Understanding Disconnection and Stabilization of Chord

Zhongmei Yao

Joint work with Dmitri Loguinov

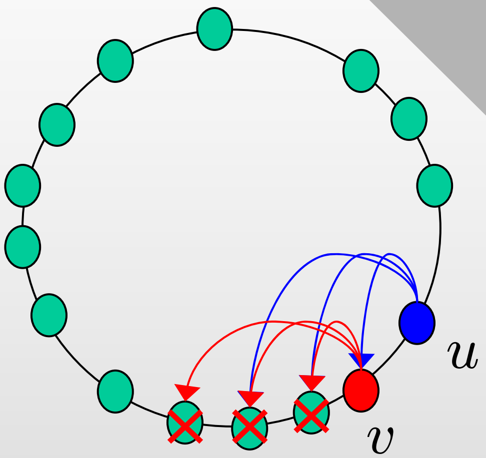
Internet Research Lab
Department of Computer Science
Texas A&M University, College Station, TX 77843

April 17, 2008

Agenda

- Background and Motivation
- Disconnection of Chord under Static Node Failure
- Disconnection of Chord under Dynamic Node Failure
  - Node Isolation
  - Graph Disconnection
- Stabilization Strategies
- Wrap-up

Successor Sets in DHTs



- We study one particular DHT, Chord
  - Each user maintains a list of r successors
  - When all of r successors of a node fail, the graph is disrupted
- Roles of Successor lists
  - Successor lists ensure correctness of lookup and connection of the graph, whereas routing tables are mainly used for reducing lookup latency
  - Unlike routing tables, a different strategy is used in Chord to stabilize successor lists
- There exists strong dependency among successor lists of consecutive users on the ring
  - As illustrated, the number of common successors shared by two consecutive nodes is $r-1$

Background

Node Failure in P2P Networks

Static node failure (i.e., p-fraction)

- Routing table
- Successor list

Dynamic node failure (i.e., churn)

- Routing table
- Successor list

Gummadi 2003, Kaashoek 2003, Leonard 2005

Stoica 2003

Krishnamurthy 2005, Leonard 2005

Krishnamurthy 2004 (TR)

A brief announcement (e.g., find r successors w.h.p.)

Studied graph disconnection for fixed $r = 2$ and Markovian systems

Agenda

- Background and Motivation
- Disconnection of Chord under Static Node Failure
- Disconnection of Chord under Dynamic Node Failure
  - Node Isolation
  - Graph Disconnection
- Stabilization Strategies
- Wrap-up

Model Basics

- Suppose that each node in Chord correctly links to its r closest successors
- Consider the resilience of Chord when each user becomes dead with an independent probability p
  - This is a one-time simultaneous failure event
- Note that Chord is connected iff each user has at least one alive successor among its r successors
- Define a Bernoulli random variable X_i for each node i:

$$X_i = \begin{cases} 1 & i \text{ is alive and its } r \text{ successors failed} \\ 0 & \text{otherwise} \end{cases}$$

  - Call the event $X_i = 1$ as isolation of node i

Disconnection under Static Node Failure

- The number of isolated nodes in a system with n users is:

$$X = \sum_{i=1}^{n} X_i$$

Sum of dependent random variables

- The probability that Chord is connected is equal to:

$$P(\text{connected}) = P(X = 0)$$

  – Using the Erdös and Rényi law, we obtain the next result

- Theorem 1: Given that $r \to \infty$ as $n \to \infty$, the probability that Chord remains connected under p-fraction node failure is:

$$\lim_{n\to\infty} \frac{P(X=0)}{e^{-n(1-p)p^r}} = 1$$

Disconnection under Static Node Failure

- Simulation results on the non-partitioning probability of Chord

$p = .933,\ r = \lceil 2\log_2 n \rceil$		
n	Simulations	(13)
1,000	.9417	.9369
5,000	.9373	.9360
10,000	.9367	.9360
20,000	.9365	.9360
30,000	.9368	.9367
40,000	.9363	.9361
50,000	.9393	.9393
100,000	.9395	.9394

$n = 50,000,\ r = \lceil 10\log_2 n \rceil$		
p	Simulations	(13)
.89	.9999	.9999
.9	.9997	.9997
.91	.9983	.9983
.92	.9919	.9918
.93	.9614	.9613
.94	.8344	.8343
.95	.4514	.4514
.96	.0368	.0371

Discussion

- Note that for $r \to \infty$ as $n \to \infty$, node isolations become rare events $P(X_i = 1) = (1 - p)p^r \to 0$
- The probability that Chord is connected can then be transformed into:

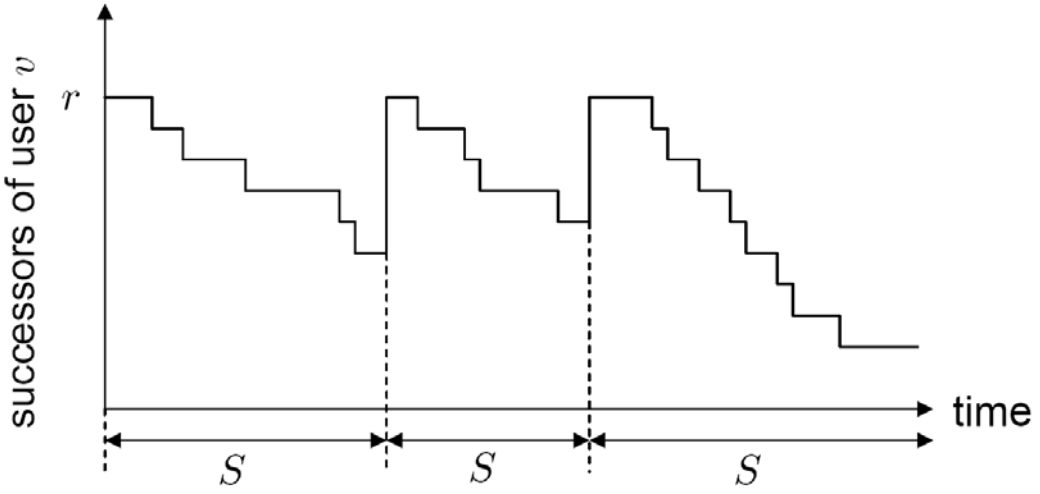
$$P(X = 0) = e^{-n(1-p)p^r} = \prod_{i=1}^{n} P(X_i = 0)$$

  - The above shows that variables X_i behave *as if* they are completely independent
  - Though dependency among successors of consecutive users is very strong, Chord exhibits the same static resilience as other P2P networks using mostly independent peers in their routing tables (see Leonard 2005)
  - The rate of convergence of $P(X = 0)$ is different

Agenda

- Background and Motivation
- Disconnection of Chord under Static Node Failure
- Disconnection of Chord under Dynamic Node Failure
  - Node Isolation
  - Graph Disconnection
- Stabilization Strategies
- Wrap-up

Model Basics



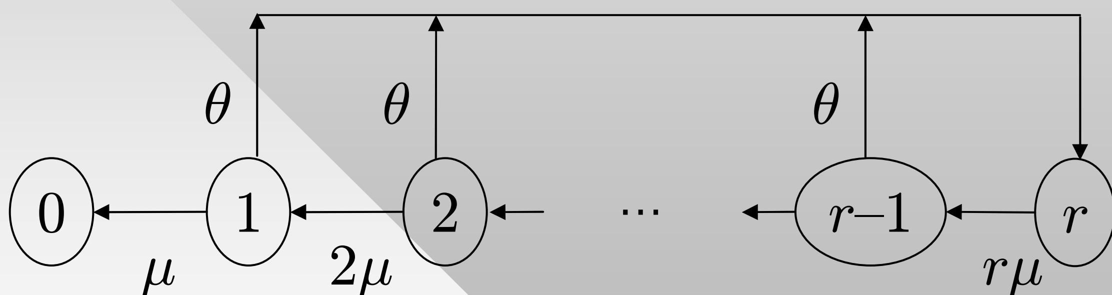
- Unlike static node failure, user arrivals/departures are not synchronized in human-based P2P networks
- Model basics
  - Each joining user obtains r closest successors and leaves the system in L time units
  - Each user performs stabilization every S time units on the entire successor list
  - At the end of each stabilization, bring the number of successors back to r (old successors + newly arriving users)
  - Denote by $Z(t)$ the number of successors of user v at time t , where time 0 is the time when v joins the system

Exact Node Isolation Model

- Define the first-hitting time T onto state 0:

$$T = \inf(t > 0 : Z(t) = 0 | Z(0) = r)$$

- Lemma 1: For exponential user lifetimes with rate μ and exponential stabilization intervals with rate θ, $\{Z(t)\}$ is a continuous-time Markov chain

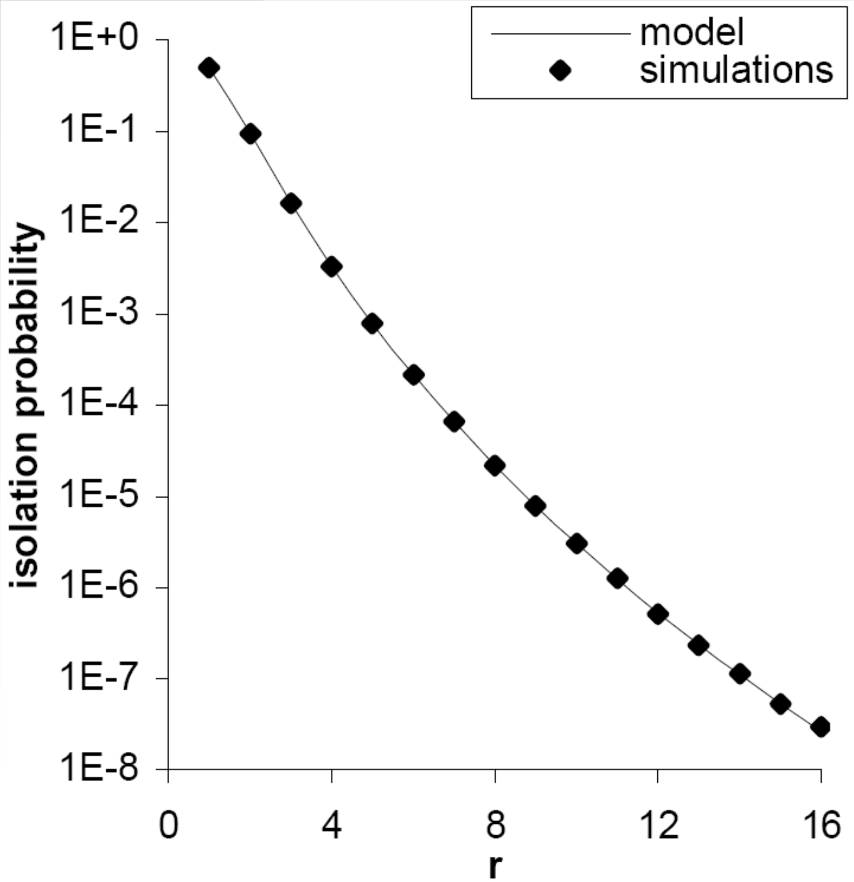


  – Using the Markov chain, we obtain the PDF $f_T(t)$ of T
  – The probability that node v is isolated due to the failure of the entire successor list within v's lifetime is then:

$$\phi = P(T < L) = \int_0^\infty P(L > t) f_T(t) dt$$

Exact Node Isolation Model

- Simulation results on isolation probability ϕ for mean user lifetime $E[L] = 30$ minutes and $E[S] = 2$ minutes



- The model is accurate for exponential L and S
- As r increases or $E[S]$ decreases, ϕ sharply decreases
- This exact model will be used to verify the accuracy of our later closed-form bounds on ϕ when other distributions of S are used

Graph Disconnection

- Recall that X_i is a Bernoulli variable indicating whether node i is isolated within its lifetime due to the failure of its successor list:

$$\phi = P(X_i = 1) = 1 - P(X_i = 0)$$

  - If $X_i = 0$, the network is said to survive the presence of use i

- Supposing that N users have joined the system, the number of isolations among N continuous join/departure events:

$$X_N = \sum_{i=1}^{N} X_i$$

Sum of dependent random variables

- Open question: how do individual node isolations affect the connectivity of Chord under churn $P(X_N = 0)$?

Graph Disconnection

- Define B_i to be the set of users who share at least one successors of user i

$$B_i = \{i - r + 1, \ldots, i, \ldots, i + r - 1\}$$

- For $|B_i| = 2r - 1$, we use the Chen-Stein method to show that X_N is asymptotically a Poisson random variable under the condition given below

- Theorem 2: Given that $Nr\phi \to 0$ as $N \to \infty$, the probability that Chord survives N user joins without disconnection approaches:

$$\lim_{N\to\infty} \frac{P(X_N = 0)}{(1-\phi)^N} = 1$$

Graph Disconnection

- Simulation results on $P(X_N = 0)$ that Chord accommodates N joining users without disconnection
  - $r = 8$, mean system size $2{,}500$, $E[L] = 0.5$ hours, and $\rho{=}E[L]/E[S]$

$\rho = 40$ ($E[S] = 45$ s)			$N = 50,000$			
N	Simul.	(46)	ρ	$E[S]$ s	Simul.	(46)
1,000	1.000	.9999	16	112.5	.4831	.4557
5,000	.9996	.9995	24	75.0	.9176	.9139
8,000	.9993	.9993	32	56.3	.9833	.9829
10,000	.9992	.9991	40	45.0	.9954	.9955
50,000	.9954	.9955	48	37.5	.9985	.9985
100,000	.9910	.9910	56	32.1	.9995	.9994
500,000	.9555	.9556	64	28.1	.9998	.9998

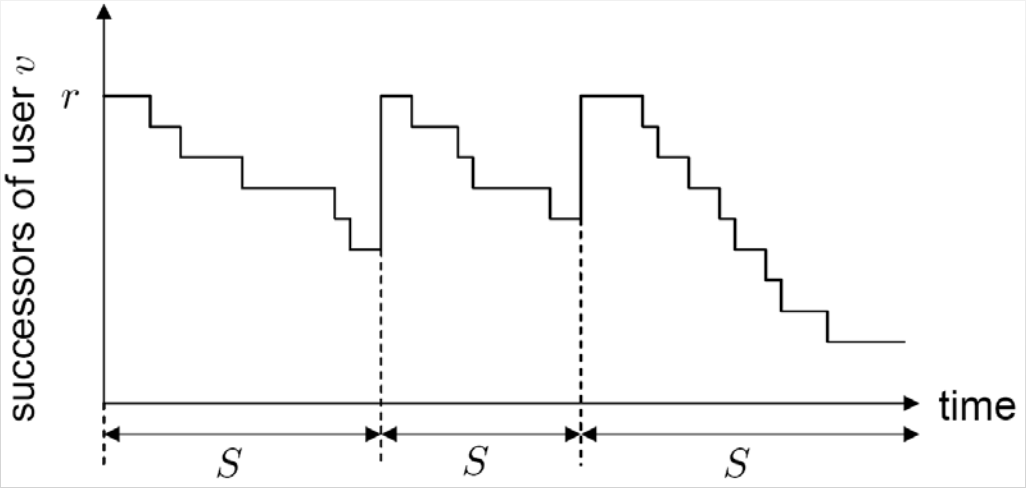
Discussion

- A similar result $P(X_N = 0) = (1 - \phi)^N$ was observed in *Leonard 2005* without proof where node isolations thought neighbor sets are considered as independent events
  - Our result is formally proven and is stronger than *Leonard 2005* since it applies to successor lists having dependency during failure
- As node isolations become rare, the probability of non-partitioning of Chord converges to that of avoiding isolation of each joining user
  - Node isolation probability provides sufficient information to predict the disconnection probability of Chord
  - Next, we focus on node isolation probability when different distributions of stabilization intervals are used

Agenda

- Background and Motivation
- Disconnection of Chord under Static Node Failure
- Disconnection of Chord under Dynamic Node Failure
  - Node Isolation
  - Graph Disconnection
- Stabilization Strategies
- Wrap-up

Closed-Form Bounds on Node Isolation Probability



- Note that the sequence of stabilization intervals forms a renewal process with cycle length S
  - The probability that all r successors fail in a particular interval S is given by:

$$f = P(\max\{L_1, \ldots, L_r\} < S)$$

remaining lifetime of the i-th successor

- Based on renewal theory and Jensen's inequality, we derive a simple bound using f for isolation probability

Closed-Form Bounds

- Theorem 3: For $L \sim \exp(\mu)$ and $S \sim \exp(\theta)$, isolation probability ϕ is upper-bounded by:

$$\phi \le \rho f \qquad f = \frac{\rho! r!}{(\rho + r)!}$$

$$\rho = \frac{E[L]}{E[S]} = \frac{\theta}{\mu}$$

Moreover, as $\rho \to \infty$, the upper bound becomes exact

ρ	$E[S]$ s	exact model	upper bound	Relative Error
10	180	1.46×10^{-4}	2.29×10^{-4}	57.05%
50	36	2.30×10^{-8}	2.61×10^{-8}	13.41%
100	18	2.66×10^{-10}	2.84×10^{-10}	6.85%
200	9	2.55×10^{-12}	2.64×10^{-12}	3.46%
500	3.6	4.74×10^{-15}	4.80×10^{-15}	1.29%
1,000	1.8	3.86×10^{-17}	3.89×10^{-17}	0.69%

Uniform Stabilization Delays

- Theorem 4: For fixed $r > 3$ and $E[L]$, and uniform intervals $S \in [0, 2E[S]]$, the ratio of isolation probability ϕ_u for uniform S to ϕ for exponential S approaches:

$$\lim_{E[S]\to 0} \frac{\phi_u}{\phi} = \frac{f_u}{f} = \frac{2^r}{(r+1)!} < 1$$

$$f = P(\max\{L_1, \ldots, L_r\} < S)$$

- Model $\phi_u/\phi = .0127$ for $E[L] = 0.5$ hours, $r = 6$

ρ	$E[S]$ s	Simulations of ϕ_u	Simulations of ϕ	ϕ_u/ϕ
20	90	2.15×10^{-6}	7.10×10^{-5}	.0303
40	45	7.59×10^{-8}	3.86×10^{-6}	.0197
60	30	9.98×10^{-9}	6.10×10^{-7}	.0164
80	22.5	2.28×10^{-9}	1.62×10^{-7}	.0141
100	18	7.18×10^{-10}	5.59×10^{-8}	.0128

Optimal Stabilization Strategy

- Theorem 5: For exponential L and the same mean $E[S]$, isolation probability ϕ_c under constant stabilization delays S is no greater than that under any random S, where:

$$\lim_{E[S]\to 0} \phi_c = \frac{\rho\rho!}{(\rho + r)!}$$

$$\rho = E[L]/E[S]$$

- The above result shows that using constant intervals is not only a simple but optimal method to stabilize successors in Chord

Computer Science, Texas A&M University

Wrap-up

- We derived formulas for the resilience of Chord's successor lists
  – Under static node failure, Chord enjoyed the same resilience through the successor list as many other P2P networks though their neighbor sets
  – Under dynamic node failure, we used the Chen-Stein method to show that isolations of individual peers can be *treated* as independent when system size and successor lists become large
- We analyzed the effect of periodic stabilizations
  – Stabilization with constant intervals was optimal and kept Chord connected with the highest probability