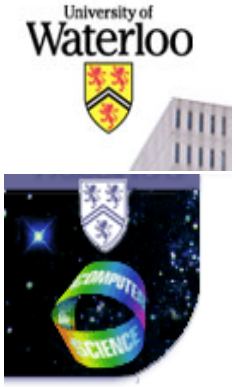
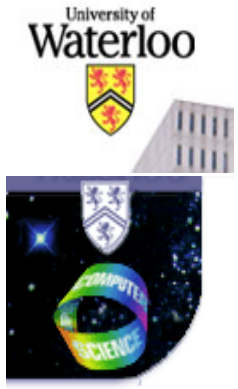


CS 856
Internet Transport Performance

Congestion Control:
TCP, ECN, and Variants

Martin Karsten

School of Computer Science, University of Waterloo
mkarsten@uwaterloo.ca



Contents

TCP Variants

Explicit Congestion Notification (ECN)

Distributed Admission Control and Rate Allocation

TCP Congestion Control - Background

Pre-Congestion Control TCP

- **reliable transport protocol**
- **window-based flow control between sender and receiver**
  - Go-back-N sliding window algorithm

Problems

- **window size negotiated between sender and receiver only**
- **round trip time (RTT) estimation insufficient**

⇒ No Consideration of Changing Network State

- **no large initial bursts after connection setup**
- **Internet congestion collapse**

TCP Congestion Control

- **slow start after connection establishment and time-out**
- **exponential timer back-off upon congestion indication**
- **improved RTT estimation**
- **dynamic window sizing**
  - minimum of receiver window and congestion window

TCP - Deployed Variants

Tahoe

- **basic version with slow-start**
- **three duplicate ACKs or time-out → congestion**
- **congestion → reset congestion window to 1 and enter slow-start**

Reno

- **fast retransmit and recovery**
- **three duplicate ACKs → fast retransmit/recovery**
  - retransmit packets clocked by ACKs without checking congestion window
- **time-out → reset congestion window to 1 and enter slow-start**

NewReno

- **improvement of fast recovery for multiple packet losses**

Sack

- **selective acknowledgements**

TCP Vegas

TCP Observations

- **requires packet loss as congestion signal**
  - packet loss caused by buffer overflow
  - relatively drastic congestion control action
- **oscillation between under-utilization and over-utilization?**

TCP Vegas

- **observe RTT and infer buffer occupation**
- **reduce sending rate when RTT grows large**
- **increase sending rate when RTT is small**

Problems

- **estimation of base RTT, i.e. propagation delay round-trip**
- **congestion on reverse path from destination to source**
- **fairness compared to traditional TCP**

Random Early Detection (RED)

Early Detection of Incipient Congestion

- **measure average queue length**
  - if greater than threshold max_{th} → mark packet
  - if between min_{th} and max_{th} → randomly mark packet
- **packet marking: explicit (ECN) or drop packet**

Goals

- **avoid synchronization between flows**
  - buffer overflow → packets from all connections would dropped
  - probabilistic dropping of packets from some connections
- **control/reduce queue length by packet dropping → better average delay**

Problems

- **parameter setting**
- **infinite variants**

Explicit Congestion Notification (ECN)

Early Packet Dropping Problems

- **drop without necessity**
- **other reasons for packet drop, e.g. corruption**
- **other transport protocols than TCP?**

Idea: Packet Marking instead of Dropping

- **packets carry congestion signal to end systems**

IETF Version: TCP/ECN

- **end systems: negotiate ECN capability at connection setup**
  - sender: mark packets a ECN-capable
  - receiver: reflect congestion indications back to sender
  - sender: react to congestion indications similar as to drop
- **router: use RED or comparable algorithm to infer incipient congestion**
  - mark ECN-capable packet or drop otherwise
- **using bits 6 and 7 in old IP TOS field**
  - 00 → not ECN capable
  - 01 & 10 → ECN capable (two codepoints)
  - 11 → congestion indication

TCP/ECN Challenges

Incremental Deployment

- **packet dropping still possible → TCP must react to drop and mark**
- **non-compliant end systems**
- **non-compliant routers**
- **restrictive firewalls**

Future-Proof Mechanisms

- **isolated congestion signal vs. persistent load signal**
  - different time-scales
- **fairness between different marking and flow control algorithms**
- **suitability of marking algorithm?**
- **interaction with tunnels and non-IP subnet layers (e.g. MPLS)**

Security

- **malicious end systems pretending to use ECN**
  - not a new threat: TCP w/o congestion control, multiple TCP flows, etc.
- **malicious routers subverting ECN**
  - not a new threat: routers can drop, insert or change packets
- **interaction with IPsec → isolation of tunnels required**

TCP/ECN - ECN Nonces

Scenario

- **sender wants to use ECN congestion control, e.g. web server**
- **receiver wants to trick sender into fast downloads**
  - receiver never reflects a congestion indication to sender

Solution: ECN Nonces

- **sender produces random stream of 01 and 10 codepoints**
- **receiver needs to reproduce stream in ACKs to prove no congestion**
- **router congestion indication destroys codepoint information**
  - congestion indication changes ECN bits to 11

Other Marking and Control Algorithms

Active Queue Management (AQM)
- send the "right" signal and use the "best" flow control algorithm

Virtual Queue (VQ)
- run virtual queue at lower speed, mark when virtual buffer is full
- stop marking when virtual buffer empties

Adaptive Virtual Queue (AVQ)
- load > target rate $\rightarrow$ reduce virtual rate
- load < target rate $\rightarrow$ increase virtual rate
- start/stop marking depending on virtual buffer compared to limit

Load-based Marking (LBM)
- marking probability derived from relative load $\rightarrow$ explicit load signal
- problem with multiple resources along transmission path

Numerous Other Suggestions

Packet Marking - Other Application Scenarios

Consider Packet Mark as Charge Indication

- **enforce congestion control cooperation from end users**
- **not only to prevent/handle congestion**
- **differentiate network/transport service → "rate control"**

System Mechanics

- **end systems control amount of traffic**
- **network controls load signal (probabilistic packet marks)**
- **higher network load → more marks → higher price for transmission rate**
- **end systems back off from sending at increasing prices**

Theory

- **distributed resource auction**
- **strictly concave utility curves**
- **mark/price represents end-to-end load situation**
  - i.e. additive price for multiple resources
- **dominance of long-lived flows**

⇒ **Stable Rate Allocation, i.e. QoS**

Perceived QoS

Utility Curve

- user's value of one-dimensional transmission performance
- fairness: user's willingness to pay
- assumption: certain typical shapes

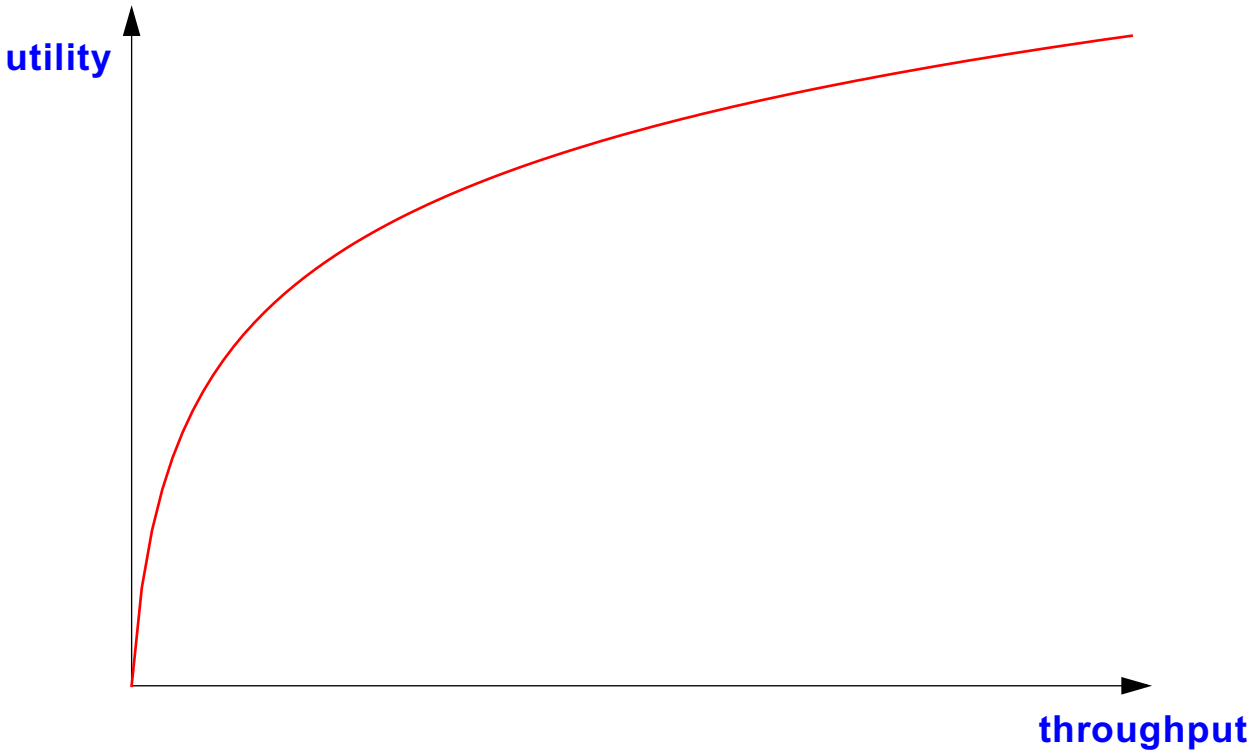
Simple Application Classification

- elastic
- inelastic

Utility in Reality

- multi-dimensional utility curves
- variation of utility curve during transaction

Elastic Applications

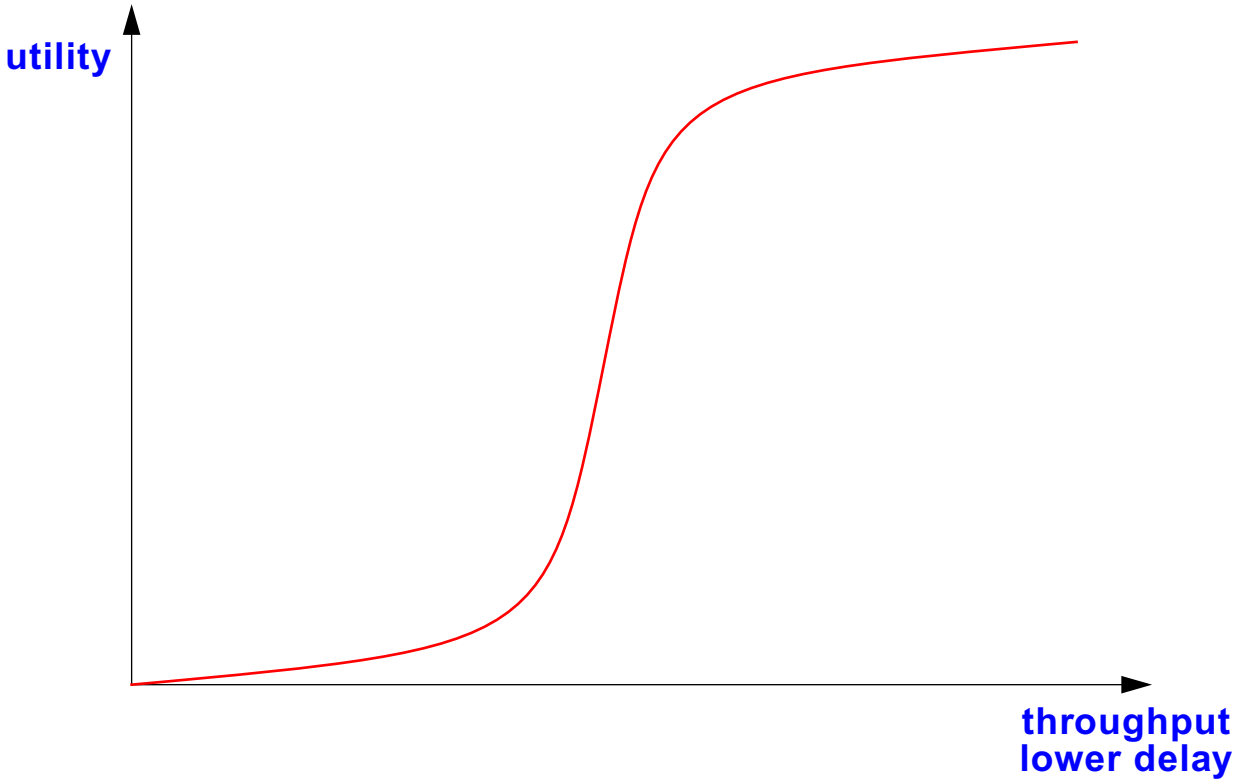


Examples

- **file transfer**
- **web**

Often modelled as strictly concave, but in reality there's an upper limit

Inelastic Applications



... if the respective other QoS constraints are met, as well

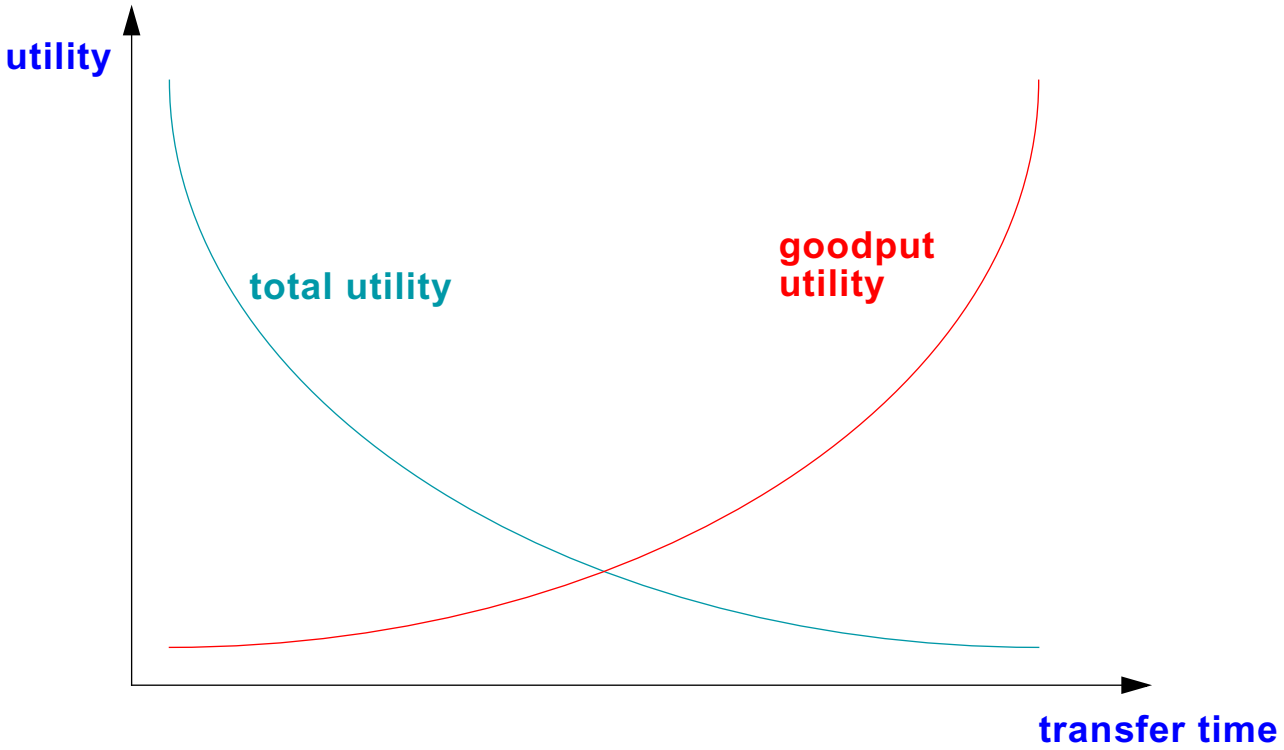
Examples

- **multimedia transmission**
- **interactive games**

Utility Curves - Example: `ftp`

Average Throughput → Concave Utility Curve

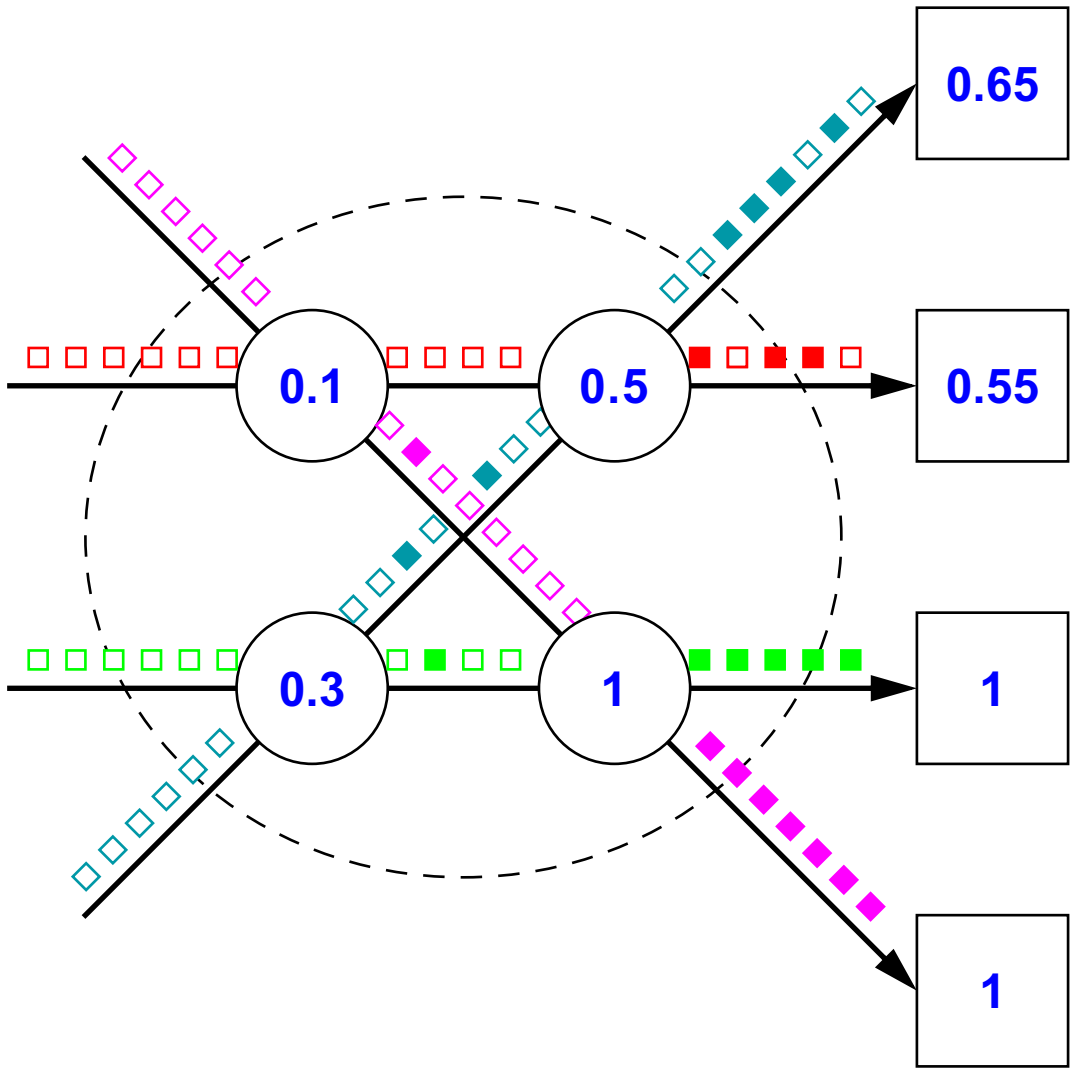
Utility and Transfer Time



Still Not Complete Picture ⇒ Complex Problem

Binary Packet Marking

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



Service Differentiation

WTP Modification to TCP
- approximate rate-based flow control by window-based congestion control
- additional parameters control speed of increase and decrease

Marking Algorithms
- RED, VQ, LBM

Service Differentiation
- good with RED & LBM, ok with VQ
- depending on TCP/WTP details

Marking Probability
- steep increase with RED, depending on RED configuration

Average Queueing Delay
- RED & LBM better than VQ

Packet Marking and Rate Control - Challenges

Modelling

- **applicability of idealized mathematical models**
- **short-lived flows → background noise?**
- **system stability in the presence of feedback delays?**

Technical

- **multiple resources vs. single-bit marking**
  - additive load signal? only approximated, if marking rate is very low
- **per packet accounting is expensive**

User Preferences

- **price may not be known ahead of time**
- **price may fluctuate**
- **incentives for providers to upgrade networks?**
  - only indirect marketing incentives

Security

- **responsibility for payment vs. responsibility for sending rate?**

Flow Control vs. Admission Control

Flow Control

- throttle sending rate according to load signal (packet marks)
- packet marking at internal nodes
- feedback from receiver
- suitable for elastic flows

Admission Control

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

Packet Marking and Admission Control

Use Packet Marking for Admission Control of Inelastic Traffic

- end systems or gateways probe network to infer load state
- or infer load from existing traffic
- no calls are accepted, if load is below threshold

Gateways: Solve Technical and Security Challenges

- controlling traffic and reacting to packet marks
- easier to account for than individual end systems
- more reliable rate control reaction
- operated by network provider
- offering traditional signalling-based service interface
- fixed price per mark vs. very high willingness-to-pay

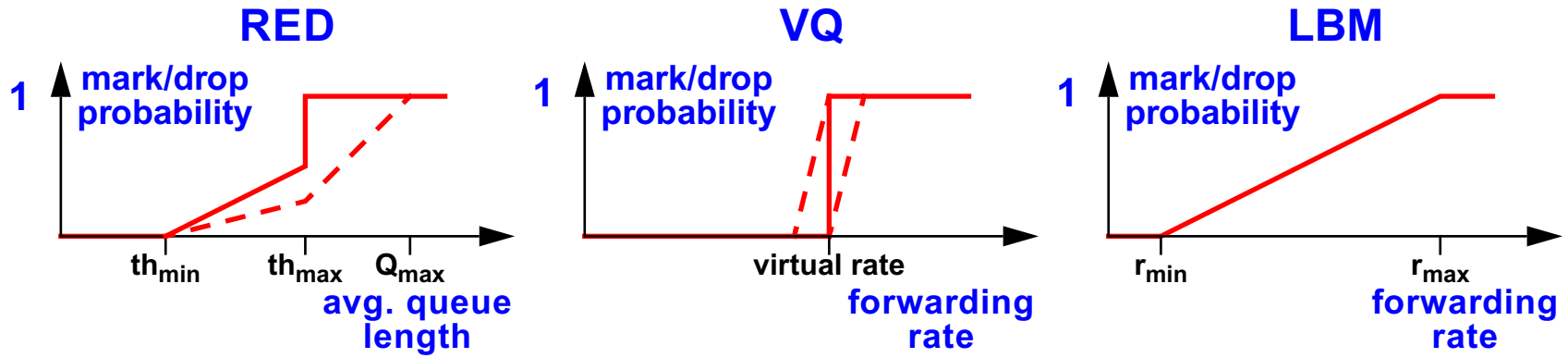
Different Level of Insight

- mathematical models and proofs for flow control
- reasoning and simulation for admission control

Integration of Elastic and Inelastic Flows?

- at least offer unused capacity to background traffic

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

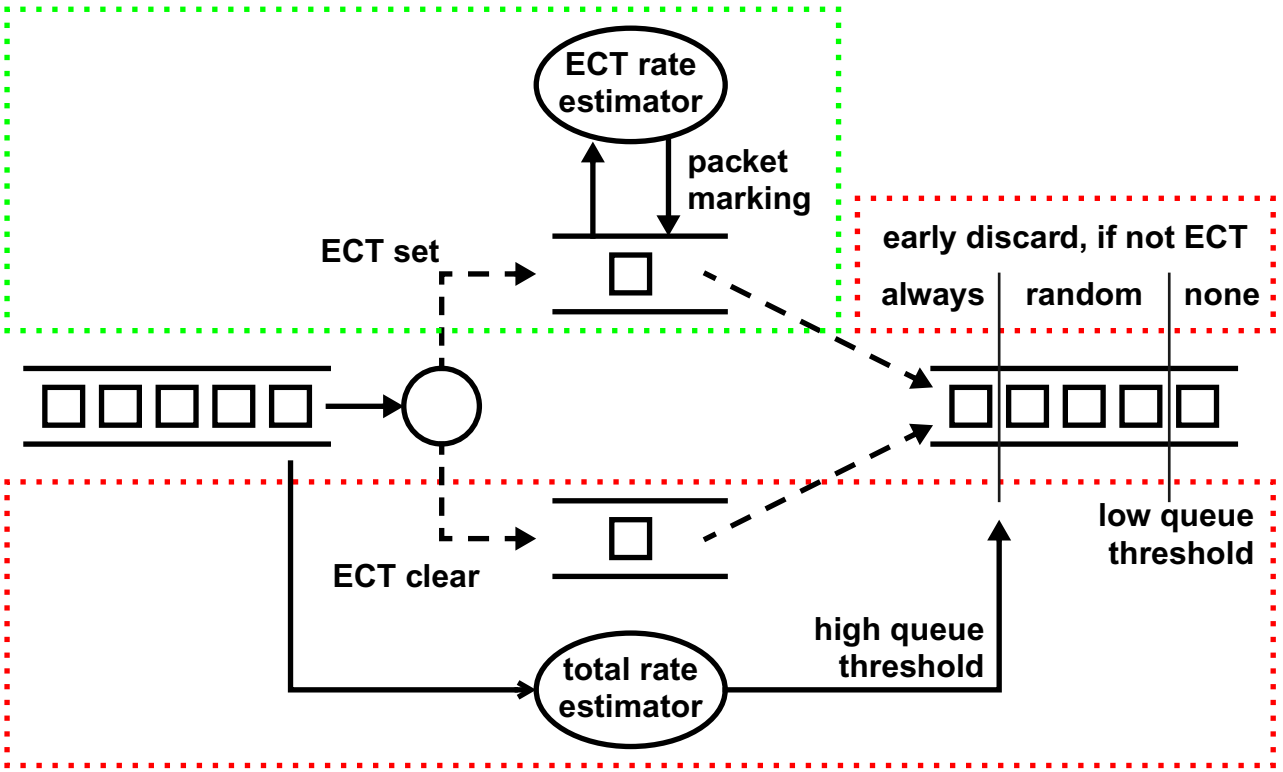
- hybrid feedback, time-scale dependent
- ineligible packets → bursty dropping (?)
- inelastic flows → binary path marking rate

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 (!)

Marking vs. Dropping

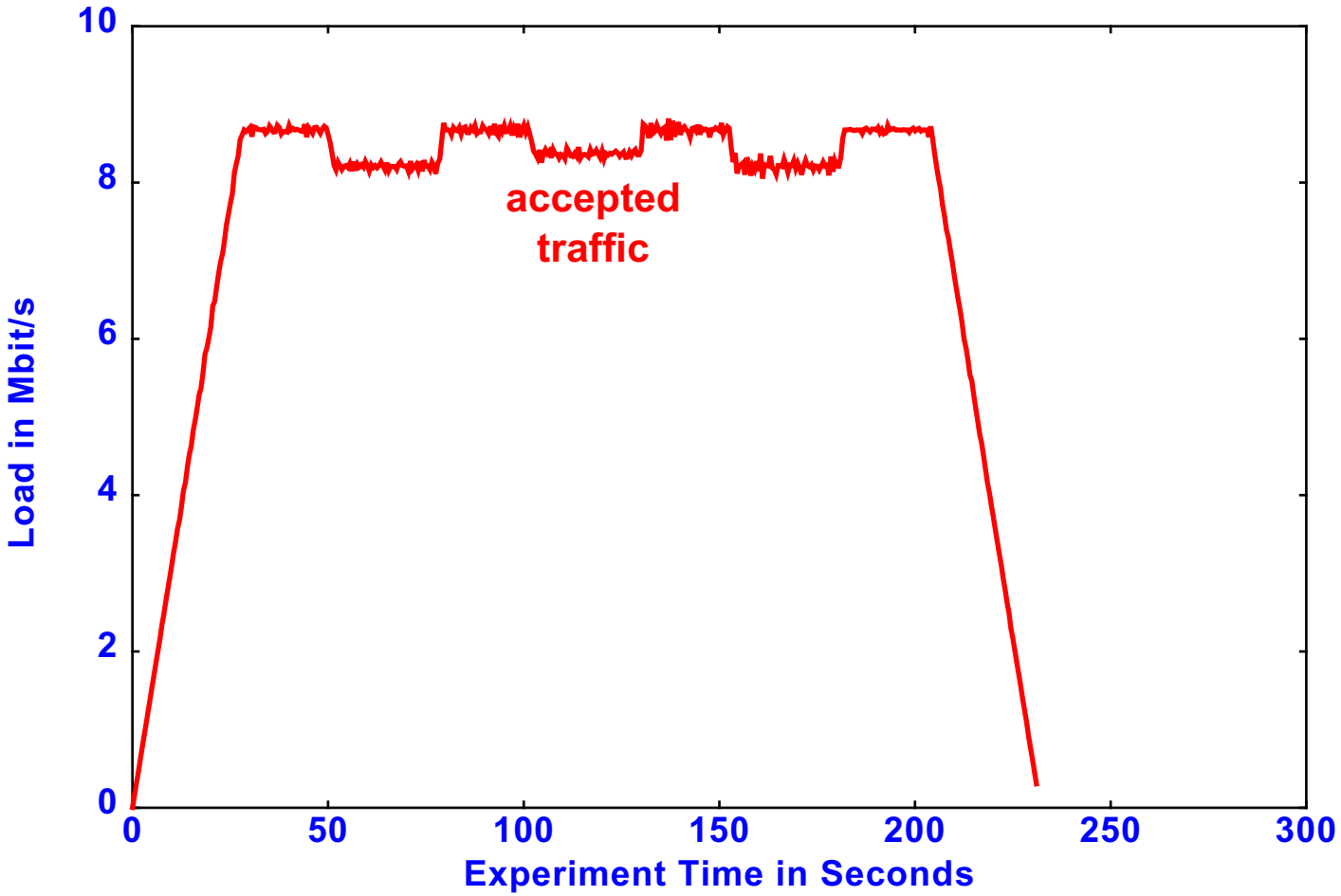
Dropping of Ineligible Packets?
Influence of Ineligible Packets on Local Load?
⇒ Differentiated Queue Management (DQM) at Internal Nodes



packet marking

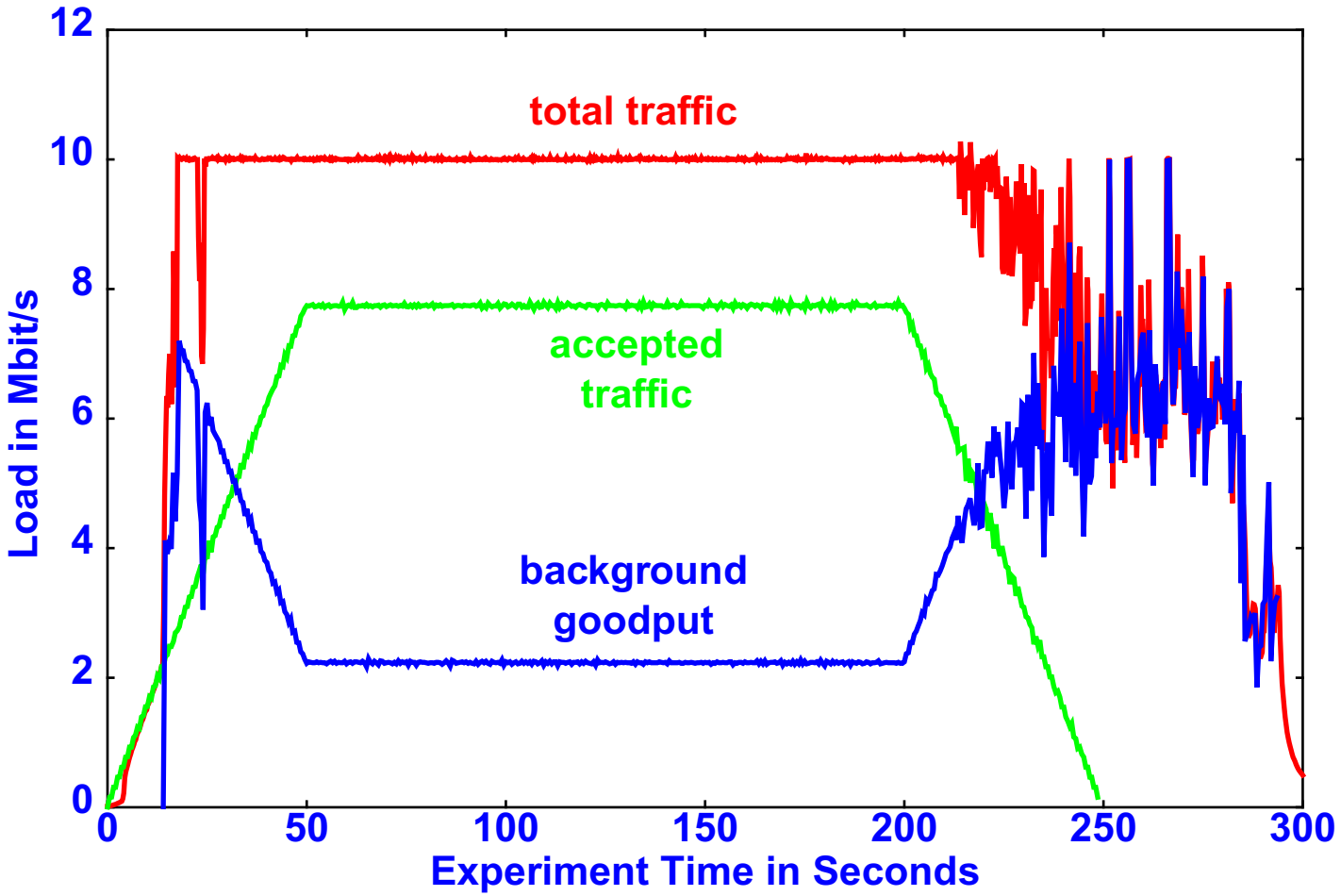
adaptive early random drop

Experimental Results - Admission Control



- admission control of VoIP flows
- VQ marking & simple admission control threshold

Experimental Results - Traffic Discrimination



- mix of VoIP flows and unresponsive UDP background traffic
- VQ marking, simple admission control threshold, DQM discrimination

Dynamic Pricing

Packet Marking, Flow Control, and Dynamic Pricing

- **economic market models**
- **market efficiency: stability and high utilization**

Packet Marking and Admission Control Gateways

- **traditional business model possible: price per time and resources**
- **dynamic pricing: calculate dynamic price from load information**
  - price formula $\frac{a}{1-x} + b$ possible for relative marking rate x
- **insurance model**
  - gateways operated by different institution than network provider
  - gateways accept the economic risk of rising load for a fixed surcharge
  - over-subscription $\rightarrow$ bad luck for everyone

Deployment

- **use gateways now, end system control later?**
- **chain of trust to solve security problems?**

Admission Control - Challenges

Service Integration

- **support elastic and inelastic flows**
- **different time-scales**
- **flow control at edge vs. flow control at end system**

Dynamic System Adaptation

- **reactive system → safety margins needed**
  - automatic parameter adaptation?
- **controlled overbooking → currently "use it or loose it"**
  - only actual traffic generates load signals

Delay Guarantees?

- **precise: delay differentiation**

Discussion

End-System Rate Control and Packet Marking Charges

- packet marking algorithms

Packet Marking and Gateway Admission Control

- packet marking algorithms
- "use it or loose it" problem
- safety margins