

## P2P ネットワークから消滅するノードの 検出手法

小林 守<sup>†</sup> 大野 博<sup>††</sup> 渋谷 進<sup>††</sup> 米倉 達広<sup>††</sup>

**概要:** 非構造型 P2P ネットワークでは、ノードがネットワークに自由に参加/離脱できるため、ファイルがネットワークから消失して、安定したファイル共有が行われにくい場合がある。本報告では、P2P ネットワークにおけるノード寿命が超指数分布に従うときのノードの孤立をモデル化し、ノードの孤立時間と孤立確率を求めている。また、指数分布、パレート分布、超指数分布の寿命を持つノードの孤立時間と孤立確率を数値的に求めて、比較している。超指数分布のパラメータの値を調整することによって、超指数分布を用いてパレート分布を近似することができる。

### Lifetime Analysis of Nodes to Disappear from P2P Network

Mamoru Kobayashi<sup>†</sup> Hiroshi Ono<sup>††</sup>  
Susumu Shibusawa<sup>††</sup> Tatsuhiro Yonekura<sup>††</sup>

**Abstract:** In unstructured P2P network, a steady file sharing might not be done easily because nodes can join and leave to the network freely and files sometimes disappear from the network according to the node leave. This report models the isolation of nodes in the P2P network when the node lifetime is under hyper-exponential distribution, and presents the isolation time and the isolation probability of nodes. We also show the numerical values of the isolation time and the isolation probability for exponential distribution, Pareto distribution and hyper-exponential distribution, and compare them. Adjusting the values of parameters of hyper-exponential distribution, Pareto distribution can be approximated by hyper-exponential distribution.

<sup>†</sup> 茨城大学大学院理工学研究科

<sup>††</sup> 茨城大学工学部

## 1. はじめに

構造型の P2P ネットワークは、より確実にノードを管理することにより安定して稼働することができる。非構造型 P2P ネットワークは、各ノードの性能が異なる上にネットワークに自由に参加/離脱することができるために、安定した稼働が得られにくい。

非構造型ネットワークで安定した稼働を行うために、大きく 3 つの手法が検討されている。1 つはネットワーク上をクエリに合わせて通過するノードに複製配置する手法である<sup>1),2)</sup>。2 つ目はノードを複数のグループに分けて、そのグループ内でノードの情報を補完することにより、予告なしに発生する離脱に対し他のノードが代行する手法である<sup>10)</sup>。また 3 つ目は、ネットワーク内のノードの余寿命あるいは孤立確率から、ノードが離脱する前にノードの情報を隣接のノードに複製する手法である<sup>3),4),5),6),8),9)</sup>。

ノード情報の複製手法により、多くの場合、ノードの離脱によってノードが有する情報が消失してしまうのを防ぐことが可能となる。しかし非構造型 P2P ネットワークでは、ノードの性能が低いために、しばしば他のノード情報を保持できず、またノードの情報を共有する規約がなく、参加や離脱が自由にできるため、ノードの情報を完全に複製配置することが難しい。これらの制限の下でノード情報の消失を防ぐために、インターネットにおけるノードの動向を分析し、離脱する可能性の高いノードを検出して、そのノードが消滅する前に複製する方法が考えられる。

本稿では、P2P ネットワークにおけるノード寿命が超指数分布に従うときのノードの孤立をモデル化し、ノードの孤立時間と孤立確率を求める。また、指数分布、パレート分布、超指数分布の寿命を持つノードの孤立時間と孤立確率を数値的に求めて比較する。本稿では、2 章で関連研究、3 章で寿命に基づいたノード離脱のモデル化、4 章で孤立ノードの検出、5 章でまとめを述べる。

## 2. 関連研究

P2P ネットワークにおける基本的な複製配置手法には、ネットワーク上のクエリに合わせて、通過するノードに複製配置する Owner Replication, Path Replication, Random Replication 法がある<sup>1)</sup>。Cohen ら<sup>2)</sup> は、非構造化 P2P ネットワークにおける一様配置と比例配置がクエリに対して等しい平均性能を示し、最適な方法は、その 2 つの中間にあることを示した。Liben ら<sup>3)</sup> は、Chord ネットワークのノードが参加/離脱するとき、ネットワークを接続するのに必要なノードの比率の下界を示した。Kageyama ら<sup>7)</sup> は、需要予測に基づいた複製配置手法を提案し、需要の少ないファイルに対し需要を予測する手法により複製配置を行っている。安部<sup>10)</sup> は、グループ内のノード間で一貫性を保つために、Paxos 同意プロトコルを用いてノードが離脱してもプロセス移送技術により別のノードに複製する手法で、グループ内のノード数を確保することにより情報の消失を防いでいる。

Sutzbach ら<sup>4)</sup> は、3 つの P2P ネットワークのモデルにおけるノードの参加/離脱について検討を行い、その結果、セッションの長さが指数分布よりワイブル分布の方が一致することを報告している。また Gnutella ネットワークが高い弾力性をもつトポロジである

ことを示した. そして Stutzbach ら<sup>9)</sup>は, ネットワークの参加/離脱問題における特徴として, セッション時間が軸の長いべき乗則分布で表わされることを示した. Leonard ら<sup>5)</sup>は, ノードの参加/離脱とネットワーク間の関係を導くために, ノードの孤立確率や時間による弾力性について, 指数分布とパレート分布を用いて提案した. Yao ら<sup>6)</sup>は, 非構造型 P2P ネットワークのノード孤立モデルを超指数分布関数を用いて提案しているが, ノードの孤立確率を求めている.

上記のように, ネットワークにおけるクエリの動きを利用して複製配置して対処する方法, 複数ノードのグループ化により互いに情報を補完することで予告なしの離脱に対処する方法, ネットワークのグラフ理論に基づいて隣接のノードに複製配置して対処する手法がある. しかし非構造型 P2P ネットワークにおいて, クエリの要求に従って通過したノードにしか複製しない, あるいは異なるノード性能のグループ内でノード情報を複製補完するだけでは, 複製配置を完全に行うことができる可能性が少ない.

### 3. 寿命に基づいたノード離脱のモデル化

P2P ネットワークにおけるノードの消滅は, ネットワークを構成するノードが制限されず自由に参加/離脱することにより発生する. 特に非構造型 P2P ネットワークでは, 各ノードが有する情報がそれぞれ異なるために, そのノードが消滅するとネットワークとしてノード情報が消失することに繋がる. Leonard ら<sup>5)</sup>は, ネットワーク内のノードの寿命モデルを仮定するために,  $k$  正則 P2P グラフを考え, 他の隣接ノードが全て離脱したためにノード  $v$  が孤立する確率を解析している. そしてノード  $v$  が接続するネットワークは十分に長く存在すると仮定している.

ノード  $v$  がネットワークに参加したときの隣接ノード数を  $k$  とするとき, 隣接ノードの離脱とともにノード  $v$  の次数が減少する. 隣接ノードを補わないとき, 時間  $t$  とともに変化するノード  $v$  の次数  $w(t)$  は再生過程となる. これを図 1 に示す.

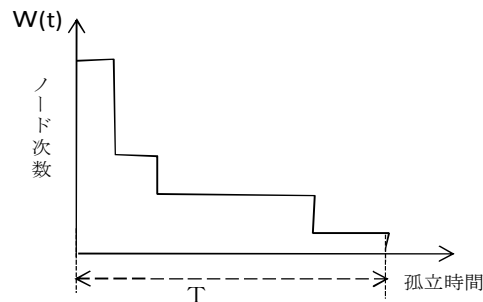


図 1 次数の変化の過程

時間  $t_v$  に隣接ノード  $v$  がネットワークに参加し, 隣接ノード  $i$  の参加時間と離脱時間をそれぞれ  $a_i$  と  $d_i$  とすると, ノード  $i$  の余寿命  $R_i$  を  $R_i = d_i - t_v$  と定義する. 参加時間と離脱時間およびノード  $i$  の余寿命  $R_i$  の関係を図 2 に示す.  $L_i$  は隣接ノード  $i$  の寿命を表す.

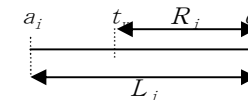


図 2 ノード  $i$  の参加時間と離脱時間および余寿命

#### 3.1 ノードの余寿命

ネットワーク内のノードの寿命は, 同じ分布関数  $F(t)$  に従うものとする. ネットワークのノード数  $n$  は, 十分大きく定常状態であると仮定する. このときノード  $v$  の隣接ノード  $i$  の余寿命  $L_i$  の分布関数は, 次のように表すことができる.

$$F_R(x) = \frac{1}{E[L_i]} \int_0^x (1 - F(z)) dz \quad (1)$$

ここに  $E[L_i]$  は, ノード  $i$  の寿命の期待値である.

ノードの寿命が指数分布の場合, 余寿命は次のように表される,

$$F_R(x) = 1 - e^{-\lambda x} \quad (2)$$

ここに  $\lambda$  は指数分布の係数である. この式は, 余寿命の分布関数が元の寿命の分布関数に等しくなることを示しており, これは指数分布の無記憶性による.

次にパレート分布の分布関数は, 次のように表される.

$$F(x) = 1 - \left(1 + \frac{x}{\beta}\right)^{-\alpha} \quad (\alpha > 1, \beta > 0) \quad (3)$$

パレート分布の寿命をもつノードの余寿命の分布関数は, 次のように表される<sup>5)</sup>.

$$F_R(x) = 1 - \left(1 + \frac{x}{\beta}\right)^{1-\alpha} \quad (4)$$

超指数分布の分布関数は、次のように表される。

$$F(x) = 1 - \sum_{i=1}^m p_i e^{-\lambda_i x} \quad (5)$$

ここに  $\lambda_i > 0$  は係数であり、 $p_i$  は各項の重みを表し、 $\sum_{i=1}^m p_i = 1$  である。

超指数分布の期待値は、次のように表される。

$$\mu = E[X] = \sum_{i=1}^m \frac{p_i}{\lambda_i} \quad (6)$$

超指数分布の寿命をもつノードの余寿命の分布関数について、次の性質が成り立つ。

**[補題 1]** 超指数分布の寿命をもつノードの余寿命の分布関数は、次のように表される。

$$F_R(x) = 1 - \frac{1}{E[L_i]} \sum_{i=1}^m \frac{p_i}{\lambda_i} e^{-\lambda_i x} \quad (7)$$

### 3.2 ノードの孤立時間と孤立確率

ノード  $v$  がネットワークに参加した後、ノード  $v$  がネットワークを離れる前に、ノード  $v$  のすべての隣接ノードが離脱する時間をノード  $v$  の孤立時間と呼び、 $T$  で表す。 $T$  は、ノード  $v$  のすべての隣接ノードの余寿命の最大値である。

孤立時間  $T$  の期待値  $E[T]$  は、次のように表される。

$$E[T] = \int_0^{\infty} (1 - (F_R(x))^k) dx \quad (8)$$

指数分布関数の寿命をもつノードの孤立時間の期待値は、次のように表される。

$$E[T] = \frac{1}{\lambda} \sum_{i=1}^k \frac{1}{i} \quad (9)$$

またパレート分布の寿命をもつノードの孤立時間の期待値は、次のように表される。

$$E[T] = -\beta \left[ 1 + \frac{\Gamma\left(\frac{1}{1-\alpha}\right) k!}{(\alpha-1) \Gamma\left(k+2-\frac{\alpha}{\alpha-1}\right)} \right] \quad (\alpha > 2) \quad (10)$$

超指数分布については、次の性質が成り立つ。

**[補題 2]** 超指数分布の寿命をもつノードの孤立時間の期待値は、次のように表される。

$$E[T] = \int_0^{\infty} \left( 1 - \left( 1 - \frac{1}{E[L_i]} \sum_{i=1}^m \frac{p_i}{\lambda_i} e^{-\lambda_i x} \right)^k \right) dx \quad (11)$$

次に、ノード  $v$  がネットワークを離れる前に、ノード  $v$  のすべての隣接ノードが離脱する確率と呼び、 $\pi$  で表す。孤立確率  $\pi$  は、寿命の確率密度関数  $f(x)$  を用いて、次のように表される。

$$\pi = \int_0^{\infty} (F_R(x))^k f(x) dx \quad (12)$$

指数分布の寿命をもつノードの孤立確率は、次のように表される。

$$\pi = \frac{1}{k+1} \quad (13)$$

またパレート分布の寿命をもつノードの孤立確率は、次のように表される<sup>5)</sup>。

$$\pi = \frac{\Gamma\left(1 + \frac{\alpha}{\alpha-1}\right) k!}{\Gamma\left(k+1 + \frac{\alpha}{\alpha-1}\right)} \quad (\alpha > 1) \quad (14)$$

超指数分布関数については、次の性質が成り立つ。

**[補題 3]** 超指数分布の寿命をもつノードの孤立確率は、次のように表される。

$$\pi = \int_0^{\infty} \left( 1 - \frac{1}{E[L_i]} \sum_{j=1}^m \frac{q_j}{\lambda_j} e^{-\lambda_j x} \right)^k \left( \sum_{k=1}^m q_k \lambda_k e^{-\lambda_k x} \right) dx \quad (15)$$

### 4. 孤立ノードの検出

ネットワーク内のノード  $v$  の次数がゼロになるとノード  $v$  は孤立する。孤立したノードは、他のノードと接続していないためにネットワークから消滅したものとみなされる。このようなノードの検出は、前記のノードの孤立時間と孤立確率を用いて行う。ここでは指数分布、パレート分布、超指数分布の寿命をもつノードについて、ノードの孤立時間と孤立確率を具体的に求める。ノード次数に対する孤立時間の期待値と孤立確率を表 1 に示す。表 1 において超指数分布  $m=3$  の場合を示しており、パラメータ  $p_i$  と  $\lambda_i$  の 3 つの値を表 1 の下に示す。

表1 ノード次数における孤立時間の期待値と孤立確率

No.	ノード次数 [k]	孤立時間の期待値(時間)			ノードの孤立確率		
		超指数分布	パレート分布	指数分布	超指数分布	パレート分布	指数分布
		$m=3$	$\alpha=3, \beta=1$	$\lambda=2$	$m=3$	$\alpha=3, \beta=1$	$\lambda=2$
1	1	1.64	1	0.5	0.22314	0.4	0.5
2	5	4.1247	3.0634	1.1416	0.04257	0.08525	0.16666

《  $E[L_i]=0.5$  時間, 超指数分布のパラメータ  $p_i = \{3/4, 1/20, 1/5\}$ ,  $\lambda_i = \{10, 2, 0.5\}$  》

表1のノード次数が1の場合、孤立時間の期待値はパレート分布が最も大きな値を示しており、これは形状パラメータ  $\alpha=3$  の影響によるものと考えられる。またノード次数が大きくなることにより、どの分布でも孤立時間の期待値が大きくなっている。逆にノード次数が大きくなることによりノードの孤立確率が小さくなっているのは、ノードが長く生存できることを示している。なお超指数分布、パレート分布、指数分布の孤立時間を比較すると、パレート分布と指数分布の間では差が大きい、超指数分布とパレート分布の間の差は、あまり大きくない。パレート分布が超指数分布で近似できることより、孤立時間の期待値と孤立確率について超指数分布を用いて分析することが可能となる。

## 5. まとめ

本稿では、P2Pネットワークにおけるノード寿命が超指数分布に従うときのノードの孤立についてモデル化し、孤立時間と孤立確率を求めた。また指数分布、パレート分布、超指数分布の寿命をもつノードの孤立時間の期待値と孤立確率を具体的に求めて比較した。今後は、P2Pネットワークのノードの孤立とノードの消滅の仕組みを明らかにする予定である。

## 参考文献

- 1) Lv, Q., Cao, P., Cohen, E., Li, K., and Shenker, S.: Search and Replication in Unstructured Peer-to-Peer Networks, *Proc. 16th ACM Int'l Conf. on Supercomputing*(2002).
- 2) Cohen, E., and Shenker, S.: Replication Strategies in Unstructured Peer-to-Peer Networks, *Proc. of ACM SIGCOMM 2002*, pp.177-190 (2002).
- 3) Liben-Nowell, D., Balakrishnan, H., and Karger, D.: Analysis of the Evolution of Peer-to-Peer systems, *Proc. the 21st Annual Symposium on Principles of Distributed Computing*, pp.233-242(2002).
- 4) Stutzbach, D., and Rejaie, R.: Understanding Churn in Peer-to-Peer Networks, *Proc. the 6th ACM SIGCOMM Conf. on Internet Measurement*, pp.189-202(2006).
- 5) Leonard, D., Yao, Z., Rai, V., and Loguinov, D.: On Lifetime-Based Node Failure and Stochastic Resilience of Decentralized Peer-to-Peer Networks, *IEEE/ACM Trans. on Networking*, Vol.15, No. 5, pp.1-13(2007).
- 6) Yao, Z., Wang, X., Leonard, D., and Loguinov, D.: On Node Isolation under Churn in Unstructured P2P Networks with Heavy-Tailed Lifetimes, in *Proc. IEEE INFOCOM*, pp. 2126-2134(2007).

- 7) Kageyama, J., Kobayashi, M., Shibusawa, S., and Yonekura, T.: A Files Replication Method Based on Demand Forecasting in P2P Networks, *Proc. IEEE/ICADWIT2009*, pp.279-285(2009).
- 8) Stutzbach, D., and Rejaie, R.: Towards a Better Understanding of Churn in Peer-to-Peer Networks, *Technical Report, CIS-TR-04-06*, Department of Computer and Information Science, University of Oregon (2004).
- 9) Stutzbach, D., and Rejaie, R.: Characterizing Churn in Peer-to-Peer Networks, *Technical Report, CIS-TR-2005-03*, Department of Computer and Information Science, University of Oregon (2005).
- 10) 安部広多: P2P システム上での安定したサービス提供基盤 musasabi, 情報処理学会研究報告, 2009-IOT-4(23), pp.131-136(2009).