情報処理学会論文誌 Vol.57 No.12 2682–2695 (Dec. 2016)

プライバシに配慮したユーザ参加型 Web観測フレームワーク

松中 隆志[1,†1,a)] 山田 明[1,†2] 窪田 歩[1,†2] 笠間 貴弘[2]

受付日 2016年3月11日, 採録日 2016年9月6日

概要：ユーザをマルウェアに感染させる主要な方法の 1 つとして，Web を媒体として Web ブラウザ経由で自動的にマルウェアをダウンロード・実行させる Drive-by Download 攻撃（以下，DBD 攻撃）が問題となっている．著者らは，DBD 攻撃など多様な Web 上の脅威を網羅的に観測するために，ユーザ参加型の攻撃対策フレームワーク（FCDBD: Framework for Countering Drive-By Download）を提案，実装した．また，著者らは FCDBD フレームワークにユーザが安心して参加できるようにフレームワークの参加者に関するプライバシ上の懸念を解消するための取り組みを実施した．本論文では，著者らが実施したプライバシに関する取り組みについて報告する．また，実際に FCDBD フレームワークを用いて 1,000 人規模の参加者を募り，データ収集および検知方式の評価のための実験を実施した結果についても報告する．

キーワード：Drive-by Download 攻撃，Web セキュリティ

A User-participating Framework for Monitoring the Web with Privacy Guaranteed

Takashi Matsunaka[1,†1,a)] Akira Yamada[1,†2] Ayumu Kubota[1,†2] Takahiro Kasama[2]

Received: March 11, 2016, Accepted: September 6, 2016

Abstract: A Drive-by Download (DBD) attack is one of the major threat on the Web. The attack forces a user to download a malware via his/her web browser. We proposed and implemented the user-participating Framework for Countering Drive-By Download (FCDBD) to monitor threats on the Web. We dealt with users' concern against their privacy in the framework in order that users can participate the framework with ease. Finally, we report the result of our field trial with over 1,000 participants to evaluate our detection methods on the framework and to collect web access data.

Keywords: drive-by download attack, web security

1. はじめに

Drive-by Download 攻撃（以下，DBD 攻撃）は，今日の主要なマルウェア拡散方法の 1 つである．この攻撃は，Web を利用してマルウェアを拡散する攻撃であり，ユーザは，攻撃が仕掛けられた Web ページにアクセスするだけでマルウェアに感染させられてしまう．**図 1** に DBD 攻撃の典型的な流れを示す．攻撃者は正規サイトを改ざんし，当該サイトを訪問したユーザを攻撃用に準備したサイト群へ誘導するためのスクリプトを挿入する．改ざんされたサイトにアクセスしたユーザは，まず入口サイト（Landing site）へ遷移される．入口サイトでは，ユーザの環境（OS，ブラウザの種類/バージョン，プラグインの種類/バージョン，Cookie の情報，IP アドレス，リファラ情報など）を

[1] 株式会社 KDDI 研究所
KDDI R&D Laboratories, Inc., Fujimino, Saitama 356–0003, Japan
[2] 国立研究開発法人情報通信研究機構
National Institute of Information and Communications Technology, Koganei, Tokyo 184–8795, Japan
[†1] 現在，KDDI 株式会社
Presently with KDDI Corporation
[†2] 現在，株式会社 KDDI 総合研究所
Presently with KDDI Research, Inc.
[a)] ta-matsunaka@kddi.com

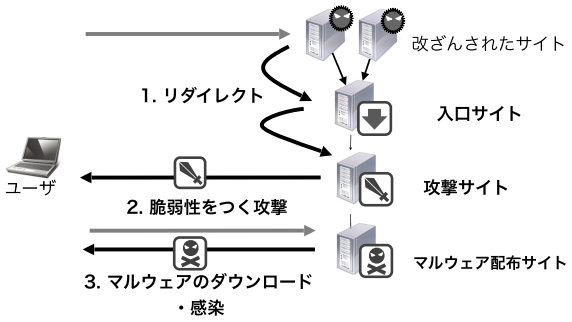

図 1 DBD 攻撃の典型的な流れ
Fig. 1 A typical flow of DBD attacks.

調査し，条件を満たす場合のみ攻撃サイト（Exploit site）へユーザを遷移させる．その際，攻撃サイトの隠ぺいのために，入口サイトから攻撃サイトの間に多くの中継サイト（Intermediate site）を介する場合がある．攻撃サイトでは，ブラウザ，プラグインなどユーザのマシンにインストールされているソフトウェアの脆弱性を攻撃するコンテンツがダウンロードされる．そして攻撃が成功すると，ユーザはさらにマルウェア配布サイト（Malware Distribution site）へ遷移され，最終的にマルウェアを強制的にダウンロードさせられ，感染させられてしまう．

Nappa ら [5] の報告によると，DBD 攻撃に係る悪性サイト（攻撃サイト，マルウェア配布サイト）は生存期間が短く，およそ 60%の悪性サイトは 1 日以内に閉鎖される．そのため発見や解析が困難である．悪性サイトを発見するための方法として，高対話型の Web クローラ（honeyclient）[6], [7] を用いて Web を巡回する方法がある．しかし，効率的な発見のために巡回する Web サイトを適切に決定しなければならない．また，クローラによる発見，検出を回避するために，攻撃サイト側で特定の IP アドレスをブロックするなどの対抗策（cloaking）がとられている場合がある．そのため，クローラによる悪性サイトの発見，検出のみでは十分といえない．

そこで著者らはユーザの協力のもと Web の状況を監視し DBD 攻撃の早期発見，対策を行うためのフレームワーク（FCDBD: Framework for Countering Drive-by Download）を提案，実装した [1], [2]．このフレームワークでは．ユーザは積極的にフレームワークに参加し，自身の Web ブラウジングに関する情報を提供する．セキュリティ分析者はユーザから提供された情報を解析し，結果として得られた脅威に関する情報をユーザにフィードバックする．

FCDBD フレームワークの実現のためには，収集された情報から DBD 攻撃など Web 上の脅威を検出する方法の確立と多くのユーザによる永続的なフレームワークへの協力が必要である．

攻撃の検出方法について，著者らはこれまで DBD 攻撃における Web サイト間のリンク遷移における挙動に着目した手法を提案し [3]，DBD 攻撃の事例を収録したデータセット [10] での評価において効果を確認した．

ユーザからの永続的な FCDBD フレームワークへの協力を得るための取り組みとして，著者らはフレームワークへの参加において障壁となるプライバシへの懸念を解消するための対策を実施した．技術面の対策として，フレームワーク参加時にユーザにインストールしてもらう Web ブラウザのプラグインソフトウェア（観測センサ）の ID を，Web ブラウザの起動ごとにランダムに変更するなどの対策を行った．また，フレームワークにおいて取得する情報の内容，利用目的，保管方法，保管期間など取得した情報にかかわる情報を開示するための文書を制定した．さらにプライバシ関連の有識者を招聘して，文書を含む FCDBD フレームワークでの情報の取得や取扱いなど実施内容を審議するための検討会を実施し，フレームワークの参加者のプライバシ上問題がないか実施内容および文書について審議した．

著者らは，実際に FCDBD フレームワークを用いて 1,000 人規模の参加者を募り，データ収集および検知方式の評価のための実験を実施した．実験により収集された Web ブラウジング情報から，本フレームワークによって従来の観測方法では観測しきれなかった範囲まで観測可能であることが分かった．今回の実験では，収集されたデータにおいて DBD 攻撃による感染事例は確認されなかった．しかしながら，DBD 攻撃の入口サイトと思われる URL へのアクセスは確認されており，これは FCDBD フレームワークによって DBD 攻撃が観測可能であることを示している．さらに参加人数およびデータの収集期間を増やすことで，感染事例を含むより多くの DBD 攻撃事例が観測できると考える．

また，実験により収集された Web ブラウジング環境（Web ブラウザの種類・バージョン，プラグインの種類・バージョン）の情報について，文献 [22] を参考に個人の識別可能性について評価した．その結果，Web ブラウザの種類およびインストールされたプラグインの種類に関する情報は，参加ユーザを識別できるほどの情報量を有することが分かった．こちらについては，たとえば利用している参加者が少ないプラグインの種類については情報を収集しない，などの対策が必要となる．

以降の本論文の構成について記載する．2 章で本論文に関連する既存技術について述べる，3 章で著者らの提案する FCDBD フレームワークについて説明する．5 章では著者らが提案した DBD 攻撃の検出手法について説明と評価結果について述べる．6 章で FCDBD フレームワークの実験について述べ，最後に 8 章で本論文をまとめる．

2. 関連研究

DBD 攻撃サイトを発見，検出する手法の 1 つとして，Web クローラ（honeyclient）を用いた Web サイトの巡回

（クローリング）がある [6], [7]．honeyclient で効率的に悪性サイトを検知するためには，クローリングの起点となる seed を適切に与える必要がある．また，攻撃者が，自身の悪性サイトの検出を防ぐために，セキュリティ関連企業，研究機関によるクローリングと思われるアクセスに対して正常の Web サイトのように振る舞う（cloaking）ような対策を行うこともあり，クローリングによる悪性サイトの発見，検知は非常に困難である．さらに，悪性サイトの生存期間は数時間程度と短命なため，その実態をリアルタイムに把握することは困難である．

FCDBD フレームワークのように，ユーザから収集した情報をもとに脅威情報を把握し，その脅威情報をユーザの保護に利用する仕組みは，たとえば Kaspersky 社の Kaspersky Security Network（KSN）[23] など各セキュリティベンダが提供するソフトウェアでも導入されている．しかしながら，検出された脅威情報は自社製品に閉じて活用され，保護の恩恵は自社製品のユーザに限定される．またプライバシの面において，収集されるデータをユーザが詳細に制御することができない（データを提供する・提供しないの設定しかできない），収集されるデータについての説明が十分でなく一般ユーザにとって分かりにくいなど，プライバシの面での対応が必ずしも十分かつ明確であるとはいえない．

悪性サイトを検出する手法として，Web ページ間の遷移関係の構造（リンク構造）に着目して，未知の悪性サイトを検出する方法が提案されている．Zhang ら [11] の手法は，DBD 攻撃事例の HTTP トラヒック情報から悪性サイトのリンク遷移元をたどり，URL の類似性などを考慮して複数の悪性サイトに共通のハブとなるサイト（central server）を検出し，そのサイトをもとにマルウェア配布のためのネットワーク（MDN: Malware Distribution Network）を検出する．そして，その central server の URL の特徴をシグネチャとして，既知の MDN に属する未知の悪性サイトを検出するものである．Stringhini ら [12] の手法は，リンク構造上の特徴に加えてユーザのマシンの環境（OS，ブラウザ，プラグインなど）も加味して，特定のマシン環境のユーザのみが到達するリンクのパスを抽出することで，膨大な Web アクセスログから DBD 攻撃の悪性サイトを検出するものである．Wand ら [13] の手法は，文献 [11] による MDN の検出の後に，入口サイトのコンテンツの内容，URL の特徴を抽出して，未知の悪性サイトを検出するものである．進藤ら [8] は DBD 攻撃のリンク遷移におけるファイルタイプの変化から特徴を抽出して攻撃の有無を検知する手法を提案している．これらの手法は，新たな MDN の検出，URL，コンテンツなどの特徴の抽出のためには膨大な攻撃事例のデータが必要である．そのため，未知の MDN に属する入口サイトなど悪性サイトの検出には即時的に対応できない．また，MDN の検出，特徴の抽出のための攻撃事例のデータ収集をいかに行うかも重要な課題となる．

3. FCDBD: Framework for Countering Drive-by Download

図 2 に FCDBD フレームワークの概要を示す．FCDBD フレームワークはユーザ側に配置される観測センサとネットワーク側に配置される大規模分析・対策センタ（以降，分析センタ）で構成される．

観測センサはユーザの Web ブラウザ（ブラウザセンサ）に，プラグインソフトウェアの形で配置され，**表 1** に示す Web ブラウジング情報およびセンサ環境情報を収集し，分析センタにレポートする．

観測センサを Web ブラウザのプラグインソフトウェアとして実装することで，Web ページ上でのマウスクリックなどユーザの挙動が観測できる．DBD 攻撃では，前述のとおりユーザの操作なしに Web サイト間の遷移が発生するため，ユーザの挙動に関する情報は悪性サイトの検出に有益である．また，ユーザが煩雑な処理を必要とせずに観測センサを導入できる，既存のアンチウイルスソフトとの併用も容易になるといった点も利点としてあげられる．

分析センタでは，各センサから提供された情報をもとに分析し，悪性サイトを検出する．分析センタは検出された悪性サイトの情報を，たとえばブラックリストとして観測

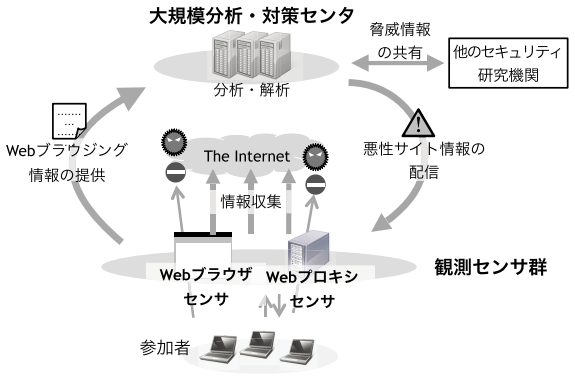

図 2 FCDBD フレームワークの概要
Fig. 2 An overview of the FCDBD framework.

表 1 Web ブラウジング情報，センサ環境情報の主な内容
Table 1 Examples of the information related to web browsing and sensor's environments.

Web ブラウジング情報
・観測センサの ID
・ユーザがアクセスした URL
・ダウンロードしたコンテンツのハッシュ値
・Web ページ遷移時のマウスイベントの有無
・HTTP Request/Response ヘッダ
センサ環境情報
・観測センサの ID
・Web ブラウザの種類・バージョン
・プラグインの種類・バージョン

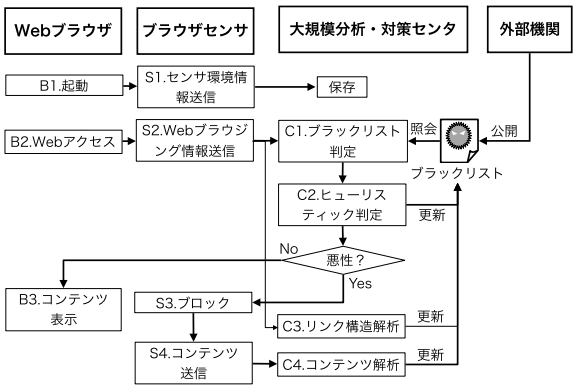

図 3 FCDBD フレームワークの処理フロー
Fig. 3 A flow sequence of the FCDBD framework.

センサ側に提供する．もしくは観測センサから当該サイトへのアクセスがレポートされた際に，そのアクセスをただちに遮断するよう観測センサに通知する．また，検出された悪性サイトの情報について他のセキュリティ関連の研究機関などと情報交換を行うことでセキュリティ包囲網の拡大，セキュリティ技術の研究開発の促進を目指す．

図 3 に FCDBD フレームワークの処理フローの概要を示す．観測センサは，Web ブラウザが起動されると同時に起動される．起動後，観測センサは自身の ID をランダムに生成し，その後，Web ブラウザの種類・バージョン，Web ブラウザにインストールされているプラグインソフトウェアの種類・バージョンといった自身のセンサ環境情報を収集して分析センタに送信する．

Web ブラウザが Web アクセスを行うと，観測センサは Web ブラウザから収集される情報をもとに Web ブラウジング情報を生成する．そして，生成した情報を分析センタに送信する．観測センサから Web ブラウジング情報を受信すると，分析センタはまず Web ブラウジング情報内のアクセスした URL とコンテンツのハッシュ値を，分析センタ内で保持しているブラックリストと照合する（C-1）．当該ブラックリストは，分析センタで過去に悪性と判定したコンテンツの URL およびハッシュ値，外部機関から取得したブラックリストにより構成される．次に分析センタは Web ブラウジング情報をもとに Web サイト間のリンク遷移の特徴に基づくヒューリスティックな判定方法を用いて当該サイトが悪性かどうかを判定する（C-2）．ここで用いるヒューリスティックな判定方法として，後述するコンテンツのダウンロードに至るページ遷移の振舞いに基づく判定方法，コンテンツのダウンロードに至るまでのリダイレクト段数に基づく判定方法を用いる．

分析センタは（C-1），（C-2）での判定結果を観測センサに返答する．観測センサでは，この結果に基づいて Web サイトから取得されたコンテンツをブロックする．ブロックされたコンテンツは，さらにコンテンツの解析（C-4）を行うために分析センタに送信される．コンテンツの解析では，動的解析 [14] と静的解析 [15] を用いてコンテンツを解析する．悪性と判定されると当該コンテンツの情報をブラックリストに登録する．良性と判定されたコンテンツがブラックリストに登録されている場合は，当該コンテンツをブラックリストから削除する．

また，送信された Web ブラウジング情報を用いて，分析サイトは Web サイト間のリンク構造を解析し，疑わしい Web サイトの URL を抽出する（C-3）．Web サイト間のリンク構造の解析による悪性が疑われる Web サイトの URL の検出手法として，たとえば Web サイトからの遷移先サイトの変化に着目して検出する手法 [16]，Web ブラウザの環境の違いによるリダイレクトパスの違いに着目して検出する手法 [17] を用いる．疑わしいとして検出された Web サイトの URL はブラックリストに登録され，再度当該 Web サイトの URL にアクセスされた際にコンテンツを収集し，上述のコンテンツ解析を行う．

4. FCDBD フレームワーク参加者のプライバシへの配慮

FCDBD フレームワークによる Web 観測を実現するためにはユーザの協力が必要不可欠である．ユーザが安心して FCDBD フレームワークに参加できるようにするためには，FCDBD フレームワークの参加者のプライバシへの配慮が必要である．フレームワーク参加者のプライバシ対策を検討するにあたり，プライバシ関連の技術および法律の有識者を招聘した検討会を実施するなどして，実験実施におけるプライバシ上の課題点について明らかにした．

4.1 実験の実施内容の検討会

実際にフレームワークの参加者の募集および実験を行うにあたり，プライバシ関連の技術および法律の有識者を招聘して実験の実施内容に関する検討会を実施し，参加者のプライバシに配慮するために行うべき対策について議論した．検討会では，主に文書によって開示すべき情報について以下のように提示された．

- 実証実験の実施主体，取得情報の管理主体は委託元の情報通信研究機構，受託者の KDDI 研究所，セキュアブレインの三者共同とする．
- 取得情報の取扱いについて，取得情報の内容，提供・共有先，保有期間，利用目的を文書に明記すること．
- 実験の目的，実験参加によるリスクを文書に明記すること．

4.2 参加者のプライバシに配慮した実験に向けた課題点

実験の実施に先駆け，参加者のプライバシに配慮するうえで課題となる点について，4.1 節の検討会での議論などをふまえて以下のとおり整理した．

- 実験の実施について

– 実験の実施主体を明確にする．
– 実験の目的を明確にする．
– 実験参加によるリスクを明確にする．

- 実験の参加について
  – オプトイン形式にする．
  – 参加者が任意のタイミングで参加を止められる．
- 取得情報について
  – 参加者による取得情報の制御を可能にする．
  – 参加者の承諾なしに情報を収集しない．
  – 認証情報など参加者の秘密に関わる情報を収集しない．
  – 取得した情報から参加者個人が特定されないようにする．
  – 取得情報の内容，提供・共有先，保有期間，利用目的を明確にする．
  – 取得情報の管理主体を明確にする．

4.3 プライバシに配慮した実験の実施に向けた対応

前節の課題点への対応として「文書による実験の実施内容の開示」「技術面での対応」の 2 通りの対応を実施した．

4.3.1 文書による実験の実施内容の開示

実験を実施するにあたり，実験の実施内容，参加・脱退の方法，さらに取得情報について開示した文書を制定した．そして，実験の参加者に対して当該文書の内容への同意を得ることとした．実際の文書の内容は付録に記載した．

4.3.2 技術面での対応

技術面においては，収集したデータから参加者個人の Web アクセス履歴が過度に追跡されないように，ブラウザセンサにおいては Web ブラウザが起動されるごとに観測センサの ID をランダムに変化させるようにした．さらに Web ブラウジング情報の収集において，参加者のプライバシに関わる情報を過度に収集することがないように，また参加者が意図しない情報を勝手に収集することがないように以下の対策を実施した．

- HTTP ヘッダから取得する情報はレスポンスコード，Content-Type，Content-Length（コンテンツの情報），Connection, Refer, Location（リンクに関する情報）のみとし，Authorized（認証情報），Cookie，Set-Cookie（クッキーの情報）は取得しないように制限した．
- 参加者が明示的に許可しない限り HTTPS での通信に関する情報は取得しないこととした．
- コンテンツ自身を収集する際は，ダイアログでそのつど通知し，参加者に承認されたときのみ，コンテンツを取得することとした．
- 収集される情報は，項目ごとに許可する/しないを選択できることとした．

5. Web サイト間のリンク遷移の挙動に基づくヒューリスティックな検出手法

5.1 手法 1：マルウェアのダウンロード時のリンク遷移における挙動に着目した悪性サイト検出手法

図 4 に DBD 攻撃におけるマルウェアのダウンロードまでのページ遷移の事例を示す．入口サイトにアクセスしたユーザは，難読化されたスクリプトによって生成される HTML タグによって，攻撃サイトもしくは中継サイトに遷移させられる．そして攻撃サイトから脆弱性をつくためのコンテンツ（Java アーカイブ，PDF ファイルなど）をダウンロードさせられる．そして攻撃が成功するとマルウェアをダウンロードさせられる．その際，上述のとおり，一連のページ遷移は難読化されたスクリプトおよび脆弱性をつく攻撃によって引き起こされるため，マルウェアのダウンロードに至るページ遷移は，一連のページ遷移の過程においてダウンロードされた HTML ファイル，JavaScript などからは容易に推測できない（特徴 1）．また，マルウェアのダウンロードは攻撃によって引き起こされるため，当該ダウンロードが引き起こされるもととなった参照元のページの URL は Referer，Location ヘッダなど HTTP ヘッダ上に記載されない（特徴 2）．以上の特徴をもとに，本手法では観測センサから送信される情報をもとに以下の 2 つの条件をいずれも満たすページ遷移によって実行ファイルがダウンロードされた場合に，マルウェアなど悪性コンテンツのダウンロードと判定する．以下，初期ページとはユーザによるリンクのクリック，ブックマークの選択などにより初めにアクセスされる Web ページのことを示す．

条件 1：初期ページから引き起こされる一連のページ遷移に係るすべての HTTP リクエスト/レスポンスヘッダに，対象となる実行ファイルの参照元となる情報（e.g., Referer ヘッダ，Location ヘッダ）が存在しない．

条件 2：初期ページから引き起こされる一連のページ遷移によってダウンロードされるすべての HTML ファイル，JavaScript 内に，対象となる実行ファイルへの遷移を示す

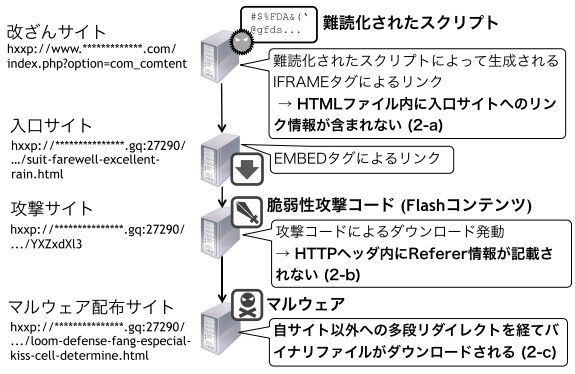

図 4 DBD 攻撃の事例
Fig. 4 An example of DBD attacks.

情報が存在しない．

5.2 手法 2：リダイレクト段数による検出手法

DBD 攻撃では改ざんされた Web ページから複数回遷移を繰り返した後にマルウェアに感染する．本手法では，上記特徴に基づきあるしきい値以上の遷移の後にバイナリデータがダウンロードされた際にそれを検知する [9]．

5.3 評価

D3M データセット [10] を用いて評価を実施した．

5.3.1 D3M データセット

D3M（Drive-by Download Data by Marionette）データセット [10] は，マルウェア対策研究人材育成ワークショップ（MWS: Anti Malware Engineering Workshop）でマルウェア対策技術の技術者の育成および研究促進を目的として配布されているデータセットである．D3M データセットには，Marionette [7] と呼ばれる honeyclient を用いて収集された DBD 攻撃に係る悪性サイトへのアクセス時のトラヒックおよびマルウェアへの感染によって引き起こされたトラヒックのキャプチャデータが収録されている．

本論文では，悪性サイトの Web アクセスデータのサンプルとして D3M データセット 2013，2014 を用いた．手法の評価には，バイナリデータ（Content-Type ヘッダの値が application/x-msdownload，application/octet-stream）がダウンロードされていた 47 URL を抽出し，そのアクセスデータを評価用データとして用いた．

正規サイトのサンプルについて，手法 1 の評価用として，実行ファイルなど（.exe，.dll，.pdf，.swf，.msi，.zip）を自動的にダウンロードさせるサイトのうちアクセス数の多かったもの 75 件について，当該サイトのコンテンツおよび当該サイトへアクセスした際の Web ブラウジング情報を収集した．手法 2 の評価用として，Content-Type ヘッダの値が application/x-msdownload, application/octet-stream のバイナリデータがダウンロードされたサイト 429 件についての Web ブラウジング情報を用いた．

5.3.2 評価結果

表 2 に手法 1 の評価結果を示す．表内の#URLs はサンプルとして用いたサイト（URL）の数，条件 1，条件 2 はそれぞれ上述の条件 1，条件 2 を満たすサイトの数，条件 1∧2 は条件 1，条件 2 の両方を満たすサイトの数を示す．表 2 より，今回の D3M データセットより抽出したサンプルにおいてはすべて条件 1 および条件 2 を満たしていた．また正規サイトにおいては，条件 1，条件 2 の両方を満たすサイトは存在しなかった．

表 3 に手法 2 の評価結果を示す．D3M データセットに収録されているデータは，図 4 の入口サイトに相当するサイトのデータが収録されていると想定される．そのため，D3M データセットのリダイレクト段数は入口サイトからの段数であると想定される．表 3 のデータを見ると，72.3%の入口サイトがバイナリデータまでに 2 段のリダイレクトを経ている．残りのサイトは段数が 1 となっているが，これらのサイトにおいても脆弱性攻撃コードと思われる PDF ファイル，Flash コンテンツがダウンロードされていた．これらのことから，悪性サイトにおいて改ざんサイトからのリダイレクト段数は 3 段以上と想定することができる．

正規サイトのリダイレクト段数を見ると，4.2%のサイトがバイナリデータまでに 3 段以上のリダイレクトを有している．そのため，手法 2 では 4.2%の false positives が発生することとなる．

表 2 手法 1 の評価結果
Table 2 The evaluations of the method 1.

サンプル	条件 1	条件 2	条件 1∧2
D3M データセット（$N = 47$）	47	47	47
正規サイト（$N = 75$）	32	0	0

表 3 手法 2 の評価結果
Table 3 The evaluations of the method 2.

リダイレクト段数	0	1	2	3	≥ 4
D3M データセット（$N = 47$）	0	13	34	0	0
正規サイト（$N = 429$）	228	151	32	16	2

表 4 実験で収集されたデータの諸元
Table 4 The profile of our field trial.

ユーザ数 (1)	1,676
全センサ ID 数 (2)	49,146
全 Web アクセス情報数	4,425,689
ユニーク URL 数	2,178,381
ユニーク FQDN 数	34,195

(1) 2015 年 10 月 21 日に 1,676 人に到達した．
(2) センサ ID は Web ブラウザの起動ごとに変化する．

6. 実際のユーザによる FCDBD の実験

開発した FCDBD フレームワークを用いて，実際に 1,000 人程度のユーザにブラウザセンサを配布してデータの収集を実施し，5 章の検出手法を評価した．さらに，実験の参加者に対してアンケートを行い，プライバシに関する意識調査ならびに実験の実施に先駆けて策定した文書の効果の評価を行った．

実験により収集されたデータの諸元を**表 4** に示す．なお，表 4 に示したデータは，2015 年 7 月 1 日～2015 年 11 月 30 日に収集されたデータに基づき集計した．

6.1 検出手法の評価結果

収集された全 URL において Google Developers で提供されている Safe Browsing API [20] を用いて調査したところ，23 件がマルウェア関連の Web サイトの URL として

表 5 実験による検出手法の評価結果
Table 5 The results of our field trial.

手法 1（ダウンロードの振舞い）	0
手法 2（リダイレクト段数）	11
Google Safe Browsing	23

検出された．しかし，検出された URL へのアクセスの後にバイナリデータがダウンロードされた形跡がないことから，今回の実験期間中においては DBD 攻撃による感染は観測されなかった．

表 5 に FCDBD フレームワークでの検出方法の評価結果を示す．今回の実証実験では，5.1 節に記載の手法で検出される Web サイトは存在しなかった．上述のとおり，今回の実証実験では DBD 攻撃による感染は観測されなかったため，本手法での誤検知数，未検知数はともに 0 であると考えられる．

一方，リダイレクト段数による検出では 11 件検出され，そのうち拡張子で明らかに画像と思われる URL 以外の URL は 7 件であった．7 件についてさらに VirusTotal [21] でコンテンツハッシュ値をもとに調査を実施したところ，悪性なコンテンツであるとは確認できなかった．以上より，検出された 11 件はすべて誤検知と推測される．

6.2 アンケートによる意識調査

実験の参加者を対象に，今回の実験に関してアンケートを実施した．その回答結果からユーザにおけるプライバシに対する意識について考察する．**表 6**，**表 7**，**表 8** にアンケートの回答結果を示す．なお，アンケートは 15〜60 歳までの男女 1,000 人に対して実施し，そのうち 506 人から回答が得られた．

表 6 は参加者が実験への参加にあたって不安に感じた点についての調査結果である．これより，「インストールするソフトウェア」「（実験の）実施期間」など実験の実施内容自体よりも収集される情報について不安に感じていた人が約半数と多いことがうかがえる．

表 7 では，参加者が実験へ参加する決め手となった項目についての調査結果である．表のとおり「謝礼の金額が妥当だった」が全体の 38.5%と最も多かった．ちなみに今回の謝礼金額は 2,000 円（2015/10/21 からの参加者には 1,000 円）であった．また，「収集される情報の説明」「実験の実施内容・目的の説明」が十分かつ明確になされていたと約 1/4 の参加者が回答していた．「収集される情報の説明」「実験の実施内容・目的の説明」と回答した参加者のうち「謝礼の金額が妥当だった」とも回答した参加者は，それぞれ 126 人中 26 人，128 人中 33 人と少なかった[*1]．このことから，今回のようなユーザの行動に関して情報を収集するような実験においては，謝礼の支払いのみではなくしっかりと説明を行うことで参加者の不安を取り除くことが重要であるといえる．

[*1] 有意水準 1%の残差分析において有意に少ない．

表 6 項目 1：実験への参加にあたって不安だった点は？
Table 6 Q1: What were your concerns about our field trial?

（複数回答可）

収集される情報	276（54.5%）
インストールするソフトウェア	258（51.0%）
実施期間	31（6.1%）
実験の実施者	47（9.3%）
その他	51（10.0%）
特にない	88（17.4%）

表 7 項目 2：実験への参加を決めた理由は？
Table 7 Q2: Why did you decide to participate our field trial?

（複数回答可）

収集される情報の説明が十分になされていた	126（24.9%）
実験の実施内容・目的の説明が明確だった	128（25.3%）
セキュリティ関連の話題・技術に興味があった	151（29.8%）
セキュリティ関連の技術開発に貢献したかった	80（15.8%）
謝礼の金額が妥当だった	195（38.5%）
実験の実施者が信頼できるところだった	83（16.4%）
その他	21（4.2%）

表 8 項目 3：規約など各文書で確認した項目は？
Table 8 Q3: What terms did you check in the agreements?

（複数回答可）
（別の設問で「文書をすべて確認した」と回答した 180 人に対して質問）

実験を実施する企業・組織	98（54.4%）
実験の目的	130（72.2%）
実験の実施内容	83（46.1%）
実験の参加者が行う事項と実施期間	66（36.7%）
収集される情報の種類	76（42.2%）
収集される情報の保管場所・期間	49（27.2%）
収集される情報の利用目的	87（48.3%）
収集される情報の提供先	57（31.7%）
収集される情報と連絡先情報との関連	47（26.1%）
保証や責任範囲	50（27.8%）
その他	0（0%）

表 8 では，各文書について実際にすべての文書を確認した 180 人に対して，確認した項目について調査した結果である．表のとおり，参加者は「実験を実施する組織・企業」「実験の目的」「実験の実施内容」とともに「収集される情報の種類」といった実験の詳細まで関心があることがうかがえる．そのため，参加者と締結する規約文書においては，詳細な内容まで記載し，参加者と実験の実施者の双方が納得した形で実験を実施することが望ましい．

7. 観測データの検証

実験によって収集されたデータをもとに，FCDBD フレームワークにおけるユーザ環境の多様性，ユーザ環境に

よる個人識別性，ならびに観測データの網羅性について検証を行った．

7.1 ユーザ環境の多様性

表 9 に参加者の利用した Web ブラウザの種類の内訳を示す．表より，センサ ID ごとの Web ブラウザの内訳は Firefox のセンサ数が Internet Explorer の約 7 倍であった．しかしながら，参加者に対して実施したアンケートでは，58.3%の参加者が Internet Explorer を利用したと回答した．この差異は，Firefox 版と Internet Explorer 版のブラウザセンサの動作仕様の違いが影響していると考えられる．Internet Explorer 版のブラウザセンサには Web ブラウザを起動したままブラウザセンサの機能を ON/OFF する機能が備わっていなかった（ブラウザセンサをアンインストールしない限り，Web ブラウザの起動時はつねにブラウザセンサの機能が ON になる）のに対し，Firefox 版ではブラウザセンサの機能を ON/OFF する機能が備わっていた．そのため，Firefox 版ではユーザが閲覧する Web ページに応じてこまめにブラウザセンサ機能の起動・停止を行うことが可能であり，結果としてセンサ ID の数が Internet Explorer より多くなったのではないかと考えられる．Web ブラウザのバージョンについては，Internet Explorer が 8 から 11 まで，Firefox が 31 から 43 まで存在した．**表 10** に 2015 年 10 月，2015 年 11 月の各ブラウザのバージョンごとのセンサ ID の数を示す．最新バージョンのブラウザを利用している参加者が大多数であるが，古いバージョンを利用している参加者が両ブラウザともに若干数みられた．Web ブラウザにインストールされていたプラグインの種類は 488 種類あり，バージョン情報を含めると全部で 1,231 種類存在した．これらと Web ブラウザの種類，バージョンとあわせると合計 4,067 種類もの組合せがみられた．

表 9 参加者の Web ブラウザの内訳
Table 9 The number of web browsers.

Web ブラウザ	センサ ID 数（$N = 49{,}146$）	アンケートの回答（$N = 506$）
Internet Explorer	5,639（11.5%）	295（58.3%）
Firefox	39,801（81.0%）	211（41.7%）
その他・不明	3,706（7.5%）	0（0%）

表 10 参加者の Web ブラウザのバージョンの内訳
Table 10 The number of versions of web browsers.

	2015/10	2015/11
Internet Explorer 11	60	71
Internet Explorer 10	0	0
Internet Explorer 9	5	6
internet Explorer 8	6	0
Firefox 43	3	8
Firefox 42	2	544
Firefox 41	494	313
Firefox 40	77	21
Firefox 39 以下	41	30

※ リリース日：Firefox 41：2015/9/22，Firefox 42：2015/11/3 [18]

7.2 ユーザ環境による識別性の検証

収集されたユーザ環境の情報から個別のユーザを特定できるかどうかを定量的に評価した．評価は，文献 [22] を参考に，各々のユーザ環境の情報量をビット数で表すことで実施した．以下に情報量の算出式を記載する．ここで $F = \{f_1, \ldots, f_n\}$ をユーザ環境の情報の集合，$P(f_k)$ をユーザ環境の情報 f_k（$k \in \{1, \ldots, n\}$）における確率分布，$I(f_k)$ をユーザ環境の情報 f_k による情報量とする．

$$I(f_k) = -\log_2(P(f_k)) \tag{1}$$

なお，今回は簡易な評価として，各ユーザ環境の情報の確率分布を一様分布であると仮定し，情報量の下限を求めて識別性を評価することとした．

まず，ブラウザの種類による情報量を計算する．今回収集されたユーザ環境の情報において，ブラウザの種類は 5 種類存在した．よってブラウザの種類による情報量は $I(f_{browser}) = -\log_2(P(f_{browser})) = -\log_2(1/5) \approx 2.32$ ビットとなる．また，ブラウザの種類およびバージョンは全部で 64 通りであった．この場合の情報量は $l(f_{browser\_ver}) = -\log_2(1/64) = 6$ ビットとなる．

次にプラグインの種類による情報量を計算する．参加ユーザの Web ブラウザにインストールされていたプラグインのパターンは全部で 858 通り存在した．よってプラグインのパターンによる情報量は $I(f_{plugin}) = -\log_2(1/858) \approx 9.73$ ビットとなる．プラグインのバージョン情報も含めると 1,546 通りとなり，その情報量は $I(f_{plugin\_ver}) \approx 10.6$ ビットとなる．

最後にブラウザ，ブラウザのバージョン，プラグイン，プラグインのバージョンの組合せは全部で 4,067 通りであったことから，これらによる情報量は $I(f_{sinfo\_ver}) \approx 12.0$ ビットとなる．バージョン情報を省いた場合は全部で 1,554 通りとなり，情報量は $I(f_{sinfo}) \approx 10.6$ ビットとなる．

今回の実験の参加ユーザ数は 1,676 人（$I(f_{user}) \approx 10.7$）であることから，ブラウザおよびインストールされているプラグインの情報は参加ユーザをほぼ識別しうる情報量を有することが分かる．さらに，ブラウザおよびプラグインの情報において，同一の組合せを有するセンサ ID 数を集計すると，同一の組合せを有するセンサ ID が複数（2 個以上）存在した組合せは 1,269 通りであった．また同一の組合せを有するセンサ ID 数の最大値は 1,475，平均で約 36 であった．このことから，もしブラウザとプラグインの組合せと参加ユーザがほぼ 1 対 1 に対応づけされるとし，参加ユーザが 1 日に割り当てられるセンサ ID 数を 1 とすると，平均で約 1 カ月分の Web ブラウジング履歴がトレー

スされうることとなる．

以上より，ユーザのプライバシ保護のためには，今回のセンサ ID のランダム化に加えて，センサ情報の匿名化も必要である．たとえば，情報を収集するプラグインを利用者の多いプラグインに限定するなどが考えられる．

7.3 FCDBD の網羅性の検証

収集されたデータから FCDBD フレームワークの網羅性について検証する．まず，Alexa [19] から取得した日本でのドメイン別アクセスランキングをもとに，Alexa の上位 100 ドメインのうち，FCDBD フレームワークの参加者によってアクセスされた初期ページのドメインに含まれる割合について調べたところ 100%であった．これより，FCDBD フレームワークによって収集されるデータは，主要な Web サイトを漏れなく網羅していると考えられる．

次に，FCDBD フレームワークによって観測される Web サイトの範囲について考察する．FCDBD フレームワークでは，センサによって観測された Web ブラウジング情報をもとに悪性サイトの URL を検出し，その後他のセンサが同じ悪性サイトの URL にアクセスするのを防ぐ．そのため，少なくとも 2 回同じ URL にアクセスされることで悪性サイトへのアクセス防止効果が得られる．このことから，2 回以上アクセスされた Web サイトの URL を，FCDBD フレームワークによって観測された Web サイトとして集計する．

表 11 に FCDBD のフレームワークの実験によって収集されたデータの内訳を示す．

実験期間中にアクセスされた Web サイト 2,178,381 URL のうち，FCDBD フレームワークで観測された Web サイトは 212,804 URL であった，そのうち，Alexa の日本での上位 100 ドメインに該当しない Web サイトは 156,112 URL であった．

表 11 FCDBD フレームワークによる収集データの内訳
Table 11 The details of obtained data by our field trial.

ユニーク Web サイト数	2,178,381
2 回以上アクセスされた Web サイト数	212,804 (9.8%)
Alexa 日本上位 100 ドメイン上の Web サイト数	56,692 (2.6%)
Alexa 日本上位 100 ドメイン以外の Web サイト数	156,112 (7.2%)
総アクセス数	4,425,689
2 回以上アクセスされた Web サイトへのアクセス数	2,460,112 (55.6%)
Alexa 日本上位 100 ドメイン上の Web サイトへのアクセス	668,537 (15.1%)
Alexa 日本上位 100 ドメイン以外の Web サイトへのアクセス	1,791,575 (40.5%)

アクセス数の割合でみると，全アクセス数のうち，観測された Web サイトへのアクセス数の割合は全体の 55.6%であり，そのうち Alexa 日本上位 100 ドメイン以外のドメインへのアクセス数の割合は全体の 40.5%であった．このことから，Alexa での上位ドメイン上の Web サイトを巡回する場合と比較して，FCDBD フレームワークによって新たに全体の 40.5%にあたるアクセスが観測され，保護されることとなる．

8. まとめ

本論文では，著者らが提案，実装したユーザ参加型のドライブ・バイ・ダウンロード攻撃対策フレームワーク（FCDBD: Framework for Countering Drive-By Download）において，実際に 1,000 人規模の参加者を募った実験を実施した．その結果，本フレームワークによって Alexa の上位ドメイン上の Web サイトの巡回のみでは観測しきれなかった範囲まで観測可能であることが分かった．また，DBD 攻撃に関連するサイトの URL が観測されていることが確認された．今後，参加者と収集期間を増やすことで DBD 攻撃の感染事例の観測，それにともなう DBD 攻撃の抑止効果が期待できる．さらに収集されたセンサ環境情報の個人識別性について定量的に評価したところ，Web ブラウザの種類とインストールされているプラグインの種類に関する情報については，今回の参加人数分（1,676 人）のユーザを識別可能な情報量を有することが分かった．対策として，たとえば情報を収集するプラグインを限定するなどが考えられる．

謝辞 本研究成果は国立研究開発法人情報通信研究機構（NICT）の委託研究「ドライブ・バイ・ダウンロード攻撃対策フレームワークの研究開発」により得られたものである．ここに深謝する．また，本委託研究の共同受託者であり FCDBD のフレームワークの実験の共同実施者である株式会社セキュアブレインの関係者各位に深謝する．

参考文献

[1] 笠間貴弘，井上大介，衛藤将史，中里純二，中尾康二：ドライブ・バイ・ダウンロード攻撃対策フレームワークの提案，コンピュータセキュリティシンポジウム 2011（CSS2011）(2011).
[2] Matsunaka, T., Urakawa, J. and Kubota, A.: Detecting and Preventing Drive-by Download Attack via Participative Monitoring of the Web, *Proc. 8th Asia Joint Conference on Information Security* (*AsiaJCIS2013*) (2013).
[3] Matsunaka, T., Kubota, A. and Kasama, T.: An Approach to Detect Drive-by Download by Observing the Web Page Transition Behaviors, *Proc. 9th Asia Joint Conference on Information Security* (*AsiaJCIS2014*) (2014).
[4] Provos, N., Mavrommatis, P., Rajab, M.A. and Monrose, F.: All Your iFRAMEs Point to Us, *Proc. 17th USENIX Security Symposium* (2008).

[5] Nappa, A., Rafique, M. and Caballero, J.: Driving in the Cloud: An Analysis of Drive-by Download Operations and Abuse Reporting, *Proc. 10th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (*DIMVA2013*) (2013).
[6] Wang, Y.-M., Beck, D., Jiang, X., Verbowski, C., Chen, S. and King, S.: Automated Web Patrol with Strider HoneyMonkeys: Finding Web Sites That Exploit Browser Vulnerabilities, *Proc. 13th Annual Network & Distributed System Security Symposium* (*NDSS2006*) (2006).
[7] Akiyama, M., Iwamura, M., Kawakoya, Y., Aoki, K. and Itoh, M.: Design and Implementation of High Interaction Client Honeypot for Drive-by-Download Attack, *IEICE Trans. Comm.*, Vol.E93-B, No.5, pp.1131–1139 (2010).
[8] 進藤康孝，佐藤彰洋，中村 豊，飯田勝吉：マルウェア感染ステップのファイルタイプ遷移に基づいた Drive-by Download 攻撃検知手法，コンピュータセキュリティシンポジウム 2014（CSS2014）(2014).
[9] 安藤槙悟，寺田真敏，菊池浩明，趙 晋輝：通信の遷移に着目した不正リダイレクトの検出による悪性 Web サイト検知システムの提案，情報処理学会研究報告，Vol.2011-CSEC-54，No.33 (2011).
[10] 神薗雅紀，秋山満昭，笠間貴弘，村上純一，畑田充弘，寺田真敏：マルウェア対策のための研究用データセット–MWS Datasets 2015，情報処理学会研究報告，Vol.2015-CSEC-70, No.6 (2015).
[11] Zhang, J., Seifert, C., Stokes, J.W. and Lee, W.: ARROW: GenerAting SignatuRes to Detect DRive-By DOWnloads, *Proc. 20th International World Wide Web Conference* (*WWW2011*) (2011).
[12] Stringhini, G., Kruegel, C. and Vigna, G.: Shady Paths: Leveraging Surfing Crowds to Detect Malicious Web Pages, *Proc. 20th ACM Conference on Computer and Communications Security* (*CCS2013*) (2013).
[13] Wand, G., Stokes, J.W., Herley, C. and Felstead, D.: Detecting Malicious Landing Pages in Malware Distribution Networks, *Proc. 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (*DSN2013*) (2013).
[14] 神薗雅紀，西田雅太，星澤裕二：動的解析を利用した PDF マルウェア解析システムの実装と評価，電子情報通信学会技術研究報告，情報通信システムセキュリティ（ICSS），Vol.110, No.475 (2011).
[15] 西田雅太，星澤裕二，笠間貴弘，衛藤将史，井上大介，中尾康二：文字出現頻度をパラメータとした機械学習による悪質な難読化 JavaScript の検出，情報処理学会研究報告，Vol.2014-CSEC-64, No.21 (2014).
[16] 松中隆志，山田 明，窪田 歩：Drive-by Download 攻撃対策フレームワーク実現に向けたリンク構造解析による Web サイトの分析，情報処理学会研究報告，Vol.2015-CSEC-68, No.48 (2015).
[17] 笠間貴弘，衛藤将史，神薗雅紀，井上大介：クライアント環境に応じたリダイレクト制御に着目した悪性 Web サイト検出手法，電子情報通信学会技術研究報告，情報通信システムセキュリティ（ICSS），Vol.114, No.71 (2014).
[18] RapidRelease/Calender – MozillaWiki, available from ⟨https://wiki.mozilla.org/RapidRelease/Calendar⟩.
[19] Alexa – Actionable Analytics for the Web, available from ⟨http://www.alexa.com/⟩.
[20] Safe Browsing API – Google Developers, available from ⟨https://developers.google.com/safe-browsing/⟩.
[21] VirusTotal – Free Online Virus, Malware and URL Scanner, available from ⟨https://www.virustotal.com/⟩.
[22] Eckersley, P.: How Unique Is Your Web Browser?, *Proc. 10th International Conference on Privacy Enhancing Technologies* (*PETS2010*) (2010).
[23] Kaspersky Lab., Kaspersky Security Network, available from ⟨http://ksn.kaspersky.com/⟩.

付 録

以下に制定した文書を記載する．文書は FCDBD フレームワークの実験への参加に関する規約を定めた「ドライブ・バイ・ダウンロード攻撃の観測・分析・対策技術の研究開発実証実験参加規約」と，FCDBD フレームワークで収集する情報の内容，保管など取り扱いについて記載した「実証実験における取得情報の取り扱いについて」からなる．

A.1 ドライブ・バイ・ダウンロード攻撃の観測・分析・対策技術の研究開発実証実験参加規約

A.1.1 本実証実験の目的

本実証実験は，ユーザ群の Web アクセスの挙動に基づくドライブ・バイ・ダウンロード攻撃の観測・分析・対策技術の評価，確認を目的とします．

A.1.2 適用範囲

株式会社 KDDI 研究所（以下，「甲 1」と言います．），株式会社セキュアブレイン（以下，「甲 2」と言います．）と独立行政法人情報通信研究機構（以下，「甲 3」と言います．）（以下，甲 1，甲 2，甲 3 を総称して「甲」と言います．）は「ドライブ・バイ・ダウンロード攻撃の観測・分析・対策技術の研究開発」を遂行する中で，実証実験を実施します．実証実験に参加する者（以下，「乙」と言います．），は，本実証実験への参加にあたって，本規約に同意し，遵守するものとします．

A.1.3 実証実験の内容

(1) 乙は，乙が使用するコンピュータにブラウザ型センサをインストールします．
(2) 乙は，ブラウザ型センサをインストールしたブラウザで Web アクセスを行います．Web アクセスを行うと，まず，ブラウザ型センサはアクセス先に関して，ブラウザ型センサが持つ機能でアクセスを遮断した方がよいか判定します．遮断した方がよいと判定した場合には，警告画面を表示します．次に，ブラウザ型センサは，甲が管理する大規模分析・対策センタに，別紙「実証実験における取得情報の取り扱いについて」に定める「ブラウザ型センサ取得情報」を送信し，大規模分析・対策センタのブラックリストとの照合を要求します．ブラックリストに登録されていた場合には，ブラウザ型センサは乙に対し警告画面を表示します．
(3) 甲 1，甲 2 は，送信されたブラウザ型センサ取得情報

を大規模分析・対策センタ内で管理し，以下のように分析します．

- Webサイトから誘導されるリンク先の変化が通常と異なるか
- Webページに用いられるスクリプトが悪性のものの特徴を備えているか
- 悪性の可能性があるファイルを実際に実行しその挙動が悪性のものの特徴を有するか

などを解析し，乙がアクセスしたWebページにドライブ・バイ・ダウンロード攻撃があるか否かを総合的に判断します．大規模分析・対策センタでドライブ・バイ・ダウンロード攻撃に関連するサイトであると判断されたWebサイトのURLをブラックリストに登録します．

(4) 甲3は，ブラウザ型センサ取得情報を大規模分析・対策センタから取得し，甲3の責任のもと管理し，以下のように分析します．

- URLやその引数，リンク先への誘導の方法を含めた特徴が，既知の悪性サイトの特徴と類似するか
- ブラウザにインストールされたプラグインやそのバージョンなどの情報をもとに誘導するリンク先を変更する手法が用いられているか
- アクセス先のホストが他の攻撃観測手法によって観測された悪性ホストと一致するか

などを解析し，乙がアクセスしたWebページにドライブ・バイ・ダウンロード攻撃があるか否かを総合的に判断します．

A.1.4 実証実験における乙の実施事項

(1) 甲が提供するブラウザ型センサをダウンロードしインストールします．
(2) ブラウザ型センサをインストールしたブラウザで，普段どおりWebアクセスを行います．
(3) 甲の指定するアンケートに回答します．

A.1.5 参加登録

乙は，本規約，並びに別途甲が提示する「実証実験における取得情報の取り扱いについて」及び「ソフトウェア使用許諾契約約款」の内容を確認し，同意した上，甲の提供するWebサイトで参加登録を行うものとします．

A.1.6 実証実験期間

(1) 本実証実験の実施期間は，2014年11月1日から2016年3月31日までの間で甲が定める期間とします．
(2) 甲は，乙の承諾を得ることなく，本実証実験の実施期間を変更することがあります．その場合には，実施期間の変更を事前もしくは事後速やかにWebサイトへ掲載するとともに，電子メールの送付により乙に通知するものとします．

A.1.7 乙の実証実験参加終了

(1) 乙は，いつでも実証実験から脱退することができます．
(2) 乙は，ブラウザ型センサをいつでもアンインストールすることができます．
(3) 乙は，実証実験から脱退する際に，甲に実証実験から脱退する旨を連絡します．甲は，乙からの連絡を受けて，「実証実験における取得情報の取り扱いについて」に定める「連絡先情報」および「実証実験参加同意書」を廃棄します．

A.1.8 乙の情報提供と保護

乙は，甲が本実証実験を実施するうえで必要となる乙の情報を甲に提供することに同意するものとします．甲は，本実証実験で取得した乙の情報に関し，別紙に定める「実証実験における取得情報の取り扱いについて」に基づき，適切な取り扱いを行うものとします．

A.1.9 免責事項

(1) ブラウザ型センサのインストール及び使用により乙のコンピュータシステムが破損する，データが消失する等の可能性があります．その場合でも，甲は，これらにより乙に生じた損害について，一切の責任を負いません．
(2) ブラウザ型センサをインストールすることにより，甲が悪性と判断し，ブラックリストに登録したWebページへのアクセスが遮断されます．乙が意図していない遮断があっても，甲は乙に一切の責任を負いません．また，すべての悪質サイト等へのアクセスが遮断されるものではありません．Webサイトへのアクセスは乙自らの責任において行ってください．甲は，悪質サイト等へのアクセスにより生じた損害について，乙に対し一切の責任を負いません．
(3) 悪質サイト等へのアクセスが遮断できない場合が発生するのは，例えば以下のような場合です．下記の例ですべてのケースを網羅しているわけではありません．

- 悪質と疑われるサイトに関する分析を大規模分析・対策センタで実施中であり，ブラックリストへの登録がまだ行われていない場合
- ブラウザ型センサが有する悪質コンテンツの検出機能で想定するパターン外の悪質コンテンツへのアクセスが行われた場合

(4) 乙は，「ソフトウェア使用許諾契約」に従うものとし，同契約に違反する使用等により生じた損害について，甲は，乙に一切の責任を負いません．
(5) 乙のブラウザ型センサの使用行為により，第三者に損害等が生じた場合，乙は，自らの責任をもって当該第

三者との間で問題を解決するものとします．この場合，甲は一切の責任を負わないものとします．

A.1.10 本規約の変更

甲は乙の承諾を得ることなく，本規約を変更することができるものとします．その場合には，変更内容を事前にWebサイトへ掲載するとともに，電子メールの送付により乙に通知するものとします．

A.1.11 準拠法および管轄裁判所

本規約の準拠法は日本法とします．また，本規約に関連して甲と乙との間で生じた紛争については東京地方裁判所を第一審専属管轄裁判所とします．

A.2 実証実験における取得情報の取り扱いについて

A.2.1 連絡先情報の利用目的

「実証実験参加同意書」およびメンバサイト登録の際に入力していただいた名前，電子メールアドレス，電話番号等（以下，「連絡先情報」といいます．）は，以下の目的でのみ利用します．

(1) 実験参加者への連絡
(2) 実験参加者からの問い合わせ対応
(3) 実験参加者への謝礼の送付

A.2.2 連絡先情報の取得および保管

(1) 実証実験参加同意書および連絡先情報は，株式会社セキュアブレイン（以下，「セキュアブレイン」といいます．）が取得，保管を行います．
(2) ご提出いただいた実証実験同意書および連絡先情報は，本研究が終了する2016年3月末をもって廃棄します．実施期間を変更する場合は，変更を事前もしくは事後速やかにWebサイトへ掲載するとともに，電子メールの送付により参加者に通知するものとします．
(3) 実証実験期間中に，参加者より実証実験から脱退する旨の連絡をセキュアブレインが受けた場合には，当該参加者の実証実験参加同意書および連絡先情報を廃棄します．

A.2.3 連絡先情報の第三者への提供

セキュアブレインは，法令の定める場合を除き，連絡先情報を第三者に提供することはありません．

A.2.4 ブラウザ型センサ取得情報の内容

(1) ブラウザ型センサから大規模分析・対策センタへ送信する情報（以下，「ブラウザ型センサ取得情報」といいます．）は，以下の項目です．（以下では，Internet Explorerを「IE」，Firefoxを「FF」といいます．）

- ブラウザデータ
  ブラウザの種類，バージョン：IEかFFか，またそのバージョン
  プラグインの名称，バージョン：ブラウザの機能を拡張するソフトウェアの名称とバージョン
  ブラウザに設定されているオプション情報：IEの場合，保護モードの有効/無効，ポップアップブロックの有効/無効，SmartScreenフィルターの有効/無効．FFの場合，ダウンロード時のアンチウィルススキャンの有効/無効，右クリック可否，gifアニメーションの表示方法
- Webサイトアクセスデータ
  センサID：ブラウザ型センサが起動するたびに新たに割り当てられる値
  タブID：1つのウィンドウに複数のタブでWebページを表示する場合に，タブごとのアクセスを区別するためにブラウザ側で付与する値
  アクセスURL：ブラウザのアドレスバーに表示されるアクセス先を指し示す文字列
  アクセスURLのハッシュ値：任意の長さの入力値を固定長の出力値に変換するハッシュ値と呼ばれる変換方式を用いてアクセスURLを変換した値．ハッシュ関数としてSHA-256を用いている
  アクセス先IPアドレス：アクセス先のWebページのファイルを持っているWebサーバのIPアドレス
  アクセス日時：アクセスを行った日．YYYY-MM-DD hh:mm:ss形式
  HTTPリクエストの送信時刻：ブラウザからサーバに対して要求を送信した時刻．1970年1月1日0時0分0秒からの秒数で表記
  HTTPレスポンス受信時刻：ブラウザからの要求に対するサーバからの応答の受信開始時刻．1970年1月1日0時0分0秒からの秒数で表記
  HTTPレスポンス受信完了時刻：ブラウザからの要求に対するサーバからの応答の受信完了時刻．1970年1月1日0時0分0秒からの秒数で表記
  センタで悪性判定を行うかどうか：センタでの悪性判定を行うかどうかの設定内容
  コンテンツのサイズ：コンテンツのバイト長
  コンテンツのハッシュ値：コンテンツをハッシュ値に変換した値
  先頭ページであるかどうか：ブラウザで一つのページを表示するために発生する最初のアクセスであるかどうか
  リダイレクトの有無：リダイレクトが発生したかどうか
  リダイレクト元URL：リダイレクトによるアクセスの場合，リダイレクト先へ誘導したリダイレクト元

URL
コンテンツダウンロードの原因となったURL：ブラウザがコンテンツを取得する場合に，コンテンツ取得を誘導したファイルのURL
マウスイベント：アクセスが参加者のマウス動作の結果起こったものか，マウス動作なしにおこったものか
HTTPプロトコルのリクエストヘッダ：ブラウザからサーバに対する要求のうちヘッダ情報．リクエストURI, メソッド, User-Agent, Referer, Accept, Accept-Languageヘッダのみを送信．Authorized（BASIC認証のパスワードなど），Cookie, Set-Cookie（クッキーの情報）は送信しません
HTTPプロトコルのレスポンスヘッダ：ブラウザからの要求に対するサーバの応答のうちヘッダ情報．レスポンスコード，Content-Type，Content-Length，Connection，Locationヘッダのみを送信
ブラウザ型センサによるコンテンツの良悪性判定結果：ブラウザ型センサが持つコンテンツの良悪判定機能による結果

- コンテンツ
PDF：PDFファイル．Content-Typeでapplication/pdfが指定されたもの
Windowsの実行形式のもの：Windows OS上で動作するソフトウェア．HTTPプロトコルのレスポンスヘッダのContent-Typeでapplication/octet-stream, application/x-msdownloadのいずれかが指定されたもの
JavaScript：HTTPプロトコルのレスポンスヘッダのContent-Typeでtext/javascript，text/js，application/javascript，application/x-javascriptのいずれかが指定されたもの
HTML：HTTPプロトコルのレスポンスヘッダのContent-Typeでtext/html, application/xhtml+xmlのいずれかが指定されたもの

(2) 次のWebサイトにブラウザ型センサ取得情報のサンプルを掲載していますので，ご覧下さい．（http://www.fcdbd.jp）
(3) ブラウザ型センサ取得情報は，項目ごとに取得対象から除外するよう参加者が設定することが可能です．
(4) 既定の設定では，HTTPSによるアクセス，プライベートIPアドレスへのアクセスについてはデータ取得を行いません．
(5) データ取得を行わないWebサイトをドメイン名を指定して設定することが可能です．
(6) コンテンツは，悪質コンテンツと疑われるものを取得対象とします．取得対象となった場合は，参加者に許可を求める確認画面を表示し，参加者の許可が得られた場合にのみ取得します．アクセスしたページの構成により，これらファイル群に個人情報が含まれる場合には，当該個人情報も取得されることになります．

A.2.5 ブラウザ型センサ取得情報の取得および保管

(1) ブラウザ型センサ取得情報は，株式会社KDDI研究所（以下,「KDDI研究所」といいます.），セキュアブレインおよび独立行政法人情報通信研究機構（以下,「NICT」といいます.）が取得，分析，保管を行います．
(2) KDDI研究所，セキュアブレインは，ブラウザ型センサ取得情報を，本研究終了（2016年3月末）後も最大1年間研究試料として保有し，ドライブ・バイ・ダウンロード攻撃の対策技術の研究開発のための分析に使用します．当該期間経過後にこのブラウザ型センサ取得情報を破棄します．
(3) NICTは，ブラウザ型センサ取得情報を，本研究終了（2016年3月末）後も最大3年間研究試料として保有し，ドライブ・バイ・ダウンロード攻撃の対策技術の研究開発のための分析に使用します．当該期間経過後にこのブラウザ型センサ取得情報を破棄します．
(4) KDDI研究所，セキュアブレイン，NICTは，参加者の承諾を得ることなく，ブラウザ型センサ取得情報を保有する期間を変更することがあります．その場合には，保有期間の変更を事前もしくは事後速やかにWebサイトへ掲載するとともに，電子メールの送付により参加者に通知するものとします．
(5) KDDI研究所，セキュアブレイン，NICTは，ブラウザ型センサ取得情報を分析して得られた研究成果を報告書，論文，学会等で公開する際には，個人を特定できない態様で，ブラウザ型センサ取得情報を記載します．なお，報告書，論文，学会等で公開した情報については(2)，(3)に定める破棄の対象外とします．

A.2.6 ブラウザ型センサ取得情報と連絡先情報の関連

連絡先情報とブラウザ型センサ取得情報は別々に取得し，関連付けを行うことは一切ありません．

A.2.7 ブラウザ型センサ取得情報を利用した個人の特定・識別

KDDI研究所，セキュアブレインおよびNICTは，ブラウザ型センサ取得情報を利用した個人の特定・識別は一切行いません．

A.2.8 ブラウザ型センサ取得情報の第三者への提供

KDDI研究所，セキュアブレインおよびNICTは，法令の定める場合を除き，ブラウザ型センサ取得情報を第三者に提供することはありません．

松中 隆志 （正会員）

2004 年北陸先端科学技術大学院大学情報科学研究科博士前期課程修了．同年 KDDI（株）入社．（株）KDDI 研究所（現 KDDI 総合研究所）で無線通信，ネットワークセキュリティの研究開発に従事．現在 KDDI（株）．

山田 明 （正会員）

2001 年神戸大学大学院自然科学研究科電気電子工学専攻博士前期課程修了．同年 KDDI（株）入社．2009 年東北大学大学院情報科学研究科博士後期課程修了．2010～2011 年 Carnegie Mellon 大学客員研究員．現在（株）KDDI 総合研究所でネットワークセキュリティの研究開発に従事．

窪田 歩 （正会員）

1995 年京都大学大学院情報工学専攻博士前期課程修了．同年国際電信電話（株）（現 KDDI）入社．2003 年～2004 年米国 California 大学 Berkeley 校客員研究員．現在（株）KDDI 総合研究所でネットワークセキュリティの研究開発に従事．

笠間 貴弘 （正会員）

2014 年 3 月横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程後期修了，博士（工学）．2011 年 4 月より情報通信研究機構に研究員として入所．マルウェア解析やネットワーク攻撃観測・分析等サイバーセキュリティの研究開発に従事．2011 年情報処理学会山下記念研究賞受賞．