エッセンス情報引継ぎ方式による OSのフォールトトレラント化

6G-6

黒羽法男, 加藤匡史, 田中茂
富士通 (株)

1. はじめに

故障(fault)発生によるシステム稼働中止を防止し，アベイラビリティを大幅に向上させるため，冗長化による方式を採用したFTC(Fault Tolerant Computer)が提案され，製品化されている. [Serlin 84]

ソフトウェアにおいてとられている処理継続の方法は，トランザクション・リカバリに見られる論理的な処理継続方法と，プロセスの二重化に見られる物理的な処理継続方法に大きく分けられる.

本稿では，OSのソフトウェアのFT化のために考案した「エッセンス引継ぎ」による論理的な処理引継ぎ方式について報告する.

2. 従来の引継ぎ方式

従来，故障発生時に処理を継続するため，とられてきた手法を大きく分けると次の2つになる.

方式1) トランザクション・リカバリによる方法

トランザクション仕掛中の現用系で故障が起きると，リカバリ機構により仕掛中のトランザクションのアトミシティを保証した後，待機系で新たなサービスを再開する. [AIM ホットスタンバイ]

方式2) チェックポイント・リスタートによる方法

チェックポイントの度に，現用系から待機系へ変更情報を転送する. 現用系で故障が起きると，待機系で仕掛中の状態から処理を再開する. [Bartlett 81]

方式1を，本来トランザクションを適用する必要のない資源に対して，そのまま適用するのは，定常性能・引継ぎ性能の点で得策ではない. 方式2は，引継ぎのためのオーバヘッドが定常性能（レスポンスとスループット）に与える影響が大きい点に問題がある. また，現用系の故障の原因となったソフトウェアのバグを待機系に引き継ぐ可能性が高い点にも問題がある.

3. エッセンス引継ぎ方式

論理的な引継ぎにより，定常性能に与える引継ぎのためのオーバヘッドを削減し，ソフトウェアの故障に対してもFTとしたエッセンス引継ぎ方式を提案する.

3.1 引継ぎ情報の整理

現用系が故障を起こし，待機系に引き継がれる際，失われるのは現用系のメモリである. 従来方式2では，メモリそのものを二重化することにより仕掛中の処理を継続することで，この問題を解決していた.

しかし，現用系の故障を要求元に見せずにサービスを継続するためには，現用系の状態をすべて引き継ぐ必要はなく，現用系が提供する個々のサービスに対して，トランザクションの性質 [Gray 81] の中の次の2つの性質を保証すればよいことが分かった.

1) 持続性(durability)

ひとたびサービスが完了すると，故障が起きてもサービスの効果は保存される.

2) アトミシティ(atomicity)

サービスは完全に為されるか，一切為されないか(all or nothing)のいずれかである.

上記1) 2)を保証するために，待機系に引き継ぐ必要のある情報をエッセンス情報と呼び，次の2つに集約できる.

α) 現用系が要求元に対して保証したサービスの効果をメモリ上に保持している場合，その情報

β) 現用系が仕掛中のサービスのアトミシティを保証するための情報

言い換えれば，上記α情報・β情報以外の以下の情報は引き継ぐ必要がない.

a) 自プログラムが走行するために必要な情報

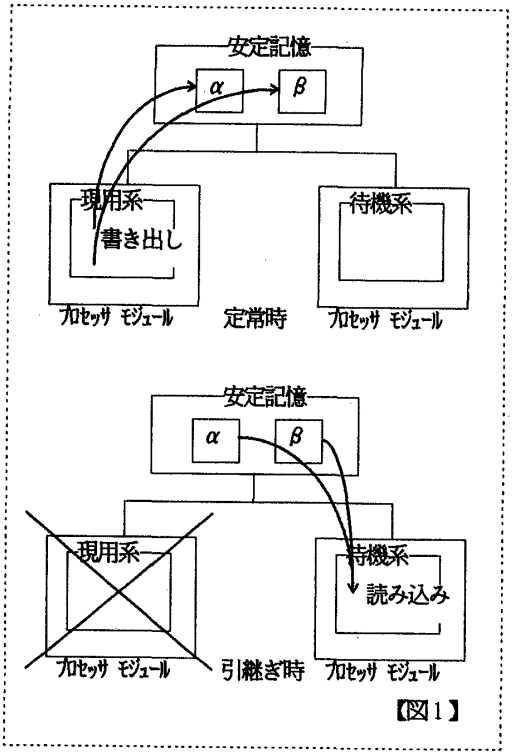
b) 処理途中で生成される中間的な情報

3.2 引継ぎ動作

次に，引継ぎの動作を一般的に説明する. (図1)

現用系は，定常処理時に上記α情報・β情報を安定記憶上に書き出す. 待機系では，引継ぎ時に安定記憶上からα情報・β情報を読み込み，α情報により現用系で保証したサービスの効果を保証し，β情報により仕掛中処理のアトミシティを保証した後，サービスを再開する. アトミシティを保証する場合，トランザクションのようにundo, redoの両方可能とする必要はなく，undo, redoのどちらか一方が可能であれば良い.

A Fault Tolerant Operating System using Essential Recovery Data
Norio KUROBANE, Tadaomi KATO, Shigeru TANAKA
FUJITSU, Ltd.



【図1】

3.3 実験システムの例

実験システムで，ファイル創成・削除等機能とアクセス機能を提供するファイルサーバを例に説明する．

3.3.1 α情報の例

ファイルのオープン依頼に対して，ファイルサーバでは，ファイルにアクセスするための権限をチェックし，以降アクセス時に使用するid（ファイルアクセスid）を要求元に返す．このファイルアクセスidとオープンしたファイルを関係付けるための情報（オープンファイル管理情報）をメモリ上に置く．このとき，主プロセスの故障により，要求元からのアクセスが不能にならないように，オープンファイル管理情報（α情報）を安定記憶に書き出す．オープンファイル管理情報は，ファイルアクセスid，ファイルidから成り，数バイトの大きさである．

3.3.2 β情報の例

ファイル創成処理では，スペースを獲得するためスペース管理情報を更新した後，名前を登録するため名前管理情報の更新を行う．スペース管理情報更新後，主プロセスの処理が中断したことにより，スペース管理情報と名前管理情報との間に矛盾を生じさせないため，引継ぎ時は，待機系で，どの管理情報を更新したかを意識してアトミシティを保証する必要がある．そのために，スペース管理情報更新に先立って，どの情報を更新するか（β情報）を安定記憶に書き出す．β情報の大きさは，数10バイトである．

3.3.3 引継ぎ情報整理の効果

実験システムのファイルサーバにおいて，使用メモリ全体に占めるα情報の割合は約6％であることが分かった．また，β情報の獲得は1回のサービスに対して約2回であり，1回の情報量が使用メモリ全体に占める割合は1％にも満たないことが分かった．

4．エッセンス引継ぎ方式の特徴

従来方式に比較して，以下の特徴がある．

1）定常処理時のオーバヘッドが小さい

エッセンス情報のみ引き継ぐことにより，引き継ぐ情報量を大幅に削減できる．

2）ソフトウェア障害に対してもFTである

ハードウェア（プロセッサモジュール）の故障に対してFT化されている他，エッセンスによる引継ぎをとっているため，ソフトウェア障害に対してもFTである．

3）引継ぎ情報取得のタイミングが容易につかめる

マルチプロセッサをサポートするためマルチスレッド処理を行っている場合でも，個々のスレッド毎に引継ぎ情報を取得するタイミングをつかめる．

4）予備プロセスが動作していることを前提としない

冗長化が主プロセスと安定記憶との間で実現されているので，予備プロセスの創成タイミングが自由になり，新しい版数のプロセスプログラムの採用が可能である．これは，ＯＳ資源の削減や，システム稼働中のソフトウェアの保守に道を開く効果がある．

5．おわりに

本稿では，引継ぎ情報を必要最小限に押さえた論理的な処理継続方式である「エッセンス引継ぎ方式」について報告し，それがＯＳソフトウェアの引継ぎ方式として優れていることを示した．

【参考文献】

[Bartlett 81] Bartlet, J. F. :A NonStop Kernel, Proc. 8th Symp. on Operating System Princioles, Dec. pp. 22-29 (1981).

[Gray 81] Gray, J. N. : The Transaction Concept : Virtues and Limitations Proc 7th Int. Conf. on Very Large Databases, Sept. (1981).

[Serlin 84] Serlin, O. :Fault-Tolerant Systems in Commercial Applications, IEEE. Computer. Aug. pp. 19-30 (1984).

[AIM ホットスタンバイ] FACOM OS IV/F4 MSP SUPホットスタンバイ説明書，富士通㈱ (1988).