様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月　 ７ 日現在

機関番号：３４３１５
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ１１３０７

研究課題名（和文）テイント解析技術を用いたデータ漏洩防止機能を有するソフトウェア実行基盤

研究課題名（英文）Software Execution Platform for Data Leakage Prevention by Taint Analysis Technique

研究代表者
毛利　公一（Mouri, Koichi）

立命館大学・情報理工学部・教授

研究者番号：９０３１３２９６

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：本課題は，人が介在する必要のない安心・安全なソフトウェア実行基盤として，次の3つの技術の研究開発を行った．（1）プログラム内におけるテイント解析とテイント解析機能をOSから操作可能とするインタフェースを有するCPUエミュレータの開発．（2）ポリシーの管理機能，データとポリシーの対応付けの管理機能，データをネットワーク・ハードディスク等外部へ出力する時に，ポリシーに基づいて出力の可否を判定する出力制御機能，出力が許可された場合，リモートホスト・ハードディスク等の外部記憶装置へポリシーを引き継ぐ機能を有するOS．（3）ソフトウェアコンテナに基づく実アプリケーション実行基盤構成技術．

研究成果の学術的意義や社会的意義
データ漏洩事故の4割は人がコンピュータを介してデータを漏洩させるもので，それを防止する技術の開発が急務である．データ漏洩のポイントは，（1）データを利用した日常業務が行われその際はデータが平文でアクセスできる．（2）プログラムはアクセス可能なデータを複製・送信することが容易である．（3）データには流通させてもよい範囲が定められるが，それを人が都度判断しており誤判断が発生しやすい．本課題は，（1）のような場合でも，OSがプログラムによるデータの複製・送信を検知し，データの流通許可範囲外に漏れることがないかを判定することで，（2）や（3）によるデータ漏洩事故を防ぐ技術の開発を目指すものである．

研究成果の概要（英文）：In this project, we researched and developed the following three technologies for a safe and secure software platform that does not require human intervention. (1) Development of a taint analysis method for inside programs and a CPU emulator with an interface for operating taint analysis mechanisms from operating systems. (2) An operating system with a policy management function, a function for managing the mapping between data and policy, and an output control function that determines whether or not data can be output based on policy when outputting data to external devices such as network hard disks, and a function for transferring the policy to external storage devices such as remote hosts and hard disks when output is permitted. (3) Application execution platform technology based on software containers.

研究分野：ソフトウェア

キーワード：オペレーティングシステム　テイント解析　データ漏洩防止

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

2016 年情報セキュリティインシデントに関する調査報告書（日本ネットワークセキュリティ協会）によると，データ漏洩事故原因のトップは「管理ミス」（34.0%）で印刷された機密データを誤廃棄するようなケースであった．第 2 位はメールに機密ファイルを誤添付して送信するような「誤操作★」（15.6%），第 3 位は外部からの「不正アクセス★」（14.5%），第 4 位は USB メモリ等の「紛失・置き忘れ」（13.0%），第 5 位は「不正な情報持出し★」（6.8%），他「設定ミス★」（4.7%），「内部犯行★」（0.9%），「目的外使用★」（0.4%）などがある．特に★を付した 6 つのケース（計 42.9%）は，人が故意または不意にコンピュータを介してデータを漏洩させるものである．これらには次のような特徴がある．

- 日常的に業務で利用しているデータが漏洩の対象となっている．そのため，業務上暗号キーを知る者がデータを持ち出す場合がある．また，データはプログラム内で処理される段階までには平文でアクセスできる状況となる．
- 一般的なプログラムでは，データが一度読み込まれるとそのデータの特性に応じたファイルやネットワークへの出力（以下，出力）を制御する手法がなく，簡単に複製・送信が実現してしまう．
- 組織のルール上は複製や外部への送信の禁止などデータの流通許可範囲（以下，ポリシー）が定められていることが多いが，その確認を都度個々人に頼っているために，判断や操作のミス，故意の漏洩に繋がってしまう．

以上から，暗号化等の有無にかかわらず（1）データに対してポリシーを事前に記述可能とし，（2）読み込まれたデータのプログラム内での流れを追跡し，（3）出力時にはポリシーと出力先に基づいてその可否を判定することで，人が介在する必要のない実用的な安心・安全なソフトウェア実行基盤の設計・構築技法を明らかにする必要がある．

２．研究の目的

本研究課題は，人が介在する必要のない安心・安全なソフトウェア実行基盤として，具体的には下記のような技術の開発を目的とする．

- 次の機能を有するプロセッサ（エミュレータ）の開発
  - プログラム内でのデータの流れを追跡する機能（テイント解析機能）
  - テイント解析機能を OS から操作可能とするインタフェース
- 次の機能を有する OS の開発
  - ポリシーの管理機能，データとポリシーの対応付けの管理機能
  - データをネットワーク・ハードディスク等外部へ出力する時に，ポリシーに基づいて出力の可否を判定する機能（出力制御）
  - 出力が許可された場合，リモートホスト・ハードディスク等の外部記憶装置へポリシーを引き継ぐ機能
- ソフトウェアコンテナに基づく実アプリケーション実行基盤構成技術

３．研究の方法

研究目的を達成すべく，具体的には次のような機能の設計・構築技法を明らかにする．

- データを任意の長さのバイト列として扱うことができる，細粒度な「出力制御」技術．
- データ単位で重要性や機密度を「ポリシー」として定義可能とし柔軟な出力制御技術．
- どのようなアプリケーションにも適用可能な方式．
- 実現方式については実用性も十分検討する．

目的達成のためキーポイントとなる技術として次の(a)～(f)を明らかにする．

(a) データの区別法

データ毎の保護方針を明確にするとともにデータを区別可能とするために，保護方針毎に「タグ」と呼ぶ番号を割り当て，バイト単位でタグを付与できる仕組みを明らかにする．

(b) データの追跡法
データは，CPU，メモリ，ハードディスク，ネットワークデバイス，リモートホストにおいて，ロード・ストア（読み書き・送受信）による移動や演算等の操作対象となる．このような場合に，データとともにタグを伝播させてデータ追跡をするためのテイント解析技術を明らかにする．

(c) 保護方針の記述法
本課題では従来にない新たなデータ保護機能を実現するため，その保護方針の記述法も提案する．例えば，「データ出力不可」「元のファイルへのみ書き戻し可」「同一 PC 内のファイルへ出力可」「USB メモリへのコピー不可」「ネットワークへの送信不可」などを意図する内容を保護方針として記述できるよう検討する．

(d) 保護方針の管理法
保護方針を安全に設定・更新・削除および保存するための管理方式について明らかにする．また，保護方針とタグの対応を管理する方式についても明らかにする．

(e) 保護方針に基づいた出力制御法
出力処理（書き込み，送信等）の際には，データの出力が保護方針によって許可されているか否かを判定しなければならない．その実現方式を明らかにする．

(f) ソフトウェアコンテナ方式による構成法
本課題で提案する方式はテイント解析技術を用いるため，オーバヘッドが大きくなることが想定され，実環境への適用が困難となる可能性がある．これを解決すべく，漏洩防止機能が必要なアプリケーションのみへ提案方式を選択的に適用し，それ以外のアプリケーションはオーバヘッドのない通常環境で動作させ，それらの間をシームレスに接続できるようなソフトウェアコンテナ方式を明らかにする．

４．研究成果

本研究課題ではデータ追跡を実現すべく，Argos と呼ばれる，CPU エミュレータ QEMU をベースとしたテイント解析システムを用いた．Argos は，ネットワークから受信したデータにタグ付けをする機能（タグは有無の 2 種類のみを表現可能），およびデータがコピーされた際にタグを伝播させる機能，タグのついたデータが命令として実行される場合に例外割込みを発生させる機能（※）を有している．
以上を踏まえ，まず，様々な種類の保護方針をタグとして取り扱うために，(1) Argos を拡張して 1 バイトごとに 255 種類のタグを付与可能とするシャドウメモリを実現した．また，ゲスト OS からシャドウメモリを参照・変更可能とするために，(2) シャドウメモリを PCI デバイスとして読み書きできるように拡張した．さらに，ハードディスクへもタグを伝播可能とすべく，(3) シャドウディスクを設け，データ書き込み時はシャドウメモリのタグをシャドウディスクへも出力する機能，データ読み込み時はシャドウディスクのタグもシャドウメモリへ読み込む機能を実現した．一方で，例外発生機能（上記※）は不要であるため，無効化した．
次に，ゲスト OS（本研究課題では Linux を使用）を拡張し，(4) データ漏洩を検出するための保護方針を記述する方式と管理する方式を定め，それをタグと対応付けるための機能を実現した．また，アプリケーションプログラムがファイル入出力・ネットワーク送受信等を行う際には，(5) そのためのシステムコールをフックし，データにタグがついているか否かをチェックした上で，その方針に違反していないかどうかを確認する機能を実現した．
以上で基本的な機能の実現が完了したので，実際に近い利用環境を想定し，samba（ファイル共有サーバ），postfix（メールサーバ），sylpheed（メールクライアント）らをインストールした上で，ファイルサーバ上にあるタグのついたファイルを samba 経由で読み出し，それをメールに添付して送信（これを操作ミスであるとみなす）しても，データ漏洩を防止できることを示した．

５．主な発表論文等

〔雑誌論文〕　計4件（うち査読付論文　4件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 福田 泰平, 明田 修平, 瀧本 栄二, 齋藤 彰一, 毛利 公一	４．巻 59-9
２．論文標題 JDWPによる動的解析を利用したAndroidアプリケーションの端末識別情報利用実態調査	５．発行年 2018年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 1678-1688
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 オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Yuki Kajiwara, Junjun Zheng, and Koichi Mouri	４．巻 E104-D
２．論文標題 Performance Comparison of Training Datasets for System Call-Based Malware Detection with Thread Information	５．発行年 2021年
３．雑誌名 IEICE Transactions on Information and Systems	６．最初と最後の頁 2173-2183
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 オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 松本 隆志, 瀧本 栄二, 齋藤 彰一, 毛利 公一	４．巻 60-12
２．論文標題 TCPによるネットワークワイドなテイント追跡を用いた情報漏洩防止システム	５．発行年 2019年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 2269-2278
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 オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 福田 泰平, 鄭 俊俊, 瀧本 栄二, 齋藤 彰一, 毛利 公一	４．巻 60-12
２．論文標題 Androidにおける端末識別情報送信検出のための動的解析システム	５．発行年 2019年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 2259-2268
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 オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計7件（うち招待講演　0件／うち国際学会　1件）

1．発表者名 梶原 友希, 鄭 俊俊. 毛利 公一
2．発表標題 対象スレッドの違いによるマルウェア検知精度の比較
3．学会等名 情報処理学会コンピュータセキュリティ研究会
4．発表年 2020年

1．発表者名 原田 隆成 , 鄭 俊俊 , 毛利 公一
2．発表標題 Linuxゲスト向けCuckoo Sandboxへのファイル保存機能の実現
3．学会等名 情報処理学会コンピュータセキュリティ研究会
4．発表年 2020年

1．発表者名 梶原 友希, 鄭 俊俊, 毛利 公一
2．発表標題 動的解析システムのネットワーク接続の有無によるマルウェア検知精度の比較
3．学会等名 情報処理学会コンピュータセキュリティ研究会
4．発表年 2021年

1．発表者名 Yuya Yamashita, Junjun Zheng, Shoichi Saito, Eiji Takimoto, Koichi Mouri
2．発表標題 Implementation of Virtual Machine Monitor-Based Stack Trace Mechanism on Windows 10 x64
3．学会等名 International MultiConference of Engineers and Computer Scientists 2019 (IMECS 2019)（国際学会）
4．発表年 2019年

１．発表者名
山下 雄也，鄭 俊俊，齋藤 彰一，瀧本 栄二，毛利 公一

２．発表標題
AlkanetにおけるMeltdown対策済Windowsのシステムコールトレース手法

３．学会等名
コンピュータセキュリティシンポジウム2018(CSS2018)論文集

４．発表年
2018年

１．発表者名
森本 康太，鄭 俊俊，齋藤 彰一，瀧本 栄二，毛利 公一

２．発表標題
動的解析においてログが取得できないマルウェアの実態調査

３．学会等名
コンピュータセキュリティシンポジウム2018(CSS2018)論文集

４．発表年
2018年

１．発表者名
村松 拓実，金城　聖，毛利 公一

２．発表標題
組込み機器におけるLinux完全性検証機能IMAの性能評価

３．学会等名
情報処理学会第85回全国大会

４．発表年
2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
立命館大学情報理工学部システムソフトウェア研究室
https://www.asl.cs.ritsumei.ac.jp/
立命館大学情報理工学部システムソフトウェア研究室
https://www.asl.cs.ritsumei.ac.jp/

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

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