様　式　Ｆ－１９－２

科学研究費助成事業　　研究成果報告書

令和　　５ 年　　５ 月 ３１ 日現在

機関番号：１２６０１
研究種目：国際共同研究加速基金（国際共同研究強化(A)）
研究期間：2019～2022
課題番号：１８ＫＫ０３１０

研究課題名（和文）準パススルー型仮想マシンモニタのための安全かつ信頼できる実行環境に関する研究

研究課題名（英文）Research on Secure and Reliable Execution Environments for para-pass-through virtual machine monitors

研究代表者

品川　高廣（Shinagawa, Takahiro）

東京大学・情報基盤センター・准教授

研究者番号：４０３６１７４５

交付決定額（研究期間全体）：（直接経費）　　11,300,000円
渡航期間：　18ヶ月

研究成果の概要（和文）：本研究では準パススルー型アーキテクチャを踏まえて、まず新しいCPUアーキテクチャの機能を活用して、仮想マシンとコンテナの利点を併せ持った実行環境の研究を実施した。また、既存のx64 CPUアーキテクチャを拡張して、ネステッドページングにおけるオーバーヘッドを削減する変換パススルーの研究を実施した。これらの研究により、クラウド環境に向けた安全かつ信頼できる実行環境の実現に向けた新たなアーキテクチャの方向性を示した。

研究成果の学術的意義や社会的意義
本研究の成果は、学術的には安全で信頼性の高い実行環境の実現に関するシステムソフトウェアとコンピュータアーキテクチャの最新の知識を広めることに寄与する。また、社会的にはクラウドコンピューティングの安全性と信頼性を向上させることにより、ビジネスやサービスにおける利用が促進され、効率化や柔軟性の向上が期待できる。さらに、クラウドサービスの利用者に安心感を与え、クラウド環境とデジタル社会の進展に貢献する。

研究成果の概要（英文）：Based on the para-pass-through architecture, this study first conducted a study of an execution environment that combines the advantages of virtual machines and containers by leveraging the capabilities of the new CPU architecture. We also conducted research on translation pass-through, which extends the existing x64 CPU architecture to reduce the overhead in nested paging. These studies provided a new architectural direction for a secure and reliable execution environment for cloud environments.

研究分野：オペレーティングシステム

キーワード：仮想化技術　セキュリティ

様　式　Ｆ－１９－２

１．研究開始当初の背景

応募者は、準パススルー型という応募者自らが考案した仮想マシンモニタの新しいアーキテクチャの研究を長年積み重ねてきた。準パススルー型仮想マシンモニタでは、ゲスト OS からハードウェアへのアクセスを仮想マシンモニタで原則としてパススルーとすることで小型軽量化を可能にしつつ、必要十分なアクセス制御を施すことで強力なセキュリティ機能を併せ持つことができるという革新的なアーキテクチャであり、その研究成果は高く評価されてきた。

実際､最初に準パススルー型アーキテクチャを発表した論文は Google Scholar によると 2023 年 5 月現在で 319 件の論文から参照されており、この分野では国際的にも高いインパクトを持っている。また、本アーキテクチャの応用論文がシステムに関する一流国際会議である ASPLOS 2015 に採択されており、その応用性の高さも国際的に高く評価されている。

一方で､様々な応用のための機能拡張を容易に実現できるようにするためには､汎用的に使える安全かつ信頼できる実行環境の構築が不可欠である｡本アーキテクチャの軽量･小型という特徴を損なわずに､そのような実行環境を構築するためには､様々な側面から検討が必要であった。

２．研究の目的

本国際共同研究の研究目的は、これまでに実施した準パススルー型仮想マシンモニタに関する研究成果も踏まえて、準パススルー型アーキテクチャを発展させ、有用性･汎用性を格段に向上させることである。従来の準パススルー型アーキテクチャでは、OS に依存せずに仮想マシンモニタ内でマイグレーションやサンドボックス、フォルトインジェクションなど様々な拡張セキュリティ機能を実現することが出来たが、その機能を動作させるための汎用的な実行環境の構成が十分に研究されておらず、目的ごとにバラバラの手法で拡張をおこなってきた。本研究では、拡張機能を実行できる汎用的で安全かつ信頼できる実行環境の研究をおこなう。

３．研究の方法

英国の Imperial College London の Large-Scale Data & Systems (LSDS) グループを率いる Peter Pietzuch 教授らと国際共同研究をおこなうことで、仮想マシンモニタと連携した安全かつ信頼できる実行環境に関する議論をおこない、準パススルー型アーキテクチャの軽量･小型といった特徴を生かしつつ、汎用的な機能拡張を提供できる仕組みの研究をおこなった。Peter Pietzuch 教授らの LSDS グループは、Intel 社製 CPU のセキュリティ拡張機能である Software Guard Extentions (SGX) を用いて信頼できる実行環境 (Trusted Execution Environment: TEE) の実現に関する研究を多数おこなっており、信頼できる実行環境の構造や、そこで実行するコードを生成するためのプログラム分割などに関する知見を持っている。従って、応募者の持つ準パススルー型アーキテクチャなどの安全な実行環境の実現に関する知見を組み合わせることで、上記の安全で信頼できる実行環境の実現を加速することを目指した。

４．研究成果

本研究は大きく分けて二つのアプローチで実施した。一つ目は新しい CPU アーキテクチャの機能を活用するアプローチ、二つ目は既存の CPU アーキテクチャを拡張するアプローチである。一つ目のアプローチでは、さらに二段階に分けて研究を実施した。以下で、それぞれについて述べる。

４．１．ケーパビリティ仮想マシン (Capability Virtual Machine: cVM)

一つ目の新しい CPU アーキテクチャの機能を活用するアプローチでは､CHERI というケーパビリティをサポートするハードウェアを活用した研究をおこなった。まず 2021 年度には、CHERI を応用した新しい保護ドメインの概念である cVM を提案した。本研究の概要は以下のとおりである。

クラウドスタックは､アプリケーションコンポーネントを隔離する一方､同じ物理ホスト上に配置されたコンポーネント間で効率的なデータ共有を可能にする必要がある。従来､MMU が分離を強制しつつページ単位での共有を可能にしてきた。しかし、MMU のアプローチは、カーネル空間に大きな TCB を持つクラウドスタックが必要であり、ページ粒度ではデータ共有のための非効率な OS インターフェイスが必要となる。今後登場するメモリ機能をハードウェアでサポートする CPU は、より細かい粒度で分離と共有を実現する新たな機会を提供する。

cVM は単一の仮想アドレス空間を安全に共有し､それぞれが自身のメモリにアクセスする機能のみを持つ。cVM は、信頼できる小さなモニターが支援する 2 つのケイパビリティベースのプリミティブを通じて、効率的にデータを交換することができる：(i) cVM 間で共有するバッファへの非同期読み書きインターフェース、(ii) cVM 間の制御を転送するコールインターフェース。これら 2 つのプリミティブを用いることで、効率的なクロス KVM 通信のための、より

表現力の高いメカニズムを構築する。CHERI RISC-V の機能を用いたプロトタイプの実装では、cVM がデータ共有を改善しながら、低いオーバーヘッドでサービス（Redis と Python）を分離することが示された。

cVM の研究により、バイト粒度での保護や軽量な保護ドメイン切り替え、保護ドメインのネストなどを容易に実現できるようにして、仮想化のオーバーヘッドを低く抑えつつバイト粒度での柔軟な保護と共有が可能になる汎用的で安全かつ信頼できる実行環境を実現することができた。本研究の成果の論文は、国際会議 OSDI 2022（採択率 19.4%）に採択された[1]。

４．２．ケーパビリティによるオブジェクト再利用（Object Reuse with Capability: ORC）

2022 年度には、cVM の機能を応用してメモリ共有効率を高める研究をおこなった。本研究の概要は以下のとおりである。

クラウド環境は多くのテナントを抱えており、テナントが実行するアプリケーションバイナリやライブラリの間には、通常、かなりの重複が存在する。このように，メモリ重複排除は，共有バイナリに一度だけメモリを割り当てることで，メモリ密度を高めることができる．しかし、既存の重複排除アプローチは、バイナリオブジェクトの重複排除を共有 OS に依存し、許容できないほど弱い分離を提供するか、ハイパーバイザーベースの重複排除をメモリページレベルで利用し、共有されるオブジェクトのセマンティクスに盲目であるかのいずれかである。

ORC は、テナント間でバイナリオブジェクトのきめ細かな共有をサポートし、テナントは小さな信頼できるコンピューティングベース（TCB）を通じて強く隔離される。ORC は、ハードウェアがサポートするメモリ機能を利用してテナントを分離するため、共有オブジェクトを複数のテナントから安全にアクセスすることができる。ORC は、ケイパビリティによって単一のアドレス空間内でバイナリオブジェクトを共有するため、共有オブジェクトのロード時に、スレッドローカルストレージを使用してテナントごとの状態を作成する新しい再配置タイプを使用する。ORC は、信頼できないゲストが TCB の外でオブジェクトをロードすることをサポートし、ロードされたデータの安全性を検証するだけである。我々の実験によると、ハイパーバイザーベースの重複排除と比較して、ORC はより低いパフォーマンスオーバーヘッドでより高いメモリ密度を達成することができる。

ORC の研究により、cVM 間でオブジェクトレベルでのメモリ共有ができるようにして、クラウド環境におけるメモリ消費量を削減して効率の良い共有ができるようになった。本研究の成果の論文は、会議 OSDI 2023（採択率 19.6%）に採択された[2]。

４．３．変換パススルー（Translation Pass-through: TPT）

二つ目の既存の CPU アーキテクチャを拡張するアプローチとしては、変換パススルーという手法の研究をおこなった。本研究の概要は以下のとおりである。

仮想マシン（VM）は、クラウドにおける統合、分離、プロビジョニングに使用されるが、大きな作業セットを持つアプリケーションは、VM におけるメモリアドレス変換のオーバーヘッドに影響される。(i)入れ子式ページングには、ページテーブルの深さに応じて増加する最悪のレイテンシがあり、(ii)準仮想化およびシャドーページングには、ゲストページテーブルを更新する際にハイパーバイザーの介入コストがかかる。

我々は、ネイティブに近い性能を達成する新しいメモリ仮想化メカニズムである TPT（Translation Pass-Through）について説明する。TPT は、1 次元ページテーブルを使用して、ゲスト仮想アドレスからホスト物理アドレスへの仮想メモリ変換を VM が制御することを可能にする。同時に、コモディティ CPU の物理メモリタギングをサポートする新しいハードウェアを利用することで、ホストによる VM 間の分離を実現する。

KVM/QEMU ハイパーバイザーを変更し、Linux ゲストを啓発することで、TPT のプロトタイプを作成した。AMD CPU のメモリタグ機構をエミュレートして評価した。控えめな性能評価では、TPT は実世界のデータセンターアプリケーションのネイティブな性能を達成し、ネストページングとシャドーページングに対してそれぞれ最大 2.4 倍と 1.4 倍のスピードアップを達成した。

TPT の研究により、従来の仮想化方式においても、ページングのおけるパススルーの度合いを向上させることで、ネステッドページングを部分的に不要にすることが可能になった。本研究の成果の論文は、国際会議 USENIX ATC 2023（採択率 18.4%）に採択された[3]。

これらの研究によって、仮想化環境において安全かつ信頼できる実行環境を実現するための新しいアーキテクチャの方向性を示すことができた。

[1] Vasily A. Sartakov, Lluís Vilanova, David Eyers, Takahiro Shinagawa, Peter Pietzuch. CAP-VMs: Capability-Based Isolation and Sharing in the Cloud. In Proceedings of 16th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2022), Jul 2022.
Acceptance Ratio: 19.4%.
[2] Vasily A. Sartakov, Lluís Vilanova, Munir Geden, David Eyers, Takahiro Shinagawa, Peter Pietzuch. ORC: Increasing Cloud Memory Density via Object Reuse with Capabilities. In Proceedings of the 17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2023), Jul 2023.
Acceptance Ratio: 19.6%, Accepted for publication.
[3] Shai Bergman, Mark Silberstein, Takahiro Shinagawa, Peter Pietzuch, Lluís Vilanova. Translation Pass-Through for Near-Native Paging Performance in VMs. In Proceedings of the 2023 USENIX Annual Technical Conference (USENIX ATC 2023), Jul 2023.
Acceptance Ratio: 18.4%, Accepted for publication.

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3件／うち国際共著　3件／うちオープンアクセス　3件）

１．著者名 Vasily A. Sartakov, Lluis Vilanova, David Eyers, Takahiro Shinagawa, Peter Pietzuch	４．巻 -
２．論文標題 CAP-VMs: Capability-Based Isolation and Sharing in Clouds	５．発行年 2022年
３．雑誌名 In Proceedings of 16th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2022)	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Vasily A. Sartakov, Lluis Vilanova, Munir Geden, David Eyers, Takahiro Shinagawa, Peter Pietzuch.	４．巻 -
２．論文標題 ORC: Increasing Cloud Memory Density via Object Reuse with Capabilities.	５．発行年 2023年
３．雑誌名 In Proceedings of the 17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2023)	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Shai Bergman, Mark Silberstein, Takahiro Shinagawa, Peter Pietzuch, Lluis Vilanova.	４．巻 -
２．論文標題 Translation Pass-Through for Near-Native Paging Performance in VMs.	５．発行年 2023年
３．雑誌名 In Proceedings of the 2023 USENIX Annual Technical Conference (USENIX ATC 2023)	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

Compartments and Cloud-Native Applications https://lsds.doc.ic.ac.uk/projects/cloudcap

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
主たる渡航先の主たる海外共同研究者	ピーツー　ピーター (Pietzuch Peter)	インペリアルカレッジロンドン・コンピューティング専攻・教授	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
英国	Imperial College London			
ニュージーランド	University of Otago			
イスラエル	Technion			