

個人情報及び個人識別子を含むファイルと
通信を
検出するための双子の環境

Zhang Shishen

1992 09 27

Introduction

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、ユーザトラッキング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住所・氏名等の個人情報、および cookie のように個人情報と結び付けられた個人識別子が送信されることがある。

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、個人情報及び個人識別子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情報を削除するツールを実装する。

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、これらの識別子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。本研究では、双子の環境を実装して、個人情報および個人識別子が含まれているファイルや通信を検出する。なお、以下では、個人情報と記載した時にも、個人情報と個人識別子の両方を含むものとする。

目 次

1	双子の環境と双子のブラウザによる個人情報の検出	3
2	コンテナによる双子の環境の実装	4
3	ファイルの差分検出	5
3.1	前処理とテキスト化	5
3.2	タイムスタンプの扱い	5
3.3	ランダム性の排除	5
4	ネットワークメッセージの差分	6
5	Conclusions	7
A	A Long Proof	8

第1章 双子の環境と双子のブラウザによる個人情報の検出

第2章 コンテナによる双子の環境 の実装

第3章 ファイルの差分検出

- 3.1 前処理とテキスト化
- 3.2 タイムスタンプの扱い
- 3.3 ランダム性の排除

第4章 ネットワークメッセージの 差分

第5章 Conclusions

...

付 録 A A Long Proof

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)。人間における双子の研究では、異なる双子が異なる環境で育てられた際にそれぞれの医学的、遺伝子的、心理学的性格を調査ことで、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。