Implementation and evaluation of OSPF Optimized Multipath Routing

Balasubramanian Ramachandran

M.S. Thesis defense

Committee:

Dr. Joseph B. Evans (chair)

Dr. David W. Petr

Dr. Susan Gauch

I T T C

Organization

- Introduction
- Motivation
- Open Shortest Path First (OSPF) Optimized Multipath(OMP) and Opaque-LSA overview
- Design and Implementation
- Performance Evaluation
- Conclusions

ITTC

Introduction

- Traffic Engineering - What is it?
- Objectives
  - Improve network performance
  - Utilize resources efficiently
    - load-balancing in presence of varying traffic patterns
- Styles
  - Off line
  - On-line

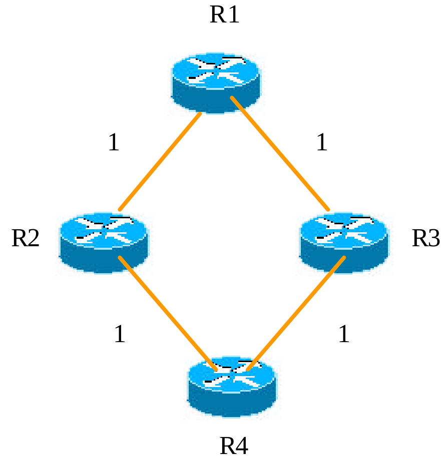
Motivation

- On-line load balancing hard
  - "It is easier to move a problem around than it is to solve it" - Ross Callon
  - Necessity for efficient algorithms imperative for online load balancing
  - Uniform link utilization in networks
  - stability concerns

I T T C

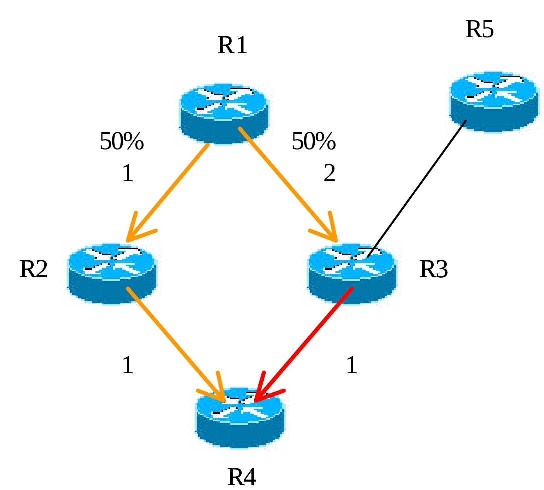
Adaptive weights method

- Given traffic demand, optimization using link metrics is not possible
- Dynamic weights
- Use of multiple equal-cost paths came into practice



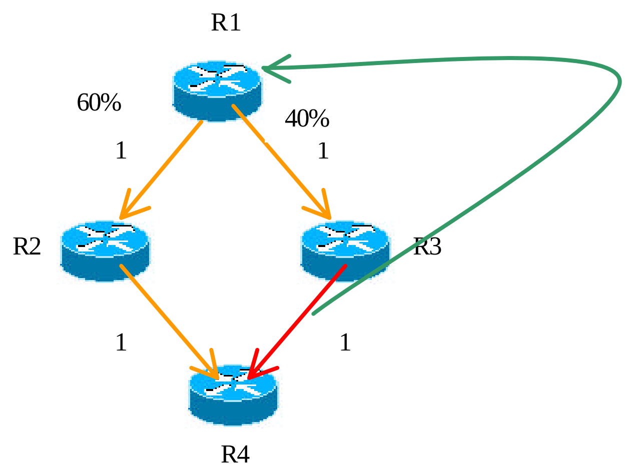
Equal Cost Multipath (ECMP)

- R1-R4 uses ECMP
- Simple and stable
- Congestion caused by overlapping of shortest paths
  - R1 Unaware of R3-R4 link utilization
- Consider cost (R1-R3-R4) just greater than (R1-R2-R4)

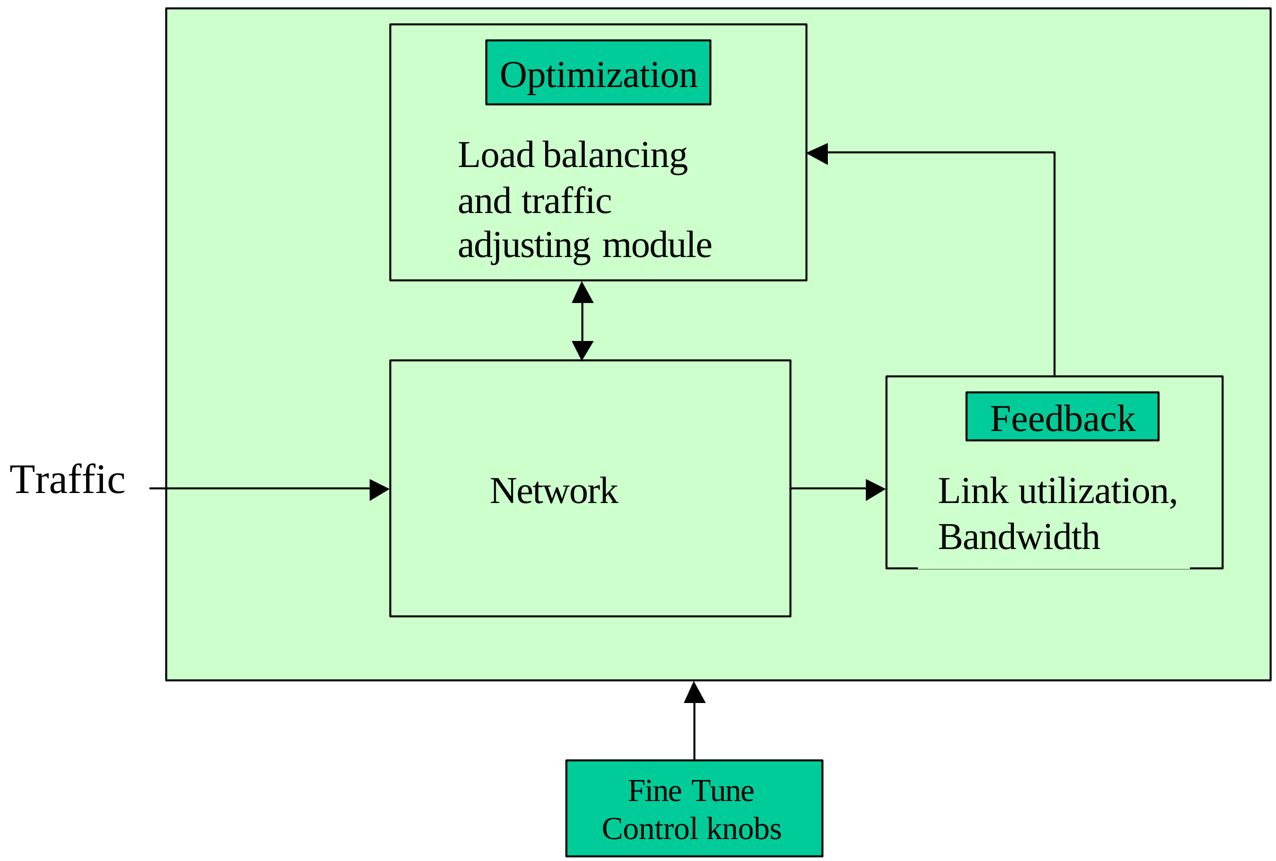


OSPF OMP Overview

- Network designed to support multiple paths between high traffic end-points
- Identify congested links
  - Use IGP to flood load statistics
- move traffic away from congested paths
  - vary traffic injection in multiple paths based on link utilization
  - soft on previous congested links
- Relaxes shortest path criteria



OMP Model in TE context



OSPF Opaque-LSA

- Facilitates dissemination of application oriented information using existing infrastructure
- Link-local, area-local, Autonomous System (AS)-local scopes
- Trade-off
  - additional traffic over-head
  - Additional memory

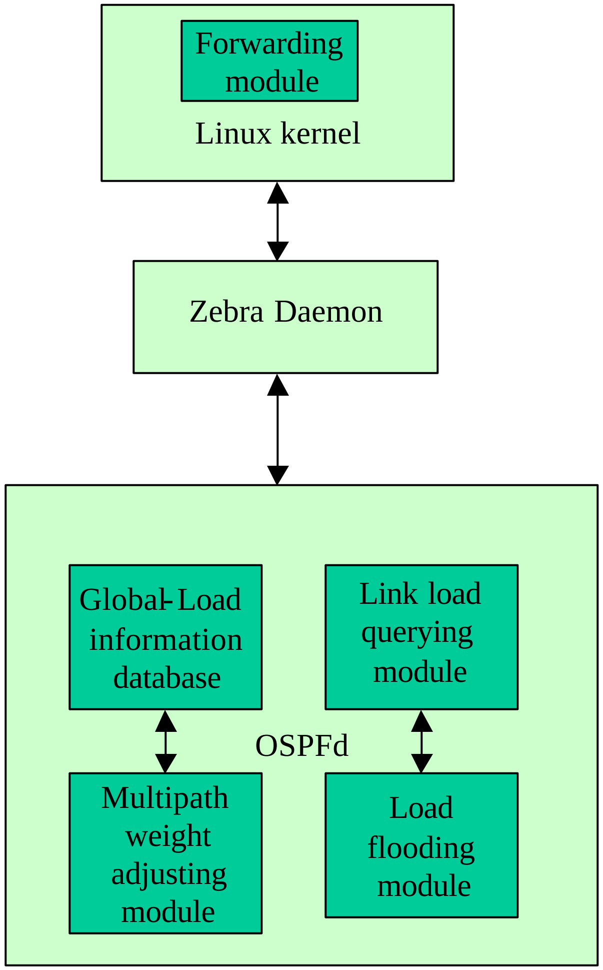
OSPF Opaque LSA packet format

<table>
<tr><td colspan="2">Link State Advertisement Age</td><td>Options</td><td>LSA Type</td></tr>
<tr><td>Opaque Type</td><td colspan="3">Opaque ID</td></tr>
<tr><td colspan="4">Advertising Router</td></tr>
<tr><td colspan="4">Link State Advertisement Sequence Number</td></tr>
<tr><td colspan="2">LSA Checksum</td><td colspan="2">LSA length</td></tr>
<tr><td colspan="4">Application oriented information</td></tr>
</table>

- Link State Advertisement (LSA) types 9,10 and 11
- opaque type/opaque id replacing LS id
- Lsa header followed by application specific info

Design and Implementation

- OSPFd
  - load query
  - load flood
  - traffic adjustments
- Kernel
  - Forwarding



Link load querying module

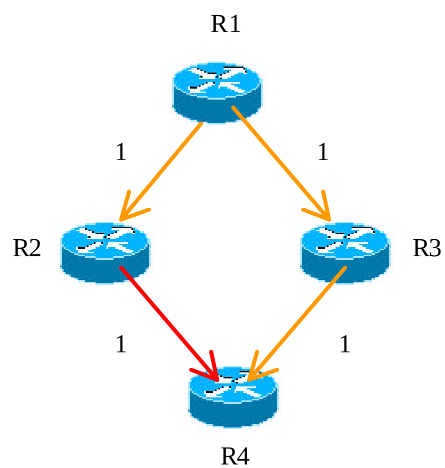
- Interface Management Information Base (MIB) parameters sampled every 15 sec
- Values are filtered using a simple filter
- Fractional Link utilization calculated

Load flooding module

- uses type-9 opaque LSA
- fractional link utilization, link bandwidth
- flooding decision based on
  – current value of the load
  – difference between current and previous loads
  – elapsed time since last flooding
- trade-off: flooding frequency and traffic overhead

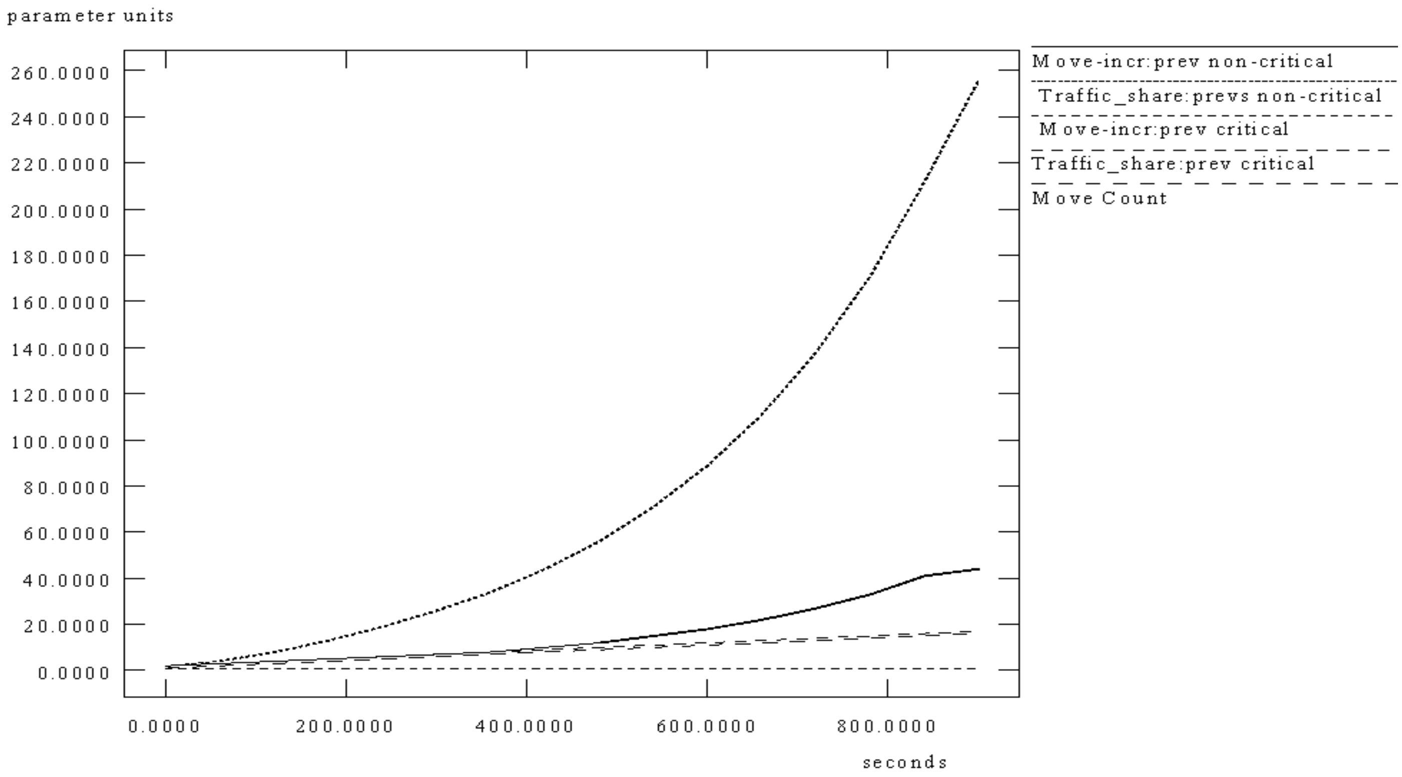
Nexthop structures

- For each multipath destination,
  - list of nodes from source to destination {R1-R2-R4, R1-R3-R4}
  - critical segment R2-R4
  - previous critical segment
  - Traffic adjustment information

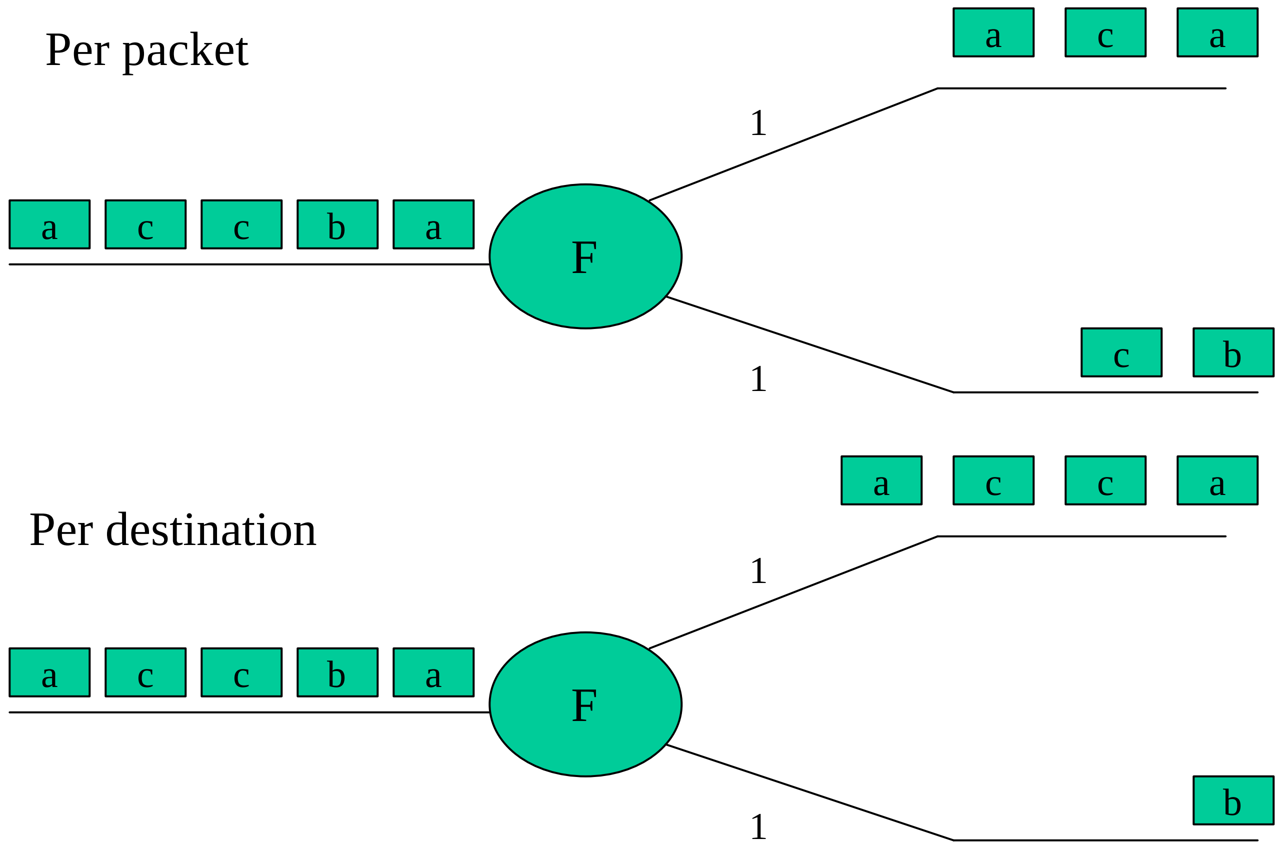


Traffic adjustments

- Traffic moved away from congested links
- move exponentially into non congested paths
  - To ease out congestion quickly

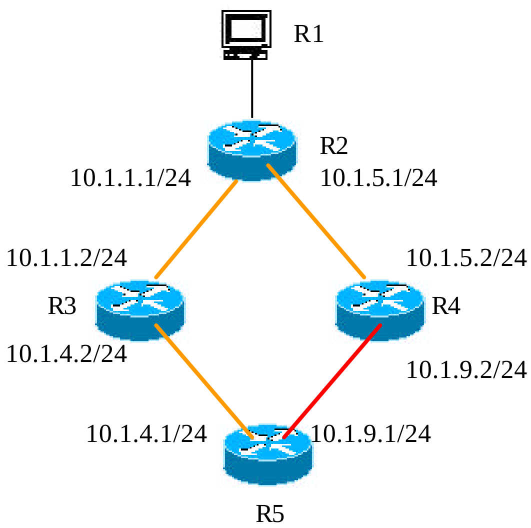


Forwarding module



Evaluation

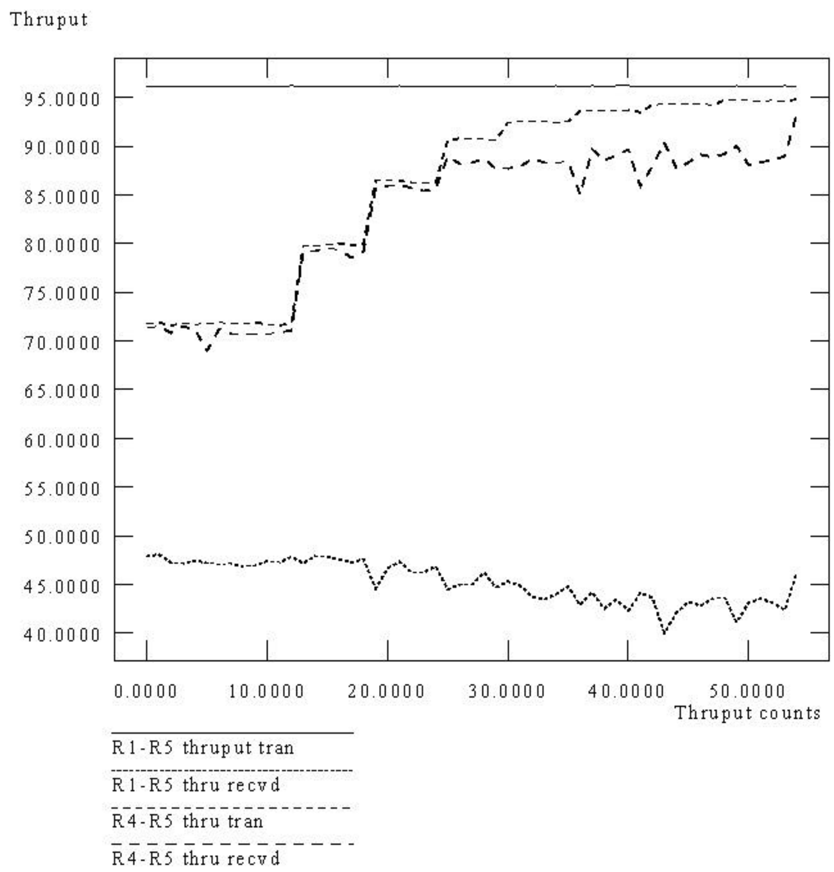
- Opaque - LSA propagation time
  - less than 2 seconds
- Per-packet load balancing tests
  - UDP burst traffic generated from R1 towards R5
  - R2->R5 have multipaths
  - R4-R5 link congested

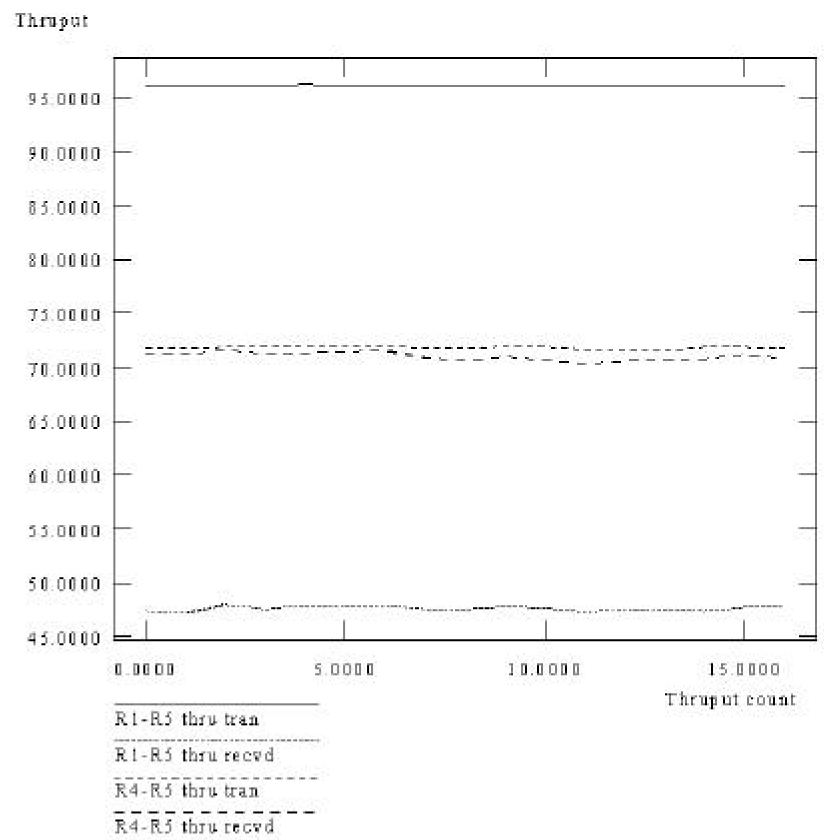


UDP traffic

- Starts with Equal traffic distribution
- On feedback R2 sends more traffic onto R3
- R3-R5 link utilization increases
- R4-R5 link utilization decreases

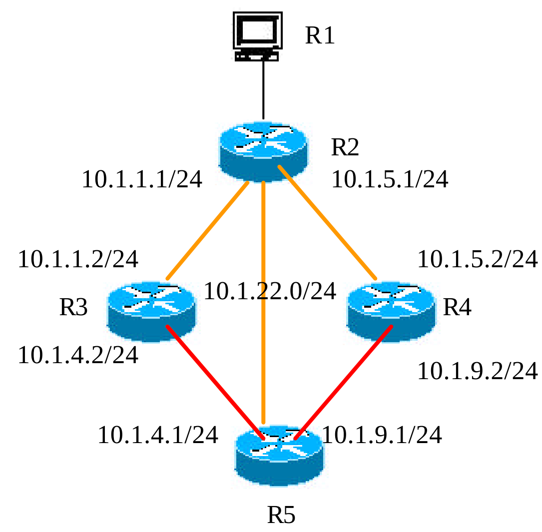
UDP Traffic contd...



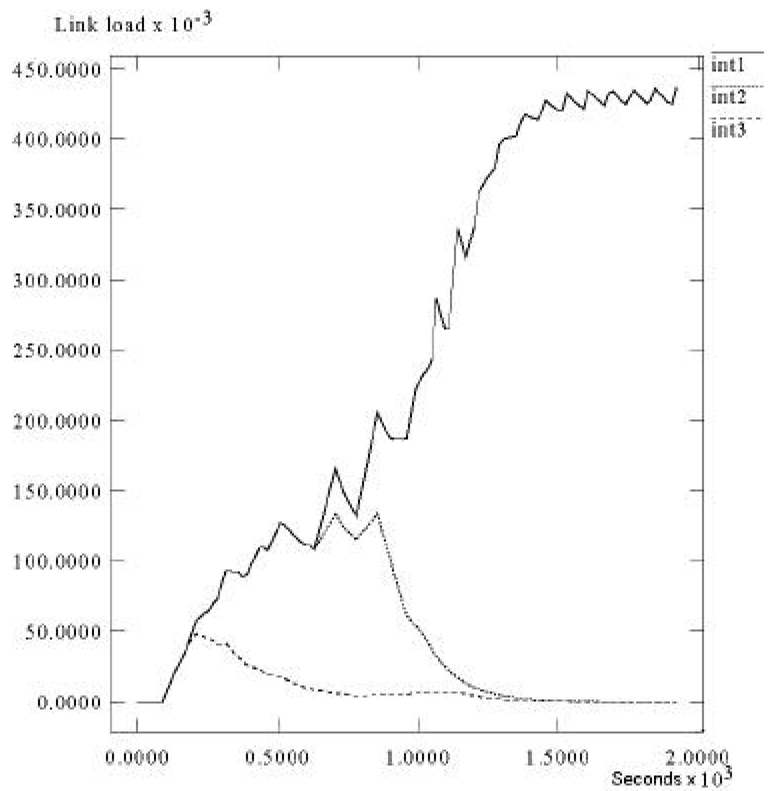


UDP Traffic contd...

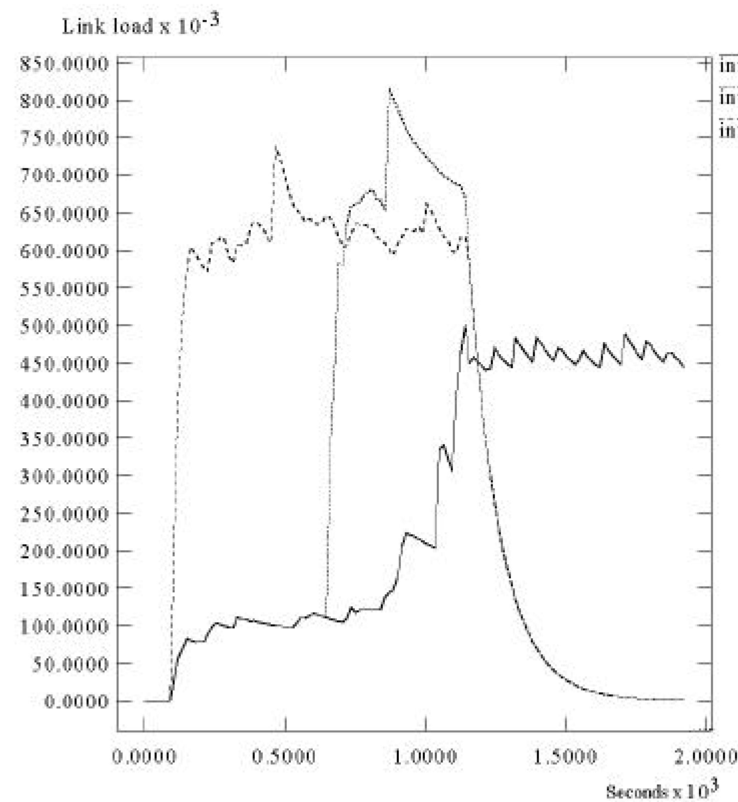
- Three paths from R2 to R5
- High link util in R4-R5
- After about 645 seconds R3-R5 link util is increased



UDP Traffic contd...



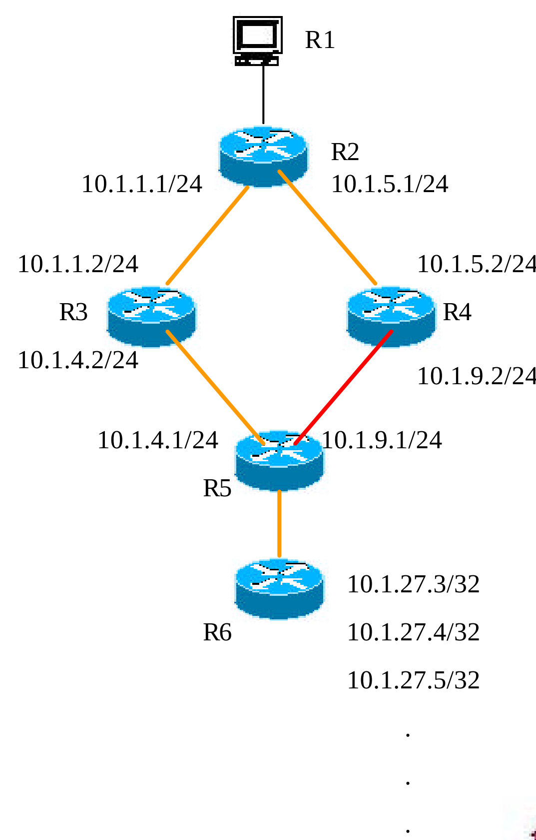
Link characteristics at R2



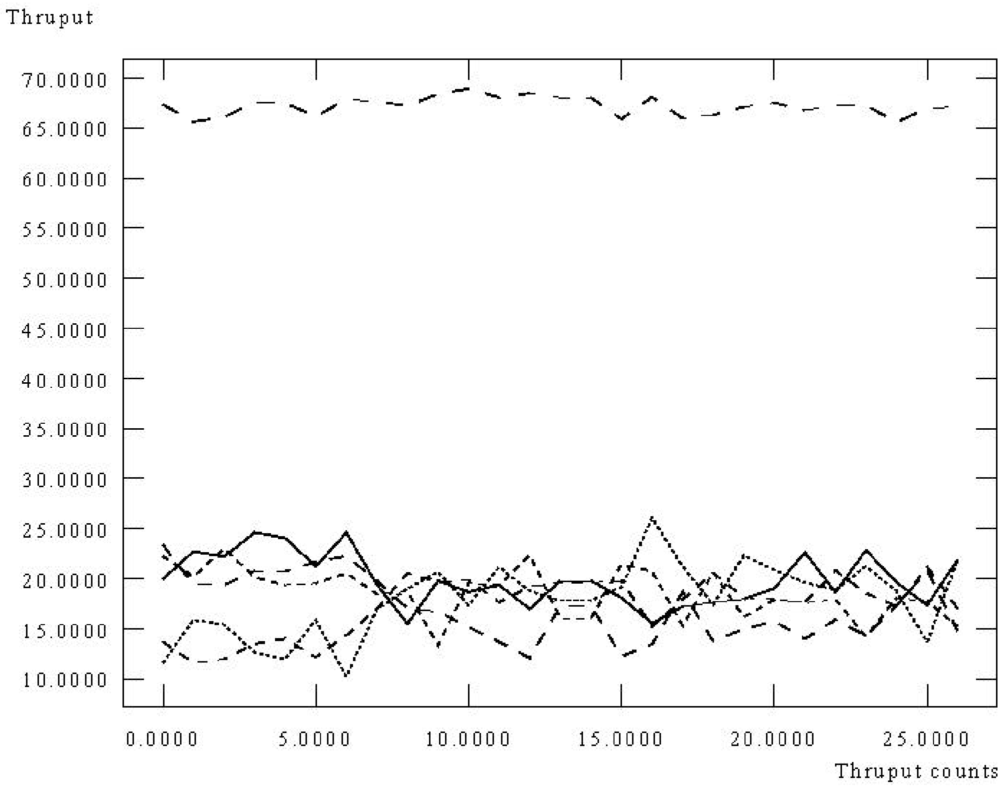
Link characteristics at R5

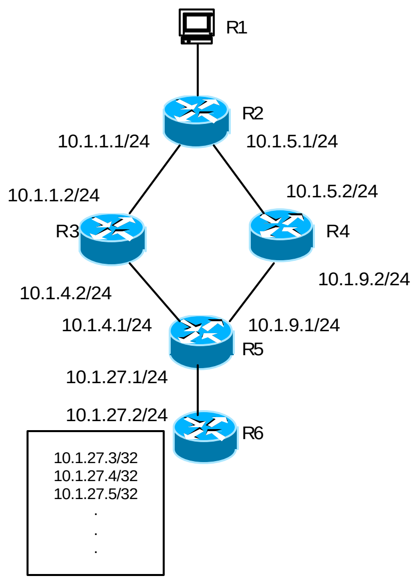
TCP Traffic

- TCP traffic
  - 1500 byte serialized in 120? sec in 100 Mbps
  - If Delay diff > 3 * serialization time, packet re-ordering can occur
  - Poor thruput in per-packet load balancing
- Per-destination load balancing
  - Traffic generated from R1 towards R6
  - R4-R5 link congested
  - Thruput low for flows taking R4 nexthop
  - Feedback shifts more flows from R4 to R3

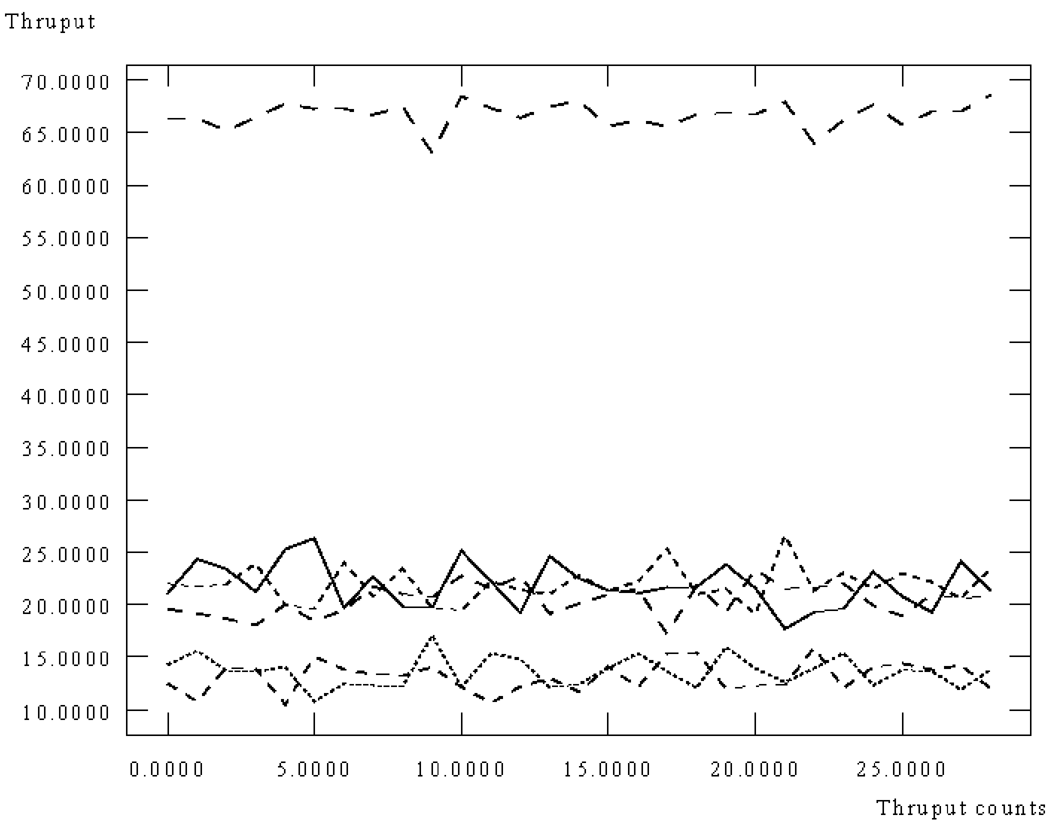


TCP Traffic contd...



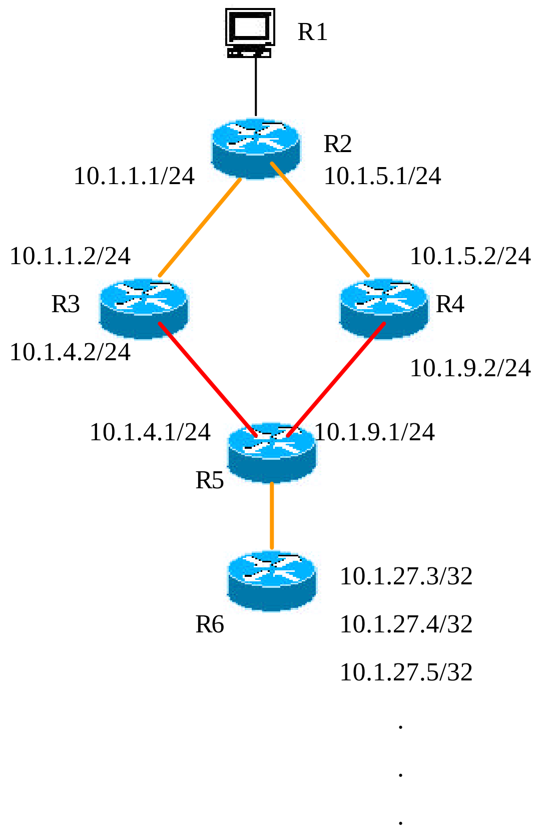


TCP Traffic contd...

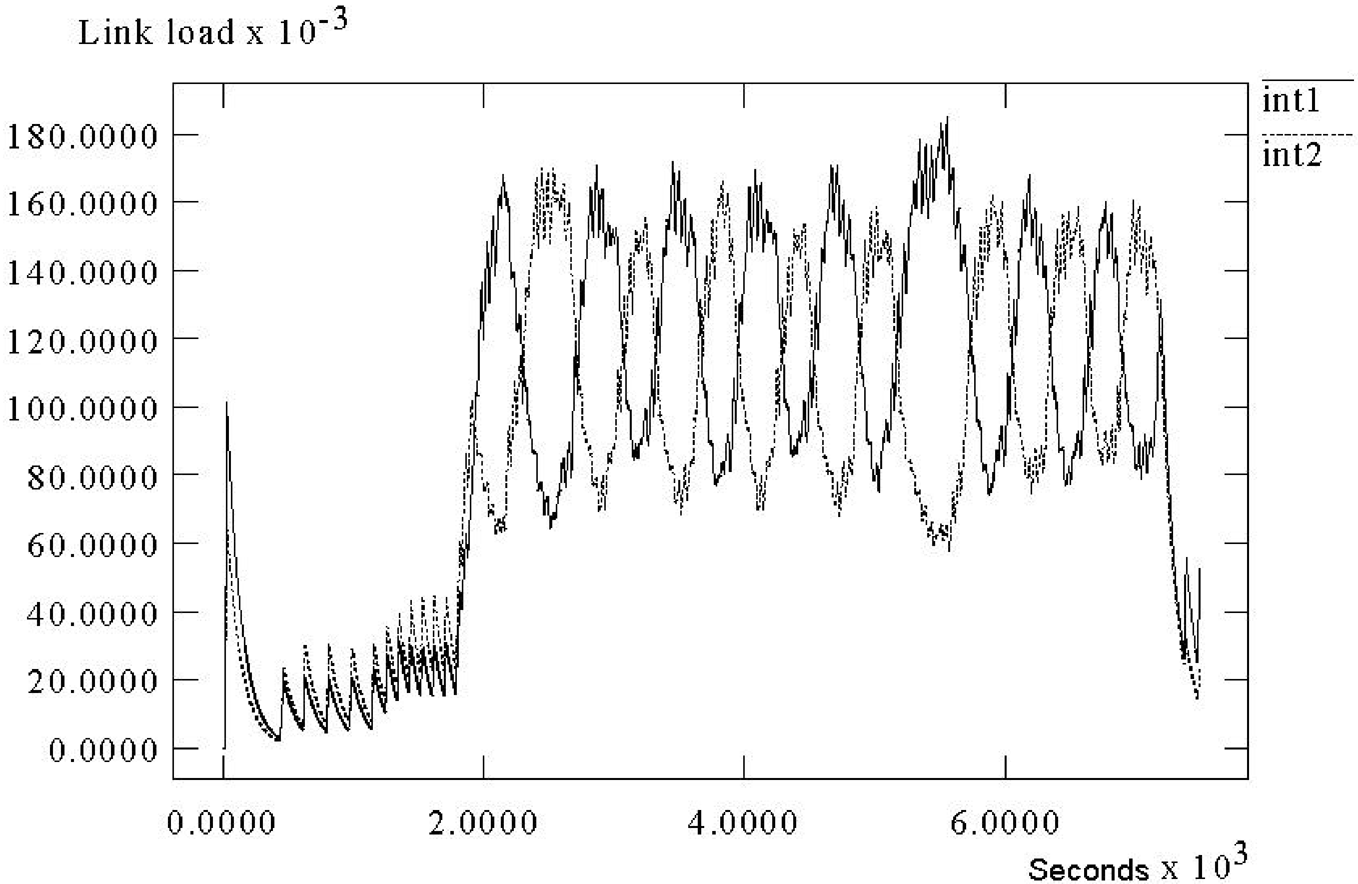


TCP Traffic contd...

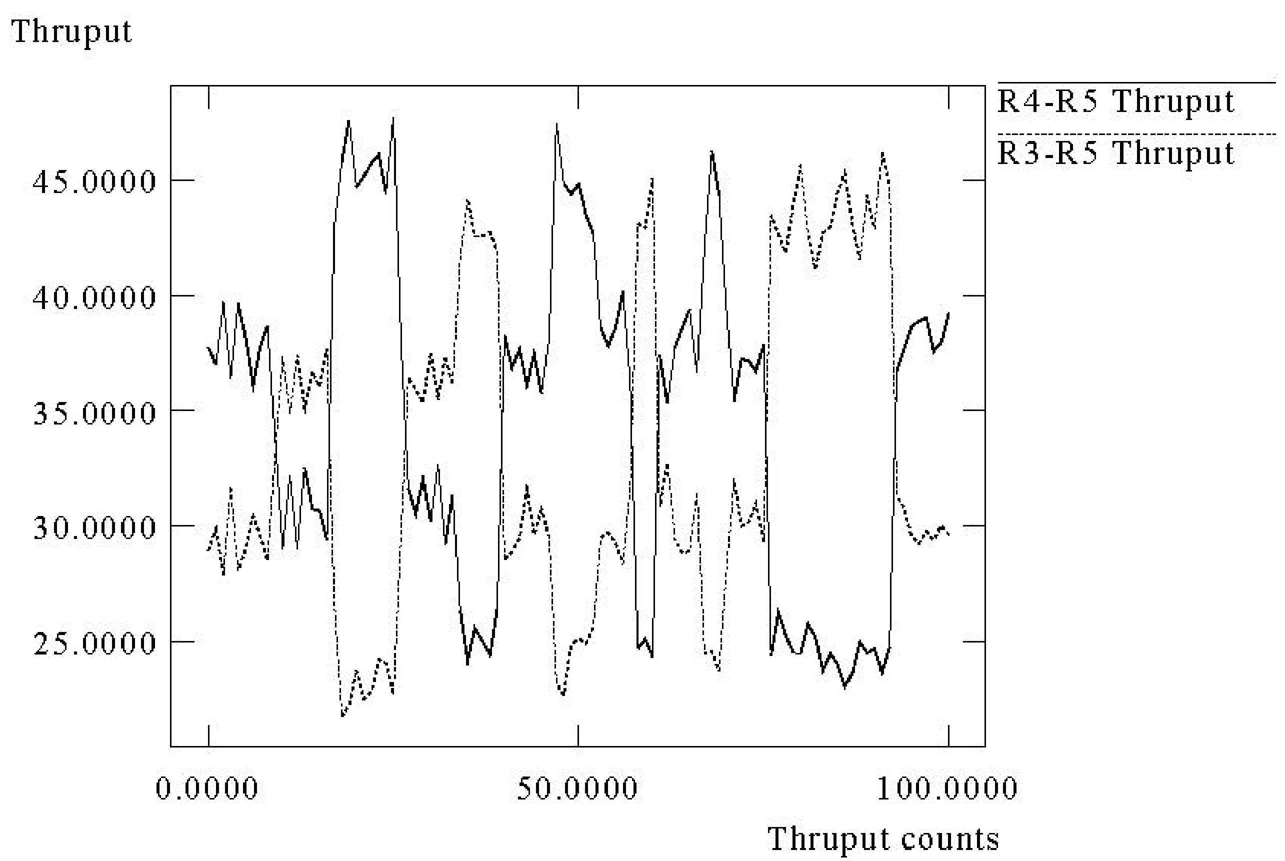
- Both R3-R5 and R4-R5 congested
- Only one critical segment
- Not enough to prove instability
- No traffic shifts in the midst
- Hash-space adjustment will dampen oscillation



TCP Traffic contd...



TCP Traffic contd...



Conclusions

- Algorithm achieves gradual traffic shift
- Rate of traffic shift into a path depends on previous congestion
- Big networks opaque-LSA propagation time ??
  - Speed of feedback, response determines success
- Trade-off
  - Complexity
  - Traffic overhead
- How effective would over-provisioning be?

Future Work

- Implementation can be extended to support other link types and to inter-area
- Relax shortest path criteria
- Framework used to evaluate MPLS-OMP

THANK YOU