マルチメディア通信と分散処理 96－1
（2000．2．9）

広域分散ネットワークシステムにおける障害対策方式の設計

秋山 康智、田中 功一、柳生 理子
三菱電機（株）情報技術総合研究所

近年、コンピュータネットワークは、計算機の処理能力の急激な上昇に伴い、従来のホスト集中型である垂直分散環境から、各ジョブを複数の計算機で分散して行う水平分散環境へと移行している。さらに高速なネットワークによる情報通信システムの進歩が目覚しく、通信インフラの整備と共に水平分散環境の広域化が進んでいる。ネットワークの規模に伴い、その管理、運営は複雑かつ困難になってしまう。特に障害発生時の対処には専門知識が必要となるため、そのネットワークの安全性が管理者の技術レベルに大きく依存していた。本稿では特に複数のドメインで構成された広域分散ネットワーク上での障害管理の課題を明確化し、特に障害発生時の障害検知、通知方式について考察し、設計を行った。その結果、階層的構成の採用により広域分散ネットワークシステムに適応した、複数のサーバおよび回線を制御することにより障害に強い、正確な障害場所の検知が可能な障害検知方式および障害情報通知方式を得た。

Failure Measure Methods for A Wide-Area Distributed Network System

Koji Akiyama, Koichi Tanaka, Riko Yagyu
Information Technology R&D Center, Mitsubishi Electric Corporation

Recently Computer network is changing from vertical distribution style to horizontal one by drastic evolution of computing power. In additional to this evolution of information communication systems with high-speed networks make distributed system's size wider. So the wider the scale of it is , the more difficult and complicated the administration becomes. And the network administrators must have expert knowledge. In this paper, we discuss methods of network trouble checking and information notification for wide-area distributed systems made plural network domains. As the result, proposed failure points detection method can apply to to wide-area distributed network system by using hierarchical structure and plural network lines.

1　はじめに

近年、コンピュータネットワークは、計算機の処理能力の急激な上昇により、従来のホスト集中型である垂直分散環境から、各ジョブを複数の計算機で分散して行う水平分散環境へと移行している。さらに高速なネットワークによる情報通信システムの進歩が目覚しく、通信インフラの整備と共に水平分散環境の広域化が進んでいる。ネットワークの管理・運営は、その規模に従い複雑かつ困難になる。特に障害発生時の対処には専門知識が必要となるため、そのネットワークの安全性が管理者の技術レベルに大きく依存していた。このような状況を受けてこれまで管理者の負荷を軽減するための様々なツールや運営方式が提案されてきたが、そのほとんどが LAN(Local Area Network)レベルのものであった。本稿では特に複数のドメイン(LAN)で構成された広域ネットワーク上での運用管理、特に障害発生時の障害検知、通知方式ついて考察する。

2　背景、目的

今回我々が目指したのは、「優れた抗堪性」、さらに「高い構成柔軟性」を兼ね備えたシステムである。具体的には、障害が発生し、ジョブの継続が不可能となった場合、そのジョブを自動的に他の計算機上で再実行し、それをユーザが全く認識せず自動代替を行うことができるシステムを目指した。特に今回は、従来あまり注目されていなかった、複数のネットワークドメインで構成されている大規模広域分散ネットワークシステムをターゲットとした。また、ユーザからの要望もあり、「高信頼」および「抗堪性」を第一に考慮した方式を検討した。

上記目的を実現するために必要な技術を図１の様に検討した。

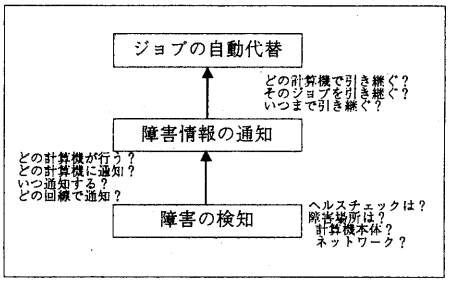

図１　障害対策の検討

ジョブの自動代替を行うためには、どの計算機がどのジョブを引き継ぐ等の問題がある。これを解決するには、どのジョブを行っていた計算機に障害が起きたのかを知る必要がある。つまり、まず障害場所を通知される必要があり、どの計算機が、どの回線を用いて、どの計算機に、どういうタイミングで通知するかが問題となる。また、障害場所を通知するためには、障害場所を検知する恒常的な監視が必要となり、ヘルスチェックにおいては、サーバをどのように設定するか、どの回線を用いるか、障害場所が計算機か回線かを明確に判別するためのチェック方式が問題となる。

本論文では、上記ヘルスチェック方式および、障害情報の通知方式において考察した結果を報告する。今回提案する方式で使用するプロトコルは、全て TCP/IP の上に位置するアプリケーション層のプロトコルである。

3　ヘルスチェック方式

ヘルスチェック方式において、以下のポリシーをに従い設計した。

・負荷分散を考慮したチェック方式
・抗堪性の高いチェック方式
・障害場所を明確に判別可能
・広域ネットワークに適用

3.1 負荷分散を考慮したチェック方式

ヘルスチェックは、障害発生時に即座にそれを判別・対応する必要があり、そのためには、ヘルスチェックサーバとなる計算機は、常にチェックをスムーズに行うための余力を維持する必要がある。一般的なネットワーク環境において、１つの計算機上で様々なジョブを実行しており、１つの計算機をサーバとして永続的に指定する場合、その計算機上で高い負荷を負う可能性のあるジョブを実行することができなくなり、また CPU 資産が使われずにいる計算機が存在している可能性も高い。よって、ヘルスチェックサーバを動的に設定する方式を採用した。本方式は以下の流れでサーバを選択する。

・ デフォルトのサーバを選択、次候補を複数優先度付で設定
・ 一定時間ごとにサーバの負荷を測定
・ サーバ負荷が設定した閾値以上であれば、次候補の計算機負荷を測定
・ 閾値以上であれば次々候補の負荷を測定、閾値以下であれば、サーバに設定
・ 新サーバが決定した際、これをサブネットワークドメイン内の各計算機に報告

また、管理するクライアント数によりサーバの負荷は一次関数的に増加してしまうという問題がある。これを解決するために、ネットワークドメイン内の管理計算機をグループに分け（サブドメイン化）、その中でヘルスチェックを行い、各サブドメイン内サーバは担当するサブドメインの障害情報を、ネットワークドメインのサーバに報告する、階層的な管理体系を取ることにした。これにより、複数のサーバで負荷を分散することができる。さらにネットワークの規模の増大に伴い、より複雑かつ困難になる管理を、各規模でのグループ化を行うことにより、より単純にかつ容易に行うことが可能となる。

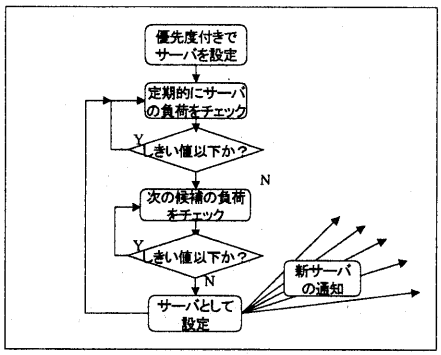

図２　サーバの動的設定

3.2 抗堪性の高いチェック方式

サーバ計算機や障害情報を伝える回線に障害が生じた場合でも、それを確実に判別し、ヘルスチェックを継続する必要がある。サーバ計算機自身に障害が生じた場合、速やかにサーバ計算機の代替を行う必要があるからである。また通信回線に障害が発生した場合、使用できる回線が他にあれば、早急にその回線に切り替える必要がある。もし使用回線が障害を起こした回線のみであった場合は、回線復旧までこの回線に接続された計算機は、ヘルスチェックのメンバーから削除されることになる。よって本方式を採用するネットワークドメイン内にお

いて各計算機は複数の接続回線を有するものとする。クライアントからサーバに障害情報を送付する際、回線が使用できなかった場合、以下の動作を行う。

・サーバに接続する他の回線を用いて障害情報を送信
・情報通知回線の登録設定
・情報を受信したサーバ上で、通信回線を再設定

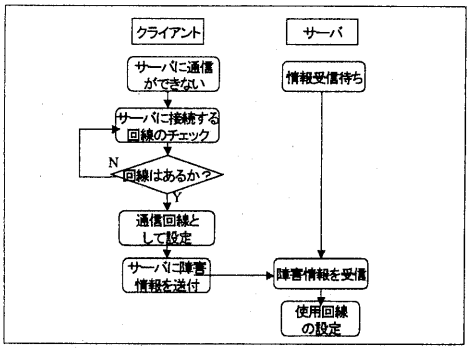

図３　ヘルスチェック回線の切り替え

また接続している各回線の使用優先度は、図４のように３つのテーブルをもとに設定される。３つのテーブルとは、回線速度、回線使用可能時間等が設定される物理要素テーブル、回線のコリジョン、使用可能の是非が設定されている回線状況要素テーブルおよび、できるだけ早く配信する、または安く配信する等のユーザ要求が設定されているユーザ要求要素テーブルである。これにより、使用時刻が限定されている等の各回線の使用条件、使用状況に応じた回線制御が可能となる。

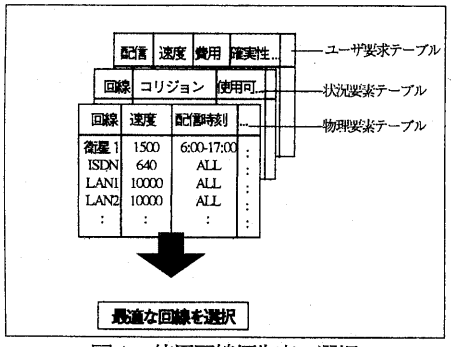

図４　使用回線優先度の選択

3.3 障害場所の明確な判別

ネットワークを用いてヘルスチェックを行う方式では、その使用回線に障害が発生した場合、サーバから見て計算機に障害があった場合と同様の振る舞いとなる場合がある。つまりサーバへの反応が無い場合、使用している回線に障害が発生している可能性と共に、接続計算機自身が停止している可能性もあるということである。前述のようにジョブの代替を考慮した場合、障害場所の明確な判別が必要となる。そのために今回は、複数の回線が接続していることを利用し、障害場所の判別のために以下の操作を行う。

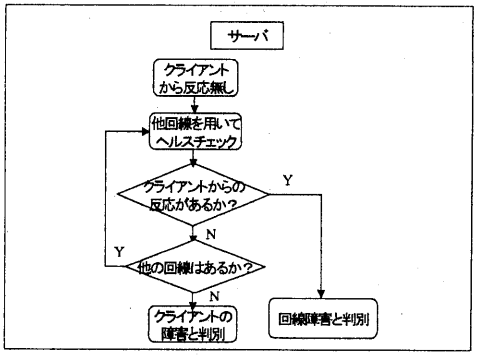

図５　障害場所の判別

・クライアントからの反応が無い場合、使用した回線以外の回線で、再度ヘルスチェックを実行
・反応が無い場合、さらに別の回線でヘルスチェックを実行
・これを全ての接続回線で確認
・クライアントの反応が無い場合は、クライアント計算機自身に障害が発生してい

るもの判断。
・他の回線で、クライアントからの反応を受信した場合は、回線の障害と判断

3.4 広域ネットワークへ対応

大規模広域分散ネットワークにおいて生じる障害は、個々の計算機レベルのものから、ネットワークドメインレベルのものまで考えられる。例えば、阪神大震災のような大地震やそれに伴う火災等により、ネットワークドメイン全体が機能不能になってしまう場合も十分考えられる。従って各ネットワークドメイン間でのヘルスチェックが必要である。

ドメイン間チェック方式は、図６に示すように、前述したドメイン内ヘルスチェックの階層化を取り入れ、１つのドメインを１つのクライアントまたはサーバとして管理する方式を採用した。これにより、ネットワークの規模が変わっても、グループ化を行うことにより容易にチェックが可能な管理方式を実現した。

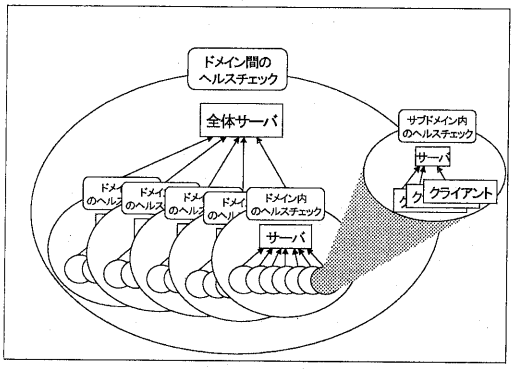

図６　階層型ヘルスチェック方式

各ドメイン内でのヘルスチェックの結果を全体サーバに報告する際、報告を行う計算機は、ドメイン内チェック時のサーバである必要は無い。ドメイン間のゲートウェイとなっているという条件に合っていれば、任意の計算機を設定できる。また、ドメイン内の各計算機同様の理由により、各ドメイン間も複数のネットワーク回線で接続されていること、１つのドメインに複数のゲートウェイ計算機が設定されていることを前提としている。これにより、例えば、あるドメインからの反応が無い場合以下の処理により障害場所の判別を行う。

・別の接続回線でヘルスチェック
・反応があれば回線の障害と判断し、切り替えた回線を使用するように設定
・反応が無ければ、他の全ての接続回線で確認
・反応が無ければ、ゲートウェイ計算機に障害があるものと判別、他のゲートウェイ計算機に対し上記と同様の動作でチェック
・反応があれば、そのゲートウェイ計算機を今後使用するよう設定する。
・全てのゲートウェイ計算機において反応が無い場合、ネットワークドメイン全体が機能していないと判別

以上のようなヘルスチェック方式を採用することにより、設計ポリシーであった
・負荷分散を考慮した方式
・抗堪性の高い方式
・　障害場所の明確な判別可能な方式
・広域ネットワークに適用した方式
を満たしたチェック方式を得ることができた。

4　障害情報通知方式

上記ヘルスチェックにより、障害場所が確定でき、それをどのように通知するかが問

題となる。今回の方式の目的は、ジョブの自動代替の実現である。これを実現するため、以下のポリシーを持って、障害情報の通知方式を設計した。

・ 確実に障害情報を伝える
・ 障害場所判別後、早急なジョブ代替処理を行うために、できるだけ早く伝える

設計における問題点を以下の3つにしぼり考察した。

・どの計算機が行うか？
・どの回線を用いるか？
・どのタイミングで行うか？

まずどの計算機が行うかについて考察する。上述のようにヘルスチェック方式では、サーバに障害情報が収集される。ヘルスチェック処理からスムーズに通知動作に移行できることから、ヘルスチェックサーバが直接通知するものとした。

次にどの回線を用いて通知を行うかについて考察する。上述したように本方式において、ヘルスチェックサーバには複数の回線が接続されているという前提で議論を行っている。確実に情報を通知するために、ヘルスチェックで用いた方式と同様に、使用優先度の高い回線から順番に使用可能かをチェックし、可能であればこれを使用する方式を採用した。ただし障害情報通知は、その緊急性が必須であるため、回線の使用優先度設定の際、回線速度のパラメータに重みをおいて、決定することとした。

最後にどのタイミングで通知処理を行うかについて考察する。障害情報通知は障害発生後、できるだけ早急に行う必要がある。通知が遅れればそれだけジョブの代替等の対策処理も遅れる事になる。よって、ヘルスチェックサーバに障害情報が届いた時点で、サーバは即座に障害情報の通知処理を行うこととした。さらにサーバでの障害情報通知処理プロセスの優先度を他のアプリケーションよりも高く設定することで、他のアプリケーションに通知処理を中断されることが無いようにした。

5. おわりに

広域大規模分散ネットワークシステムにおける障害時のジョブ代替のための、ヘルスチェック方式および障害情報通知方式について考察し、設計を行った。今後は実システムへの実装・評価を行い、実際の運用上での問題点を明確にし、より現実的な方式を提案していきたい。また、提案した方式の目的である障害時の自動ジョブ代替方式において、上記問題点を参考にし、現実的な方式を提案、実装および評価を行い、高信頼広域分散ネットワークシステムでの総合的な障害対策方式を提案していく。

参照文献

[1] George Coulouris, Jean Dollimore, Tim Kindberg; Distributed Systems Concepts and Design 2nd Edition, Addison-Wesley Publishing Company 1994

[2] 秋山, 田中,笠井; 衛星利用データ配信システムの実装と評価, DPS91-1(1998)

[3] 秋山, 田中; 衛星利用データ配信システムの評価, DPS80-29(1997)

[4] Andrew S. Tanenbaum, Modern Operating System, Prentice-Hall inc, 1992