ダークウェブ内の違法物品取扱サイトのミドルウェアの特徴に着目した実態調査

新井悠[†1†2]　吉岡克成[†3†4]　松本勉[†3†4]

概要：近年，違法薬物，児童ポルノ，銃などの火器，あるいはサイバー攻撃ツールや攻撃代行サービスといった違法物品ならびにサービスが，ダークウェブ上に構築された仮想取引所などで取引されており，これを利用することにより誰でもそれら違法物品を手に入れることが可能になってきている．このため，法執行機関によってこうした違法物品取扱サイトをテイクダウンする試みが継続的に行われてきている．加えて，ダークウェブをクローリングすることで，それらサイトの分類や，取り扱われている物品の状況などを確認する試みが行われてきている．他方で，ダークウェブのインフラの解析に焦点を置いた研究は少ない．本研究ではダークウェブ上に構築されているこれらの仮想取引所における，インフラを司るミドルウェアに係る実態調査を行い，データを収集した結果について調査・考察する．

キーワード：CSS2019，ダークウェブ，ミドルウェア，Tor

A Study on the Features of Middleware for Illegal Goods Trading Sites in the Tor Hidden Service

YU ARAI[†1]　KATSUNARI YOSHIOKA[†2†3]
TSUTOMU MATSUMOTO[†2†3]

Abstract: In recent years, illegal goods and services such as illegal drugs, child pornography, firearms such as guns, cyber attack tools and cyberattack attack outsourcing services are being traded at virtual exchanges built on the dark web. It has become possible for anyone to obtain such illegal goods. For this reason, law enforcement agencies have been continuously trying to take down such illegal goods trading sites. In addition, by crawling the dark web, attempts have been made to confirm the classification of those sites and the status of the goods being handled. On the other hand, few studies have focused on the analysis of the dark web infrastructure. In this research, we investigate the actual condition of middleware that operates the infrastructure of these virtual exchanges built on the dark web, and then analyze/make consideration for the results of collecting data.

Keywords: CSS2019, Dark Web, Middleware, Tor

1. はじめに

近年，違法薬物，児童ポルノ，あるいはサイバー攻撃ツールや攻撃代行サービスといった違法物品ならびにサービスが，Tor ネットワーク内にホストされる Hidden Service，いわゆるダークウェブ内に構築された Web サイトで取引されている．こうした取引所を利用することで，違法薬物販売者との接点を持つことなく違法薬物を入手することが可能であることや，あるいは特別な情報処理の知識を持たずとも，販売されているツールやサイバー攻撃のアウトソーシングサービスを，仮想資産等を使用して購入することで，サイバー攻撃を実行することが可能となってきている．

†1 トレンドマイクロ(株)
Trend Micro Ltd.
†2 横浜国立大学大学院環境情報学府
Graduate School of Environment and Information Sciences, Yokohama National University
†3 横浜国立大学先端科学高等研究院
Institute of Advanced Sciences, Yokohama National University
†4 横浜国立大学大学院環境情報研究院
Faculty of Environment and Information Sciences, Yokohama National University

このため，海外では 2014 年に FBI や Europol が中心となり「Operation Onymous」を実施し，400 以上のダークウェブ内のサイトを停止させたと明らかにしている[1]. また 2017 年 2 月には，「Freedom Hosting II」と呼ばれるダークウェブ専用のホスティングサービスが何者かによって不正侵入され，その結果 74GB 分のファイルと，2.3GB 分の MySQL データベースが漏洩したという[2]．こうした漏洩の結果，10,000 以上のサイトが閉鎖され，これは当時のダークウェブ全体の 15%から 20%に相当する数であり，さらにその 1 ヶ月後にはダークウェブ全体の 85%に相当するサイトが消失したという[3]．その後，縮小したとはいえダークウェブ内の違法物品取扱サイトは継続的に運営されており，2019 年 5 月には，ふたたび Europol がダークウェブ内の違法物品取扱サイトの運営者を検挙したことを明らかにしている[4]．日本国内においてもこうした違法行為が問題となっており，京都府警が 2018 年の 6 月に児童ポルノサイトをダークウェブ内において運営していた被疑者を検挙している[5]．加えて，2019 年 5 月には，経済産業省の職員がダーク

ウェブ内の違法物品取扱サイトを使用することで，米ロサンゼルスから成田空港に国際郵便で到着したファッション雑誌の袋とじの中に，覚せい剤を入れたものを同年 4 月に受け取っていた容疑で検挙されている[6].

このようにダークウェブで売買されている物品の社会問題が浮上していく一方で，ダークウェブを構成しているソフトウェアは，どのようなミドルウェア等で構成されているのかはよくわかっていない.

本研究では，かかる不明瞭な実態を解明するために，次のような手順を用いて，ダークウェブの違法物品取扱サイトで使用されているミドルウェアの実態調査を行った.

1. ダークウェブで使用されている.onion ドメインを持つ URL を，ダークウェブ内の検索サイト，Wiki ページ等から収集し，クローリング先 URL の初期巡回先リストを作成する
2. 同リストの URL に対してクローリングを行い，HTTP ヘッダのデータを蓄積する
3. 蓄積したデータにアノテーションを実施する
4. 同データの分析を行い，違法物品取引サイトと他のサイトで使用されているミドルウェアの特徴に差異がないかといった観点での，詳細な調査を行う

本論文では，上述の手順により，まず初期巡回先リストとして 5763 サイト分の.onion ドメインを持つ URL を収集した．次に，クローラを使用して初期巡回先リストの URL を巡回し，うち 4340 サイトから HTTP レスポンスデータを得ることができた.

結果の得られた 4340 サイトの HTML データを確認し，同時に，可能であれば当該サイトに実際に接続して目視による確認を行う，アノテーションを行った．アノテーションを行う上で，違法物品取引サイトであると判断したのは，次のようなサイトである.

・違法薬物の取扱
・サイバー攻撃サービスの提供
・重火器の販売
・偽造クレジットカードや偽造身分証明書の販売
・暗号資産ミキサー
・詐欺サイト
・児童ポルノ
・著作権侵害コンテンツの取扱
・その他，犯罪活動に結びついていることが思料されるサイト

なお「その他」には，たとえばマネーロンダリングを企図していると思料される，異常に安価な販売価格のスマートフォンなどの電子ガジェット等の販売サイト等も含まれる.

本論文の実態調査により，以下のような実態が明らかになった.

・クローリング結果の 41%にあたる 1799 サイトが違法物品取扱サイトであった
・ダークウェブの違法物品取扱サイトの HTTP 応答から OS 情報を類推できたサイトでは，およそ 9 割が Linux を使用している
・同サイトでは，1%と僅かではあるが Raspbian を使用していると考えられる
・Web サーバの類推に使用できる HTTP Server ヘッダを応答したダークウェブ内のサイト数は 4075 であり，クローリング結果全体の 93%であった
・ダークウェブの違法物品取扱サイトの HTTP 応答から Web サーバ情報を類推できたサイトでは，およそ 92%が nginx もしくは Apache を使用している

本論文では，2 章で関連研究についてまとめ，3 章でダークウェブのクローリングによるミドルウェアの特徴に着目した実態調査の内容と結果について述べる．さらに 4 章で調査結果の考察について述べ，5 章でまとめと課題について整理する.

2. 関連研究

ダークウェブに対するクローリング

ダークウェブにおける Hidden Service 提供ホストを調査するための手段として，クローリングならびにスクレイピングを行ってデータ収集を行った先行研究として，脆弱性などの脅威情報の収集を目的としたもの[7]や，収集した HTML データの分類を目的としたもの[8]がある．しかし，いずれの研究も，Web ページのテキストに着目したものであり，ダークウェブを構成している各ホストのミドルウェアの特徴等については示されていない.

3. 調査手法

3.1 クローリングのための初期巡回先リストの作成

ダークウェブのクローリングを行うためには，.onion という Hidden Service で使用されている，特有のドメインを持つ URL を事前に収集し，かかる URL に対してアクセスをしなくてはならない．このため，こうしたダークウェブの URL を Hidden Service 専門の検索サイト[9]，およびダークウェブ内に設けられた Wiki ページ等から収集し，クローリング先 URL の初期巡回先リストを作成した．その結果，初期巡回先リストとして 5763 サイト分の.onion ドメインを持つ URL を収集した．その上で，クローラに初期巡回先リストを使用して，クローリングを実施した.

3.2 クローリングの実施

先の手順で作成した初期巡回先リストを使用し，クローリングを行った．クローラーには Python3.6 を使用し，クローリング結果の出力には JSON を用いてファイルに出力した．クローリング結果の JSON 出力例を**図 1** に示す.

```
{
  "headers": {
    "Server": "nginx",
    "Date": "Fri, 14 Jun 2019 01:26:06 GMT",
    "Content-Type": "text/html",
    "Content-Length": "162",
    "Connection": "keep-alive"
  },
  "snapshot": "<html>\r\n<head><title>403 Forbidden</title></head>
r><h1>403 Forbidden</h1></center>\r\n<hr><center>nginx</center>\r\
  "forum": 0
}
```


図 1: クローリング結果の JSON 出力例

それぞれの JSON 要素については表 1 の通りである．なお forum 要素は，後のアノテーションで使用するため，収集時点ではデフォルト値として 0 を設定している．

表 1: JSON 要素の詳細

headers	収集した HTTP ヘッダをオブジェクトで格納
snapshot	収集した HTML ページを文字列で保存
forum	違法物品取扱サイトかどうかのフラグ

クローリングを実施した結果，4340 サイトから HTTP レスポンスデータを得ることができた．なお，Sarah Jamie Lewis の報告[3]によれば，2017 年の 3 月の時点で，ダークウェブ全体のサイト数はおよそ 4400 程度であったという．また、本研究のクローリングを実施したのと同時期に，Fresh Onions[10]と呼ばれる，ダークウェブをクローリングし，その結果を表示することのできるオープンソースツールが運用され，ダークウェブ上に蔵置されていた．その結果によると，同サイトが本研究のクローリングを行ったのと同時期にアクセス可能とリスト表示されたダークウェブのサイト数は 1490 であった．また，OSINT のためのソフトウェアを開発している Hunchly[11]によると，本研究のクローリングを行ったのと同時期にアクセス可能なダークウェブのサイト数は 4584 であった．したがって，Hunchly の収集結果と比較して本研究のクローリング結果は，単純なサイト数の比較として 94.67%と，やや少ないものの，ダークウェブのサイトは頻繁に停止したり，アクセス不能になったりする傾向があるため，ダークウェブのクローリング結果の網羅性としては充分であるという蓋然性は高いと考えられる．

3.3 アノテーションの実施

クローリング結果のデータに対して，目視で違法物品取扱サイトかどうか，アノテーションを行った．アノテーションを行うにあたっては，アノテーション専用 Web アプリケーションを独自に開発し，収集した JSON データをサイト毎にロードして確認したうえで，アノテーションできるようにした．こうした作業を行った結果，ダークウェブのクローリングの結果得られた 4340 サイトのうち，41%にあたる 1799 サイトが違法物品取扱サイトと確認できた．

なお，違法物品取扱サイトではない，クローリング結果の 59%に含まれるサイトはおよそ次のとおりである．

・個人のブログサイト
・反政府的な主張を掲載しているサイト
・新聞社などの報道機関が運営しており，投稿者は Tor を使用することで自身の匿名性を維持したまま，報道機関に対して情報提供ができるサイト
・自作の詩や小説，パロディなどの紹介
・ダークウェブのリンクサイト
・一般的な掲示板サイト
・法執行機関によってテイクダウンされ，閉鎖されたというメッセージが記載されたサイト

3.4 調査結果

3.4.1 全体の傾向

まず全体的な傾向を把握するため，クローリングの結果として得られたデータのうち，各サイトから応答された HTTP ヘッダの行数をもとにした分析を行った．**図 2** は，得られたデータについて，HTTP 応答ヘッダの行数を度数としたヒストグラムである．全体的に，違法物品取扱サイトも，そうでないサイトも似たような傾向にあるが，非取扱サイトのほうは分布が広く，HTTP ヘッダの行数が多いサイトが散見される．また，違法物品取扱サイトの HTTP ヘッダの平均行数は 9.408 であり，非取扱サイトは 9.899 であったことから，違法物品取扱サイトはそうでないサイトと比較して HTTP ヘッダの行数がやや少ない傾向がみられた．

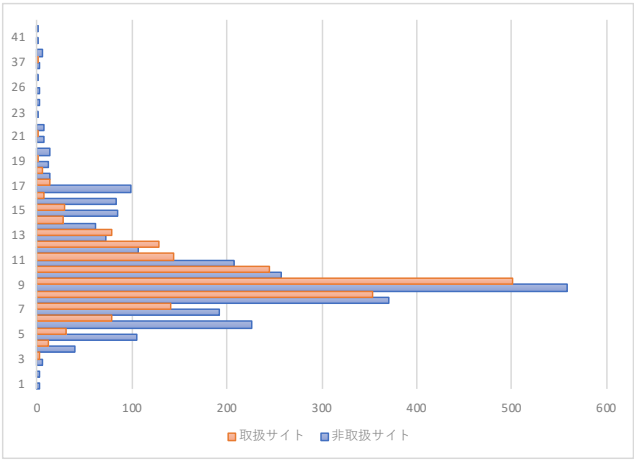

図 2: HTTP 応答ヘッダの行数を度数とした，サイト数のヒストグラム

3.4.2 Server ヘッダによる OS・Web サーバの推定

サーバプログラムに接続し，バナー情報を取得することでサーバプログラムのバージョンを推定し，脆弱性の有無

などを類推する手法は，情報セキュリティにおけるペネトレーションテストなどで恒常的に使用されているものである．なかでも，Web サーバの HTTP 応答に含まれる情報を取得することは，情報セキュリティのガイドライン等でも紹介[12]されており，さらには，情報通信研究機構が実施している，NOTICE プロジェクトにおいてもバナー情報の取得が行われて[13]おり，こうした行為は非常に一般的なものであると解釈されうる．Web サーバの HTTP 応答の具体例を図 3 に示す．

この図のように，Web サーバをデフォルトインストールした場合，HTTP 応答に当該ホストの OS 名称ならびに Web サーバの名称・バージョンが含まれることがある．前述の通り，一般にこうした情報はペネトレーションテストなどで使用されるが，本研究はこの情報を確認することで，ダークウェブのミドルウェアの使用実態を明らかにしていく．

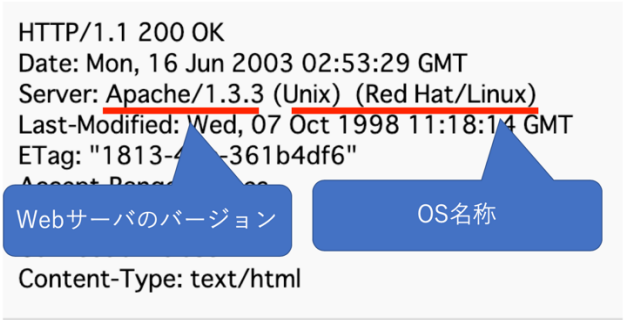

図 3: Web サーバの HTTP 応答に含まれる情報の取得例

OS の類推に使用できる HTTP Server ヘッダを応答したサイト数は 842 であり，クローリング結果全体の 18%であった．その内訳を図 4 に示す．

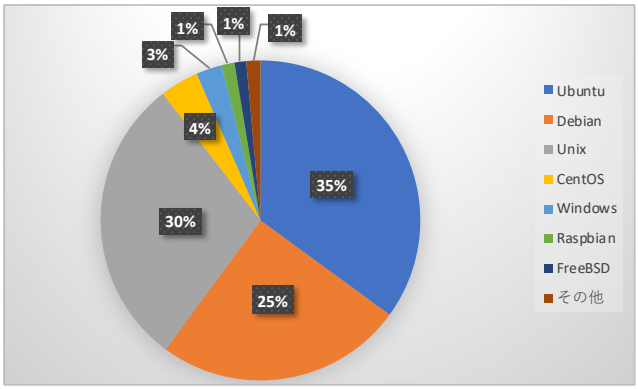

図 4: クローリング結果全体から，OS 情報が確認できた HTTP Server ヘッダの内訳

次に，違法物品取扱サイトにおいて，OS の類推に使用できる HTTP Server ヘッダを応答したサイト数は 275 であり，違法物品取扱サイト全体に占める割合は 15%であった．その内訳を図 5 に示す．このように，Debian と Ubuntu が全体の 8 割以上を占めており，また，わずかながら Raspbian が確認(3 サイト)できた．

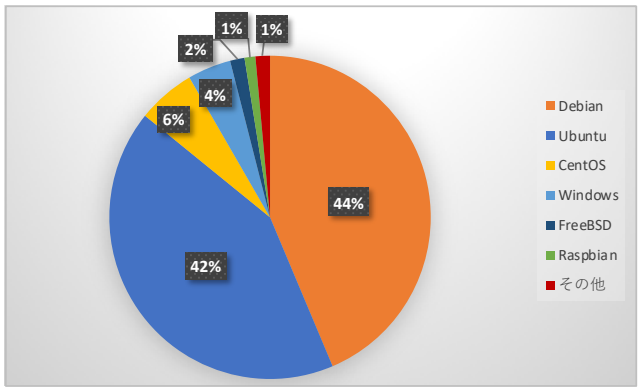

図 5: 違法物品取扱サイト全体から，OS 情報が確認できた HTTP Server ヘッダの内訳

次に，Web サーバの類推に使用できる HTTP Server ヘッダを応答したサイト数は 4075 であり，クローリング結果全体の 93%であった．その内訳を図 6 に示す．なお，図中の「空」は，Server ヘッダは存在するものの，その設定値はブランクであったことを意味する．

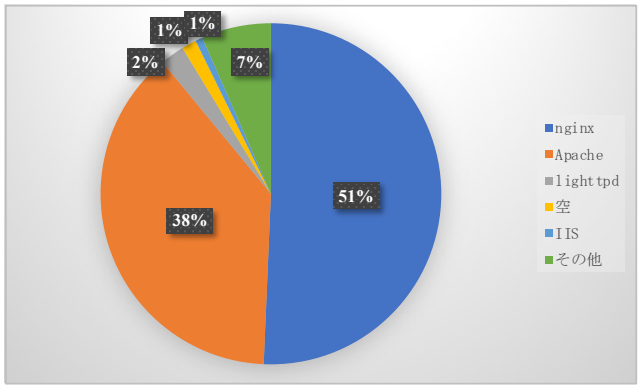

図 6: クローリング結果全体から，Web サーバ情報が確認できた HTTP Server ヘッダの内訳

このように，nginx と Apache が全体のほぼ 9 割を占めていることが確認できた．次に，違法物品取扱サイトにおいて，Web サーバの類推に使用できる HTTP Server ヘッダを応答したサイト数は 1701 であり，違法物品取扱サイトの 94%であった．その内訳を図 7 に示す．

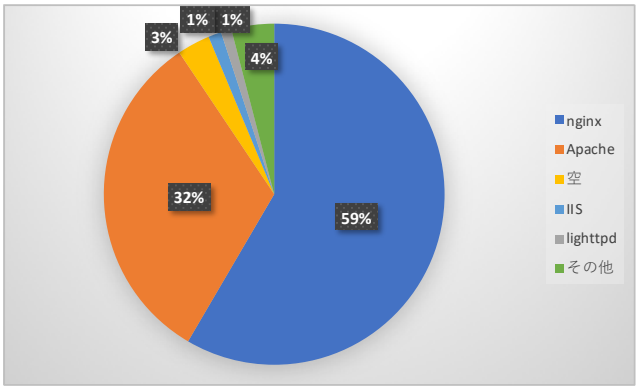

図 7: 違法物品取扱サイト全体から，Web サーバ情報が確認できた HTTP Server ヘッダの内訳

nginx と Apache が全体のほぼ 9 割を占めていることは，クローリング結果全体の傾向とほぼ変わりがなかった．

4. 考察

インターネット(いわゆる表層 Web)で使用されている Web サーバの統計情報を，長期間にわたって記録している Netcraft によれば，2019 年 8 月の時点で表層 Web の Web サーバの使用動向は**図 8** の通りである[14]．このように，nginx と Apache の使用率があわせて 61%あることに加えて，IIS の使用率も 14%あることがわかる．一方で，前章で紹介したダークウェブのクローリング結果では，IIS は全体の 1%であった．

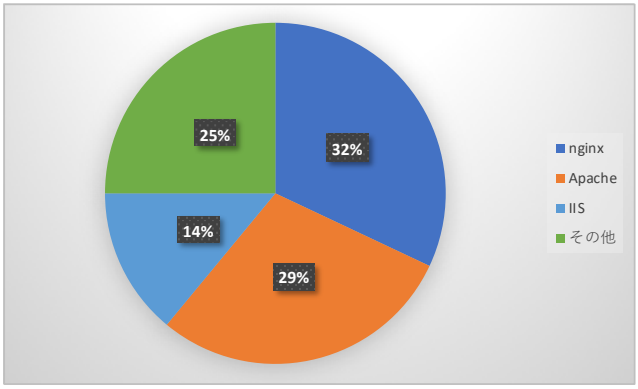

図 8: Netcraft による Web サーバの集計(2019 年 8 月)

このように，表層 Web とダークウェブの Web サーバの使用傾向が異なるのには，ダークウェブのサイト運営者，ひいては違法物品取扱サイトの運営者側に，何らかの理由があると考えるのが自然である．その理由を検討すると，次のような可能性が考えられる．

(1) フィードバック機構による，ホスト情報漏洩の防止

現在，ソフトウェア製品やオンラインサービスには，フィードバック機構が含まれていることがあり，一般的にはオプトインで提供されている．その目的は利用者の動向の把握であり，その結果として，よりよい製品の開発の目的のために収集される．たとえば Windows では，インターネット経由でのライセンス認証の際に「プロダクトキー，ソフトウェア，デバイスに関するデータ」を Microsoft に送信することが明示されている[15]．また，ライセンス認証が行われなければ，修正プログラムを適用することはできない．よってダークウェブのホスト運営者や違法物品取扱サイトの運営者は，ライセンス認証を実施しなければ自らのホストに脆弱性を残したままの状態にしてしまうことになる．一方で，こうしたフィードバック機構によって自らのホスト情報が送信されてしまうことから，ダークウェブのホスト運営者や違法物品取扱サイトの運営者は，そうではない選択肢を選択している可能性がある．

(2) ダークウェブのホスティングサービスの存在

Cloudflare は，コンテンツデリバリーネットワークなどを提供している米国の企業である．同社は Cloudflare Onion Service という，ホスティングサイトの運営者が Tor 利用者に.onion ドメインのホストとしてアクセスできる手段を提供している[16]．このように，正当なホスティング会社がダークウェブのホスティングサービスを提供していることもあるが，ただし利用規約などによって，その利用者は違法物品取扱サイトなどを運営することはできない．その一方で，ダークウェブ内には，ダークウェブのホスティングサービスを提供している事業者も存在する．図 9 に，その一例を示す．このようなホスティングサービスを，暗号資産を使用することで利用できる．そしてこれらのホスティングサービスの多くが，いわゆる LNMP/LAMP 環境であるがゆえに，ダークウェブのクローリング結果に占める，HTTP Server ヘッダからの Web サーバ類推結果に nginx と Apache が全体のほぼ 9 割を占めた結果につながっている可能性もある．

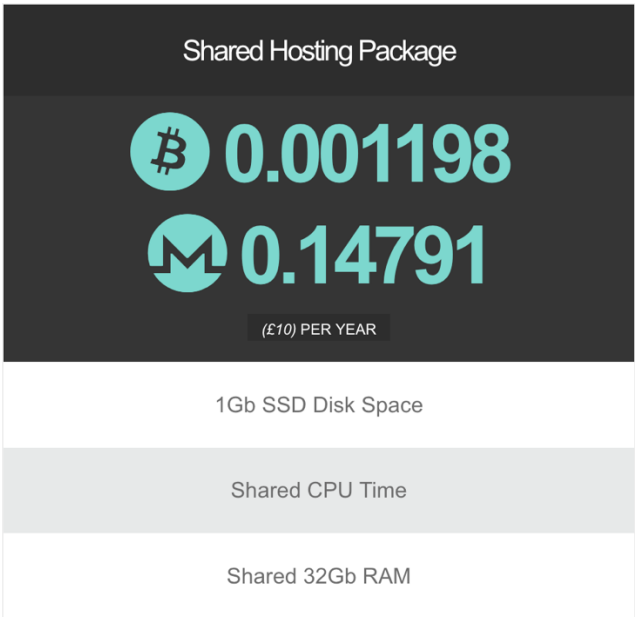

図 9: ダークウェブの特定サイト内で示されたホスティングサービスの提供価格例

5. まとめと今後の課題

本研究では，ダークウェブの違法物品取扱サイトで使用されているミドルウェアの実態調査を行った．調査には，まずダークウェブの URL をダークウェブ内の検索サイト，Wiki ページ等から.onion ドメインをもつ URL を収集し，クローリング先 URL の初期巡回先リストを作成した．そして同リストの URL に対してクローリングを行い，HTTP ヘッダのデータを蓄積した．蓄積したデータにアノテーションを行った上で，同データの分析を行い，調査と考察を行った．調査の結果から，一般的な表層 Web よりも，ダークネットではミドルウェアにはオープンソースのソフトウェアが広く使用されているという実態が判明した．

一方で，今回の調査はダークウェブの HTTP 応答のみに着目した調査であり，クローリング結果の HTML データの調査には着手していない．今後はこの HTML データにも調査をすすめ，より対策となりうるような手段の案出について研究を続けたい．

参考文献

[1] “Operation Onymous | Europol“. https://www.europol.europa.eu/activities-services/europol-in-action/operations/operation-onymous
[2] COMPUTERWORLD. “「闇ウェブ」の利用者データがハッキングで大量流出”. https://project.nikkeibp.co.jp/idg/atcl/idg/14/481542/020800330/
[3] Sarah Jamie Lewis. “OnionScan Report: Freedom Hosting II, A New Map and a New Direction.”. https://mascherari.press/onionscan-report-fhii-a-new-map-and-the-future/
[4] Europol. “DOUBLE BLOW TO DARK WEB MARKETPLACES”. https://www.europol.europa.eu/newsroom/news/double-blow-to-dark-web-marketplaces
[5] 産経新聞. “匿名化ソフト「Ｔｏｒ」使い児童ポルノ公開疑い　京都府警が初摘発”. https://www.sankei.com/west/news/180605/wst1806050108-n1.html
[6] 現代ビジネス. ” 経産省 20 代キャリア官僚「覚せい剤密輸」にちらつくダークウェブの影”. https://gendai.ismedia.jp/articles/-/64579
[7] E. Nunes et al. "Darknet and deepnet mining for proactive cybersecurity threat intelligence", Intelligence and Security Informatics (ISI) 2016 IEEE Conference on. IEEE, pp. 7-12, 2016.
[8] Daniel Moore et al. Cryptopolitik and the darknet. Survival, 58(1):7–38. 2006.
[9] AHMIA. https://ahmia.fi/
[10] Fresh Onions. https://github.com/dirtyfilthy/freshonions-torscraper
[11] Hunchly. https://www.hunch.ly/
[12] OWASP Testing Guide v4. “Fingerprint Web Server (OTG-INFO-002)”. https://www.owasp.org/index.php/Fingerprint_Web_Server_(OTG-INFO-002)
[13] 総務省. “国立研究開発法人情報通信研究機構法（平成 11 年法律第 162 号）附則第８条第２項に規定する業務の実施に関する計画の認可申請の概要”. http://www.soumu.go.jp/main_content/000595925.pdf
[14] Netcraft. “August 2019 Web Server Survey”. https://news.netcraft.com/archives/2019/08/15/august-2019-web-server-survey.html
[15] “Microsoft のプライバシーに関する声明”. https://privacy.microsoft.com/ja-JP/privacystatement
[16] “Introducing the Cloudflare Onion Service”. https://blog.cloudflare.com/cloudflare-onion-service/