システムソフトウェアと
オペレーティング・システム 88－4
（２００１．７．２６）

「ネットワークを渡り歩けるコンピュータ」のチェックポイント機能

須崎 有康

k.suzaki@aist.go.jp
http://www.etl.go.jp/~suzaki/NTC

産業技術総合研究所
情報処理研究部門

「ネットワークを渡り歩けるコンピュータ」は仮想計算機を介して OS の実行途中の状態 (スナップショット) を撮り、別の計算機に転送／再開を可能にするシステムである。今までの実装ではスナップショットをハイバネーションの機能で撮っているので一度仮想計算機を停止する必要があった。新しい実装では OS(Linux) にチェックポイントの機能を拡張することで OS の実行を停止することなしに転送可能にした。

Checkpoint for Network Transferable Computer

Kuniyasu SUZAKI

National Institute of Advanced Industrial Science and Technology,
Information Technology Research Institute
Tsukuba Central 2, Umezono 1-1-1, Tsukuba, Ibaraki, 305-8568, Japan

"NTC: Network Transferable Computer" is a system which enable to transfer the running OS image (Snapshot) to another machine using virtual machine. The previous version has to stop the virtual machine to get the snapshot, because the snapshot is taken by hibernation. We develop checkpoint on the OS (Linux) and the new version enables to get snapshot without stopping the virtual machine.

1 はじめに

「ネットワークを渡り歩けるコンピュータ」 [1] とは、物理的な計算機を運ばなくてもオフィスで使っている計算機の実行イメージを自宅の計算機でそのまま継続できるシステムである (図 1)。これは telnet のようにネットワークを介してオフィスの計算機を使うのではない。VNC [2] のようにディスプレイのイメージのみ別の計算機で射影するものでもない。オフィスで使っていた計算機のスナップショットを撮り、その実行イメージをもとに自宅の計算機で再現するのである。実行イメージは仮想計算機を介してスナップショットとして撮られる。スナップショットを撮ったり、再現したりするメカニズムはノートパソコンで使われているハイバネーション機能を用いる。通常のハイバネーションでは計算機自体が止ってしまってディスクを操作することができないが、仮想計算機上のハイバネーションならば仮想計算機を動かしている OS は停止していないのでディスクを操作可能である。現在、「ネットワークを渡り歩けるコンピュータ」は仮想計算機ソフト の vmware [3]、Free OS の Linux、ハイバネーションソフト SWSUSP [4] の組合わせで実現されている。この環境で X window 上の QuickTime 動画再生を中断し、別の計算機上で継続することが可能になっている。

「ネットワークを渡り歩けるコンピュータ」環境では二つの目標がある。一つはソフトウエア開発者にコラボレーションのプラットホームとして「ネットワークを渡り歩けるコンピュータ」を提供し、オープンソースでは得られなかった共通の実行／デバッグ環境を与えること、もう一つはエンドユーザに計算機の実行状況が転送可能である機能を利用し、情報家電時代のトラブルシューティングに使えることをである。

ソフトウエア開発者は「ネットワークを渡り歩けるコンピュータ」によりコラボレーションのための共通プラットホームが与えられる。ソースコードのみの静的情報交換から、計算機の実行イメージのコピーを配布して同じ状態を共有することで、デバック途中のダイナミックな情報や X window 上のグラフィカルな情報など、より具体的な情報が得ることができる。これによりソフト開発がより容易に、且つ加速することができる。

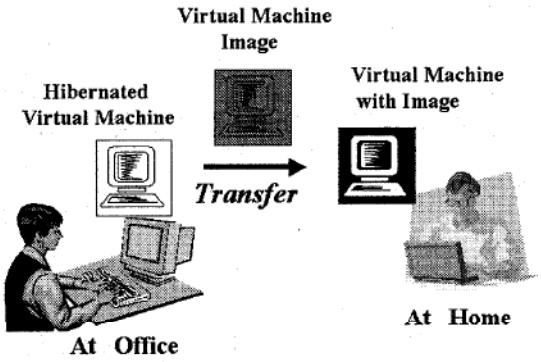


図 1: 実行イメージの転送（オフィスの環境を自宅に転送）

また、エンドユーザにはソフトウエアのトラブルの状態をそのままサービスセンターにネットワーク経由で転送可能とする。今までソフトウエアのトラブルがあるとハードウエアをサービスセンターに持ち込み、その状態をブートから再現する必要があった。これは手間と再現性に問題がある。今後、情報家電が広く普及すれば、手間と再現性の問題は深刻になると予想されるが、「ネットワークを渡り歩けるコンピュータ」はその解決策の一つになるものと期待している。

残念ながら今まで開発されていた「ネットワークを渡り歩けるコンピュータ」では、OS のスナップショットを取るために一旦仮想計算機を停止する必要があった。これはスナップショットがハイバネーションにより取られていたためである。このことはサーバなどに適用する場合、ネットワークのコネクションが切断されることを意味する。この問題を解決するために実行途中でもスナップショットが取れるチェックポイント機能を開発した。本論文ではその詳細を報告する。

2 ネットワークを渡り歩けるコンピュータ

「ネットワークを渡り歩けるコンピュータ」が行っていることは、ある計算機の内蔵ハードディスクを

取り出し、そのハードディスクを他の計算機に接続してブートすることと同一である。この際、移した先で問題なくブートするには計算機が同じアーキテクチャであり、接続デバイスや BIOS が同じである必要がある。

「ネットワークを渡り歩けるコンピュータ」は図 2 に示すように仮想計算機上を渡り歩く。仮想計算機は仮想計算機ソフトによって提供され、異なるＯＳ上でもデバイスが共通の環境が提供される。具体的に例をとると仮想計算機ソフト vmware [3] は Linux や Windows NT 上で動作可能である。仮想計算機ソフトが実行できる OS は「ホスト OS」と呼ばれ、仮想計算機上にインストールされる OS は「ゲスト OS」と呼ばれる。ゲスト OS は共通化されたデバイスをもとに仮想計算機ソフトが提供する仮想ディスクにインストールされる。ゲスト OS が停止してもホスト OS は停止しないので、ゲスト OS がインストールされたディスクイメージはゲスト OS の停止後も操作可能であり、このディスクイメージを転送すれば他のマシン上の仮想計算機でもブート可能である。転送先のマシンで仮想計算機ソフトが実行できれば、デスクトップ PC でもノート PC でもかまわない。

「ネットワークを渡り歩けるコンピュータ」はだた単にディスクイメージの転送することにとどまらず、OS が走っている環境をそのまま別の計算機で再開することができる。これはノート PC のハイバネーション機能と同じである。アプリケーションを終了せずに一旦 OS を停止した後に OS を再開し、アプリケーションの処理を続行する。ハイバネーションと仮想計算機を組み合わせることでどの計算機においてもアプリケーションの処理を継続することを可能にする。

「ネットワークを渡り歩けるコンピュータ」はネットワークあるいはリムーバブル記憶メディアでデータ転送可能な環境で、計算機の実行イメージのスナップショットを撮り、別の計算機でスナップショットを撮った状態から再開する。この「ネットワークを渡り歩けるコンピュータ」を実現するための基本機能は仮想計算機と仮想計算機上の OS のハイバネーションである。以下の章でその詳細と現在の実装状況について述べる。

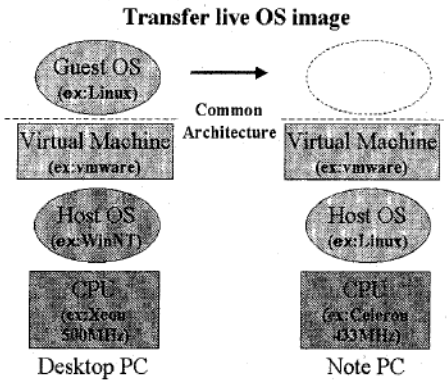


図 2: ネットワークを渡り歩けるコンピュータの実現方法

2.1 仮想計算機

仮想計算機はもともと OS 開発を行うために作り出された。仮想計算機は計算機の状態をトレースし、デバックを容易にする。最近では異る OS のアプリケーションを実行するために用いられている。この用途ではサン・マイクロシステムの wabi やフリーソフトの wine,dosemu などのエミュレーションソフトが多い。CPU パワーの向上、ディスク容量の拡大に伴って仮想計算機ソフトも一般的になりつつある。

「ネットワークを渡り歩けるコンピュータ」では異る計算機で実行イメージを再現したいのでアプリケーションのエミュレーションのみのソフトは適さない。本当の仮想計算機を提供する Connextix 社の virtual PC [5] や VMware 社の vmware [3] などが適する。virtual PC は Machintosh 上で動作する仮想計算機ソフトであり、vmware は Linux, Windows NT 上で動作する仮想計算機ソフトである。「ネットワークを渡り歩けるコンピュータ」ではソースコードがオープンな Linux 上で動作する vmware を採用した。

vmware が提供する仮想計算機は CPU が Intel Pentium 相当、512M までのメモリ、SVGA のグラフィックス、IDE ハードディスク、AMD PCnet のネットワークカード、Sound Blaster16、PhoenixBIOS 4.0R6 である。この仮想計算機ソフトが実行できる OS (Linux, WindowsNT) は「ホスト OS」と呼ばれ、

仮想計算機上にインストールされる OS は「ゲスト OS」と呼ばれる。ゲスト OS には Linux、FreeBSD、Windows95,98,NT,2000 がサポートされている。ゲスト OS は vmware が提供する仮想ディスクにインストールされる。SWAP 領域も同じ仮想ディスクに取られる。「ネットワークを渡り歩けるコンピュータ」ではこの仮想ディスクが実行イメージのスナップショットとして渡される。

2.2 ハイバネーション

ハイバネーションは計算機の実行状態の保持して電源をセーブする機能である。この機能はノート PC にインストールした OS が電源をセーブするために CPU を止めたり、ハードディスクを止めても元の状態から再開するために作り出された。ハイバネーションは OS によって状態遷移の用語が異なるので本論文では以下のように定義する。

スタンバイ状態 メモリに実行状態を残し、CPU やハードディスクを停止する。電源供給を続けてメモリの内容を保持しなければならない。

サスペンド状態 実行状態をすべて不揮発記憶（ハードディスク）に移し、電源を停止する。電源供給の必要はない。

「ネットワークを渡り歩けるコンピュータ」では実行状態を転送しなければならないので、ハイバネーションした場合にすべてのデータがハードディスクに存在するサスペンド状態になる必要がある。

通常、ハイバネーションは BIOS の省電力のためのインターフェース規格 (APM: Advanced Power Management や ACPI: Advanced Configuration and Power Interface) と OS が連係して各デバイスの電力を制御する。Linux では apm コマンドにより実現できることになっているが、BIOS に依存してサスペンド状態になれないことがある。幸いにも Linux には apm コマンドの欠点を補う SWSUSP (Software Suspend) ソフトがある。これは BIOS に依存せずに電源を切れる状態に移行可能にするソフトである。

SWSUSP は Linux カーネルパッチと終了／起動に関係するソフト (shutdown,init 等) へのパッチで構成される。正確には SWSUSP でのハイバネーションはサスペンド状態への移行ではない。シャットダウン（shutdown プロセス）とブート（init プロセス）との間で、実行プロセスの状態の SWAP 領域への退避とメモリへの復旧を行うのである。

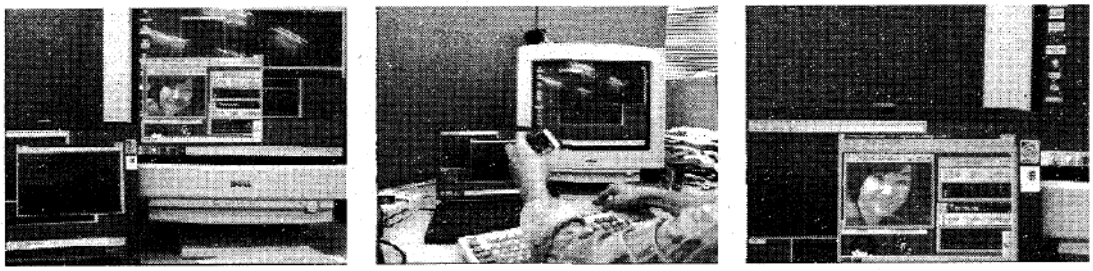
SWSUSP のパッチを当た shutdown プロセスでは、実行中のプロセスに SIGKILL を送ってプロセスを殺す前に、 bdflush プロセスを呼び出し、実行中すべてのプロセスが持つ dirty とマークされたバッファを SWAP 領域に退避する。退避が完了した段階で sync を実行して OS を halt させる。次のブート時に SWSUSP のパッチを当た init プロセスが SWAP 領域を利用可能にする際 (swapon が実行される際) に退避した実行イメージをメモリに戻し、プロセスを再開させるのである。

SWSUSP は特殊な終了／起動をハイバネーションに模している。このため計算機がマルチ OS に対応していれば Linux が SWSUSP によるハイバネーションを行った後に別の OS を立ち上げ、その OS のシャットダウン後に SWSUSP で停止した Linux の状態から再開することも可能である。

2.3 実装

現在、「ネットワークを渡り歩けるコンピュータ」は Linux をベースに運用されている。まず、Linux(ホスト OS) 用の仮想計算機ソフト vmware を利用した。この vmware が提供する仮想計算機にハイバネーションソフト SWSUSP のパッチを当てた Linux をゲスト OS としてインストールした。ゲスト OS の Linux は vmware が提供する仮想ディスクに SWAP 領域を含めてインストールされている。SWSUSP によってスナップショットを撮られた実行イメージは仮想ディスク上の SWAP 領域に格納される。この仮想ディスクがホスト OS によって他の計算機に運ばれる。運ばれた先の vmware で仮想ディスクの Linux がブートされ、SWSUSP によって撮ったスナップショットから再開される。

この環境でデスクトップ PC 上の xanim による QuickTime 動画の再生を中断させ、実行イメージをネットワークあるいは PC カードで運んでノート PC で続行することが可能である (図 3)。このデモン



デスクトップPCで動画の再生。この実行イメージを保存する。

リムーバブルメディアで転送。

ノートPCで再開。

図 3: デモンストレーション風景

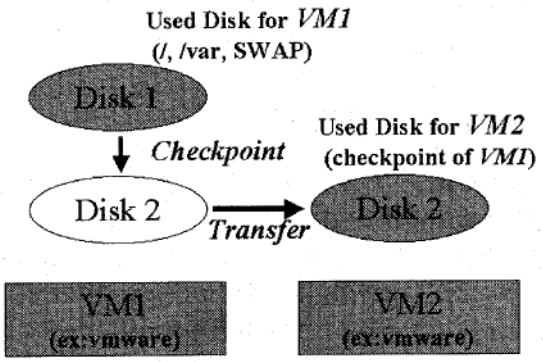


図 4: チェックポイントを使った「ネットワークを渡り歩けるコンピュータ」

ストレーションの様子とインストール手順は「ネットワークを渡り歩けるコンピュータ」のホームページ (http://www.etl.go.jp/~suzaki/NTC) で公開している。

3 チェックポイント機能

チェックポイント機能はプロセスの実行を止めず、途中の状態を保存できる機能である。チェックポイントは主にフォールトトレラントのために開発された。これに対してハイバネーションはノートPCの消費電力を抑えるためにOSの実行を中断する機能である。このため状態の停止/再開はできるが、状態を保存する機能はない。

NTCはもともとハイバネーションをベースに開発されたため、OSのスナップショットを取るためには仮想計算機を一旦停止しなくてはならない。再開によりプロセスは再実行可能であるが、ネットワークのコネクションを張っていた場合は切断されてしまう。この問題を解決するために仮想計算機を停止せずに実行途中のスナップショットが撮れるようなチェックポイント機能を開発した。

チェックポイントの開発はLinuxのハイバネーションソフトであるSWSUSP (Software SUSPend) をベースに改良を行った。基本的には図4のように二つのハードディスクを用意し、一つはオリジナルOS用 (Disk1)、もう一つはスナップショット用 (Disk2) に用いる。Disk1で動作しているOSがチェックポイントを起こすとDisk1のコピーがDisk2に撮られる。チェックポイントによりスナップショットが撮られた後は、オリジナルOSはそのまま実行を続け、別の仮想計算機がDisk2を用いてチェックポイントを起こした時点から再開可能になる。

開発したチェックポイントが行うディスク操作の詳細を図5に示す。二つのディスクは同じパーティション構成をしている。この構成は更新がないRead Onlyのパーティション、更新があるRead/Writeのパーティション、SWAP領域のパーティションに分けられる。Disk1のSWAP領域は通常のスワップとして用いられ、Disk2のSWAP領域はスナップショット用にもちいられる。チェックポイントを起こす際にはメモリとDisk1のSWAP領域にある実行中のプロセスをDisk2のSWAP領域にセーブする。その際にセーブする必要のない部分 (Dirtyでないページ) は

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.