Link Lifetimes and Randomized Neighbor Selection in DHTs

Zhongmei Yao

Joint work with Dmitri Loguinov

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

April 15, 2008

Agenda

- Background and Motivation
  - Terminology and related work
- Link Lifetime Model for switching systems
  - General DHT space, neighbor dynamics, A semi-Markov chain
- Lifetimes of Deterministic Links
- Lifetimes of Randomized Links
- Wrap-up

Terminology

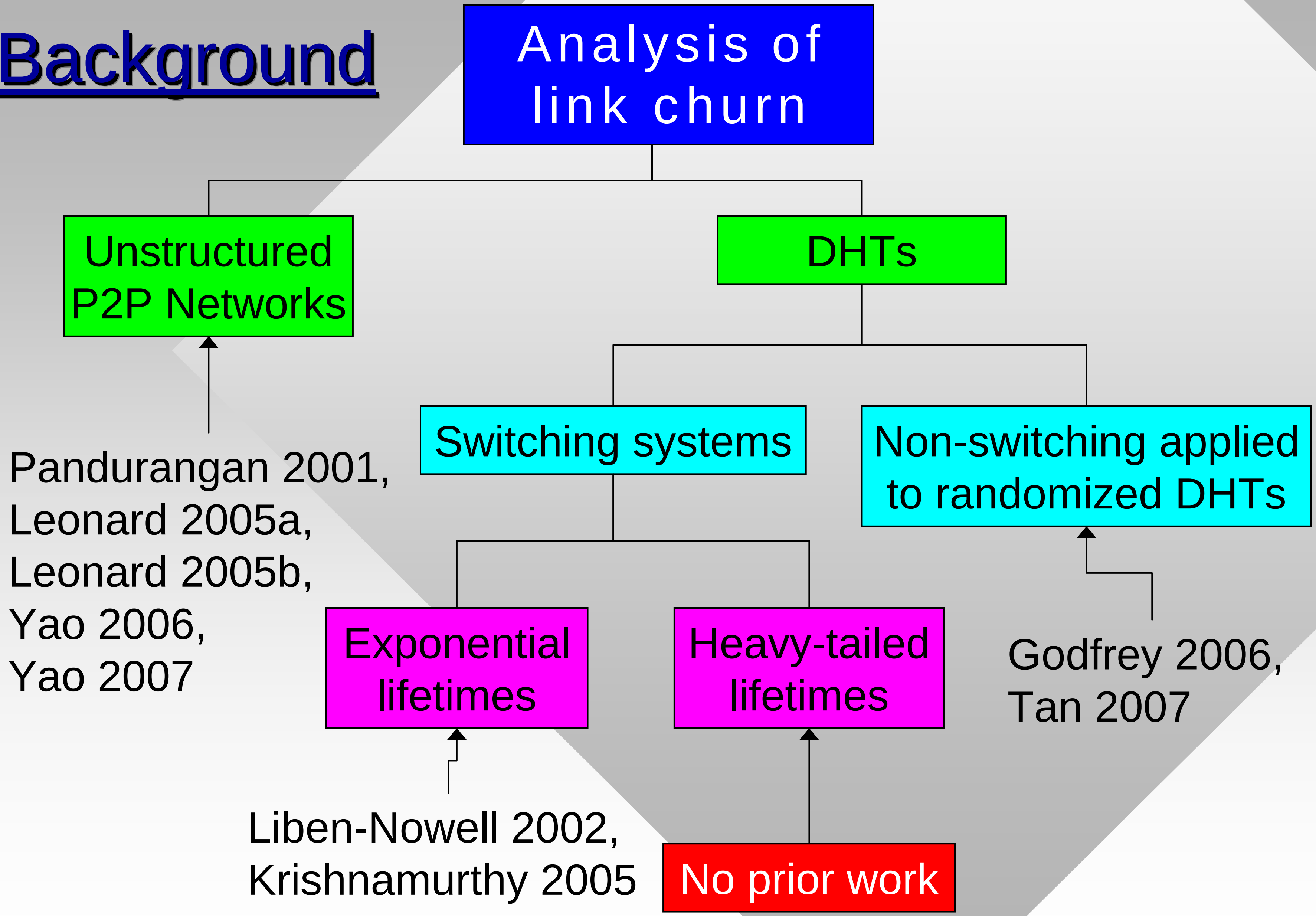
- User churn
  - User arrivals and departures are not synchronized
- Link creation in routing tables
  - Each user generates k out-links pointing to its neighbors
- Non-switching systems (e.g., Kad and Gnutella)
  - The link points to the same neighbor until it fails
- Switching systems (classic DHTs)
  - Links switch to new neighbors before the current neighbor dies
- Link lifetimes
  - Time duration when the neighbor adjacent to the link is alive

Terminology

- Repair of failed links
  - Detect failed links and replace with alive peers within S time units
- Link churn
  - The dynamic behavior of links

Computer Science, Texas A&M University

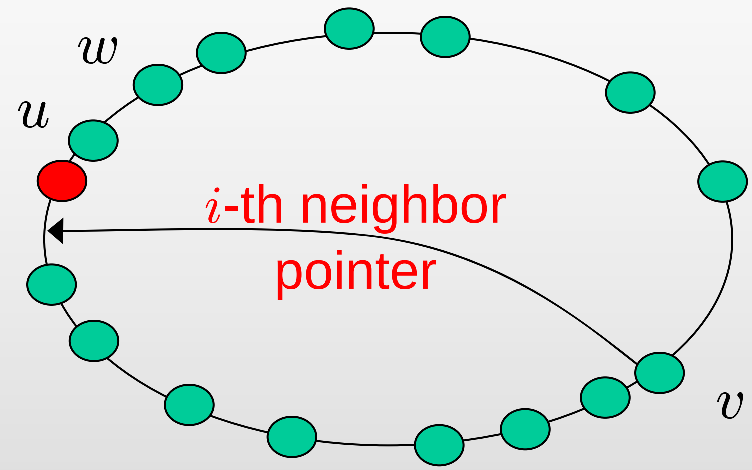
Background



Agenda

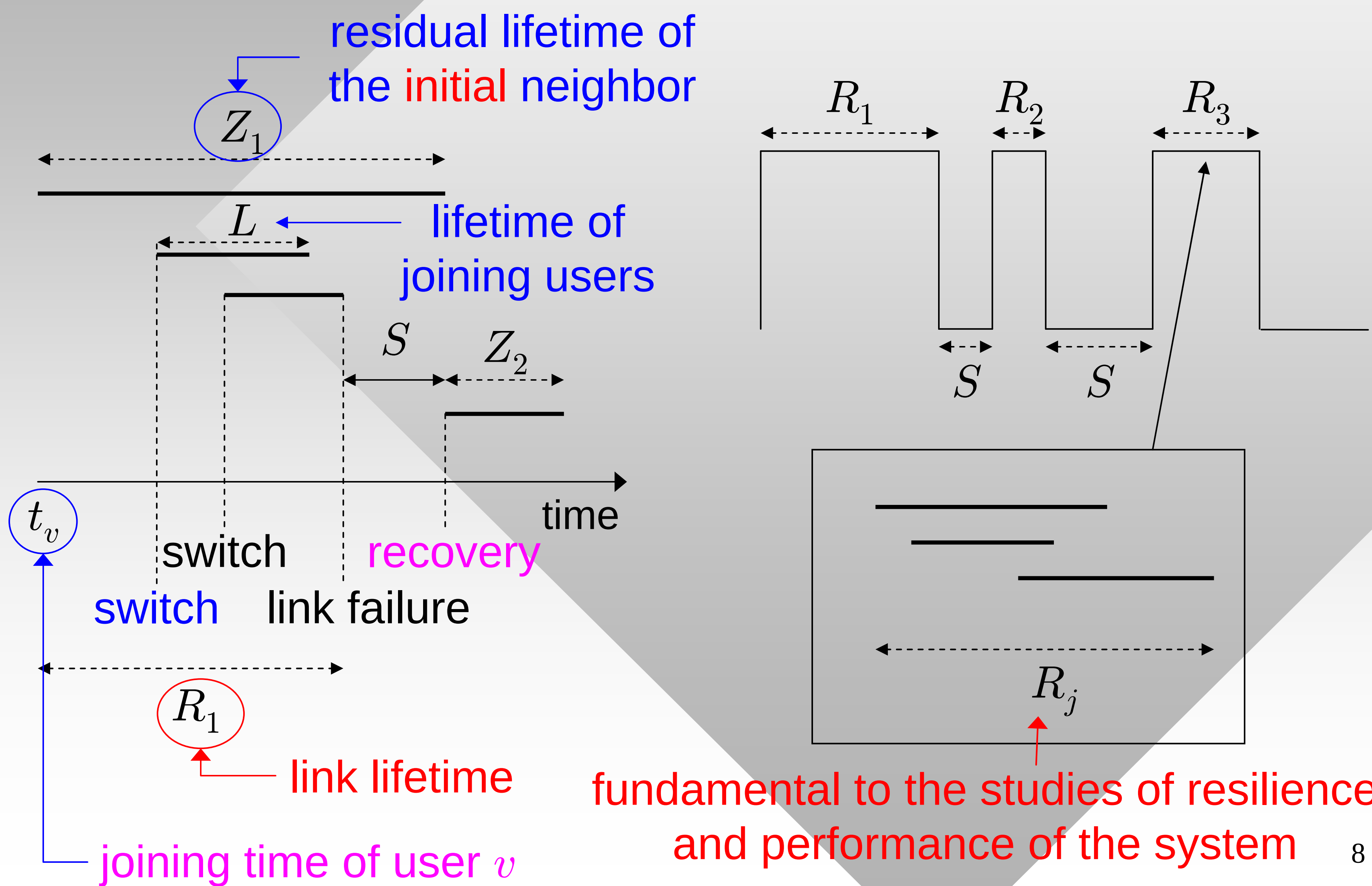
- Background and Motivation
  - Terminology and related work
- Link Lifetime Model for switching systems
- Deterministic Links
- Randomized Links
- Wrap-up

Links in DHTs



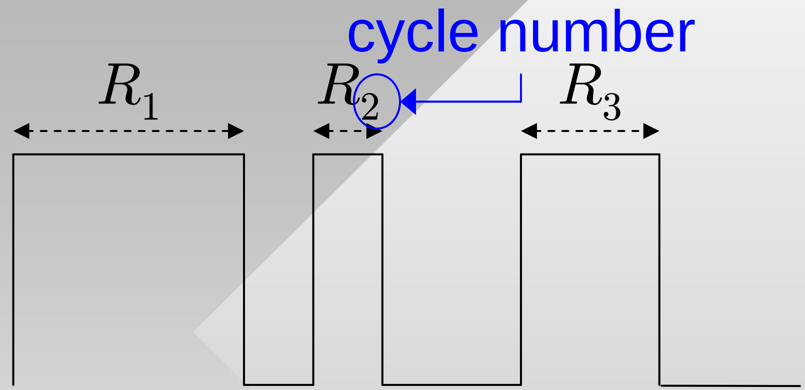
- The DHT space: consider a unit ring where hash indexes of users are uniform in $[0, 1)$
  - Random-split: zones are split at the hash indexes of joining users
- Fixed rules for selecting neighbors in routing tables
  - The successor of each neighbor pointer is v's neighbor
- Link ownership changes under churn
  - Recovery: an existing neighbor dies and the ownership is assigned to the successor of the failed neighbor
  - Switch: a link switches to new users who arrive into the zone before the current successor fails

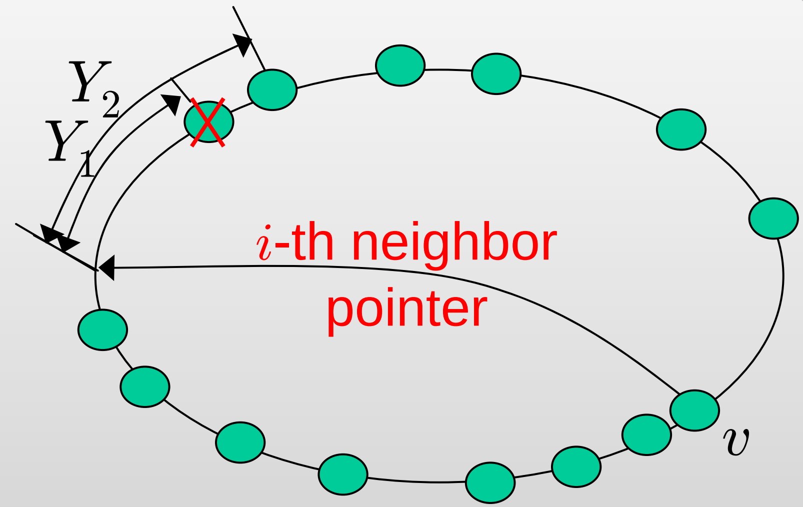
Link ON/OFF Behavior



fundamental to the studies of resilience and performance of the system

Zone Size





- Denote by Y_j is the zone size from the neighbor pointer to the initial neighbor who starts the jth cycle
  - Variable Y_1 is the zone size of the initial neighbor obtained when user v joins the system
- It determines the arrival rate of new users that split the zone and become the owner of the neighbor pointer
  - Large Y_j implies that more users arrive into the zone
- Other Y_j correspond to link recovery: the initial neighbor is found in replacement of the failed neighbor for $j = 2, 3, \ldots$

Link Lifetime Model Overview

- Define the conditional link lifetime $R(y)$ as the link ON duration conditioned on the zone size $Y_j = y$
  - We use a semi-Markov chain to study $R(y)$
- Properties of link lifetimes can then be examined:

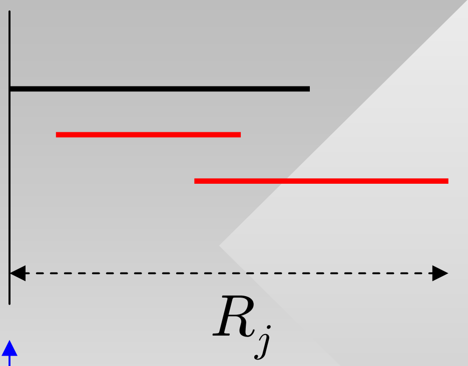
$$P(R_j < x) = \int_0^1 P(R(y) < x) f_{Y_j}(y) dy$$

the PDF of Y_j

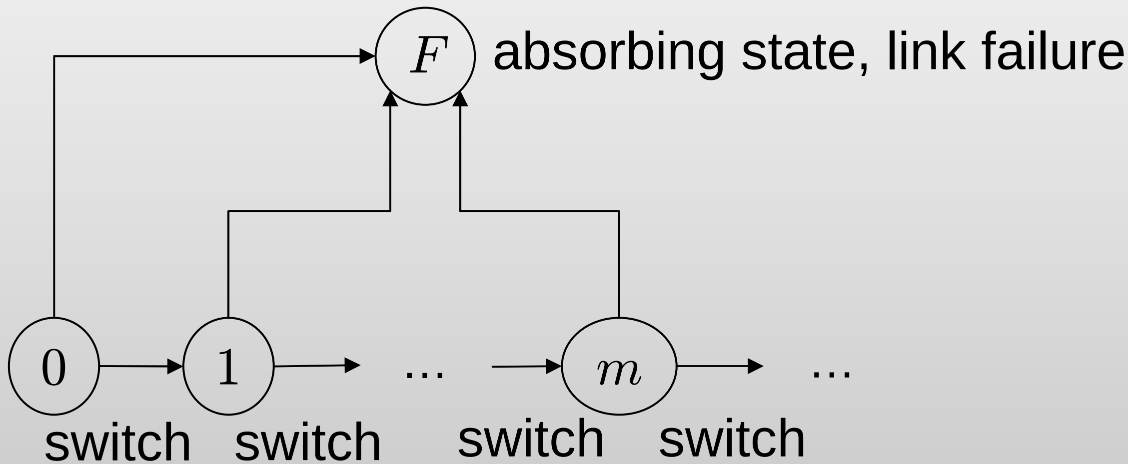
  - Compute the distribution of Y_j for deterministic DHTs and randomized DHTs accordingly

Computer Science, Texas A&M University

Conditional Link Lifetimes



time 0: the beginning of the current cycle



- Denote by A_t^y the number of switches (to new users) that have occurred in $[0, t]$ for given zone size $Y_j = y$
- Using notation A_t^y, we describe:

$$R(y) = \inf\{t > 0 : A_t^y = F \mid A_0^y = 0, Y_j = y\}$$

Conditional link lifetime

Conditional on zone size

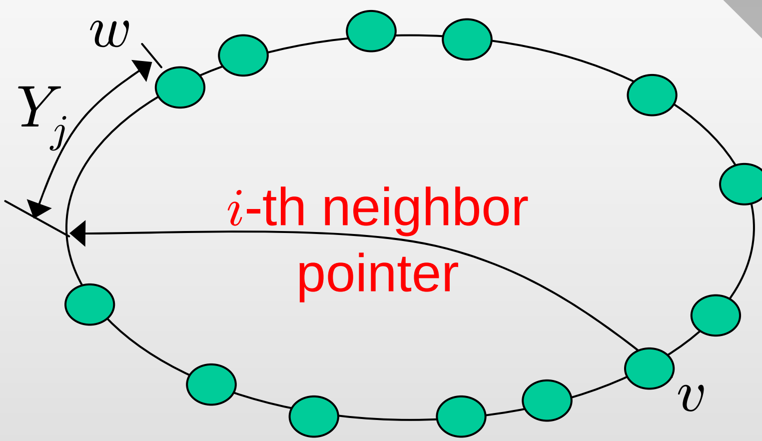
Conditional Link Lifetimes

- Theorem 1: Process $\{A_t^y\}$ is a semi-Markov process where the sojourn time τ_i in each state $A_t^y = i$ follows a general distribution
  - This process is fully determined by the distribution of residual lifetime Z of the initial neighbor, the distribution of user lifetimes L, and the arrival process of new users splitting the given zone
  - Based on this semi-Markov chain, one can obtain the distribution $P(R(y)<x)$ and expectation $E[R(y)]$
- Next, we focus on the distribution of zone size Y_j to get final results on link lifetimes

Agenda

- Background and Motivation
  - Terminology and related work
- Link Lifetime Model for switching systems
- Deterministic Links
- Randomized Links
- Wrap-up

Deterministic DHTs



- Theorem 2: In deterministic DHTs, the limiting distribution of Y_1 is exponential with mean $1/E[N]$ and that of Y_j for $j = 2, 3, \ldots$ is Erlang-2 with mean $2/E[N]$ as system population N becomes sufficiently large

- The mean link lifetime is given by:

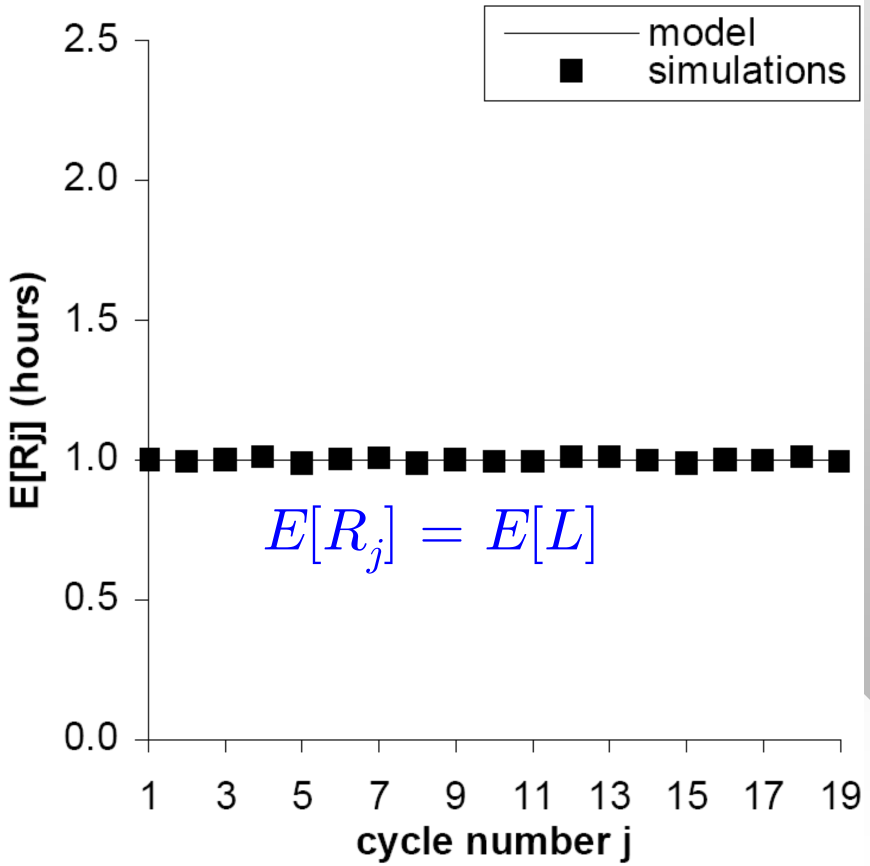
$$E[R_j] = \int_0^\infty E[R(y)] f_{Y_j}(y) dy$$

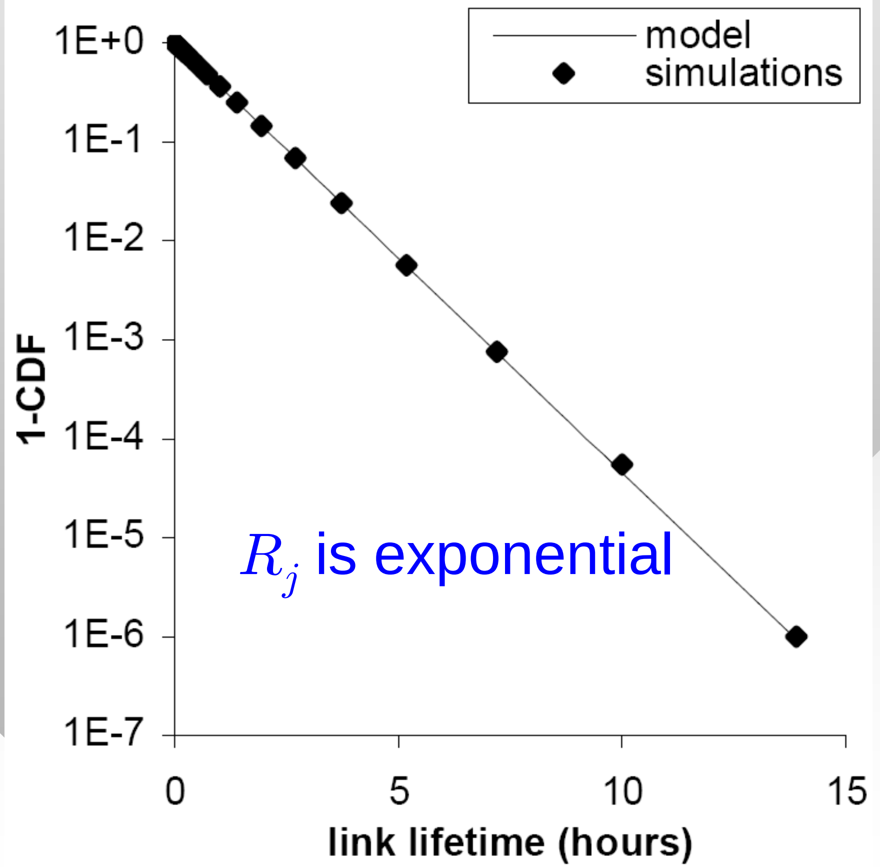
the mean conditional link lifetime

the PDF of Y_j

Exponential User Lifetimes

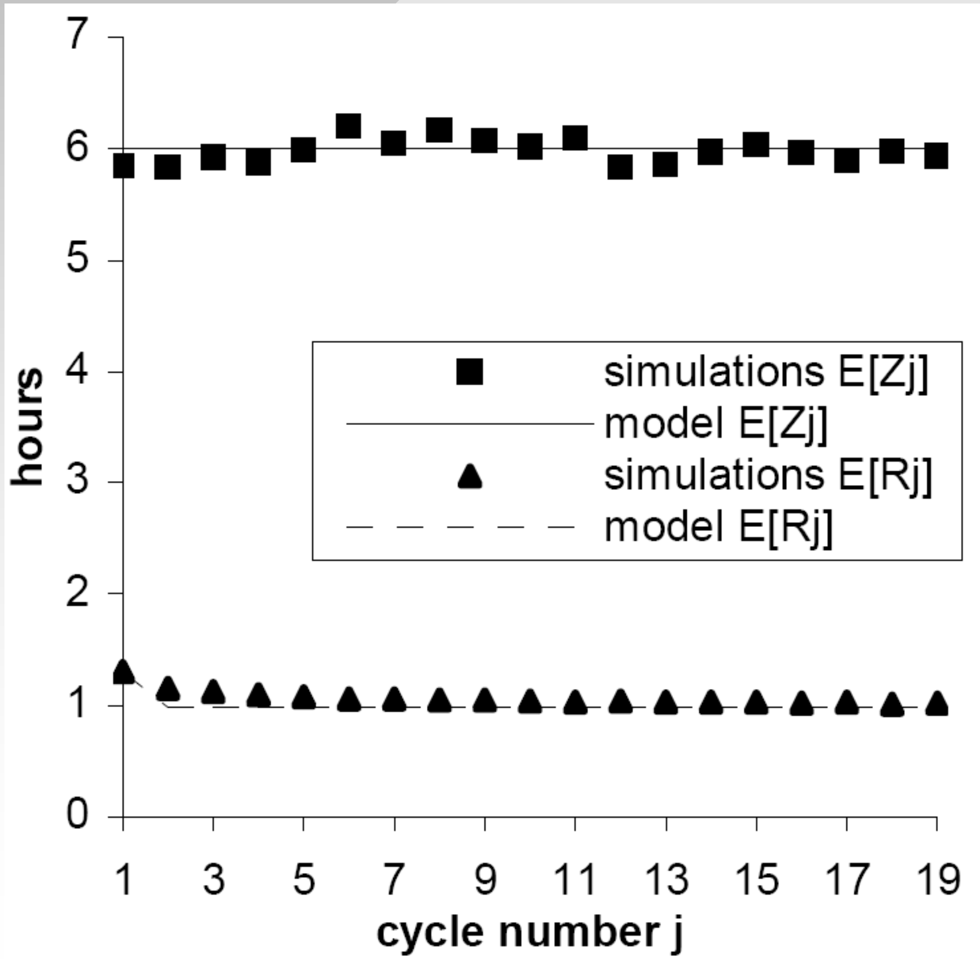
- Properties of link lifetimes R_j for exponential user lifetimes with $E[L] = 1$ hour





Pareto User Lifetimes

Pareto L with $\alpha = 2.2$ and $E[L] = 1$ hour



- The initial neighbor is reliable since $E[Z_j] > E[L]$
- $E[R_j]$ is very close to $E[L]$ for $j = 2, 3, \ldots$
- $E[R_1] > E[R_2]$ since Y_1 is stochastically smaller than Y_2
  - A smaller zone size leads to a larger mean link lifetime

Discussion

- Our model shows that link lifetime R in deterministic DHTs is stochastically smaller than residual lifetime Z of the initial neighbor holding the link
  - Switching to newly arriving users makes R smaller
  - Unlike non-switching systems, classic DHTs do not obtain benefits from heavy-tailed L
- Abandon switching systems?
  - Non-switching DHTs create inconsistence in routing tables and may expect longer routing delay
- We propose a new method that not only retains the advantage of switching systems, but increases link lifetimes

Agenda

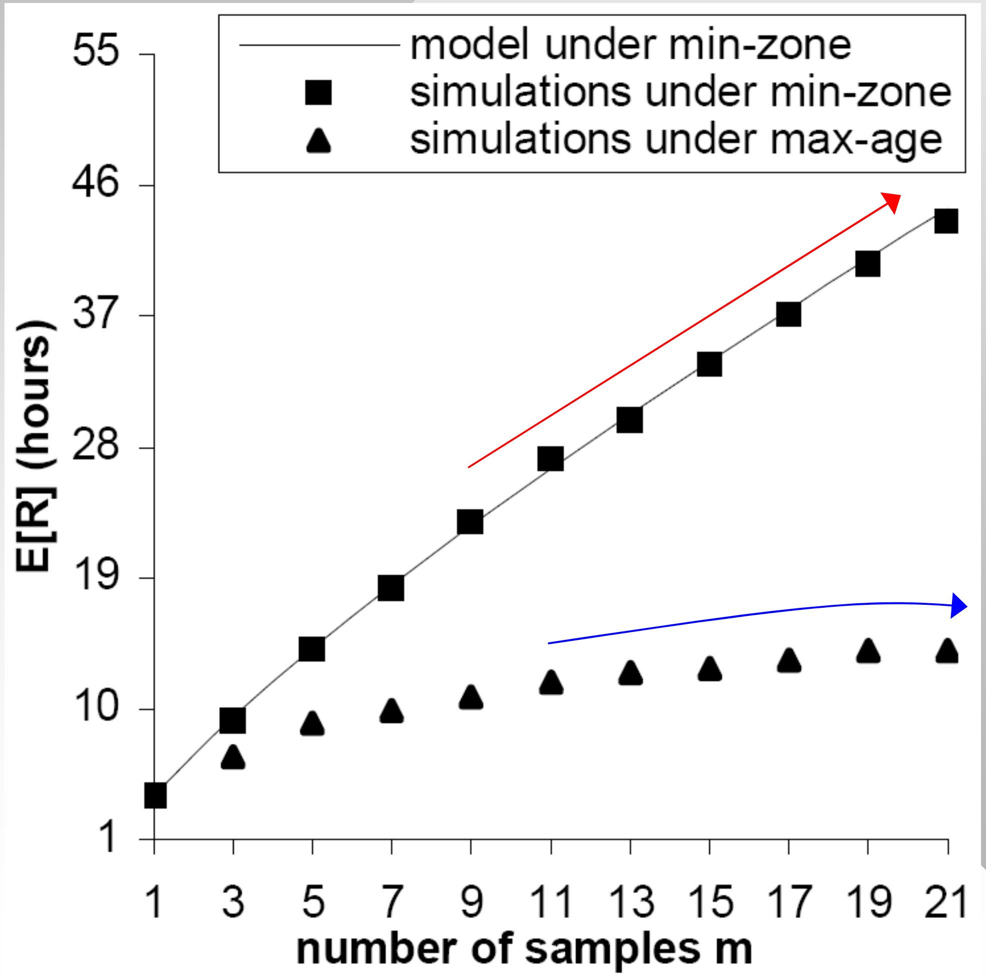
- Background and Motivation
  - Terminology and related work
- Link Lifetime Model for switching systems
- Deterministic Links
- Randomized Links
- Wrap-up

Improvement: Randomized Links

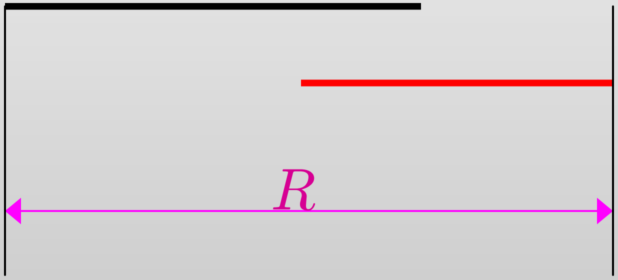
- We utilize the freedom of selecting links in randomized DHTs to propose min-zone selection
  - User v uniformly samples m points in $[\text{id}(v) + 2^i/2^{64}, \text{id}(v) + 2^{i+1}/2^{64}]$, and then selects the point whose successor has the minimum zone size
  - Upon link failure, user v uses the same strategy to find a replacement
  - Zone size Y_j is exponential but has a smaller mean $E[Y_j] = 1/(mE[N])$, where N is system population, for all j
- For comparison purpose, we also examine max-age selection
  - The only difference is that age is used as selection criteria

Link Lifetimes under Min-Zone Selection

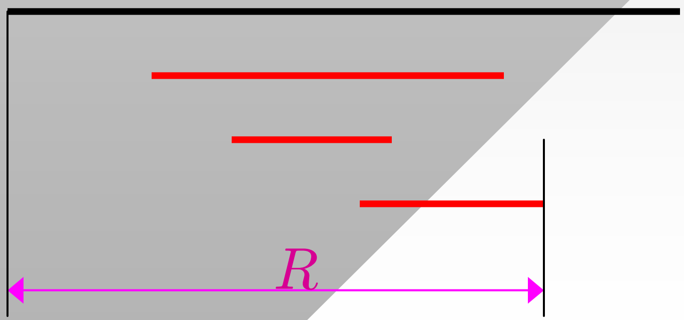
Pareto L with $\alpha = 1.2$ and $E[L] = 1$ hour



- Min-zone:

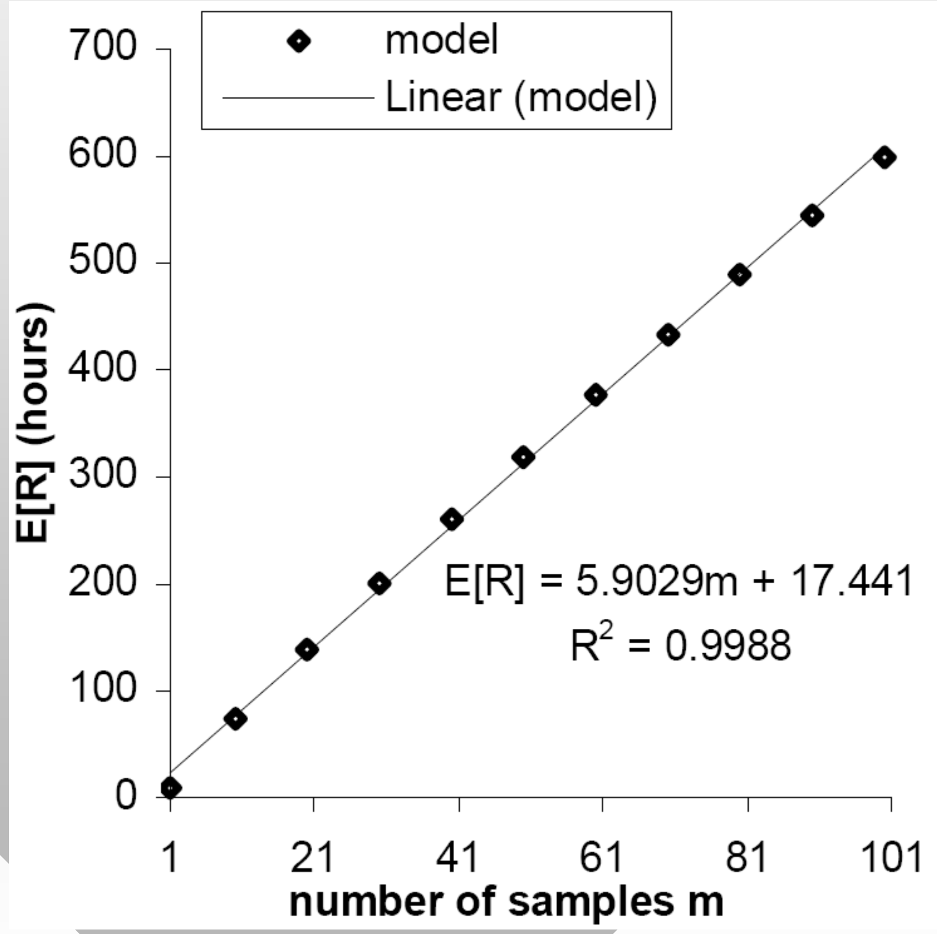


- Max-age:



Link Lifetimes for $\alpha \leq 2$

- Theorem 3: For Pareto user lifetimes with shape $1 < \alpha \leq 2$ and min-zone selection, $E[R] \to \infty$ as the system size and m approach ∞. For max-age selection and any $\alpha > 1$, $E[R]$ is finite.



Pareto L with $\alpha = 1.09$ and $E[L] = 1$ hour

Wrap-up

- We developed a model for link lifetimes R in DHTs
  - The mean link lifetime in deterministic DHTs is very close to the mean user lifetime
  - Switching leads to smaller link lifetimes
- We proposed min-zone selection which sufficiently increases R for heavy-tailed user lifetimes
  - It allows us to achieve a spectrum of neighbor selection strategies while keeps routing tables consistent
  - For $m = 1$, it is the regular switching in DHTs
  - For $m = \infty$, the probability of switching is reduced to be 0
  - Additionally, it benefits DHTs by balancing load such that users with smaller zone sizes are responsible for fewer keys while forwarding more queries

Computer Science, Texas A&M University