自己安定分散システム

研究ノート

増　澤　利　光*

Self-stabilizing Distributed Systems

Key Words：Distributed System, Fault-Tolerance, Self-Stabilization

1．はじめに

インターネットのような地球規模の計算機ネットワーク環境が整備されており，この計算機ネットワークを活用した大規模分散システムの構築が期待されている．分散システムの長所の一つは，一部の計算機で障害が発生しても分散システム全体の機能は停止せず，正常な部分でサービスを継続できる故障耐性を実現できることである．ネットワークが大規模化するほどネットワーク中の計算機が故障等による障害を受けることは避けられない．また，不特定多数のユーザが使用する場合，(悪意のある)ユーザが計算機を停止させたり，誤動作させたりするおそれもある．従って，実用的な分散システムには高度な故障耐性が要求される．

分散システムの故障耐性を実現するアプローチは，マスク型と非マスク型に大別できる．マスク型故障耐性は故障をユーザから隠蔽する．つまり，故障が生じてもユーザは故障に気づくことなく，分散システムの利用を継続できる．一方，非マスク型故障耐性では，正常動作に復帰するまでの間，ユーザに故障の影響が及ぶことを許す．一般に，非マスク型故障耐性はマスク型に比べ，容易に(低コストで)実現できる．そのため，一時的な動作の乱れが許容できる分散システムには非マスク型故障耐性が適し，そのような乱れが許容できない場合はマスク型故障耐性を提供する必要がある．

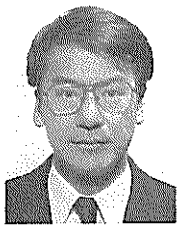
我々の研究室では，分散システムの故障耐性を実現するアルゴリズム的アプローチについてさまざまな研究をおこなっているが，本稿では非マスク型故障耐性の有力なパラダイムである自己安定(self-stabilization)を紹介する[1]．

2．自己安定分散システム

通常の分散システムでは，あらかじめ定められた初期状況から実行を開始することを前提としている．つまり，分散システム実行開始時には，すべての計算機の状態が初期状態にリセットされていることを前提とする．一方，自己安定分散システムはネットワークのどのような状況から実行を開始しても，いずれ正常動作に復帰することを保証する．この性質から，計算機の一時的な故障のためにネットワークがどのような状況に陥っても，それらの故障が復旧し，十分に長い間新たな故障が発生しなければ，自己安定分散システムは自動的に正常動作に復帰する．つまり，自己安定分散システムは，任意の数と種類の一時的な故障に対する非マスク型故障耐性を有する．

ここでは自己安定分散システムの簡単な例として，データやサービスを提供するサーバ計算機がいくつか存在するときに，各計算機からサーバ計算機までの最短路を求める方法を紹介する．計算機間は適当に通信線で接続されており，各通信線にはそれを利用するのに必要なコスト(例えば、通信費、遅延時間など)が与えられている(図1)．そして経路の長さは，その経路に現れる通信線のコストの総和とする．

図2に各計算機からサーバ計算機までの最短路を求める自己安定システムを示す．図2では，計算機P_i，P_j間に通信線があるとき，P_iからP_jへの通信はP_iがレジスタ$R_{i,j}$に書込んだデータをP_jが読出



*Toshimitsu MASUZAWA
1959年11月生
1987年大阪大学大学院・基礎工学研究科・物理系専攻・博士後期課程了
現在，大阪大学・大学院情報科学研究科・コンピュータサイエンス専攻，教授，工学博士，計算機科学
TEL 06−6850−6580
FAX 06−6850−6582
E-Mail masuzawa@ist.osaka-u.ac.jp

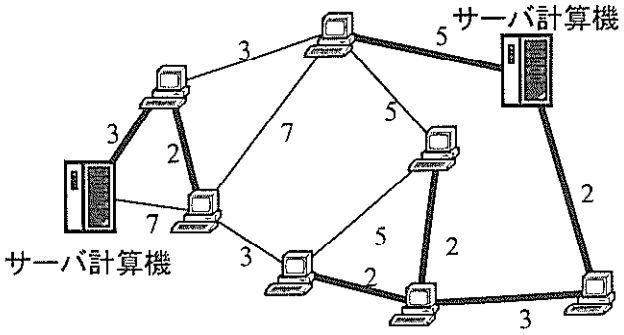


図1　サーバまでの最短路

サーバ計算機P_iのプログラム
while true **do**
$dist_i$：＝0；
for each neighbor P_j **do** $R_{i,j}$：＝$dist_i$；

サーバ以外の計算機P_iのプログラム
while true **do**
for each neighbor P_j **do** $r_{j,i}$：＝$R_{j,i}$；
$dist_i$：＝$\min\{r_{j,i}+c_{i,j} | P_i, P_j$は隣接$\}$；
for each neighbor P_j **do** $R_{i,j}$：＝$dist_i$；

図2　サーバへの最短路を求める自己安定分散システム

すことによって行なっている．同様に，P_jからP_iへの通信は$R_{j,i}$を用いて行なう．また，計算機P_i，P_j間のコストを$c_{i,j}$と表し，各計算機の変数$dist_i$にはP_iからサーバ計算機までの最短距離が求まる．図2の分散システムでは，各P_iは$dist_i$をすべての隣接計算機に知らせ，隣接計算機の中で$dist_j+c_{i,j}$の最小値を$dist_i$とするという動作を繰返す．ただしサーバ計算機は距離として，常に値0をすべての隣接計算機に繰返し知らせる．なお，図2では各P_iはサーバ計算機までの最短距離を$dist_i$に求めるだけで，サーバ計算機までの最短路を求めてはいない．しかし，P_iに対して$dist_i=dist_j+c_{i,j}$なる隣接計算機P_jがサーバ計算機への最短路上の隣接計算機であることは明らかである．

図2の分散システムは，サーバ計算機の変更，通信線のコスト変化，ネットワークの形状変化などが生じた場合，その変化に応じて新たな最短路を自動的に再計算する．つまり，自己安定分散システムは，動的に変化するネットワークにおいて変化に応じて自動的に解を更新する分散システムである．

3．強安定分散システム

これまでに提案されている自己安定分散システムは，その計算量や計算モデルの仮定のために，実際の大規模分散システムの構築への利用が困難なものが多い．そこで，実際の大規模分散システム構築への利用可能性を高めるための拡張について簡単に紹介する．

自己安定分散システムを実際に運用している状況を考えると，一部の計算機の一時故障により不都合な状況に陥ることがあるが，その状況は正しい状況からかけはなれた状況ではなく，わずかに変動(小変動)した状況になることがほとんどである．小変動からの復旧は効率的に行われることが期待される．しかし，多くの自己安定分散システムでは小変動の影響が分散システム全体に波及してしまうため，復旧を効率的に行えない．例えば，図2の自己安定分散システムでは，最短路を求めた状況である計算機P_iの変数$dist_i$の値が故障により不当に小さな値になってしまうと，その隣接計算機も変数$dist$を変更してしまう．同様のことが繰り返されると，たった1つの計算機の故障の影響がネットワーク全体に広がってしまうことがある．このような自己安定分散システムでは，小規模な一時故障の影響が分散システム全体に波及し，正常動作への復帰に要する計算量(時間，通信量など)が分散システム全体のサイズに依存するため，特に，大規模分散システムではその有用性が損なわれる．

そこで，小変動が生じた場合，その影響を制限する自己安定分散システムが重要である．このような特性は強安定，故障封じ込めと呼ばれるが，これらの特性を有する自己安定分散システムの研究を行っている[2,3,4]．これらの自己安定分散システムは小規模な故障の影響を制限するという意味で，小規模な故障に対する部分マスク型故障耐性を実現していると言える．

4．むすび

本稿では，高度な故障耐性を有する自己安定分散システムを紹介し，我々の研究室で取り組んでいる研究の一端について述べた．我々の研究室では今後，自己安定分散システムの高度な適応性に着目し，モバイル分散システムやエージェントシステムなど動

的な分散システムへの自己安定の適用について研究を進めていく予定である.

参考文献

[1] S. Dolev：“Self-Stabilization”, The MIT Press (2000).

[2] 片山, 増澤：“重み最小生成木を構成する故障封じ込め自己安定プロトコル”, 電子情報通信学会論文誌(DI), J84-D-I(9), pp.1307-1317 (2001).

[3] Y. Katayama, E. Ueda, T. Masuzawa and H. Fujiwara；“A latency optimal superstabilizing mutual exclusion protocol in unidirectional rings”, Journal of Parallel and Distributed Computing, 62(5), pp.865-884 (2002).

[4] D. Kondou, H. Masuda and T. Masuzawa：“A self-stabilizing protocol for pipelined PIF in tree networks”, Proc. of ICDCS 2002, pp.181-190(2002).