Load Balancing Inbound Traffic in Multihomed Stub Autonomous Systems

Ashok Singh Sairam and Gautam Barua
Department of Computer Science and Engineering
IIT Guwahati

Agenda

- Motivation
- Problem Description
- Our Approach
  I. Offline Models
  II. Online Models
- Experimental Analysis
- Conclusion.

Agenda

- Motivation
- Problem Description
- Our Approach
  - I. Offline Models
  - II. Online Models
- Experimental Analysis
- Conclusion.

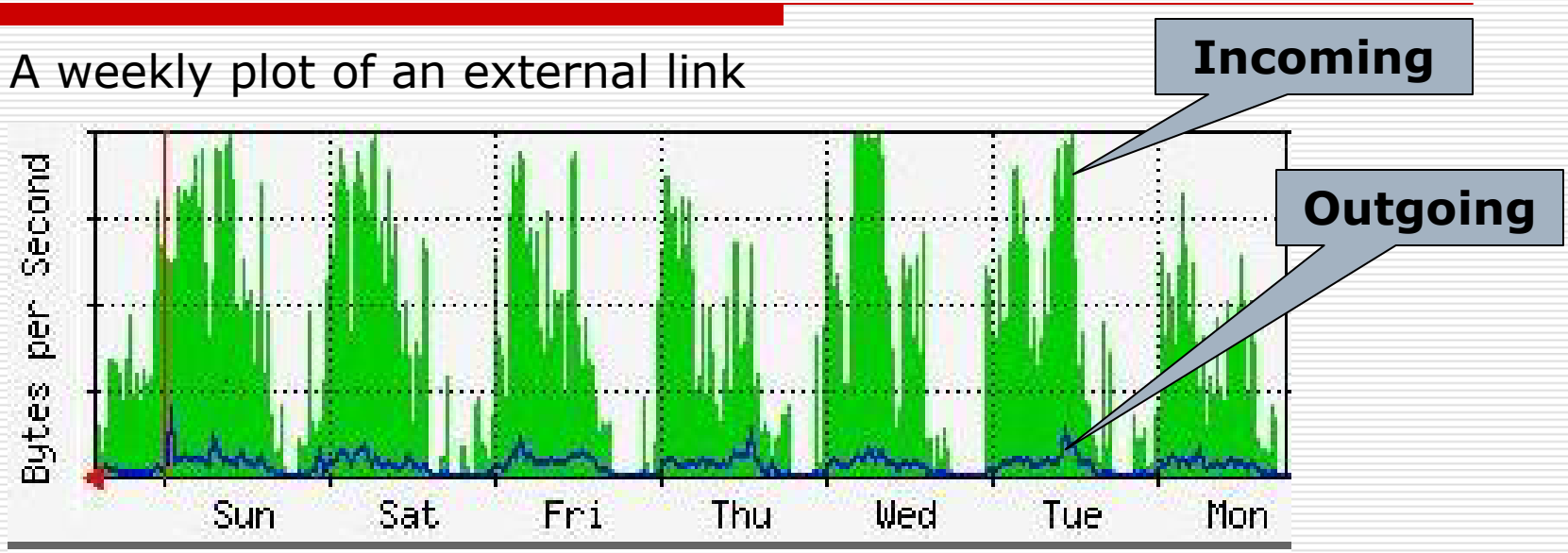
Factors Affecting Internet Performance

- Local Factors – Last mile connectivity of source; ill-configured software stack etc. Can be handled by locally upgrading the system

- Remote Factors – Last mile connectivity of destination; congestion between source and destination ISP; performance of remote server. Over-coming these factors require global co-ordination.

Internet Traffic Characteristics

Fig: A weekly plot of an external link



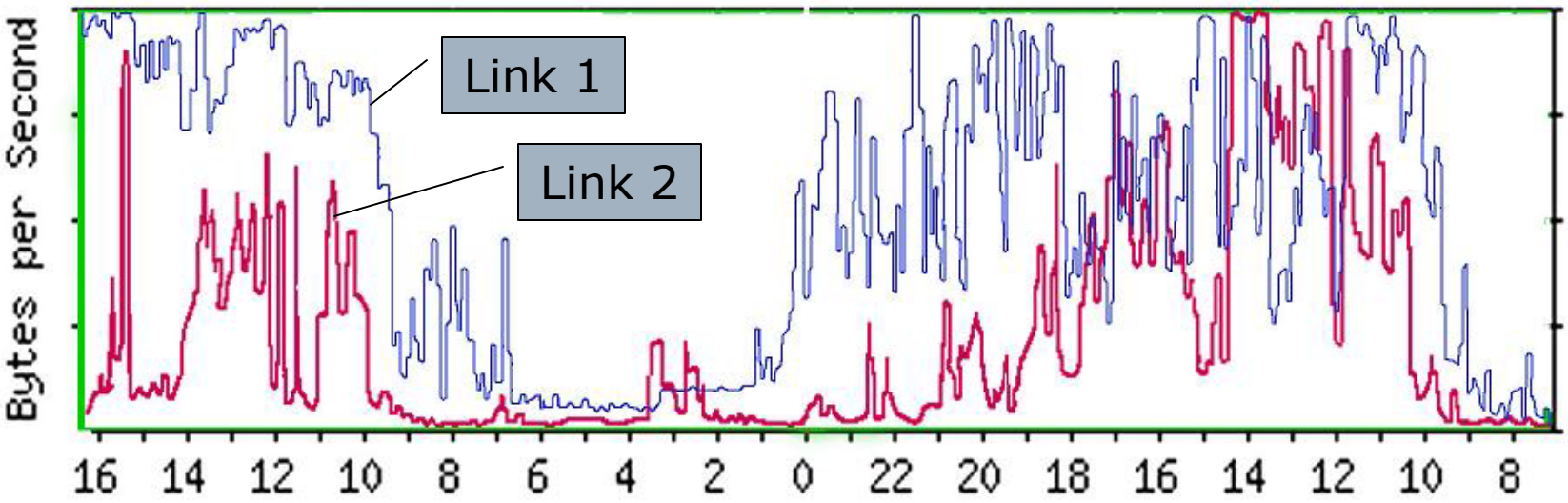
- Observation 1: Internet traffic bi-directional but asymmetric. Focus is on Content Receivers.
- Observation 2: Internet traffic has heavy-tailed distribution but chaotic for small time-scales.

Internet Traffic Characteristics – contd.

- Observation 3: Majority traffic contributed by few users. 10% users contribute 90% traffic.
- Observation 4: The trend toward multihoming is prevalent not only in stub ASes but also in small end networks . Multihoming is increasingly used to provide redundancy as well as performance benefits.

Internet Traffic Characteristics – contd.

Fig: Comparison of Incoming Traffic of a 2 ISP Multihomed Network.



- Observation 5: In multihomed networks, utilization of the links is skewed.

Internet Traffic Characteristics – contd.

- Observation 6: Although significance of inbound traffic control have been felt, the issue has not been addressed properly.
- Observation 7: Controlling of inbound traffic is difficult since it involves influencing the remote destination. Inbound traffic cannot be controlled by directly acting on the traffic.

Objective & Goal

- How to load balance incoming traffic in a multihomed, stub network to improve the overall Internet experience?
- The goal is to control incoming traffic by regulating the corresponding outgoing traffic

Agenda

- Motivation
- Problem Description
- Our Approach
  - I. Offline Models
  - II. Online Models
- Experimental Analysis
- Conclusion.

COMSNETS 2009

Problem Statement

We define the following Sets/Variables

- External Links, $E=\{e_1,e_2,\ldots,e_N\}$
- Users, $I=\{i_1,i_2,..,i_S\}$
- Time Instants, $T=\{t_1,t_2,\ldots\}$
- $A_e(t)$ = Available Bandwidth of ***e*** (Mbps)
- $U_e(t)$ = Utilization (Mbps) of link ***e***
- $b_i(t)$ = Incoming traffic of user ***i*** (Mbps)
- $x_{ie}(t)$ = 1, if user ***i*** assigned ***e,*** else 0

Problem Statement – contd.

- Ideal utilization of link ***e,***

$$iU_e(t) = \frac{\sum_{l=1}^{N} U_l(t)}{\sum_{l=1}^{N} A_l(t)} A_e(t)$$

Problem Statement – contd.

- ***Objective 1:*** Keep link utilization as close as possible to their Ideal value

$$f1: \min |iUe(t) - Ue(t)| \forall e \in E$$

- ***Objective 2:*** Minimize user re-assignments

$$f2: \min \sum_{e=1}^{N} \sum_{i=1}^{S} | x_{ie}(t) - x_{ie}(t-1) |$$

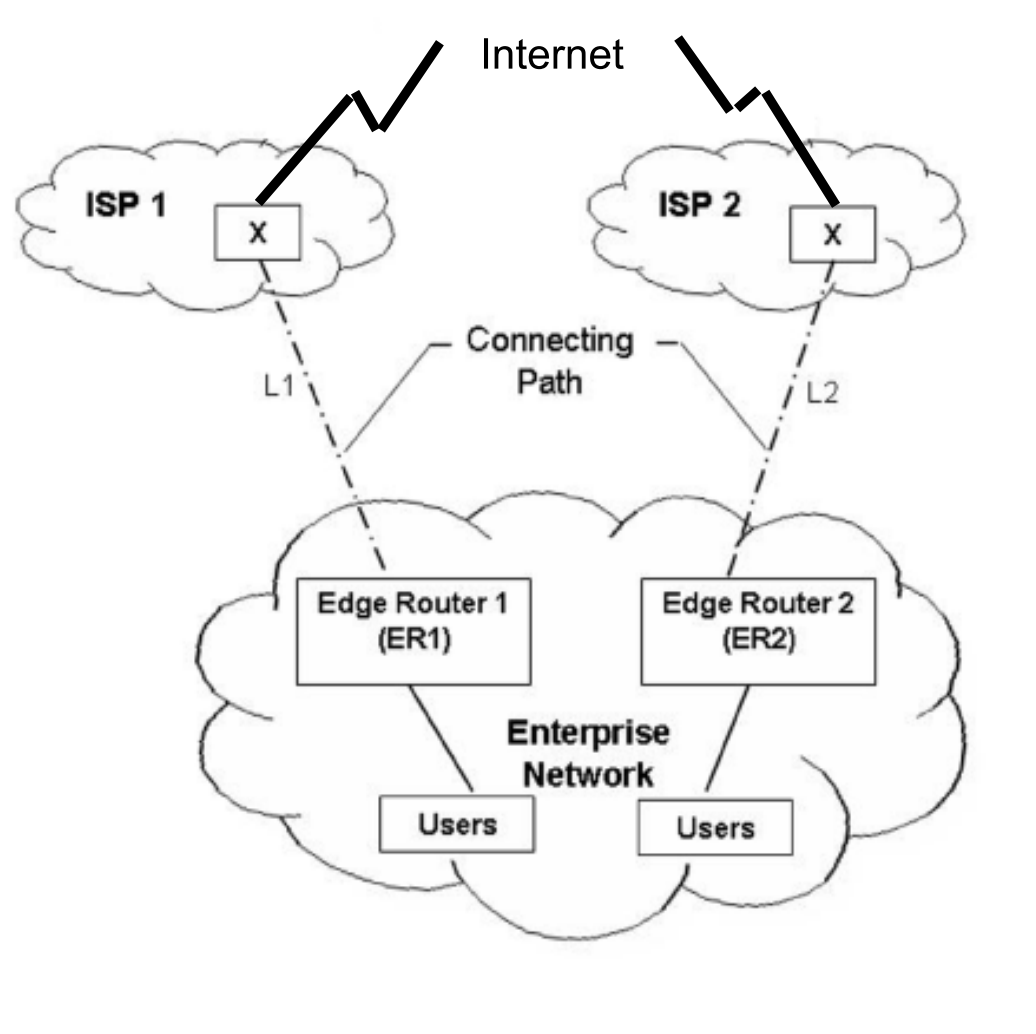
Problem Statement – contd.

- ***Objective 3:*** In case of more then one equally good external routes select the one with least intra-domain cost. Let d(i,e) denote the intra-domain cost from user ***i*** to link ***e*** and $E' \subseteq E$ denote the set of equally good routes

$$f3 : d(i,e) \leq d(i,e'), \forall e, e' \in E'$$

- Handle the problem at two levels – offline & online.

Measuring Available Bandwidth



Connecting Link

- There is a point X to which all external packets travel. Assume beyond point X, there is infinite bandwidth.
- Point X is to be discovered and available bandwidth between source network and X needs to be discovered.
- We call such paths connecting path or connecting link.

Connecting Link – contd.

- If length of connecting link = 1, then SNMP probe the edge router and measure the utilization (as well as available bandwidth).
- Are Tools Available for the general case?
  - Survey a no. of open-source active/passive tools.
  - Most tools measure end-to-end bandwidth and require cooperation at both ends.
- Pathneck [Ningning Hu et. al.]: Light-weight, single-ended control.
- Probing source sends recursive packet trains. Identifies choke points and provides an upper-bound on the available bandwidth of the path.

Agenda

- Motivation
- Problem Description
- Our Approach
  - I. Offline Models
  - II. Online Models
- Experimental Analysis
- Conclusion.

Offline Models

- *First Assumption*: Input traffic is known, i.e. $b_i(t)$'s are known.
- Compute total incoming traffic & utilization

$$T\_traffic \leftarrow \sum_{i=1}^{S} b_i(t);$$

$$U_e(t) \leftarrow U_e(t) + b_i(t), \text{ if } x_{ie}(t) = 1, \ \forall i = 1..S$$

- Available Bandwidths ($A_e(t)$) can also be measured. Therefore, compute $iU_e(t)$.
- Rank of link e, $rank_e(t) \leftarrow iU_e(t) - U_e(t)$

Re-Assignment Problem

- If $rank_e(t) > 0$, link e is under-utilized and if $rank_e(t) < 0$, it is over-utilized.
- **"Re-allocate some users from over-utilized to under-utilized link such that the traffic load is balanced".** We call it Re-assignment problem.
- Re-assignment Problem is NP-Complete.

Relation between Over-utilized and Under-utilized link

- Theorem: If there are over-utilized links then there must exist one or more under-utilized links. Absolute sum of ranks of over-utilized link equals absolute sum of ranks of under-utilized link. Proof

Offline Models - contd

- *Second Assumption:* Input traffic known and bounded

$$\text{if } x_{ie}(t) = 1 \text{ then } 0 \le b_i(t) \le A_e(t).$$

- The problem is pseudo-polynomial which technically means it is exponential too!!!
- We further restrict the values of b_i to two – 0 or 1. We can have an exact solution of O(S) time.

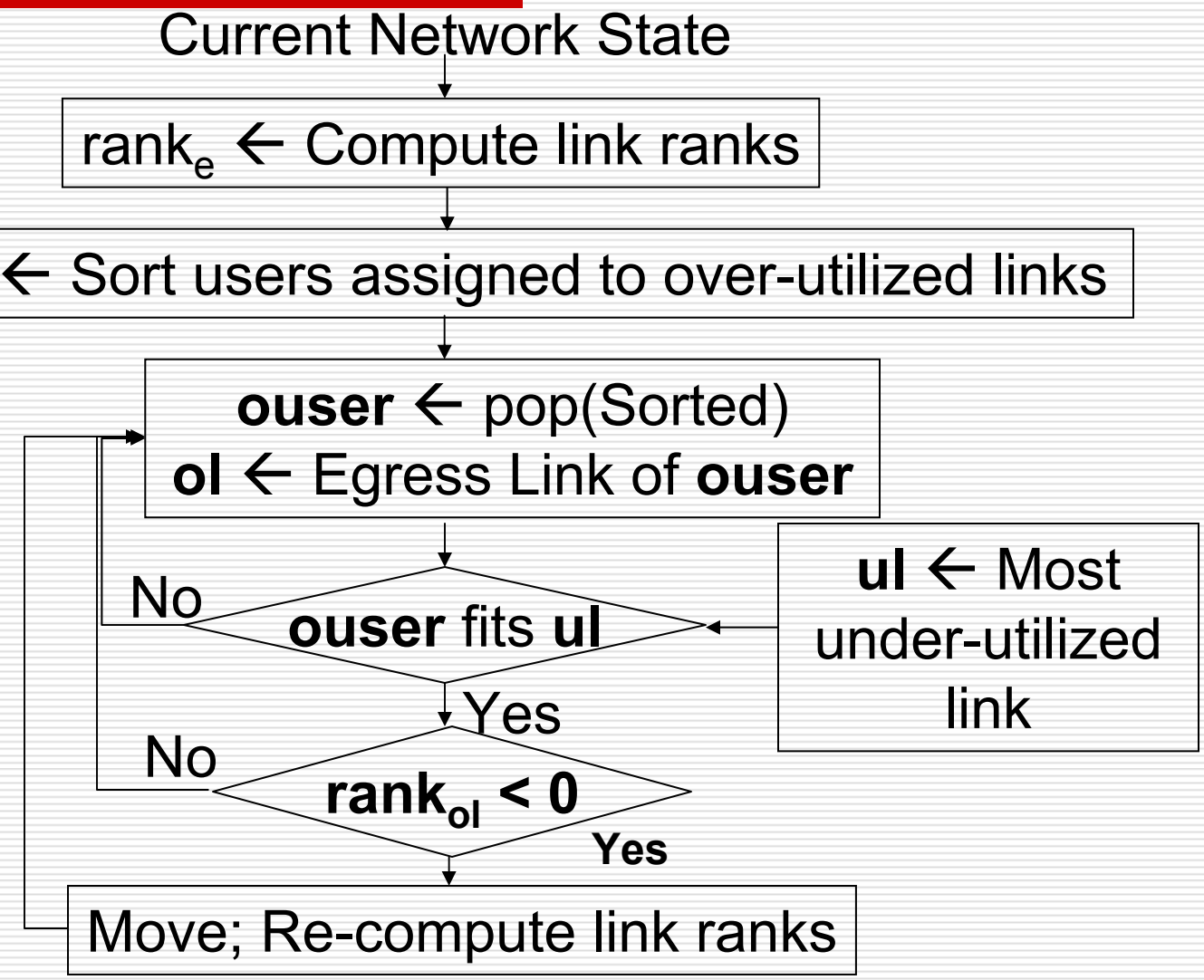
Agenda

- Motivation
- Problem Description
- Our Approach
  - I. Offline Models
  - II. Online Models
- Experimental Analysis
- Conclusion.

Online Models: Greedy Approach

- *Assumption:* No restrictions on Input Traffic but Intradomain traffic static.
- Global approach, bandwidth manager sees traffic dynamics of the whole network.
- Modeled as a *recursive* two stage recourse problem. The second stage acts as the first stage for the next run.
- Based on current measurement make provisions for the next period.

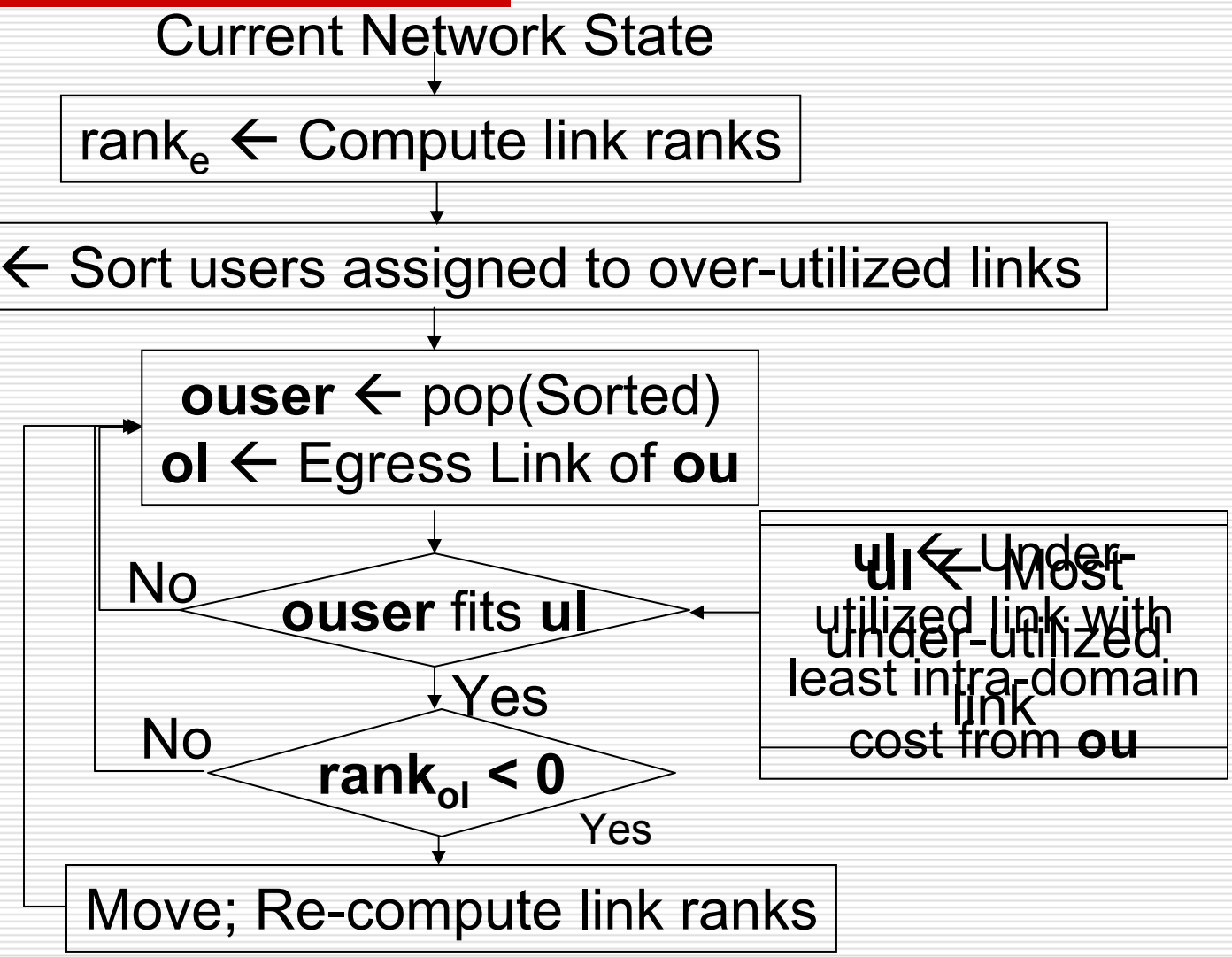
Greedy Approach



Greedy with Intra-domain cost.

- In Internet architecture when the interdomain protocol learns several routes to the same destination, intra-domain protocol is used to break the tie, select one with least cost.
- Drawing an analogy, all under-utilized links that fits an user are equally good routes. Select the one with least cost

Greedy with Intra-domain cost – Algorithm



Agenda

- Motivation
- Problem Description
- Our Approach
  - I. Offline Models
  - II. Online Models
- Experimental Analysis
- Conclusion.

COMSNETS 2009

Experiment setup - Data

- Experiments were conducted using both synthetic as well as actual data.
- Collected actual traces from the institute's various egress points.
- Analysed the traces:
  - Majority of the connections have low throughput and are short-lived (mice).
  - Small fraction of users account for major traffic volume. They are high throughput, long duration flows (elephants).

Experiment 1

- Extracted the distinct users (source IP addresses) present in each of the trace. Next, merged the traces chronologically.
- Read the merged trace. Timestamp of first packet marks beginning of the simulation.
- Sum link-wise, the payload of all packets that fall within a period (default utilizations). Duration of a period is 5 mins.

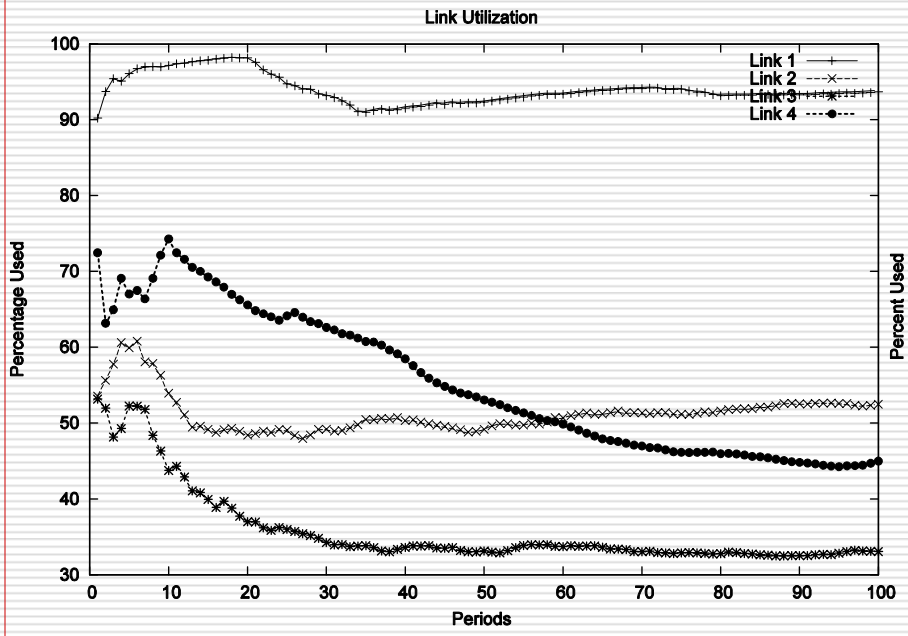
Experiment 1 Results

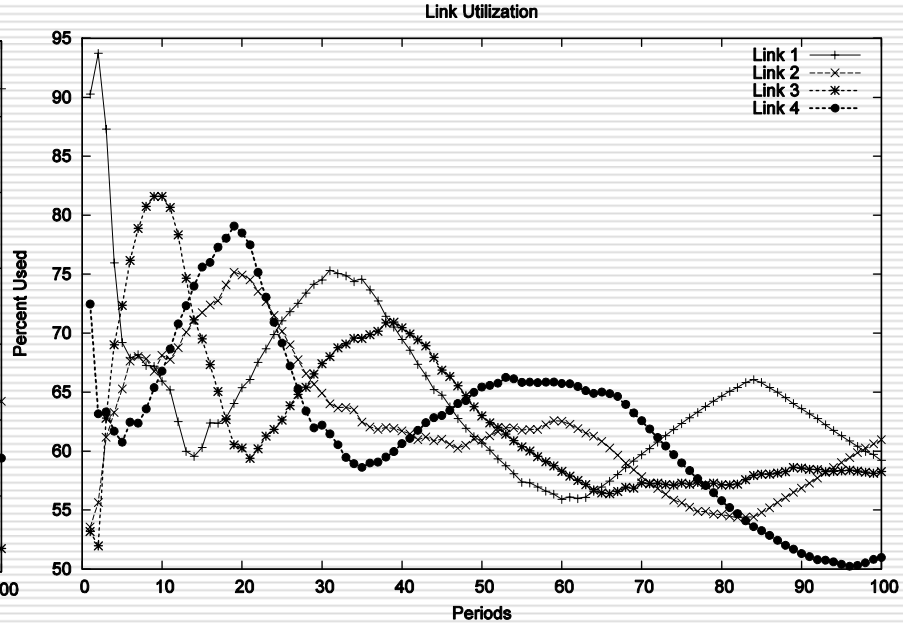
- Simulated greedy approach on the merged trace. Tool written in perl.
- No. of edge links = 4; Trace Duration =4 hours; Three runs – Default, No cross-traffic, and with cross-traffic.
- Rank of a link is a measure of the deviation of a link from its ideal value. Compared ranks.
- Improvement in load balancing 53 percent when no cross-traffic was considered and 43 percent when cross-traffic was considered.
- No. of user re-assignments was about 2 percent

Experiment 2: Validating Using Synthetic Data

- Created basic topology (using BRITE). Later enhanced the topology by dynamically changing the link attributes.
- Used three different traffic models of ns-2: packmime, PagePool/WebTraf and FTP server
- Topology: 25 nodes; 50 Intradomain links; 4 Edge links; 100 period duration
- Improvement in RTT 7 percent

Experimental 2- Results





Experiment 3 – Validating using actual data.

- Processed the traces so that they can be re-played on ns-2.
- For each distinct IP address, record the payload and interarrival time. As many ns2 traces as users present in the trace.
- For each ns2 trace, create a user node and destination node. Stream trace from destination
- To handle route changes create additional links from the destination.

Experiment 3: Network Model

Cost= Capacity/
Available Bandwidth

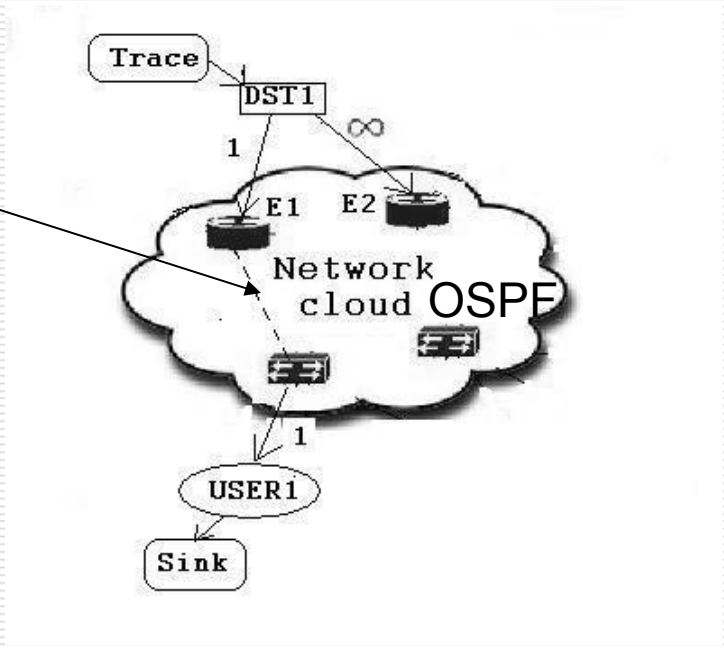


Fig : Network Topology

Experiment 3 – With Actual Traffic

- Topology: 50 Nodes, 300 Edges, 8 Egress Nodes, 17 access nodes. User Base 1500. Duration – 2 Hrs.
- Network Traffic state: Medium; Performance:

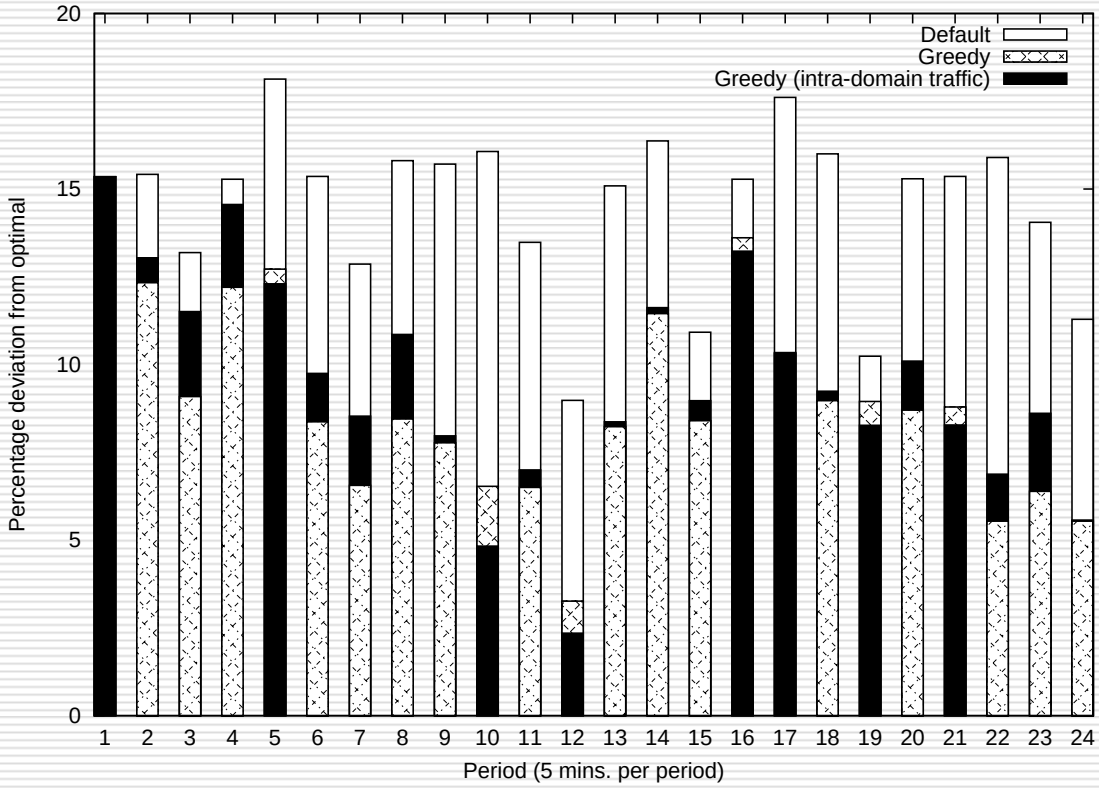


Fig. 12: Comparison of Ranks

Experiment 3 – contd.

- Plot of Users re-assigned
- No. of users moved lower.

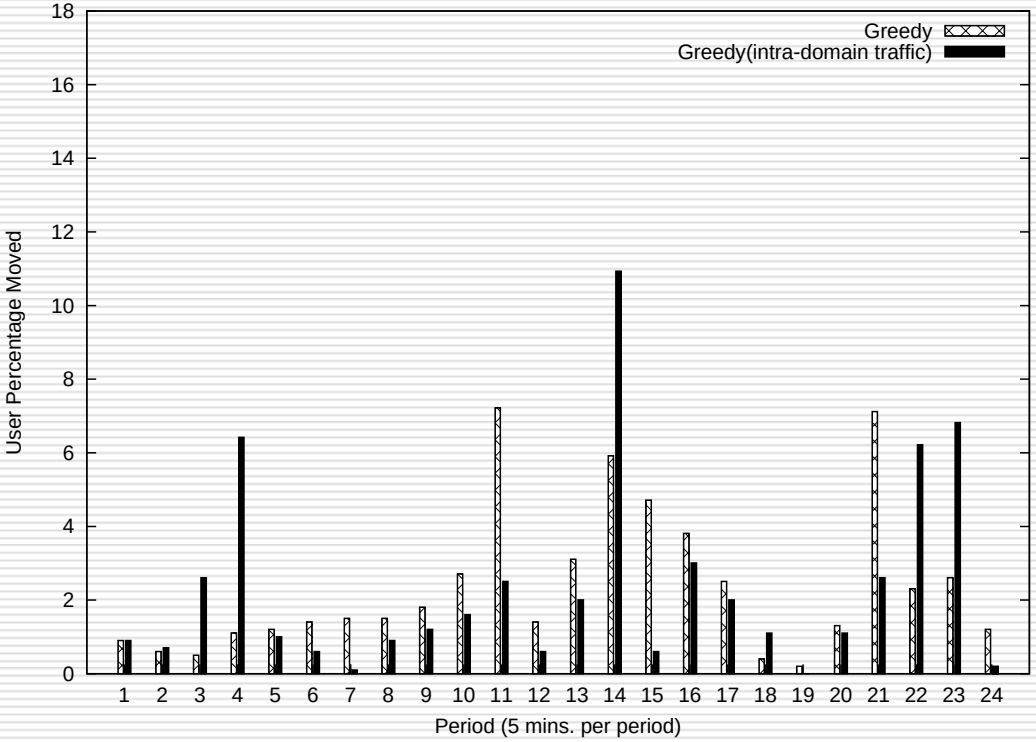


Fig. 13: Percentage of users re-assigned, about 2 percent

Experiment 3 – contd.

- In actual network rate of incoming traffic will depend on delay incurred by traffic within network and delay incurred by corresponding outgoing requests/acks.
- We cannot see the effect of such delays in our simulation.
- Therefore, computed the frequency of difference in intradomain route paths followed by the two greedy approaches.
- Above 20% route changes is detected.

Agenda

- Motivation
- Problem Description
- Our Approach
  - I. Offline Models
  - II. Online Models
- Experimental Analysis
- Conclusion.

Conclusions and Future Work

- Proposed a scheme to load balance traffic by keeping link utilizations in proportion to their available bandwidth.
- The scheme does not require cooperation from uplink providers.
- Improve Internet performance by making best use of the resources available.
- Proposal was tested under different traffic loads and network scenarios.

Conclusions – contd.

- Validated using real traffic traces and different topologies.
- Results show significant improvement with load balancing
- Future Work: Study the interaction of our route control mechanisms with that of the provider network and other competing networks.
- Propose a distributed version of our approach.


1969
The End

Theorem: If there are over-utilized links then there must exists one or more under-utilized links. The absolute value of sum of ranks of over-utilized links must equal sum of ranks of under-utilized links.

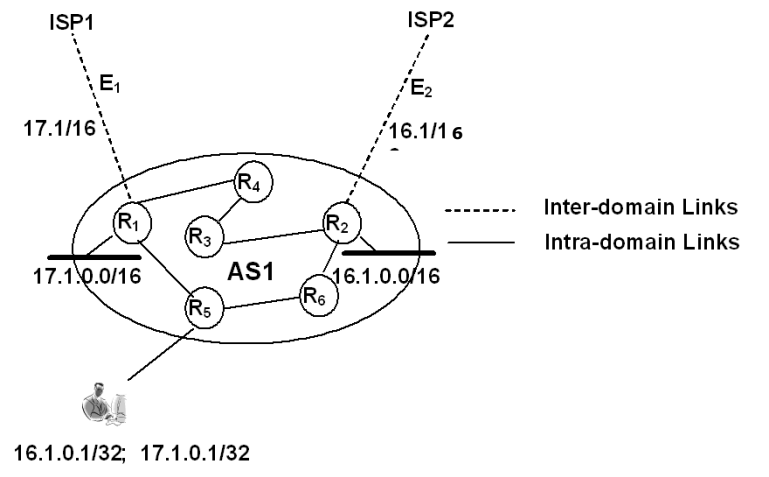
Proof: Let ideal utilization of link e, $iU_e(t)$ & utilization $U_e(t)$.

$$\sum_{e=1}^{N} U_e(t) = \sum_{e=1}^{N} iU_e(t) = \text{Total incoming traffic at t.}$$

Suppose ***l*** and ***m*** are over-utilized links at time t. Means $iU_l < U_l$ and $iU_m < U_m$. As the sum of all iU's and U's must be same, it implies there must exists at least one (under-utilized) link n at time t such that $|(iU_l - U_l) + (iU_m - U_m)| = (iU_n - U_n)$ Back

Network Model

- Requirement: Outgoing and incoming traffic follow the same edge router. NAT is the most practical solution for inbound traffic engineering.
- Selective Advertisement



- Waste 50% IP Address; May require changes to the node s/w.

Network Model – contd.

- Selective sub-prefix advertisement – 17.1/26, 17.1.0.64/26, 16.1/26 etc.
- Once the egress route of a user is changed, withdraw the route from old link and announce on the new link.
- Adv: (i) No wastage of IP addresses; (ii) All traffic including ongoing traffic sessions will start following the new route.

Network Model – contd.

- Disadvantages: (i) Many ISPs filter out small IP prefixes. However, the use of sub-prefixes is prevalent in practice [Cisco ISP essentials].
- (ii) Frequent route changes may cause convergence problem. However, our solution requires that once a user is moved, it is not selected for re-assignment for the next few periods. Moreover tools are available to predict traffic flow due to route changes.