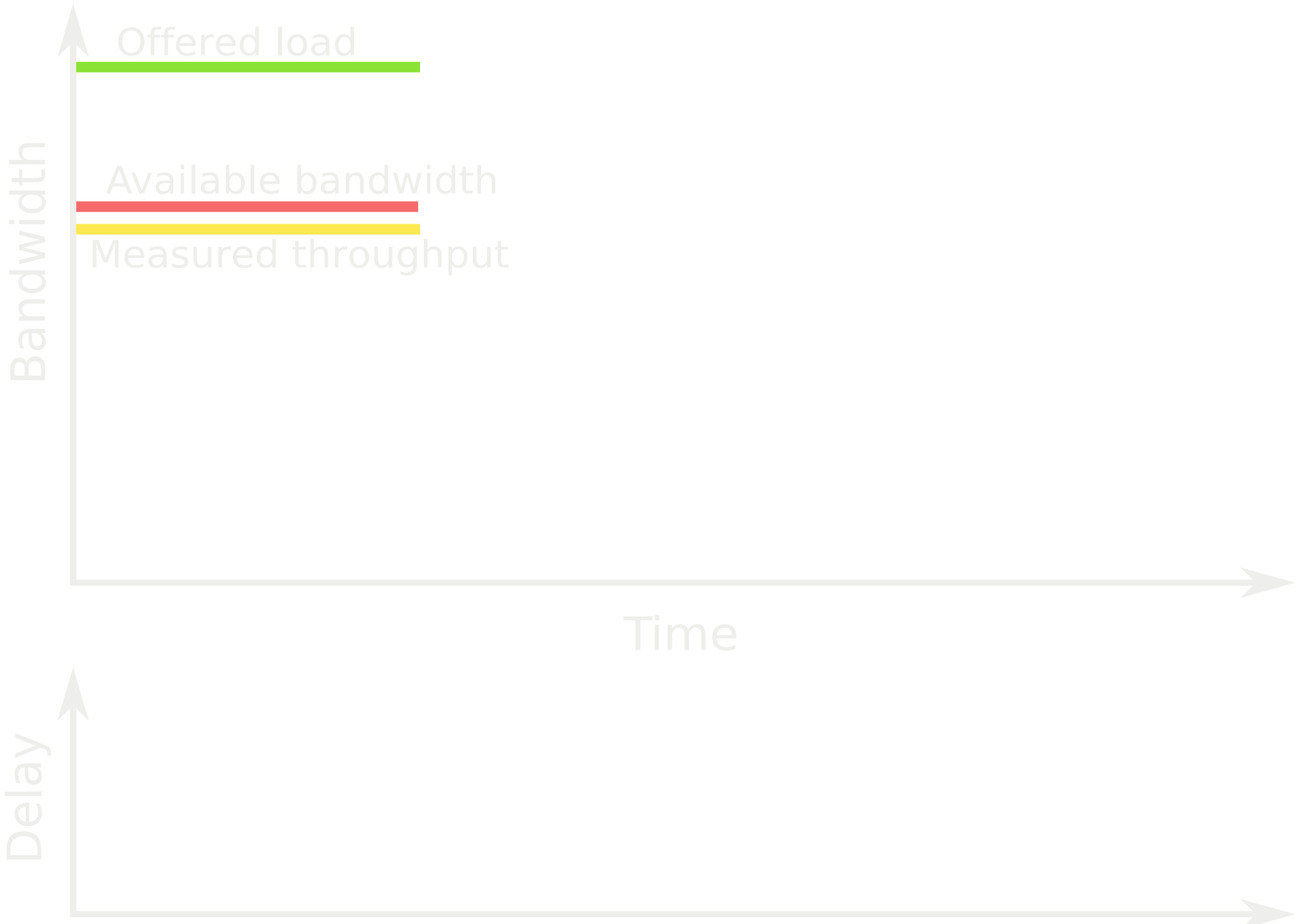
- Caused by full queue at bottleneck link
  - Difficult to measure directly
- So measure queue length in time: Max recent RTT - Base RTT

Throughput and ABW

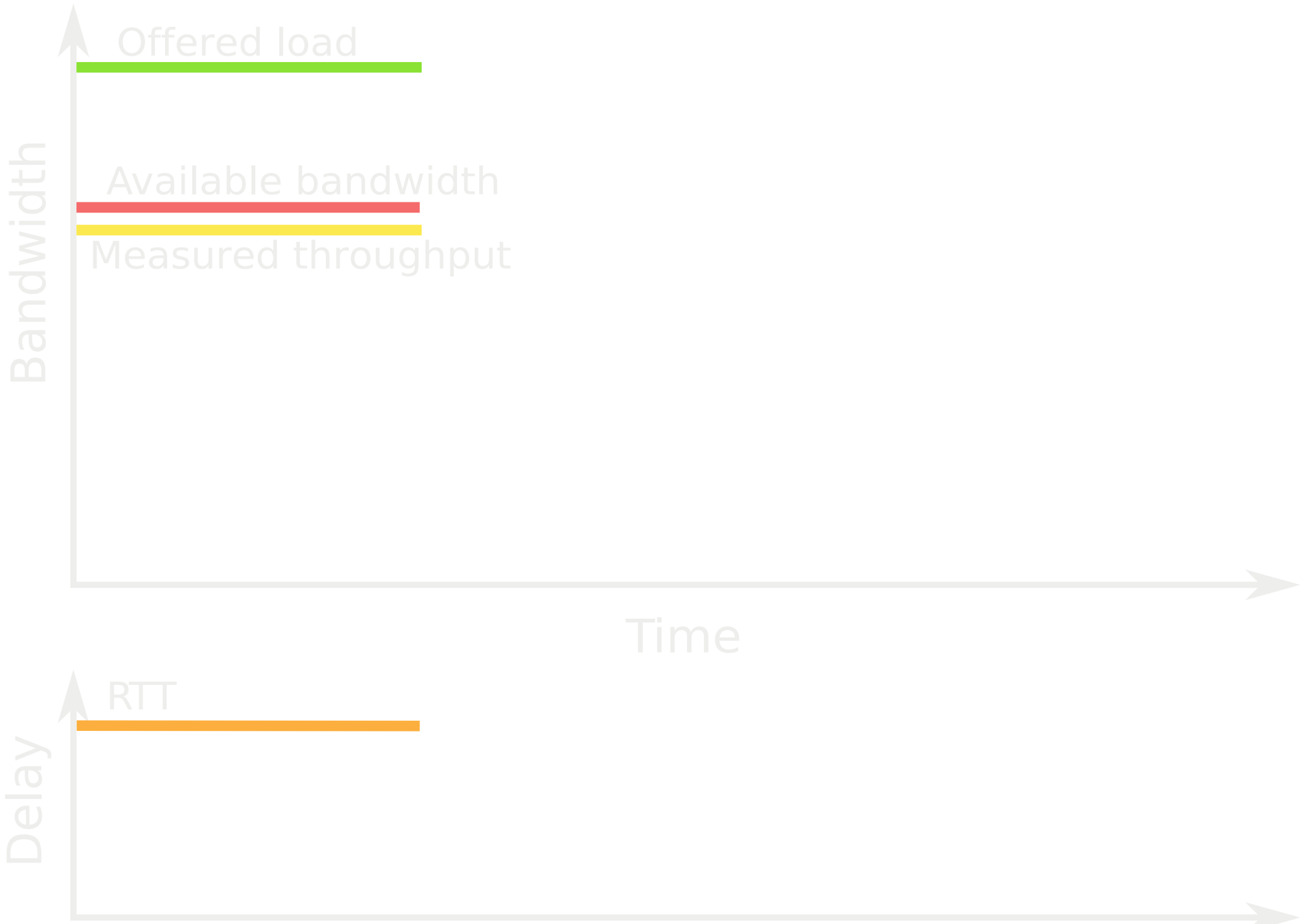
Throughput and ABW



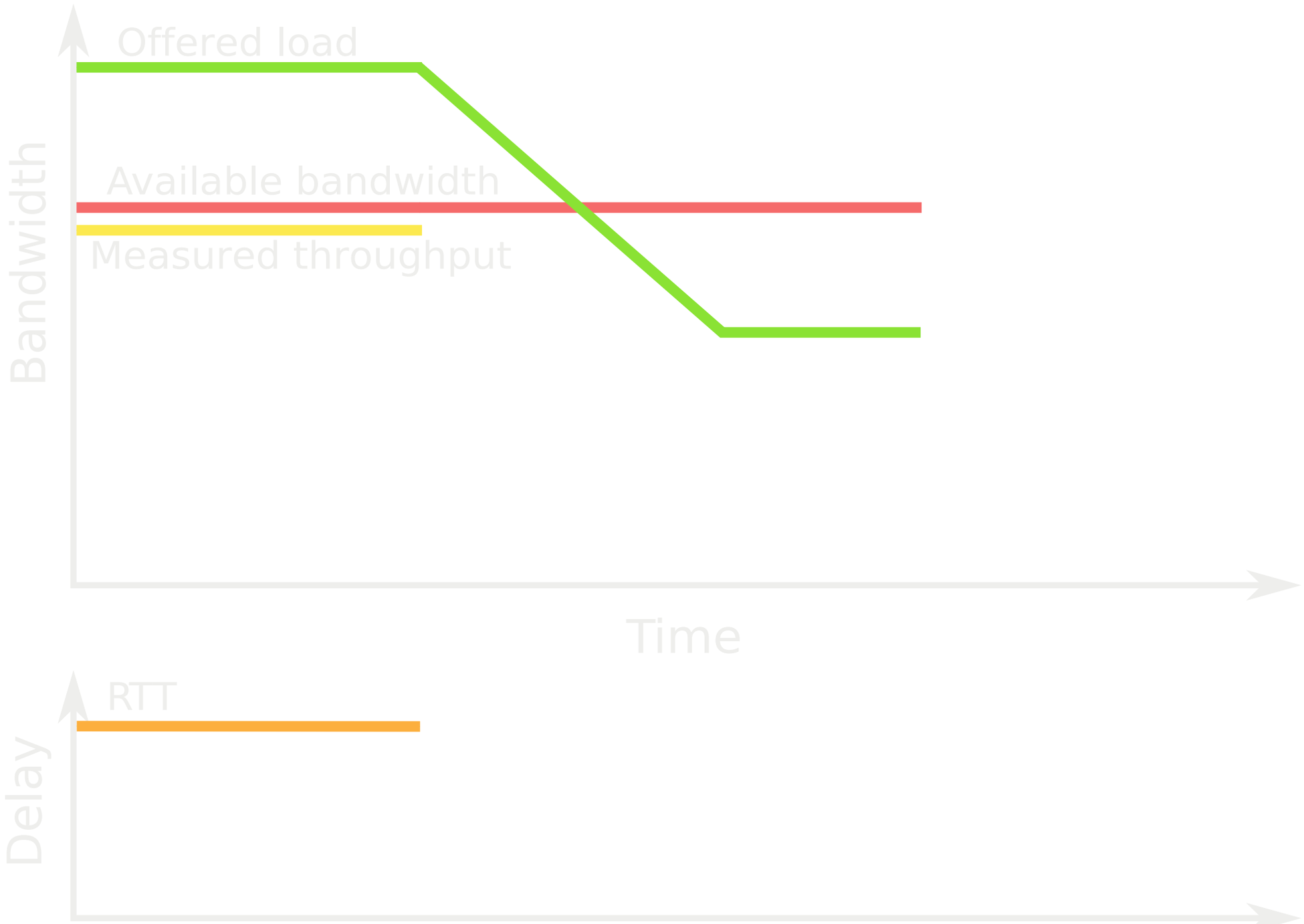
Throughput and ABW



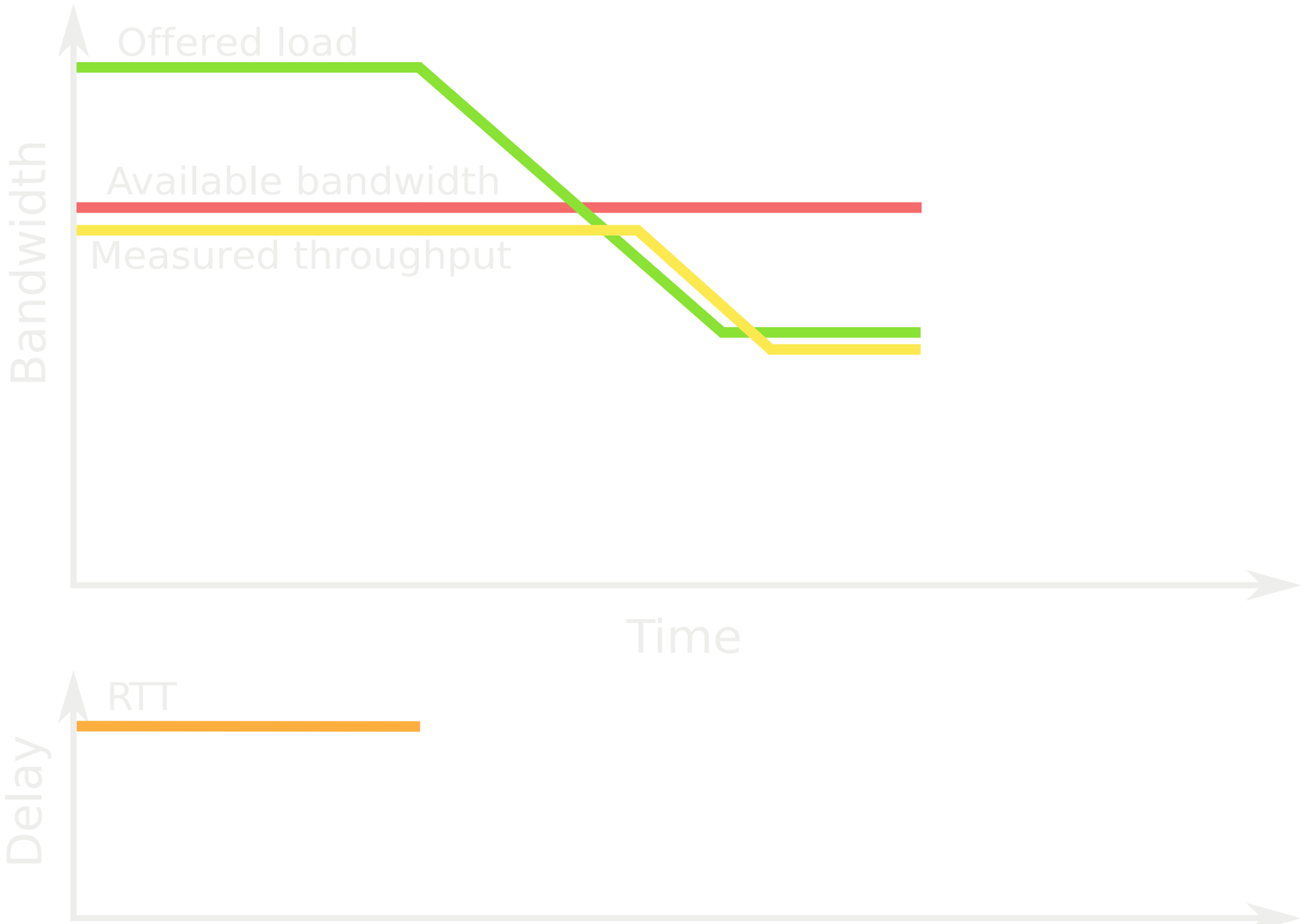
Throughput and ABW



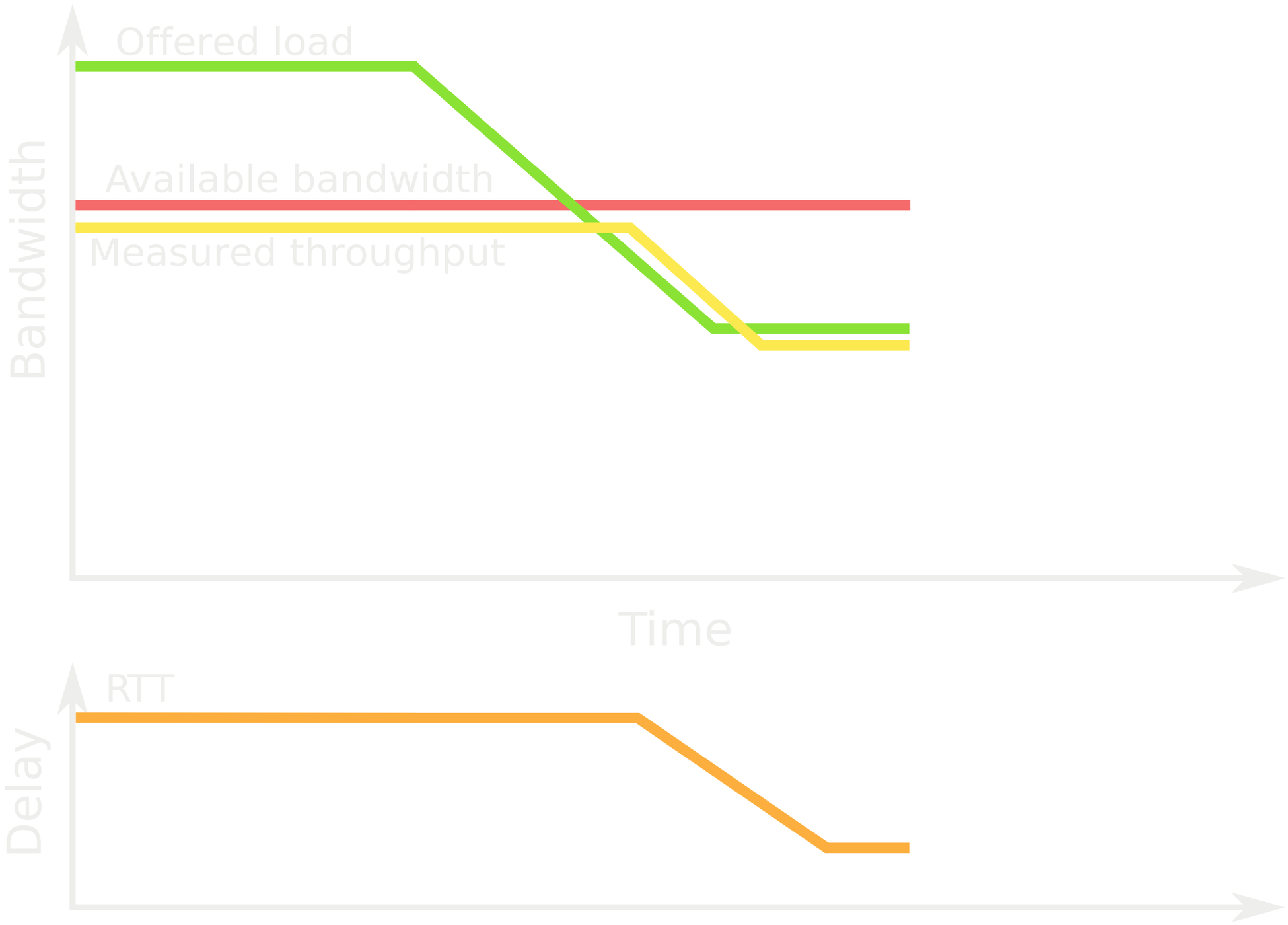
Throughput and ABW



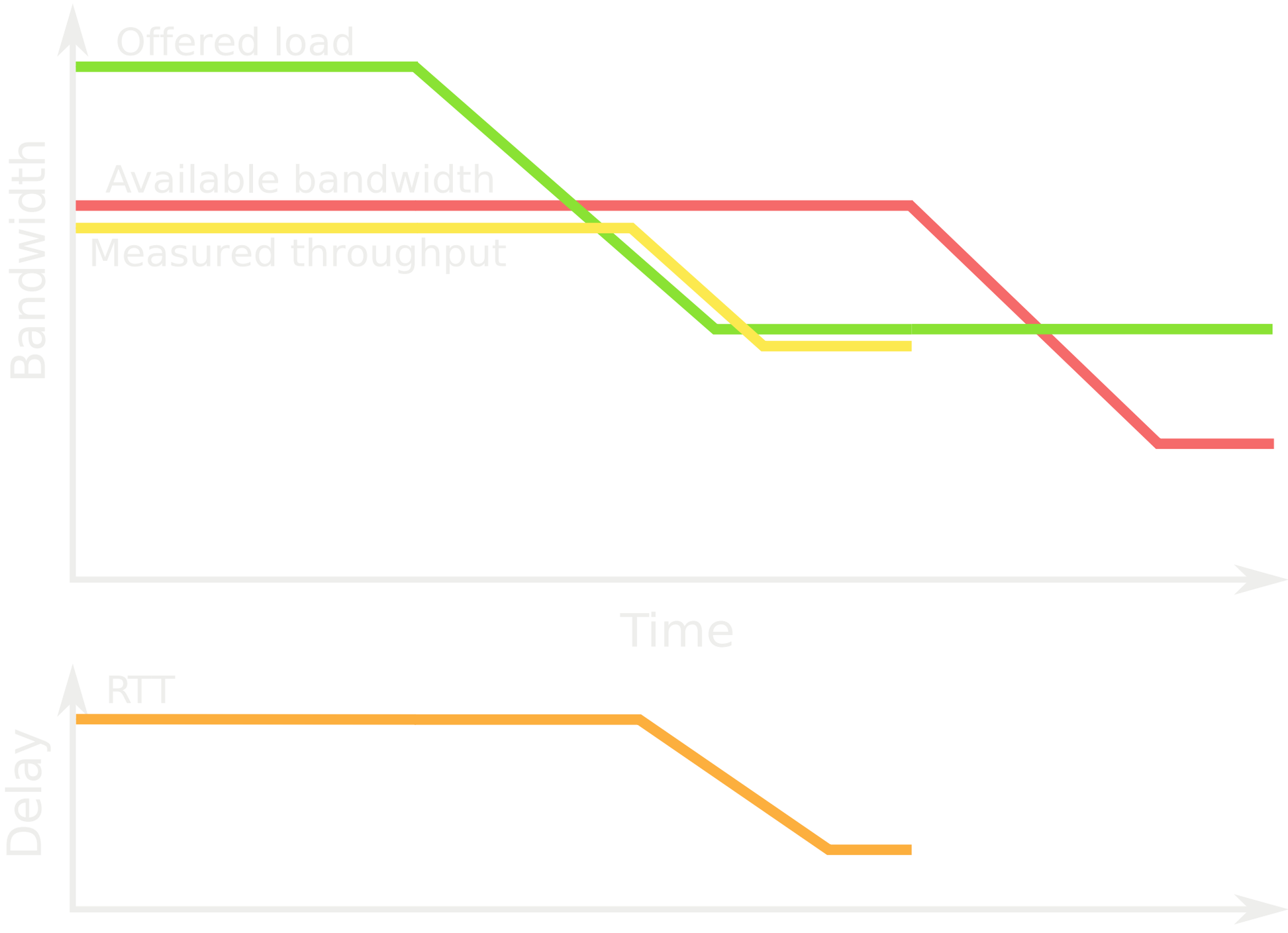
Throughput and ABW



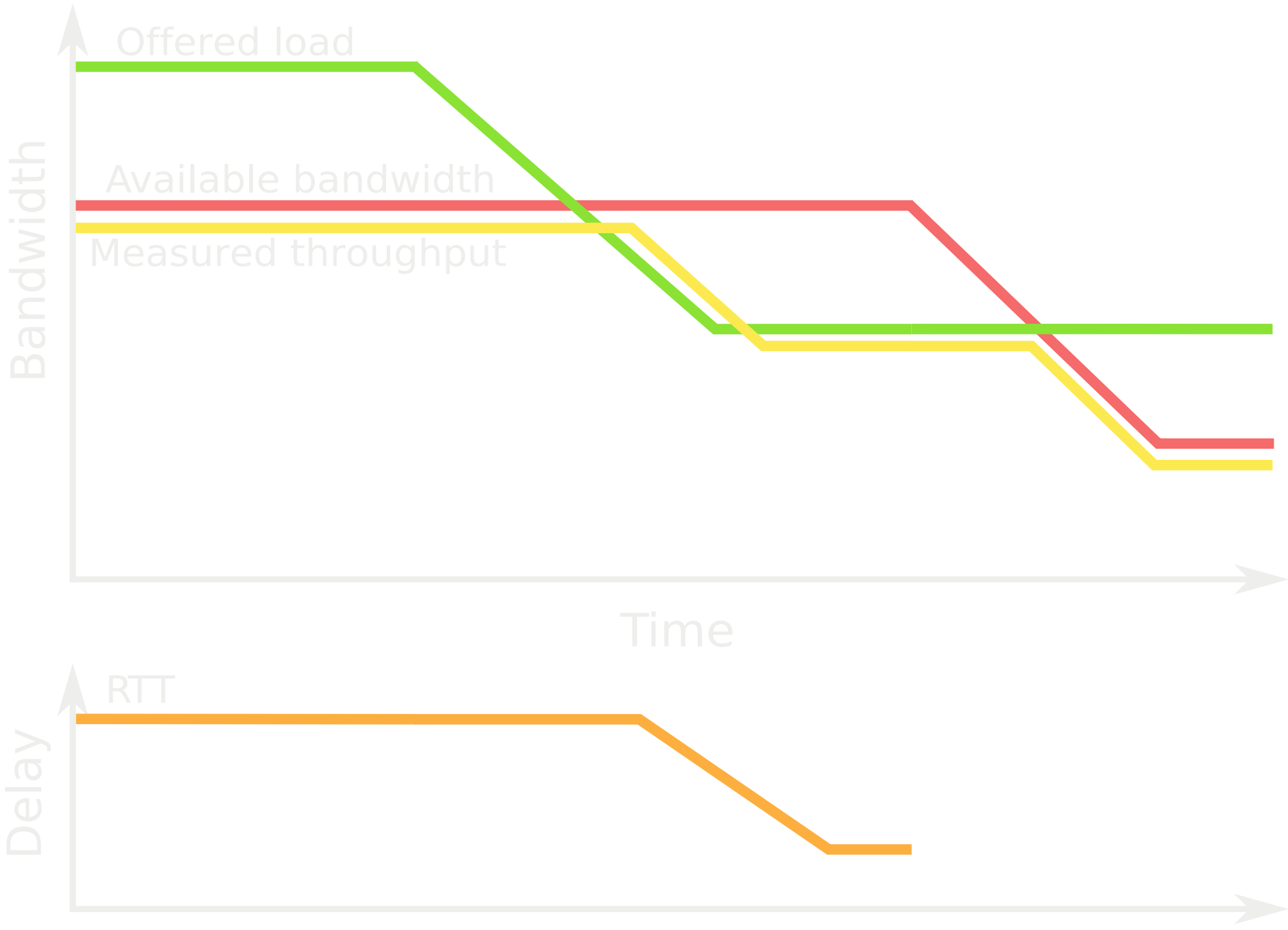
Throughput and ABW



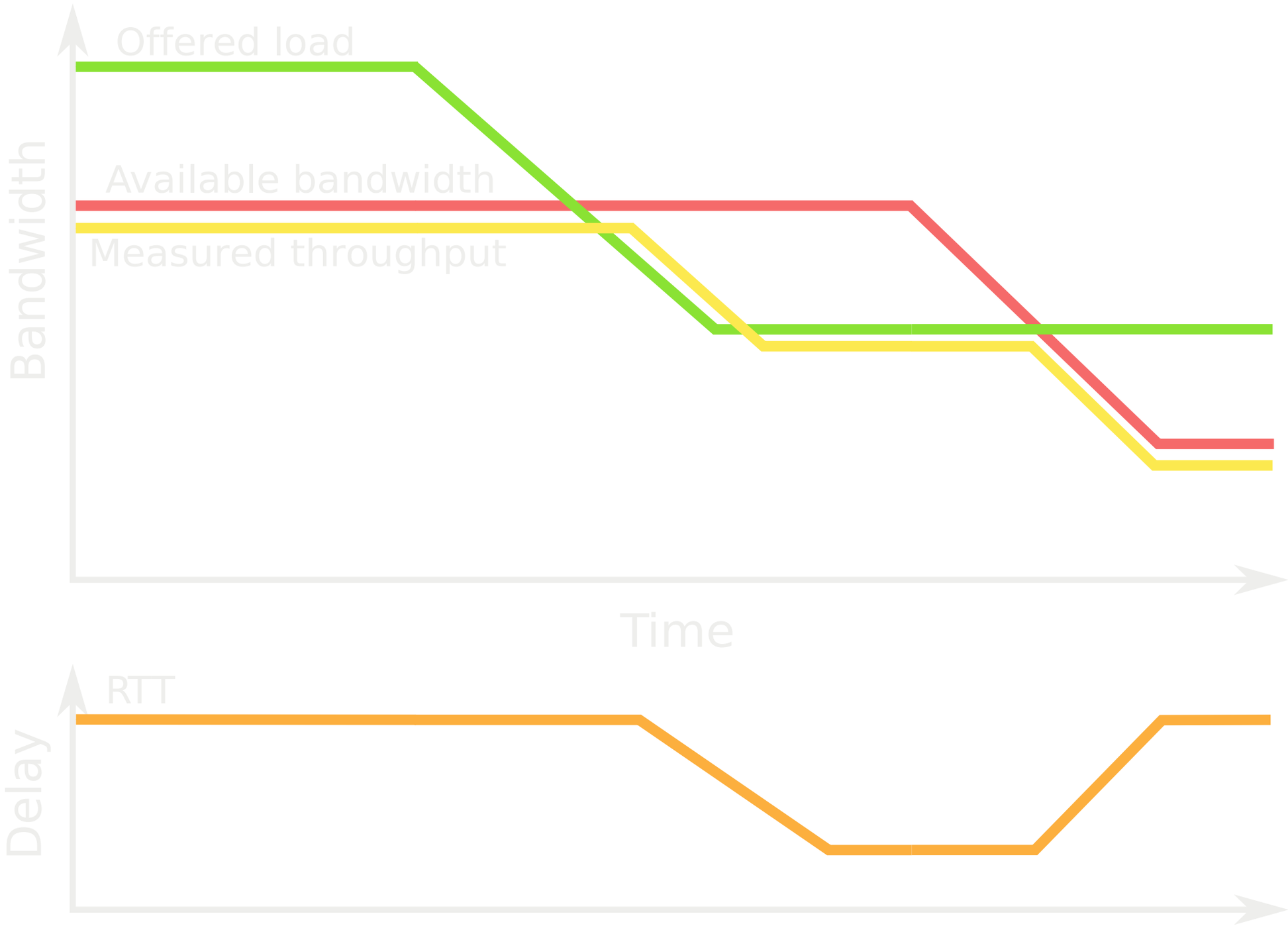
Throughput and ABW



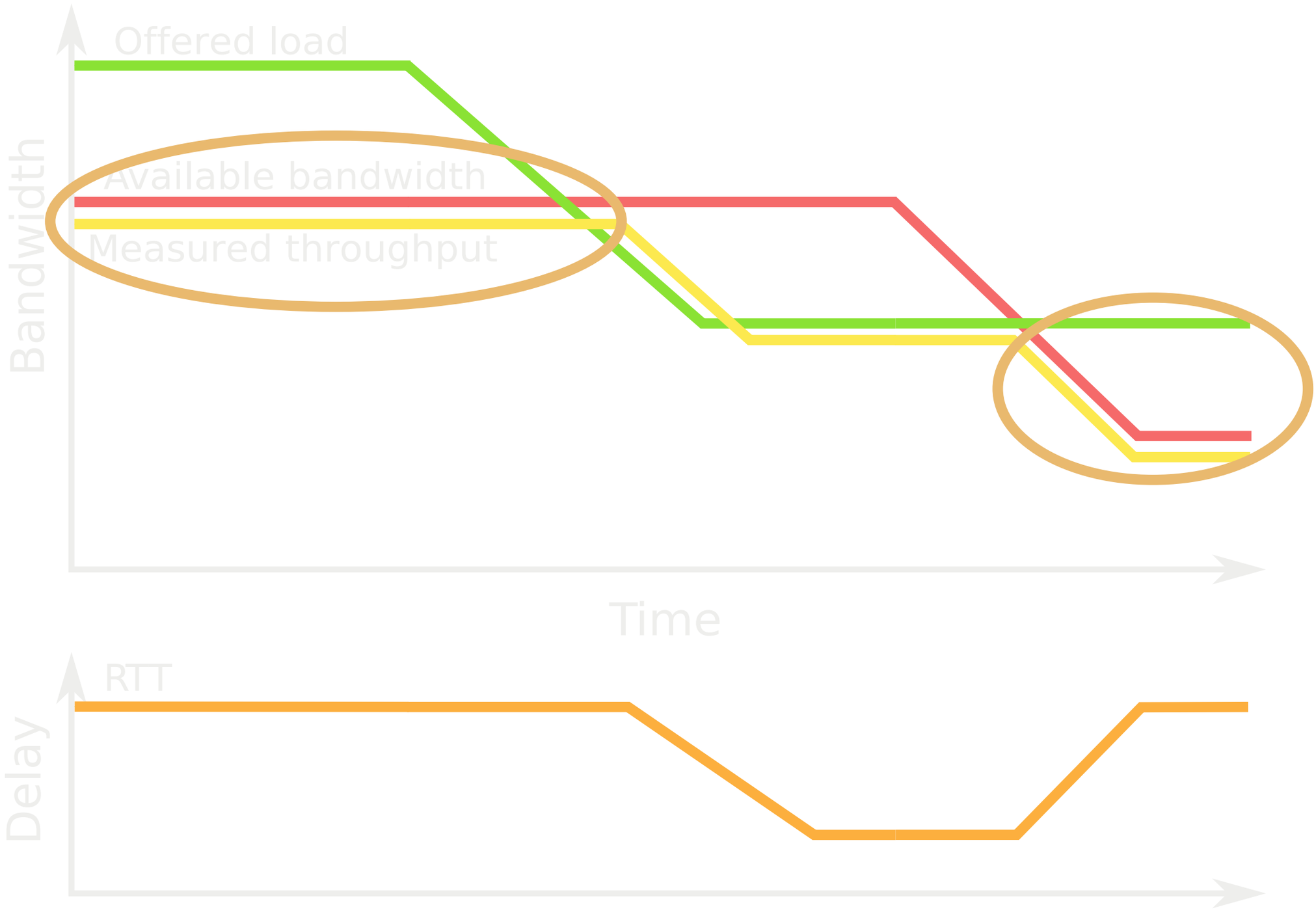
Throughput and ABW



Throughput and ABW



Throughput and ABW



Throughput and ABW

- If (throughput > last ABW measurement), use new value
- Else, look for indications that throughput has reached ABW
  - Socket buffer is filling up AND
  - Recent RTTs have been increasing
    - Using linear regression

ACIM Features

- Precise: assesses only relevant parts of the network
  - Scales in nodes and paths
- Complete: automatically captures all relevant network behavior
  - Simpler to measure e2e effects than find causes
  - Detects rare and transient effects
  - Evokes all reactive network behaviors (except content-based)
  - Rapidly tracks conditions

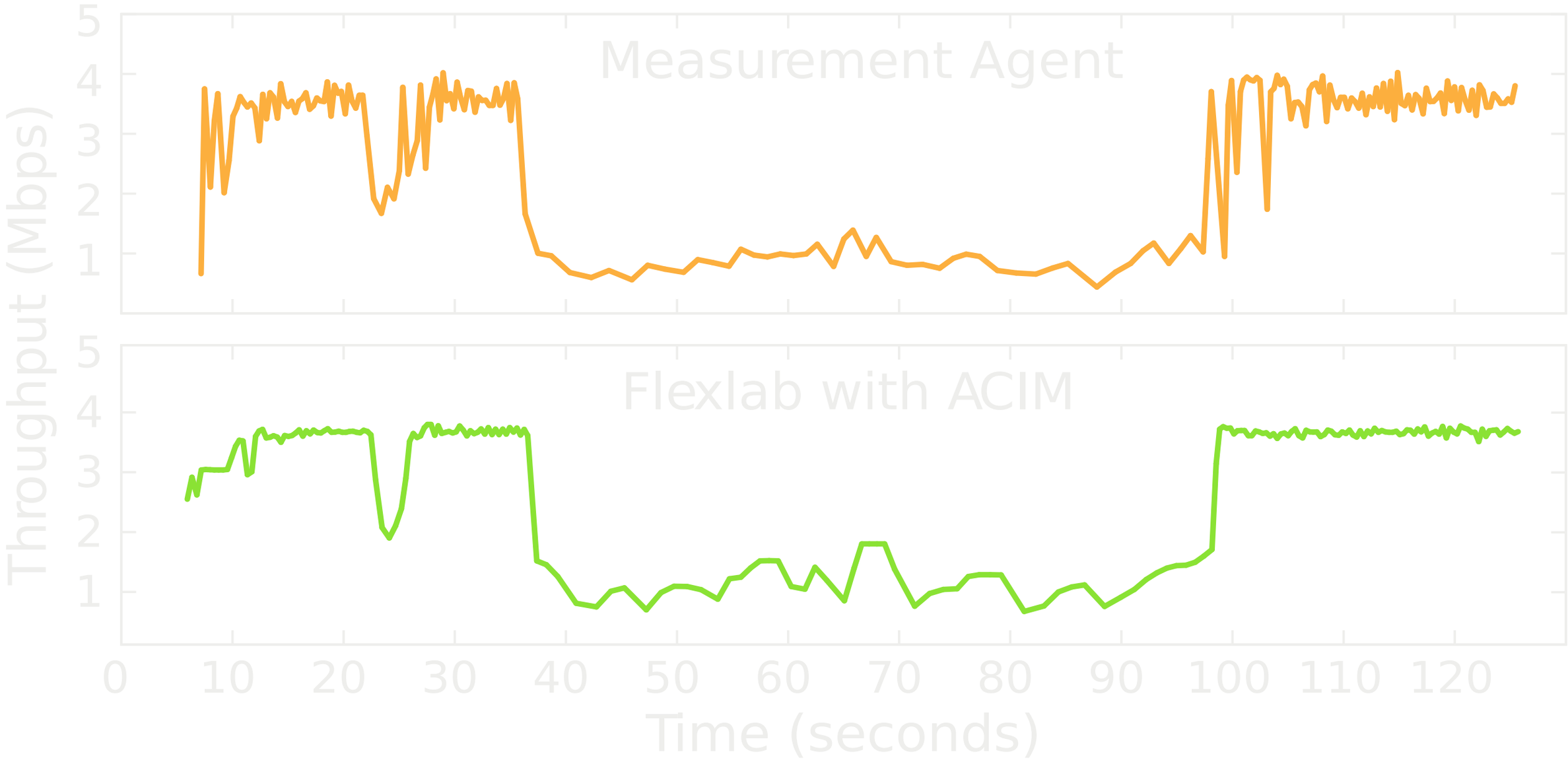
ACIM Accuracy

- Is ACIM path emulation accurate?
- Is it accurate at fine granularity?

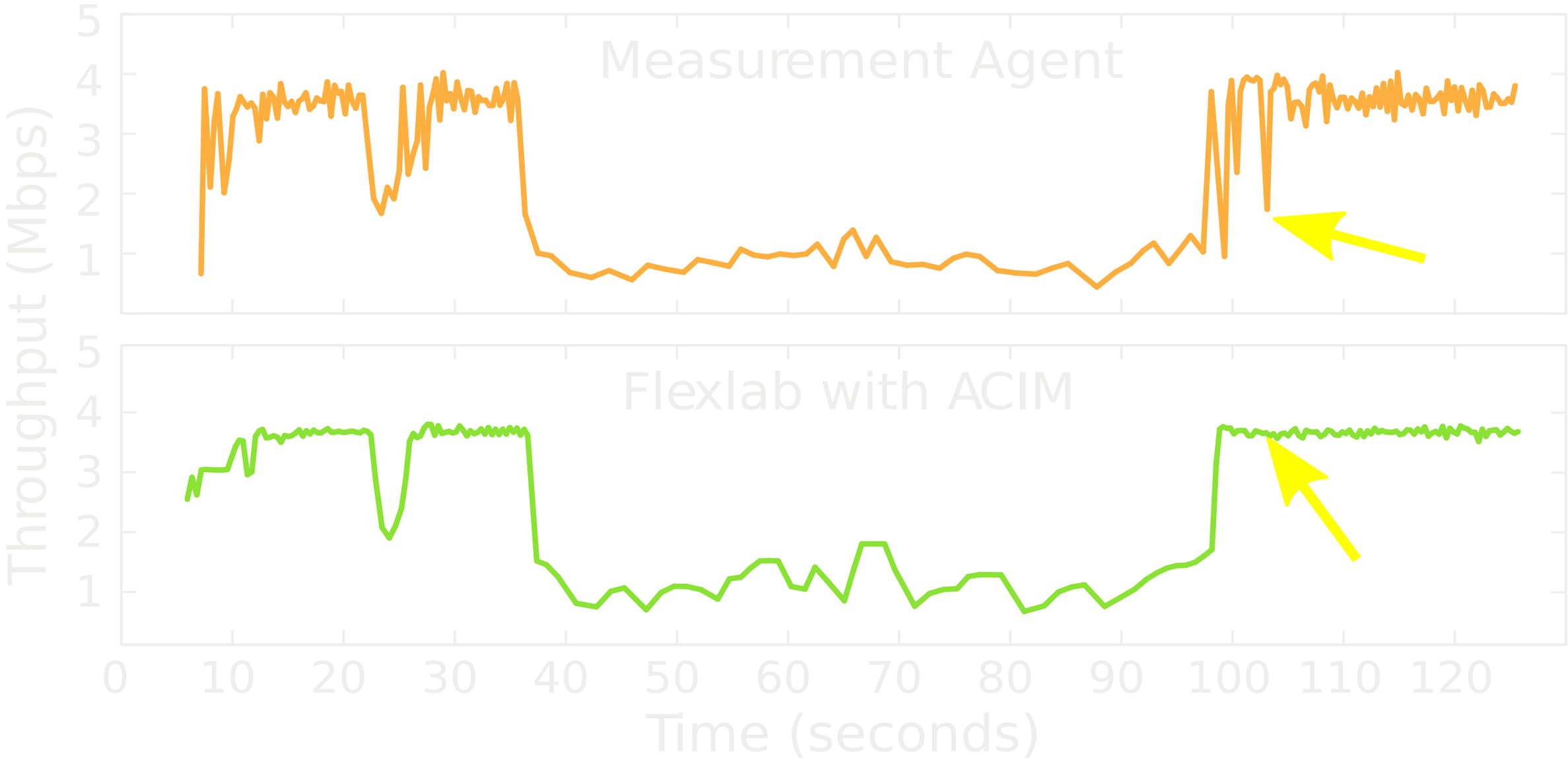
Methodology

- iperf runs in Emulab
- Measurement Agent runs on PlanetLab at UT Austin and AT&T Research
- We added transient TCP cross traffic between these sites

TCP iperf Throughput



TCP iperf Throughput



A Real Application

- Does ACIM give accurate results for a real,complicated application?

A Real Application

- Does ACIM give accurate results for a real,complicated application?
- … does PlanetLab?

A Real Application

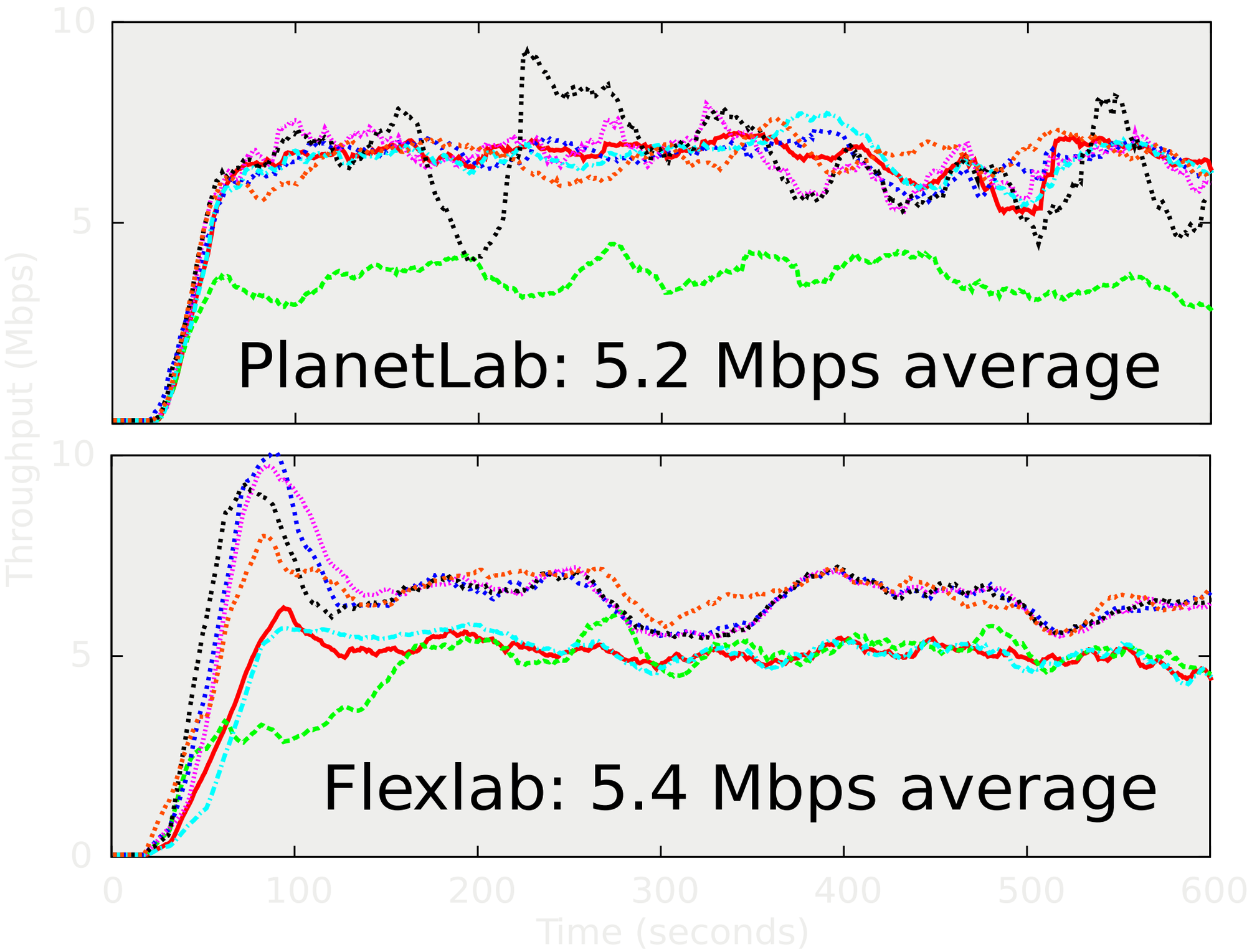
- Does ACIM give accurate results for a real,complicated application?
- … does PlanetLab?
- Can we discover ground truth?

Methodology: BitTorrent

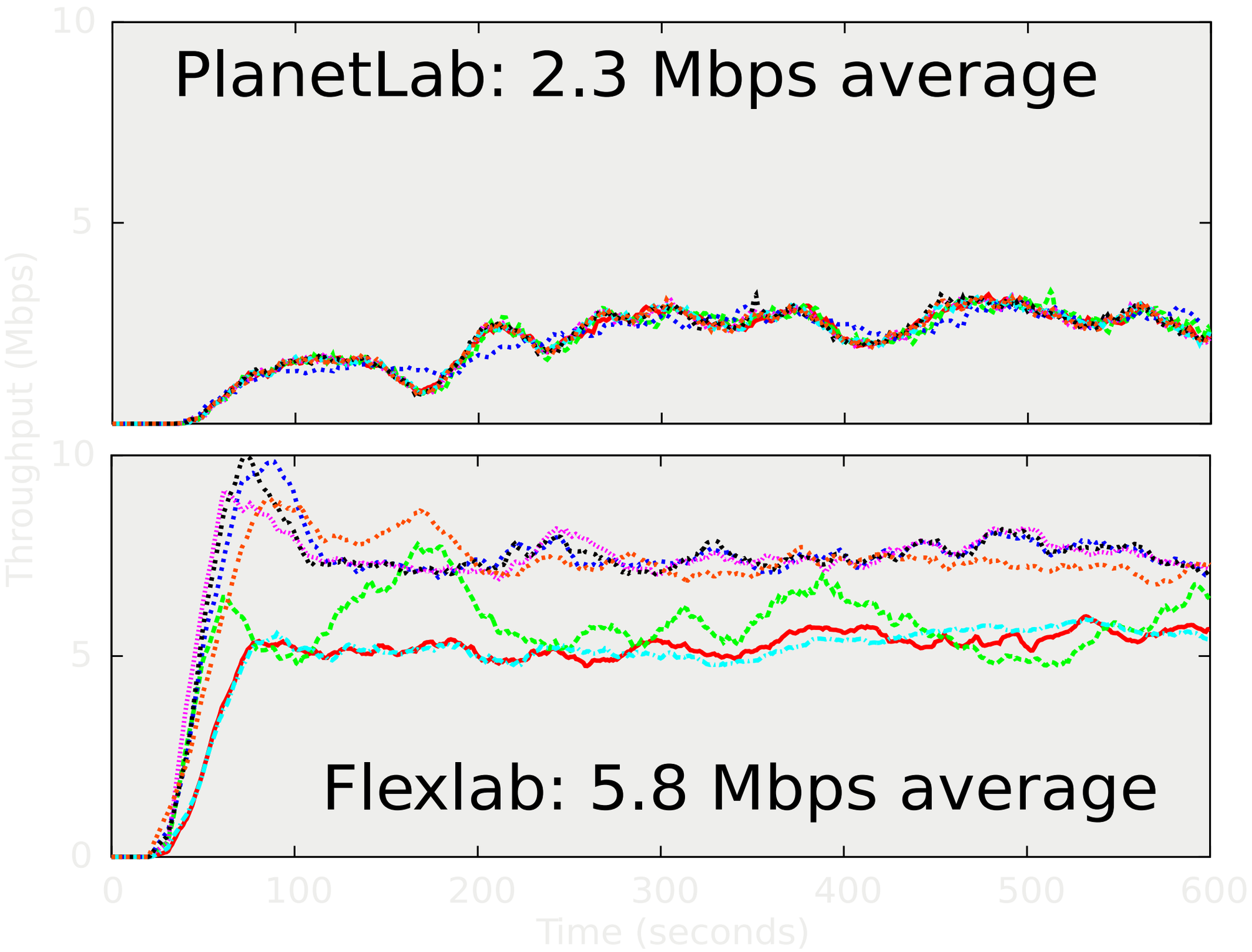
- Two simultaneous instances of reference BitTorrent:
  - One on PlanetLab
  - One in Flexlab
- Eight nodes in US and Europe: One seed, seven clients
- We reduced randomness in BT … but some still remains

BitTorrent w/ CPU Reservation

BitTorrent w/ CPU Reservation



BitTorrent w/o CPU Reservation



BitTorrent Bottom Line

- Conclusion: For this experiment, both Flexlab and PlanetLab with CPU reservations give accurate results
  - PlanetLab alone does not
- CPU availability on PlanetLab hurts BitTorrent
- ACIM reduces host resource needs on PlanetLab for this experiment
  - BitTorrent: 36-76% CPU
  - ACIM Agent: 2.6% CPU
  - Factor of 15 - 30 CPU
  - Factor of 4 memory

The Future?

- No need to perfect in PlanetLab:
  - Full resource isolation
  - Total control over hosts
  - Orthogonal control network
- … use in the emulators that already have them
- Use PlanetLab nodes as NICs
- Conserve resources for deployed services with end users

Conclusion

- New approach to evaluating networked systems
- Separates the network model
- Designed to leverage vibrant measurement and modeling community
- Couples an emulator to an overlay testbed
- ACIM high fidelity emulation technique
- Contact testbed-ops@emulab.net to use

Backup Slides

Why not just add more nodes to every PlanetLab site?

- Remaining problems:
  - Poor repeatability
  - Hard to develop/debug
  - No privileged operations
- Some malicious traffic cannot be tested
- Some Flexlab network models reduce network load
- Emulab node pool stat muxed and shared more efficiently than per-site pools
- Overload can (will?) still happen with PL's pure shared-host model
- Major practical barriers: admin, cost

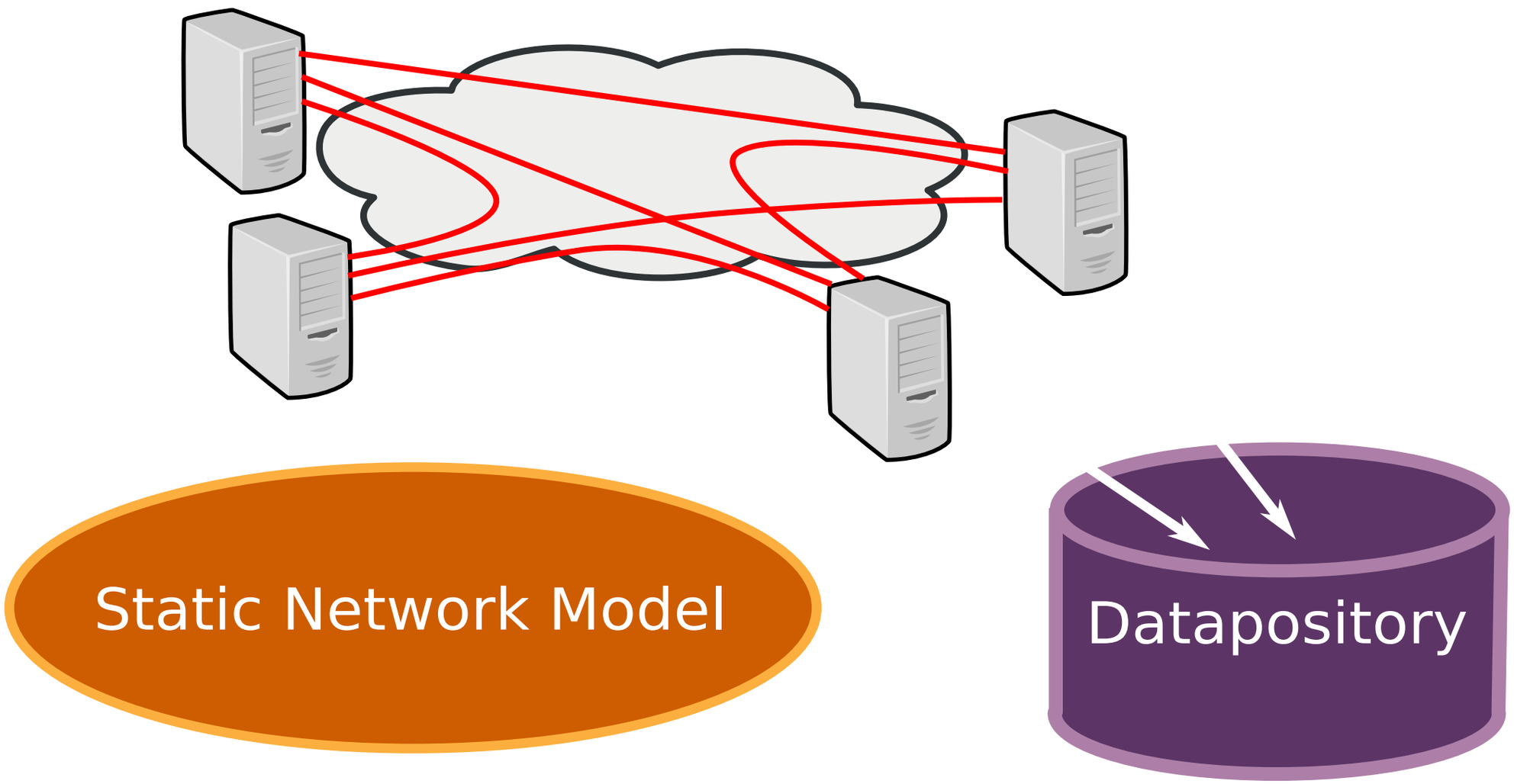
Flexlab and VINI

- Entirely different kinds of realism and control
- Flexlab: passes "experiment" traffic over shared path
  - Real Internet conditions from other traffic on same path, but app. traffic is not from real users
  - Control: of all software
  - Environment: friendly local dev. environ, dedicated hosts
- VINI: can pass "real traffic" over dedicated link
  - Real routing, real neighbor ISPs, potentially traffic from real users, but network resources are not realistic/representative
  - Dedicated pipes with dedicated bandwidth, that insulate experiment from normal Internet conditions
  - Control: restricted to VINI's APIs (Click, XORP, etc.)
  - Environment: distributed environ; shared host resources

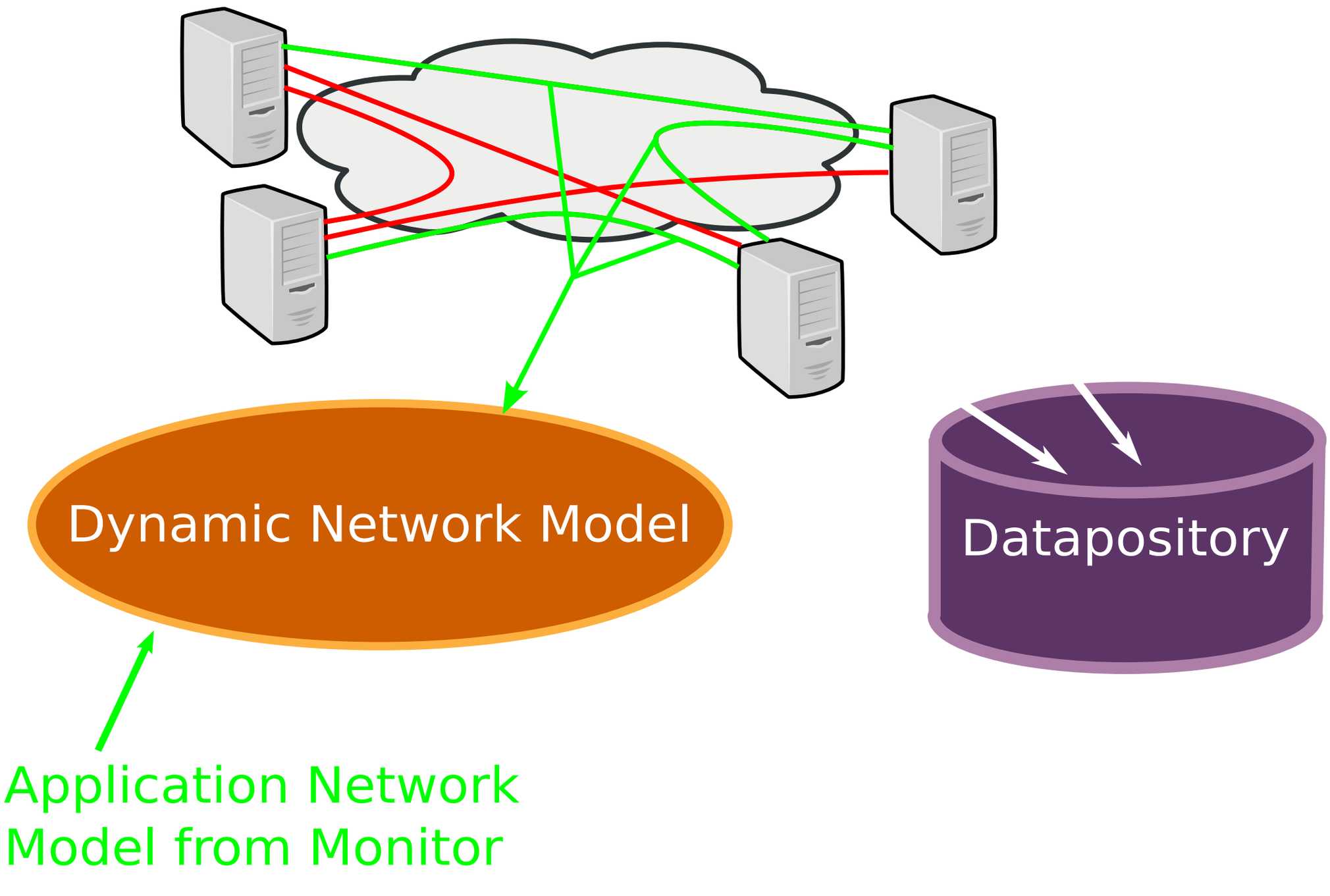
Change Point Analysis

Path	High	Low	Change
	39%		
	15%		
			12%

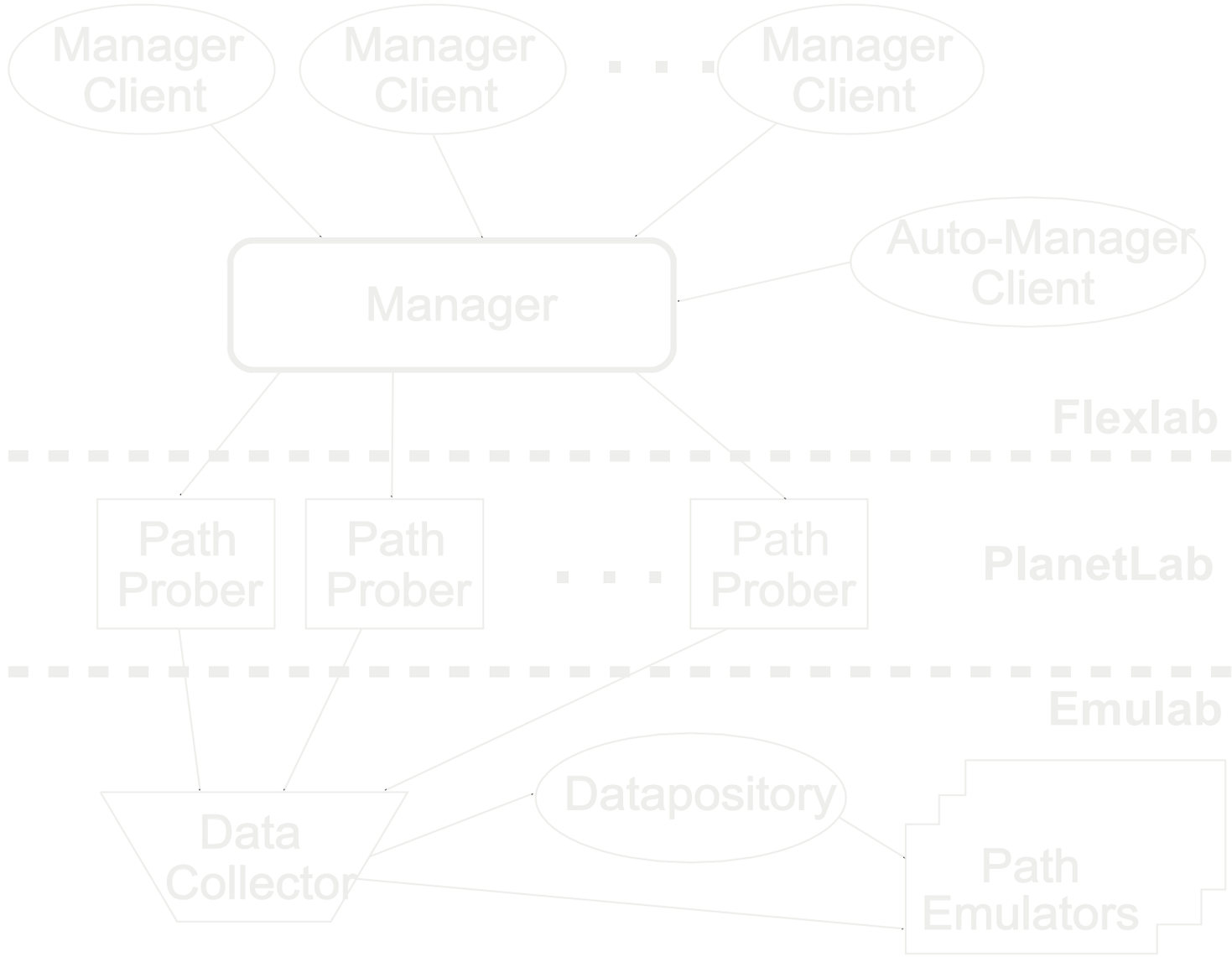
Simple Static Model



Simple Dynamic Model



Flexmon Architecture



- Shared
- Reliable
- Safe

- Adaptive
- Controllable
- Accommodates high-performance data retrieval

CPU Starvation on PlanetLab

- Host Artifacts
  - Long period when agent can't read or write
  - Empty socket buffer or full receive window
  - Solution: Detect and ignore
- Packet loss from libpcap
  - Long period without reading libpcap buffer
  - Many packets are dropped at once
  - Solution: Detect and ignore

Reverse Path Congestion

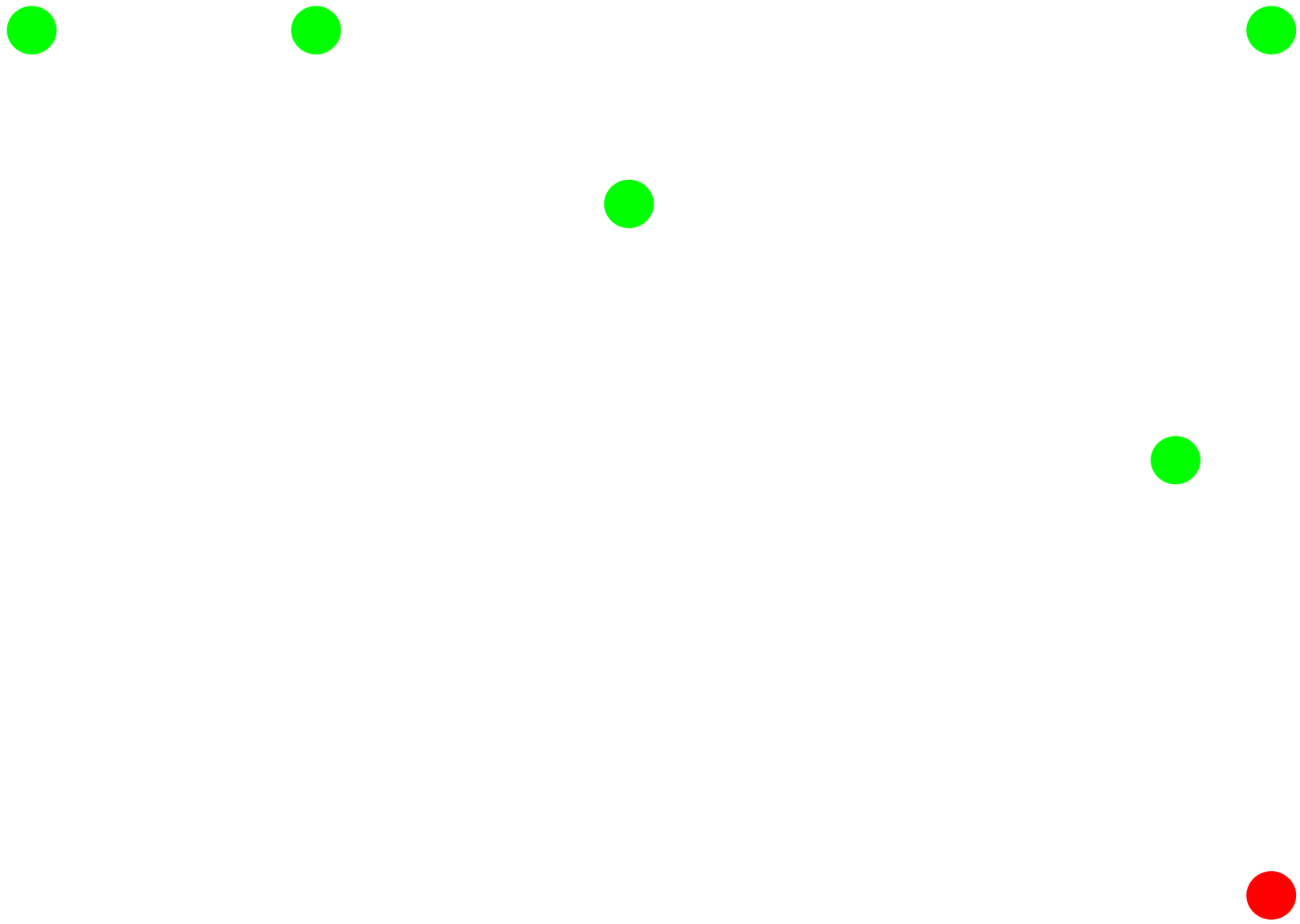
- Can cause ack compression
- Throughput Measurement
  - Throughput numbers become much noisier
  - We abuse the TCP timestamp option
  - PlanetLab: homogeneous OS environment
  - Extending it would require hacking client
- RTT Measurement
  - Future work

Initial Conditions

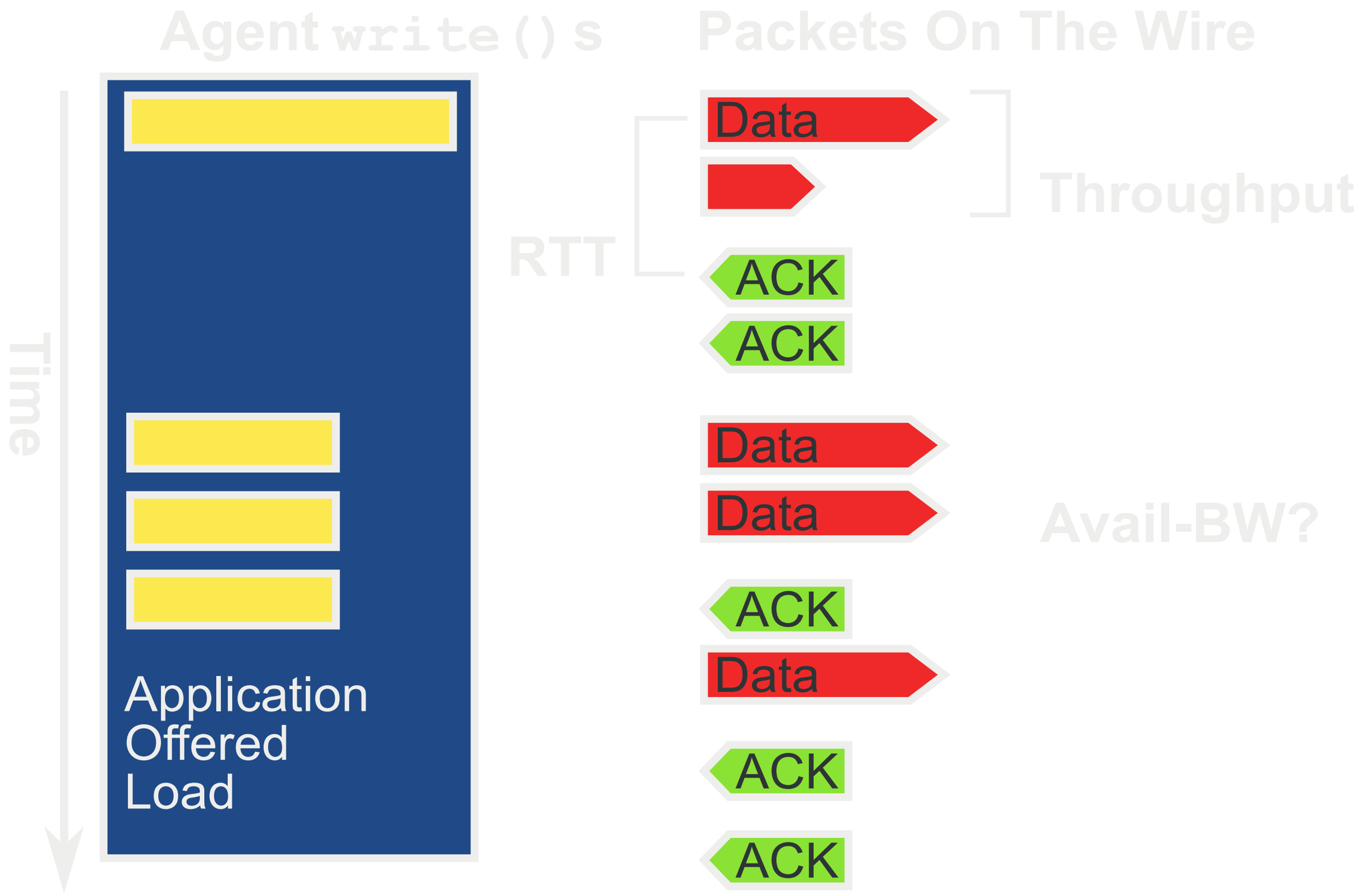
- Needed to bootstrap ACIM
  - ACIM uses traffic to generate conditions
  - But conditions must exist for first traffic
- We created a measurement framework
  - All pairs of sites are measured
  - Put data into measurement repository
  - Set initial conditions to latest measurements

Simultaneous TCP iperf

Repeatability vs. Fidelity

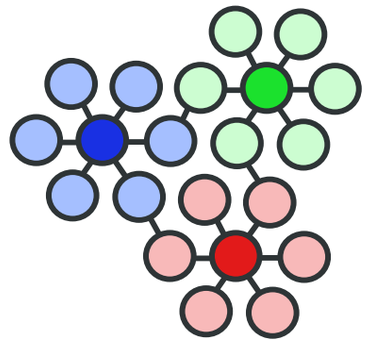


Throughput and ABW

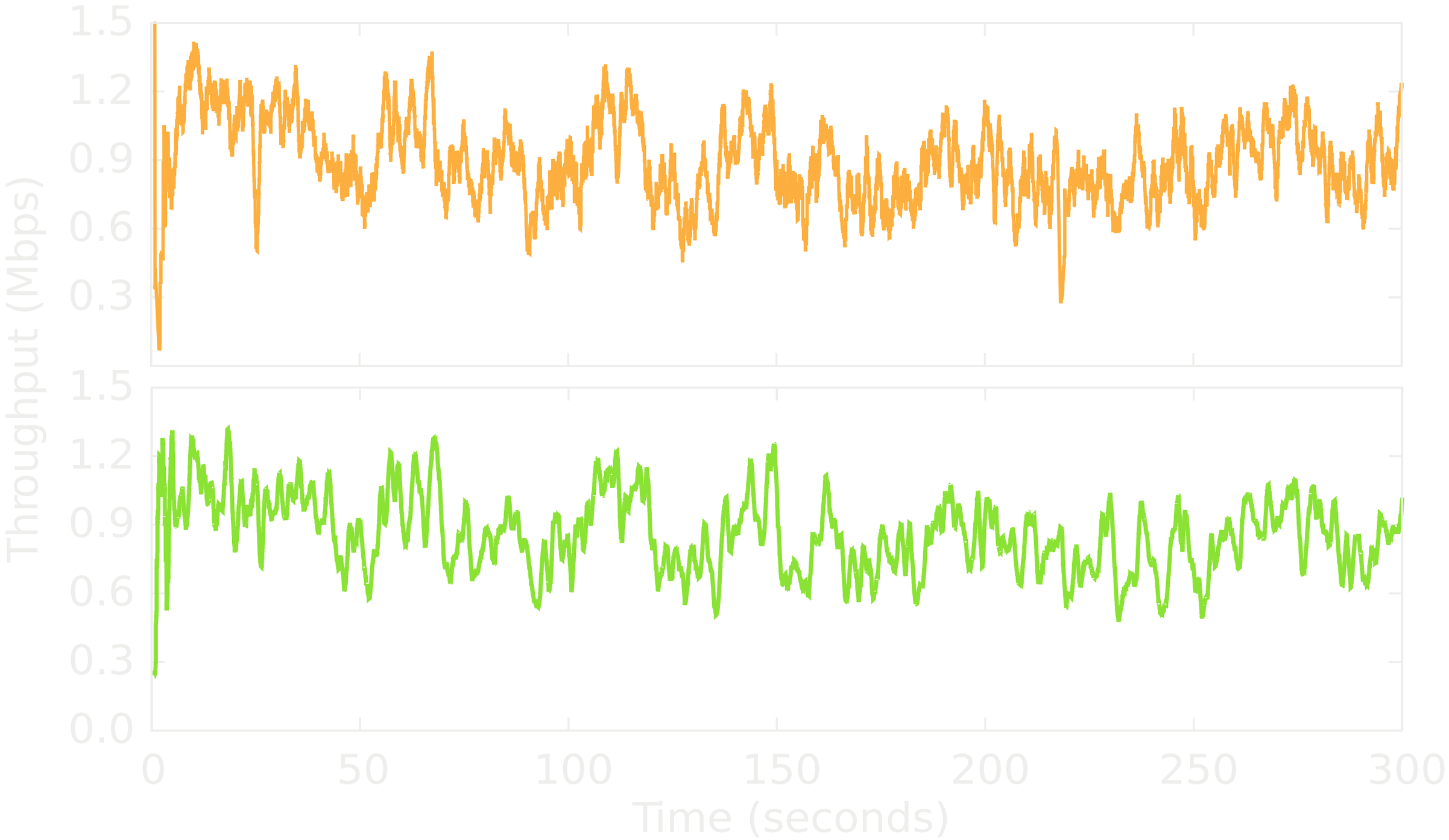


Currently available for
Beta Testing

http://www.flux.utah.edu/flexlab



UDP Streaming Video



Opens Up New Questions

- Further validation
- Accuracy tests at runtime
  - Similar in spirit to Emulab's linktest
- Use to compare models
  - Find which models most appropriate for different classes of applications
- Replay for ACIM
- Study fidelity of different software combinations
  - Different TCP implementation or OS in Emulab