On Static and Dynamic Partitioning Behavior of Large-Scale Networks

Derek Leonard, Zhongmei Yao, Xiaoming Wang,
and Dmitri Loguinov

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

November 9th, 2005

Outline

Outline

1. Introduction
   - Background
2. Global Resilience
   - Generic Global Resilience Model
   - Static Node Failure
   - Dynamic Node Failure
3. Local Resilience
   - Static Node Failure
   - Dynamic Node Failure

Background I

Resilience of Peer-to-Peer Networks

The resilience of P2P networks to disconnection (partitioning) is important to their design and performance

Background I

Resilience of Peer-to-Peer Networks

The resilience of P2P networks to disconnection (partitioning) is important to their design and performance

Two Types of Resilience

- The *local resilience* of a graph G we define to be the probability that a node v is isolated (has no neighbors) from the system

Background I

Resilience of Peer-to-Peer Networks

The resilience of P2P networks to disconnection (partitioning) is important to their design and performance

Two Types of Resilience

- The *local resilience* of a graph G we define to be the probability that a node v is isolated (has no neighbors) from the system
- *Global resilience* refers to the probability that G is a connected graph

Background II

Node Failure

To study the two kinds of resilience, some sort of random node failure is imposed on G

Two Types of Node Failure

- The most popular approach we call *static node failure*, where each node is simultaneously removed from the system with independent probability p

Background II

Node Failure

To study the two kinds of resilience, some sort of random node failure is imposed on G

Two Types of Node Failure

- The most popular approach we call *static node failure*, where each node is simultaneously removed from the system with independent probability p
- A more realistic approach is to study P2P networks under random arrival/departure of nodes, which we call *dynamic node failure*

Background III

Random Graph Connectivity

- Erdös and Rényi demonstrated that *almost every* (i.e., with probability $1 - o(1)$ as $n \to \infty$) random graph is connected if and only if it has no isolated vertices

Background III

Random Graph Connectivity

- Erdös and Rényi demonstrated that *almost every* (i.e., with probability $1 - o(1)$ as $n \to \infty$) random graph is connected if and only if it has no isolated vertices
- Define $\Phi(G)$ to be the probability that graph G remains connected under node or edge failure

$$\Phi(G) = P(G \text{ has no isolated nodes})$$

Background III

Random Graph Connectivity

- Erdös and Rényi demonstrated that *almost every* (i.e., with probability $1 - o(1)$ as $n \to \infty$) random graph is connected if and only if it has no isolated vertices
- Define $\Phi(G)$ to be the probability that graph G remains connected under node or edge failure

$$\Phi(G) = P(G \text{ has no isolated nodes})$$

- Burtin [1977] and then later Bollobas [1983] prove that this holds under static node failure for the hypercube

Background IV

Combinatorial Result

Najjar and Gaudiot [1990] proposed a combinatorial model of connectivity for k-regular graphs under static node failure:

$$\Phi(G) = \sum_{i=0}^{n} Q_i \binom{n}{i} p^i (1-p)^{n-i},$$

where

$$Q_i = \prod_{j=1}^{i} \left[1 - \frac{k(n-k-1)!(j-1)!(n-j)}{(n-1)!(j-k)!} \right].$$

Outline

Generic Global Resilience Model I

Properties for Connectivity

- Recall that $\Phi(G)$ is defined as the probability that G remains connected under node or edge failure:

$$\Phi(G) = P(G \text{ has no isolated nodes})$$

- To satisfy this expression, the number of edges leaving *each* set of nodes S must be an *increasing* function of set size $|S|$

Generic Global Resilience Model I

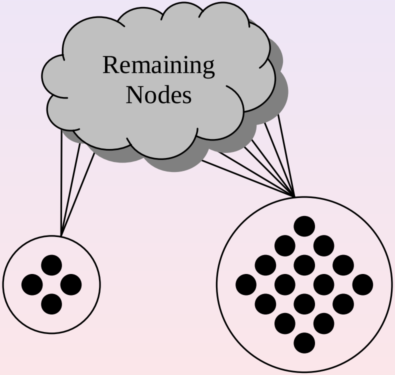
Properties for Connectivity

- Recall that $\Phi(G)$ is defined as the probability that G remains connected under node or edge failure:

$$\Phi(G) = P(G \text{ has no isolated nodes})$$

- To satisfy this expression, the number of edges leaving *each* set of nodes S must be an *increasing* function of set size $|S|$
- Burtin [1977] showed that hypercubes have this property by proving that larger sets S are *always* better connected than smaller sets

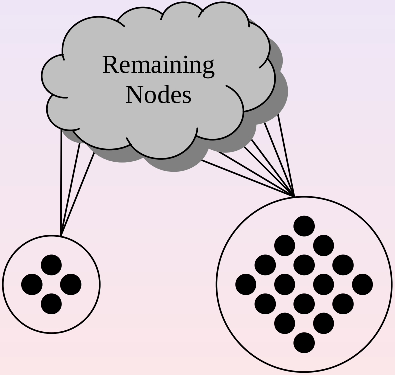
Generic Global Resilience Model II



Illustration

- The figure illustrates the connectivity of node sets in hypercube and similar random graphs

Generic Global Resilience Model II



Illustration

- The figure illustrates the connectivity of node sets in hypercube and similar random graphs
- The probability of any large subgraph disconnecting after node failure is negligible compared to that of an individual node

Generic Global Resilience Model III

Proposition

If a graph G has node expansion properties no worse than those of hypercubes or random graphs of the same size, it will remain almost surely connected under random node failure if it has no isolated nodes

Generic Global Resilience Model III

Proposition

If a graph G has node expansion properties no worse than those of hypercubes or random graphs of the same size, it will remain almost surely connected under random node failure if it has no isolated nodes

Applicable Graphs

- All DHTs that can be reduced to the hypercube

Generic Global Resilience Model III

Proposition

If a graph G has node expansion properties no worse than those of hypercubes or random graphs of the same size, it will remain almost surely connected under random node failure if it has no isolated nodes

Applicable Graphs

- All DHTs that can be reduced to the hypercube
- Graphs with better expansion than the hypercube

Generic Global Resilience Model III

Proposition

If a graph G has node expansion properties no worse than those of hypercubes or random graphs of the same size, it will remain almost surely connected under random node failure if it has no isolated nodes

Applicable Graphs

- All DHTs that can be reduced to the hypercube
- Graphs with better expansion than the hypercube
- Random graphs where users rely on random selection of neighbors during join

Generic Global Resilience Model IV

Summary

- For graphs that have properties described in the previous slide:

$$\Phi(G) = P(G \text{ has no isolated nodes})$$

Generic Global Resilience Model IV

Summary

- For graphs that have properties described in the previous slide:

$$\Phi(G) = P(G \text{ has no isolated nodes})$$

- This is certain for extremely large graphs, but we next verify its applicability to *smaller, finite* graphs through simulations
- We later derive accurate models for the probability of isolation under both static and dynamic node failure

Outline

Static Node Failure I

Simulations

- Recall that static node failure is done by removing each node from the graph independently with probability p

Static Node Failure I

Simulations

- Recall that static node failure is done by removing each node from the graph independently with probability p
- We simulated a number of graphs using $100,000$ failure patterns for each value of p

Table: Chord with $n = 16384$ under static node failure

p	$\Phi(G)$	P(no isolated nodes)
0.5	0.99996	0.99996
0.6	0.99354	0.99354
0.7	0.72619	0.72650
0.8	0.00040	0.00043

Outline

Dynamic Node Failure I

Lifetime-based Node Failure

- What can be said about node-failure in real-world P2P systems?
- Nodes arrive/depart dynamically instead due to individual behavior

Dynamic Node Failure I

Lifetime-based Node Failure

- What can be said about node-failure in real-world P2P systems?
- Nodes arrive/depart dynamically instead due to individual behavior
- *Model*: we assign each user a random lifetime L_i from a distribution $F(x)$ the reflects the behavior of the user and represents the duration of his/her service to the system

Dynamic Node Failure II

Model Assumptions

- *Arrival*: nodes arrive randomly according to any process; however, their arrival times are uncorrelated with lifetimes of existing nodes

Dynamic Node Failure II

Model Assumptions

- *Arrival*: nodes arrive randomly according to any process; however, their arrival times are uncorrelated with lifetimes of existing nodes
- *Departure*: nodes deterministically die (fail) after spending L_i time units in the system

Dynamic Node Failure II

Model Assumptions

- *Arrival*: nodes arrive randomly according to any process; however, their arrival times are uncorrelated with lifetimes of existing nodes
- *Departure*: nodes deterministically die (fail) after spending L_i time units in the system
- *Neighbor selection*: neighbors are picked from among the existing nodes using any rules that do not involve node lifetimes or age (e.g., based on random walks, etc.)

Dynamic Node Failure II

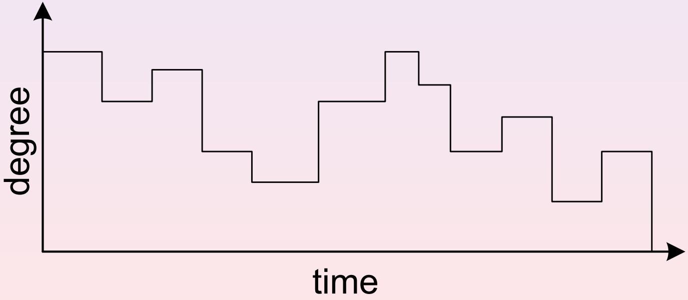
Model Assumptions

- *Arrival*: nodes arrive randomly according to any process; however, their arrival times are uncorrelated with lifetimes of existing nodes
- *Departure*: nodes deterministically die (fail) after spending L_i time units in the system
- *Neighbor selection*: neighbors are picked from among the existing nodes using any rules that do not involve node lifetimes or age (e.g., based on random walks, etc.)
- *Neighbor replacement*: once a failed neighbor is detected, a replacement search is performed

Dynamic Node Failure III

Definition

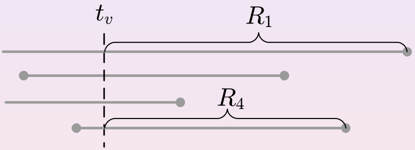
A node becomes isolated when all of the neighbors in its table are in the failed state



Dynamic Node Failure IV

Lifetimes of Neighbors

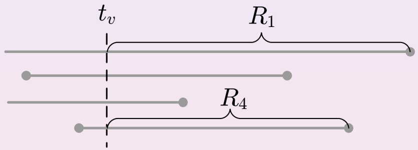
The k neighbors of node v are represented by residual lifetimes



Dynamic Node Failure IV

Lifetimes of Neighbors

The k neighbors of node v are represented by residual lifetimes



Definition

Let R_i be the remaining lifetime of neighbor i when v joined the system

Dynamic Node Failure V

Formalizing Search Time

- When a neighbor fails, there is usually some detection mechanism (e.g., periodic probing)
- Systems often repair the failed zone of a DHT or find a random replacement neighbor in unstructured systems
- We allow this process to be arbitrary as the technique employed has no effect on our results

Dynamic Node Failure V

Formalizing Search Time

- When a neighbor fails, there is usually some detection mechanism (e.g., periodic probing)
- Systems often repair the failed zone of a DHT or find a random replacement neighbor in unstructured systems
- We allow this process to be arbitrary as the technique employed has no effect on our results

Definition

Let S_i be a random variable describing the total search time for the i-th replacement in the system

Dynamic Node Failure VI

Definition

Define Z to be the random time (in terms of user joins) when a graph G disconnects for the first time under the lifetime model

Dynamic Node Failure VI

Definition

Define Z to be the random time (in terms of user joins) when a graph G disconnects for the first time under the lifetime model

Proposition

The probability that the graph *stays connected* for more than N user joins is almost surely:

$$P(Z > N) = (1 - \phi)^N,$$

where ϕ is the probability of isolation for a single node

Dynamic Node Failure VII

Simulations

Simulations results are below with ϕ computed empirically

Table: Simulations for 12-regular CAN with $N = 10^6$ and degree-irregular Chord with $k \approx 13$ and $N = 50,000$

Search time	CAN	Model	Chord	Model
6 mins	.9732	.9728	.6295	.6251
7.5 mins	.8118	.8124	.3284	.3184
8.5 mins	.5669	.5659	.2189	.2206
9 mins	.4065	.4028	.1460	.1483
9.5 mins	.2613	.2645	.1211	.1274
10.5 mins	.0482	.0471	.0493	.0493

Dynamic Node Failure VIII

Implications

- Using 6-minute replacement delays, CAN has a 97% chance of surviving *1 million* user joins without disconnecting

Dynamic Node Failure VIII

Implications

- Using 6-minute replacement delays, CAN has a 97% chance of surviving *1 million* user joins without disconnecting
- For reasonably small search delays, network partitioning in lifetime-based systems almost surely effects only *a single node* in the system

Dynamic Node Failure VIII

Implications

- Using 6-minute replacement delays, CAN has a 97% chance of surviving *1 million* user joins without disconnecting
- For reasonably small search delays, network partitioning in lifetime-based systems almost surely effects only *a single node* in the system
- Note that in these simulations ϕ is calculated empirically, we later derive a model for ϕ and show that it is very accurate as well

Outline

Static Node Failure Model I

Preliminaries

- We now develop a simple closed-form model for the probability that there are no isolated nodes under *static node failure*

Static Node Failure Model I

Preliminaries

- We now develop a simple closed-form model for the probability that there are no isolated nodes under *static node failure*
- Let p_i be the probability that node i is isolated and alive after the failure

Static Node Failure Model I

Preliminaries

- We now develop a simple closed-form model for the probability that there are no isolated nodes under *static node failure*
- Let p_i be the probability that node i is isolated and alive after the failure

Proposition

The number of isolated vertices X tends to a Poisson distribution with mean $\lambda = \sum_i p_i$ and the probability $\Phi(G)$ of having a connected graph converges to $e^{-\lambda}$ with probability 1 as $n \longrightarrow \infty$

Static Node Failure Model II

Table: Simulation results and model for two regular graphs.

p	Chord $n = 16384$, $k = 27$			de Bruijn $n = 20736$, $k = 24$		
	$\Phi(G)$	Model	Najjar	$\Phi(G)$	Model	Najjar
.5	.9999	.9999	.9982	.9993	.9994	.9940
.55	.9992	.9993	.9976	.9944	.9945	.9892
.6	.9935	.9933	.9916	.9618	.9615	.9550
.65	.9500	.9503	.9463	.7954	.7907	.7750
.7	.7262	.7239	.7055	.3199	.3037	.2737
.75	.1788	.1766	.1501	.0079	.0055	.0033
.8	.0004	.0004	.0002	0	10^{-9}	10^{-10}

Static Node Failure Model III

Observations

- The model is very accurate for all values of disconnection probability p
- The more complex result of Najjar and Gaudiot is less accurate than the model

Static Node Failure Model III

Observations

- The model is very accurate for all values of disconnection probability p
- The more complex result of Najjar and Gaudiot is less accurate than the model

Next step

We now show that our model also applies to *irregular* graphs, which has not been attempted previously to our knowledge

Static Node Failure Model IV

Table: Simulation results and model for three irregular graphs.

p	Symphony		Gnutella		Randomized Chord	
	$\Phi(G)$	Model	$\Phi(G)$	Model	$\Phi(G)$	Model
.45	.9998	.9996	.9661	.9666	.9999	.9999
.5	.9977	.9977	.8626	.8646	.9997	.9997
.55	.9875	.9875	.5804	.5829	.9975	.9976
.6	.9391	.9394	.1708	.1700	.9844	.9845
.65	.7552	.7535	.0055	.0053	.9162	.9151
.7	.3115	.3107	0	10^{-7}	.6375	.6372
.75	.0127	.0122	0	10^{-15}	.1299	.1282
.8	0	10^{-7}	0	10^{-34}	.0003	.0002

Outline

1. Introduction
   - Background
2. Global Resilience
   - Generic Global Resilience Model
   - Static Node Failure
   - Dynamic Node Failure
3. Local Resilience
   - Static Node Failure
   - Dynamic Node Failure

Dynamic Node Failure Model I

Exact Numerical Solution

- What is the probability ϕ that a node will become isolated from the network during its lifetime?

Dynamic Node Failure Model I

Exact Numerical Solution

- What is the probability ϕ that a node will become isolated from the network during its lifetime?
- For exponential lifetimes and exponential search delays, the probability of isolation is:

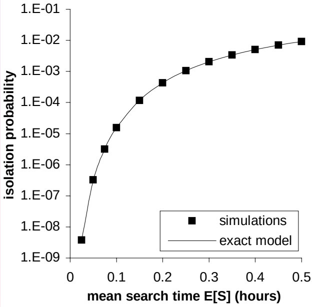
$$\phi = \sum_{j=1}^{k} \frac{(\delta \mathbf{v}_j)(\mathbf{u}_j^T \mathbf{r})}{\mu + \xi_j},$$

where $\mu = 1/E[L_i]$, ξ_j is the j-th eigenvalue of a scaled rate matrix R, $\delta = (0, 0, \ldots, 1)$, and $\mathbf{v}_j$ and $\mathbf{u}_j$ are eigenvectors of R

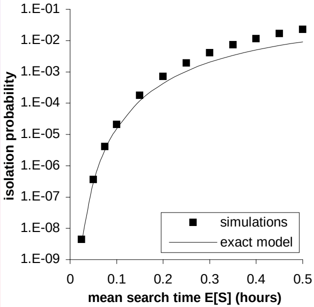
Dynamic Node Failure Model II

Simulations

Simulations with $E[L_i] = 0.5$ and $k = 8$

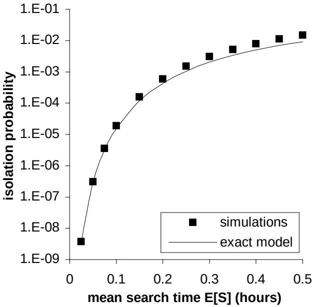


(a) exponential S_i

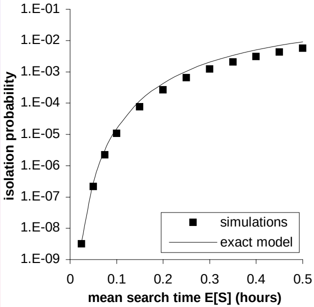


(b) constant S_i

Dynamic Node Failure Model III



(c) uniform S_i



(d) Pareto S_i with $\alpha = 3$

Dynamic Node Failure Model IV

Asymptotic Expansion

This result requires numerical manipulation to calculate, so we simplify the model for ϕ to the following

Dynamic Node Failure Model IV

Asymptotic Expansion

This result requires numerical manipulation to calculate, so we simplify the model for ϕ to the following

Proposition

For asymptotically small exponential search delays and exponential lifetimes:

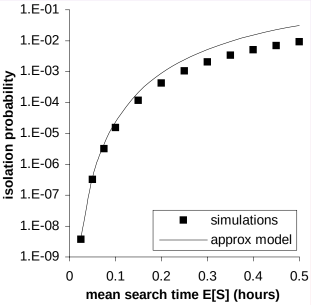
$$\phi = \frac{\rho k}{(1+\rho)^k + \rho k - 1} + o(1),$$

where $\rho = E[L_i]/E[S_i]$

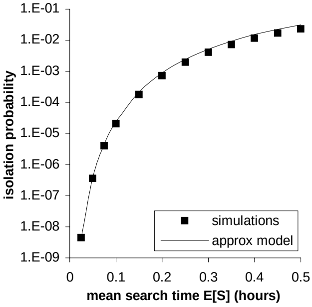
Dynamic Node Failure Model V

Simulations

Simulations with $E[L_i] = 0.5$ and $k = 8$

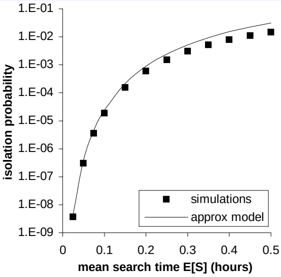


(e) exponential S_i

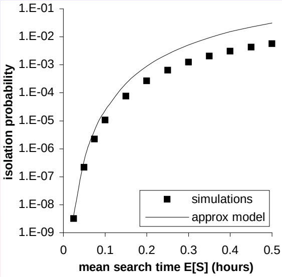


(f) constant S_i

Dynamic Node Failure Model VI



(g) uniform S_i



(h) Pareto S_i with $\alpha = 3$

Simulations

As $E[S_i]$ becomes small the simulations converge to the model

Dynamic Node Failure Model VII

Convergence of Asymptotic Expansion to Numerical Result

As search delay decreases, the asymptotic expansion for ϕ approaches the numerical solution

Table: Exponential search delays, $E[L_i] = 0.5$ and $k = 8$

$E[S_i]$	Numeric ϕ	Asymptotic ϕ	Ratio
1 hour	3.2480×10^{-2}	1.3971×10^{-1}	4.3017
6 min	1.5379×10^{-5}	2.3814×10^{-5}	1.5485
36 sec	8.2856×10^{-12}	8.7397×10^{-12}	1.0548
3.6 sec	1.0023×10^{-18}	1.0078×10^{-18}	1.0054
360 ms	1.0218×10^{-25}	1.0224×10^{-25}	1.0006

Dynamic Node Failure Model VIII

Heavy-tailed Lifetimes

This result also applies to heavy-tailed lifetime distributions

Dynamic Node Failure Model VIII

Heavy-tailed Lifetimes

This result also applies to heavy-tailed lifetime distributions

Proposition

For an arbitrary distribution of search delays and any lifetime distribution $F(x)$ with an exponential or heavier tail, the following holds:

$$\phi \leq \frac{\rho k}{(1+\rho)^k + \rho k - 1},$$

where $\rho = E[L_i]/E[S_i]$

Global Resilience under Dynamic Failure I

Global Resilience Model

Using this result, we now revisit global resilience under dynamic node failure previously demonstrated with simulated ϕ

Global Resilience under Dynamic Failure I

Global Resilience Model

Using this result, we now revisit global resilience under dynamic node failure previously demonstrated with simulated ϕ

Proposition

The global resilience of a graph G is lower bounded by:

$$P(Z > N) = (1 - \phi)^N \geq \left(1 - \frac{\rho k}{(1+\rho)^k + \rho k - 1}\right)^N,$$

where Z is the number of user joins before the first disconnection of the system

Global Resilience under Dynamic Failure II

Simulations

12-regular CAN with exponential lifetimes, $E[L_i] = 0.5$ hours, $n = 4096$, and $N = 10^6$ user joins

Table: Comparison of $P(Z > N)$ in CAN

Fixed search time (min)	Actual $P(Z > N)$	Model
6	.9732	.9728
7.5	.8118	.8215
8.5	.5669	.5666
9.5	.2613	.2419
10.5	.0482	.0424

Global Resilience under Dynamic Failure III

Example

- Consider a system with $k = 12$, $E[S_i] = 1$ minute, and $E[L_i] = 0.5$ hours

Global Resilience under Dynamic Failure III

Example

- Consider a system with $k = 12$, $E[S_i] = 1$ minute, and $E[L_i] = 0.5$ hours
- The probability of isolation $\phi = 4.57 \times 10^{-16}$

Global Resilience under Dynamic Failure III

Example

- Consider a system with $k = 12$, $E[S_i] = 1$ minute, and $E[L_i] = 0.5$ hours
- The probability of isolation $\phi = 4.57 \times 10^{-16}$
- If 35 million users join and leave the system each week, the probability that the network survives for *10,000 years* without disconnecting is *at least 99.2%*

Global Resilience under Dynamic Failure III

Example

- Consider a system with $k = 12$, $E[S_i] = 1$ minute, and $E[L_i] = 0.5$ hours
- The probability of isolation $\phi = 4.57 \times 10^{-16}$
- If 35 million users join and leave the system each week, the probability that the network survives for *10,000 years* without disconnecting is *at least 99.2%*
- The further implication is that the mean delay between disconnections is lower bounded by $1/\phi$ user joins, or *1.2 million years*

Wrap-up

Wrap-up

- Classical results on graph disconnection apply to most P2P networks under both static and dynamic node failure

Wrap-up

Wrap-up

- Classical results on graph disconnection apply to most P2P networks under both static and dynamic node failure
- In the case of both static and dynamic node failure, global resilience has been effectively reduced to local resilience

Wrap-up

Wrap-up

- Classical results on graph disconnection apply to most P2P networks under both static and dynamic node failure
- In the case of both static and dynamic node failure, global resilience has been effectively reduced to local resilience
- *P2P systems are more resilient than we thought*