仮想マシン環境における障害解析支援手法 - 実行トレース再生機構の検討 -

近江 雅紀　片山 吉章　國分 俊介　樋口 毅　松本 利夫
三菱電機（株）　情報技術総合研究所

1. はじめに

近年、サーバリソースの有効活用や運用コストダウンを目的に、仮想化技術によってサーバの集約を行う事例が増加している。しかし、1台の物理サーバに機能が集約されると、当該サーバで障害が発生した場合に影響が大きくなるため、より高い信頼性が求められている。

このような背景のもと、我々はシステム開発の段階で信頼性を確保するための手法として、障害発生時の動作検証を支援する、仮想障害発生機構[1]を開発した。さらに、システム稼働後に発生したアプリケーション障害に対し、早期原因究明・対策を支援する、サーバ仮想化技術を利用したアプリケーション障害再現システムの提案[2]をしている。

本提案システムは、障害再現に必要な情報を収集する機構と、収集した情報をもとに障害を再現する機構から構成されている。本稿では、本提案システムにおける障害再現機構である、実行トレース再生機構の実現方式について述べる。

2. アプリケーション障害再現における課題

サーバアプリケーションの場合、オペレーション等の入力はネットワーク経由となる。そこで、システム稼動中に収集した仮想マシンのスナップショット情報と、ネットワークパケット情報を利用することで障害再現が可能となると考えた。サーバアプリケーションの動作を再現するには、収集したネットワークパケット情報からネットワーク通信を再現する必要があるが、以下に示す課題がある。

(1) コネクション型である TCP 通信の場合、コネクション接続処理で、ランダムに決定した初期シーケンス番号を交換し、この番号を用いた ACK 番号を返すことで接続確認を行っている。そのため、収集したネットワークパケット情報をそのまま送信すると、ランダムに決定される初期シーケンス番号に対応できないためコネクション接続に失敗し、ネットワーク通信を再現できない。

(2) TCP 通信では、送信済みパケットのシーケンス番号とデータ長から受信可能な ACK 番号の範囲が決まるため、範囲外の ACK 番号を持つ受信パケットは破棄される。そのため、ターゲットからのパケット受信を待たずに送信するとネットワーク通信の再現に失敗する。

(3) 仮想マシンの再生完了を待たずにネットワークパケット情報を送信すると、必要なパケットをアプリケーションが受信できないためネットワーク通信の再現に失敗する。

3. 実行トレース再生機構

本章では、前章で示した課題を解決する、TCP/IP モデルのトランスポート層以下のネットワーク通信の再現と、仮想マシンの再生完了タイミングとネットワークパケット情報の送信開始タイミングの同期に対応した実行トレース再生機構を提案する。本機構を用いることで、事前に収集したスナップショット情報とネットワークパケット情報を利用して、スナップショット情報収集時点の状態からネットワーク通信を再現可能となり、サーバアプリケーションの動作が再現可能となる。

本機構は、ネットワークに接続された検証用サーバとパケット再送サーバから構成される。検証用サーバでは、仮想マシンモニタのスナップショット再生機能によって、スナップショット情報から障害発生サーバを再生する。パケット再送サーバでは、実行トレース再生モジュールによって、検証用サーバ上でのスナップショット情報からの障害発生サーバ再生指示、及び、ネットワークパケット情報の送信処理を行う。提案する機構のソフトウェア構成を図1に示す。

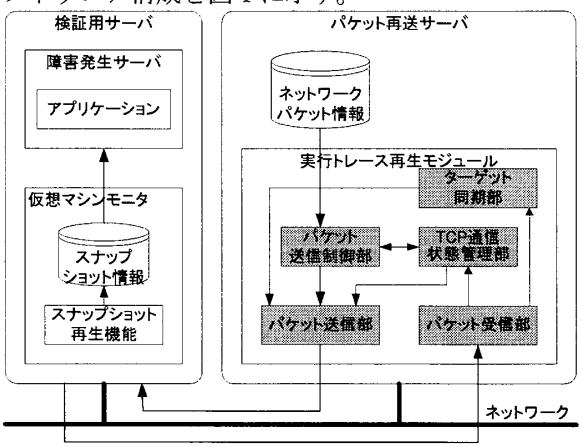

図1　実行トレース再生機構のソフトウェア構成

"A Support Technique of Trouble Analyses for Virtual Machine Environment"
Masanori Omi, Yoshiaki Katayama, Shunsuke Kokubu, Tsuyoshi Higuchi and Toshio Matsumoto
Information Technology R&D Center, Mitsubishi Electric Corporation

仮想マシンモニタ：VMWare[3]、Xen[4]等のサーバ仮想化ソフトウェア。
スナップショット情報：ある瞬間の仮想マシンのCPU/メモリ/ディスク等の情報。
スナップショット再生機能：仮想マシンモニタが持つ、スナップショット情報から仮想マシンを再生する機能。
ネットワークパケット情報：スナップショット情報収集以降の、障害発生サーバが通信していたネットワークパケット情報。
実行トレース再生モジュール：ネットワーク通信の再現機能と、仮想マシンの再生完了に同期してネットワークパケット情報の送信を開始する機能を持つモジュール。

4. 実現方式

本章では、前章で示した実行トレース再生モジュールの実現方式について述べる。

4.1 ネットワーク通信の再現機能

ネットワーク通信の再現は、コネクションレス型とコネクション型の両方の通信方式に対応することで実現する。

コネクションレス型の場合、ネットワークパケット情報に含まれるタイムスタンプの間隔に従ってパケットを送信することで実現する。

次に、コネクション型の場合、TCP 通信を模擬してネットワークパケット情報を送信することで実現する。TCP 通信を模擬したネットワーク通信の再現処理は、TCP 通信状態管理部、パケット送信制御部、パケット送信部、パケット受信部から構成される。

TCP 通信状態管理部：TCP 通信の状態を管理するため、クライアント・サーバのコネクション毎に以下に示す情報を TCP 通信管理情報として管理する。

(1) クライアントの IP と PORT 番号
(2) 障害発生サーバの IP と PORT 番号
(3) ネットワークパケット情報中の初期シーケンス番号
(4) コネクション接続時の初期シーケンス番号
(5) 受信済みパケットのシーケンス番号
(6) 受信パケットのデータ長

パケット送信制御部：コネクション接続処理とパケットの編集を行う。送信パケットが TCP の SYN パケットの場合、コネクション接続処理を行う。コネクション接続に成功後、TCP 通信状態として TCP 通信管理情報の(1)～(4)の情報を登録する。送信パケットがコネクション接続済みの場合、TCP 通信管理情報をもとに、送信パケットの TCP ヘッダ情報の ACK 番号とチェックサムの書換えを行う。
パケット送信部：ネットワークパケット情報に含まれるタイムスタンプの間隔に合わせて待った後、TCP 通信管理情報を確認する。送信パケットの ACK 番号が、障害発生サーバが受信可能な ACK 番号の範囲となるまでパケット受信を待ち、その後パケットを送信する。
パケット受信部：TCP パケットを受信するたびに TCP 通信管理情報の(5)、(6)の情報を更新する。TCP 通信管理情報と比較し、古いパケットを受信した場合は更新しない。

最後に、TCP 通信を再現するためのパケット送信フローを図 2 に示す。

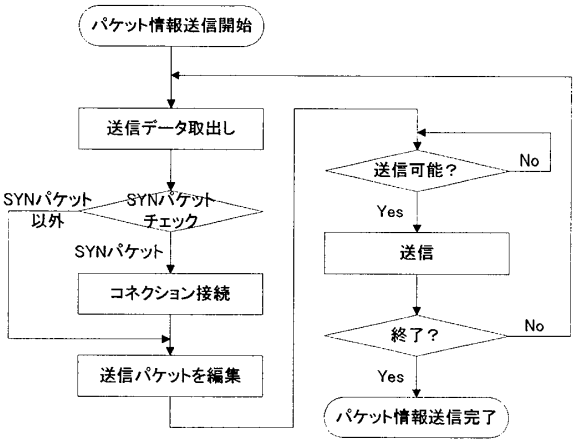

図 2　TCP パケットの送信フロー

4.2 同期機能

仮想マシンの再生完了とネットワークパケット情報の送信開始タイミングの同期は、**ターゲット同期部**にて、以下に示す方法で実現する。

(1) ネットワークパケット情報が送信パケットから始まる場合、サーバに対して ICMP パケットを送信しポーリングを行い、仮想マシンの再生完了を待つ。
(2) 受信パケットから始まる場合、サーバからのパケット受信を待つ。

5. おわりに

本稿では、実行トレース再生機構の実現方式について述べた。本機構を用いることにより、事前に収集したスナップショット情報とネットワークパケット情報を利用し、スナップショット情報収集時点の状態から障害発生までのネットワーク通信が再現可能となるため、サーバアプリケーションの障害再現を短時間で行うことが可能となる。

参考文献

[1] 片山、國分、樋口、松本、相浦：仮想マシン環境における障害模擬手法～仮想障害発生機構、電子情報通信学会 2008 年総合大会（2008）
[2] 樋口、國分、片山、近江：サーバ仮想化技術を利用したアプリケーション障害システムの提案、情報処理学会研究報告（2009-OS-110）
[3] VMware Inc.：http://www.vmware.com/
[4] Xen：http://www.xen.org