分散型データベースにおけるコピー配置に関する一考察

6C−1

坂上 雄一
広島大学大学院工学研究科

宮尾 淳一
広島大学総合科学部

1 まえがき

本論文では，分散型データベースシステムにおける通信コストを考慮したファイルの複数コピー配置問題の考察を行う．これまでの研究において，任意形状のネットワークにおける最適ファイル配置を求める問題がNP完全であること，及び，ネットワーク形状に制約を加えた場合の最適配置可能性が証明された[1]．しかし，ネットワーク形状がリングで，各枝の通信コストが一定などの制約が加えられていた．本研究では，リングネットワークの各枝の通信コストが任意の場合の最適データ配置手法に関する考察を行う．

2 準備

本稿で用いる用語を定義する（[1] 参照）．**ネットワーク**をグラフ $G=(V,E)$ により定義する．ここで，V はプロセッサ集合を表し，E はプロセッサを結合するリンク集合を表す．G 上でデータ x の配置されたプロセッサ集合を x に関する**配置集合** (residence set) と定義し，$R(x)(\subseteq V)$ と表す．G に対する $V'(\subseteq V)$ の誘導部分グラフを $U(V';G)=(V',E')$ と定義する．位数２の正則な連結グラフを**リングネットワーク** $G_R=(V_R,E_R)$ と定義する．ただし，$V_R=\{v_1,v_2,\ldots,v_n\}$，$E_R=\{(v_1,v_2),(v_2,v_3),\ldots,(v_n,v_1)\}$ とする．これ以降は，すべてリングネットワークについて考える．

$G_R=(V_R,E_R)$ における連結部分グラフを**ストリング** (string) と定義し，$S=(V_S,E_S)$ と表す．ここで，$V_S\subseteq V_R$，$E_S\subseteq E_R$となる．G_Rにおける配置集合 $R(x)$ に対して，$R(x)$ に属していない G_R上のプロセッサから導出される極大連結部分グラフの集合を Q とし，Q に属する任意の要素を**ホール** (hole) と定義し，$H=(V_H,E_H)$ と表す．$G_R=(V_R,E_R)$ の各リンク $e\in E_R$ の重みを $w(e)$ と定義する．G_R上の任意のプロセッサ集合 $P\subseteq V_R$に対して，Pのすべてのプロセッサを含むストリングで，$min\sum_{e\in E_S}w(e)$ を満足するものを Pの**最小ストリング** $SS(P)$ と定義する．

リングネットワーク G_R上の任意のホール $H=(V_H,E_H)$ を考える．この時，ホールの重み $|H|$ を

$$|H|=\sum_{e\in E_H}w(e)\quad+\quad(H\text{と隣接する配置集合との境界に存在する２本の枝の重みの和})$$

と定義する．また，各プロセッサの読み出し回数$\#R$，書き込み回数$\#W$とし，**読み書き比** αを$\alpha=\#R/\#W$と定義する．

3 定式化

プロセッサ v_iにおけるデータ x の読み出しコストは $R(x)$ に含まれる１つのプロセッサまでの最短距離とし，rc_iと表す．また，書き込みコストは $R(x)$ に含まれるすべてのプロセッサで書き込みを行うための最短距離とし，wc_iと表す．wc_iの値は次の性質１により求められる．

[性質1] G_Rをリングネットワーク，Rを配置集合，そして，iを G_R上のプロセッサとする．この時，プロセッサ iの書き込みコスト wc_iは，最小ストリング $SS(R\cup\{i\})$ の枝の重みの総和に等しい．□

以降では，各データ項目に対する最適配置集合はそれぞれ独立に求められるので，配置集合 $R(x)$ を単に R と表す．

[問題] RSPR(Residence Set Problem forRing Network)
入力：リングネットワーク $G_R=(V_R,E_R)$，
　実数 α（読み書き比）
出力：配置集合 R_{opt}
目的関数：$Cost(R_{opt})=\sum_{v_i\in V_R}(\alpha\times rc_i+wc_i)\rightarrow min$ □

この問題 $RSPR$は，すべてのプロセッサでの読み書き比αを一定とし，すべてのプロセッサの書き込み回数が等しい場合の最適化問題となっている．

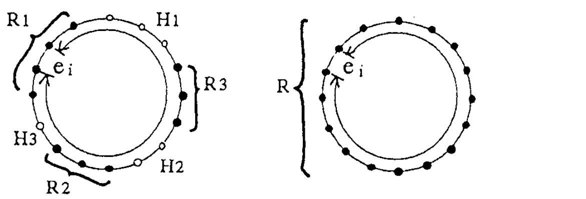
4 提案手法

以下では，N個のストリングからなる配置集合が存在する場合，各ストリングに対するプロセッサ集合を $R_1,R_2,\ldots,R_N$ とする．この時，N個のホールが存在することになる．重みに関して非増大順となるようにホールに番号付けを行い，$H_1=(V_1,E_1),H_2=(V_2,E_2),\ldots,H_N=(V_N,E_N)$（ただし，$|H_1|\geq|H_2|\geq\cdots\geq|H_N|$）とする．

A Study of Replicated Data Residence Method in Distributed Database Systems
Yuichi SAKAUE and Jun'ichi MIYAO
Hiroshima University

[補題 1] リングネットワーク $G_R=(V_R,E_R)$ の最適な配置集合がホールを1個以上持つ場合，G_Rの枝集合 E_R には $|H_1|<w_i$ となる最大重みを持つ枝 e_iが存在しない．

(証明) 背理法で証明する．今，G_R上で最適配置集合 Rがあり，G_Rの枝集合に $|H_1|<w_i$となる最大重みを持つ枝 e_iが存在すると仮定する．また，$N(N\geq 1)$ 個のホールが存在すると仮定する．枝 e_iは当然配置集合 Rの誘導部分グラフ $U(R;G_R)=(R,E')$ の枝集合 E'に含まれることになる．ここでは，誘導部分グラフ $U(R_1;G_R)=(R_1,E')$ の枝集合に含まれるとする．すべてのプロセッサにおける書き込みは図1(a)の矢印のようになる．この時，$\sum_{i=1}^{N}V_i$ のプロセッサを配置集合に含めても書き込みコストは変化しない．また，$\sum_{i=1}^{N}V_i$ のプロセッサの読み出しコストは0に減少する．よって，すべてのプロセッサを配置集合とすることで (図1(b))，よりコストの小さな配置集合を求めることができる．したがって，Rの最適性に矛盾するので，このような Rは存在しない． □



(a)　(b)

図1：$|H_1|<w_i$ の枝が存在するリングネットワーク

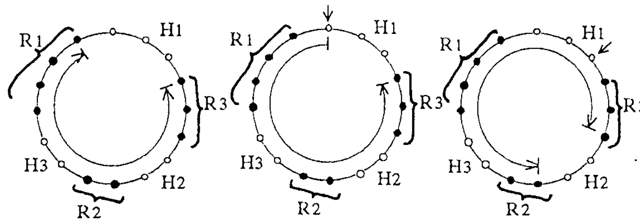
補題1より，最適配置集合は，その誘導部分グラフの枝集合に $|H_1|$ より大きな重みを持つ枝を含まないものを考えれば十分であることが理解できる．

[補題 2] リングネットワーク G_Rの最適な配置集合のホールは2個以下である．

(証明) 背理法で証明する．今，G_Rの最適な配置集合 Rがあり，ホールを $N(N\geq 3)$ 個もつと仮定する．ここで，$R'=(\bigcup_{i=3}^{N}V_i)\cup R$とおき，配置集合を新しく R'としたときのコストの変化について考える．まず，書き込みコストについて考える．$(\bigcup_{i=2}^{N}V_i)\cup R$のプロセッサによる書き込みは補題1より図2(a)の矢印のようになる．よって，配置集合を R'に更新したとしても書き込みコストは変化しない．また，V_1のプロセッサによる書き込みは図2(b)と(c)の2通りがある．この場合も，配置集合を R'に更新したとしても書き込みコストは変化しない．よって，配置集合を R'とした時の書き込みコストは変化しない．次に，読み出しコストについて考える．$(\bigcup_{i=1}^{2}V_i)\cup R$のプロセッサの読み出しコストは変化せず，$(\bigcup_{i=3}^{N}V_i)$ のプロセッサの読み出しコストは $(\bigcup_{i=3}^{N}V_i)$ が配置集合に含まれることになるので0に減少する．よって，配置集合を $R'=(\bigcup_{i=3}^{N}V_i)\cup R$ としたとき，総コストは減少する．よって，Rの最適性に矛盾するので，このような Rは存在しない． □

補題2より高々2つのホールからなる配置集合を調べることで最適配置集合を求めることができる．また，ホールが丁度2個の場合は次の補題が成立する．

[補題 3] ホール2個の最適な配置集合においては $|H_1|\leq 2|H_2|$ を満たす． □



(a)　(b)　(c)

図2：2個以下のホールを持つ配置集合

[定理 1] 任意の枝重みを持つリングネットワークに対して，問題 $RSPR$は $O(n^4)$ で解ける．

(証明) リングネットワーク上のプロセッサ数を n とする．当然，ネットワーク上の枝の数も n となる．補題2より，高々2個のホールをもつ配置集合を考えれば十分である．よって，1つの連続した配置集合を選ぶ選び方は高々$O(n^2)$ 通り存在する．同様に，2つのそれぞれ連続した配置集合を選ぶ選び方は高々$O(n^4)$ 通り存在する．このすべての配置集合のパターンに対するコストを求め最小コストのものを選択するという列挙法により最適配置を求めることができる． □

5 あとがき

本稿では，リングネットワークの各枝の通信コストが任意の場合の最適データ配置手法に関する考察を行った．また，列挙法に基づくアルゴリズムを SUN-4/IPX 上に実現し実験を行い，20ノードからなるリングネットワークに関して，約2500秒で最適配置が求まった．今後の課題としては，ツリー，ハイパーキューブ等の形状を持ったネットワークに関するコピーの最適配置について考察を行うことがある．

文献

[1] O. Wolfson and A. Milo, "The multicast policy and its relationship to replicated data placement," *ACM Trans. Database Syst.*, Vol.16, no.1, pp.181-205(1991).