Optimizing Capacity-Heterogeneous Unstructured P2P Networks for Random-Walk Traffic

Chandan Rama Reddy
Microsoft

Joint work with Derek Leonard and Dmitri Loguinov

Internet Research Lab
Department of Computer Science and Engineering
Texas A&M University, College Station, TX 77843

Sep 09, 2009

Computer Science, Texas A&M University

Agenda

- Introduction
  - Terminology, related work, and motivation
- Proposed System
- Proposed Metrics
- Evaluation
- Wrap-up

Introduction – Key Components

- Overlay topology
  - Determined by the neighbor selection policy
  - Influences the distribution of traffic among nodes
- Search methodology
  - Defines how queries are propagated in the overlay
  - Directly influences the outcome of queries
- Replication strategy
  - Determines how a node selects peers in the network for replication
  - Ensures availability of files in the search path

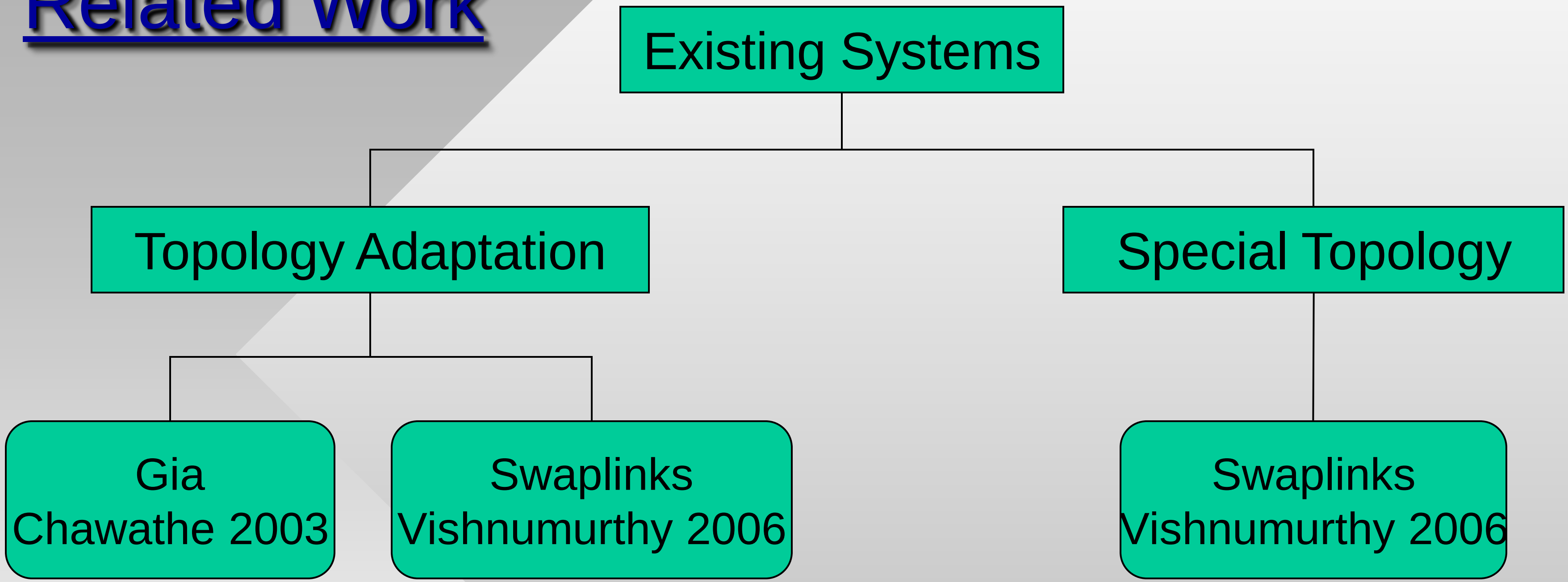
Introduction – Random walks

- Build walk
  - Node seeking a neighbor starts a walk with $TTL=k_b$
  - Peer at the end of the walk is selected as a neighbor
- Search walk
  - Node looking for a file starts a walk with $TTL=k_s$
  - Required file is searched on nodes at every hop
- Replication walk
  - Node i starts a random walk with $TTL=k_r$
  - Nodes along the walk are selected as replicas

Introduction – Capacity-Heterogeneity

- P2P networks rely on collaboration between nodes
- Capacity indicates the amount of service provided by a node to other peers in the network
- Node capacity can be defined in terms of the available bandwidth and local resources of a node such as its processing power
- Measurement studies show that P2P networks are capacity-heterogeneous

Related Work



- Involves replacing existing neighbors with better ones to satisfy capacity constraints
- Creating overlays with node degree linearly proportional to capacity

Motivation

- Existing systems have following drawbacks
  - Rely on topology adaptation which causes high overhead during churn
  - Use node degree linearly proportional to capacity which results in high overhead to maintain large neighbor set
- The objective of this research is to design an unstructured P2P system which overcomes the above drawbacks and utilizes capacity-heterogeneity to improve search performance in the system

Agenda

- Introduction
- Proposed System
- Proposed Metrics
- Evaluation
- Wrap-up

Proposed System – Optimal Network

- Maximum rate of processing of messages at node i = C_i
- If incoming traffic $T_i > C_i$, messages are added to an infinite queue
- Given number of nodes n, set of capacities $\{C_1,\ldots,C_n\}$, fixed average degree d, and random walk length k
- We define a network N with a search algorithm S to be throughput-optimal if (N, S) achieves the maximum rate of completion of random walks $M = \sum_i C_i/k$
- Majority of the traffic in the network is due to random walks and hence throughput of the system is rate of completion of random walks

Proposed System – Optimal Network

- Lemma 1: Assume that random walks are started at each node with rate λ_i and proceed according to a positive irreducible Markov chain with transition matrix P. If k is larger than the mixing time of P, the following holds

$$T_i = \pi_i \sum_{i=1}^{n} \lambda_i k$$

  where π is the unique solution to $\pi = \pi P$

- Theorem 1: Assuming k is sufficiently large, the optimal stationary distribution of random walks is

$$\pi_i = \frac{C_i}{\sum_{j=1}^{n} C_j}$$

Proposed System – CPMH Framework

- A framework to achieve the optimal π
- Applies the Metropolis-Hastings algorithm to find transition probability of random walks having optimal π
  - We first choose candidate transition probability $q(i,j)$

$$q(i,j) = \frac{C_j}{\sum_{x \in N(i)} C_x}$$

  - Random walk then makes this transition with probability

$$\alpha(i,j) = \min\left(1, \frac{\sum_{x \in N(i)} C_x}{\sum_{x \in N(j)} C_x}\right)$$

- Achieves Capacity-Proportional π using Metropolis-Hastings algorithm, hence the name CPMH

Proposed System – CSOD Topology

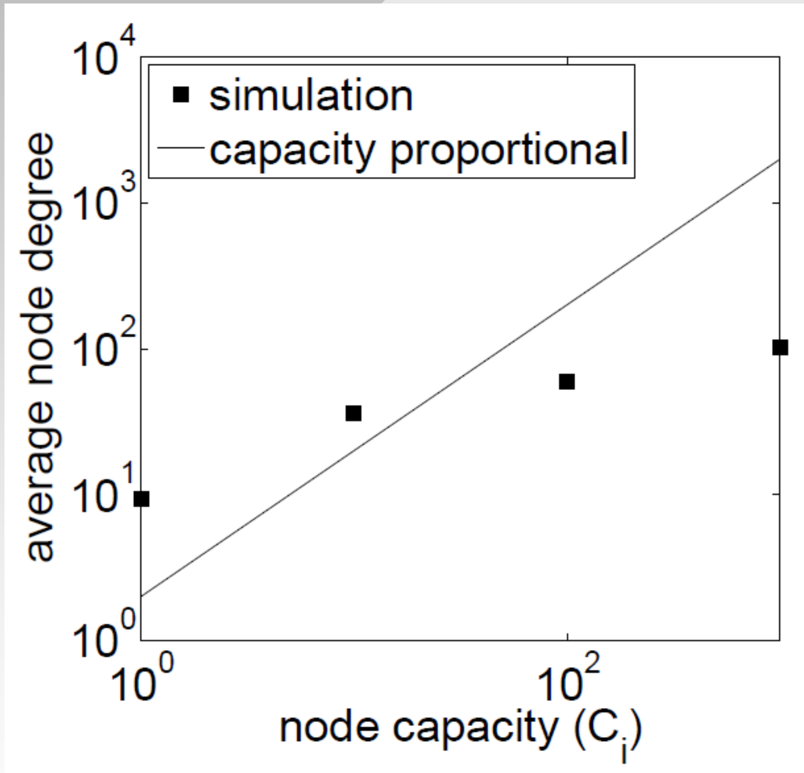
- A new node i joining the system will start $d_{out}(i)$ unbiased random walks for selecting its neighbors
- Desired out-degree $d_{out}(i) = a + \lfloor b \log_{10} C_i \rfloor$
  a, b are constants ($a = 4$ and $b = 15$ in simulations)
- Out-degree is scalable with capacity hence called Capacity Scalable Out Degree (CSOD)
- CSOD achieved fastest convergence to optimal π

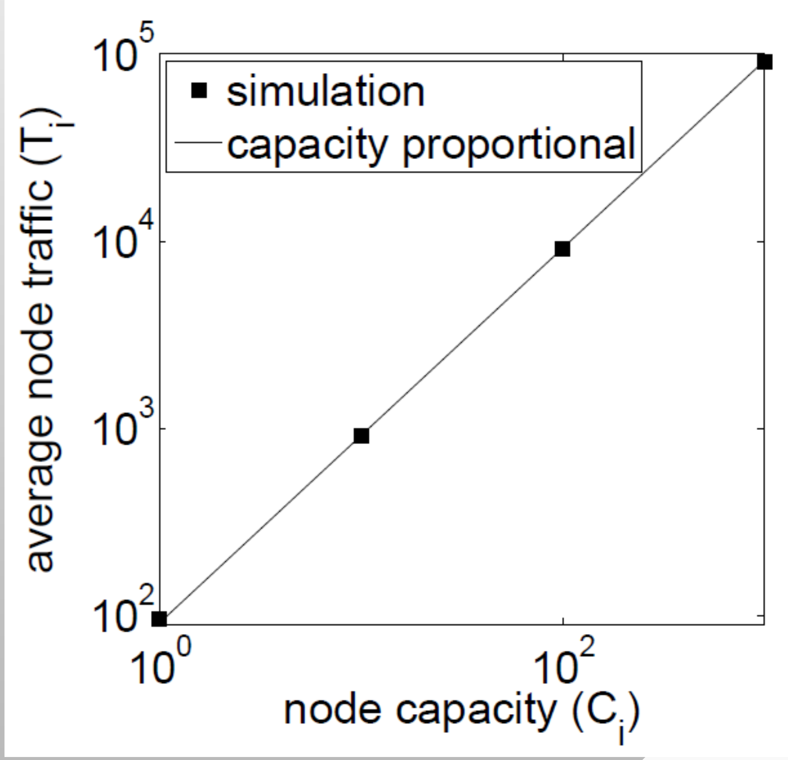
Topology	TTL
CSOD	50
Gnutella	620
BA	640

Capacity-unaware networks (Gnutella, BA)

Proposed System – CSOD Topology

- CPMH walks on CSOD





- 10,000 node network was constructed
- Walks of $TTL = 1024$ started at $\Lambda = 50$ qps

Proposed System – CPMH Search

- CPMH walks are used for query propagation
  - Achieve capacity-proportional traffic
- A query is propagated for k_s hops or till the specified number of query-hits are achieved
- CPMH queries are run on CSOD topology hence the proposed system is called CSOD-CPMH

Proposed System – CPMH Replication

- Maximum number of replicas stored in a node $= C_i$
- We propose random walk replication scheme using CPMH walks
- To achieve a replication factor r, a node starts one CPMH walk with $TTL = k_r$
  - Walk transitions for first h_f hops without replication
  - Subsequently, replication is done at every unique node visited till r replicas are created or TTL is reduced to 0
- In simulations, we use $k_r = 200$, $h_f = 50$ and $r = 20$

Agenda

- Introduction
- Proposed System : CSOD-CPMH
- Proposed Metrics
- Evaluation
- Wrap-up

Proposed Metrics

- End-to-end query metrics such as success rate additionally depend on replication strategy
- Need metrics to evaluate the topology of the system for supporting random walks
- Propose 2 topology metrics
  - Build Saturation Point (BSP)
  - Search Saturation Point (SSP)

Proposed Metrics – BSP

- BSP quantifies an overlay's ability to handle churn
- Churn involves nodes leaving the network and new ones joining the system
- Churn rate $r_c =$ departure rate of nodes
- BSP is defined as the maximum r_c for which the expected queue length $E[Q] < c$, after certain fixed time t, where c is a constant

Proposed Metrics – SSP

- Quantifies an overlay's ability to handle search walks
- Consider an overlay graph G with n nodes, capacity distribution $\{C_i\}$ and average degree d
- Random walks of length $k \geq 2$ are started from randomly selected nodes
- As input rate Λ of walks increases, completion rate M increases till the network is saturated
- Beyond saturation, message backlog increases and M decreases
- SSP: Unique maximum completion of walks achieved

$$SSP = \max_{\Lambda}[M]$$

Proposed Metrics – OPT Network

- Acts as an upper bound while comparing overlays using SSP
  - To add a comparative measure to SSP numbers
- We propose a centralized algorithm for construction of OPT
- Node i in OPT has $d_i = 2C_i$
- On OPT, unbiased random walks are run to get capacity-proportional traffic through nodes

Agenda

- Introduction
- Proposed System : CSOD-CPMH
- Proposed Metrics : BSP, SSP
- Evaluation
- Wrap-up

Evaluation

- Evaluate the proposed CSOD-CPMH against OPT-unbiased, Gia-biased, CSOD-biased
- Naming convention: {topology}-{search walk}
  - e.g., Gia-biased has Gia topology and uses capacity-biased search walks

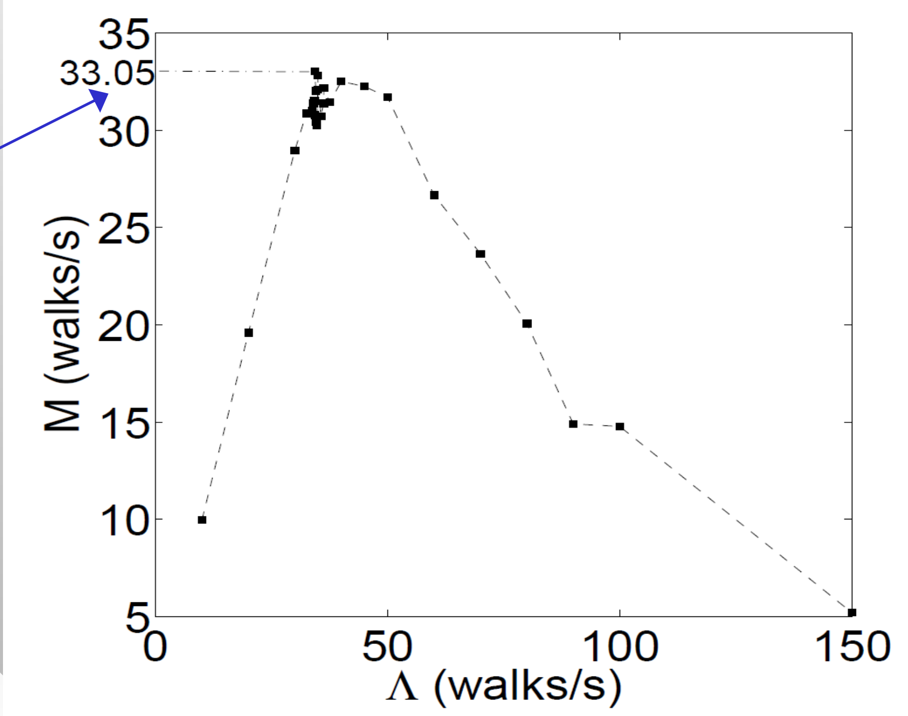
$$p(i,j) = \frac{C_i}{\sum_{j \in N(i)} C_j}$$

- CSOD-biased is considered to show the need for CPMH search walks

Evaluation – Static Network

- SSP is the maximum rate of completion M of search walks

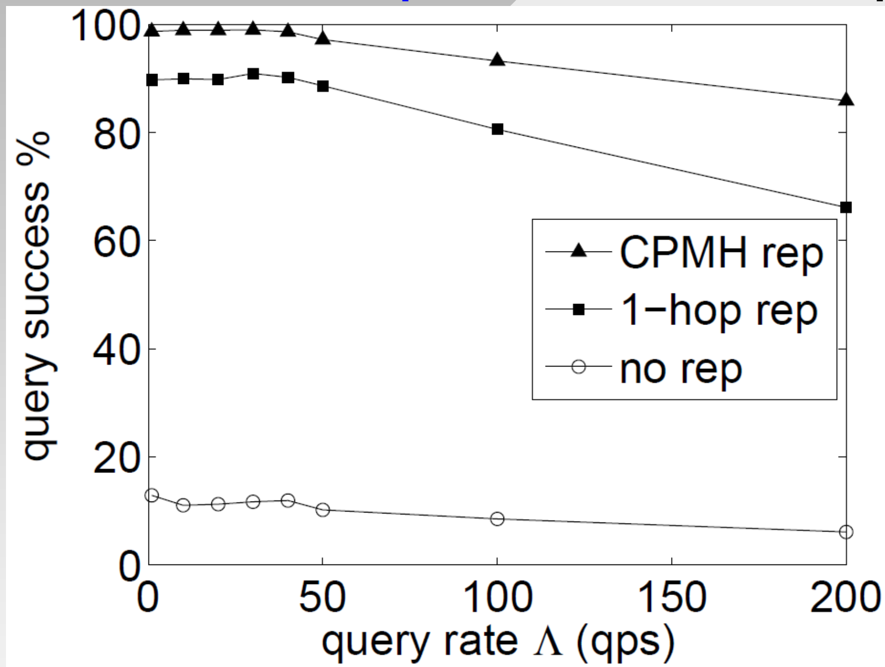
Name	SSP
OPT-unbiased	33.05
CSOD-CPMH	27.75
Gia-biased	6.60
CSOD-biased	5.94



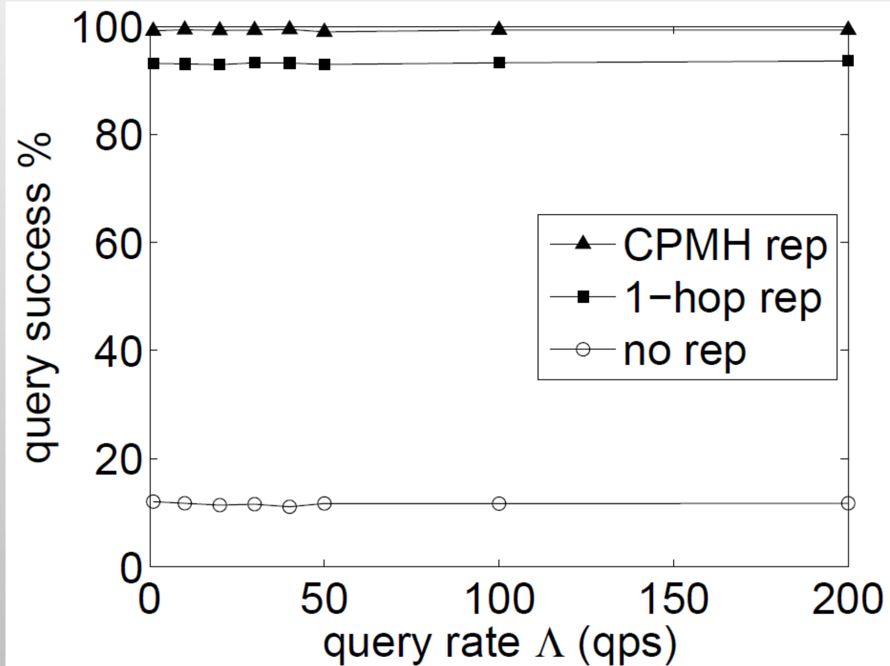
- TTL k_s=1024

Evaluation – Static Network

- CPMH replication is compared with 1-hop



Gia-biased



CSOD-CPMH

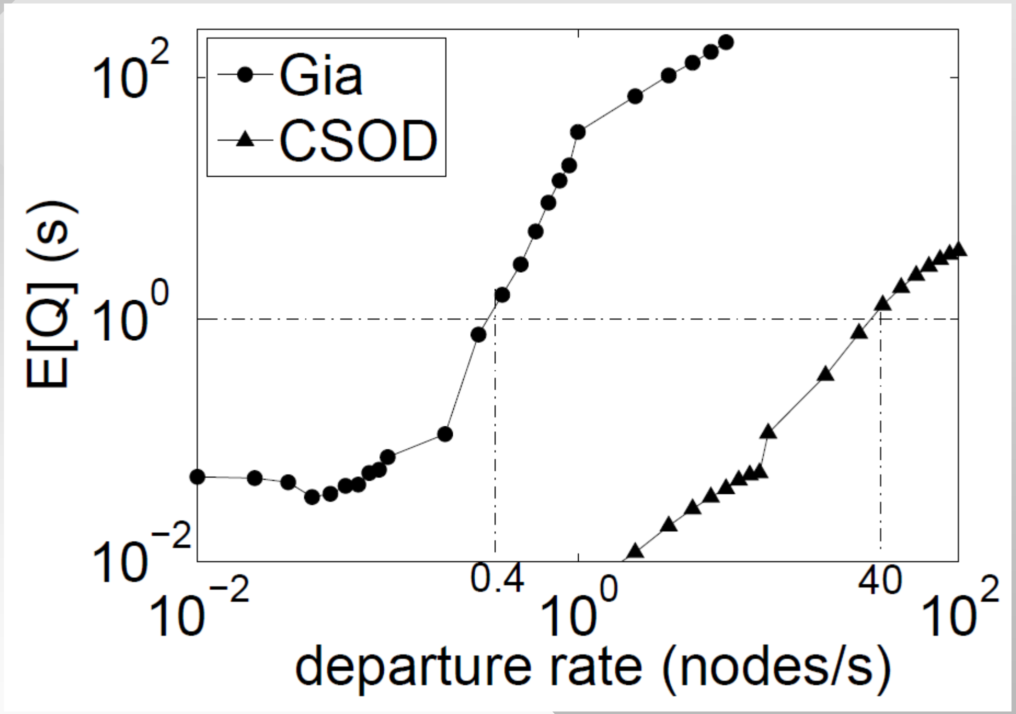
- TTL $k_r = 200$, $h_f = 50$, $r = E[d_i] = 20$
- CPMH-rep is up to 20% better than 1-hop in Gia-biased

Evaluation – Churn Model

- Nodes have Pareto lifetime L
  - Shape parameter $\alpha=3$, $E[L]=10000$ s
- New nodes arrive as a Poisson process
  - Arrival rate = Departure rate
  - Inter-arrival delay X, $E[X] = 1/r_c = E[L]/n$

Evaluation – Dynamic Network

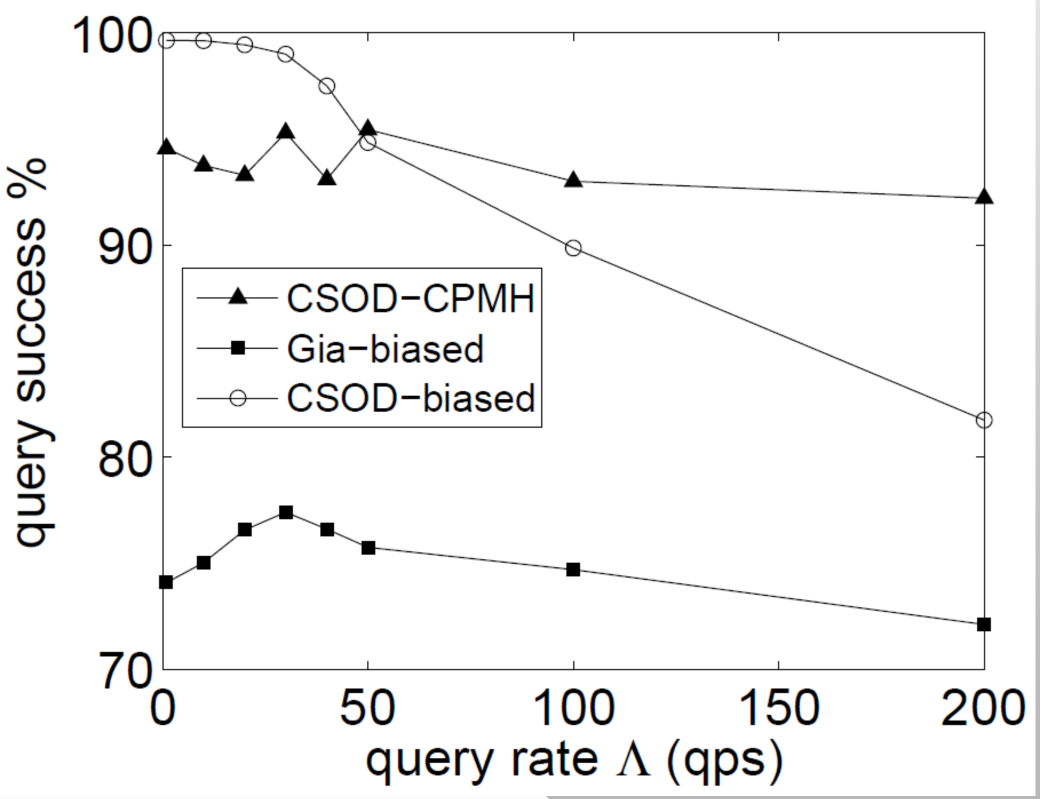
- BSP is the maximum μ for which $E[Q] \leq c$, after time t
  - In Simulation, backlog threshold $c = 1$ s and $t = 1000$ s



- CSOD node starts 10 unbiased build walks with $k_b = 8$
- Gia undergoes continuous topology adaptation

Evaluation – Dynamic Network

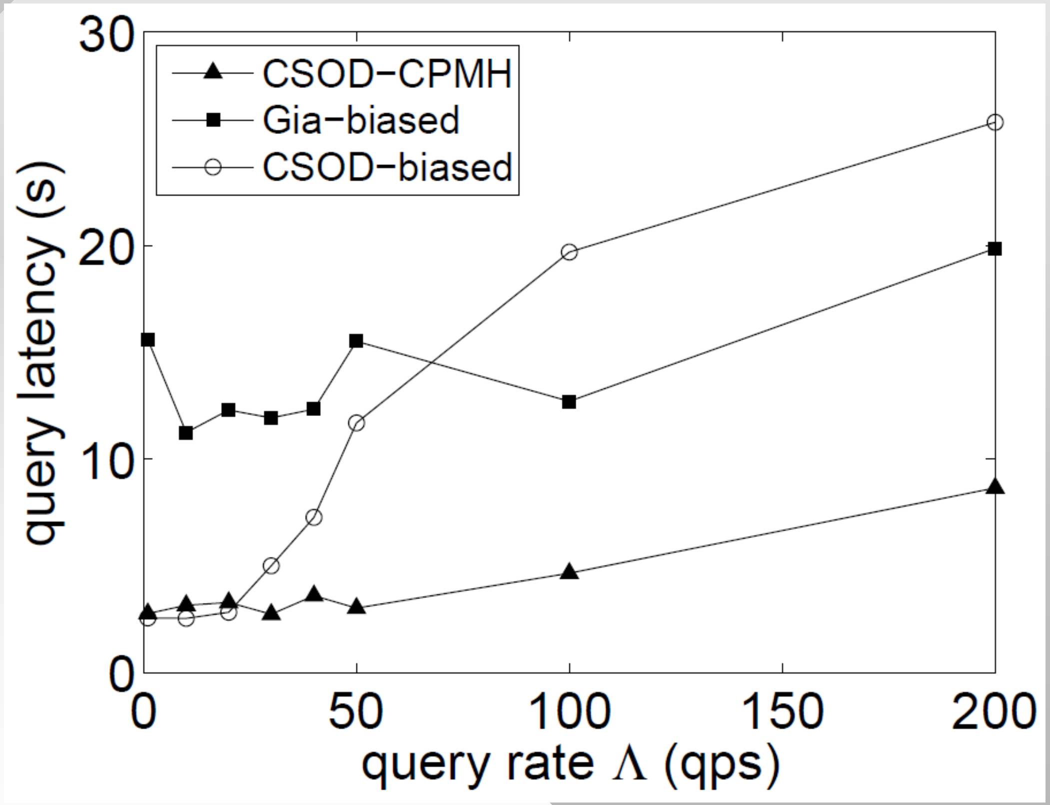
- Query success rate is the percentage of queries with one or more query-hits



- CSOD-CPMH has 20% higher success rate than Gia-biased

Evaluation – Dynamic Network

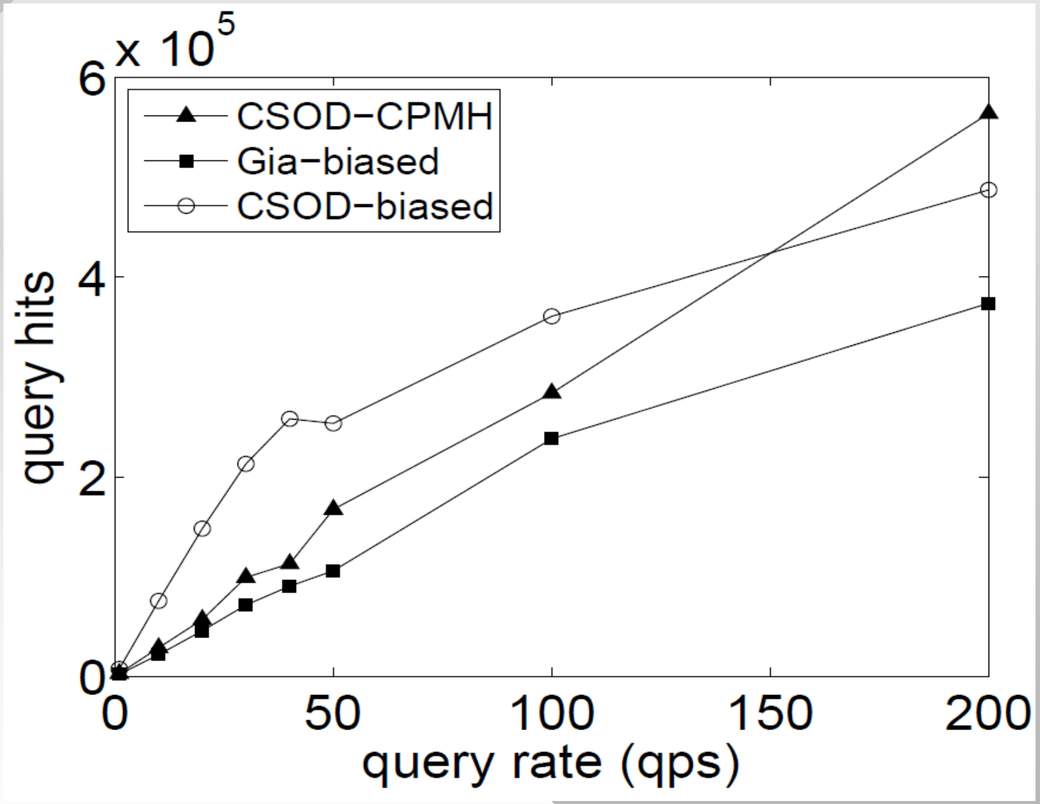
- Query Latency is the time to get the first query result



- CSOD-CPMH has 50% lower query latency than Gia-biased

Evaluation – Dynamic Network

- Query Hits is the total number of query responses



- CSOD-CPMH gets 50% more query hits than Gia-biased

Wrap-up

- Capacity-heterogeneity in a P2P network can be utilized without
  - Performing topology adaptation
  - Constructing topologies with $d_i = C_i$
- CSOD-CPMH performs better than Gia under both saturation metrics and all the end-to-end query parameters
- CPMH framework is topology-agnostic
  - Enables incremental deployment of proposed system into existing networks such as Gnutella