Locating network monitors: complexity, heuristics, and coverage

Kyoungwon Suh, Yang Guo, Jim Kurose, and Don Towsley

kwsuh, kurose,towsley@cs.umass.edu

Yang.Guo@mathworks.com

presented by Kyoungwon Suh

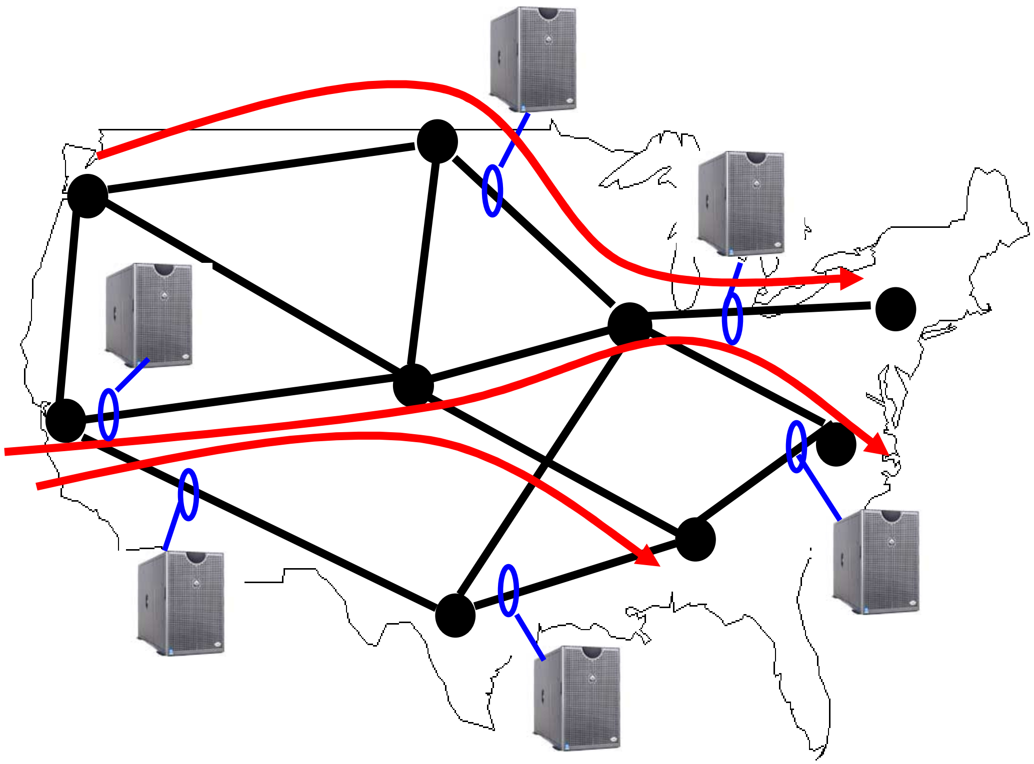
UMASS

University of Massachusetts
Department of Computer Science

Outline

- Motivation
- Related Work
- Monitoring cost and reward models
- Passive monitoring: problems and solutions
- Evaluating our heuristics
- Summary and future work

Challenges of distributed monitoring



- Need to observe large fraction of traffic
- Single link measurement point: NOT enough
- Monitoring cost versus monitoring reward

Related Work on monitor placement

- Number and location of active monitors
  - NP-hardness proof and approx. algorithm
    - Jamin (INFOCOM '00), Adler (NGC '01), Horton (IMC '03), Barford (IMW '01) Bejerano (INFOCOM '03), and Li (INFOCOM '03)
- Distributed passive monitors
  - Locations were chosen in ad-hoc basis
    - Sprint IPMon Project (PAM '01)
- Adaptive sampling strategies
  - Concerned with "single" passive monitor
    - INMON's sFlow, Choi (SIGMETRICS '02) and Duffield (IMW '02)

No previous work on location and sampling strategies for passive monitors

Monitoring cost and reward models

- Deployment cost:

$$C_D = \sum_{i \in L} f_i y_i$$

L: set of all links
f_i: monitor deployment cost at link *i*
y_i: 1 if a monitor is placed at link *i*

- Operating cost:

$$C_O = \sum_{i \in L} y_i c_i \sum_{j \in D} \rho_j m_{ij}$$

D: set of all flows
c_i: unit sampling cost of monitor *i*
ρ_j: traffic demand of flow *j*
m_{ij}: fraction of flow *j* sampled by monitor *i*

Monitoring cost and reward models - Continued

- Monitoring reward:
  - Benefit of traffic monitoring
  - e.g., monitored traffic demand (i.e., coverage)

$$C_M = \sum_{j \in D} u_j(M_j)$$

M_j: fraction of flow *j* sampled by monitors
$u_j(M_j)$: benefit gained by monitoring flow j; non-decreasing and concave utility function

Examples :

$$M_j = 1 - \prod_i (1 - m_{ij})$$

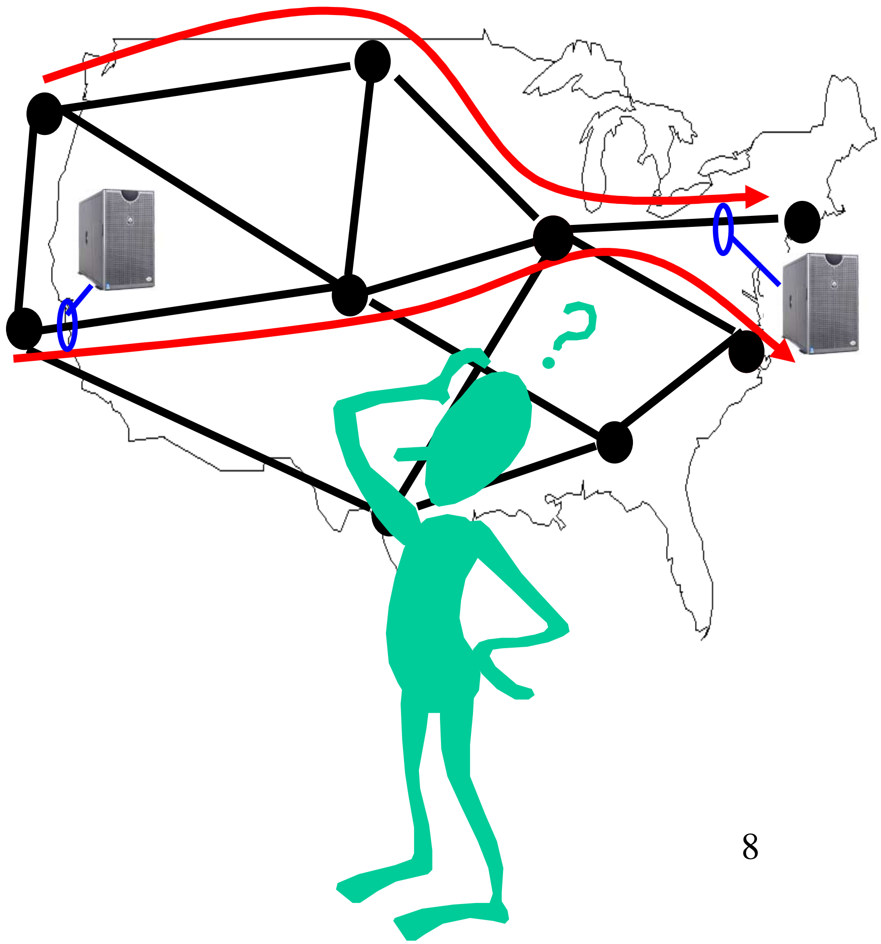
$$u_j(M_j) = \rho_j M_j \;\; ; \;\; u_j(M_j) = (1 - \exp(-5.0 * M_j))$$

Monitoring Problems

- Monitoring problems without sampling
  - BCMCP: Budget Constrained Maximum Coverage Problem
  - MDCP: Minimum Deployment Cost Problem
  - MDOCP: Minimum Deployment and Operating Cost Problem
- Monitoring problem with sampling
  - BCMCP-S: Budget Constrained Maximum Coverage Problem w/ Sampling

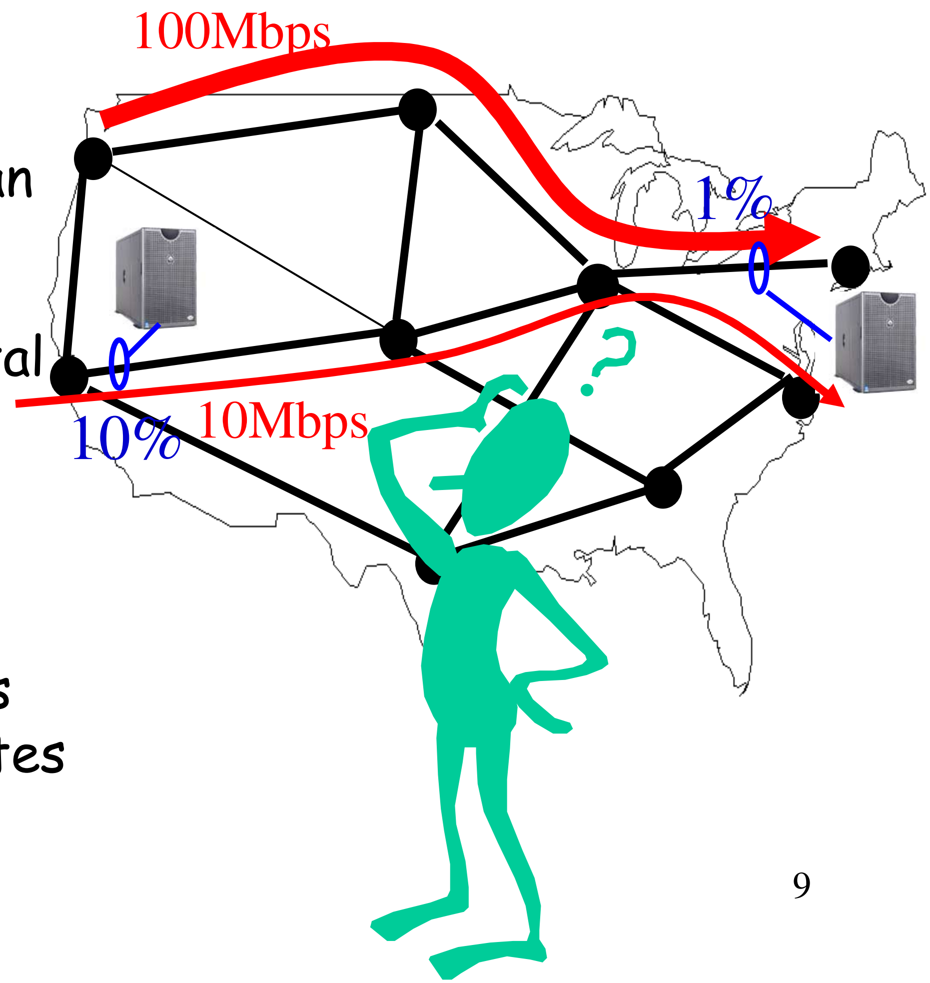
BCMCP - Example
(Budget constrained maximum coverage problem)

- Constraints:
  - Monitor captures all packets seen on link
  - Constraint: K monitors
- Objective:
  - Maximize total rate of monitored packets
- What to determine:
  - Optimal monitor locations



BCMCP-S – Example
(Budget Constrained Maximum Coverage Problem with Sampling)

- Constraints:
  - Placed monitor *samples* packets carried by link
  - Per-flow sampling rate can be chosen
  - Constraint: K monitors
  - Constraint: R Mbps in total
- Objective:
  - Maximize sum of flow sampling rates
- What to determine:
  - Optimal monitor locations
  - Optimal flow sampling rates



BCMCP: problem formulation and solution

- Objective
  - Maximize monitoring reward w/o violating total deployment cost constraint
- Complexity: NP-Hard
  - ILP formulation
  - Reduction from budgeted maximum coverage problem
- Proposed greedy heuristics
  - Greedy: choose links maximizing marginal gains per unit cost
  - Approximation ratio: 1-1/e

BCMCP-S: problem formulation and solution

- Objective
  - Maximize monitoring reward w/o violating deployment and operating cost constraints
- MINLP formulation
  - Difficult to obtain optimal solution directly
- Proposed algorithm
  - Greedy algorithm for location
  - Gradient projection method for sampling rates

Questions

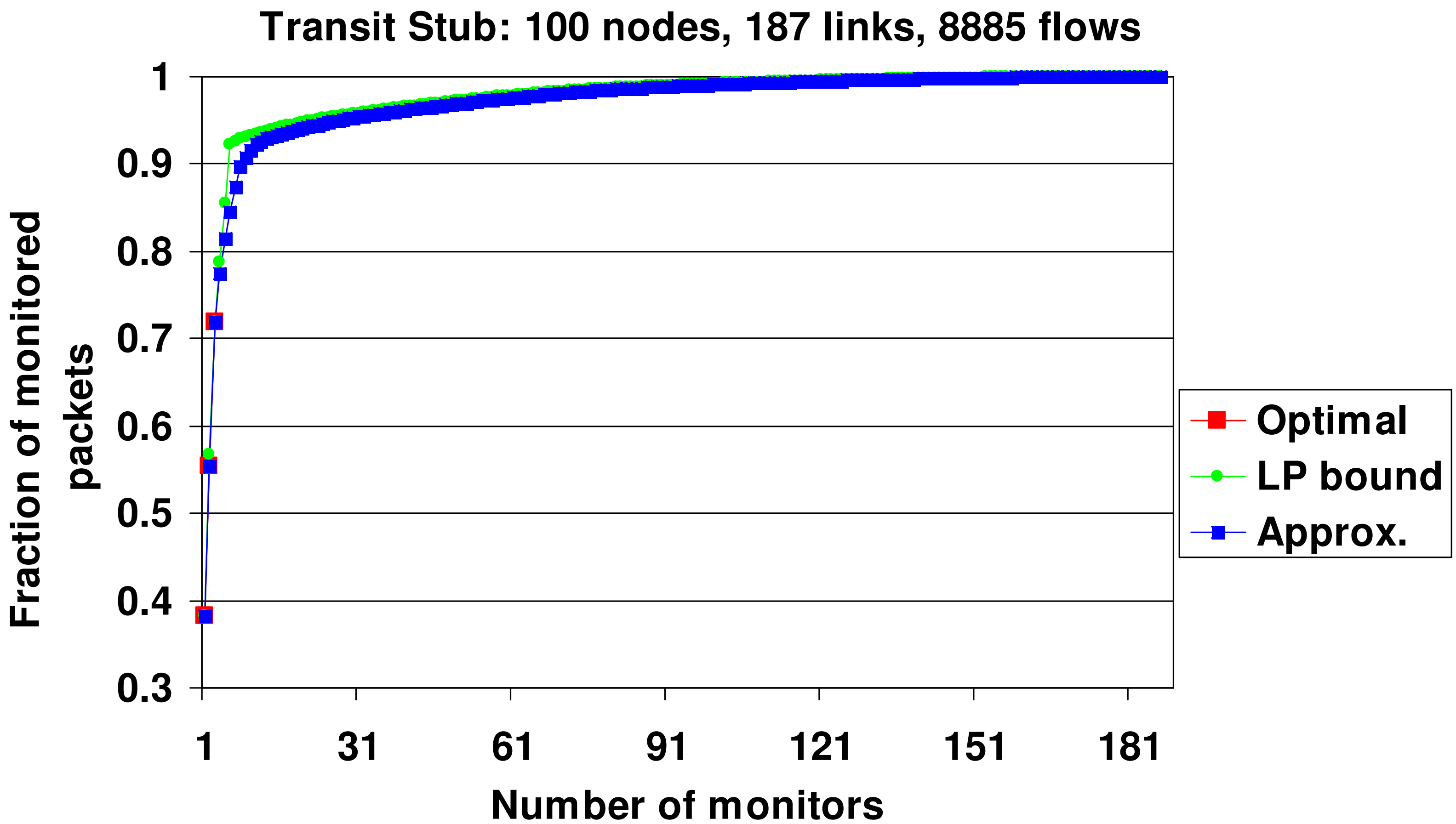
- How close are approximation solutions to optimal solutions?

- How many monitors are required to monitor “large” fraction of traffic?

Evaluating our heuristics

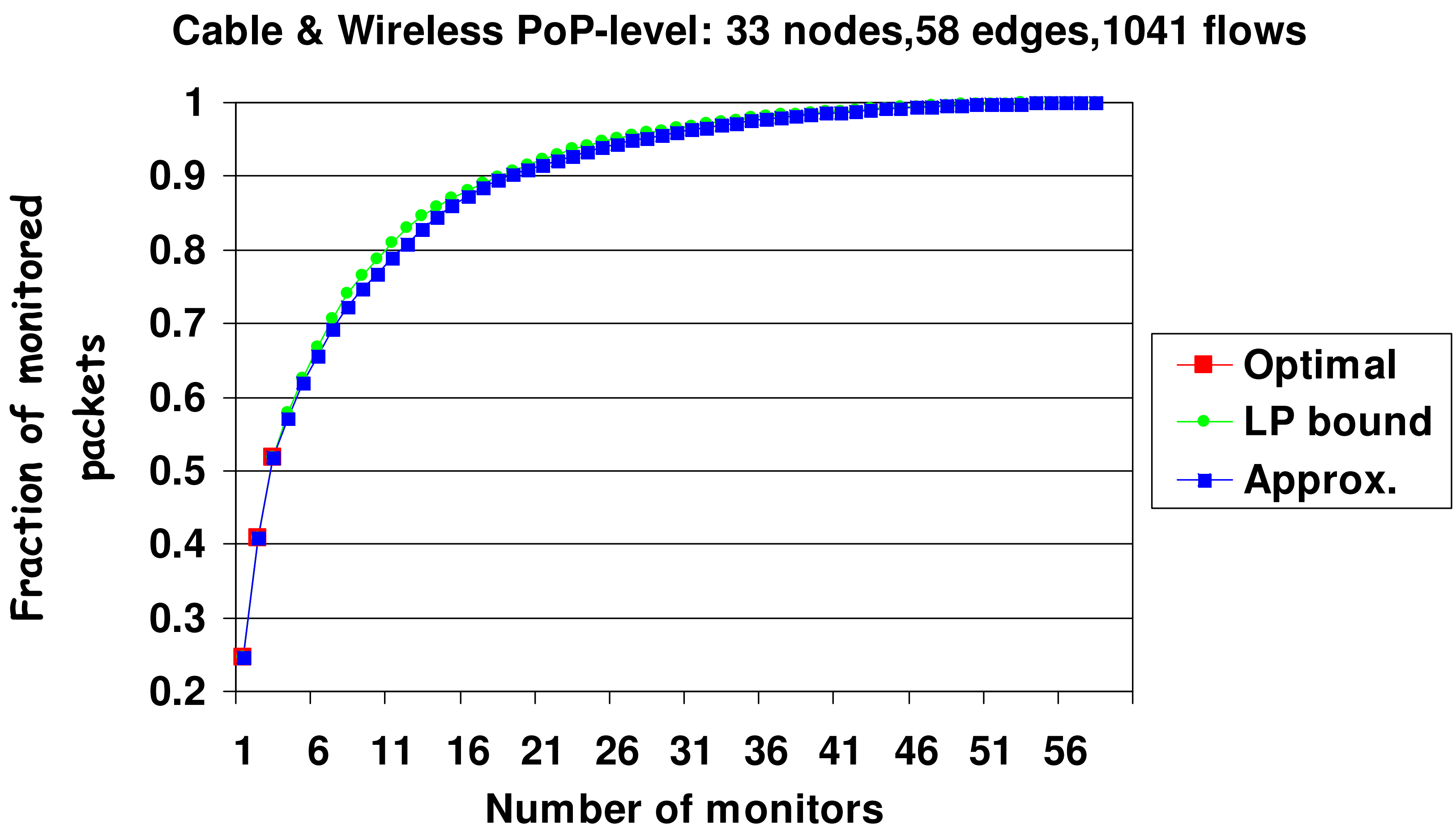
- Simulation settings:
  - Random and transit-stub topologies generated by GT-ITM (10 and 100 nodes)
  - Cable & Wireless PoP-level topology (33 nodes)
  - Node: abstraction of a router or a PoP
  - Utility function: linear or exponential
  - Cost functions: equal deployment cost, equal operating cost
  - Traffic matrices: Fortz and Thorup's traffic generation formula (INFOCOM '00)

BCMCP



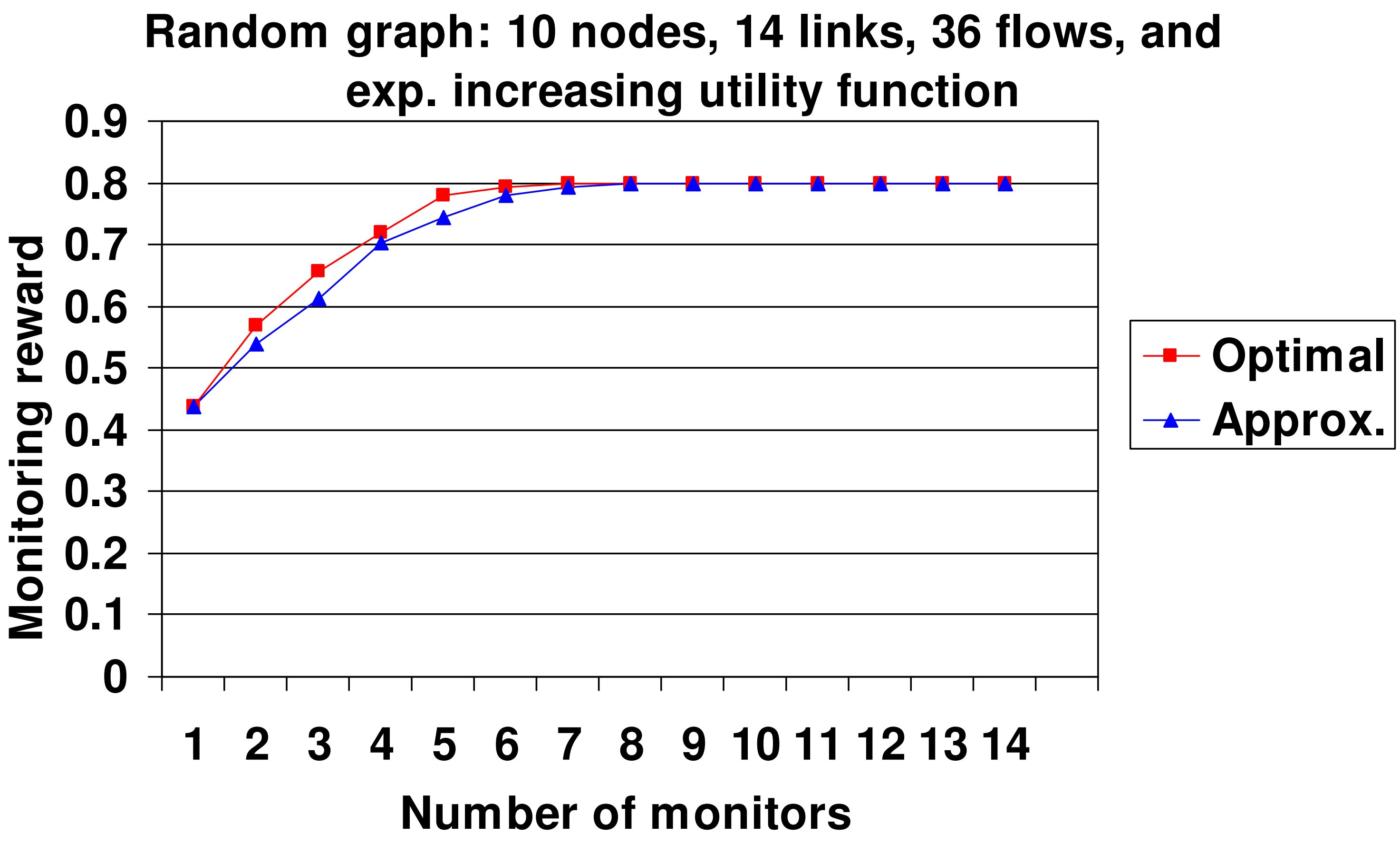
8 monitors (4% of possible locations) can monitor more than 90% of packets

BCMCP



20 monitors (34% of possible locations) can monitor 90% of packets

BCMCP-S



Marginal increase in monitoring reward is decreasing

Questions/Answers

- How close are approximation solutions to optimal solutions?
  - Much better than theoretical worst-case bounds
  - Often, same as optimal solutions
  - Why: shortest-path routing, all-pair traffic demands, and network topology
- How many monitors are required to monitor "large" fraction of traffic?
  - Marginal increase in monitoring reward is decreasing
  - Only a few monitors are enough!

Summary and Future work

- Proposed novel monitor location models
- Proposed monitor placement and operating strategies
- Experiment results show that:
  - Very good approximation ratio can be achieved
  - Only a small fraction of links is often enough
- Extension of our models
  - Routing changes caused by link/router failures
  - Stochastic traffic demand models

The End