Packet Marking for Integrated Load Control

IM 2005, Nice, France

Martin Karsten

School of Computer Science, University of Waterloo
200 University Ave W, Waterloo, Ontario N2L 3G1, Canada
mailto:mkarsten@uwaterloo.ca
http://www.uwaterloo.ca/~mkarsten

Presentation Overview

Introduction

System Design

Admission Control

Load Estimation

System Prototype

Evaluation

Wrap Up

Introduction

QoS in the Internet

- long debated, many research proposals, limited deployment
- concerns: end-to-end scope, technical complexity, scalability
- technical tools
  - admission control and traffic regulation
  - differentiated packet scheduling

Our Goals

- admission control and load-based routing for network domain
- ultra-scalability: low complexity and little functionality in core
- modularization of control functions

Basic Mechanism: Binary Packet Marking

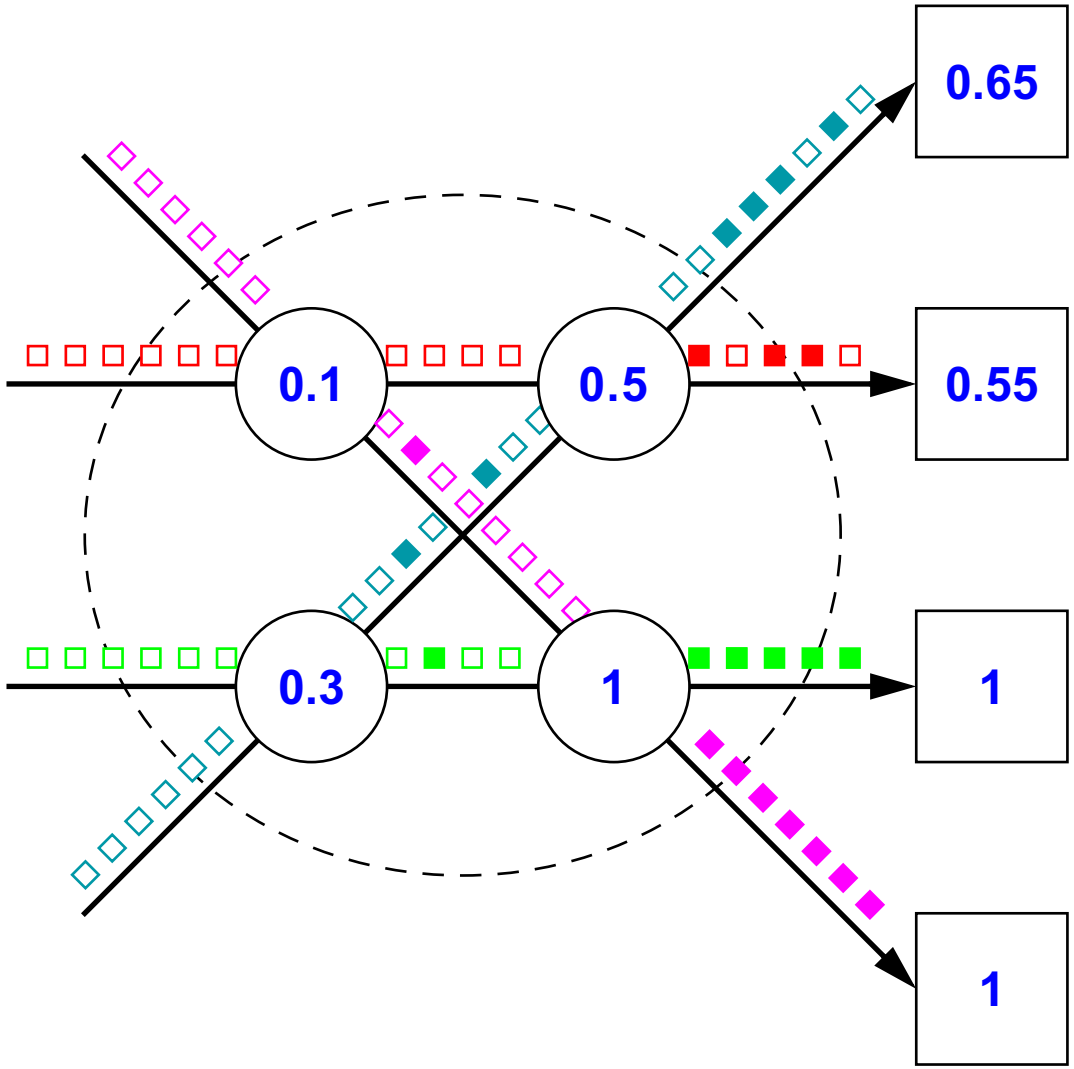
- Kelly et al. - results for strictly concave utility curves
- TCP, TFRC and ECN
- other admission control proposals, e.g. RMD

Key Aspect

- support admission AND load-based routing

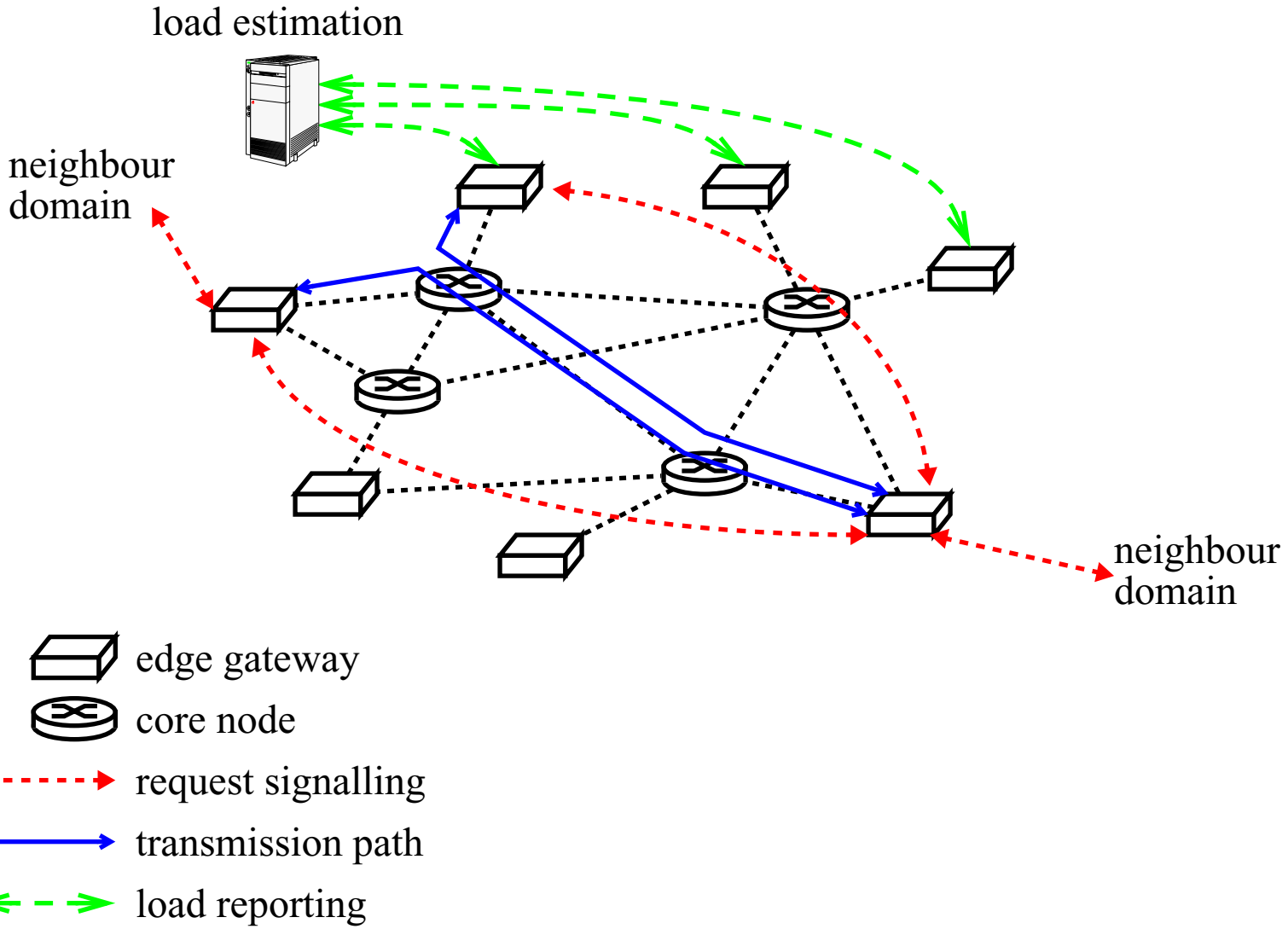
Binary Packet Marking

Single Marking Bit $\Rightarrow$ Path Marking Rate: $M_p = 1 - \prod_{i \in p} (1 - m_i)$



System Design

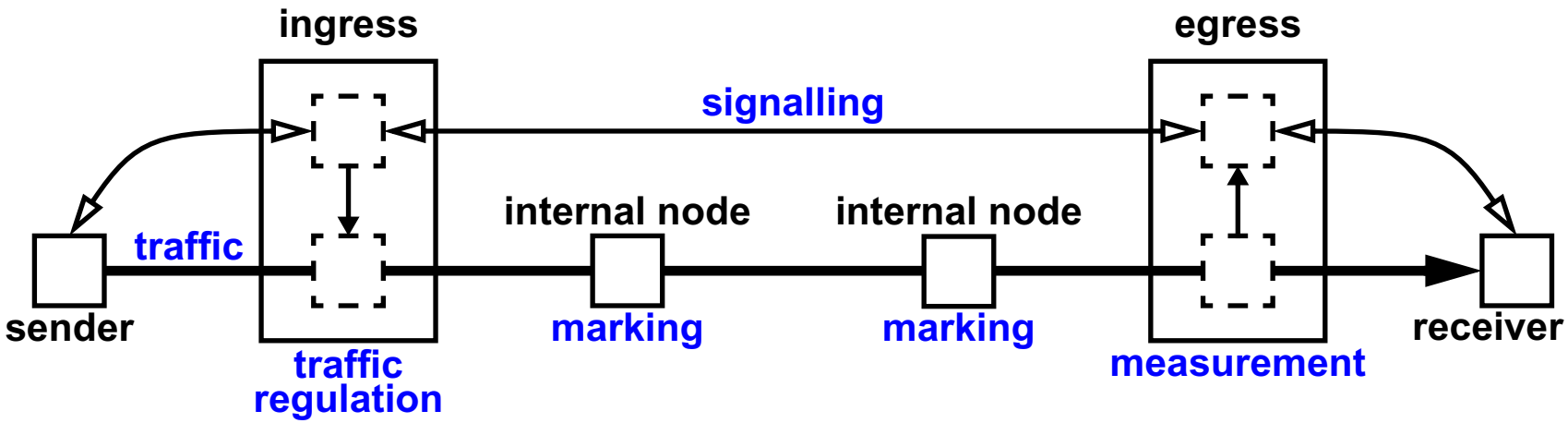
Network Domain Overview



System Design - Details

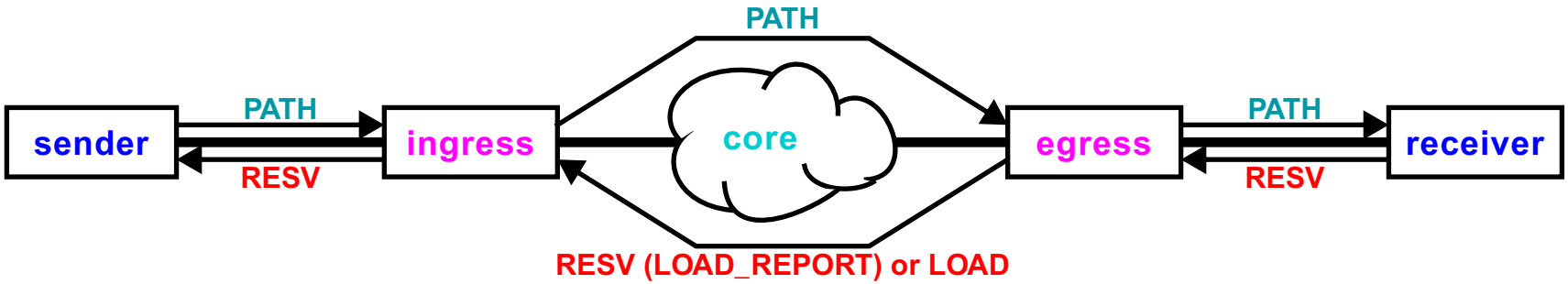
Combination of Proactive & Reactive Resource Allocation

- proactive resource allocation based on network load feedback



- inherent feedback delay between egress and ingress

Signalling Design: RSVP Extension



- purely local extension (LOAD_REPORT object & LOAD message)

Admission Control

Admission Control

- **decide about connection acceptance**
- **inherent per-domain concept**
  - reliability of end systems
- **feedback between edge gateways**
- **suitable for inelastic flows**

vs. Flow Control

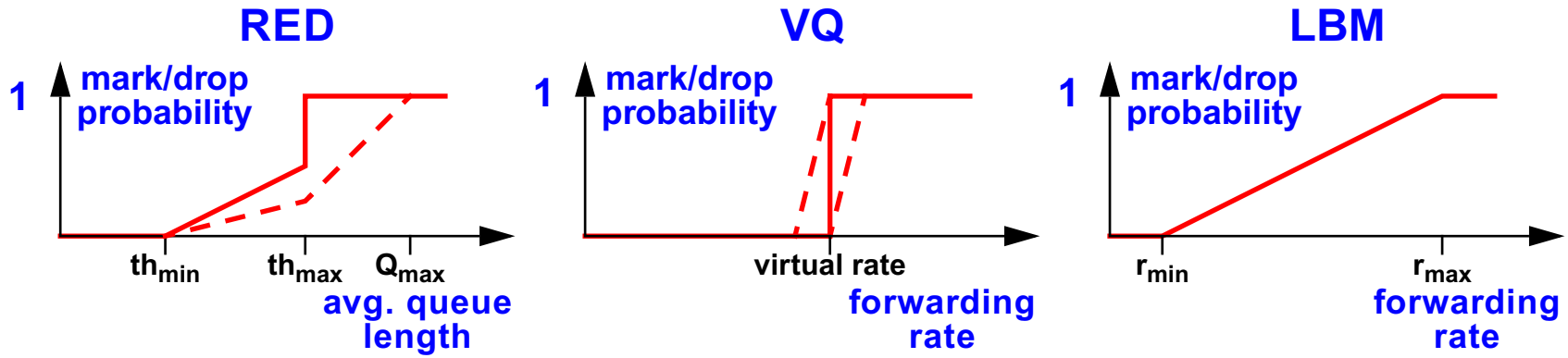
- **throttle sending rate according to load signal**
- **feedback from receiver**
- **suitable for elastic flows**

$\Rightarrow$ Support for Both Needed

Feedback Delay

- **traffic load $\rightarrow$ measurement $\rightarrow$ marking/transmission $\rightarrow$ measurement**
- **admission control at ingress vs. egress $\rightarrow$ no difference**
  - measuring always at egress
  - traffic control always at ingress
- **safety margin in resource utilization**
- **vs. arrival of service requests?**

Packet Marking Algorithms



Random Early Detection (RED) & Variants

- queue-based feedback
- ineligible packets → random drop (ok)
- meaning of path marking rate for inelastic flows (?)

Virtual Queue (VQ) & Variants

- hybrid feedback, time-scale dependent
- ineligible packets → bursty dropping (?)
- inelastic flows → binary path marking rate

Load Based Marking (LBM)

- rate-based feedback
- ineligible packets → continuous random dropping (?)
- path marking rate is product of local load values
- use relative load of link or node (!)

Load Estimation

Goals

- provide load information to constraint-based routing
- decouple centralized routing from resource allocation system

Capacity-oblivious Load Estimation

- service class capacity adaptation by independent allocation system
- wireless or overlay links with varying capacity
- complex notion of load, e.g. combination of processing and link load
- heterogeneous notion of load at different nodes

Hybrid Load Estimation

- marking-based load estimation: multiplicative error propagation
- usage → capacity estimation: only additions, less sensitive to errors

Packet Marking

- see Binary Packet Marking → system of equations → individual load
  - need continuous marking function
  - usually over-specified → use last N load reports
- real-world engineering challenge: three 'signals' in two bits

System Prototype

Lab Prototype

- **FreeBSD, Linux, Solaris**
- **signalling → KOM RSVP engine (user-level daemon)**
- **packet handling (internal & edge) → FreeBSD/ALTQ (kernel modules)**
- **traffic generation & measurement**

Simulation

- **ns-2**
- **RSVP-based signalling - fully shared code basis with prototype**
- **packet handling - partially shared code basis with prototype**
- **traffic generation & measurement - mostly separate code**
- **load estimation - simulation only**

Packet Marking

- **using ECN bits**
- **RED, VQ, AVQ, LBM**
- **threshold-based marking (TBM) → simplification of VQ**
  - forwarding rate > threshold → mark or drop

Evaluation

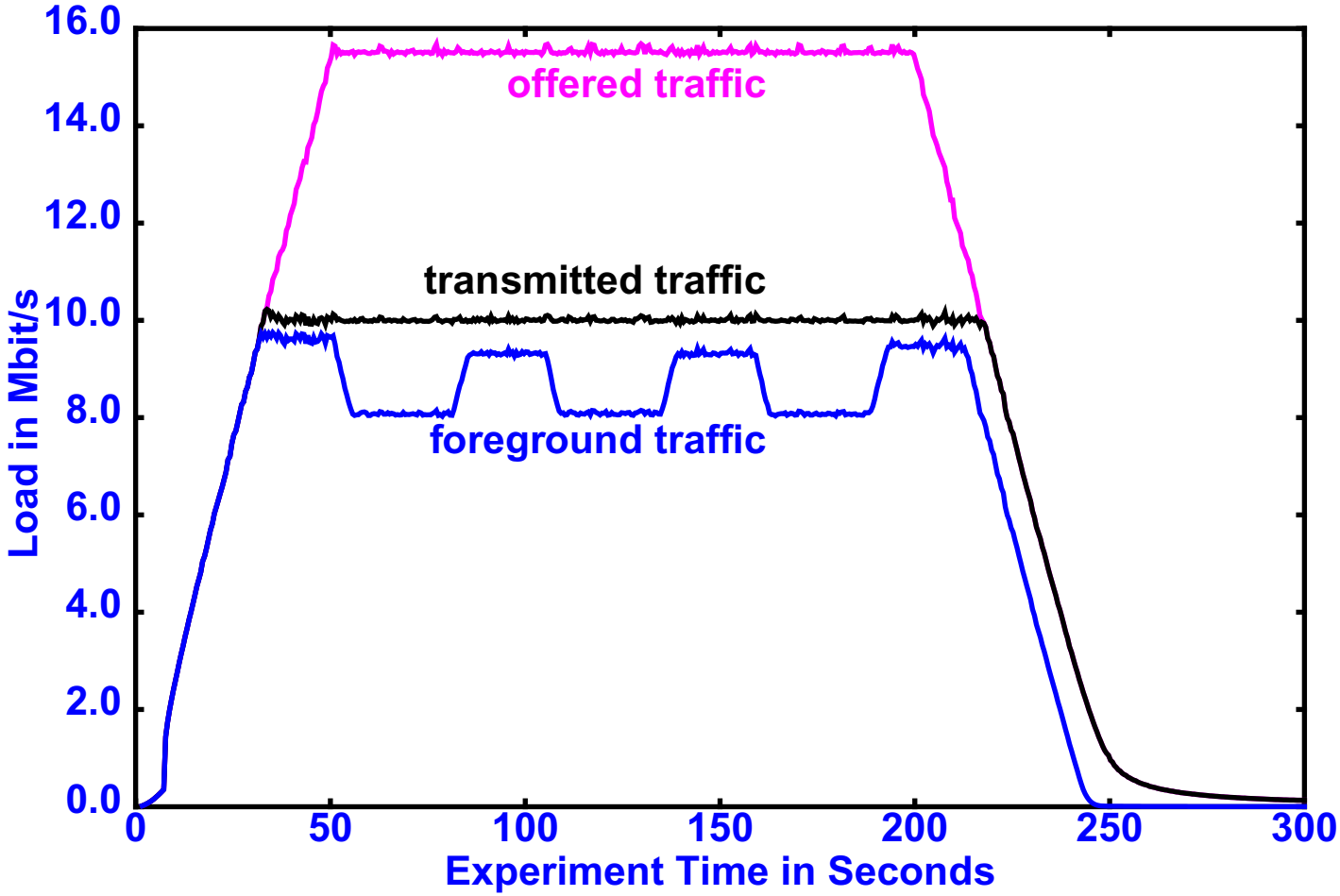
Admission Control

- **lab experiment with real prototype**
  - reconciliation/calibration for larger simulation experiments
- **load system and test whether admission control works**
- **assessment: reaction delay**
- **assessment: traffic discrimination**

Load Estimation

- **simulation experiments**
- **load system and compare measured local load with estimated load**
- **assessment: reporting delay**
- **assessment: estimation precision**

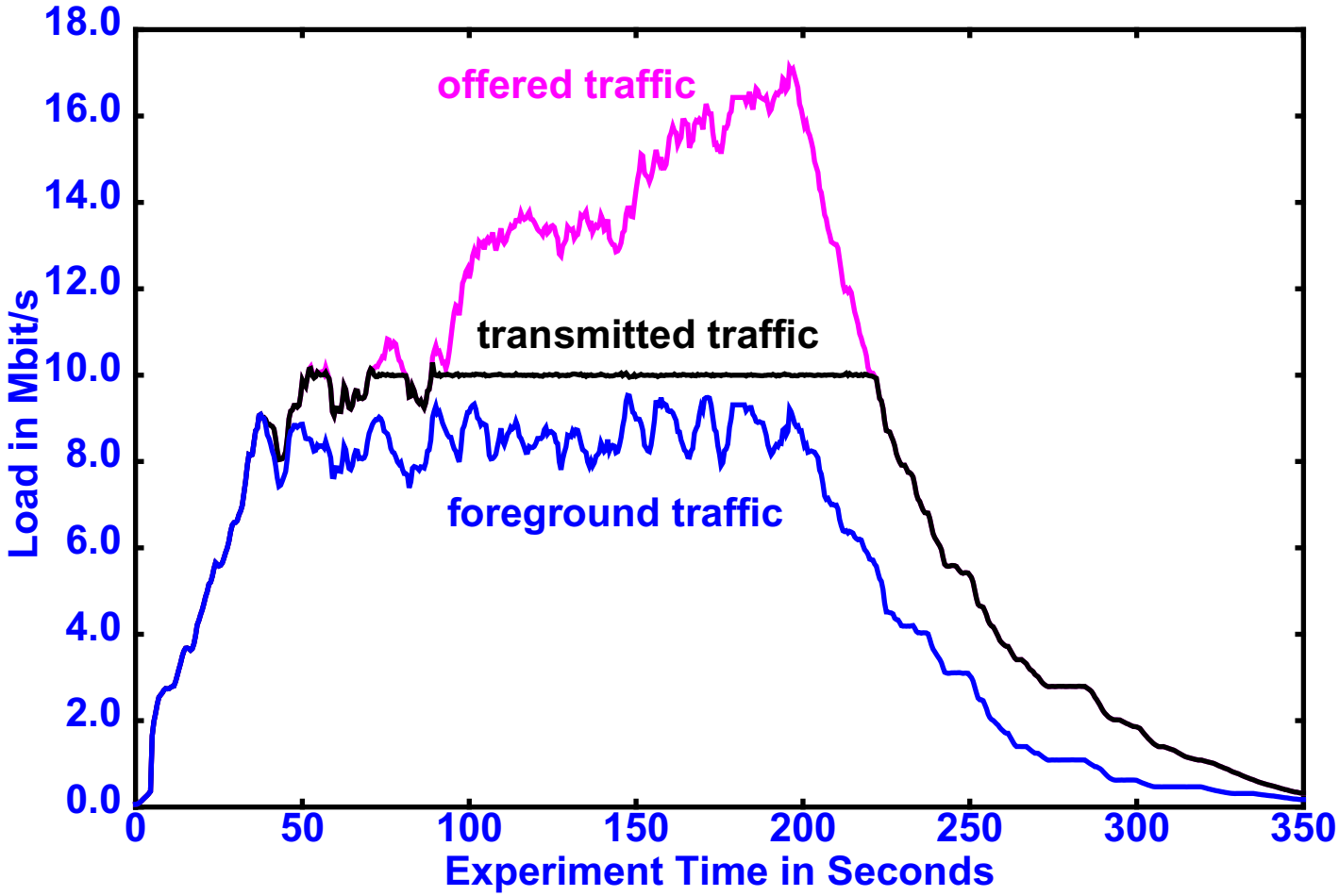
Admission Control - Deterministic Session Arrival



Assessment

- **reaction delay**
- **resource utilization**

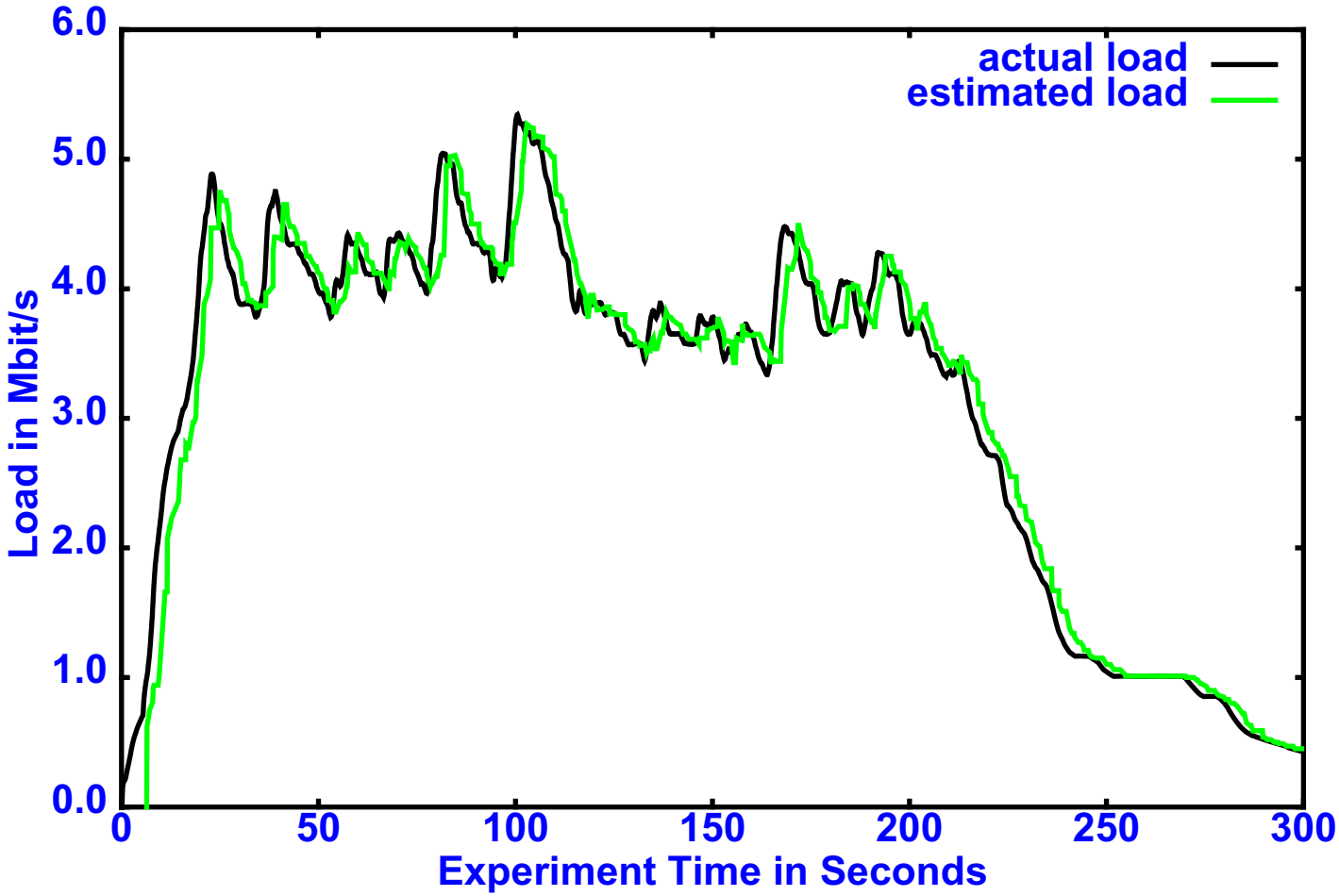
Admission Control - Persistent Random Session Arrival



Assessment

- **reaction delay**
- **resource utilization**

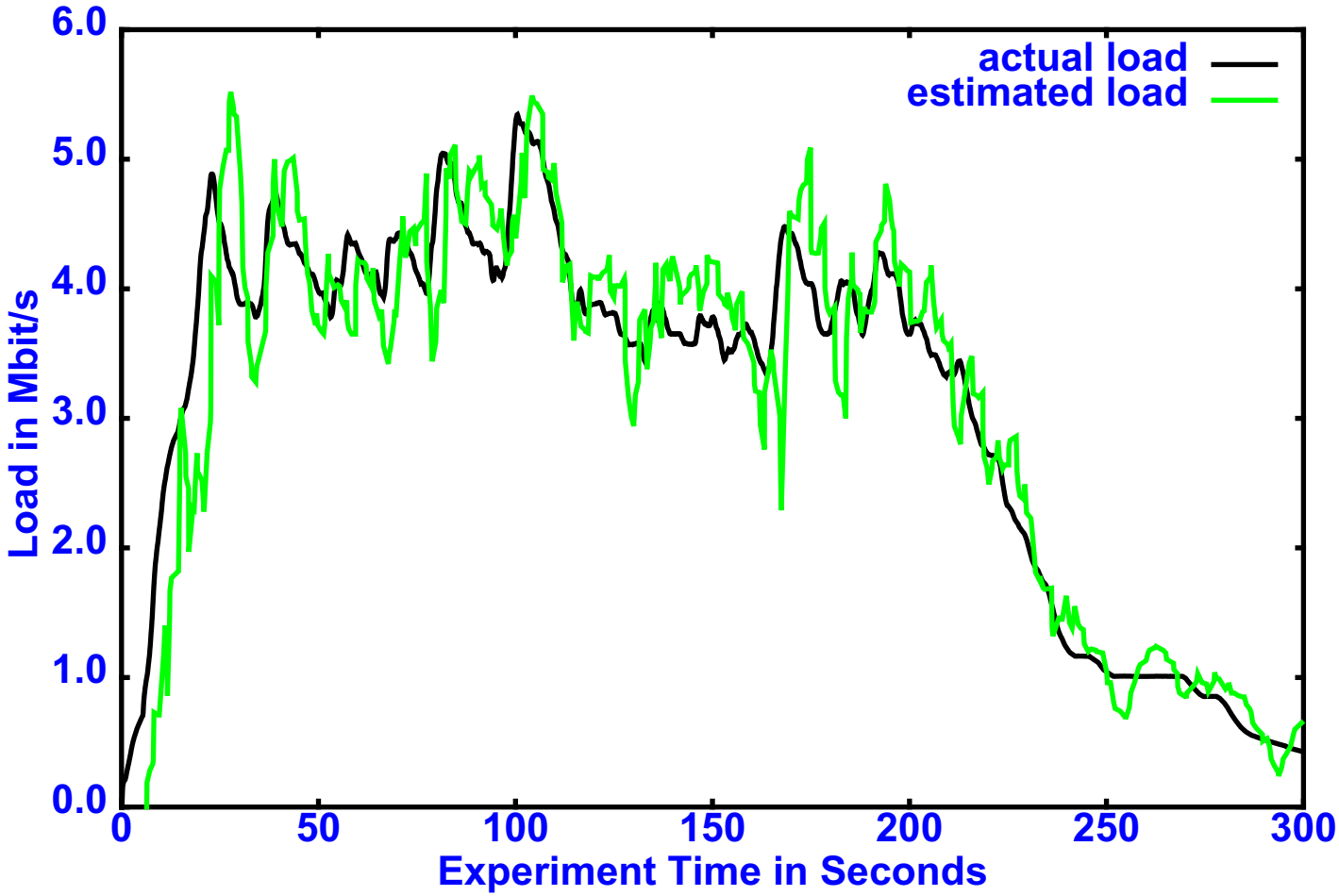
Load Estimation - Usage Based



Assessment

- reporting delay
- estimation precision

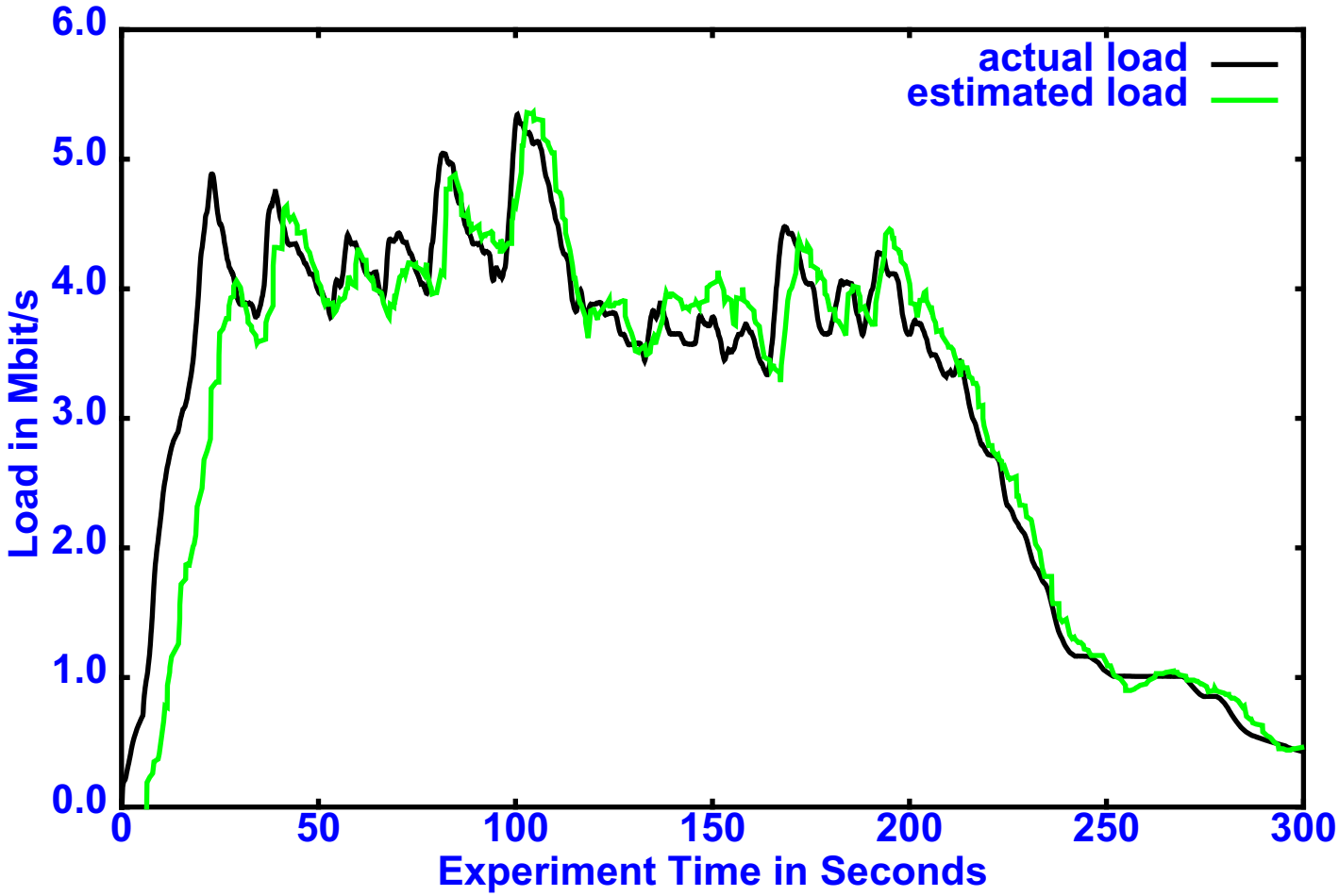
Load Estimation - Load Based



Assessment

- **reporting delay**
- **estimation precision**

Load Estimation - Hybrid



Assessment

- **reporting delay**
- **estimation precision**

Wrap Up

Integrated Load Control System

- **admission control and load estimation**
- **packet marking and evaluation at network edge**
- **system design, architecture, and prototype (simulation and lab)**
- **experimental validation**

Admission Control

- **feedback signalling & admission control → RSVP extensions**
  - fundamentals about admission control & feedback delay
- **various packet marking algorithms**

Load Estimation

- **early, speculative work → basic proof of concept**
- **enabling technology**

Remarks

- **system design is orthogonal to DiffServ**
- **deployment path: RED/ECN & DiffServ for dedicated service class**
  - per-node/per-path deployment possible