A Scalable Framework for Segment Routing in Service Provider Networks: The Omnipresent Ethernet Approach

Sarvesh Bidkar, Ashwin Gumaste and Arun Somani
Gigabit Network Laboratory, Indian Institute of Technology, Bombay, Mumbai, India 400076 Dept. Electrical and Computer Engineering, Iowa State University, Ames 50011
Email: sarvesh@cse.iitb.ac.in and ashwing@ieee.org arun@iastate.edu

2014 IEEE 15th International Conference on High Performance Switching and Routing (HPSR)

Problem in Service Provider Networks

Challenge 1 : Explosion of hand-held **devices** and **next generation data-services**

- Legacy transport networks that were built using SONET/SDH **circuit-switched paradigms**
- **Packet-based networking solution** that would be flexible, yet provide for **service-oriented communication**

Challenge 2: Operations Administration Maintenance and Provisioning (OAM&P)

- IP at the network layer and MPLS as a transport solution have been adopted
- Despite the ability to differentiate flows and provision services, the IP/MPLS layer falls short to provide full OAM&P

Background

Segment Routing SR

- IETF use source routing as an ebabler towards simplifing networks
- Segment routing facilitates implementation of source routing within a closed domain by **encoding end-to-end route information within the packet header** as a sequence of segments.

Carrier Ethernet (CE)

- CE is a service-oriented version of Ethernet with **no spanning tree protocol, no MAC learning**
- Forwarding is based exclusively on control plane defined **identifiers**
  - provider backbone bridged traffic-engineering (IEEE)
  - MPLS-TP (MPLS-transport profile) (IETF)

Introduction

- The focus of this paper is to present a pragmatic method to implement segment routing based on CE advances
- **Omnipresent Ethernet (OE)** was shown in as a **source routing+binary routing** solution for **software defined networks (SDNs)**
  - **OE as an SR framework serves providers effectively**
  - **source routing :** multi-parameter service provisioning is possible
  - **binary routing :** the lookup at a node is reduced making OE carrier-class with deterministic latency
- H-SR : improve scalability ofsegment routed networks based on a hierarchical segment routing (H-SR) framework

SR Concept

- Reduces the per-flow state maintenance, route computation complexity and is service oriented
- Nodes are labeled with globally unique identifiers (called node-segment IDs)
- Links are labeled with locally unique identifiers (called adjacency-segment IDs)

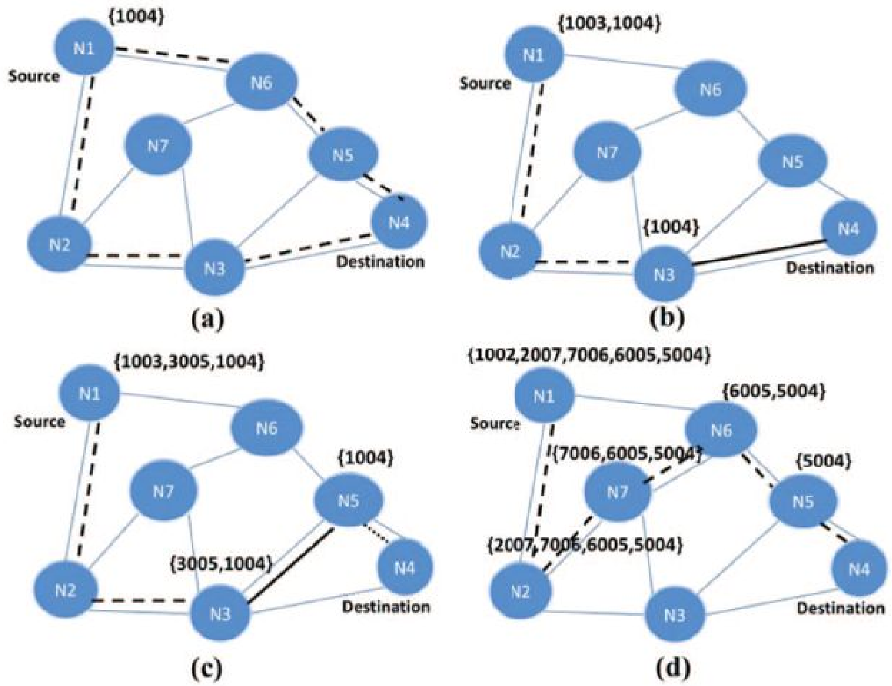
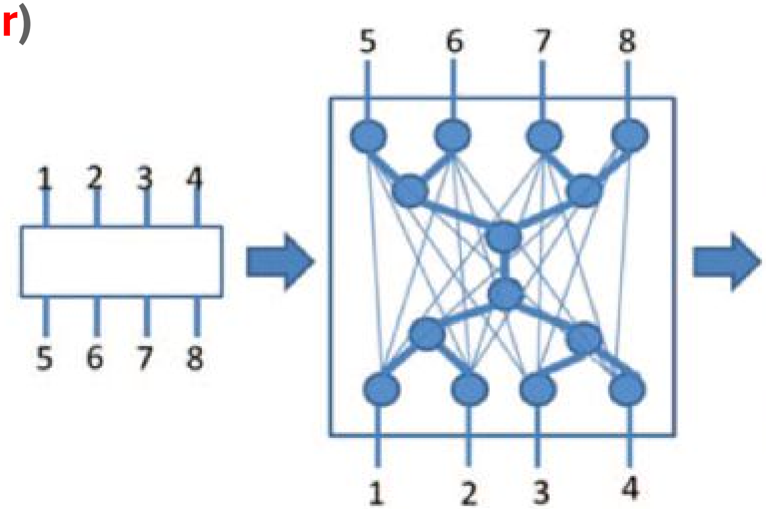


Fig. 1. Different routing schemes possible using Segment Routing

OE approach SR

- OE-domain specific identifier facilitates fast transport through the OE domain
- A **logical binary graph is created at each node** such that each physical port of the node is a leaf node of the logical binary graph.
- Internal Route Map (IRM) is computed for each node using binary routing
- Entire network becomes an **interconnection of logical binary graphs** and we can reach every node from every other node using **binary routing** and **source routing (long header)**



Internal Route Map (Port 1)	
Output Port	Binary Route
2	0
3	100
4	101
...	...
8	1100

Fig. 2. Logical binary tree representation of an 8-port switch and example Internal Route Map.

Route-label

- We use MPLS-TP labels stacked in an Ethernet frame for encoding the binary route to the destination
- 16 MSBs of the route-label to encode the binary route and the remaining 4 bits as the start-bit pointer
- Node further fetches the 2log2D bits to determine the out port

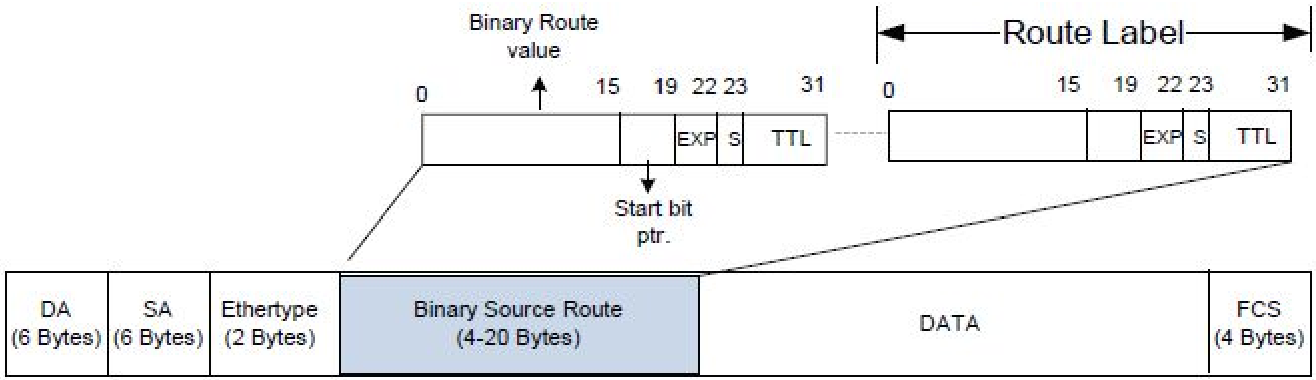
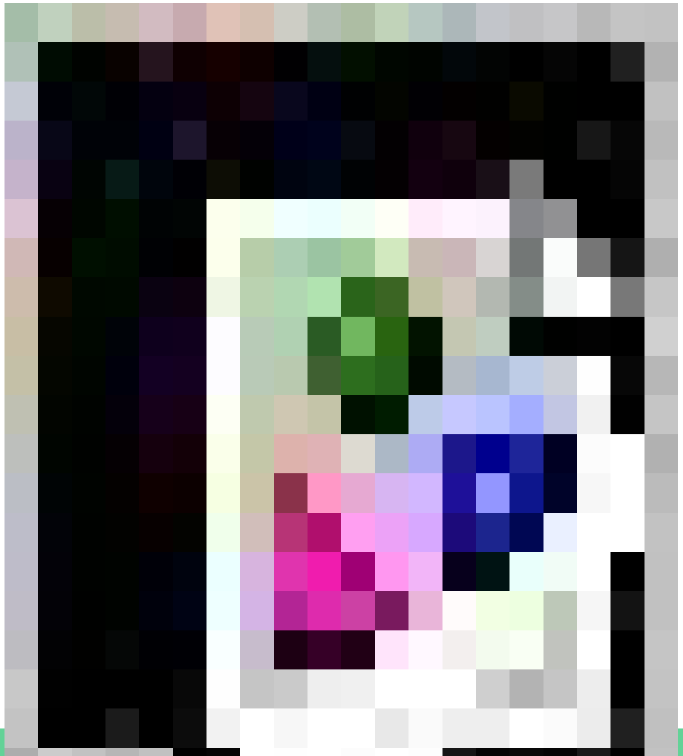


Fig. 3. Protocol Data Unit of OE.

Scalability challenges of Omnipresent Ethernet

- The binary routing and source routing principles of OE facilitate **low-latency switching** by **eliminating lookup operations** at intermediate nodes
- segment mapping table (SMT) maintains the mapping of service identifiers to binary routes
- With network size, the average number of hops between a source and destination increases.
- hierarchical segment routing (H-SR)



Hierarchical segment routing (H-SR)

- neighborhood (NH) of a swap node as the set of nodes that are **within connection proximity** of a swap node and may have similar address prefixes
- For simplicity, the NH of a swap node is analogous to an **Autonomous System** (AS)
- Internal Segment Mapping Table (I- SMT)
  - stores the routing information of all the nodes within a NH
- External Segment Mapping Table (E-SMT)
  - stores the routing information of all the other swap nodes in the network at a swap-node

Communication

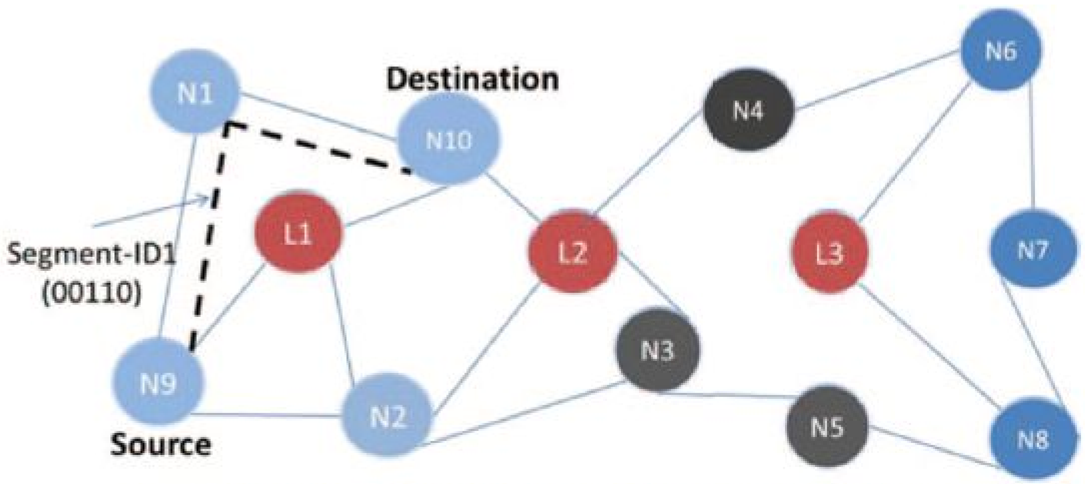
- Intra-NH Communication
- 

Fig. 4. H-SR policy for intra-NH communication.

Communication

- Intra-NH Communication

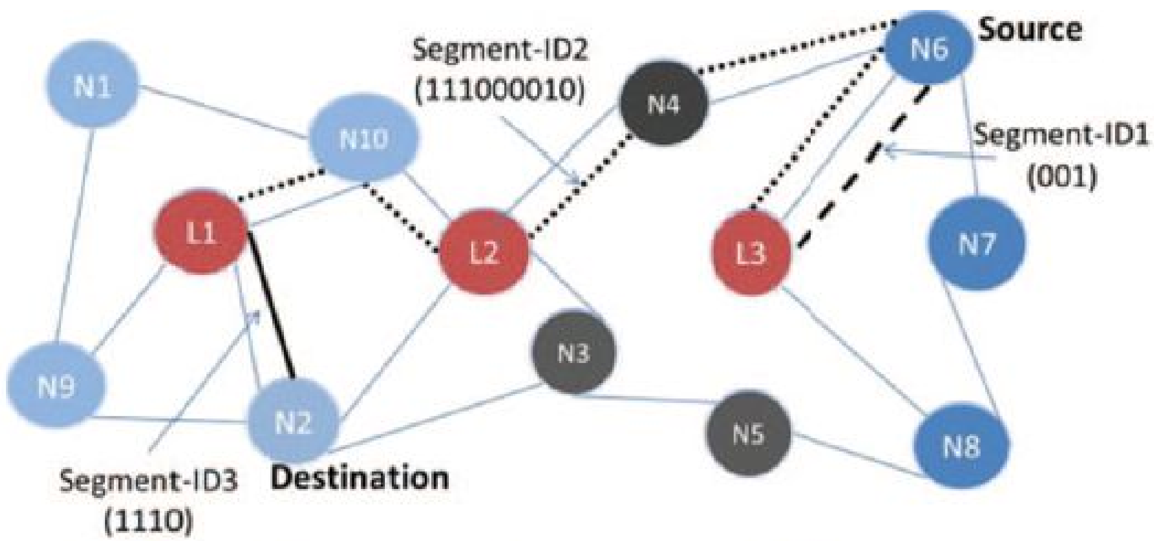


Fig. 5. H-SR routing policy for inter-NH communication.

Swap node selection techniques

- H-SR routing scheme for the inter-NH communication can potentially obtain a longer path than shortest path routing.
- To evaluate routing schemes a well- known metric called stretch is defined as:

$$\text{stretch} = \frac{\text{routing scheme (H−SR) path distance}}{\text{shortest path distance}}$$

- minimize the stretch of each path and hence the aspects related to formation of the NH

Swap node selection scheme

- Our swap node selectioon schemes are based on the **centrality indices of nodes** in a graph
- Three types of centrality indexs are defined:
  - Degree Centrality: Degree centrality index is the normalized degree of a node in a network

$$C_D(v) = D_v / \sum D_j$$

  - Betweenness Centrality: Betweenness centrality index is proportional to **the number of shortest paths in the network that pass through a particular node**

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

  - Closeness Centrality: Closeness centrality index is a measure of how close a node is to all other nodes in the n twork

Testbed

- A test-bed is built using OE supporting routers designed by us
- 22 nodes and 23 fiber links (RailTel Cop. India)
-

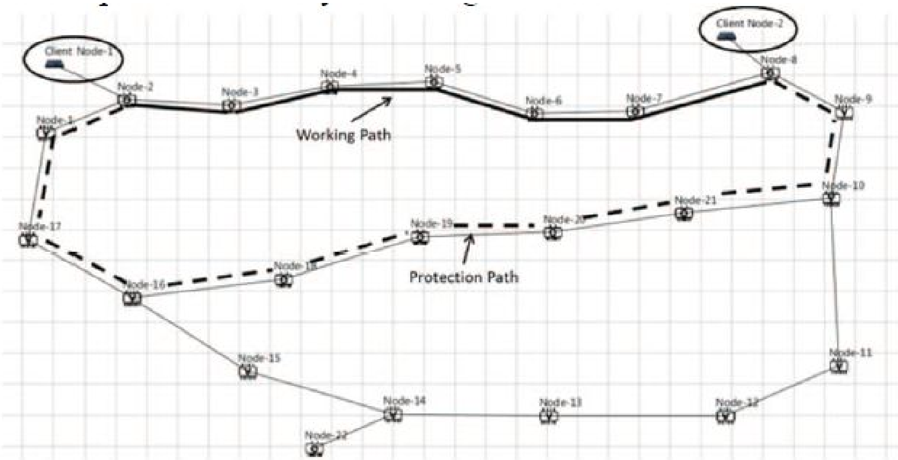


Fig. 6. Snapshot of the network topology of the test-bed.

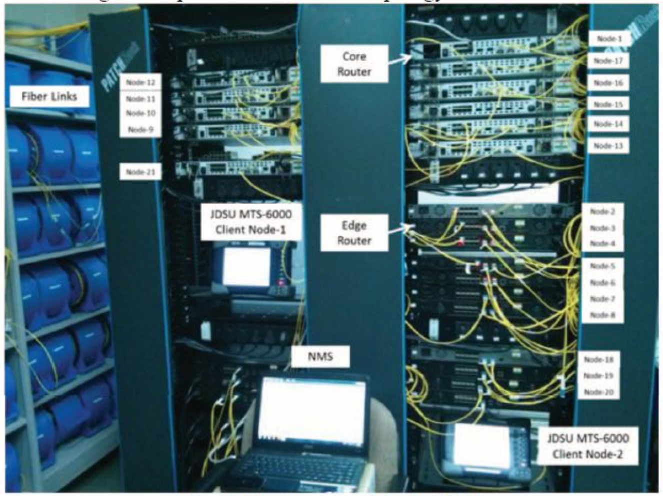


Fig. 7. Photograph of the test-bed.

Results

- throughput remains 100% for loads up to 90% and then reduces to around 90-91% for 100% load.
- This change in throughput is due to the **OE encapsulation** of the native Ethernet packet

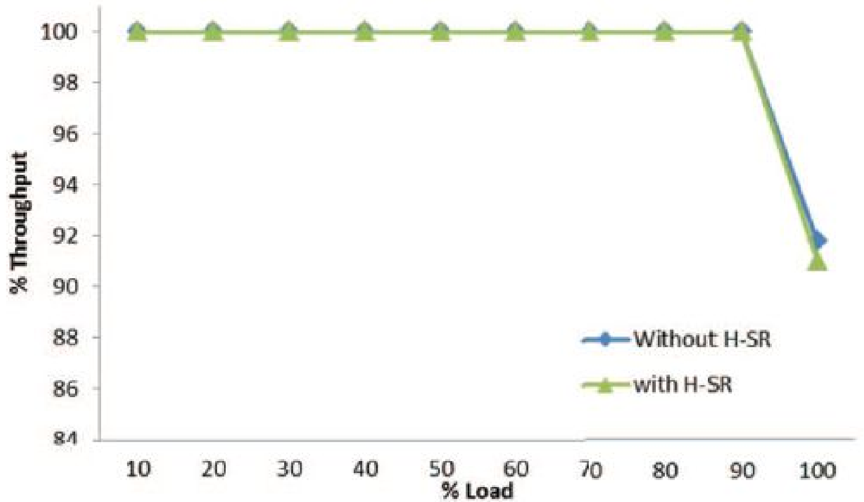


Fig. 8. Throughput of 1Gbps service for different load values.

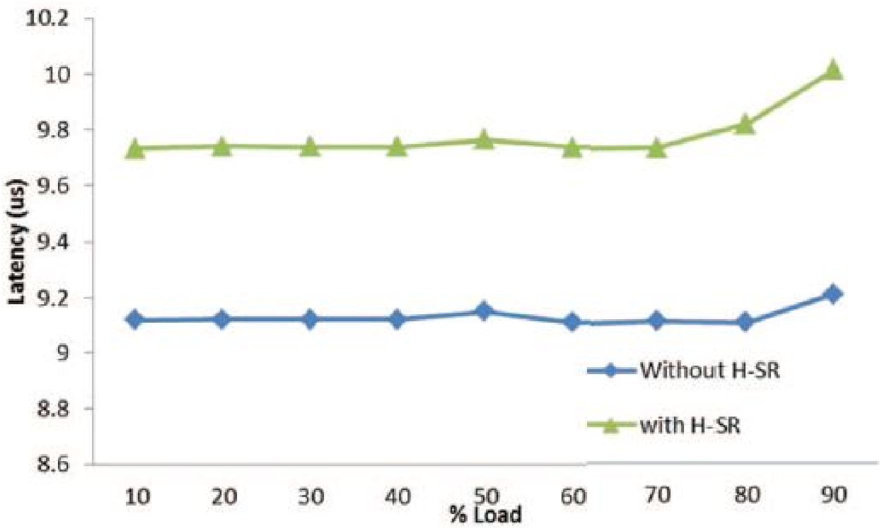


Fig. 9. Latency of 1Gbps service for different load values.

- latency is almost constant for load up to 70%

Simulation setup and results

- Network sizes with nodes ranging from 1000-10000 in the increments of 1000
- Highest Degree 1-Swap” in all subsequent graphs represents key results from compact routing which prefers highest degree nodes as swap node
- Average stretch values for betweenness centrality based swap node selection scheme is close to 1

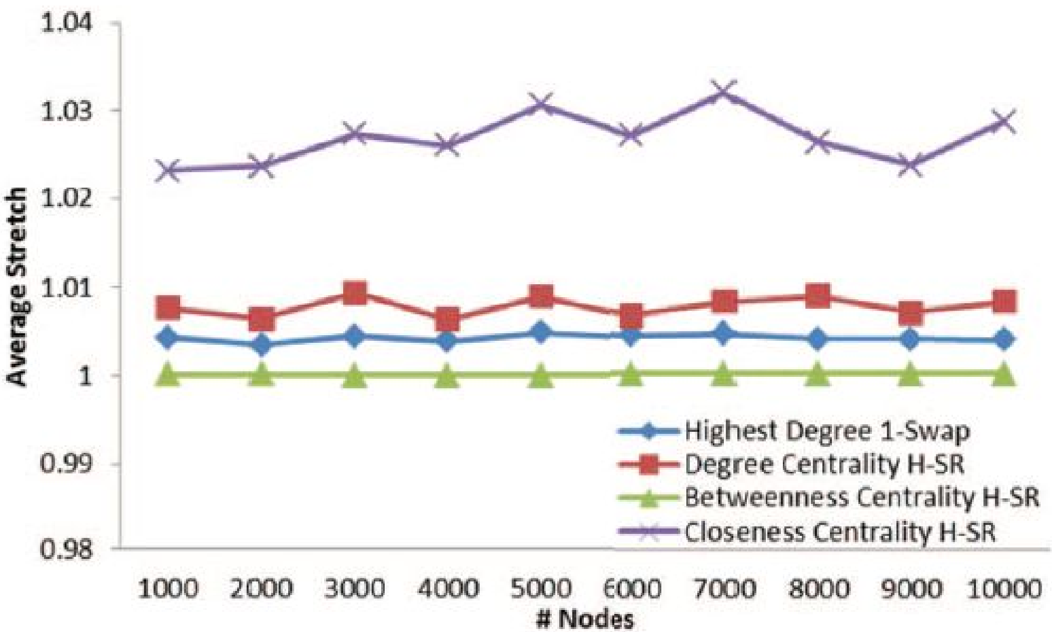


Fig. 10. Average stretch.

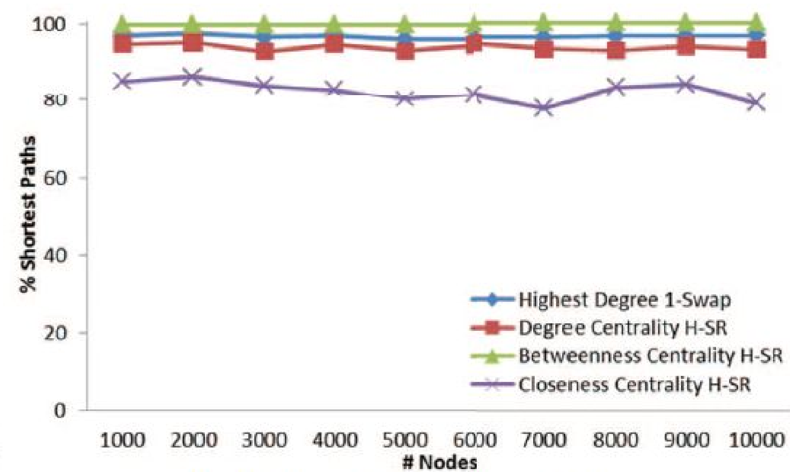


Fig. 12. Percentage of shortest paths.

number of routing table entries

- There is no comparable difference in the schemes for number of routing table entries
  - the same neighborhood formation approach for each scheme

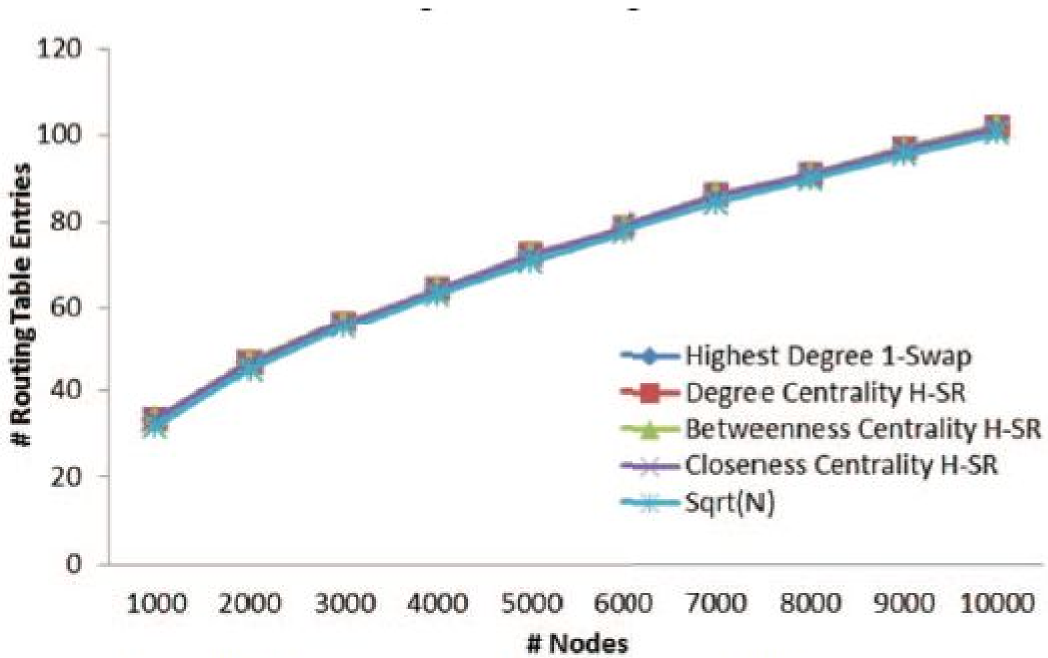


Fig. 11. Average number of routing table entries.

- propose a new metric called “Latency-Stretch” (LS) which is the product of **average end-to-end latency** and the **average path stretch** for a routing scheme
- In service oriented environments, **low-latency paths with deterministic latency but more hops are favorable** than the high-latency shortest paths with probabilistic latency

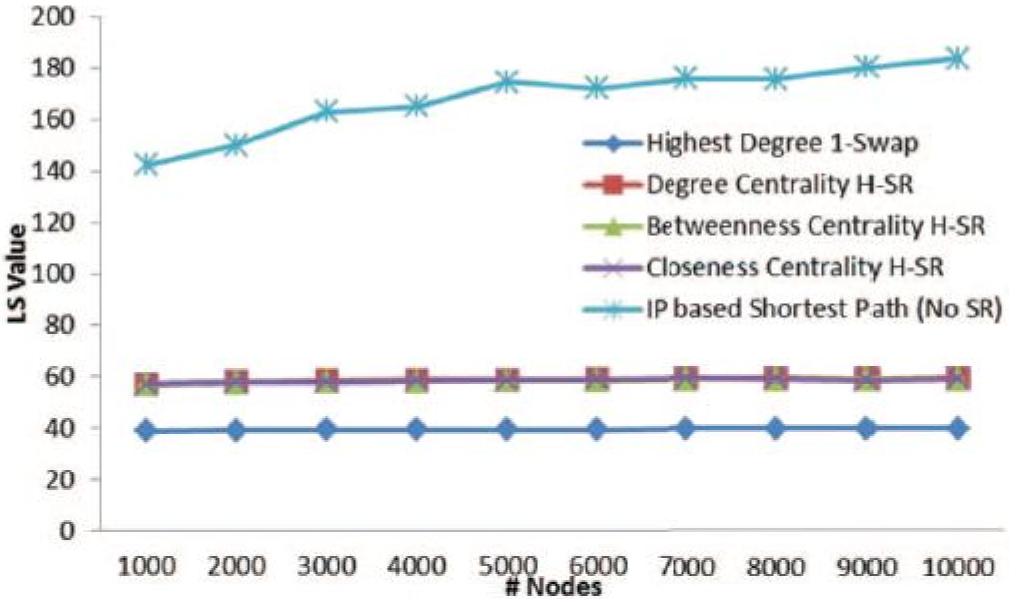


Fig. 13. Latency Stretch value as a function of network size.

label size

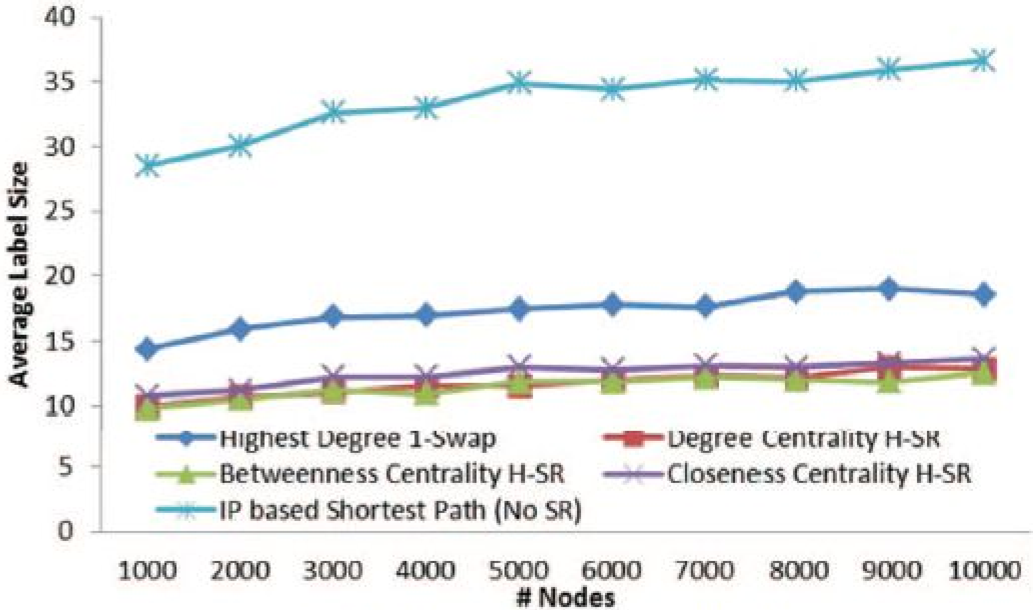


Fig. 14. Label size in the header.

CONCLUSION

- Integrated SR concepts with our recently proposed paradigm of Omnipresent Ethernet within the realm of Carrier Ethernet
- Allows providers to implement a packet-based carrier-class core network
- hierarchical-SR , three techniques for swap nodes, selection