

ホスト計算機の高速切替方式とそのプロトコルの提案†

平田俊明^{††} 田中光一^{††} 田中一喜^{†††}

本論文では、データベースを共有する複数の現用機および1台のホットスタンバイ機からなるホスト計算機システムと複数の利用者端末が接続される分散計算機システムからなる通信システムを対象とし、ホスト計算機障害時の交代計算機への高速切替方式およびこれを実現するホスト計算機-分散計算機間の通信プロトコルに関して考察する。従来のホットスタンバイ方式では、システムの大規模に伴い、(1)ジャーナルによる状態引継ぎ処理時間、(2)ホットスタンバイ機-分散計算機間の回線接続時間、が増大し利用者から見た業務停止時間が無視できなくなっている。本論文では、上記問題点(1)を解決するため、次の2段階の切替方式を提案する。現用機障害発生時、まず、別の業務を実施している他の現用機が状態引継ぎ処理を実施することなく高速に障害の発生した現用機上での業務を引き継ぎ、肩代わり処理を行う。次に、ホットスタンバイ機での状態引継ぎ処理終了時、さらにホットスタンバイ機が前記業務を引き継ぐ。ここで、第2段階の切替を早急にかつ利用者から見て業務が継続しているように見せることによりトータルとして利用者から見た業務停止時間を無視できる程度にできる。問題点(2)は、ホットスタンバイ機-分散計算機間の回線をあらかじめ接続しておくことにより対処する。提案方式をオンラインシステムの典型的なモデル上で数値的に評価した結果、提案方式の有効性が確認できた。

1. 緒言

近年、大規模化かつ複雑化するネットワークシステムにおいては、障害に対する社会的影響は重大であり、システムに障害が発生したときの回復時間の短縮は、最重要課題の1つとなっている。特に、銀行・証券等においては、ネットワークシステムに障害が発生した場合、その業務停止時間の大きさによっては、取引の機会を失うことにより重大な損失を招くこととなる。

本論文では、上記ニーズを踏まえ、ホスト計算機(センタ設置の汎用大型計算機、以降ホストと略記)、分散計算機(支店等に設置される部門計算機、以降分散機と略記)からなるオンライントランザクション処理を対象とした分散処理システムにおいて、ホスト計算機障害時の交代計算機への高速切替方式およびこれを実現するためのホスト-分散機間の通信プロトコルに関して考察する。

本論文では、データベースを共有する複数の現用機、1台のホットスタンバイ機からなるホスト計算機システム、および複数の利用者端末が接続される分散計算機システムとからなる通信システムを対象とする。ホスト計算機システムは、通常運用時、現用機、

ホットスタンバイ機間で、相互に監視し、また、各現用機では各種障害に備えシステム履歴情報(以下、ジャーナルと記す)を取得する。

コンピュータシステムの高信頼化のため多くの手法が提案されている¹⁾。これらには、(1)三重多数決、(2)現用系に対して待機系を持ち、現用系障害時待機系に切り替える待機交換手法、(3)二重化法による誤り検出と待機交換手法を組み合わせたペア・アンド・スベア手法等がある。これらは、特に業務の停止が許容できない制御用計算機、また一部のトランザクション処理向け商用フォールトトレラントシステムにおいて実用化されているが²⁾⁻⁵⁾経済的理由から1 MIPS程度のマイクロプロセッサを多重化した形態が多い。

一方、高速、大容量の汎用大型計算機では、多くの場合、上記制御用計算機のような厳しい信頼性は要求されないこと、また経済的理由から一般に待機交換手法を用いている。このうち、待機系を現用系と同期させて動作させ、処理の中断なく待機系に切り替える方式も提案されているが⁶⁾実用化レベルではまだ普及していない。現状では、複数の現用系に対して1台の待機系を持ち、現用系障害時、ジャーナルを用いた状態回復処理の後、待機系を動作させる方式(以降、本方式をホットスタンバイ方式と記す)が一般的である⁶⁾⁻¹⁰⁾。

本論文では、上記ホットスタンバイ方式を対象とし、従来方式の問題点を示し、これに対する解決策を提案する。

従来のホットスタンバイ方式では、現用機障害時、

† A Study for Methods of Switchover to Backup and Standby Computer by TOSHIKI HIRATA, KOICHI TANAKA (Systems Development Laboratory, Hitachi, Ltd.) and KAZUKI TANAKA (Software Works, Hitachi, Ltd.).

†† (株)日立製作所システム開発研究所

††† (株)日立製作所ソフトウェア工場

ホットスタンバイ機は状態回復処理（ジャーナルによる状態引き継ぎ処理）後、分散機と回線を再接続することにより業務を再開している。しかし、本方式では、システムの大規模化に伴い(1)ホットスタンバイ機での状態回復処理完了までの時間、(2)ホットスタンバイ機-分散機間の回線接続時間、が増大し利用者から見た業務停止時間が無視できなくなっている。

本論文では、上記問題点(1)を解決するため次の2段階の切替方式を提案する。現用機障害発生時、まず、別の業務を実施している他の現用機が状態回復処理を実施することなく高速に障害の発生した現用機上での業務を引き継ぎ、肩代わり処理を行う。次に、ホットスタンバイ機での状態回復処理終了時、さらにホットスタンバイ機が前記業務を引き継ぐ。

すなわち、上記肩代わり処理を実施する現用機（以下、「一時バックアップホスト」と記す）へは、ジャーナルを用いた状態回復処理を実施することなく高速に切り替えることにより、利用者から見た業務を短時間で回復する。一時バックアップホストでの肩代わり処理中は、利用者端末から見た応答が悪くなった状態でありかつジャーナルが障害ホストと異なるためこの間のファイル破壊等の障害には対処できない等の制限があるがホットスタンバイ機での状態回復処理終了時、一時バックアップホストからホットスタンバイ機へ切替を高速に行い、利用者から見て業務が継続しているように見せることによりトータルとして利用者から見た業務停止時間を無視できる程度にすることができる。

次に、前記(2)の問題に対処するため、ホットスタンバイ機-分散機間、一時バックアップホスト-分散機間ともあらかじめ回線を接続しておくこととした。

以下、第2章では対象とするシステム構成を示し、第3章では本論文で提案する切替方式を示し、第4章では第3章で示した切替方式を実現する通信プロトコルを提案し、これに対しオンラインシステムの典型的なネットワークモデル上で数値的評価を行う。

2. システム構成

本論文で対象とする分散処理システムは、図1に示すようにデータベースを共有する複数の現用機と1台のホットスタンバイ機からなるホスト計算機システムと複数の利用者端末が接続される分散計算機システム間の通信システムである。ホスト計算機システムは、通常運用時、現用機、ホットスタンバイ機間で相互に

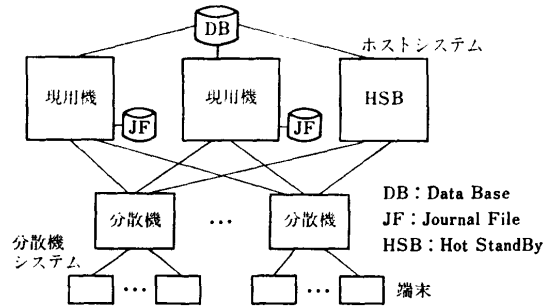


図1 対象とする分散処理システムの構成
Fig. 1 Configuration of distributed processing system.

監視し、また、各現用機では各種障害に備えジャーナルを取得している。ジャーナルは一般に次のような情報からなる⁷⁾。

(1) ファイルジャーナル

データベース更新に関する情報（更新前、更新後情報）

(2) 同期点ジャーナル

データベース更新の一連の処理の完了/未完了を示す情報

(3) 送信完了ジャーナル

端末へのメッセージ送信完/未完了を示す情報

これらのジャーナルの取得方法の一例を図2、図3に示す⁷⁾。図2では、ファイルの更新に関するすべての情報を得た後、ファイルを実際に更新する方式である。図3は、送信完了ジャーナルの取得方法を示す。

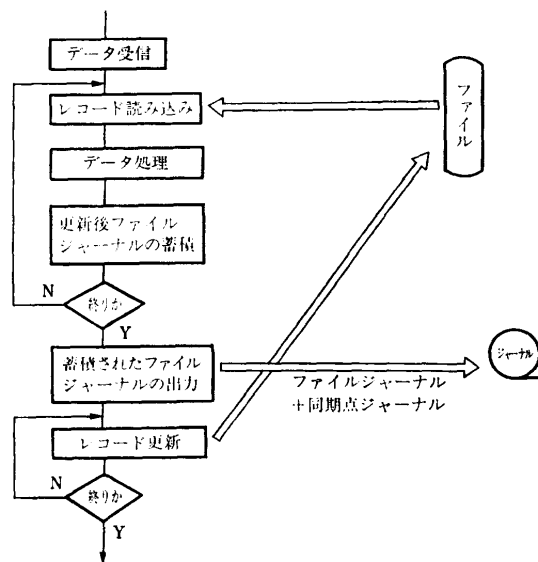


図2 ファイルジャーナル+同期点ジャーナル
Fig. 2 File journal and synchronous point journal.

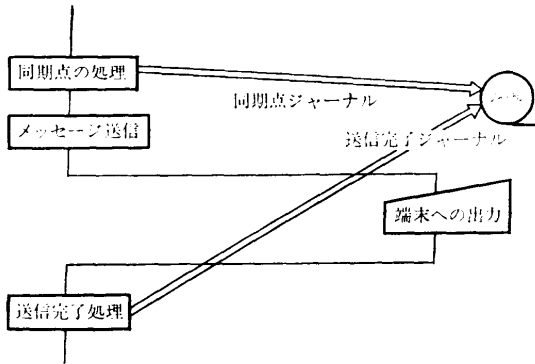


図3 送信完了ジャーナルの取得
Fig. 3 Transmission completion journal.

現用機障害時、ホットスタンバイ機では当該現用機のジャーナルを引き継ぎ（ジャーナルを格納する I/O 機器を物理的に切り替える）、これを用いてトランザクション処理の未完/完了を判断し、回復動作を行う。さらに、これらのジャーナルはデータベース障害時の回復処理にも使用可能であり、このとき、定期的に更新されるバックアップファイルの最新情報との差分情報（バックアップファイル更新以降のジャーナル）を保持しておく必要がある。

3. 切替方式

3.1 従来方式の問題点

従来のホットスタンバイシステムでは、図4(b)に示すように現用機障害時、ホットスタンバイ機は状態回復処理（ジャーナルによる状態引継ぎ処理）後、分散機と回線を再接続することにより業務を再開する^{6)~10)}。本方式ではシステムの大規模化に伴い下記の切替時間が増大し、利用者から見た業務停止時間が無視できなくなっている。

(1) ホットスタンバイ機での状態回復処理完了までの時間

(2) ホットスタンバイ機-分散機間の回線接続時間

3.2 提案方式

本論文では、3.1節で述べた問題点(1)を解決するため次の方式を提案する。図4(a)に示すように現用機障害発生時、ホットスタンバイ機での状態回復処理終了までの間、別の業務を実施している他の現用機（一時バックアップホスト）で障害の発生した現用機の業務を一時的に肩代わりし、

上記処理終了時、さらにホットスタンバイ機へ切り替える。すなわち、一時バックアップホストへの切替は、ジャーナルを用いた状態回復処理を実施せず高速に行うことにより利用者から見た業務を短時間で再開する。このとき、必要ならば中断したトランザクション処理を再開するために必要な情報（2章参照。ただし、仕掛り中トランザクションに関する情報のみ）は、複数ホストから参照可能なファイル（図1参照）を用いて引き継ぐ。

一時バックアップホストで時間稼ぎしている間にホットスタンバイ機の状態回復処理を実施する。一時バックアップホストによる肩代わり処理中は障害ホストと異なるジャーナルを用いているためこの間のファイル破壊等の障害には対処できない。また一時バックアップホストの負荷が大きくなるため利用者から見た応答時間が若干長くなる、という制限はある。しかし、ホットスタンバイ機での状態回復処理後は、4章で示す方法によりホットスタンバイ機へ直ちに業務を移すことができるため、利用者から見た業務への実質的な影響をなくせる。すなわち、ホットスタンバイ機への切替時、利用者には業務が継続しているように見せて、状態回復処理に係る時間を見かけ上無視できる程度とすることにより、高速切替を実現している。具体的には、現用機から一時バックアップホストへの切替は障害により中断された業務を再開するため切替対象端末

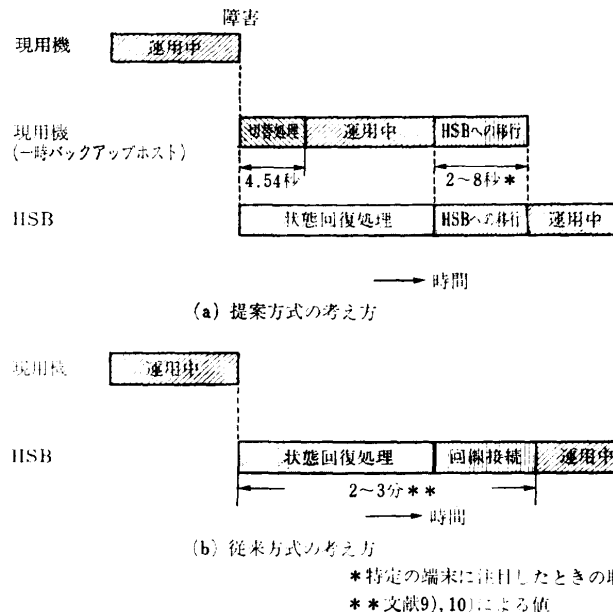


図4 提案方式および従来方式の考え方

Fig. 4 The idea of proposed method and past method.

* 特定の端末に注目したときの取引停止時間
** 文献9), 10)による値

の一括切替とする。また、一時バックアップホストからホットスタンバイ機への切替は端末利用者の業務を中断させないため端末ごとに一連の業務（トランザクション）の切目での切替とする。

一方、3.1節(2)の問題に対処するため一時バックアップホスト-分散機間、およびホットスタンバイ機-分散機間はあらかじめ回線を接続しておく。

4. 切替手順

本章では第3章で提案した切替方式を実現するためのホスト-分散機間の通信プロトコルを述べる。

4.1 ホスト-分散機間通信の論理的構成

本論文では、各端末は特定のホスト計算機とオンライン業務を行う負荷分散形態での運用および、以下のようなホスト-分散機間通信の構成を前提とする。

図5に示すように、ホスト-分散機間は回線で接続され、回線上には複数の論理的通信路が多重化されている構成とする。論理的通信路には、オンライン業務に使用する業務用論理的通信路と制御情報のやり取りに使用する（業務用論理的通信路の確立・解放等）制御用論理的通信路がある。前者は端末ごとに、後者は回線上に1本設定する。業務用論理的通信路は複数を一括して確立可能であり、また、確立後オンライン業務にて使用可能な状態または使用不可能な状態に設定可能である。図6に上記論理的通信路の状態遷移と使用するメッセージの種類を示す。一方、制御用論理的通信路は、回線接続時に確立する。

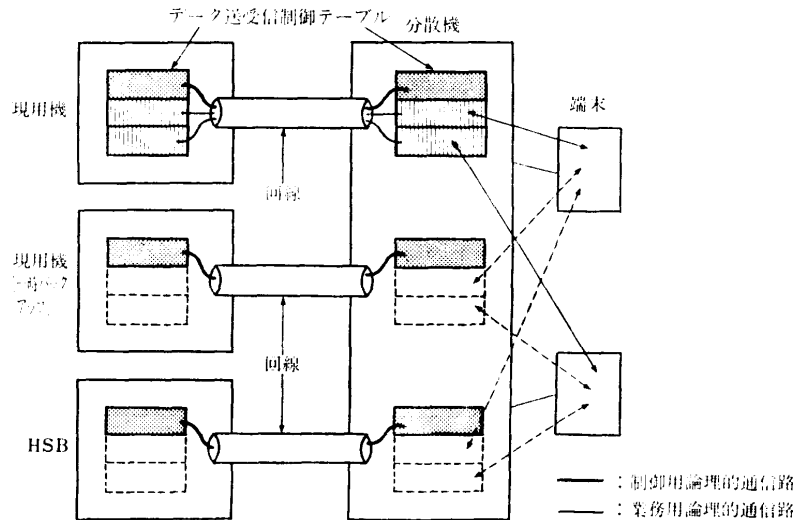


図5 ホスト分散機間の論理的構成
Fig. 5 Logical configuration between host computers and distributed processing computers.

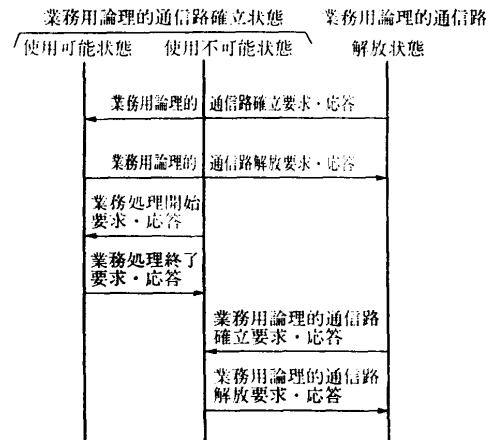


図6 論理的通信路の状態遷移
Fig. 6 State diagram of connection.

第3章で述べた切替方式では、通常運用時、分散機は、現用機、一時バックアップホストおよびホットスタンバイ機との回線を接続し、同時に前記回線上の制御用論理的通信路を確立しておく。一方、業務用論理的通信路はリソースの節約のため現用機-分散機間の回線上のみを確立しておくものとする。ここで、分散機は、配下の端末を2つのグループに分け、各グループで異なるホストに接続する運用を行えば、これらのホストで相互に一時バックアップを行う構成に設定でき、バックアップ用の回線を有効に利用できる。

4.2 切替手順

第3章で示した提案方式による現用機障害発生時の切替手順を次の2段階に分けて説明する。

(1) 現用機から一時バックアップホストへの切替

一時バックアップホストは、現用機の障害検出時、分散機との間で切替対象端末に対応する業務用論理的通信路の一括確立要求・応答メッセージをやり取りし、オンライン業務を再開する。

(2) 一時バックアップホストからホットスタンバイ機への切替

3.2節で示したように、この第2段階(2)の切替では、端末ごとに一連の業務の切目で切り替える方式として、2方式を比較検討する。

下記的方式2は、各端末ごとに個別に切替に必要なメッセージを

やり取りする方式であるのに対して、方式1は、ホットスタンバイ機での状態回復処理終了時、一連の業務処理が継続中でない端末は一括して切替に必要なメッセージをやり取りする方式である。

(a) 方式1

図7に示すように次の2フェーズからなる切替方式である。

(i) 一括切替フェーズ

本フェーズでは、一連の業務処理の切目にある端末を一括して切り替える。このため、分散機には、配下の各端末ごとに一連の業務処理が継続中か否かを示す制御ビットを設ける。

(イ) ホットスタンバイ機は状態回復処理終了時、分散機に対して切替対象端末に対応する業務用論理的通信路の一括確立要求メッセージを送信する。(図7①)

(ロ) 分散機は、業務用論理的通信路を確立し、以下の処理後、ホットスタンバイ機へ端末ごとに使用可/不可の状態を付加して応答メッセージを送信する(図7③)

●前記制御ビットを参照し、一連の業務処理が継続中でない端末(以降、非仕掛り端末と略記)に対しては、一時バックアップホストとの間で確立している業務用論理的通信路の一括解放要求・応答メッセージをやり取りする。(図7②)

●非仕掛り端末に対する業務用論理的通信路は使用可能状態とし、そうでない端末(以降、仕掛り端末と略記)については使用不可能状態とする。

(ii) 個別切替フェーズ

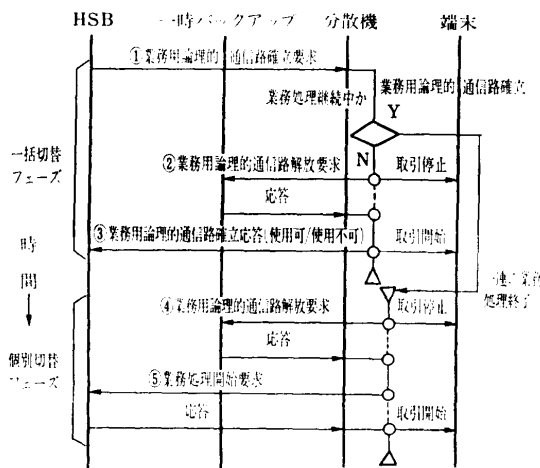


図7 切替手順(方式1)
Fig. 7 Sequence of method 1.

(i)の(ロ)で仕掛り端末については、仕掛り解除時(一連の業務処理終了時)、端末ごとに下記の処理を行う。

(イ) 一時バックアップホストとの間で確立している業務用論理的通信路の解放要求・応答メッセージをやり取りする。(図7④)

(ロ) ホットスタンバイ機と新たに確立し使用不可能状態となっている業務用論理的通信路について業務開始要求・応答メッセージをやり取りする。(図7⑤)

(b) 方式2

図8に切替手順を示す。

(i) ホットスタンバイ機は、状態回復処理終了時、ホットスタンバイ機から分散機へ切替対象端末に対する業務用論理的通信路の一括確立要求メッセージを送信する。(図8①)

(ii) 分散機では業務用論理的通信路を確立し、すべて使用不可能状態としてホットスタンバイ機へ応答メッセージを送信する。(図8②)

(iii) 端末ごとに、一連の業務処理の切目で方式1(ii)の(イ)、(ロ)と同様の処理を行う。(図8③、④)

4.3 切替時間の評価

本論文で提案した切替方式の有効性を図9に示すネットワーク構成例について、主として切替時間に注目して数値的に評価する。

4.3.1 前提条件

(1) システム構成とトラヒック量

図9にシステム構成を示す。ホストシステムは、現用機:ホットスタンバイ機=3台:1台とし、現用機は、他のいずれかの現用機の一時的バックアップとなる。分散機群は、専用線(4800 bps)でフロントエンドプロセッサ(Front End Processor, 以下 FEP と

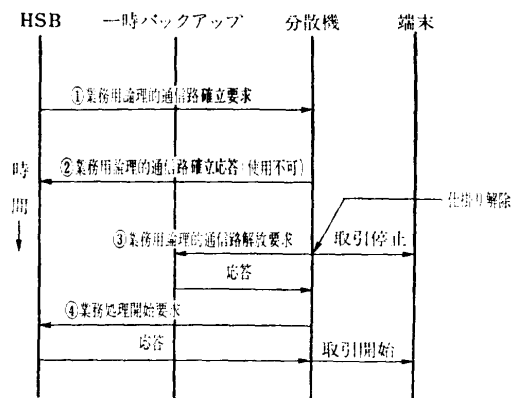
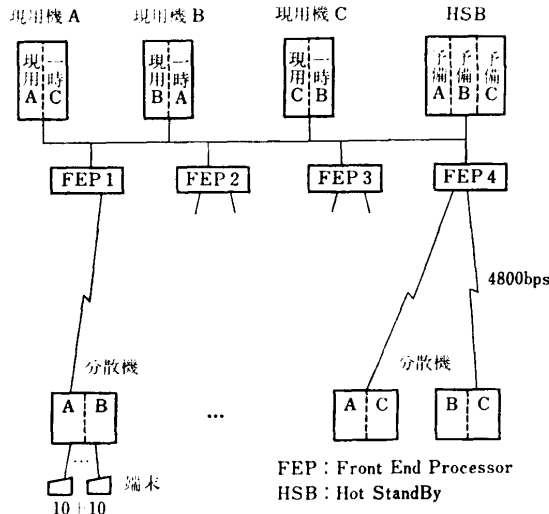


図8 切替手順(方式2)
Fig. 8 Sequence of method 2.



- 1) 分散機：300台
- 2) 端末：20台/分散機
- 3) 現用機は、1台当り分散機100台の処理を行う。
- 4) FEPは、1台当り分散機75台の処理を行う。

図9 システム構成例
Fig. 9 Example of system configuration.

略記) を介してホストシステムと接続される。なお、ホスト-FEP間はチャンネル(50KByte/s)で接続される。

(a) 分散機・端末数

分散機数：300
端末数：300×20=600

1ホストに接続される分散機は100台とする。また、1分散機に接続される端末は10端末ずつ異なるホストへ接続するものとする。

(b) 通常運用時のトラヒック量

1分散機あたり 0.5件/秒

(2) 評価対象切替処理

(a) 現用機から一時バックアップホストへの切替

(b) 一時バックアップホストからホットスタンバイ機への切替

(1)の前提の下では全分散機の2/3が切替の対象となる。また、対象分散機内の1/2の端末が切替の対象となる。

(3) 切替時のデータの流れ

障害が発生したホストについても回線への通常の負荷があるとする。

(4) メッセージ長

表1に示すメッセージ長を仮定する。

4.3.2 方法

1メッセージの分散機-ホスト間の平均転送時間 T_0

表1 メッセージ長
Table 1 Message length of the system.

メッセージ種別		メッセージ長
通常メッセージ	問合せ (端末→ホスト)	500 byte
	応答 (ホスト→端末)	500 byte
切替に使用するメッセージ	業務用論理的通信路確立要求・応答	300+50n byte
	業務用論理的通信路解放要求	300+50n byte
	業務用論理的通信路解放応答	300 byte
	業務処理開始要求・応答	300 byte

n: 1分散機当りの切替対象となる端末数

を計算する。 T_0 は、 T_1 を回線転送に関する時間、 T_2 をチャンネル転送に関する時間として、

$$T_0 = T_1 + T_2 \tag{1}$$

で計算する。ここで T_1, T_2 は、 t_1, t_2 をそれぞれ回線、チャンネルの通常トラヒックでの転送時間、 w_1, w_2 をそれぞれ回線、チャンネルでの転送待ち時間として、

$$T_1 = w_1 + t_1 \tag{2}$$

$$T_2 = w_2 + t_2 \tag{3}$$

となる。

なお、ホスト、FEP、分散機上でのメッセージ処理時間は上記要素に対して無視できる程度と判断し省略した。(4.4節参照)

次に、 w_1, t_1 を計算する。このため、回線をサービス窓口、回線上の平均トラヒックを窓口への平均到着時間間隔、回線の転送時間をサービス時間として M/M/1 モデルを適用する。 w_1 は、メッセージの平均到着時間間隔を $1/\lambda_1$ 、サービス時間を $1/\mu_1$ 、回線の利用率を ρ_1 とすると、

$$w_1 = \frac{\rho_1^2}{1-\rho_1} \times \frac{1}{\lambda_1} \tag{4}$$

$$\rho_1 = \frac{\lambda_1}{\mu_1} \tag{5}$$

$$\frac{1}{\mu_1} = \frac{L}{K} \tag{6}$$

となる¹³⁾。ここで、 L はメッセージ長、 K は回線速度とする。式(4)、(5)、(6)より

$$w_1 = \frac{\rho_1}{1-\rho_1} \times \frac{L}{K} \tag{7}$$

となる。また、

$$t_1 = \frac{1}{\mu_1} = \frac{L}{K} \tag{8}$$

である。なお、 w_2, t_2 も同様に計算できる。

4.3.1項の前提のもとで回線およびチャンネルの通常

表 2 データトラフィックと利用率
Table 2 Data traffic and utilization on line and channel.

種 別	データトラフィック (Byte/秒)	利用率
回 線	250	0.40
チャンネル	6250	0.12

トラフィックでの利用率を式(5)より計算した結果を(チャンネルの場合も同様)表2に示す。なお、表2は、チャンネルでの1メッセージの平均到着時間間隔は分散機からの全トラフィックの1/12が1本のチャンネルのトラフィックである(3台ある各現用機ごとに4台のFEPから均等のトラフィックがあるものとする)として計算した。

4.3.3 計算結果

従来方式および提案方式の現用機障害時の業務停止時間(切替時間)を計算する。各メッセージのホスト-分散機間の転送時間は、4.3.1項の前提のもとで式(1)に表1のメッセージ長を適用して計算した。

(1) 従来方式

従来のホットスタンバイ方式での業務停止時間は、次のようになる。

(業務停止時間)

$$= (\text{ホットスタンバイ機での状態回復時間}) \\ + (\text{回線接続時間}) \\ + (\text{業務用論理的通信路一括確立時間})$$

(ホットスタンバイ機での状態回復時間)+(回線接続時間)は、2~3分程度である^{9),10)}。ここで、業務用論理的通信路一括確立時間とは、ホットスタンバイ機-分散機間の業務用論理的通信路一括確立要求・応答メッセージのやり取りの時間であり、前記前提のもとでは、4.54秒となる。

(2) 提案方式

(a) 現用機から一時バックアップホストへの切替切替時間は、前記業務用論理的通信路一括確立時間のみで4.54秒となる。

(b) 一時バックアップホストからホットスタンバイ機への切替

4.2節で述べた2方式を比較する。

表3に方式1と方式2のそれぞれについてホットスタンバイ機の状態回復処理終了時、1分散機当りの仕掛り中の端末数に対する非仕掛り端末の切替完了までの時間(方式1では、図7③完了までの時間、方式2では、全非仕掛り端末の図8④完了までの時間)および使用メッセージ数を示す。なお、方式2における端

表 3 一時バックアップホストからホットスタンバイ機への切替時間、メッセージ数

Table 3 Time and message number of switchover from temporary backup computer to hot-standby computer.

仕掛り端末数	方式 1		方式 2	
	切替時間	メッセージ数	切替時間	メッセージ数
0	7.57	4	22.93	44
1	7.43	8	21.25	44
2	7.29	12	19.57	44
3	7.15	16	17.89	44
4	7.02	20	16.21	44
5	6.86	24	14.53	44
6	6.72	28	12.85	44
7	6.58	32	11.17	44
8	6.54	36	9.49	44
9	6.31	40	7.81	44
10	4.45	44	4.45	44

切替時間: 秒

末ごとのメッセージのやり取りは、1本の制御用論理的通信路上での送受信制御を単純にする意味から、端末ごとに順に行うものとして計算した。また、使用メッセージ数は、FEP-分散機間に公衆網を用いたときの課金に影響する。以上より、切替時間、メッセージ数とも非仕掛りの端末数が多いほど方式1が有効であることが分かる。方式1は、方式2に比べて端末ごとの仕掛り状態の管理が必要であるが、一般にオンライン業務においては、ある時点で一般にかなりの端末が業務処理の切目にある点に注目し、本論文では、方式1を提案する。

4.3.4 従来方式との比較

上記評価システムでは、一時バックアップホストからホットスタンバイ機への切替に方式1を用いた場合の特定の端末に注目したときの業務停止時間(現用機から一時バックアップホストへの切替時間+一時バックアップホストからホットスタンバイ機への切替時間)は、4.3.3項の計算方法より以下ようになる。

(1) 最小の場合

第2段階の切替で該当端末が個別切替フェーズでの切替対象となる場合で、4.54秒+1.68秒=6.22秒となる。

(2) 最大の場合

第2段階の切替で該当端末が一括切替フェーズでの

切替対象となりかつ 10 端末を一括して切り替える場合で、4.54 秒+7.57 秒=12.11 秒となる。

従来方式での業務停止時間は 2~3 分程度^{9),10)} 要していたことから提案方式では、業務停止時間が大幅に改善される可能性を示している。

4.4 一時バックアップホストでの運用中の制限事項に対する考察

提案方式では、一時バックアップホストで運用中、通常より負荷が大となり利用者から見た応答時間が若干長くなる恐れがあること、また、この間のファイル破壊には対応できないという制限がある。以下、上記 2 点の一時バックアップホストでの運用中の制限事項に対し実用上は問題ないことを数値的評価により示す。

4.4.1 応答時間に関する考察

本項では、一時バックアップホストでの運用中の負荷増大による端末利用者から見た応答時間への影響をホスト計算機上の処理時間を考慮し、4.3 節のモデルのもとで定量的に評価する。

(1) 前提条件

- (a) ホスト計算機の処理能力
30 MIPS
- (b) ディスクの入出力処理時間
15 msec/回
- (c) 1 トランザクション当りの負荷
200 Kstep+ディスクの入出力 5 回

その他の条件は、4.3 節のモデルと同一とする。

(2) 評価方法

端末利用者から見た平均応答時間 U を計算する。 U は、ホスト-分散機間の平均転送時間とホストでの平均処理時間の和として計算する。ホスト-分散機間の平均転送時間は、端末からホストへの問い合わせメッセージ長とホストから端末への応答メッセージ長は同一であるとしたことから式(1)の T を用い、 $2T$ となる。よって、ホスト計算機での平均処理時間を S とすると、端末利用者から見た応答時間 U は、

$$U = 2T + S \quad (9)$$

となる。 S は、 T_3 を CPU の処理に関する時間、 T_4 をディスク入出力に関する時間として、

$$S = T_3 + T_4 \quad (10)$$

で計算する。ここで、 T_3 、 T_4 は t_3 を CPU の処理時間、 w_3 を CPU の処理待ち時間、 t_4 を 1 回のディスクの入出力時間として、

$$T_3 = w_3 + t_3 \quad (11)$$

$$T_4 = t_4 \times 5 \quad (12)$$

と計算できる。

次に、 w_3 、 t_3 を計算する。このため、CPU をサービス窓口、単位時間当りの平均発生トランザクション数を窓口への平均到着時間間隔、CPU の処理時間をサービス時間とした M/M/1 モデルを適用する。 w_3 は CPU の利用率を ρ_3 、トランザクションの平均到着時間間隔を $1/\lambda_3$ 、サービス時間を $1/\mu_3$ とすると、

$$\frac{1}{\mu_3} = \frac{P}{M} \quad (13)$$

より、4.3.2 項と同様の手順により

$$w_3 = \frac{\rho_3}{1 - \rho_3} \times \frac{P}{M} \quad (14)$$

となる。ここで、 P は 1 トランザクション当りの処理ステップ数、 M は CPU の MIPS 値である。また、

$$t_3 = \frac{1}{\mu_3} = \frac{P}{M} \quad (15)$$

である。

(1) の前提条件のもとで計算したホスト計算機の負荷および CPU 利用率 ρ_3 を表 4 にまとめる。ここで、一時バックアップ時の負荷は、障害の発生した現用機の負荷がそのまま一時バックアップの負荷に加えられるものとした。また、CPU 利用率は 4.3.2 項と同様の方法で計算した。

(3) 評価結果

表 5 にホスト計算機上での処理時間とチャネル、回線上の転送時間に分けて評価結果を示す。これより、本評価システムでは、一時バックアップ中の性能劣化は、ほとんど問題にならない。一方、ホスト計算機の負荷は、通常時とピーク時では、大きな差があり現用

表 4 ホスト計算機の負荷
Table 4 Load of host computer.

	通常時	一時バックアップ時
負荷	50 件/秒	100 件/秒
利用率	33%	66%

表 5 1 トランザクション当りの応答時間
Table 5 Response time per a transaction.

	通常時	一時バックアップ時
ホスト計算機上の処理時間 (S)	85 msec	95 msec
回線、チャネルの転送時間 ($2T$)	2.80 sec	2.81 sec
合計	2.89 sec	2.90 sec

機に障害が発生した場合、一時バックアップホストが、過負荷となる可能性がある。しかし、本方式ではたとえ過負荷となっても応答時間が悪くなるだけで業務が中断されることはない。例えば、1分散機当りのトラヒックの発生量を0.74件/秒とすると一時バックアップ時のCPU利用率が99%となり、式(9)より応答時間は、9.98秒となる。

4.4.2 ファイル破壊に関する考察

提案方式では、一時バックアップホストでの運用中にファイル破壊が発生した場合、ジャーナルが障害ホストとは異なるため、障害発生時点では回復処理を行うことができない。しかし、一時バックアップホストでの運用中は、一時バックアップホストにてジャーナルを取得するため、この間のジャーナルを粉失することはない。よって、ホットスタンバイ機での状態回復処理終了後、各ホストで取得したジャーナルを用い、ファイル破壊の回復処理を実施することにより、業務停止時間を最悪でも従来方式と同程度にできる。

5. 結 言

ホスト計算機、分散計算機からなる分散処理システムにおいて、ホスト計算機障害時の高速切替方式およびこれを実現するためのプロトコルを提案した。また、提案方式をオンラインシステムの典型的なネットワークモデル上で数値的な評価を行った。これにより、提案方式は、従来方式と比較して業務停止時間が1/10~1/30程度となることを示した。

一方、提案方式における一時バックアップホストでの運用中は、一時バックアップホストの負荷が大となるが、特に、負荷のピーク時と重ならないかぎり利用者から見た応答時間にはほとんど影響しないことを示した。また、たとえピーク時と重なり過負荷となっても応答時間が悪くなるだけで業務が中断されることはない。

なお、一時バックアップホストでの切替処理では(1)仕掛り中の端末に対する応答電文の復旧処理、(2)仕掛り中のトランザクションで更新途中にあるレコードの復旧処理等を行う必要があるが、これらは基本的には2章で示した方式⁷⁾において複数ホストから参照可能なファイルを用いた場合と同様であると考えられる。しかし、状態回復処理などの詳細についてはさらに検討の必要があり、これらは今後の課題である。

本論文で提案したホスト計算機の高速切替方式は、ホスト計算機にかぎらず同種の計算機システムに対し

て広く利用することができると考えられる。

謝辞 日頃御指導頂く日立製作所システム開発研究所堂免所長、大町部長、柳生主任研究員、ソフトウェア工场上田部長に感謝します。

参 考 文 献

- 1) グレイ J. ほか：フォールトトレラントシステム、pp. 11-15、マグロウブック、東京 (1986)。
- 2) Wensley, J.H. et al.: SHIFT: Design and Analysis of a Fault Tolerant Computer for Aircraft Control, *Proc. IEEE*, Vol. 66, No. 10, pp. 1240-1255 (1978)。
- 3) Siewiorek, D., Bell, C.G. and Newell, A.: *Computer Structure: Principle and Examples*, McGraw-Hill Book Co. (1982)。
- 4) Katzman, J.A.: System Architecture for Non-Stop Computing, *Proc. 14th Computer Society Int. Conf. (Compcon 77)*, pp. 77-80, San Francisco (1977)。
- 5) Serlin, O.: *Fault Tolerant System*, ITOM International Co. (1982)。
- 6) Bahadra, K. and Schiffman, S.M.: Back up and Recovery in IBM Information Network, *AFIPS*, Vol. 52, pp. 671-674 (1983)。
- 7) 上田恭雄：オンラインシステム、pp. 74-120、昭晃堂、東京 (1987)。
- 8) 日経データプロ、汎用計算機信頼性向上の手法・システム編、ホットスタンバイシステム (1986)。
- 9) 石崎純夫編：実例コンピュータバンキング、No. 9, pp. 161-168、近代セールス社、東京 (1983)。
- 10) 石崎純夫編：実例コンピュータバンキング、No. 8, pp. 217-226、近代セールス社、東京 (1982)。
- 11) 森村英典、大前義次：応用待ち行列理論、pp. 24-32、日科技連出版社、東京 (1975)。

(昭和63年7月13日受付)

(平成元年4月11日採録)



平田 俊明 (正会員)

昭和36年生。昭和58年慶応義塾大学工学部計測工学科卒業。昭和60年同大学院修士課程修了。同年、(株)日立製作所入社。現在、同社システム開発研究所勤務。コンピュータネットワークに関する研究開発に従事。



田中 光一 (正会員)

1942年生. 1967年立教大学理学部物理学科卒業. 1988年8月まで(株)日立製作所システム開発研究所勤務. HNAの性能評価, ネットワーク管理, CCPの研究等に従事.

現在, (株)日立情報ネットワークに勤務. 電子情報通信学会会員.



田中 一喜 (正会員)

昭和27年生. 昭和46年佐賀県立塩田工業高校機械科卒業. 同年(株)日立製作所ソフトウェア工場入社. 金融機関用のオンラインシステムのシステム設計を皮切りに, 現在までDB/DCシステムのソフトウェアの開発, 設計に従事.