# 個人情報を含むファイルと通信を検出するための双子の環境の提案

張　世申 †　新城　靖 †　三村　賢次郎 †

## 1. はじめに

PC で動作するアプリケーションは、Web ブラウザのように明示的に通信を行うものだけでなく、オフィスツールのように、ユーザの意図しない通信を行うものがある。Web ブラウザであっても、キャプティブポータルの検出や危険なサイトのリストの取得のために暗黙的に通信を行うことがある。このような意図しない通信により個人情報が漏れる可能性がある。

本研究では、コンテナという OS 層の仮想実行環境を利用し、個人情報が含まれているファイルと通信を検出することを提案する。そして、ファイルや通信から不要な個人情報を削除するツールを実装する。

## 2. 双子の環境による個人情報の検出

本研究では、次の 2 つを対象として個人情報を検出する。

- ファイルに保存された個人情報
- ネットワーク通信に含まれる個人情報

たとえば、Web ブラウザは、個人情報として訪問履歴や、フォームの内容、Cookie などを保存する。ユーザトラッキングでよく利用される Cookie にはサーバが配るユニークな ID やセッション ID が含まれている。しかし、現在の Web ブラウザやオフィスツールは、非常に複雑であり、これらの情報がどのファイルにどのような形式で保存されているかを調べることは容易ではない。そこで、本研究では、双子の環境を利用して、個人情報が含まれているファイルや通信を検出することを提案する。

### 2.1 双子の環境

双子の環境とは、プログラムファイルやデータファイル等の内容がほとんど同じであるような 2 つの仮想実行環境である (図 1)。双子の環境は、人間の双子の研究を参考にして考案した。人間の双子の研究では、双子の医学的、遺伝子的、心理学的性格を調査する。双子が異なる環境で育てられた際に、どのような違いが生まれるかを調査する場合が多い。本研究で提案する双子の環境では、類似の 2 つの環境で同一のプログラムをそれぞれ実行し、同一の入力を与える。そして、2 つのプログラムの動作上の相違点を検出する。

### 2.2 双子のブラウザ

双子のブラウザとは、双子の環境で動作する組になったブラウザである。本研究では、双子のブラウザを用いてサーバによるユーザトラッキングを検出したいと考えている。ユーザトラッキングの手法としては、Cookie を使う方法や URL にタグを埋め込む方法がある。それ以外に、Flash Cookie や HTML5 の IndexedDB などのストレージを使う方法もある。本研究では双子の環境を用いて全てのファイルの差分を調査する。その差分にユーザトラッキングのための情報が含まれる可能性が高い。その差分を削除、または修正することでユーザトラッキングを阻止することができると思われる。

### 2.3 コンテナによる双子の環境の実装

双子の環境を実現するためには、様々な手法が考えられる。本研究では、環境内のファイルを調査するが、一般的な仮想マシンではホストはゲスト OS のファイルシステムを直接的アクセスできないという問題がある。そこで本研究では、コンテナという OS 層の仮想マシンを使う。

本研究はコンテナを実装する仕組みとして Docker[1] を使う。Docker では、Overlay File System というファイルシステムが利用可能である。このファイルシステムではゲスト OS のファイルをホスト OS から観測でき、またファイルの差分を取得することが容易である。本研究では、さらに 2 つのゲスト OS から Overlay File System で取得したファイルに対して差分を取る。

ファイルがテキストデータのみを含むものであれば、diff コマンド等で簡単に差分を調査することができる。しかし、ファイルの中身にはテキストデータではないものもある。そこで本研究では、表 1 に示すようなツールを用いて差分を取る。

† 筑波大学
University of Tsukuba

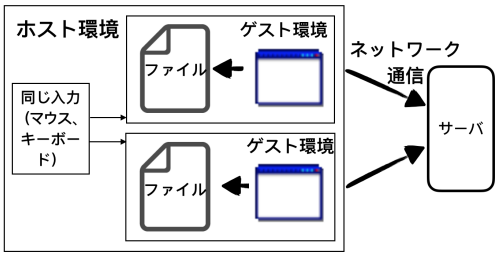


**図 1** 双子の環境

**表 1** ファイルの保存によく使われるフォーマットと差分を取るツール

| フォーマット | ツール |
|---|---|
| text | diff |
| JSON | json-diff |
| XML | Canonical XML, diff |
| SQLite | sqldiff |
| Berkeley DB | db_dump, diff |

本研究では、コンテナから発信される通信の内容を検査する。HTTPS のように情報が暗号化されていることもある。そこで本研究では Man-In-The-Middle Proxy[2] を使って HTTP と HTTPS の情報をキャプチャする。

メールリーダやテキストエディタのようなプログラムにはユーザ入力がある。本研究ではホストでのキーボードやマウスなどの入力を 2 つのコンテナに送信する。また、双子のブラウザの実装には Selenium[3] というブラウザの自動テストツールを用いることを検討している。

## 3. ランダム性の排除

ファイルや通信の差分は、個人情報の違いだけでなく、その他のコンテナの動作の違いによっても生じる。そのような個人情報とは関係がない差分が大量に存在した場合、個人情報の検出が困難になる。本研究では、そのような動作の違いを排除したい。

個人情報とは関係ない動作の違いは、様々なランダム性に起因する。アプリケーションの乱数生成器は/dev/random を読み込む。またマルチスレッドプログラムのスケジューリングもランダム性を引き起こす。本研究では、コンテナエンジンや OS カーネルの一部を変えることで、これらのランダム性を排除したいと考えている。

## 4. 関 連 研 究

Qubes-OS[4] は高いセキュリティを実現するための OS である。Qubes-OS は仮想計算機 (Xen) を用いて、アプリケーションを隔離されたドメインで実行する。Qubes-OS には、複数のドメインのファイルを比較する機能はない。

Blink-Docker[5] はコンテナ中でブラウザを実行することで、Canvas Fingerprint を利用したユーザトラッキングからユーザを保護する。Canvas Fingerprint はハードウェアや OS のわずかな違いによってクライアントを特定する。Blink-Docker はコンテナ技術を利用し、毎回 Fingerprint を変えることができる。本研究では、通信内容を検査し、そのようなユーザトラッキングを検出したいと考えている。

## 5. ま と め

本研究では人間の双子の研究に参考した双子の環境を提案した。本研究では Docker コンテナを利用し、軽量で類似の仮想環境を作り、ファイルと通信内容に含まれる個人情報を検出する。検出の妨げとなるランダム性を排除するために、コンテナエンジンや OS カーネルを修正することを考えている。

現在までに、Web ブラウザ Firefox を手動で 2 つのコンテナ内で実行し、それらが生成するファイルの差分を調査している。また、Man-In-The-Midle Proxy を用いて、通信内容を比較することができている。今後の課題は、提案方式を実装することである。そして、ファイルや通信内容から不用な個人情報を自動的に削除するツールを実装したいと考えている。

## 参 考 文 献

1) Docker - Build, Ship, and Run Any App, Anywhere: https://www.docker.com/, accessed: 2017-11-22.
2) Man-In-The-Middle proxy - an interactive man-in-the-middle proxy for HTTP and HTTPS https://mitmproxy.org/, accessed: 2017-11-22.
3) Selenium - Web Browser Automation: http://www.seleniumhq.org/, accessed: 2017-10-18.
4) Qubes OS - A reasonably secure operating system: https://www.qubes-os.org/, accessed: 2017-11-22.
5) P Laperdrix, W Rudametkin, B Baudry, "Blink: A moving-target approach to fingerprint di- versification", 37th IEEE Symposium on Security and Privacy, Poster session, 2016 [Online] http://www.ieee-security.org/TC/SP2016/poster-abstracts/59-poster_abstract.pdf