

2H-5

疎結合計算機システムのシステム異常対策 (II)

志賀浩一, 橋詰保彦

(富士通株式会社)

1. はじめに

AIM/SRCFは、フォールト・トレラント・データベース・システムを実現する機能である。すなわち、AIM/SRCFは、システム間通信路で結合された複数の計算機を制御することにより、一台の仮想的な計算機を構築する。そして、この仮想的な計算機の連続運用を可能とする。以下、この計算機群を、“制御網”と呼ぶことにする。

AIM/SRCFでは、システム間共用データベースの排他方式として、多数決方式を採用しているため、制御網内のある計算機がシステムダウン等の状態に陥っても、他の正常計算機では、業務処理が継続される。

本講演では、AIM/SRCFを利用してホットスタンバイシステムを構築した時に、現用システムのシステム異常が、制御網のトランザクション処理量に及ぼす影響について述べる。

2. 異常計算機の切離し処理

ここでは、異常計算機の切離し処理の要点について述べる。“切離し”処理とは、制御網内のある計算機が、ある計算機をダウンしたと見なした場合、その計算機を制御網外へ追出す処理を言う。

(1) AIM/SRCFシステムでは、制御網内の計算機は一定間隔 (T1秒) で相互に通信を行い、相互監視を行う。

(2) ある計算機Xに対し、T1秒以上の通信途絶を感知した場合、さらに、通信途絶に関して総ての計算機と同期をとる。これにT2秒要する。

(3) 計算機Xに対する通信途絶の同期が取れた場合、その計算機Xを異常計算機と見なし、制御網外の計算機とする。

異常計算機が使用していた共用データベースの部分は、切離し処理において閉塞される。この部分は復旧作業が完了するまでアクセスできない。

3. システム間排他方式

ある計算機がAIM/SRCFの制御対象となり、制御網の一計算機となる事を、制御網に“参入”と呼び、その逆、AIM/SRCFの制御対象外となり制御網の一計算機でなくなる事を、制御網から“離脱する”と呼ぶ。

制御網内の計算機は、ある計算機から送信されて来たある共用資源に関するアクセス要求に対し、自分の共用資源の管理情報を参照して次の投票をする；

- ・承認の時、OK投票を1票
- ・否認の時、NG投票を1票

投票論理 (“代行投票プロトコル”)

- ・ 離脱処理途中の状態と認識された計算機に対しては、OK投票0.5票、NG投票 0.5票 を、正常な計算機が代行投票する。
- ・ 参入、切離し処理途中の状態と認識された計算機に対しては、NG投票 1票を正常な計算機が代行投票する。

再審論理

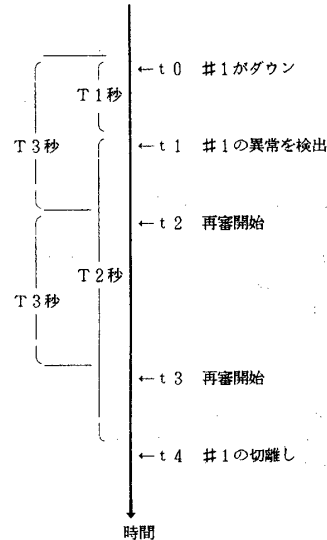
- ・ T3秒間に多数決判定がなされない時には、T3秒毎に占有要求を繰返す。

4. 切り離し処理中のトランザクション数見積り

簡単のため、次の様なモデルを設定する；

- (1) 制御網を構成する計算機は#1、#2、#3の3台とし、各計算機は同等の性能を持ち、同等の処理を行うものとする。
- (2) T1、T2、T3の間の大小関係は、次の通りとする；
 - ・ $T1 < T3 < T2$
- (3) 1トランザクションにおいて、必要なデータの占有回数は2回とする。

計算機#1がダウンする次の場合を考える；



この時、計算機#1に対する通信途絶の同期が完了するまでは、多数決回覧は確率的に送信される。従って、 t_0 と t_2 の間、計算機#2、#3が処理するトランザクション件数は、通常と比べて、次のようになる；

$$(1/2) \times (1/2) = 1/4$$

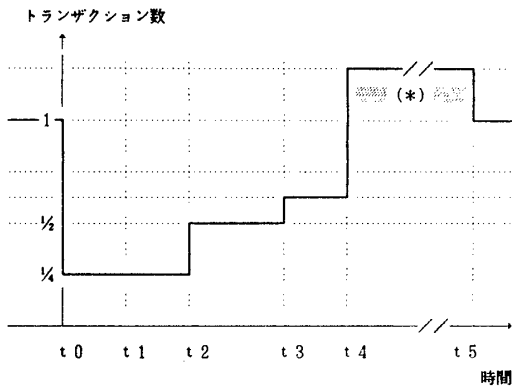
#1のダウンから T 3秒経過した t_2 からは、再審処理が開始される。再審される占有要求は、1回目と2回目の多数決回覧のうちの#1へ送信された占有要求である。従って、 t_2 と t_3 の間、計算機#2、#3が処理するトランザクション件数は、通常と比べて次のようになる；

$$(1/4) + (1/8) + (1/8) = 1/2$$

さらに T 3秒経過した t_3 からは、前回の再審処理で判定されなかった占有要求が処理対象となる。上と同様の計算を行えば、 t_3 と t_4 の間の計算機#2、#3が処理するトランザクション件数は、通常と比べて、次のようになる；

$$5/8$$

t_4 からは、代行投票プロトコルが働き、#1に多数決回覧が送信されることはない。
以上を、グラフに表せば次のようになる。



上図の(*)は、 t_4 から t_5 の間では、切離し処理が開始されるまでの間に滞留したトランザクションが、処理されることを示している。

各計算機の通常の処理におけるCPU使用率が80%であるとすると、制御網として、通常の状態に戻る時間が次のように計算される；

$$(3 \times t_4 + t_3 + 2 \times t_2 - 6 \times t_0) / 2$$

より性能の高いフォールト・トレラント・システムを実現するためには、システム異常の検出時間の短縮化と、システム切替え時間の短縮化を目指すなければならない。

5. 制御網の再構成時のCPU使用率

多数決論理による各種の判定処理、及び通信処理量を、多数決論理処理量と呼ぶことにする。

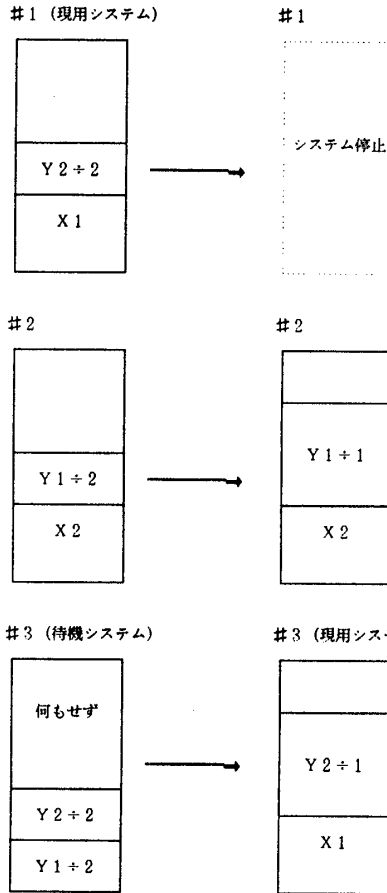
制御網の再構成処理が行われたとき、多数決論理処理量に変化がある。最後に、CPU処理能力の異なるシステムで制御網を構成したときの、多数決論理処理量の変化について、報告する。

X_i をシステム#Iで、ある資源を占有するために要する多数決論理処理量とする。

Y_i を他のシステムからの占有要求を処理するために要する多数決論理処理量とする。

制御網を、#1、#2、#3の3台で構成していて、#3は#1のホットスタンバイシステムであるとする。このとき、#1がシステムダウンしたと仮定する。

多数決論理処理量は、次のように変化する；



6. おわりに

今後も、ユーザのニーズを踏まえて、より信頼性が高く、より高性能を持ったフォールト・トレラント・データベース・システムへと発展させて行く所存である。