Vol.2014-CSEC-65 No.12
Vol.2014-IOT-25 No.12
2014/5/22

デジタルフォレンジクスの為の Web 閲覧履歴可視化方式の提案

松本晋一[†1]　鬼塚雄也[†2]　川本淳平[†3]　櫻井幸一[†3]

ユーザの多くが常時携行し，活用するモバイル端末は必然的にプライバシ情報を集約しており，デジタルフォレンジクスにおいては当該端末の調査は非常に有効となる．スマートフォンに代表されるモバイル端末は HTML5 に対応した Web ブラウザをアプリケーション実行環境とする HTML5 プラットフォーム化が進んでおり，HTML5 のクライアント側記憶機能はフォレンジクスにおける調査対象として今後重要が増すものと考えられる．本研究では，当該記憶領域の内容を，オフラインで獲得し，構造化，グラフィカルに可視化する方法を提案する．

A Proposal on Method to Visualizing Web Browsing History for Digital Forensics

SHINICHI MATSUMOTO[†1]　YUUYA ONITSUKA[†2]
JUNPEI KAWAMOTO[†3]　KOUICHI SAKURAI[†3]

Mobile devices accumulates much private information of the user. This is due to the carried by user and utilized for communication. Therefore, digital forensics that targets mobile devices will be deemed to be more effective. Mobile terminals, especially smartphones is transforming a HTML5 platform. This is what the application execution environment that supports HTML5 Web browser. One of the significant feature of HTML5 is WebStorage. It is intended to enable the storage feature that is easy to handle and large capacity to Web browser. This feature will be deemed to be important in the digital forensics. In this research, we propose the method to acquire this storage area offline and structuralize and graphically visualize it.

1. はじめに

パーソナルコンピュータやインターネットの普及により生活の利便性は飛躍的に向上している反面，これを利用した情報の不正取得，情報の不正流失，端末に対する不正な遠隔操作等のサイバー犯罪もまた増加している．またメールや, Web ページ上での掲示板や SNS でのやりとりの上で行われる不正も存在する. こうした犯罪を立証する上で，デジタルデータを証拠として確保し裁判で用いることは非常に重要となる．しかし，対象がデータであるがゆえ偽造や改ざん，削除が容易にされうることから，実際の証拠として弱くなってしまう．

この問題を解決するための調査法として，デジタルフォレンジクスの需要が高まってきている[1]．これは通信ネットワーク上やパソコンなどの端末上から証拠となるデータを発見し，改ざんされることなく証拠能力を保ったまま裁判に提出するための調査法である．デジタルフォレンジクスを実施する際，調査員は

- エビデンスの取り出し
- エビデンス間の関連付け
- エビデンスが改ざんされていないことの証明

の三点に留意する必要があり，これらによりデータを証拠性のあるものとして扱うことができるようになる．

また近年，OS や機器に依存せずインターネットに接続させようという流れが高まってきており，このようなウェブアプリケーションのプラットフォームとして，HTML の最新版であり，現在策定中の HTML5 が考案された．

HTML5 対応のブラウザでは WebStorage の利用が可能で，これに調査を適用することで，より多くの情報を収集することが可能となる．しかしこの調査には，正しくデータを取り出すための手法の設定，実際のデータの収集，収集したデータ同士の結びつけによる証拠の発見という過程が存在し，これらを目視で行うことには非常に長い時間がかかり非現実的であるという問題が存在する．

ユーザのブラウジングにより WebStorage に保存される情報は本来ブラウザが参照することを前提としているため，人間にとってはそのままのデータは見づらいという難点がある．また，ここに保存されているデータはブラウジング

†1 公益財団法人 九州先端科学技術研究所
Institute of Systems, Information Technologies and Nanotechnologies (ISIT)
†2 九州大学
Kyushu University.
†3 九州大学大学院システム情報科学研究院情報学部
Department of Informatics, Faculty of Information Science and Electrical Engineering, Kyushu University

などを補助するために保存されたもので，フォレンジクス調査に必要となりうるデータのみがあるわけではない．それゆえに WebStorage に保存されている大量のデータから必要なデータのみを取得する必要がある．

本研究では，調査効率を上げ，調査にかかる時間を短縮することで，デジタルフォレンジクスにかかる時間全体の短縮を目指す．そのために，調査に有益な情報取得の効率を上げ，取得した情報同士の関連付けの半自動化による簡易化，高速化について提案する．

2. デジタルフォレンジクス

フォレンジクスという単語には「法科学」「鑑識」などの意味がある．フォレンジクスの対象としては，筆跡鑑定や DNA 解析，指紋採取などの鑑識調査が存在する．これに対しデジタルフォレンジクスはフォレンジクスから派生した言葉で，デジタルデータを情報通信技術の観点から調査し証拠を確定させ裁判所に提出していく手法である．これはデータの収集や解析，保管を行い，その法的な証拠性を明らかにする調査を表し，フォレンジクスが法科学という意味を持つように，犯罪に対し適切に法的な処理を行うための調査である[2][3].

ここで対象となるのは，パソコン，サーバ，ネットワーク機器，携帯電話，情報家電などのデジタルデータを扱う機器である．調査として PC のハードディスクから証拠に当たるデータを探し出す，ログファイルから不正なアクセスの記録を割り出す，破損したまたは消去されたデータの復旧，データの捏造や改ざんの防止，またその検証が行われる．デジタルフォレンジクス調査は，犯罪を立証するだけでなく，容疑者の無実を証明するためにも用いることができる．

2.1 ネットワークフォレンジクスとコンピュータフォレンジクス

デジタルフォレンジクスは調査対象や調査方法から「ネットワークフォレンジクス」と「コンピュータフォレンジクス」の大きく 2 つに分けられる．

(1) ネットワークフォレンジクス

ネットワークフォレンジクスとはコンピュータネットワーク上を流れるデータを対象とした調査である．この調査では，まずネットワーク上を流れるすべてのパケットを取得し保管する．そして，すべての通信の断片から元のデータを復元することで，どの端末からどのネットワークを介して何が行われたのかを解析を行う．これによりマルウェアの挙動の解析や，情報流失の経路の解明などを行うことが可能である．ネットワークフォレンジクスで対象となり得るデータの例として，Web ページアクセスのログ，メールのやりとりのログ，ファイルの送受信ログなどが想定される．また，ネットワークフォレンジクスはインシデントが発生する前の予防としても活用可能である．内部ネットワーク上の通信データを取得し，不審な挙動を行っている PC を特定し，管理者に警告することができる．これによりその端末の操作記録を追跡できるため，ネットワークフォレンジクスの存在を組織内に周知させることで，内部からの情報流失を抑制することができる．

(2) コンピュータフォレンジクス

コンピュータフォレンジクスはコンピュータ内に保存されているデータを対象とした調査である．コンピュータフォレンジクスでは，まず不正を行ったとされるコンピュータの確保を行い，ハードディスクを取り出す．これを，実際に法廷での証拠用のハードディスクと，調査員が使用するための解析用のハードディスクに, 完全な複製を行う．証拠用のハードディスクは証拠性を証明する書類とともに厳重に保管されることで，証拠に一切の変更が加えられていないことを示し，法的な証拠性を保全する．次に解析用ディスクに対して実際の調査を行い, ファイル内容の分析，ファイルへのアクセス時刻の調査，改ざんされたデータの発見と修復，また削除されているデータの復元と解析などが行われる．この調査で対象となり得るデータの例としては，Web ページ閲覧履歴データ，パスワード履歴データ，フォーム履歴データ，メール送受信データ，削除されたファイルなどが考えられる．

3. HTML5

3.1 HTML5 概要

HTML は 1990 年代前半に導入されてから何度も改訂されており，仕様の変更や導入，タグの追加や削除が行われている. 現在 HTML の最新版として HTML5 が存在する[4]. HTML5 は Web アプリケーションを OS に依存させずに動かすためのプラットフォームを目指して考案されたものである．また HTML5 はすでに広く使われているコンテンツを取り扱う方法において，下位互換性が保たれるように規定されている．

3.2 WebStorage

HTML5 で扱うことのできる機能に WebStorage がある[5]. これは Web 上でのデータのやり取りの際にデータをクライアント側に保存する機能である．データをクライアント側で保存する方法としては Cookie が一般的であったが，WebStorage は従来使われている Cookie よりも簡単に, また大量のデータを長期間クライアント側に保存することができる．以下に WebStorage の特徴を上げる．

Vol.2014-CSEC-65 No.12
Vol.2014-IOT-25 No.12
2014/5/22

(1) **保存容量**

Cookie の最大容量 4KB に対し，WebStorage は約 5MB の容量を保存可能である．これによりクライアント側でサイズの大きなデータを取り扱うことが可能となる．

(2) **有効期限**

Cookie には有効期限が設定され，当該有効期限が期限切れとなると消去されるのに対し，WebStorage ではデータの永続的記録が可能となる．

(3) **送信**

データを毎回サーバに送受信せずとも格納，参照が可能となる．これによりネットワーク負荷が軽くなる．また自動的に送信されないためユーザにとってセキュリティ的に安全となる．ブラウザによって一旦 Cookie を保存すると，その Cookie の提供元の Web ページはユーザの許可無く保存した Cookie から情報を得ることができる．

(4) **保存形式**

データ保存の方法として，任意の保存したいデータに対して，対応する一意の標識を設定し，これらをペアで保存するキーバリューストアという形式で保存を行っている．

WebStorage には sessionStorage と localStorage という 2 種類のストレージが用意されており，目的によってそれぞれ使い分けられている．

3.3 localStorage

localStorage ではオリジン(origin)単位でデータを保存，管理している[6]．オリジンは``プロトコル”，``ドメイン”，``ポート番号”から構成され，オリジン毎に格納情報が隔離される．このため現在のホストで保存したデータを異なるホストで読み込むということは不可能となるが，localStorage は異なるタブやウィンドウ同士でも同じオリジン内であればデータを共有可能である．

また localStorage に保存されたデータはブラウザを閉じても保持され，明示的に削除されるまで保持される．

例えば Web ページ閲覧者が選択した背景色や文字色のテーマなどを localStorage に保存し維持する，ということが考えられる．

3.4 sessionStorage

sessionStorage もまたオリジン単位でデータの管理を行っている．しかし localStorage と違い，複数のタブやウィンドウ間でのデータの共有はされないようになっている．

sessionStorage はブラウザが起動している間のみ有効となるストレージである．sessionStorage に保存されたデータはブラウザを閉じたタイミングで破棄される．

4. 関連研究：プライベートブラウジング，ポータブルブラウザで残される情報の調査

ブラウザが備えるプライベートブラウジング機能は，セッションの終了時にブラウジング中にたまったデータを破棄する，またはセッション中にデータを蓄積させない機能である．これは，

- ユーザのインターネットでの行動履歴をウェブサイトのサーバから追跡されることなくブラウジングを行うこと．
- 使用している端末に，ユーザのインターネットでの閲覧履歴を残さずにブラウジングを行うこと．

の二つの目的で用いられる．後者の目的は例えば複数人で共有のパソコンを使用する場合などに特に必要となり，この研究では後者の目的に焦点を当てている．

[7]では，四種の Web ブラウザ Internet Explorer，Google Chrome，Firefox，Safari での通常のブラウジング後，プライベートブラウジング後それぞれに対しディスクを完全分析することでクライアント側に残されているデータを収集している．また三種のポータブル Web ブラウザ Opera Portable，Firefox Portable，Google Chrome Portable でのブラウジング後にも同様の調査を行っている．

通常の閲覧，プライベートブラウジング，ポータブルブラウジングに対し実験が行われた．手順として，プライベートブラウジング時に一連の決められた行動を行い，プライベートブラウジング終了時にメモリをダンプし，レジストリファイル，システムファイルを取得し，RAM のイメージを作成する．これと同様の操作をそれぞれのブラウザで行っていく．また，ポータブルブラウザを USB フラッシュメモリから起動し，プライベートブラウジング時と同様の操作を行う．その後，フォレンジクスツールでの解析をそれぞれのデータを保存したハードディスクに対し行うことで，それぞれのブラウジングについての実験を行う．

以上のような実験の結果，各ブラウザに関し，表 1, 2 に示すエビデンスの獲得に成功している．

表 1 プライベートブラウジングモードからのエビデンス抽出結果

	IE 8.0 In-private Browsing	Chrome 23.0.1271.95 Incognito	Firefox17.0.1 Private Browsing	Safari 5.1.7 Private Browsing
Browsing Indicators	○	○	○	○
Browsing History	○	○	○	○
Username/ E-mail accounts	○			
Images	○	○	○	○
Videos				

表 2 ポータブルブラウザからの
エビデンス抽出結果

	Chrome Portable	Opera Portable	Firefox Portable
Browsing Indicators	○	○	○
Browsing History	○	○	○
Username/E-mail accts	○		○
Images	○	○	○
Videos		○	

5. 提案手法

WebStorageに保存されている情報は機械可読なものであり人の目による閲覧性は考慮されていない．そのため，WebStorage に保存されているデータを調査する場合，そのままのデータでは見づらく，情報の精査や結びつけを目視で行うことは困難であるという難点がある．

本研究では，ブラウザの sessionStorage 内のデータの構造化手法を提案する．これは WebStorage に残されている大量のデータから，ユーザのページ既読情報を取得し，閲覧したページ情報を整理するための手法である[8]．今回，ユーザの閲覧行動の履歴の追跡を容易とするための，閲覧履歴のツリー構造に基づくグラフィカルな表示により，情報の精査をより容易とする方式について提案を行う．

5.1 事前調査

Windows 版の，主要なブラウザにおける WebStorage の保存場所をまとめたものを，表 3 に示す．

表 3 WebStorage の保存パス

ブラウザ種別	対応バージョン	保存パス
Internet Explorer	8 以降	未確認
Firefox	3.6 以降	C:¥Users¥<username>¥AppData¥Roaming¥Mozilla¥Firefox¥Profiles¥<profileFolder>
Google Chrome	8 以降	C:¥Users¥<username>¥AppData¥Local¥Google¥Chrome¥UserData¥Default
Opera	11 以降	C:¥Users¥<username>¥AppData¥Roaming¥Opera Software¥Opera Stable
Safari	5 以降	C:¥Users¥<username>¥AppData¥Local¥Apple Computer¥Safari

今回，Windows 版の Firefox 26.0 を対象として，当該ブラウザを用いてブラウジングを行った後，sessionStorage に保存されたデータを削除される前に確認することで，Firefox の sessionStorage に格納される情報の構造を調査した．結果として，図 1 のようなキーの構造が見られることがわかった．

図に示した key-value 構造の特徴として以下が挙げられる．[windows][tabs][entries]下に[url]，[title]，[ID]，[referrer]等が存在し，これらの属性で単一の Web ページを表し，単一の[entries]の下にこれら属性の組が複数存在しうる．

[url]は当該 Web ページの URL を収容しており，[title]は当該 Web ページのタイトル，[referrer]は当該 Web ページのリンク元となった URL を値として収容する．

また [children]は，この Web ページに埋め込まれている

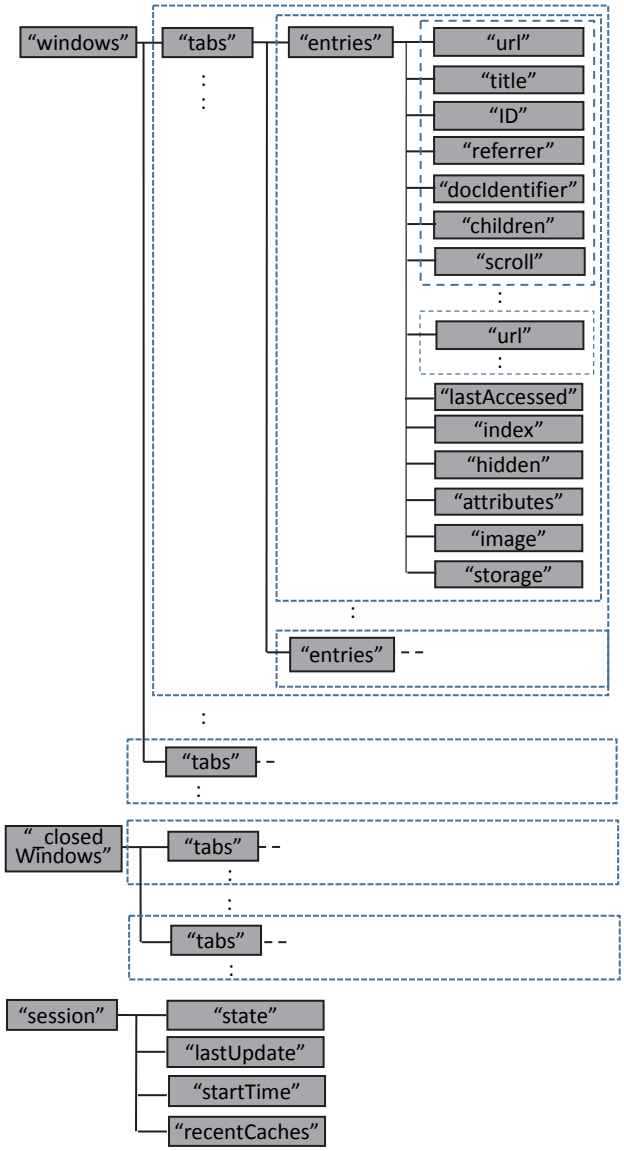

図 1 sessionStorage の key 構造

広告の Web ページ情報などを持っている。この時，1 つの[entries]下に存在する複数の Web ページ情報の集合で 1 つのタブに対応している．即ち，1 つのキー[entries]に対するバリューとして，1 つのタブで閲覧した Web ページの集合に対応する情報が格納されている．

[entries]キーが [tabs]下に複数存在する場合は，当該セッション内に複数のタブが含まれることを表す．

sessionStorage はブラウジング終了時に保存されたデータを消去するが，Firefox ではどのタイミングで消去されているのかを実際にブラウジングを行い sessionStorage の変化を見ていくことで調査した．ブラウザ起動時，ブラウジング途中, ブラウジング終了時, ブラウザプロセス終了時，次回ブラウザ起動時に調査した結果，sessionStorage に保存されたデータはブラウザプロセス終了時ではなく次回ブラウザ起動時に消去されていることがわかった．これはブラウザの仕様により，前回のタブの復元を任意で行えるようにするためであると考えられる．

5.2 提案方式

フォレンジクス調査により sessionStorage から蓄積されているデータを収集し犯罪捜査を行うことを仮定する．sessionStorage の内容はユーザがブラウジングで自動的に蓄積したデータであるが，閲覧したページ上の広告情報やページの表示サイズの情報，ウィンドウサイズの情報などのデータも大量に存在し，調査に必要なデータのみが存在するわけではない．ここで，事件捜査に必要となるデータのみを抽出する必要がある．目視のみでは調査に時間のかかってしまうこの過程において，必要な情報のプログラムによる半自動的な構造化を提案する．

提案手法で今回は Firefox の sessionStorage に対し調査を行う．sessionStorage は複数ウィンドウが開かれている場合そのすべてのウィンドウが閉じられるまで情報を貯め続けることが事前調査で分かった．このことから sessionStorage からは最後に閉じたウィンドウの情報のみではなく，同じセッション中にすでに閉じられていたタブやウィンドウでの情報も取得することができる. また sessionStorage はバックアップファイルとして常に 1 つ前のセッションでの情報を保存している．このファイルと，Firefox の sessionStorage の内容を消去するタイミングの仕様により，閉じられたブラウザでのセッション情報，またバックアップファイルとしてさらにその一つ前でのセッションの情報を取得することができる．

5.3 実装内容

プログラミング言語 Python で提案手法の実装を行っていく．処理手順としては，

1) ディレクトリを指定し sessionStorage の内容を実験環境に読み込む
2) 収集した Web ページ情報の群を，木構造の根となり得る Web ページとそれ以外の Web ページに分類
3) 根となるページ情報に含まれる[url]と一致する url を[referrer]に持つページ情報を探索
4) 見つかれば次はそのページ情報の[url]と一致する[referrer]を持つページ情報を探索
5) 見つからなければその根での探索は終了し，そこまでのページ情報の[url], [title], [ID], [docshellID], [referrer]の値を表示する
6) 根からどれ程離れたノードかを``*”の数で表現する
7) 次の根から探索を開始する
8) すべての根での探索を終えたら終了

を行う．また，sessionStorage には直接閲覧した Web ページ以外に, Web ページに埋め込まれている広告などの Web ページ情報は解析対象外とする．

6. 考察

実際に動作の評価実験を行った．ブラウザ Firefox26.0 を起動し，スタートページからブラウジングを行った．その際にどの Web ページからどの Web ページへ移動したか，どの順番で Web ページが開かれたかを記録した．一定時間ブラウジングを行い, その後 sessionStorage に対し調査を行った．

評価の結果, [url]と[referrer]が一致しない現象が見られた．これはキーワード検索を用いて，検索結果からページを読み込んだ際に，表示したページ情報の[referrer]が変化しているためである．これは特定の検索サイトにのみ見られる現象である．

7. まとめ

本稿では sessionStorage に保存される Web ページ情報に対する調査の為の，対象データの構造化の提案と調査を行った. これにより sessionStorage に残された情報からフォレンジクス調査に必要となるデータの確認と取得に成功した．

今後の課題としては，例外時の処理法について，検討を進める必要がある．また今回の調査，実装では未解析のキーが表すバリューに対する調査を行う必要がある．また，Internet Explorer や Chrome などの他のブラウザに対しても適用可能な方式を検討，実装する必要がある．

また，今回は sessionStorage を対象としたが，localStorage についても検証を行う必要がある．用途の異なるストレージ領域間の情報を突合することで，新たなエビデンスを得られる可能性がある．

Vol.2014-CSEC-65 No.12
Vol.2014-IOT-25 No.12
2014/5/22

謝辞

第一著者は, JSPS 科研費 26330169 の助成を受けています.

参考文献

1) 辻井 重男監修，特定非営利活動法人 デジタルフォレンジック研究会編集:デジタル・フォレンジック辞典, 株式会社日科技連出版社(2006)
2) Cory Altheide, Harlan Carvey: DIGITAL FORENSICS WITH OPEN SOURCE TOOLS, Syngress, (2011)
3) John Sammons,:THE BASICS OF DIGITAL FORENSICS The Primer for Getting Started in Digital Forensics", Syngress,(2012)
4) Robin Berjon, et al.: HTML5 A vocabulary and associated APIs for HTML and XHTML W3C Candidate Recommendation 6 August 2013, http://www.w3.org/TR/2013/CR-html5-20130806/
5) Ian Hickson. : Web Storage W3C Recommendation 30 July 2013, http://www.w3.org/TR/2013/ REC-webstorage-20130730/
6) A.Barch,: The Web Origin Concept, RFC6454.
7) Donny Jacob Ohana et.al,: Do Private and Portable Web Browsers Leave Incriminating Evidence? A Forensic Analysis of Residual Artifacts from Private and Portable Web Browsing Sessions”, Security and Privacy Workshop (SPW), 2013 IEEE 23-24 May 2013, p135-142.
8) 鬼塚雄也, 松本晋一, 川本淳平, 櫻井幸一: フォレンジクス支援の為の Web 閲覧履歴構造化手法, 火の国情報シンポジウム 2014 (2014).