

ダークウェブの経年的変化に関する考察

新井 悠^{1,2,3} 吉岡 克成⁴ 松本 勉⁴

概要: 近年, 様々な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いる. 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. 本研究では, 2019 年 6 月と 2020 年 7 月にダークウェブのクロールを行った結果についての詳細な分析と, 比較を行った. その結果, ダークウェブのサイト数がほぼ倍に伸張していた一方で, 2019 年 6 月にアクセス可能であったサイトの 85.1%は消失していたことが判明した. この間に新たに誕生したダークウェブ上のサイトは 8131 ののぼった. 更に調査すると, その内訳として, 児童ポルノサイトと思われるサイトが 30 倍近く増加している可能性があることが判明した.

キーワード: ダークウェブ, クローリング, Tor

A Study on the Chronological Change Of Darkweb

YU ARAI^{1,2,3} KATSUNARI YOSHIOKA⁴
TSUTOMU MATSUMOTO⁴

Abstract: In recent years, various illegal goods and services have been traded on virtual exchanges built on the dark web. Such exchange mechanisms make it possible for anyone to obtain those illegal goods. In this study, we provide a detailed analysis of the results of dark web crawling in June 2019 and July 2020, as well as a comparison. The results show that while the number of sites across the dark web had nearly doubled in growth, 85.1% of the sites that were accessible in June 2019 had disappeared. The number of new sites on the dark web reached 8131 during this period. Further investigation reveals that the number of sites considered to be child pornography sites may have increased nearly 30 times.

Keywords: Darkweb, Crawling, Tor

1. はじめに

近年, 違法薬物や児童ポルノ, あるいはサイバー攻撃ツールやサイバー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, 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. ダークウェブには, 厳密には Tor[1]以外に I2P[2]や Freenet[3]といったオーバーレイネットワークが含まれるが, 本研究ではこうした取引等が活発に行われている Tor の Hidden Service のみを研究の対象としてダークウェブと呼称している. こうしたダークウェブで取引が行われた結果, 違法薬物やサイバー攻撃ツール等によって社会的な問題が引き起こされてしまっている. たとえば, アルコールや薬物に関する世界的な調査である Global Drug Survey の 2016 年の調査[4]によると, オンラインで行われた調査の回答者のうち, インターネットで薬物を購入したと答えた人の 3 分の 2 が平均年齢 28.7 歳の男性であったという. また全回答者のうち 9.3%がダークウェブを通じて MDMA や大麻, 危険ドラッグや LSD といった違法薬物を購入したと回答したという. また, Buskirk ら[5]に

よると, 2016 年にオーストラリアにおいて実際の違法薬物使用者 800 人と対面調査を行った結果, 約 9%にあたる 68 名が過去 1 年間にダークウェブを通じて違法薬物を購入したと答えたという.

こうした違法物品がダークウェブを通じて売買されている実態に対して, 法執行機関も取り締まりを行っている. たとえば 2019 年には[6], ダークウェブを通じた違法物品の販売者および購入者に対して, ヨーロッパ, アメリカ, カナダの法執行機関が連携して捜査を行ったという. その結果, 61 名を検挙し, 299.5 キログラムの違法薬物, 51 丁の銃器, 620 万ユーロ以上 (約 400 万ユーロ相当の暗号通貨, 220 万ユーロの現金, 3 万 5 千ユーロ相当の金) を押収したという. ほかに, 同じく 2019 年に FBI とオハイオ州の法執行機関の共同捜査により, ダークウェブを通じて合成オピオイドやコカインなどの違法薬物を販売し, 総額で 280 万ドル以上の売上を得ていた兄弟が検挙されている[7].

日本においても同様の検挙事例が存在しており, たとえば 2016 年の 3 月に, 13 回にわたってダークウェブ上の掲示

1 (株)NTT データ

NTT DATA Corporation

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 国立研究開発法人情報通信研究機構

National Institute of Information and Communications Technology

4 横浜国立大学大学院環境情報研究院/先端科学高等研究院

Graduate School of Environment and Information Sciences, Yokohama National
University / Institute of Advanced Sciences, Yokohama National University

板に銀行口座の売買を持ちかける書き込みをしたとして、愛知県警が男を検挙している[8]。この男は、実際に約30人に口座を販売して、計約110万円を得ていたという。ほかにも、2019年に京都府警サイバー犯罪対策課と東山署などは児童買春・ポルノ禁止法違反の疑いで、東京都の男を検挙した[9]。ダークウェブ上で会員制の児童ポルノサイトを自ら運営し、その会員らが児童ポルノ画像を繰り返し投稿していたという。

本研究では、このように社会問題化しているダークウェブについて観測を行った上で、調査を行うことによって、ダークウェブの性質や変容を明らかにすることで、こうした社会問題の解決に寄与することを目的とする。かかる目的のために、2019年と2020年にダークウェブのクロールを行い、インデックスページを収集した。その上で、両者の調査と比較を行うことで、経年的な変化の有無や傾向について明らかにした。

2. 関連研究

2.1 ダークウェブのクロールについての関連研究

ダークウェブのクロールを行った先行研究が複数存在している。Guillon は[10]ダークウェブの利用者の分析を行うために、ダークウェブ上の1171サイトを調査し、23のカテゴリに分類した。さらに2165件の投稿を調査し、ダークウェブに非倫理的なサイトが蔓延していることに警鐘を鳴らした。

Moore らによる研究[11]では、ダークウェブのダークウェブのクロールを実施し、5615サイトを巡回して12のカテゴリ分類を行っている。この研究の目的はクロール時のダークウェブのスナップショットを提供することであった。なお、ダークウェブのクロールにおける注意事項として、児童ポルノなどの法令に抵触するデータや、公序良俗に反するデータに不用意にアクセスしてしまう可能性があるため、当該研究の著者らはクロール先の各サイトの文書データのみに限定してクロールを作成している点は、同種の研究を倫理的に行う上で重要であると考えられる。

Fu ら[12]は、ダークウェブ上の掲示板サイトの発見と収集に特化したクロールを開発した。このクロールでは90%以上の確率でフォーラムを発見した。同時に、北米、中東、ラテン語圏を含む109のフォーラムを継続的に監視することに成功している。

これらの先行研究は、分類が目的であり、ダークウェブの性質や変容の解明を目的にはしていない。

2.2 ダークウェブの長期観測についての関連研究

Aldridge らは[13]Silk Road という、かつて存在していたダークウェブの匿名マーケットに対する長期観測を行った。

さらに、同マーケットの利用者の購入履歴などを調査することで、その市場規模を推定した。これによると、2012年半ばの時点での同マーケットの年間市場規模は推定で1440万ドルであった。しかし2013年9月時点での年間市場規模は8970万ドルに増加しており、1年間で6倍以上の伸び率を示したという。

さらには、Soska ら[14]による研究では、2013年から2015年にかけて前述の Silk Road を含む35のダークウェブ上の匿名マーケットの長期観測を行っている。その調査結果によると、2013年時点でのダークウェブの匿名マーケットにおける1日あたりの総取引額は20万ドル程度であった。これに対し、2015年には1日あたり50万ドルから60万ドルの規模に増大し、年間取引高に換算すると1億ドルから1億8000万ドルの規模になっていたという。

このように、ダークウェブ上でなされる取引を確認し、市場規模を推定する目的で匿名マーケットを中心に監視対象を限定し、長期観測を行った先行研究が存在している。しかし、その観測範囲はダークウェブの匿名マーケットのみに限定されており、他の不特定多数のアクセスを想定しているサイトを含むダークウェブの変容などについては調査されていない。

3. 調査手法

ダークウェブのクロールを行うためには、リンク構造をたどっていく Web の手法とは異なり、.onion という、ダークウェブ特有のドメインを持つ URL を事前に収集しておき、かかる URL に対してアクセスをしなくてはならない。本研究では、不特定多数の訪問を前提としている、いわばメインストリームにあるダークウェブのサイトの変化を調査することにした。このため、たとえば暴露型ランサムウェアのグループが企業などから窃取したデータを暴露するサイト[15]や、telegram などを通じて招待された者のみがアクセスできるサイトといった、不特定多数の訪問を前提としていないサイトは対象としていない。2019年6月12日に、こうした不特定多数の訪問を前提としているダークウェブの URL を、ダークウェブ専門の登録型検索エンジン ahmia[16]から収集し、クロール先 URL の初期巡回先リストを作成した。その結果、2019年には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。2020年7月14日においても同一のサイトで11980サイト分の.onion ドメインを持つ URL を収集した。他方で、OSINT のためのソフトウェアを開発している Hunchly[17]によると、本研究のクロール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、2019年は4584であり、2020年は8200であった。Hunchly のサイト数よりも ahmia の網羅しているサイト数が多いこと、さらに、サンプリングした結果では Hunchly のクロール結果には

ダークウェブの登録型
検索エンジン

巡回先リスト

巡回結果データ

データ生成
クローラ

```
graph LR; A[ダークウェブの登録型  
検索エンジン] --> B[巡回先リスト]; B --> C[巡回結果データ]; C --> D[データ生成  
クローラ];
```

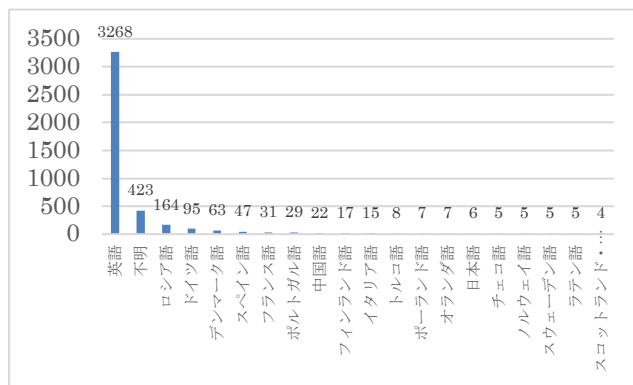
```
{
  "headers": {
    "Date": "Sun, 19 Jul 2020 04:43:30 GMT",
    "Content-Type": "text/html",
    "Content-Length": "162",
    "Server": "nginx",
    "Connection": "keep-alive"
  },
  "snapshot": "<html>\r\n<head><title>403 Forbidden</title></head>\r\n<body bg\n\ncolor='white'>\r\n<center>ch1403 Forbidden</h1></center>\r\n<hr><center>nginx\n</center>\r\n</body>\r\n</html>\r\n"
```

クローリング結果は図 2 のように JSON 形式で保存した。なお、前項の先行研究における注意事項を踏襲し、法令に抵触する可能性のある画像等の保存はせず、テキスト情報のみを保存した。なお、ページ中のリンク構造をたどってさらに別のページを発見するといったことは行わず、インデックスページのみを保存対象とした。その事由としては、インデックスページが当該サイトの性質を最も示していると仮定したためである。

4.1 ダークウェブのサイトの消失と誕生の状況確認

トが消失しやすい性質があるかどうかを確認するためである。この確認を行った結果、2020 年も応答が確認できたサイト数は 4312 サイト中 644 であった。したがって、2019 年 6 月にアクセスが確認できたサイトのうち 85.1% は 2020 年 7 月までに消失したことになる。他方で、2020 年にクロールできたダークウェブのサイト数は 8775 サイトあったということは、8131 サイトは、この間に新たに生まれたということになる。

次に、クロール結果を使用して各サイトで使用されている言語の判定を行った。言語の判定には客観的な判断を下すために Python の polyglot モジュール[18]を使用した。polyglot モジュールは、文書の言語検出のほか、形態素解析や感情分析などにも使用できる自然言語処理モジュールである。2019 年と 2020 年のダークウェブのクロール結果を使用し、各サイトで使用されている言語を推定した結果をそれぞれ図 3 と図 4 に示す。

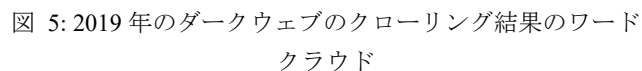


Language	Number of Speakers
英語	8598
ロシア語	65
不明	32
ドイツ語	17
デンマーク語	13
フランス語	13
スペイン語	6
イタリア語	5
マルタ語	5
韓国語	4
トルコ語	3
オランダ語	3
ポルトガル語	2
ポーランド語	2
スウェーデン語	1
フィンランド語	1
ウクライナ語	1
ソマリ語	1
ノルウェー語	1
日本語	1
ロシア語	1

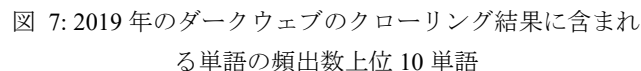
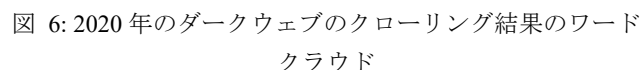
2019年のダークウェブのクローリング結果では、英語サイトが全体の75%を占めた。ついで言語判定できないものが423サイト存在した。そこで、言語判定できないものの

一方、2020 年のダークウェブのクロール結果では、英語サイトが全体の 97%を占めた。ついでロシア語が 0.7%を占めており、その次が不明であった。不明にカテゴリ化されたサイトをいくつかサンプリングしてみた限りでは、やはりリダイレクトのみを行うサイトや、画像のみで構成された、ホスティングの宣伝サイトが内訳として確認できた。なお、日本語のサイトは 1 サイトのみに減少していた。

さらに 2019 年と 2020 年のダークウェブの概況を把握・比較するために、ワードクラウドによる可視化を行った。ワードクラウドとは、文章中で出現頻度が高い単語を複数選び出し、その頻度に応じた大きさで図示する手法のことである。Web ページ等で頻出する単語を自動的に並べることで、テキストデータを可視化することが可能である。今回は、それぞれのクロール結果全体にワードクラウドを適用することで、不特定多数の訪問を前提にしているダークウェブの変化がこの 1 年間に発生していないか、あるいは傾向などはないか、といったことを可視化した。なお、ワードクラウドの作成には、Python の wordcloud モジュール[19]を使用した。同時に、ワードクラウドの対象とするクロール結果データからは、HTML タグなどは削除して、テキスト情報のみを対象とした。図 5 に 2019 年のクロール結果のワードクラウドを示し、同じく図 6 に 2020 年のクロール結果のワードクラウドを示す。



は、'Credit'、'Bitcoin'、'Card'、'Account'、'Market'といった単語が目立っている。その理由としては、ダークウェブの匿名マーケットで売買に使用されている'Bitcoin'が頻出であったことや、同じく匿名マーケットで販売されているクレジットカードの情報が頻出した可能性がある。同様に、'Account'や'Market'は匿名マーケット内のアカウント登録やマーケットそのものを指す単語が頻出したものと考えられる。



- 47 -

かる。しかし、たとえば‘porno’の頻出数は 42189 件であり、あるいは‘porn’の頻出数は 30217 件であった。これらは巡回して結果を得られたサイト数の 8775 よりも多い。すなわち、2020 年のクローリング結果には、特定のキーワードを大量に含むサイトが存在し、そのサイトでのみ大量に使用されている単語が、結果としてクローリング結果に占める割合を大きく変化させている可能性もある。

そこで、次の 2 つの条件を同時に満たすサイトを確認することにした。

(ア) 大文字小文字を区別せず単語‘child’もしくは‘childs’をテキストに含む

(イ) 大文字小文字を区別せず‘porn’もしくは‘porno’をテキストに含む

すると、2019 年のクローリング結果のうち、(ア)と(イ)の両方の条件を満たすサイト数は 69 あった。他方で、2020 年のクローリング結果のうち、(ア)と(イ)の両方の条件を満たすサイト数は 2114 サイト存在した。したがって 2019 年から 2020 年にかけてこうしたサイトが 30.63 倍に増加していることが判明した。よってこうした児童ポルノサイトを扱っていると思われるサイトの増加そのものが、特定の単語の頻出数に寄与している蓋然性が高いことが確認できた。

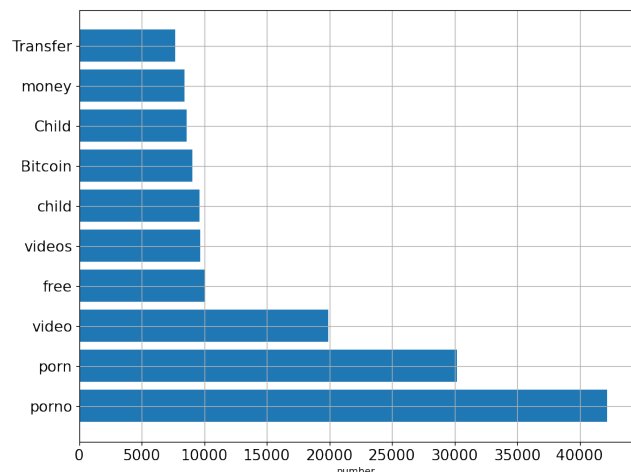


図 8: 2020 年のダークウェブのクローリング結果に含まれる単語の頻出数上位 10 単語

5. まとめと考察

本研究では、ダークウェブの経年的な変化や傾向について調査するためにクローリングを行った。2019 年 6 月と 2020 年 7 月にダークウェブのクローリングを行い、それぞれ 4312 サイトと 8775 サイトから応答が得られた。2019 年に確認されたサイトは、2020 年には 85.1%が消失していた一方で、8131 サイトが新たに誕生していた。また、使用さ

れている言語を調べたところ、2019 年の結果では 75%が英語を使用しており、2020 年の結果では 97%が英語であった。さらにダークウェブにおける単語の頻度を調査したところ、2019 年は匿名マーケットに関連した語句が頻出した。一方で、2020 年の結果では児童ポルノに関連した語句が頻出した。この傾向は、ダークウェブにおける児童ポルノサイトの増加を示唆させ、実際に大文字小文字を区別せず単語‘child’もしくは‘childs’をテキストに含み、かつ、大文字小文字を区別せず‘porn’もしくは‘porno’をテキストに含むサイトが 2114 サイト存在した。

法執行機関のこの 1 年間の活動をみると、たとえば 2019 年 10 月には、法執行機関の国際的な連携により、史上最大規模のダークウェブ上の児童ポルノサイトの運営者を含む、38 ヶ国の 337 名の被疑者が検挙された[20]。また、2020 年にはオランダ警察を中心とした法執行機関の国際連携により、ダークウェブと表層ウェブの両方で児童ポルノサイトを運営していた被疑者を検挙している[21]。ほかにも、2019 年 11 月にダークウェブを経由して児童ポルノサイトを入手していた男が検挙された[22]事例や、2020 年 6 月に、ダークウェブを通じて児童ポルノを販売していた男が検挙された事例がある[23]。このような大規模な検挙事案や、利用者の検挙など法執行機関が継続的に摘発を行っているにも関わらず、児童ポルノサイトの拡大がダークウェブ上で起こっているということであれば、市場原理などの因果関係を明らかにするなど、さらなる研究の試みを行いたい。

参考文献

- [1] “Tor Project” . <https://www.torproject.org/>, (参照 2020-08-02)
- [2] “I2P 匿名ネットワーク” . <https://geti2p.net/ja/>, (参照 2020-08-02)
- [3] “Freenet” . <https://freenetproject.org/index.php?page=index>, (参照 2020-08-02)
- [4] “Key findings from the Global Drug Survey 2016” . <https://www.globaldrugsurvey.com/wp-content/uploads/2016/06/TASTER-KEY-FINDINGS-FROM-GDS2016.pdf>, (参照 2020-08-02)
- [5] VAN BUSKIRK, Joe, et al. Characterising dark net marketplace purchasers in a sample of regular psychostimulant users. *International Journal of Drug Policy*, 2016, 35: 32-37.
- [6] “GLOBAL LAW ENFORCEMENT ACTION AGAINST VENDORS AND BUYERS ON THE DARK WEB” . <https://www.europol.europa.eu/newsroom/news/global-law-enforcement-action-against-vendors-and-buyers-dark-web>, (参照 2020-08-02)
- [7] “Brothers Built A \$2.8 Million Dark Web Drug Empire With Help From Mom’s Amazon, FBI Claims” . <https://www.forbes.com/sites/thomasbrewster/2020/01/16/brothers-used-moms-amazon-to-help-build-28-million-dark-web-drug-empire-fbi-claims/>, (参照 2020-08-02)
- [8] “ダークウェブで銀行口座売買か” . <http://archive.is/dWSz2>, (参照 2020-08-02)
- [9] “漫画家を逮捕、児童ポルノサイト開設容疑 「ダークウェブ」に画像投稿” . <https://www.kyoto-np.co.jp/articles/-/72666>, (参照 2020-08-02)
- [10] Guitton, Clement. (2013). A review of the available content on

- Tor hidden services: The case against further development.
Computers in Human Behavior. 29. 2805-2815.
10.1016/j.chb.2013.07.031.
- [11] MOORE, Daniel; RID, Thomas. Cryptopolitik and the Darknet. Survival, 2016, 58.1: 7-38.
 - [12] FU, Tianjun; ABBASI, Ahmed; CHEN, Hsinchun. A focused crawler for Dark Web forums. Journal of the American Society for Information Science and Technology, 2010, 61.6: 1213-1231.
 - [13] ALDRIDGE, Judith; DÉCARY-HÉTU, David. Not an'Ebay for Drugs': the Cryptomarket'Silk Road'as a paradigm shifting criminal innovation. Available at SSRN 2436643, 2014.
 - [14] SOSKA, Kyle; CHRISTIN, Nicolas. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In: 24th {USENIX} security symposium ({USENIX} security 15). 2015. p. 33-48.
 - [15] “Here's a list of all the ransomware gangs who will steal and leak your data if you don't pay” .
<https://www.zdnet.com/article/heres-a-list-of-all-the-ransomware-gangs-who-will-steal-and-leak-your-data-if-you-dont-pay/>, (参照 2020-08-06)
 - [16] “AHMIA” . <https://ahmia.fi/>, (参照 2020-08-06)
 - [17] “Hunchly” . <https://www.hunch.ly/>, (参照 2020-08-06)
 - [18] “polyglot 16.7.4” . <https://pypi.org/project/polyglot/>, (参照 2020-08-06)
 - [19] “wordcloud 1.7.0” . <https://pypi.org/project/wordcloud/>, (参照 2020-08-06)
 - [20] “South Korean National and Hundreds of Others Charged Worldwide in the Takedown of the Largest Darknet Child Pornography Website, Which was Funded by Bitcoin”.
<https://www.justice.gov/opa/pr/south-korean-national-and-hundreds-others-charged-worldwide-takedown-largest-darknet-child>, (参照 2020-08-06)
 - [21] “DARK WEB CHILD ABUSE: ADMINISTRATOR OF DARKSCANDALS ARRESTED IN THE NETHERLANDS” .
<https://www.europol.europa.eu/newsroom/news/dark-web-child-abuse-administrator-of-darkscandals-arrested-in-netherlands>, (参照 2020-08-06)
 - [22] “Man charged with 32 counts of child porn; tells deputies he got videos from ‘dark web’ ” .
<http://veronews.com/2019/11/20/man-charged-with-32-counts-of-child-porn-tells-deputies-he-got-videos-from-dark-web/>, (参照 2020-08-06)
 - [23] “California man ran a photography studio for years. He was just charged with massive child porn distribution” .
<https://www.mercurynews.com/2020/06/03/ca-man-ran-a-photography-studio-for-years-he-was-just-charged-with-massive-child-porn-distribution/>, (参照 2020-08-06)