Graph-Theoretic Analysis of Structured Peer-to-Peer Systems: Routing Distances and Fault Resilience

Presented by Seong-Ryong Kang

Dmitri Loguinov, Anuj Kumar,
Vivek Rai, and Sai Ganesh

Computer Science
Texas A&M University
College Station, TX 77843

Overview

- Motivation
- Optimal-diameter graphs
- Routing analysis
  - Shortest path distributions
- Resilience analysis (brief overview)
- Incremental construction of de Bruijn graphs
- Conclusion

Motivation

- Peer-to-peer (P2P) networks are important elements of the existing Internet
- Many recent proposals address the issue of constructing efficient DHTs (Distributed Hash Tables)
- However, two important pieces of analysis are missing from current work:
  - Comparison of existing methods with each other
  - Full understanding of their "optimality"
- Our work aims to fill this void

Motivation 2

- Traditional DHTs (CAN, Chord, Pastry, Tapestry) are graphs with logN diameter and logN degree
  - N is the number of peers in the network
- Can these logarithmic bounds be improved?
- Other important questions:
  - Which existing proposal is "best?"
  - What is the best possible diameter for a given degree?
  - Can fault resilience of existing methods be improved?
  - Can resilience and diameter be optimized at the same time?

Optimal Diameter

- Consider a problem of building a fixed-degree graph on N nodes with the smallest diameter
  - Assume k is the fixed degree of each node
  - Since DHTs treat all peers equally, constant node degree is a realistic assumption
  - Heterogeneous DHTs are beyond the scope of this work
- The smallest diameter is achieved in directed Moore graphs and equals:

$$D_M = \left\lceil \log_k \left(N(k-1)+1 \right) \right\rceil - 1$$

Optimal Diameter 2

- One big problem with Moore graphs
  - Non-trivial Moore graphs do not exist
- Generalized de Bruijn graphs have the best known diameters
  - Diameter $\lceil \log_k N \rceil$
  - Asymptotically optimal
- Several versions exist, but only one allows efficient (greedy) routing rules
  - Imase and Itoh, 1981

Optimal Diameter 3

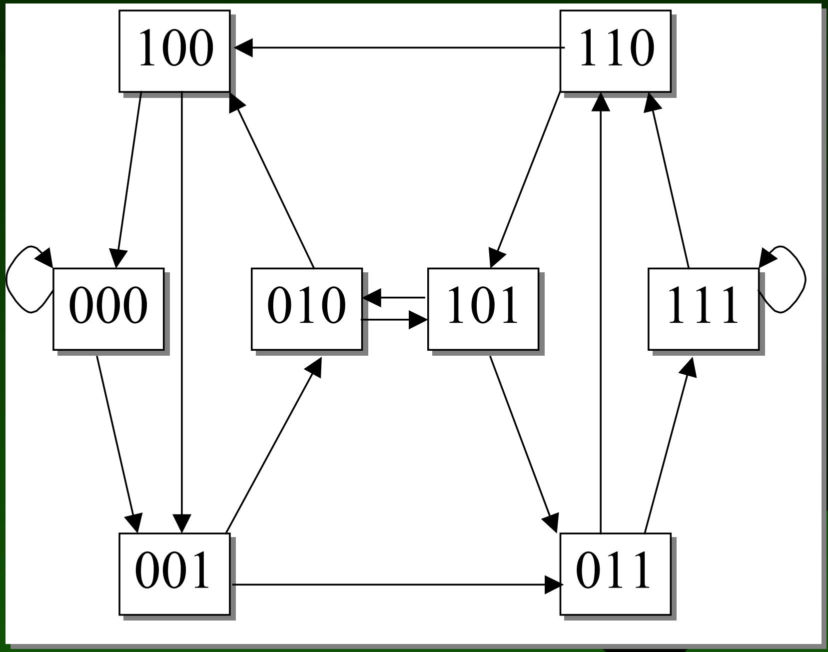
- Do we really care that de Bruijn graphs route faster than Chord/CAN/Pastry?
- Assume N = 1 million nodes
  - Chord's degree is $\log_2 N = 20$, diameter $\lceil \log_2 N \rceil = 20$
  - De Bruijn's degree k = 20, diameter $\lceil \log_{20} N \rceil = 5$
  - Moore graph of degree 20: diameter also 5 hops
- Improvement by a factor of 4 is significant

De Bruijn Graphs

- Recall that de Bruijn graphs have very simple linking rules:
  - Each node is a D-character string in some alphabet Σ
  - D is the diameter of the graph
  - Each node $(a_1, \ldots, a_D)$ links to all nodes $(a_2, \ldots, a_D, x)$, for all $x \in \Sigma$
  - Self-loops are acceptable
- For now assume that each graph is fully populated with N nodes
  - Incremental and distributed construction will be discussed later

De Bruijn Graphs 2

- Shortest-path routing is very simple and greedy
  - See the paper for details
- Classical de Bruijn graph on N = 8 nodes, degree k = 2 and diameter D = 3:



De Bruijn Graphs 3

- Next study degree-diameter tradeoffs of P2P graphs with $N = 10^6$ users:

Degree	de Bruijn	Chord	CAN	Pastry	Butterfly
2	20	–	500,000	–	31
3	13	–	–	–	20
10	6	–	40	–	10
20	5	20	20	20	8
50	4	–	–	7	7
100	3	–	–	5	5

De Bruijn Graphs 4

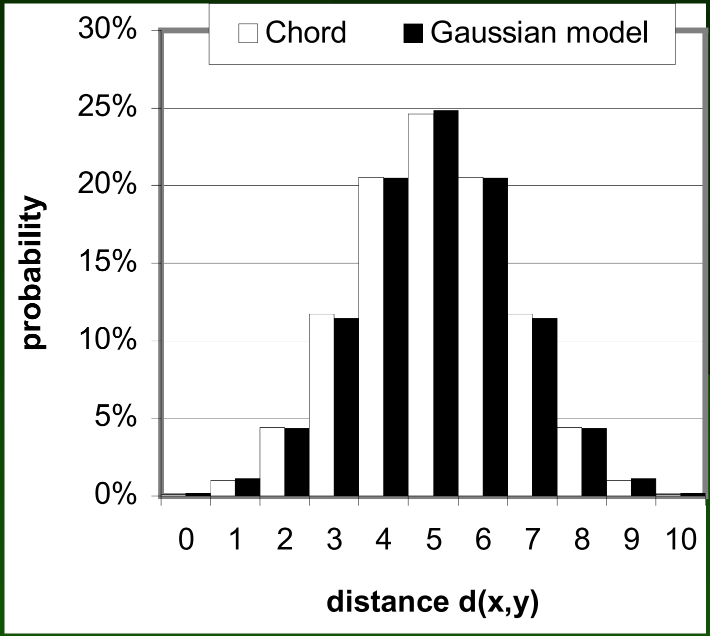
- Improvement in the diameter is significant over all existing structures
  - Even the butterfly networks offer diameter 50-60% larger than that of de Bruijn graphs
- The improvement is most noticeable in low-degree networks ($k < 20$)
  - Large neighbor tables require substantial maintenance and keep-alive traffic when peers frequently fail
  - Thus, small-degree graphs are often desirables
- *Asymptotically* (for very large N), de Bruijn graphs offer diameter D twice as small as any other graph in related work

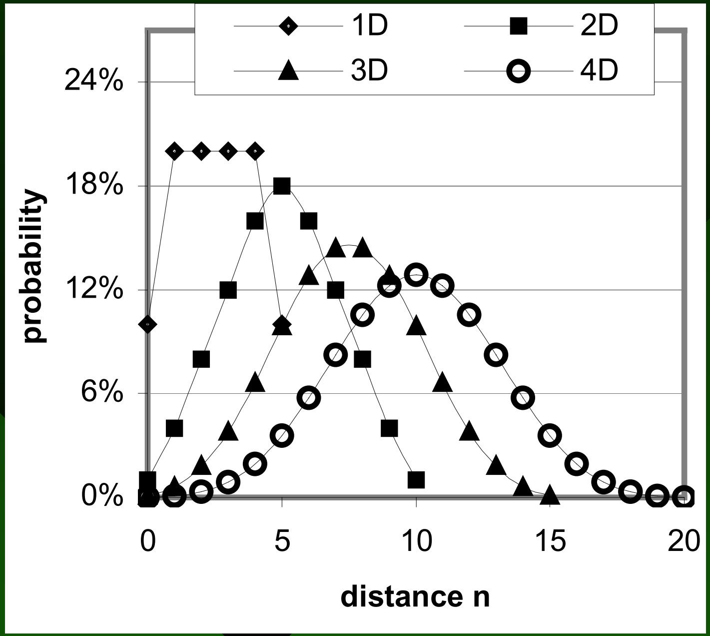
Routing Distances

- Next we analyze the average distance in each graph
  - This is the expected number of hops that each query must travel
- An important metric since there are graphs with diameters smaller than Chord's, but larger average distance
  - Xu *et al*., IEEE JSAC 2003
- We also compare Chord and CAN in this study
  - Which one is better?

Routing Distances 2

- Chord's distribution of shortest distances is known to be bell-shaped and appears Gaussian (left)
- CAN's distribution progressively becomes Gaussian as well (right)





Routing Distances 3

- Lemma 1: Chord's distribution of shortest distances is binomial with p = q = ½
  - Appears Gaussian for large D
- Lemma 2: CAN's distribution of shortest distances is a d-fold convolution of this simple 1D distribution (d is the number of dimensions):

$$p_1(n) = \frac{1}{N}\begin{cases} 1, & n = 0 \\ 2, & 0 < n < D \\ q(N), & n = D \\ 0, & otherwise \end{cases}$$

$$q(N) = \begin{cases} 1, & N = even \\ 2, & N = odd \end{cases}$$

Routing Distances 4

  - According to the Central Limit Theorem, self-convolution of $p_1(n)$ also appears Gaussian
- Now notice that if the number of dimensions d is $\log_2 N/2$, CAN's degree and diameter are the same as Chords
  - However, there is more to it
- Lemma 3: When $d = \log_2 N/2$, distribution of shortest distances in CAN and Chord are *identical*
  - Both graphs offer the same routing performance

Routing Distances 5

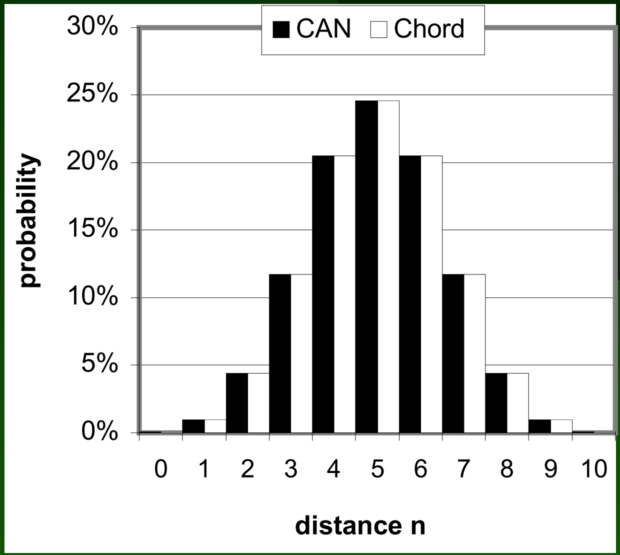
- De Bruijn graphs have a completely different routing structure
  - These graphs expand exponentially
- Lemma 4: The distribution of shortest distances (PMF) in de Bruijn graphs is:

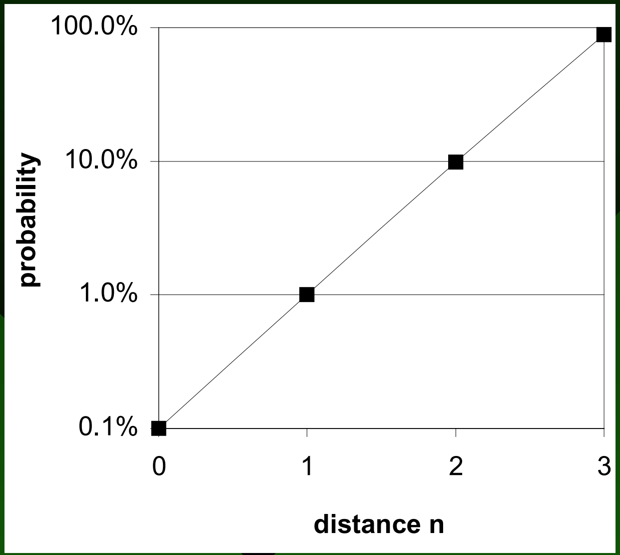
$$p(n) \approx \frac{k^n}{N} - \frac{k^{2n-1}}{N^2} \geq \frac{k^n - k^{n-1}}{N}.$$

- The number of nodes at distance n is approximately $k^n - k^{n-1}$

Routing Distances 6

- Simulations confirm that CAN and Chord for the same degree are identical (from the routing view) (figure below, left, N = 1,024)
  - However, they are not isomorphic
- De Bruijn graphs indeed expand exponentially (figure below, right, N = 1,000, k = 10)





Routing Distances 7

- Additional examples
  - The average distance μ_d in graphs of size $N = 10^6$

Degree	Moore	de Bruijn	Chord	CAN	Butterfly
2	17.9	18.3	–	250,000	22.4
3	11.7	11.9	–	–	14.7
10	5.8	5.9	–	19.8	7.3
20	4.5	4.6	10	10	5.7
50	3.5	3.5	–	–	4.3
100	2.98	2.98	–	–	3.65

Routing Distances 8

- Exponential expansion in de Bruijn graphs leads to
  - Small diameter
  - Very few short cycles
  - Low clustering
- Non-existence of short cycles means that alternative (parallel) paths to destinations do not overlap
- This further leads to better resilience to edge and node failure as the graph is tightly packed
  - We verify this in the paper

Routing Distances 9

- Additional advantage of smaller average distance is the increased capacity of the network
  - For each useful request, peers need to forward (on average) μ_d other requests
  - Thus, the capacity of the graph is inverse proportional to the average distance (similar to wireless networks)
- De Bruijn graphs offer $\log_2 \log_2 N/2$ times more capacity than Chord/CAN
- Asymptotically, 50% more than the butterfly
  - For N = 10^6, 22% more

Omitted Material

- We derive clustering coefficients of each graph
- We perform a simple expansion analysis of each graph and generalize clustering to become global
- We further show that de Bruijn graphs have bisection width larger than Chord's by a factor of $\log_2 \log_2 N/2$
- All these findings point toward higher resilience and better performance of de Bruijn graphs under node/edge failure
- We finally study the probability that a vertex appears in multiple parallel paths, per-node distribution of the number of non-overlapping shortest paths, and routing performance of these graphs under adversarial failure

ODRI

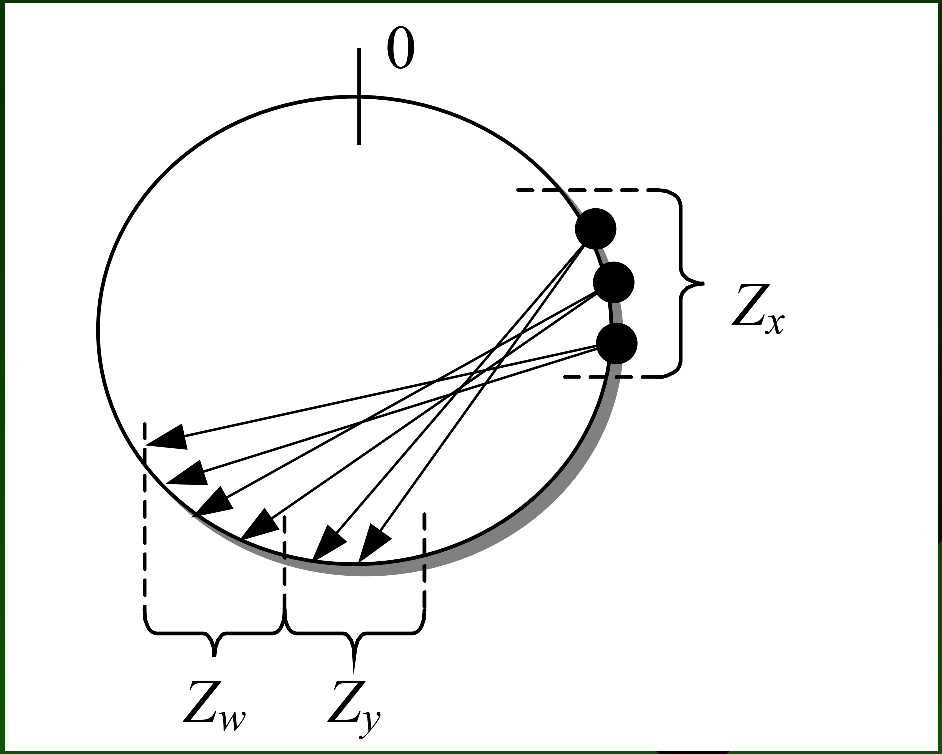
- We finish this talk by discussing incremental construction of de Bruijn graphs
  - ODRI – Optimal Diameter Routing Infrastructure
- Several other papers concurrently proposed de Bruijn graphs
  - Koorde, Kaashoek *et al*., 2003
  - Distance Halving, Naor *et al*., 2003
  - D2B, Fraigniaud *et al*., 2003
- Our construction is not substantially different

ODRI 2

- Organize all peers into a modulo-N_{max} circle
  - N_{max} is some upper limit on the number of users
- This circle represents the underlying de Bruijn graph that is split into zones by arriving users
  - Assume that degree k is known and N_{max} is some power of k
- Each zone $Z_x = [z_1, z_2]$ held by peer x contains a certain number of de Bruijn vertices (all integers between z_1 and z_2)
  - Each vertex $v \in [z_1, z_2]$ links to k other de Bruijn vertices

ODRI 3

- Peer x then links to all peers holding the other end of each edge originating in Z_x
  - In the figure, degree k = 2 and x links to peers y and w



ODRI 4

- It is easy to demonstrate that if all zones are the same, then the *application-layer* diameter is optimal and the degree of each peer is exactly k
- Under a uniform hashing function, zone distributions are not equal
  - However, the diameter is still asymptotically optimal
- Simple join method (e.g., Chord, CAN): a joining peer generates a random number and joins the ring at that location (splitting an existing node in half or otherwise)
  - Imbalance by a factor of logN with high probability

ODRI 5

- "Power of two choices" method: sample d locations in the graph and split the largest peer
- If the number of sampled locations is ~logN, then it can be guaranteed that the imbalance stays within a constant factor (usually 2) from the optimal
  - This method is implemented in Distance Halving (d = 8logN peers) and D2B (unspecified d)
- ODRI has its own variation of this method
  - Start from a random location and then walk through the graph searching for the largest node to split
  - Reduced join latency as d messages can sample d·k peers (where k is the degree as before)

ODRI 6

- Example
  - N = 30,000, k = 8
  - Traditional methods require over 400 messages to sample 82 peers, while ODRI needs only 10
- To further improve the search, ODRI is biased towards the largest neighbor at each step
  - Larger nodes “cover” more DHT space with their edges and are thus more likely to “know” other large nodes
- Loops are prevented by appending the entire path to each request packet

ODRI 7

- Node departure can re-introduce imbalance in zone distributions and actually make it worse
- Thus, each departing node x performs a d-walk searching for the smallest node to take its place
  - Once found, this smallest node y will take over x’s zone
  - Successor/predecessor of y will take over its zone
- The d-walk is still biased towards the largest neighbor at each step
  - Same reasoning as before
  - Performs very well in practice

Conclusion

- Details of these algorithms and probabilistic analysis will be presented in the next paper
  - "Evolution of Massive P2P Graphs: Zone Distribution Perspective"
- Our results in the current paper indicate that de Bruijn graphs offer an appealing framework for P2P networks
- Their diameter and average distance are smaller than that of any alternative graph
  - Their bisection width and expansion are higher than that of Chord/CAN and no worse than that of the butterfly

Conclusion 2

- De Bruijn graphs are much easier to construct incrementally than other fixed-degree graphs (e.g., the butterfly)
- They exhibit very little path overlap, clustering, and susceptibility to node failure
- Nevertheless, the bisection width of de Bruijn graphs is far from optimal
  - Thus, one final question remains: is it possible to simultaneously optimize resilience (e.g., bisection width) and diameter?