クラスタ分割されたネットワークのノード故障対策について

5X-4

石田　賢治　　宮尾　淳一　　菊野　亨　　吉田　典可

広　島　大　学

1．まえがき

本稿では，クラスタ分割されたコンピュータネットワーク[1]-[2]上のコントロールノードにおける故障対策の内で，特に障害回復に関する基礎的考察を行う．

2．クラスタ分割されたネットワーク

コンピュータネットワークの大規模化に伴い，ネットワークの効率良い管理は益々困難になりつつある．1つの解決策に，ネットワークをクラスタに分割し，クラスタごとに管理する方法がある(図1参照)．

各クラスタiにはコントロールノードC_iと幾つかのボーダノードが存在する．クラスタi内のノードはC_iによって集中的に管理される．但し，各クラスタ間は分散的に管理さる．

これらのネットワーク管理に必要な情報は（1）広域的なネットワーク情報を含む表TGと，(2)クラスタiに固有なネットワーク情報を含む表TL_iとして保持される．表の保持方法に関して次の仮定をおく．

H1　クラスタiのコントロールノードC_iは表$TG \cup TL_i$を保持している．

H2　クラスタiの各ボーダノードは表TL_iを保持している．

3．故障回復

最近，ソフトウェアの誤りやオペレータの誤動作が問題となりつつある[3]．ここでは，コントロールノード上の表$T=TG \cup TL_i$が消失した場合を想定し，表の回復方法について議論する．

今の場合，他のコントロールノードからTGを転送し，クラスタi内のボーダノードからTL$_i$を転送すれば，c_iでは表$TG \cup TL_i$を再構成できる．ここでは，TG（あるいはTL$_i$)それ自体も複数個の表から構成されていると仮定する．

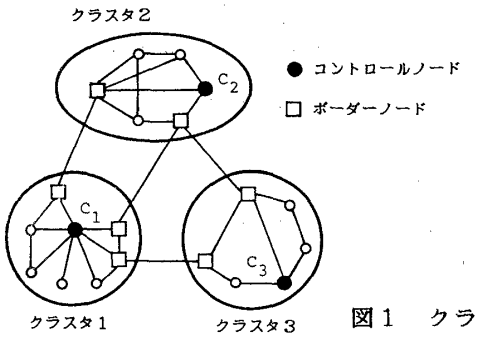

図1　クラスタ分割

4．問題の定式化

以降，ネットワーク上のノードv_0の表Tが消失し，ノードv_1，v_2，…，v_n上に（コピーとして）重複して置かれているTを利用して回復するものとする．

ノードv_0とv_iの間のリンクをℓ_iで表し，ℓ_iに関し次のパラメータを導入する．

$r(\ell_i)$ …ℓ_iを設定するのに要するコスト(時間)

$d(\ell_i)$ …単位パケットをℓ_i上で転送するのに要するコスト

更に，表Tを$T=\{t_1, t_2, \cdots, t_m\}$と仮定する．ここで，各$t_j$はいずれも表である．各$t_j$に対し，次のパラメータを導入する．

$s(t_j)$ …表t_jの転送に必要なパケットの総数

ここでは，各ノードv_iへの表(の集合)の割当て$A(v_i)(\in 2^T)$を決定する問題として定式化する．すなわち，各v_iからv_0

Reconstruction of tables in a control node for cluster-based computer networks

Kenji ISHIDA, Jun'ichi MIYAO, Tohru KIKUNO and Noriyoshi YOSHIDA

Hiroshima University

に表$A(v_i)$をリンクℓ_i上を転送することで，v_0上に表Tを再構成する．ここでは，Aを評価するため，次の評価関数D(A)を導入する．

$$D(A)=\max_{1\le i\le n}\{d(\ell_i)\cdot\sum_{t_j\in A(v_i)}s(t_j)+r(\ell_i)\}$$

次に，回復問題RTの定式化を与える．

［回復問題RT］ 次の①～④が与えられたとき，評価関数D(A)の値が最小となる割当てAを求めよ．

① $L=\{\ell_i \mid 1\le i\le n\}$

② $r(\ell_i),\ d(\ell_i)\in Z^+$

③ $T=\{t_j \mid 1\le j\le m\}$

④ $s(t_j)\in Z^+$

［例１］ 回復問題RTの例として次の①～④について考えてみる．

① $L=\{\ell_1,\ \ell_2,\ \ell_3,\ \ell_4\}$

② $r(\ell_1)=5,\ r(\ell_2)=6,$
$r(\ell_3)=1,\ r(\ell_4)=3,$
$d(\ell_1)=3,\ d(\ell_2)=4,$
$d(\ell_3)=1,\ d(\ell_4)=2.$

③ $T=\{t_1,\ t_2,\ t_3,\ t_4,\ t_5\}$

④ $s(t_1)=9,\ s(t_2)=10,$
$s(t_3)=8,\ s(t_4)=12,$
$s(t_5)=20.$

これに対し最適解として$A=(\{t_2\},\ \phi,\ \{t_4,\ t_5\},\ \{t_1,\ t_3\})$が求まる．このときD(A)＝37となる．

５．ヒューリスティック解法

回復問題RTの時間計算複雑さに関し，次の命題が導ける．

［命題１］ 回復問題RTはNP-困難である．

命題１より，回復問題RTに対するヒューリスティック解法の開発が重要になる．以降では，簡単化のため，$d(\ell_1)<d(\ell_j),\ j\neq 1$と仮定する．更に，$A(v_i)$を系列として考え，例えば$A(v_i)=(t_s,\ t_{s+1},\ \cdots,\ t_p)$と表す．この系列に対し，先頭の要素をTOP(A($v_i$))，要素$t_s$を削除する操作を$A(v_i)-t_s$，要素$t_s$を系列の最後に追加する操作を$A(v_i)+t_s$で表す．

［アルゴリズムRT］

ステップ１（初期解） …集合T上の系列$\overline{T}=(t_{i1},\ t_{i2},\ \cdots,\ t_{im})$を構成する．但し，$s(t_{ij})\le s(t_{ik})$，$j<k$とする．次に，初期解としてA＝($\overline{T}$，λ，…，λ）とおく．

ステップ２（逐次改良） …この時点での解を$A'=(a'_1,\ a'_2,\ \cdots,\ a'_n)$とする．

2.1（候補の決定） 関数$H(j)=r(\ell_j)+d(\ell_j)*(\sum_{t\in A'(v_j)}s(t)+s(TOP(a'_1)))$，$2\le j\le n$を計算し，最小の値を与える添字を１つ選ぶ(今，それをj^*とする)．

2.2（解の更新） 新しく$A=(a'_1-TOP(a'_1),\ a'_2,\ \cdots,\ a_{j^*}+TOP(a'_1),\ \cdots,\ a'_n)$を解とする．

2.3（判定） D(A)＜D(A')ならステップ２を繰り返し，それ以外ならA'を最終解として停止する．

［例２］ 例１と同じ問題に対し，上述のアルゴリズムRTを適用すると，解としてA＝($\{t_1\}$，ϕ，$\{t_4,t_5\}$，$\{t_2,t_3\}$)が求まる．このときD(A)＝３９（最適解は３７）となる．

６．むすび

アルゴリズムRTを日本・データゼネラル社のECLIPSE MV/4000上でC言語を用いて実現した．４０個のデータについてシミュレーション実験を行った結果，最適解（分岐限定法による）との相対誤差の平均値は約１９％であった．

現在，耐故障性を向上させるための表の分散配置方法，故障時のネットワーク再構成の方法について検討中である．

文献

[1] Kikuno, T., et al.:"A cluster-based approach to fault-tolerant computer networks", Proc. ISCAS-86, pp.630-633 (1986).

[2] Perlman, R.:"Hierarchical networks and subnetwork partition problem", Computer Networks and ISDN Systems, 9, 4, pp.297-303 (1985).

[3] 当麻喜弘:"フォールトトレラント技術の新フロンティア", 信学技報, FTS85-1, pp.1-8 (1985).