Real-Time Reconfiguration of MPLS/WDM Networks

Mark Shayman

LTS/UMIACS Review

February 15, 2005

Contributors: P. Fard, R. La, Y. Xin, S. Marcus, M. Shayman

Goals

- Develop an integrated management and control framework for MPLS/WDM networks that
  - Reconfigures both optical lightpaths and MPLS label switched paths in real-time to accommodate changes in traffic demands
  - Adapts proactively to both deterministic (time-of-day) and random traffic variations
  - Minimizes disruption to existing traffic due to adaptation
  - Performs well when network is congested

Outline of Approach

- Control at multiple timescales
- Real traffic data
- Incremental reconfiguration using branch exchanges
- Formulation of Markov decision process
- Approximate solution of MDP using rollout

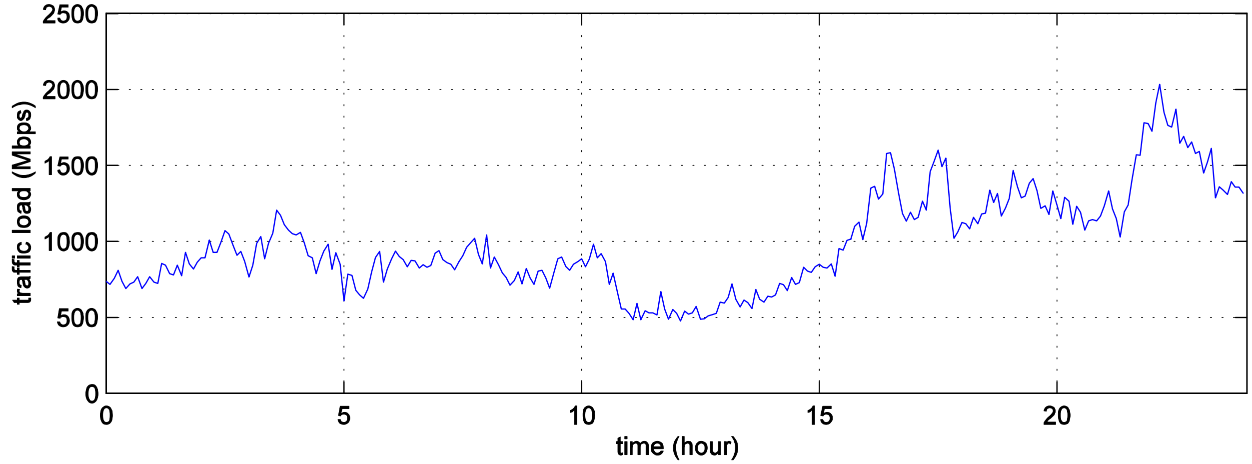
Multiple Timescales

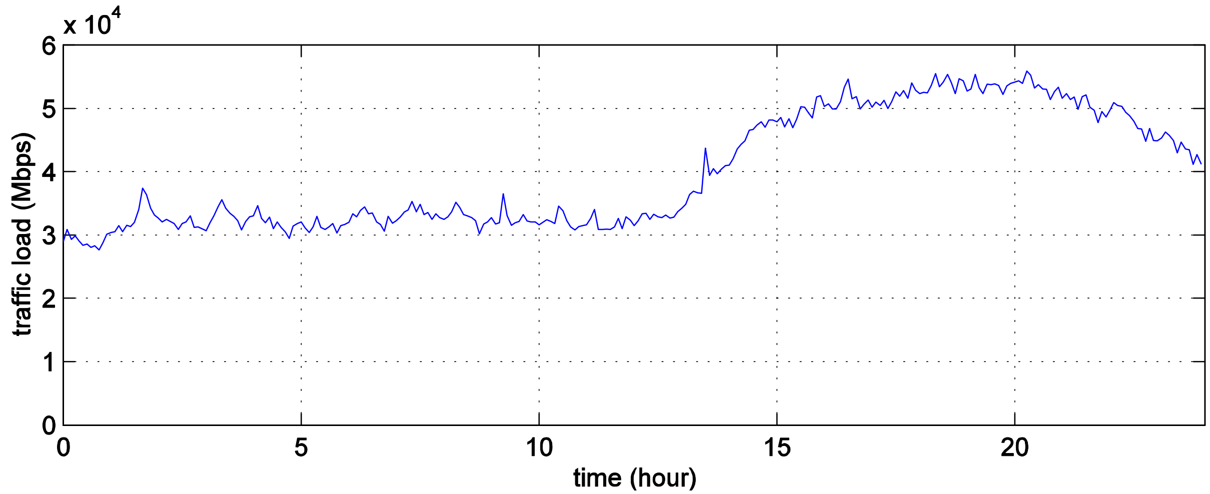
- Slow timescale: reconfiguration of logical topology (lightpaths) 2.5 minutes
- Moderate timescale: reconfiguration of MPLS label switched paths (LSPs): 30 seconds
- Fast timescale: mapping of arriving packets onto LSPs: 1 second

Traffic Model

- Based on real traffic from Abilene network.
- The Source-Destination traffic was collected for 6 weeks from this network.
  - Averaged 5 weeks of data to use as daily trend.
    - Used in conjunction with current traffic load to predict future demand
  - Used 6th week for the simulations.
- Fluid model is assumed.

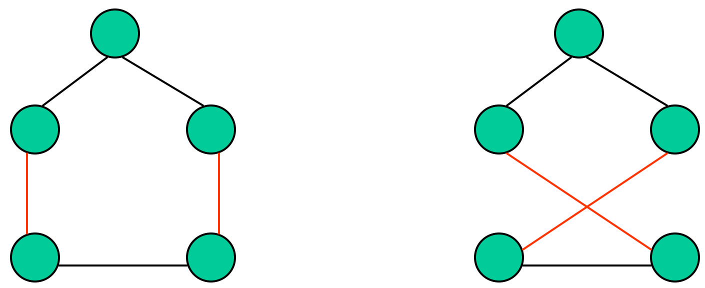
Sample traffic for a S-D pair (top) Total traffic (bottom)





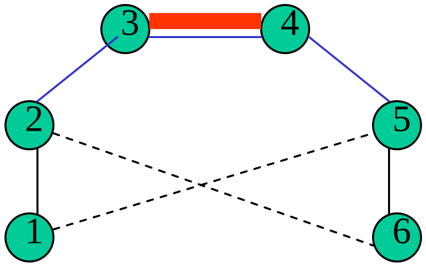
Branch Exchanges

- Double branch exchanges permit new lightpaths to be formed when there are no free router interfaces
  - Two lightpaths are torn down and two are created
- Incremental topology changes minimize disruption of existing traffic



Admissible Branch Exchanges

- Reduce search space by restricting admissible branch exchanges to those that have realistic chance of relieving congestion
- Say that a lightpath is congested if its utilization exceeds 80% (configurable) of capacity
- Say that a source-destination pair ij contributes to a congested lightpath if its traffic contributes load corresponding to more than 5% (configurable) of lightpath capacity
- Branch exchange g is admissible if at least one of the lightpaths it creates directly connects a source-destination pair that contributes to a congested lightpath



S-D pair 1,5 contributes to congested lightpath 3-4. BE creates lightpaths 1-5 and 2-6 offloading traffic from 3-4.

Why MDP Formulation?

- Sequence of decisions
  - At each slow timestep, need to make decision as to which branch exchange (or none) to perform
- Decisions have longterm effect
  - Current action determines new configuration, which determines which configurations may be reached in future
- State of system evolves with uncertainty
  - State consists of lightpath configuration, MPLS configuration and traffic matrix.
  - Traffic matrix evolves as random process (assumed Markov) with unknown transition probabilities but known mean
- Performance of system can be quantified
  - Expressed as additive cost function that reflects link utilization and drops

Cost Function

- Cost is accumulated based on utilization of the lightpaths and amount of dropped traffic

$$J^f(r) = L(r) + \sum_i \left(\frac{100u_i(r)}{C}\right)^2$$

r is fast timescale step

$L(r)$ is loss ratio in network

$u_i(r)$ is traffic rate on lightpath i

C is lightpath capacity

- Cost for moderate and slow timescale is sum of costs for all fast timescale steps in corresponding moderate or slow time scale step

Number of LSPs per S-D pair

- This is a variable that could be configured.
- If set to 1, we have a shortest path policy. We set this variable to 1 to investigate the load balancing properties of slow time scale.
- If set to > 1, at the moderate and fast time scale we have load balancing.

Moderate Timescale Policy

- Reserve BW for the existing traffic in each LSP
- Estimate the traffic for each source-destination pair ij
  - Use daily trend and the traffic measurements from the last step to predict the traffic in the next step.
  - Add the measurements from the last step to the difference in the daily trend between the last step and next step.
- If both measured BW and predicted BW are less than provisioned BW, decrease provisioned BW
- If predicted BW exceeds provisioned BW, increase provisioned BW of LSPs for given S-D pair
  - Source-destination pairs are considered in order of decreasing difference in number of hops between their best two LSPs
  - LSPs are considered in order of increasing number of hops

Fast Timescale Policy

- Traffic is assigned to LSPs in order of increasing number of hops subject to the limit of provisioned BW
- When the number of LSPs per S-D pair is set to 1, fast time scale policy is reduced to a capacity check.
  - Traffic exceeding provisioned BW is dropped

Slow Timescale Heuristic Policy

- For each admissible branch exchange, perform the following hypothetical computation using the new topology
  - Determine the 3 minimum hop LSPs for each source-destination pair
  - If a previously existing LSP no longer exists, move traffic to the active LSPs
  - Reprovision the bandwidth of the active LSPs based on the estimated traffic
  - Based on the estimate of the traffic determine the expected cost over the next slow time step
- Choose the branch exchange that gives the minimum expected cost over the next slow time step

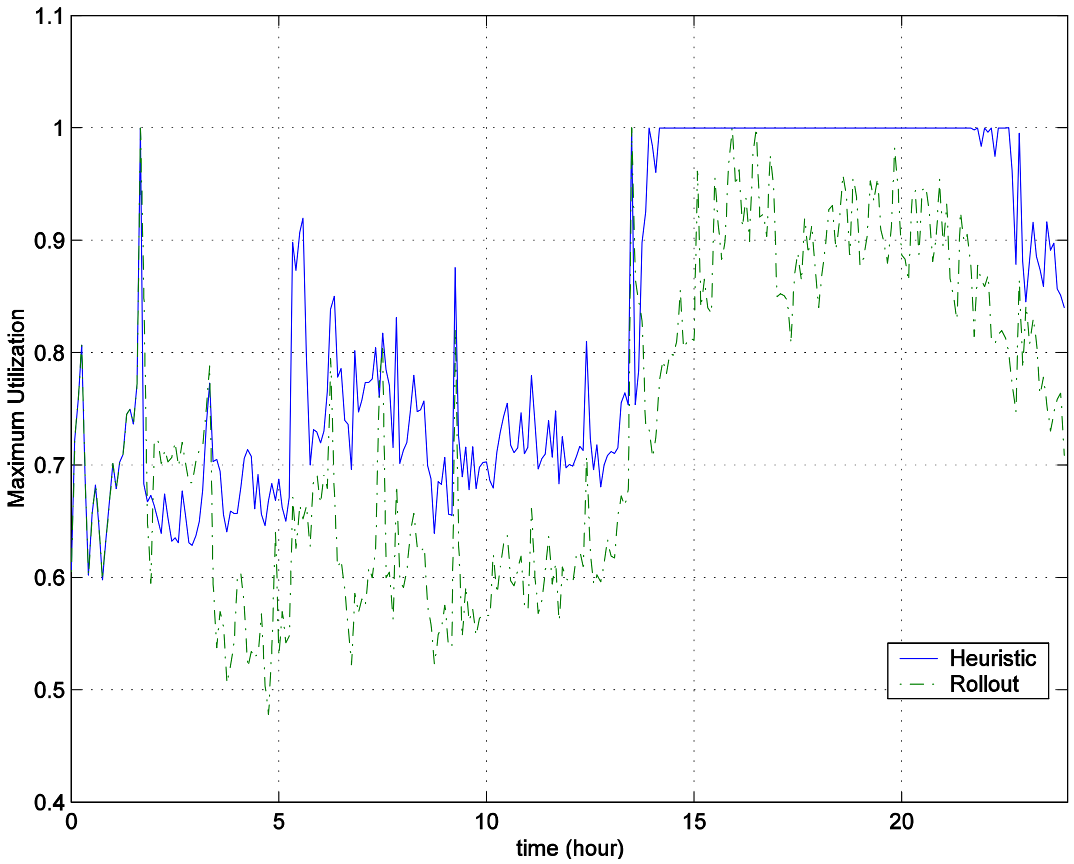
Concept of Rollout

- Starts with a heuristic policy $u = \pi(x)$ where u is the action in state x
- Creates an improved ‘rollout’ policy $u = \pi_r(x)$ as follows
  - For each possible action u in state x, evaluate the expected cost of taking u in x and following policy π starting in the next state x’.
  - Let $\pi_r(x)$ be the action that minimizes the expected cost.
- The action $\pi_r(x)$ can be computed online when state x is visited using simulation to estimate the expected cost for each possible action

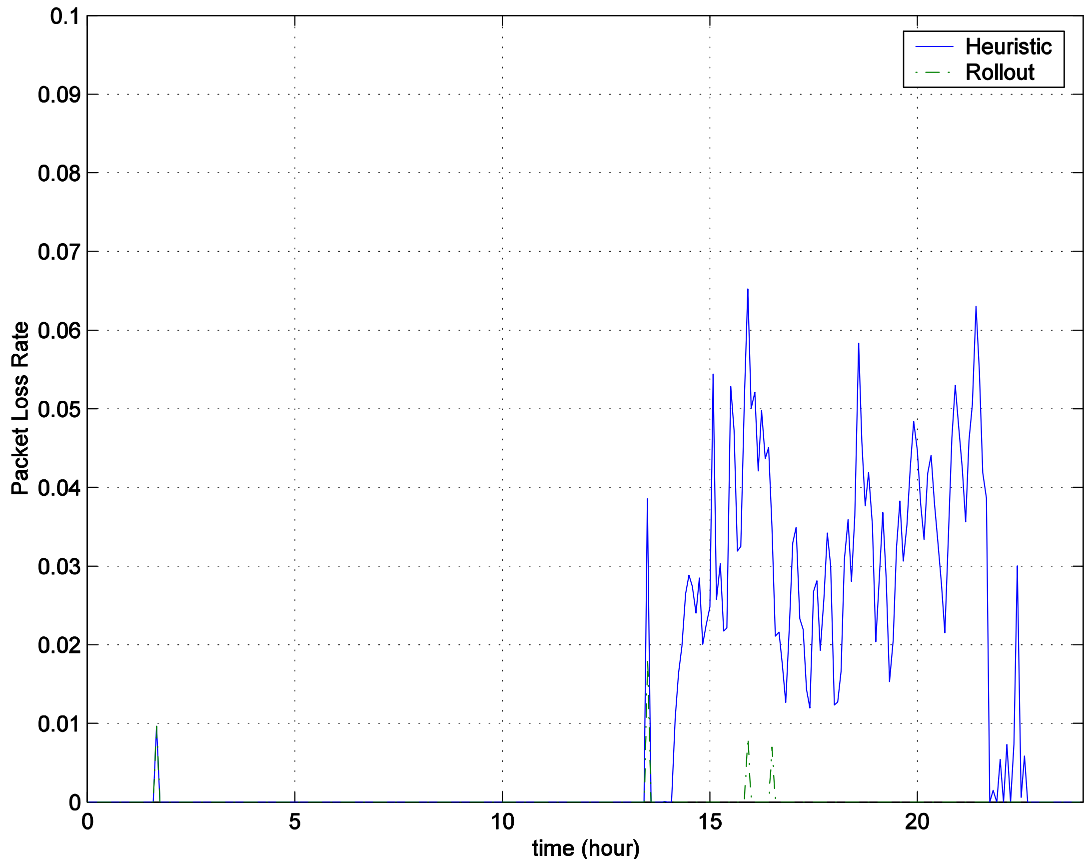
Rollout vs. Heuristic

- Heuristic is a greedy algorithm
  - Chooses branch exchange that is expected to give best performance during the next slow time step
  - Such a branch exchange may preclude reaching an advantageous topology in the future
- Rollout may choose branch exchange that is not optimal for the current time step but may lead to better performance over several steps
- Since the only information about future traffic is the average trend, rollout is performed using a single "average sample path" (certainty equivalence)

Maximum utilization for Heuristic vs. Rollout



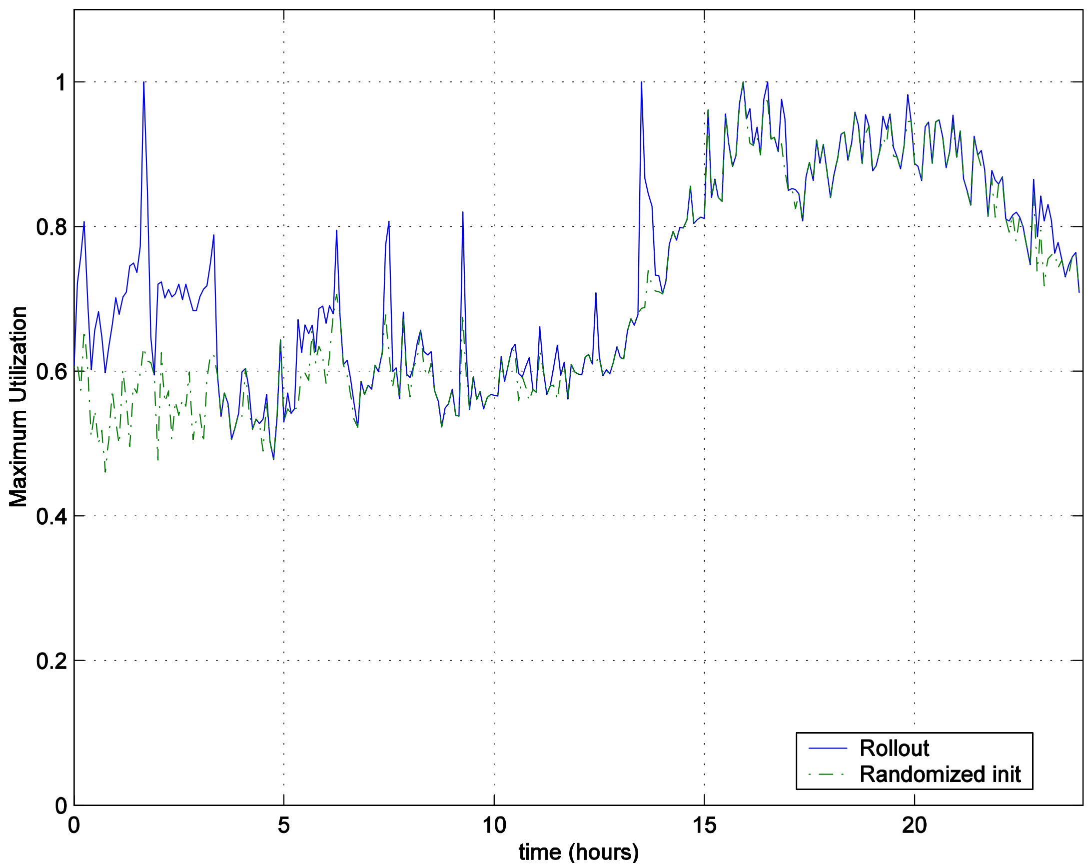
Packet loss ratio for Heuristic vs. Rollout



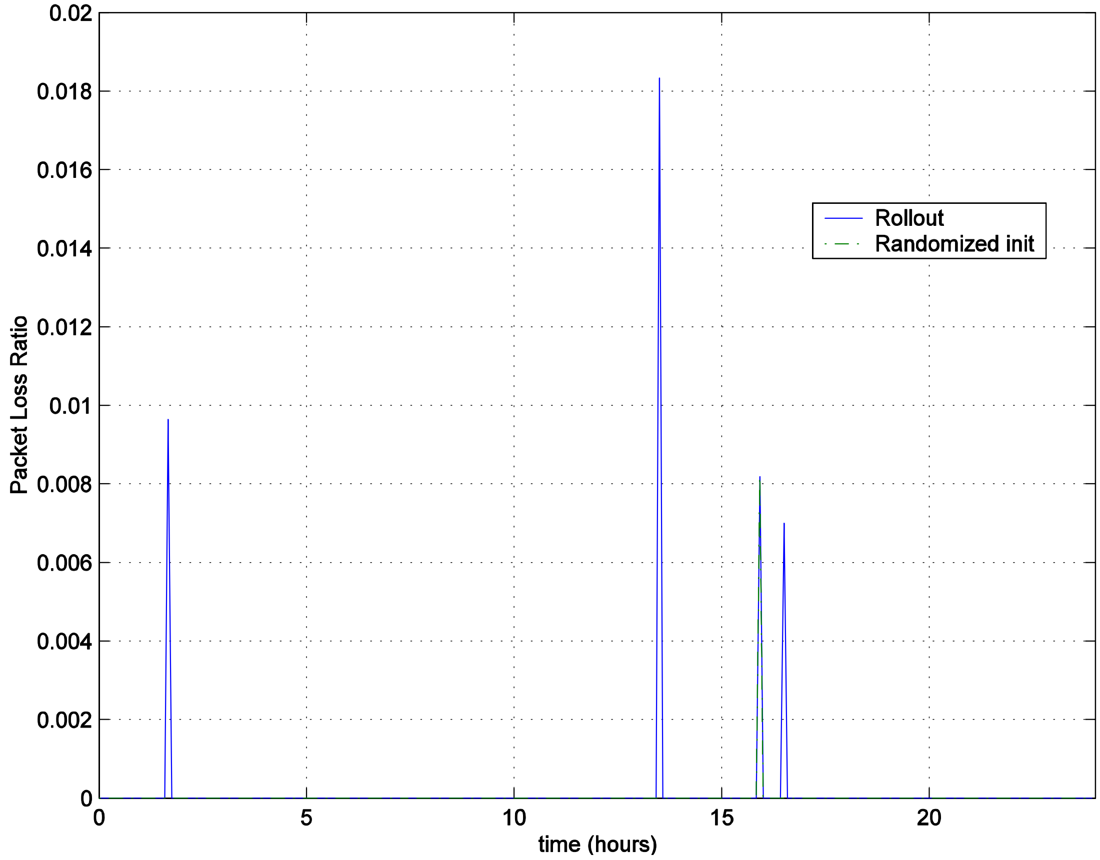
How Close to Optimal?

- Problem of optimizing topology for given traffic matrix is NP-complete
- Idea: for each traffic matrix, choose multiple random initial topologies and apply rollout action repeatedly to obtain sequence of branch exchanges until locally optimal topology is found.
  - Use best locally optimal topology as reference topology
  - Compare performance of rollout algorithm to performance of reference algorithm that chooses reference topology corresponding to current traffic matrix at each step
    - Reference algorithm is permitted to perform complete topology reconfiguration at each step

Rollout vs. Reference Algorithm



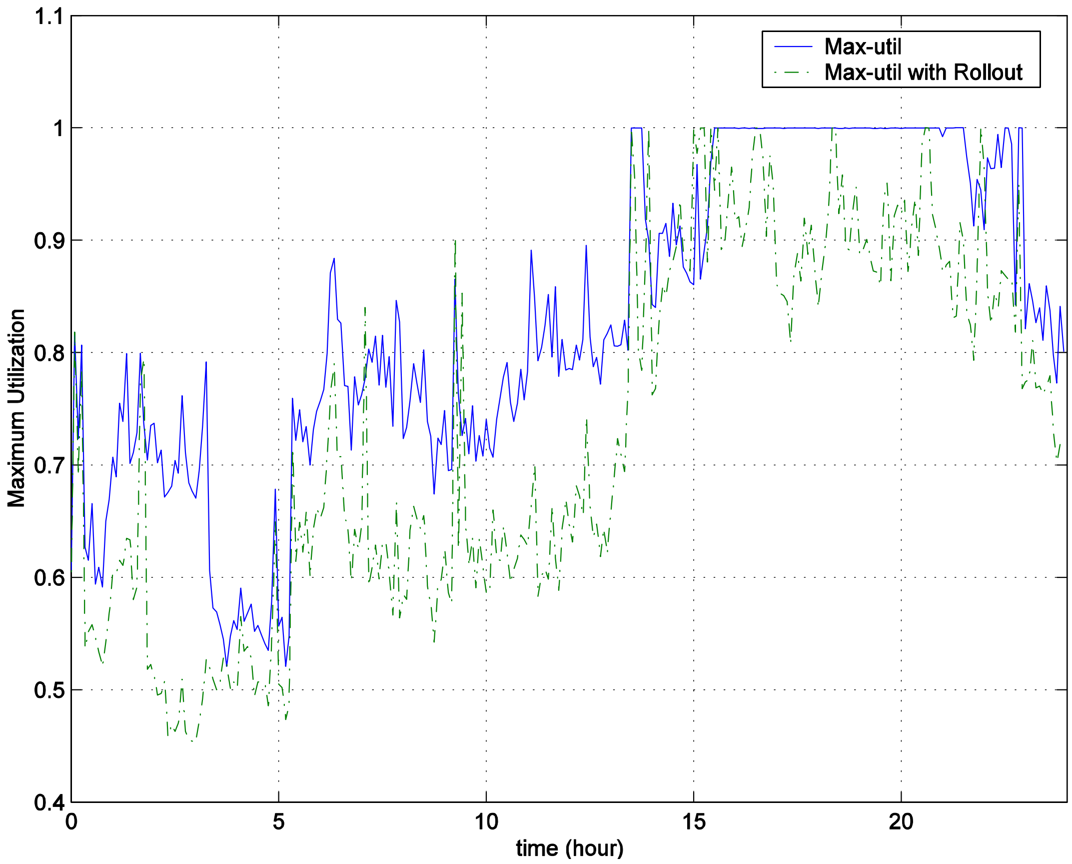
Packet loss ratio for Rollout vs. Reference Algorithm



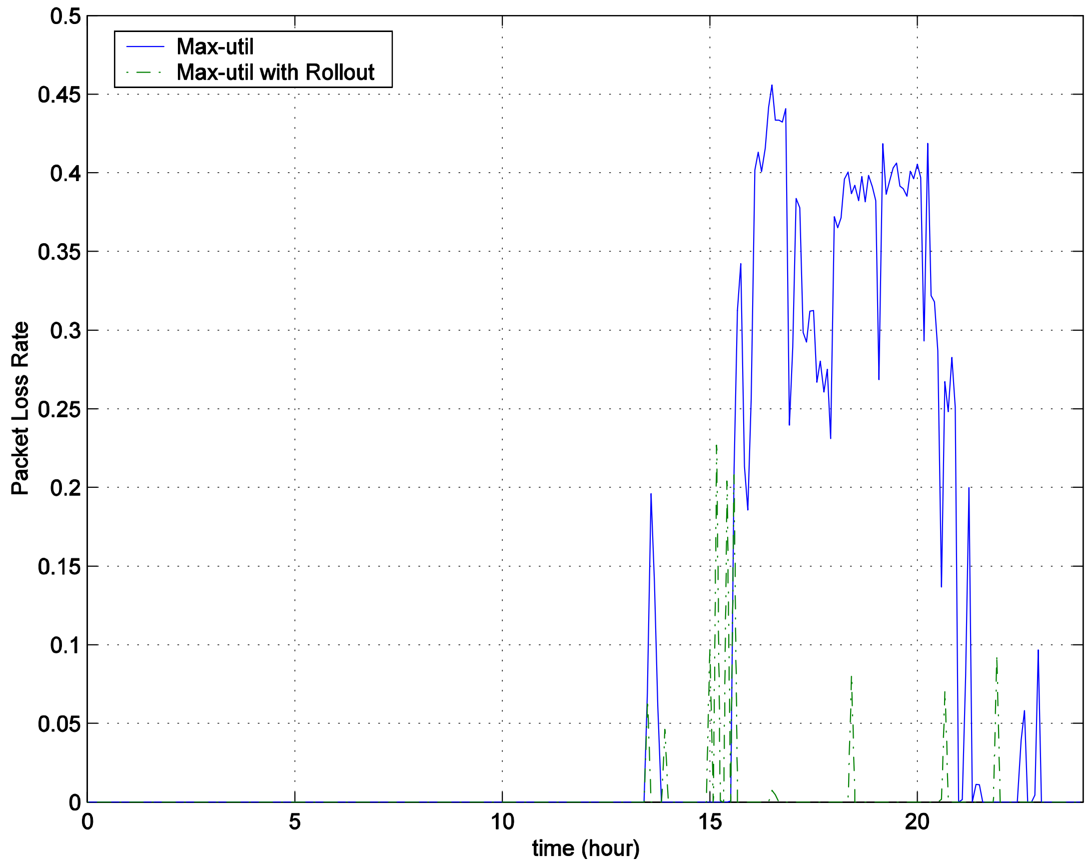
Using Rollout to Improve Modiano's Algorithm

- Modiano's algorithm
  - At each step, choose the branch exchange that minimizes the maximum link utilization
    - Minimum hop routing is used
- Used Modiano's algorithm as the heuristic in rollout

Max-util (Modiano's) algorithm vs. its rollout



Packet loss ratio for Max-util vs. its rollout



Comparing our Algorithm with Modiano's Algorithm and Its Rollout

- Our heuristic differs from Modiano's
  - Takes sum of link utilizations into account, rather than only max utilization
  - Takes drops into account
- Our rollout algo. significantly outperforms Modiano's algo. both in terms of max. link utilization and drops
- Both rollout algorithms have similar performance in terms of max. utilization, but ours performs much better in terms of drops

Proposed Plans

- Completed work does not model optical resources
  - Can be implemented using <u>overlay model</u>
    - All decisions can be made by MPLS service provider
    - Branch exchanges are requested from WDM service provider
- Future work
  - Add model of optical network: physical topology, wavelength converters
  - Develop control algorithm for <u>augmented model</u>
    - WDM SP informs MPLS SP of number of wavelengths available between each pair of routers
  - Develop control algorithm for <u>peer model</u>
    - Using integrated extended routing algorithm, MPLS is aware of the wavelengths available on each fiber and the physical path and wavelengths used on each lightpath