様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 5 月 15 日現在

機関番号：１２１０２
研究種目：挑戦的萌芽研究
研究期間：2011～2012
課題番号：２３６５００１０

研究課題名（和文）　仮想計算機を前提としたオペレーティングシステムの設計

研究課題名（英文）　Designing operating systems based on virtual machines

研究代表者
板野　肯三（ITANO, Kozo）
筑波大学・システム情報系・教授
研究者番号：２０１１４０３５

研究成果の概要（和文）：仮想計算機における「アウトソーシング」と「ゲスト・ソーシング」という独自技術を元にして、オペレーティングシステムの新たな設計原理を提案した。それに基づき、Linux におけるメモリ管理のアウトソーシング、および、Linux と Solaris の間のファイルシステムのゲスト・ソーシングを実現した。

研究成果の概要（英文）：In this research, we have proposed a new operating systems design principle based on two our own techniques, called outsourcing and guest-sourcing. Based on the principle, we have implemented outsourcing of the memory management in Linux and guest-sourcing of the file system between Solaris and Linux.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	2,800,000	840,000	3,640,000

研究分野：計算機科学
科研費の分科・細目：ソフトウエア
キーワード：オペレーティングシステム、仮想計算機、メモリ管理、ファイルシステム、アウトソーシング、ゲスト・ソーシング

１．研究開始当初の背景

OS (Operating System) は、上位層の応用プログラムと下位層のハードウェアに挟まれ、両者の変化に応じるべく常に進化が求められている。現在の OS の対応が不十分であり、かつ、重要な構成要素として仮想計算機がある。VMware Workstation に代表されるホスト型仮想計算機では、ホスト OS の上に応用プログラムとして仮想計算機モニタを動作させ、それにより作られる仮想計算機でゲスト OS を実行する。ホスト OS から見た時、仮想計算機とゲスト OS は新たなタイプの応用プログラムであるが、それを実行するために有用な機能を既存のホスト OS は提供していない。また、ゲスト OS から見た時、ホスト OS は全く新たなタイプのハードウェアであるが、その有効な活用方法は知られていない。

２．研究の目的

本研究の目的は、仮想計算機を新たに支援すべき応用プログラム、および、新たに利用すべきハードウェアとしてとらえ、それに適した OS を設計することである。研究期間内に、既存のオペレーティング・システム・カーネル Linux のメモリ管理、および、Solaris のファイルシステムを再設計する。OS をスクラッチから設計することも考えられるが、その場合、応用プログラムの用意や開発環境の整備に時間を要してしまう。本研究では、イ

ンクリメンタルに既存のオペレーティングシステムを改変しながら次世代の仮想計算機を前提とした OS の設計原理を明らかにする。

３．研究の方法

本研究では、アウトソーシングという、既存の独自に考案した手法、および、ゲスト・ソーシングという、本研究で新たに提案する手法を用いて、オペレーティング・システム・カーネルを再設計する(図 1)。

アウトソーシングとは、ゲスト OS の高水準のモジュールがホスト OS の高水準のモジュールの機能を積極的に利用することである。高水準でホスト OS の機能を利用することは、従来の準仮想化(paravirtualization)に基づきデバイス・ドライバのような低水準で利用することと大きく異なる。たとえば、ネットワーク・プロトコル・スタックを対象としたアウトソーシングを考える。ゲスト OS 内ではシステム・コールの処理に際して、Socket API (Application Program Interface) と類似の手続き群が呼ばれる。この時、これらの手続き群は、自分では処理を行わずにホスト OS のネットワーク・プロトコル・スタックを呼び出す。

アウトソーシングを実現するために、Host RPC (Host Remote Procedure Call) という独自に考案した仕組みを用いる。Host RPC は、分散システムの構築で多く用いられている RPC (Remote Procedure Call) を、仮想計算機環境に特化したものである。Host RPC では、ゲスト OS の高水準のモジュールがクライアント、ホスト OS の高水準のモジュールがサーバとなる。Host RPC は、分散システムの RPC の利点である、機能と意味を手続き呼び出しの形で明確に記述できることを引き継ぐ。Host RPC は、分散システムの RPC とは異なり、共有メモリや vmmcall 命令などの CPU の仮想化支援機能を利用して効率的に実装できる。

従来のアウトソーシングに加えて、本研究で新たに「ゲスト・ソーシング」という仕組みを導入する。ゲスト・ソーシングとは、アウトソーシングとは逆に、ホスト OS がゲスト OS のモジュールを高い水準で利用するものである。また、Host PRPC とは逆に、ホスト OS 上のクライアントがゲスト OS サーバを呼び出す PRC を、Guest RPC と呼ぶ。

本研究では、アウトソーシングとゲスト・ソーシングを用いて、次世代の OS の設計原理を明らかにする。すなわち、アウトソーシング／ゲスト・ソーシングを OS のどのモジュールに対して適用すればよいか、また、どのようなインタフェースが適切かを明らかにする。

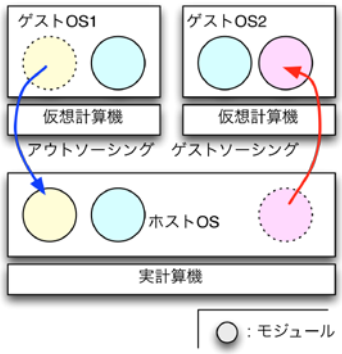

図 1 アウトソーシングとゲストソーシング

４．研究成果

本研究の成果として、まず、メモリ管理のアウトソーシングを実現したことがあげられる。アウトソーシングは、もともとは入出力を高速化するための手法として考案された。本研究ではアウトソーシングの技法をメモリ管理のモジュールという、非入出力の OS 根幹部分に適用した。これにより、ホスト型仮想計算機においては、ホスト OS とゲスト OS でメモリ管理が重複することで生じる問題のいくつかを解決した。具体的には、大域的なメモリ割り当ての最適化が難しいこと、同じページを 2 重にページアウトしようとしてしまうこと、および、ゲストのページをメインメモリにピン止めができないことである。この問題を解決するために本研究では、ゲスト OS のメモリ管理をホスト OS にアウトソースすることで、メモリ管理の重複を無くする。本研究では、ホスト OS とゲスト OS ともに Linux でメモリ管理のアウトソーシングを実現した。まず、ZONE_HOST と呼ばれるメモリ管理領域をもうけ、ゲスト・ユーザ・プロセスのメモリ割当てや mmap() によるホスト OS ファイルのマッピングを実現した。次に、ゲストのページをメインメモリにピン止めすることを、実現した。実験により、ゲスト OS のメモリを効果的に増減させることができることを示した。

本研究では、ゲスト・ソーシングのために必要なホスト・ゲスト間通信、すなわち、Guest RPC を実現した。Guest RPC では、サーバは、ゲスト OS のモジュール、クライアントはホスト OS のモジュールとなる。Guest RPC は、ソケットアウトソーシングの技術を Unix Domain Socket に対して適応することで実現した。これにより、ゲスト OS のプロセスとホスト OS のプロセスは、任意のデータをホスト OS の Unix Domain Socket を使

って交換できるようになった。この仕組みを使って、RPC を実現することで、Guest RPC を実現した。

本研究では、Guest RPC を用いて、ホストでは利用できないファイルシステムを、ゲスト OS の機能を利用して利用可能にした。具体的には、Linux ホスト OS において利用できない ZFS を、Solaris ゲスト OS を使って利用可能にした。ZFS では、カーネル内の VFS 層の機能の他に、カーネル外で様々なコマンドが必要になる。本研究では、カーネル外のコマンドについても、Solaris のものを利用できるようにした。ホスト OS で ZFS 関連のコマンドを実行すると、その引数を RPC の要求としてゲスト OS へ送り、そこで実行する。そしてそのコマンドの実行結果を、ホスト OS へ RPC の結果として返す。

研究期間全体を通じて、アウトソーシング、および、ゲスト・ソーシングという仮想計算機間の相互作用により、ある OS のカーネル、および、カーネル外のコマンドを他の OS から利用できることがわかった。また、メモリ管理についても、仮想計算機の境界を超えて１つのホスト OS で統一的な管理を行なうことが可能であることがわかった。これらのことから、今後の OS 設計の仕組みとして、仮想計算機を用いて様々な OS をつなぎ合わせることが有効であることがわかった。

今後は、ファイルシステム以外のオペレーティング・システム・カーネルのモジュールについてゲスト・ソーシングが有効か確認したいと考えている。このようなモジュール化を行なうと、モジュール間通信のオーバヘッドが大きいことが分かった。今後は、このオーバヘッドを特化やハードウェア支援により減らしたいと考えている。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 9 件）

[1] 大橋宏樹, 新城 靖, 齊藤 剛: "仮想計算機におけるソケットアウトソーシングを用い IPv4/IPv6 変換の実現と評価", 情報処理学会論文誌:コンピューティングシステム(ACS), Vol.5, No.3 (ACS 38), pp.30-41 (2012). 査読有. http://id.nii.ac.jp/1001/00082472/

[2] 小柳 光生, 吉田 一, 海野 裕也, 新城 靖: "LOUDS トライのオンライン構築のためのブルームフィルタ構築法", 情報処理学会論文誌:コンピューティングシステム(ACS), Vol.5, No.2 (ACS 37), pp.1-9 (2012). 査読有. http://id.nii.ac.jp/1001/00081509/

[3] Shingo Takada, Akira Sato, Yasushi Shinjo, Hisashi Nakai, Akiyoshi Sugiki and Kozo Itano: "A P2P Approach to Scalable Network Booting", The Third International Conference on Networking and Computing, pp.201-207 (2012). 査読有. http://dx.doi.org/10.1109/ICNC.2012.38

[4] 三戸 健一, 齊藤 剛, 新城 靖, 佐藤 聡, 中井 央, 板野 肯三: "ソケットアウトソーシングを適用した仮想計算機でのライブマイグレーションの実現", 情報処理学会コンピュータシステム・シンポジウム(ComSys2012),ポスターセッション, 2 pages (2012). 査読無. http://www.ipsj.or.jp/sig/os/index.php?plugin=attach&refer=ComSys2012%2Fposter&openfile=12-602-1.pdf

[5] 小柳 光生, 吉田 一, 海野 裕也, 新城 靖: "簡潔データ構造のオンライン構築とブルームフィルタによる検索性能の向上", 情報処理学会論文誌:データベース(TOD), Vol.4, No.4 (TOD 52), pp.1-10 (2011). 査読有. http://id.nii.ac.jp/1001/00079661

[6] Jinpeng Wei, Feng Zhu, Yasushi Shinjo: "Static Analysis Based Invariant Detection for Commodity Operating Systems", The 7th International Conference on Collaborative Computing: Networking,Applications and Worksharing (CollaborateCom 2011), pp.287-296 (2011). 招待論文.査読無. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6144814

[7] 大橋宏樹, 新城 靖, 齊藤 剛: "仮想計算機におけるソケットアウトソーシングを用いた IPv4/IPv6 変換の実現", 情報処理学会コンピュータシステム・シンポジウム(ComSys2011), Vol.2011, pp.95-104 (2011). 査読有. http://id.nii.ac.jp/1001/00079098/

[8] 小柳 光生, 吉田 一, 海野 裕也, 新城 靖: "LOUDS トライのオンライン構築のためのブルームフィルタ構築法", 情報処理学会コンピュータシステム・シンポジウム(ComSys2011), Vol.2011, pp.33-41 (2011). 査読有. http://id.nii.ac.jp/1001/00079091/

[9] 戸祭 要, 新城 靖, 齊藤 剛, 豊岡 拓, 板野 肯三: "ホスト型仮想計算機におけ

るメモリ管理のアウトソーシングの提案", 情報処理学会研究会報告,システムソフトウェアと オペレーティングシステム研究会(OS),2011-OS-119(9), 7 pages (2011). 査読無.
http://id.nii.ac.jp/1001/00078651/

〔その他〕
ホームページ等
http://www.softlab.cs.tsukuba.ac.jp/

６．研究組織
(1)研究代表者
板野　肯三（ITANO, Kozo）
筑波大学・システム情報系・教授
研究者番号：20114035

(2)研究分担者
新城　靖（SHINJO, Yasushi）
筑波大学・システム情報系・准教授
研究者番号：00253948