Carnegie Mellon

LIRA: Service Differentiation for Traffic Aggregates With Large Spatial Granularities

Ion Stoica Hui Zhang

School of Computer Science
Carnegie Mellon University

Carnegie Mellon

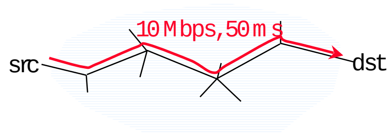
Outline

- Introduction
- LIRA - Location Independent Resource Accounting
- Simulation experiments
- Conclusions and future work

Carnegie Mellon

Motivation

- Traditional QoS models
  - per-flow end-to-end



- Appropriate for
  - long duration and steady traffic
  - video/audio traffic, virtual lease line
- Not appropriate for
  - short duration and bursty traffic, e.g., web traffic
  - aggregate traffic over multiple destinations

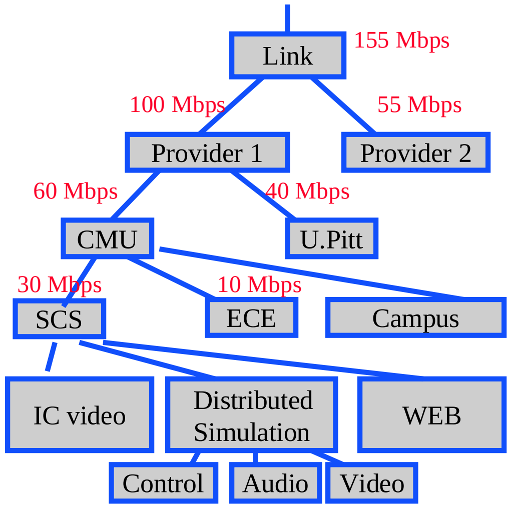
Carnegie Mellon

Service Differentiation for Traffic Aggregates

- All traffic from Hui’s workstation vs. Ion’s workstation
- All traffic from CMU vs. University of Pittsburgh
- Service can be defined for both a local link and a network

Carnegie Mellon

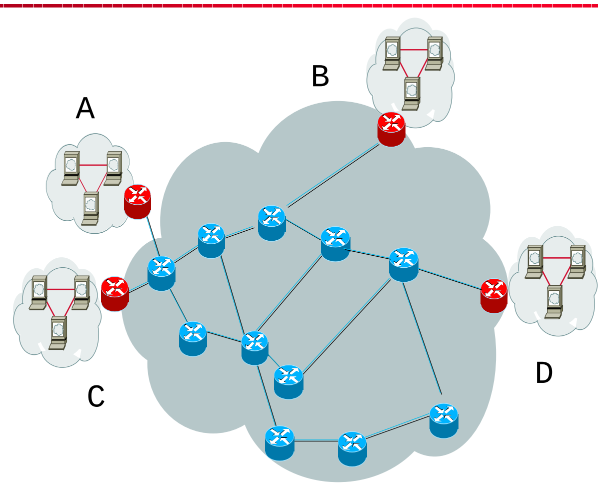
Hierarchical Link Sharing



- QoS for traffic classes with different granularities
- Defined over a single physical link
- Models and algorithms
  - Class-Based Queueing (CBQ)
  - Hierarchical Packet Fair Queueing (H-PFQ)
  - Hierarchical Fair Service Curve (H-FSC)

Carnegie Mellon

QoS For Traffic Aggregates Over a Network



- User A pays $1000
- User B pays $500
- User C pays $1000
- User C pays $100
- What should be the right service?

Carnegie Mellon

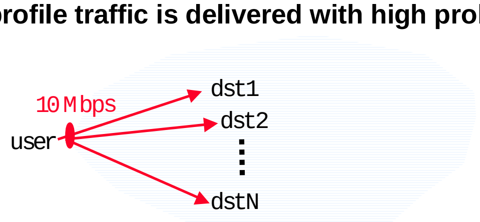
Question

- What is an appropriate QoS or service differentiation model for aggregate traffic with large spatial granularity?
  - Traffic aggregate with a large set of destinations

Carnegie Mellon

Example: Assured Service

- Proposed by Clark & Wroclawski
- Each user is associated a traffic profile independent of destination
  - usually defined in terms of absolute bandwidth
- Traffic is of two types:
  - marked (in-profile)
  - unmarked (out-of profile)
- In-profile traffic is delivered with high probability



Carnegie Mellon

Potential Problem

- Worst-case provisioning needs to assume all marked packets traverse slowest link
- No obvious optimistic provisioning algorithm
- Cannot achieve high assurance and high utilization simultaneously

Carnegie Mellon

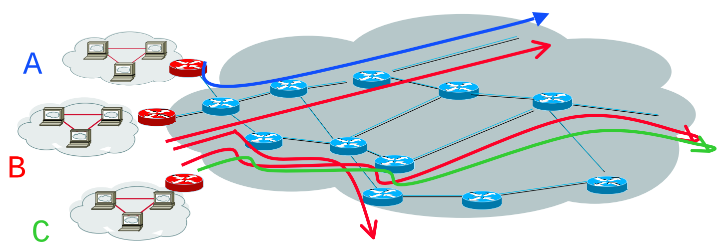
Conflict

- Profile definition: large spatial and temporal granularity
  - space: defined over a large set of destinations
  - time: defined over periods much larger than flow duration
- Achieving high service assurance requires congestion avoidance for any link, at any time
  - dynamic and local phenomenon

Carnegie Mellon

Another Example

- User Share Differentiation (USD) by Zheng Wang
  - each user is allocated a share
  - at each congested link, bandwidth is allocated proportionally to each user according to its share
- Undesirable property



1997 Hui Zhang

11

Carnegie Mellon

Our Proposal: LIRA

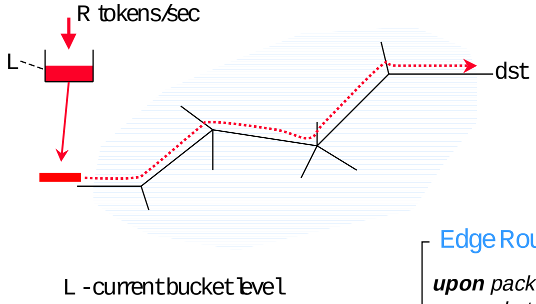
- Define service in relative rather than absolute terms
  - service defined in resource tokens instead of fixed amounts of bandwidth
- Associate to each marked packet a cost as a function of
  - congestion level of path it traverses
  - packet length
- Mark a packet only if user covers its cost
- Key properties:
  - users receive service in proportion to their assigned token rates
  - high assurance - by appropriately choosing the cost function
  - high utilization - by using dynamic feedback and load balancing

1997 Hui Zhang

12

Carnegie Mellon

Packet Marking Algorithm



Edge Router/source Alg

upon *packet* arrival:
packet_cost = f(path, packet, …);
if (preferred(packet) && L > packet_cost*)*
L = L - packet_cost;
mark*(packet);*

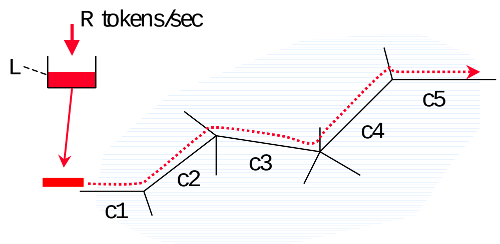
1997 Hui Zhang

13

Carnegie Mellon

Packet Cost

- **Packet cost - product between packet length and path cost**
- **Path cost - sum of costs of all links on the path**
- **Link cost - cost to forward a marked bit on that link**



packet_cost = packet_length*(c1 + c2 + c3 + c4 + c5)

1997 Hui Zhang

14

Carnegie Mellon

Link Cost

- **Objective**
  - **no** marked packet is ever dropped
- **Implication**
  - when link utilization approaches unity the cost should exceed the total number of tokens in the system
    - in general, if number of tokens is unbounded, link cost should approach infinity
- **Our choice**

$$c(t) = \frac{a}{1 - u(t)}$$

  - ***a*** - fixed cost
  - ***u(t)*** - link utilization at time *t*

link cost - reflects link congestion

Carnegie Mellon

Link Cost Computation

- **In a real system information is obsolete. This leads to**
  - system oscillations
  - inaccuracies in cost computation
- **Solution**
  - make cost function more robust - use an iterative formula

$$c(t_i) = a + c(t_{i-1}) \times u(t_{i-1}, t_i)$$

  - account for "unexpected" variations by using only 85-90 % of link's capacity for marked traffic

Carnegie Mellon

Load Balancing

- Maintain the *k-th* shortest paths
- Select among alternate paths based on their cost
- Potential concerns
  - oscillations
  - packets reordering within the same flow

1997 Hui Zhang

17

Carnegie Mellon

Avoid Oscillation and Reordering

- Solution
  - bind probabilistically a flow to a path; probability depends on path's cost
- Binding technique
  - each path is encoded by a label - XOR over (IP) addresses of all hops to destination
  - label is stored in forwarding table entry and packet header
  - forwarding based on match of labels in packet header and forwarding table
  - router updates the label in packet's header by XOR-ing it with its address

1997 Hui Zhang

18

Carnegie Mellon

Example

ROUTING TABLE

DST	COST	LABEL
$adr5$	7	$adr2 \otimes adr3 \otimes adr5$
	8	$adr2 \otimes adr4 \otimes adr5$
⋮	⋮	⋮

FORWARDING TABLE

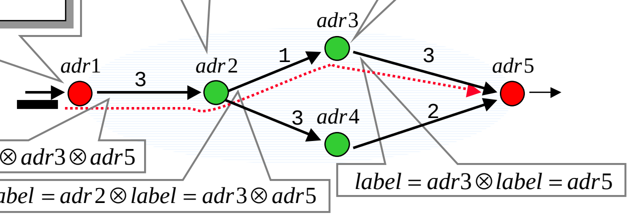
LABEL	DST	NEXT HOP
$adr2 \otimes adr3 \otimes adr5$	$adr5$	$adr2$
$adr2 \otimes adr4 \otimes adr5$	$adr5$	$adr2$
⋮	⋮	

FORWARDING TABLE

LABEL	DST	NEXT HOP
$adr3 \otimes adr5$	$adr5$	$adr3$
$adr4 \otimes adr5$	$adr5$	$adr4$
⋮	⋮	

FORWARDING TABLE

LABEL	DST	NEXT HOP
$adr5$	$adr5$	$adr5$
⋮	⋮	



Carnegie Mellon

Implementation Issues

- **Path cost computation and distribution**
  - **link cost computation : *O(1)* space/time complexity**
  - **leverage existing routing protocols to compute/distribute path cost**
    - **link-state protocols - make cost part of link state**
    - **distance-vector protocols - embed link cost in routing messages**
- **Packet marking and forwarding**
  - **per-user state at edge; no state inside network**
  - **O(1) time complexity**
- **Load balancing**
  - **extend routing protocols to compute the *k*-th shortest paths**
  - **limit *k* to avoid routing/forwarding tables explosion**
  - **integrate *CIDR***

Carnegie Mellon

Simulation Experiments

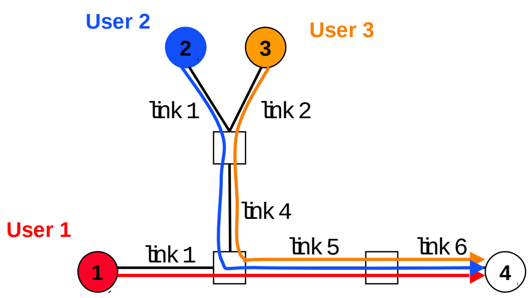
- Packet level simulator implementing both DV and SPF protocols
  - extended to compute the *k*-th shortest paths
- Traffic generation
  - self-similar -many ON-OFF flows with ON and OFF periods drawn from a Pareto distribution [Willinger et al]
  - shape parameter *a = 1.2*
- Largest simulation
  - 30 nodes
  - 15000 flows
  - over 8 millions packets

Carnegie Mellon

Experiments Setting

- Schemes
  - BASE - single path routing
    - model today's best effort Internet
    - use DV or SPF algorithms
  - STATIC - single path routing + LIRA
  - DYNAMIC-*k* - *k* shortest path routing + LIRA
- Metrics
  - user overall throughput - aggregate rate of user's traffic delivered to all destinations
  - user in-profile throughput - aggregate rate of user's in-profile traffic delivered to all destinations
- Simulation parameters
  - simulation time - 200 sec
  - routing update interval - 5 sec
  - link capacity: 10 Mbps, buffer size - 256 KB; threshold - 64 KB

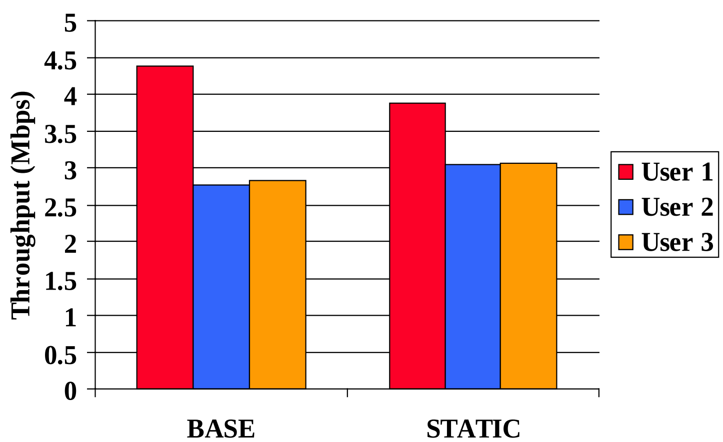
Exp. 1 - Local Fairness and Service Differentiation



1997 Hui Zhang

23

Exp. 1: Overall Throughput

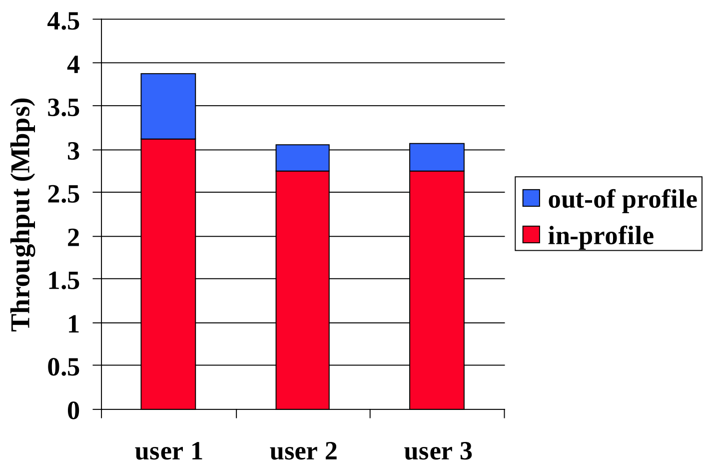


1997 Hui Zhang

24

Carnegie Mellon

Exp.1: STATIC



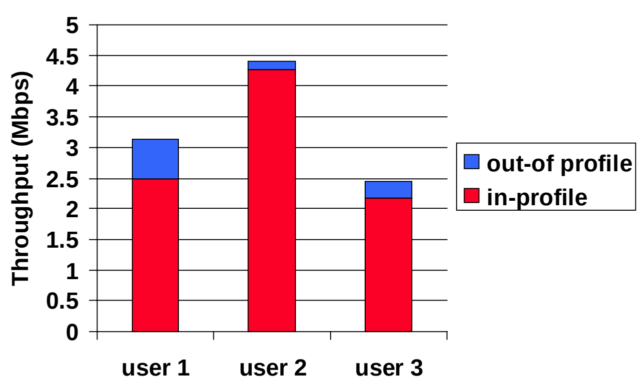
1997 Hui Zhang

25

Carnegie Mellon

Exp. 1: Service Differentiation

- User 2 receives tokens at twice the rate of other two

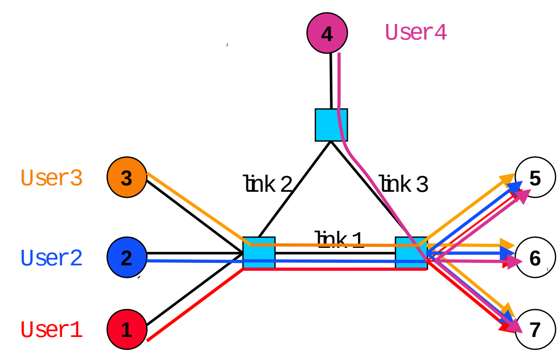


1997 Hui Zhang

26

Carnegie Mellon

Exp. 2 -Global Fairness and Load Balancing

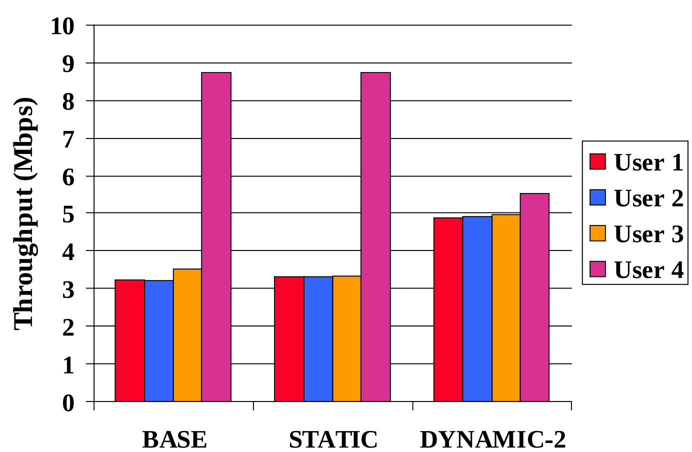


1997 Hui Zhang

27

Carnegie Mellon

Exp. 2: Overall Throughput

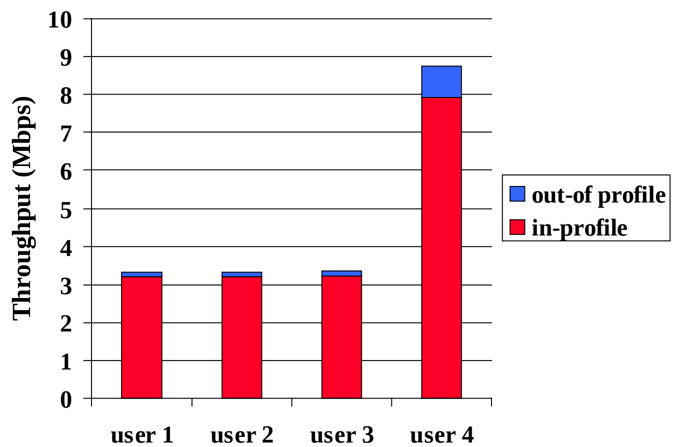


1997 Hui Zhang

28

Carnegie Mellon

Exp. 2: STATIC

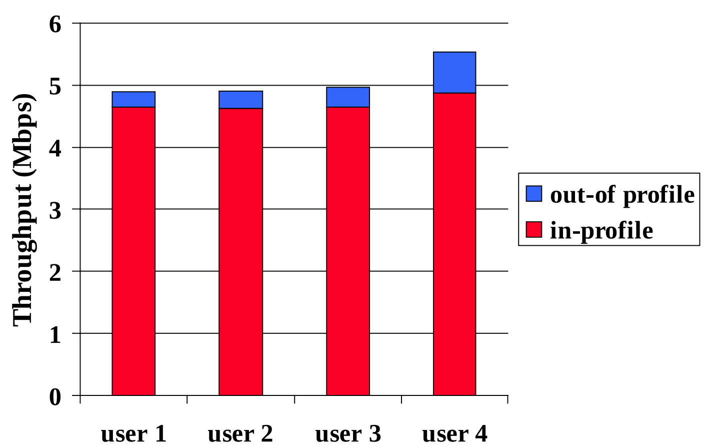


1997 Hui Zhang

29

Carnegie Mellon

Exp. 2: DYNAMIC-2



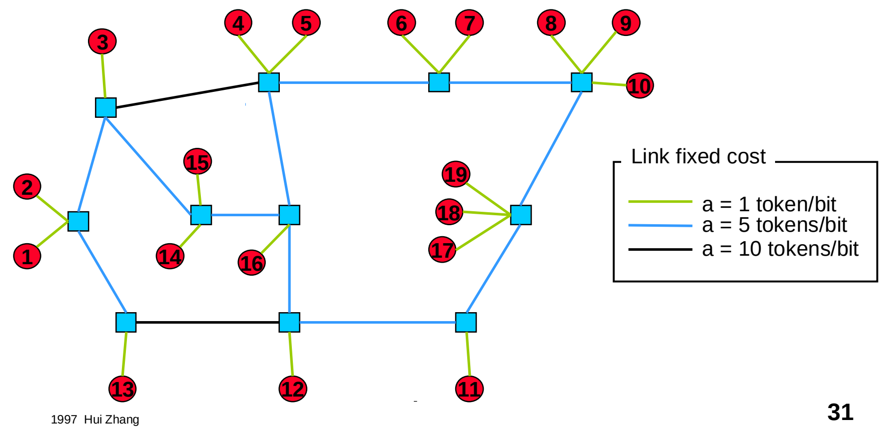
1997 Hui Zhang

30

Carnegie Mellon

Exp3: Complex Topology

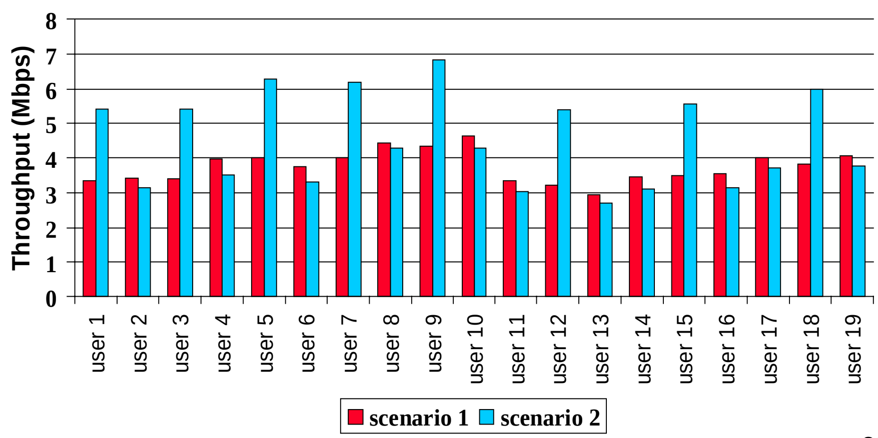
- Similar to T3 NSFNET backbone
- Link capacity: 10 Mbps; User's sending rate: ~13 Mbps
- Use DYNAMIC-3 scheme



Carnegie Mellon

Balanced Load: In-profile Throughput

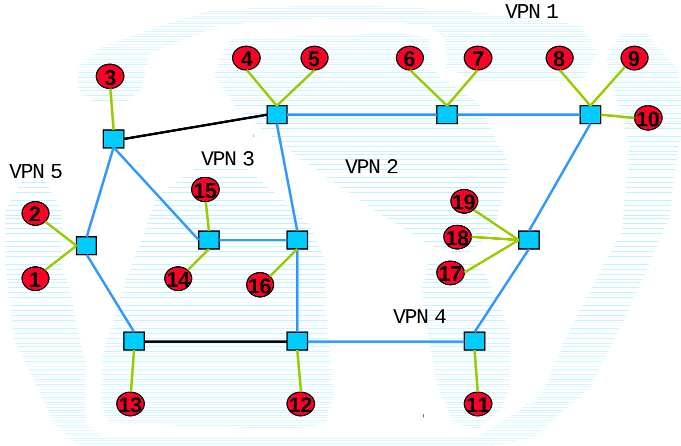
- Scenario 1: token rate of each user: $0.5*10^8$ tokens/sec
- Scenario 2: token rate of users 1,3,5,7,9,12,15,18 changed to: 10^8 tokens/sec



Carnegie Mellon

VPN Experiment

- Token rate allocated per VPN
- Each VPN divides its rate equally among its nodes



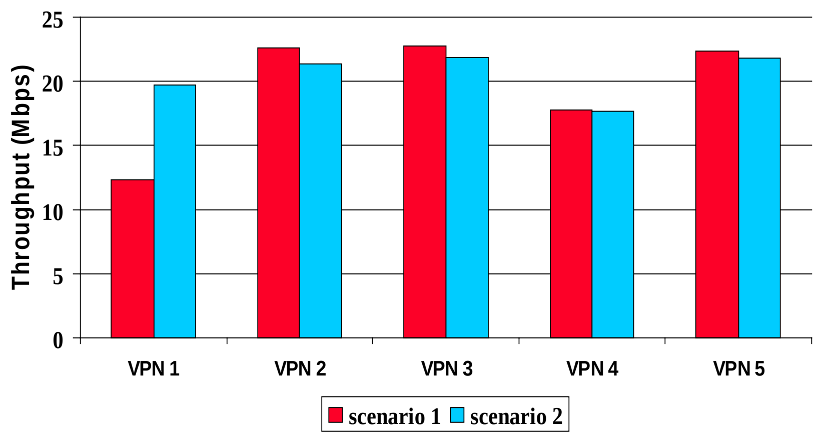
1997 Hui Zhang

33

Carnegie Mellon

VPN Experiment: VPN In-profile Throughput

- Scenario 1: token rate of each VPN: $2.4*10^8$ tokens/sec
- Scenario 2: token rate of VPN 1 changed to: $4.8*10^8$ tokens/sec



1997 Hui Zhang

34

Carnegie Mellon

Other Results

- Service assurance
  - no marked packets dropped in small simulations
  - around 0.1% market packets dropped in large simulations
  - 60% of unmarked packets are dropped

Carnegie Mellon

Related Work

- Assured service [Clark & Wroclawski]
- User-Shared Differentiation (USD) [Wang]
  - little correlation between user's share and its throughput
- Resource allocation, e.g., [Waldspurger & Weihl], [Ferguson et al],
  - do not consider problem of allocating resources for traffic aggregates
- Smart markets [MacKie-Masson & Varian]
  - relation between packet's priority and user throughputs not clear
  - difficult to achieve high service assurance
- Routing and load balancing
  - LIRA is first work to combine routing, load-balancing and congestion control
  - when all links have the same capacity, path cost is within a constant factor of *shortest-dist(P,1)* cost [Ma & Steenkiste]

Carnegie Mellon

Summary

- QoS model for aggregate traffic with large spatial granularity
  - LIRA: Location Independent Resource Accounting
  - a service not supported by current Intserv framework
- Global fairness reference model
  - each packet carries resource tokens
  - each router implements WFQ
  - packet pays token to every router traversed, weight proportion to amount of tokens paid
- Mechanisms that implements the model
  - leveraging on existing routing infrastructure to propagates the cost along a path
  - compatible in spirit with existing proposals (token-bucket based profile at edge, packet marking, semantics of marked packets)

Carnegie Mellon

Summary

- Properties of LIRA
  - high service assurance
  - high resource utilization
- Key ideas: separation of two levels of differentiation
  - user level differentiation: destination/path independent
  - packet level differentiation : destination/path dependent
  - service profile defined in relative terms instead of absolute bandwidth bridges the gap

Carnegie Mellon

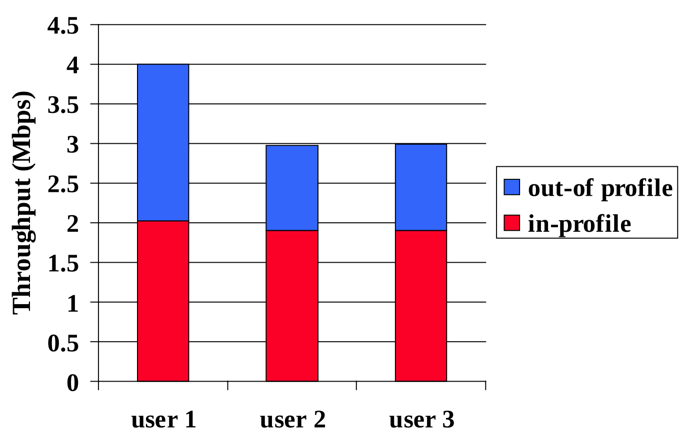
Future Work

- Extend LIRA to support
  - receiver based payment, multicast
- Utilize path cost at
  - egress nodes for active queue management
  - end systems for better end-to-end congestion management

Carnegie Mellon

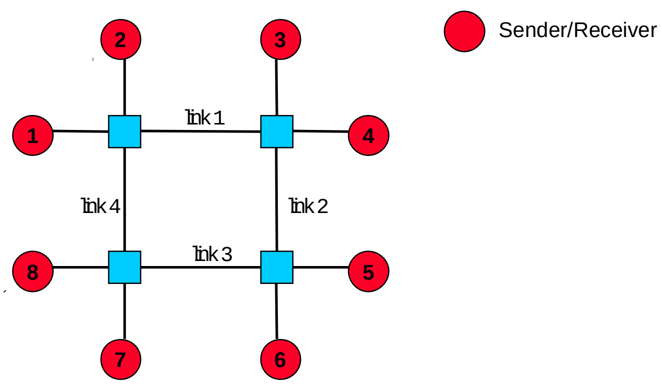
Exp. 1: STATIC

- costs of links 5 and 6 are are increased five times



Carnegie Mellon

Exp. 3 - Load Distribution and Load Balancing

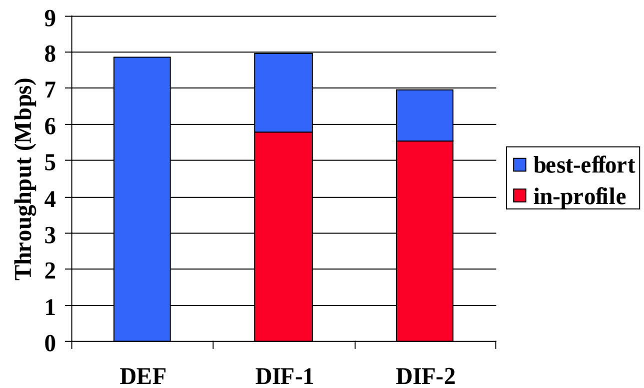


1997 Hui Zhang

Carnegie Mellon

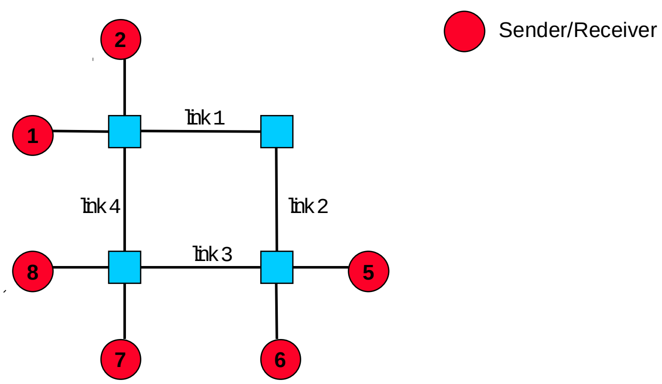
Exp. 3 (cnt'd)

- Balanced Load - average user total and in-profile throughputs



1997 Hui Zhang

Exp. 3 - Load Distribution and Load Balancing

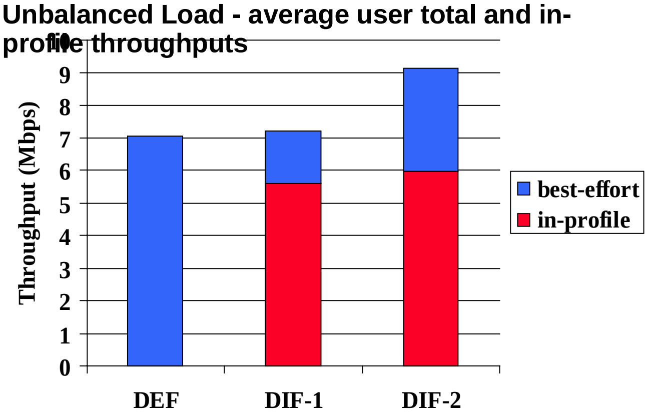


1997 Hui Zhang

43

Exp. 3 (cnt'd)

- Unbalanced Load - average user total and in-profile throughputs



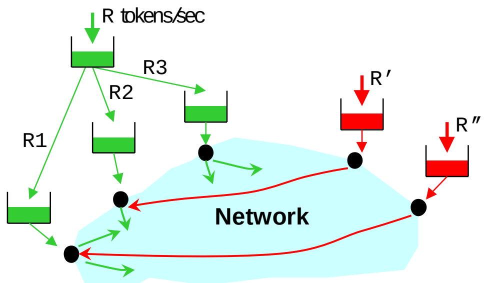
1997 Hui Zhang

44

Carnegie Mellon

Distributed Model

- Sender payment scheme
  - each user distributes its shares to access points
- Receiver payment scheme
  - ISP credits each marked packet received by a user up to a negotiated profile
    - – packet should carry its cost (due to route asymmetry it can't be determined by the receiver)



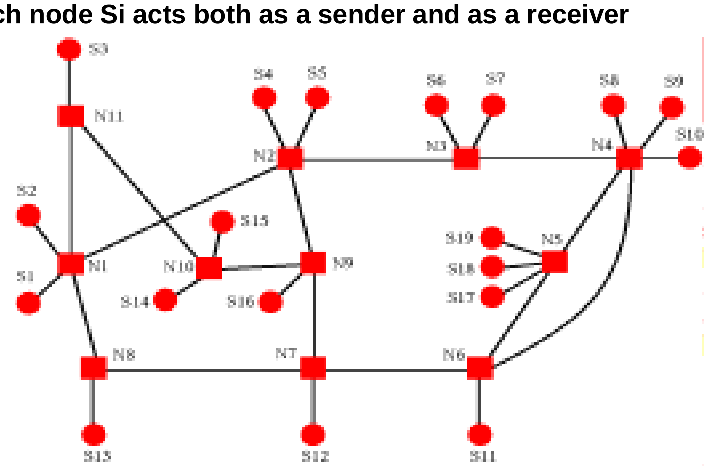
Edge Router Alg.

```
upon packet arrival:
  if (marked(packet))
    if (credited(packet))
      L -= length(packet)*cost_per_bit;
    else
      if (L < length(packet)*cost_per_bit)
        unmark(packet);
      else
        L -= length(packet)*cost_per_bit;
```


Carnegie Mellon

Exp. 3: Large Scale Example

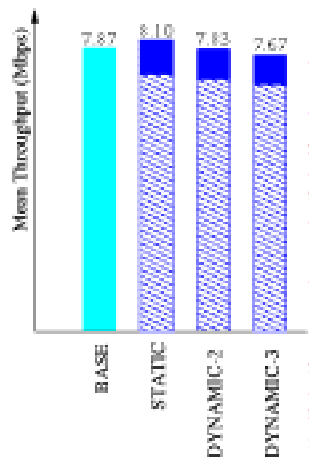
- Topology similar to NSFNET backbone
- Each node Si acts both as a sender and as a receiver



Carnegie Mellon

Exp. 3: Balanced Load

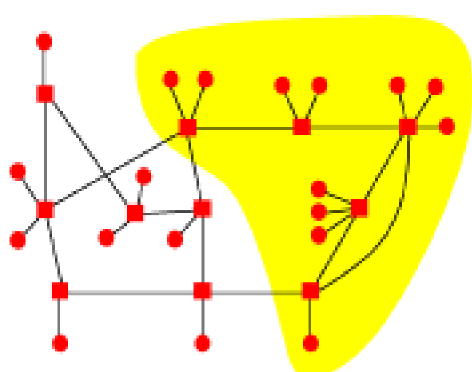
- Nodes communicate with each other with equal probability

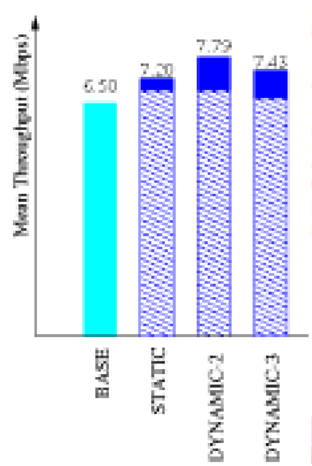


Carnegie Mellon

Exp. 3: Unbalanced Load

- Nodes inside island are 10 times more active than the other ones

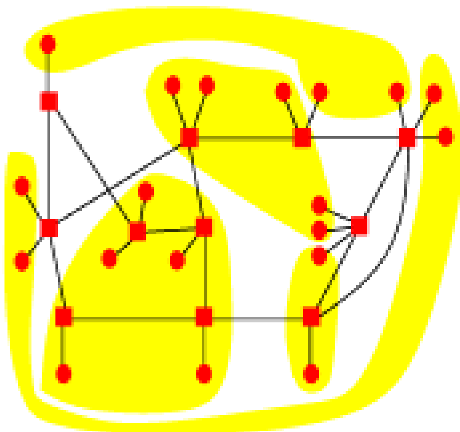


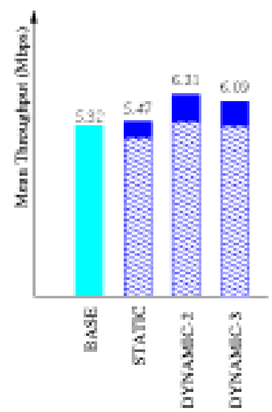


Carnegie Mellon

Exp. 3: Virtually Partitioned Network

- Only nodes within the same island communicate among them



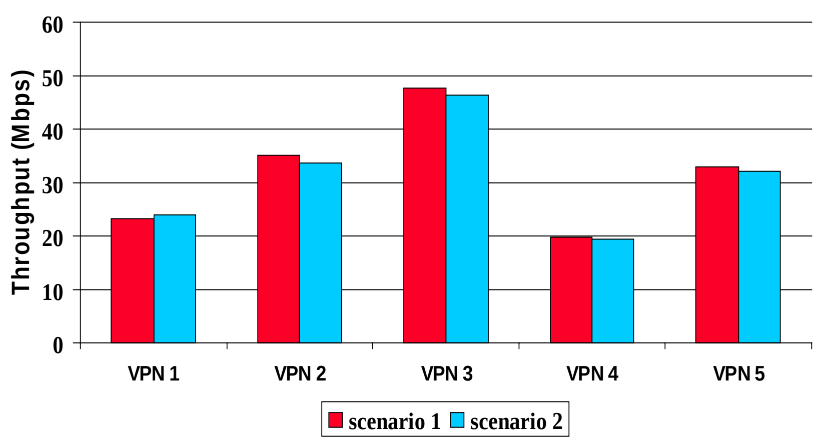


1997 Hui Zhang

Carnegie Mellon

VPN Experiment: VPN Total Throughput

- Scenario 1: token rate of each VPN: $2.4*10^8$ tokens/sec
- Scenario 2: token rate of VPN 1 changed to: $4.8*10^8$ tokens/sec

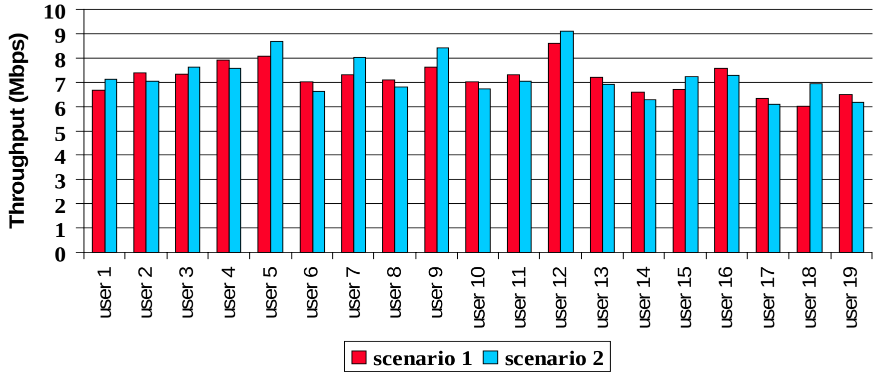


1997 Hui Zhang

Carnegie Mellon

Balanced Load: Total Throughput

- Scenario 1: token rate of each user: $0.5*10^8$ tokens/sec
- Scenario 2: token rate of users 1,3,5,7,9,12,15,18 changed to: 10^8 tokens/sec



1997 Hui Zhang

51