

Paxos の信頼性計算

渡辺 典孝^{†a)} 久保 正業[†]

Paxos[1]の信頼性(Reliability)について、システムの信頼性、リーダー交替及びデータの堅牢性の確率を計算する。

故障率 = $m C_{[m/2]} r^{[m/2]} \approx O(r^{[m/2]})$ で運用できることを示す。

また、Paxos の信頼性は、 $m \cdot n$ 冗長系システムモデルの信頼性に相当するが、多数決決定素子の故障がない点で優れている。

1. はじめに

情報系クラウドの大規模システムに於いては、現行/待機系システムあるいは二重化システムによる信頼性確保では不十分なため、レプリケーション技術による信頼性確保を行っている。例えば、Google は Paxos 過半数方式、Amazon は R+W>N 方式を採用している。

これらのレプリケーション技術の信頼性計算に関する言及が少ないので、ここに検討してみた。

信頼性工学の観点からは、複数個のレプリケーションの内一定数以上が稼働している状態、 $m \cdot n$ 冗長系システムに相当する。

ところで、今年、東日本大震災に関連して福島原発事故が発生した。福島原発事故についての信頼性について述べておきたい。

原発については、約 40 年前に「1 年間に発生する重大事故は 1 万基に 1 件以下」の目安が米国の専門家により既に示されていた[2]。集合平均と時間平均が同じであるという仮定では、「1 基あたり 1 万年に 1 回の重大事故」と言える。この観点から、「平均的に 1 万年に 1 回」という奇妙な安全神話が作られ独り歩きをした。原子力専門家はこの安全神話を推進することはあれ批判することはなかった。科学者であれば、「確率的には、明日発生するかも知れないし、10 万年後かも知れない」、あるいは「世界中で 250 基稼働していれば、平均的に 40 年に 1 回発生する」、と言わねばならない。

不幸にも、福島原発事故は信頼性工学では全くシンプルに予測されていた、いうことができる。

そこで、災害について、筆者は「故障 x 被害」を提案

したい。

原発事故は、無限小の故障 x 無限大の被害($0x\infty$)である。無限小 x 無限大の次元は不明である。

火力発電所の事故は、無限小 x 有限であれば無限小、あるいは有限 x 有限であれば有限とできる。

原発事故の被害は有限なのか、無限なのか。さて、読者は、原発事故の特殊性をどのように考えるのであろうか。

2. 二項係数による信頼性

一般に、 m 台 (セル) 中 y 台以上の稼働率は、1 台の故障率を r とすると、二項展開で記述でき、稼働率は次のようになる。

$$\sum_{i=y}^m m C_i r^{m-i} (1-r)^i$$

$$= \sum_{i=y}^m m C_i r^{m-i} \sum_{j=0}^i i C_j (-r)^j$$

$$= \sum_{i=y}^m \sum_{j=0}^i m C_i i C_j (-1)^j r^{m-i+j}$$

$i = m - k (0 \leq k \leq m - y)$ とすると、

$$= \sum_{k=0}^{m-y} \sum_{j=0}^{m-k} m C_{m-k} m-k C_j (-1)^j r^{k+j}$$

[†]株式会社 トライテック
a) E-mail: nw@tritech.co.jp

次に、 r の次数毎の和を取る。

k	定数	0次	1次	n次	
0	${}_m C_m$	${}_m C_0$	$-{}_m C_1$	$(-)^n {}_m C_n$	
1	${}_m C_{m-1}$	×	${}_{m-1} C_0$	$(-)^{n-1} {}_{m-1} C_{n-1}$	
...		×	×		
k	${}_m C_{m-k}$		×	$(-)^{n-k} {}_{m-k} C_{n-k}$	
m-y				${}_y C_0$	
				×	${}_{y-1} C_0$

r の n 次の係数は、

$$\begin{aligned} & \sum_{k=0}^l {}_m C_{m-k} {}_{m-k} C_{n-k} (-1)^{n-k} \\ &= \sum_{k=0}^l \frac{m!}{(m-k)! k!} \frac{(m-k)! (-1)^{n-k}}{(n-k)! (m-n)!} \\ &= \frac{m!}{(m-n)! n!} \sum_{k=0}^l \frac{n!}{(n-k)! k!} (-1)^{n-k} \\ &= {}_m C_n \sum_{k=0}^l {}_n C_k (-1)^{n-k} \end{aligned}$$

となる。

n が 0 次では、 $l=0$ であり、

$$= 1$$

n が 1 次以下では、 $l=n$ として、

$$= {}_m C_n \sum_{k=0}^n {}_n C_k 1^k (-1)^{n-k} = {}_m C_n (1-1)^n = 0$$

n が $l+1$ 次以上では、 $l < n$ となり、

$$\begin{aligned} &= {}_m C_n \sum_{k=0}^l {}_n C_k (-1)^{n-k} \\ &= {}_m C_n (-1)^n \sum_{k=0}^l {}_n C_k (-1)^k \\ &= {}_m C_n (-1)^n (-1)^l \sum_{k=0}^l {}_n C_k (-1)^k \\ & \left(\because \sum_{k=0}^l (-1)^k {}_n C_k = (-1)^l {}_{n-1} C_l \right) \\ &= (-1)^{n-1} {}_m C_{n-1} C_l \end{aligned}$$

となる。

したがって、稼働率は、

$$1 + \sum_{n=l+1}^m (-1)^{n-1} {}_m C_{n-1} C_l r^n$$

であり、 $l=m-y$ とすると、

$$= 1 + \sum_{n=m-y+1}^m (-1)^{n-m+y} {}_m C_{n-1} C_{m-y} r^n$$

となる。 $r \ll 1$ であれば、 $n=m-y+1$ 項のみで、

$$\begin{aligned} &= 1 + (-1)^{-1} {}_m C_{m-y+1} {}_{m-y} C_{m-y} r^{m-y+1} \\ &= 1 - {}_m C_{y-1} r^{m-y+1} \end{aligned}$$

である。

そこで、稼働率及び故障率は、

$$\text{稼働率} = 1 - {}_m C_{y-1} r^{m-y+1}$$

$$\text{故障率} = {}_m C_{y-1} r^{m-y+1} \approx O(r^{m-y+1})$$

となる。

ここで、 $y = \lceil m/2 + 1 \rceil$ とすると、過半数では

$$\text{稼働率} = 1 - {}_m C_{\lceil m/2 \rceil} r^{\lceil m/2 \rceil}$$

$$\text{故障率} = {}_m C_{\lceil m/2 \rceil} r^{\lceil m/2 \rceil} \approx O(r^{\lceil m/2 \rceil})$$

となる。

従って、Paxos 化により信頼性が巾乗で飛躍的に改善されることが分かる。

なお、簡便には以下のように計算できる。

二項定理

$$(a+b)^m = \sum_{i=0}^m {}_m C_i a^{m-i} b^i$$

で $a=r$ 、 $b=1-r$ とすると

$$1 = \sum_{i=0}^m {}_m C_i r^{m-i} (1-r)^i$$

となる。したがって、

$$\sum_{i=y}^m {}_m C_i r^{m-i} (1-r)^i$$

$$= 1 - \sum_{i=0}^{y-1} {}_m C_i r^{m-i} (1-r)^i$$

$r \ll 1$ とすると、 r の最も低い次数は、 $y-1$ の項であるから、

$$\approx 1 - {}_m C_{y-1} r^{m-y+1}$$

となる。

3. リーダー消滅の確率

Paxos ではリーダーが前提とされ、リーダーが消滅した後、リーダーが選ばれリーダー交替が行われる。

そこで、リーダーが選定されている時のリーダー消滅の確率計算を行う。

リーダーが稼働し、リーダーを除いたサーバー群が故障を起こしリーダーが消滅する確率は、 $m-1$ 台中で $[m/2 + 1] - 1 = [m/2]$ の故障であり、

故障率 = $m-1 C_{[m/2]-1} r^{[m/2]+1} \approx 0(r^{[m/2]+1})$
となる。

従って、リーダー障害を含めると、リーダー交替の確率は、

$r + (1-r) \cdot m-1 C_{[m/2]-1} r^{[m/2]+1} \approx 0(r)$
である。つまり、 $0(r)$ の確率でリーダー交替が発生することになる。

本故障は、処理の進行に関わるのみでシステムにとって致命的ではない。

なお、当社 Paxos ソリューションは故障検知機能を有し、検知時には瞬時にリーダー交替が行われる。

4. キャッシュデータ消失の確率

過半数以上のサーバーが稼働し、そのうち Paxos 合意に参加した過半数が同時に故障を起こす確率であるから、 $r^{[m/2]}$ となる。本故障は、データが消失するのでシステムにとって致命的である。

Paxos 合意に参加した過半数を構成するサーバー全てが同時に故障を起こすと、キャッシュデータが消失する。

同時故障は、独立して稼働しているサーバーがある瞬間に同時に故障する場合、プログラムの論理バグの場合、サーバーが同一電源に接続している時の電源喪失の場合等が考えられる。1 番目の独立稼働しているサーバーの同時瞬間の故障は $r^{[m/2]}$ である。対象としているのは、確率的に、宇宙線によるソフトウェア、タイミング等によるランダムな故障である。本故障の復旧のためにはキャッシュデータの永続化をインスタンス毎に行う必要がある。これを行うと著しい性能低下となる。さらに、永続化の信頼性について考察する必要がある。2 番目はバグであり、プログラムのバグ検出に使える。論理バグは修正されるとする。3 番目は、外部起因によるので外部メカニズムでの対処したいとなる。

なお、電源等の喪失については検知可能であり UPS を備えていればキャッシュデータを保存できる。また過半数故障なので処理の進行は停止する。したがって、電源等の喪失を検知し UPS の動作時間内にキャッシュデータを保存できる機能を備えていれば本致命的故障を回避できる。

5. データの堅牢性

全サーバーが停止しない限り、データは保持されるので信頼性は、

$$1 - r^m$$

となる。

なお、前節のキャッシュデータ喪失は回避されている。

6. 計算例

近似式

$$\text{故障率} = m C_{[m/2]-1} r^{[m/2]+1}$$

で $m=3, 5, 7$ 台構成で計算すると以下ようになる。

1台の故障率(r)	3台	5台	7台
5%	0.75%	0.125%	0.021875%
3%	0.27%	0.027%	0.002835%
1%	0.03%	0.001%	0.000035%

ところで、1 台あたりの故障率が 10 倍向上すれば、セルの故障率は巾乗で改善される。3、5、7 では、それぞれ、2、3、4 の巾乗であり、100、1000、10000 倍の改善となる。

1%が 10 倍向上の 0.1%となれば、

$$3 \text{ 台 } \quad 0.03\% \quad \Rightarrow 0.0003\% \quad (-5 \text{ 乗})$$

$$5 \text{ 台 } \quad 0.001\% \quad \Rightarrow 0.000001\% \quad (-7 \text{ 乗})$$

$$7 \text{ 台 } \quad 0.000035\% \Rightarrow 0.0000000035\% \quad (-10 \text{ 乗})$$

となり、飛躍的に改善される。

7. 信頼性の再吟味

信頼性の定義については少しく疑問が生じる。

100 個の製品の内、不良品が 1 個であれば 1%の不良品率は非常に解りやすい。しかしながら、100 個の製品を 1 年間使用して、1 個に故障が発生する、という期間を限定した時間依存の故障率が一般的である。これについて検討する。

7.1. 信頼性

数理的には、事象 ω に対し、実時間 t に写像する確率変数 $T(\omega)$ を考え、 $T(\omega) \in (t, \infty]$ となる T に $R(t) = P(T > t) = P(\{\omega | T(\omega) > t\})$ を信頼度関数とし、事象 ω が t まで存続する、と定義する。即ち、 $T(\omega) > t$ となる ω の集合の確率が $R(t)$ なのだが、 ω と t の明確な関係は曖昧であり、 ω を基底にある抽象的事象とし、一般に $R(t)$ 関数を仮定する。

MTTF(Mean Time To Failure)は、

$$MTTF = \int_0^{\infty} R(t)dt$$

と定義される。

$t=0$ では n 台が稼働し順次 $n-1$ 台、 $n-2$ 台と移行していく。そこで n 台稼働している場合の $MTTF$ を $MTTF_n$ とすると、 $MTTF_1 > MTTF_2 > \dots > MTTF_n$ であり、 k 台以上稼働している場合の $MTTF$ は、

$$MTTF^k = MTTF_n + MTTF_{n-1} + \dots + MTTF_k$$

とすることができる。

直列の場合は、 $MTTF_n$ のみであり、明らかに $MTTF_1$ より小さい。並列の場合は、 $MTTF_1$ を含むので明らかに $MTTF_1$ より大である。 k 台以上については不明である。

例えば、一般に、 $t=0$ で信頼度 1、時間経過で信頼度が指数関数的に下がるとして

$$R(t) = \exp(-\lambda t)$$

を仮定すると、明らかに、

$$MTTF_n = 1/\lambda n$$

であり、 k 個以上の稼働は、

$$MTTF^k = \frac{1}{\lambda} \left(\frac{1}{n} + \dots + \frac{1}{k} \right)$$

となる。

すなわち、直列では $MTTF^n = MTTF_n = MTTF_1 \frac{1}{n}$ で 1 台の n 分の 1 となり、並列では $MTTF^1 = MTTF_1 \left(\frac{1}{n} + \dots + \frac{1}{1} \right)$ であるので 1 台より大であるが一台の追加は $1/n$ の効果しかない。そして、 k 台では不明である。

Paxos の場合、 k を過半数とするので、 $n=5$ 、 $k=3$ とすると、

$$\frac{1}{5} + \frac{1}{4} + \frac{1}{3} = \frac{47}{60}$$

であり、1 台の $MTTF$ より悪い。

これは何を意味しているのだろうか。すでに、故障率はべき乗で改善されることを示したが、 $MTTF$ は改悪となっている。

これには、二つ問題がある。

- ① $MTTF$ の計算では t の無限大まで積分する。実際には、使用期間（ミッション期間）があり無限大までの積分は現実的ではない。Paxos は、 $\lambda t \ll 1$ の範囲を対象にしており、即ち、 $MTTF$ より十分に小さい範囲でも、確率的には障害が発生するのでこれを回避するデータ保全と処理の進行を目的としている。
- ② 指数関数的分布が問題である。より正確にはワイブル分布であろう。極端には、ステップ分布でもよい

であろう。

したがって、 $MTTF$ は参照のための平均的目安と考えるべきであろう。

7.2. 保全性

瞬間故障率はべき乗で改善されるが、目安となる指数分布での $MTTF$ は改悪されることが示された。そこで、保全が要請される。

つまり、故障の検知されたマシンを速やかに修理し、戻すことができれば、 $MTTF$ を延ばすことができる。

即ち、 $MTTF_n = MTTF_1 \frac{1}{n}$ で速やかに修理することができれば、 $MTTF^n$ を永遠化できる。

8. 過半数原理

8.1. Paxos の過半数の意味

旧過半数と新過半数に少なくとも 1 つの共通サーバーが存在する。このサーバーが旧過半数の状態を新過半数に引き継ぐことにより、旧から新への連続性が維持される。

したがって、新で過半数を維持できなければ旧を引き継げない。すなわち、新と旧には共通部分がなく旧の過半数を構成するサーバーがダウンしたことを意味する。

つまり、過半数は共通部分を有し、この共通部分が旧と新を媒介する。

8.2. R+W>N 方式

参照の台数 R 、更新の台数 W 、全台数 N として、書き込み時には W 個以上が実行し、読み時には R 個以上からデータを読み込む。

全台数が N 個なので、 W 個に同じデータを書き込み、 R 個以上からデータを読み込むと、少なくとも 1 個には書き込んだデータがある。というのは、 $R+W>N$ なので過半数と同様に少なくとも 1 個以上の共通部分があるので、最新の状態を得ることができる。

信頼性計算の観点からは、 R あるいは W の数の多いほうを y とすることができる。

8.3. リーダー選択への適用

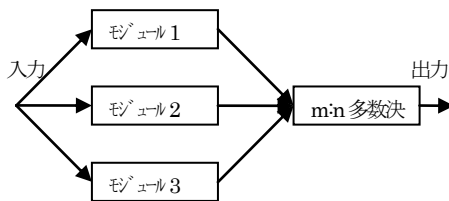
過半数原理を素直に適用したのが当社のリーダー選択アルゴリズムである。本アルゴリズムは各サーバーがリーダー候補を提案し、各サーバーは他のサーバーからの提案で過半数が同じ提案をしていれば自律的にリーダーと判断する。つまり、過半数が推薦しているのでリーダーとしてよい、と自律的に判断する。本アルゴリズムでは候補を提案し、提案が過半数であることに妙味が

ある。

しかしながら、候補を選ぶ絶対的基準が必要であり、これを誕生時刻とするのは自然であり、もって長老方式と称している¹。

9. m:n 冗長系システム

実は、本件の信頼性の議論は、信頼性工学の m:n 冗長系システムに相当する。



同じ入力を複数のモジュールに与え、それぞれの結果の多数決で出力を決定すれば、1つのモジュールの故障を隠蔽できる。

しかし、多数決決定素子は1つであり、これに故障が発生するとシステム障害となる。

これに対し、Paxos はモジュール間の協調で多数決を決定するので多数決決定素子の故障 (single point of failure) が無い。

10. まとめと効果

Paxos の信頼性計算は、従来の m:n 冗長系システムモデルの信頼性計算そのものである。従来の m:n 冗長系システムモデルでの信頼性計算では、一般に多数決決定素子の故障を無視している点で、難があった。

Paxos は、m:n 冗長系のソフトウェア的拡張と考えることができ、多数決決定を自律分散で行うので決定素子の故障がなく、致命的システム障害を劇的に (べき乗) 改善する。

その効果については、以下を列挙できる。

- ① Paxos セル構成で、システムに故障がないとみなすことができれば(ナイン 9)、back end 永続化装置を含めてキャッシュとみなすことができ、高速化を図れる。
- ② Paxos は、通信遅延を前提としたアルゴリズムなので、広域展開による災害対応ができる。
- ③ 複数台構成なので目安となる MTTF で修理を行えばシステム運転を中断することなく全体の寿命を

¹ サーバーが起動された時刻、リーダーを降りた時刻を誕生時刻とし、時刻が同じであれば若い ID による。

永遠に延ばすことができる。

付録 (メモ)

念のために、その他の表記についても記載しておく。

不完全ベータ関数による表記

不完全ベータ関数は、

$$B_x(a, b) = \int_0^x y^{a-1} (1-y)^{b-1} dy$$

で定義される。また、

$$B_1(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

であり、不完全ベータ関数比は、

$$I_x(a, b) = \frac{B_x(a, b)}{B_1(a, b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x y^{a-1} (1-y)^{b-1} dy$$

で定義される。

部分積分を繰り返すと、

$$I_x(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x y^{a-1} (1-y)^{b-1} dy \\ = \frac{\Gamma(a+b)}{\Gamma(a+1)\Gamma(b)} x^a (1-x)^{b-1} + I_x(a+1, b-1)$$

$$= \sum_{i=0}^n \frac{\Gamma(a+b)}{\Gamma(a+1+i)\Gamma(b-i)} x^{a+i} (1-x)^{b-1-i}$$

$$+ I_x(a+1+n, b-1-n)$$

となる。

$a=y$ 、 $b=m-y+1$ 、 $n=m-y-1$ とすると、

$$= \sum_{i=0}^{m-y-1} \frac{\Gamma(m+1)}{\Gamma(y+1+i)\Gamma(m-y+1-i)} x^{y+i} (1-x)^{m-y-i} + I_x(m, 1)$$

$$= \sum_{i=0}^{m-y-1} \frac{m!}{(y+i)!(m-y-i)!} x^{y+i} (1-x)^{m-y-i} + x^m$$

$$= \sum_{i=y}^m {}_m C_i x^i (1-x)^{m-i}$$

$$= I_x(y, m-y+1) = y {}_m C_y \int_0^x z^{y-1} (1-z)^{m-y} dz$$

となる。

ここで、

$$I_x(a, b) = 1 - I_{1-x}(b, a) \quad (t=1-z \text{ とすればよい})$$

から、

$$\sum_{i=y}^m {}_n C_i r^{m-i} (1-r)^i$$

$$= 1 - y_m C_y \int_0^r z^{m-y} (1-z)^{y-1} dz$$

を得る。

本表記は、

$$y_m C_y \int_0^r z^{m-y} (1-z)^{y-1} dz$$

であり、二項展開をすると、

$$= y_m C_y \int_0^r z^{m-y} \sum_{i=0}^{y-1} y_{-1} C_i (-z)^i dz$$

$$= y_m C_y \sum_{i=0}^{y-1} y_{-1} C_i \int_0^r z^{m-y+i} (-1)^i dz$$

$$= y_m C_y \sum_{i=0}^{y-1} y_{-1} C_i \frac{r^{m-y+i+1}}{m-y+i+1} (-1)^i$$

$n=m-y+i+1$ とすると、

$$= y_m C_y \sum_{n=m-y+1}^m y_{-1} C_{(n-1)-(m-y)} \frac{r^n}{n} (-1)^{(n-1)-(m-y)}$$

$$= y_m C_y \sum_{n=m-y+1}^m y_{-1} C_{m-n} \frac{r^n}{n} (-1)^{(n-1)-(m-y)}$$

$$= \sum_{n=m-y+1}^m \frac{m!}{(y-1)!(m-y)!(m-n)!(y-1-m+n)! n} (y-1)! r^n (-1)^{n-m+y-1}$$

$$= \sum_{n=m-y+1}^m \frac{m!}{n!(m-n)!(m-y)!(y-1-m+n)! n} n! r^n (-1)^{n-m+y-1}$$

$$= - \sum_{n=m-y+1}^m (-1)^{n-m+y} {}_m C_{n-1} C_{m-y} r^n$$

となり、二項係数による直接的な導出と同じになる。

個々の時間依存の信頼性を $R(t)$ とすると、システム全体の信頼性は、

$$\begin{aligned} R(t) &= y_m C_y \int_0^{R(t)} z^{y-1} (1-z)^{m-y} dz \\ &= I_{R(t)}(y, m-y+1) \end{aligned}$$

となる。

また、瞬間故障率は、

$$\lambda(t) = - \frac{dR(t)}{dt} = - \frac{R(t)^{y-1} (1-R(t))^{m-y}}{\int_0^{R(t)} x^{y-1} (1-x)^{m-y} dx} \frac{dR(t)}{dt}$$

である。

MTTF は、

$$MTTF = \int_0^{\infty} R(t) dt$$

であり、 $R(t) = e^{-\lambda t}$ のときは、

$$MTTF = \frac{1}{\lambda} \sum_{i=y}^m \frac{1}{i}$$

となる。

実際、

$$\begin{aligned} I_x(a, b) &= \frac{\Gamma(a+b)}{\Gamma(a+1)\Gamma(b)} x^a (1-x)^{b-1} \\ &\quad + I_x(a+1, b-1) \end{aligned}$$

の漸化式で、 $x = e^{-\lambda t}$ とすると

$$\begin{aligned} \int_0^{\infty} x^a (1-x)^{b-1} dt &= \frac{1}{\lambda} \int_0^1 x^{a-1} (1-x)^{b-1} dx \\ &= \frac{1}{\lambda} B_1(a, b) = \frac{1}{\lambda} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \end{aligned}$$

となる。したがって、

$$\begin{aligned} \int_0^{\infty} I_x(a, b) dt &= \frac{\Gamma(a+b)}{\Gamma(a+1)\Gamma(b)} \frac{1}{\lambda} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ &\quad + \int_0^{\infty} I_x(a+1, b-1) dt \\ &= \frac{1}{\lambda a} + \int_0^{\infty} I_x(a+1, b-1) dt \end{aligned}$$

を得る。

$a=m, b=1$ では $\frac{1}{\lambda m}$ であり、 $a=y, b=m-y+1$ まで成立

したとすると、 $a=y-1, b=m-y+2$ でも成立する。

マルコフ過程モデルによる表記

マルコフ過程モデルは、システムに状態を想定し時間依存の状態の遷移を確率的に記述した微分方程式となる。

m 個が稼働している状態を S_m 、 $m-1$ 個が稼働している状態を S_{m-1} 、 \dots 、 0 個が稼働している状態を S_0 とすると、状態は $S_m \rightarrow S_{m-1} \rightarrow \dots \rightarrow S_0$ と遷移していく。このようなシステムを複数想定した時、同時刻であるシステムは S_m 、別のシステムは S_{m-1} 等と考えられる。システムの数を十分に大きいとすると、状態 S_i に存在する確率 P_i を考えることができ、

$$\frac{dP_m}{dt} = -\mu_m P_m$$

$$\frac{dP_i}{dt} = \mu_{i+1} P_{i+1} - \mu_i P_i$$

$$\frac{dP_0}{dt} = \mu_1 P_1$$

$$t=0 \text{ で } P_m = 1, P_{m-1} = \dots = P_0 = 0$$

とする。

$$\begin{aligned} dy/dx + Py = Q \text{の解は} \\ y = e^{-\int P dx} \left\{ \int Q e^{\int P dx} dx + C \right\} \text{であることから、} \\ P_m = e^{-\mu_m t} \end{aligned}$$

$$P_{m-1} = \mu_m \left\{ \frac{e^{-\mu_m t}}{\mu_{m-1} - \mu_m} + \frac{e^{-\mu_{m-1} t}}{\mu_m - \mu_{m-1}} \right\}$$

...

$$P_{m-k} = \prod_{i=0}^{k-1} \mu_{m-i} \sum_{i=0}^k \frac{e^{-\mu_{m-i} t}}{\prod_{j=i}^k (\mu_{m-j} - \mu_{m-i})}$$

を得る。

ところで、1台の故障率を λ とすると、最初に m 台が稼働し $m-1$ 台への遷移は $m\lambda$ 、次に $m-1$ 台から $m-2$ 台への遷移は $(m-1)\lambda$ であるので、

$$\mu_m = m\lambda, \mu_i = (m-i)\lambda$$

とすることができる。

また、 $a = e^{-t}$ とすると、

$$\begin{aligned} & \frac{1}{\prod_{j=i}^k (\mu_{m-j} - \mu_{m-i})} \\ &= \frac{1}{\lambda^k (i-1) \cdots (1)(-1) \cdots (i-k)} \\ &= \frac{1}{\lambda^k i! (k-i)!} = \frac{1}{\lambda^k k!} k C_i (-1)^{k-i} \\ & \sum_{i=0}^k \frac{e^{-\mu_{m-i} t}}{\prod_{j=i}^k (\mu_{m-j} - \mu_{m-i})} = \frac{1}{\lambda^k k!} \sum_{i=0}^k k C_i (-1)^{k-i} a^{m-i} \\ &= \frac{a^{m-k}}{\lambda^k k!} \sum_{i=0}^k k C_i (-1)^{k-i} a^{k-i} = \frac{a^{m-k}}{\lambda^k k!} (1-a)^k \\ P_{m-k} &= \prod_{i=0}^{k-1} \mu_{m-i} \sum_{i=0}^k \frac{e^{-\mu_{m-i} t}}{\prod_{j=i}^k (\mu_{m-j} - \mu_{m-i})} \\ &= m(m-1) \cdots (m-k+1) \lambda^k \frac{a^{m-k}}{\lambda^k k!} (1-a)^k \\ &= \frac{m!}{(m-k)! k!} a^{m-k} (1-a)^k \\ &= {}_m C_k a^{m-k} (1-a)^k \end{aligned}$$

となる。

これは、 $a=1-r$ とすると時定数のない二項展開の k 項に相当する。

時定数導入では、 t を λt と置き換える。

奇数台のMTTFは単調減少

$$mttf^{2n+1} = \frac{1}{2n+1} + \cdots + \frac{1}{n+1}$$

とする。和は過半数個である。

次に

$$\begin{aligned} mttf^{2(n+1)+1} &= \frac{1}{2n+3} + \frac{1}{2n+2} + \frac{1}{2n+1} + \cdots + \frac{1}{n+2} \\ &= \frac{1}{2n+3} + \frac{1}{2n+2} - \frac{1}{n+1} + mttf^{2n+1} \\ &= \frac{(2n+2)(n+1) + (2n+3)(n+1) - (2n+3)(2n+2)}{(2n+3)(2n+2)(n+1)} \\ &\quad + mttf^{2n+1} \\ &= \frac{-n-1}{(2n+3)(2n+2)(n+1)} + mttf^{2n+1} \end{aligned}$$

であり、単調減少となっている。

したがって、 $n=0$ で1なので1を超えることがない。

また、偶数台の場合には、 $\frac{1}{2n+2}$ が付加される。5台で

は $\frac{47}{60}$ であり、6台では $\frac{1}{6}$ が付加される。これは1を超えないので、1台、3台そして5台以上での過半数では1台のMTTFを超えることはない。

故障率と時間依存の故障率

故障率は時間依存では、

$$r(t) = 1 - R(t) = 1 - e^{-\lambda t}$$

が仮定される。

瞬間での $r(t)$ に対し、レプリケーション技術ではべき乗で故障率は改善されることは言うまでもないが、時間を経るにつれ故障率は悪化していくことになる。

例えば、ナイン9が時間を経るにつれてシックス9になるということになる。つまり、確率論的に拡張したのがMTTFである。実際には、ミッションクリティカルでは、 $\lambda t \ll 1$ とし、MTTFの1/5あるいは1/10以上で議論する必要がある。即ち、MTTFは平均的であるので、MTTFまで待っていては、50%で故障が発生することになる。

そこで、改めて原発事故に立ち返ってみると、確率論的な「1基1万年の重大事故」で、1/10のオーダーとして保守修理をしたところで、事故は発生するのであり、無限大の甚大な被害の前では無力である。何故、確率論が「原子力安全神話」に寄与したのであろうか。

[1] L.Lampert. Paxos Made Simple.

<http://research.microsoft.com/en-us/um/people/lampor/t/pubs/paxos-simple.pdf>

[2] 吉本光一. 「文芸春秋」2011年12月号、269頁

[3] 河田龍夫 「確率と統計」朝倉書店

[4] 南谷崇 「フォールトトレラントコンピュータ」

オーム社

[5] 市川昌弘 「信頼性工学」 裳華房
等

[文責] 渡辺典孝