情報処理学会論文誌 Vol.66 No.3 513–522 (Mar. 2025)

Onionサイトの HTTPレスポンスを用いたダークウェブの大規模分析

木村 悠生[1,a)] 穐山 空道[2,b)] 猪俣 敦夫[3,c)] 上原 哲太郎[2,d)]

受付日 2024年5月20日, 採録日 2024年12月10日

概要：匿名ネットワーク技術の実装の中で多用されるものとして，Tor（The Onion Router）がある．Torの Onion Service を用いることで，IP アドレスなどの運用元特定につながる情報を秘匿しながら Web サイトが運用可能である．Onion Service は，違法な Web サイトの運営に用いられることがあるため，Onion Service を用いて運用されている Web サイトの内容についての分析は多くの研究で行われているが，メタデータについての大規模な分析を行った研究は少ない．本研究では，Onion Domain を大規模に収集し，それらの Onion Domain にリクエストして得られた HTTP レスポンスを分析することで，Onion サイトの特徴分布を明らかにすることを試みた．結果として，全体の約 4 分の 1 にあたる約 20 万件の Onion Domain を収集し，それらの HTTP レスポンスには，サーフェスウェブとは明らかに異なる分布があることが分かった．さらに，少数の運用元がバーチャルホスト機能を用いて，大量の Onion サイトをホストしている可能性が高いことが明らかになった．

キーワード：Tor，Onion Service，Onion Domain，HTTP レスポンスヘッダ

Large-scale Analysis of the Dark Web Using HTTP Responses from Onion Sites

Yuuki Kimura[1,a)] Soramichi Akiyama[2,b)] Atsuo Inomata[3,c)] Tetsutaro Uehara[2,d)]

Received: May 20, 2024, Accepted: December 10, 2024

Abstract: Tor (The Onion Router) is an anonymous network technology that is widely used in the implementation of anonymous networks. Using Tor's Onion Service, websites can be operated while keeping IP addresses and other information that can lead to the identification of the operating source secret. Since the Onion Service is sometimes used to operate illegal Web services, many studies have analyzed the contents of Web sites operated using the Onion Service. However, few studies have conducted large-scale analysis of metadata. In this study, we attempted to clarify the feature distribution of Onion Service by collecting Onion Domain on a large scale and analyzing HTTP responses. As a result, we collected approximately 200,000 Onion Domains, or about a quarter of the total, and found that their HTTP responses had a distinctly different distribution from the surface web. Farthermore, it is suggested that a small number of operators may be hosting a large number of Onion sites using the virtual host function.

Keywords: Tor, Onion Service, Onion Domain, HTTP Response Header

1. はじめに

インターネットではその普及にともない，様々なサービ

[1] 立命館大学大学院情報理工学研究科
Graduate School of Information Science and Engineering, Ritsumeikan University, Ibaraki, Osaka 567–8570, Japan
[2] 立命館大学情報理工学部
College of Information Science and Engineering, Ritsumeikan University, Ibaraki, Osaka 567–8570, Japan
[3] 大阪大学 D3 センター
D3 Center, The University of Osaka, Ibaraki, Osaka 567–0047, Japan

[a)] ykimura@cysec.cs.ritsumei.ac.jp
[b)] s-akym@fc.ritsumei.ac.jp
[c)] inomata.atsuo.cysec@osaka-u.ac.jp
[d)] t-uehara@fc.ritsumei.ac.jp

スが提供されている．その中でも Web サイトは SNS に次いで利用されており，インターネットの中核を担っている [1]．一般的な Web サイトは世界各国の法令を遵守し運営されているが，一部の Web サイトは，著作権を侵害しているコンテンツの掲載や，不法に入手した個人情報などの違法な物品の売買，法令に抵触するアダルトコンテンツの掲載など，違法性を帯びながら運用されている場合がある．違法な Web サイトの摘発には Web サイトの運用元の特定が必要となるが，IP アドレスを秘匿したりドメイン名に運用者自身の情報を紐づけないことで，Web サイトの運用元を特定することを困難にできる．

IP アドレスを秘匿しながら Web サイトを運営する手法として，Tor [2] の Onion Service [3] がある．Onion Service を用いるとき，Web サイトの運用者は ".onion" で終わるドメイン（以下，Onion Domain）を取得し，その Onion Domain を用いて Web サイトを運用する．Onion Domain は Onion Service の公開鍵をもとに生成されるため，サーフェスウェブのドメインと異なり，運用元情報と紐づかない．Onion Service では，IP アドレスを含めた多くの情報が秘匿されているが，HTTP レスポンスヘッダのように秘匿されていない情報も存在する．我々は以前の研究で，Onion サイトとサーフェスウェブのサイトとの間で HTTP レスポンスヘッダを比較することで，Onion サイトの運用元を特定できる場合があることを示した [4]．Onion Service の運用状況は Tor Metrics [5] で公開されているが，Onion Domain の総数などの概観のみであり，個々の Onion サイトの情報は明らかでない．また，Onion サイトの HTTP レスポンスのメタデータについての大規模な分析を行った研究も少ない．

そこで，本研究では，Onion サイトのメタデータに着目し，大規模な調査分析を行う．Onion サイトの調査のためには，Onion Domain を知る必要がある．先行研究 [6], [7] では現在ではポリシーで禁止されている Hidden Service Directry の脆弱性を用いて Onion Domain を収集しているが，本研究では，Tor のポリシーに違反せずに Onion Domain を大規模に収集する．

本研究では，複数の Onion サイトを少数の運用者が運用している可能性に着目する．大規模な調査分析によって，Onion Domain の数にとらわれず，実際の Onion サイト運用者数などを推定することに寄与し，Onion サイトの特徴傾向をより明らかにすることを目指す．また，レスポンスヘッダに着目することで，複数の Onion サイト間に共通の項目を発見し，複数 Onion サイトを運営している運用者が存在する可能性を調査する．

Onion Domain の収集の結果，調査期間に存在した Onion Domain の約 25%にあたる約 20 万個の Onion Domain を収集した．分析の結果，ダークウェブにおけるメタデータの分布は，サーフェスウェブとは異なり，大きな偏りがあることが判明した．また，メタデータから，複数の Onion サイトについて，少数の運用者が大量の Onion サイトを運用している傾向があることが示唆された．

2. 研究背景

2.1 Onion Service

Onion Service は，Tor のネットワークを利用することで，サーバの IP アドレスを秘匿したまま TCP 接続のサービスを提供することができる，Tor の機能である．サーバとクライアントは Tor ネットワークを介することで，お互いの IP アドレスを秘匿したまま通信することができる．Onion Service の運用数などの統計情報は，Tor Metrics [5] で公開されている．

Onion Service はその性質上，サーバの IP アドレスを秘匿できる．しかし，レスポンスに含まれる情報は運用者が意図的に秘匿しない限り，秘匿されない．そのため Onion Service では，レスポンスヘッダは秘匿せず，匿名化されていない固有の情報が含まれる場合がある．

2.2 レスポンスヘッダ

HTTP レスポンスにおけるヘッダフィールドは，HTTP レスポンスのメタデータを表す．各ヘッダは，ボディに関する情報を含むエンティティヘッダと，レスポンスの状態やレスポンスを提供するサーバに関する情報およびキャッシングポリシーなどを含むレスポンスヘッダに分類される．ヘッダは，ヘッダ名とそれに続くコロン（:），値で構成される．ヘッダ名は，英数字とハイフン（-）から構成され，大文字小文字を区別しない．

本研究では，レスポンスにおけるヘッダフィールドを特に区別するため，"レスポンスヘッダフィールド" と呼ぶ．また，レスポンスヘッダフィールドに含まれるヘッダをすべて "レスポンスヘッダ" と呼び，特にエンティティヘッダを区別する必要がある場合は "レスポンスエンティティヘッダ" と呼ぶ．さらに，本研究では，レスポンスヘッダの件数のことを "行数" と呼び，単位を "行" とする．本研究における各語の指す要素を，**図 1** に示す．

本研究では，レスポンスヘッダのうち行数，レスポンスヘッダの種類，Server ヘッダの値，加えてレスポンスボディのうち HTML の `<title>` タグの内容をメタデータと定義する．**表 1** に，本研究で用いるメタデータの一覧を示す．

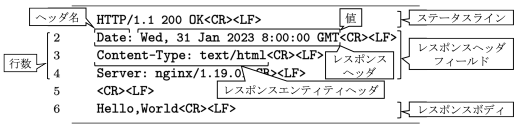


図 1 HTTP レスポンスの例における各語の指す要素

Fig. 1 Elements referred to in the example of an HTTP response.

表 1 本研究で用いるメタデータの一覧
Table 1 List of Metadata Used in This Study.

	メタデータ	例
レスポンスヘッダ	行数	6
	レスポンスヘッダの種類	Date Content-Type Server Connection Vary
	Server ヘッダの値	Apache
レスポンスボディ	取得した HTML の <title>タグの内容	Welcome to nginx!

2.3 名前ベースのバーチャルホスト

名前ベースのバーチャルホストは，1 つのサーバで複数のドメインを運用する技術および機能である．名前ベースのバーチャルホスト機能を用いることで，単一の運用元が 1 つのサーバで複数の Web サイトを運用することができ，クライアントがリクエストしたホスト名に応じて，異なるコンテンツを返すことができる．

サーフェスウェブでは，2 つ以上のホストが同一の IP アドレスを持つ場合，名前ベースのバーチャルホスト機能を用いて複数の Web サイトを運用している可能性がある．このことから，2 つ以上のホストが同一の運用元によって運用されていると容易に推測できる．しかし Onion サイトでは，Onion Domain から IP アドレスを特定することができず，同一のサーバで運用されているか否かを判断することは難しい．

3. 関連研究

杉生ら [8] は，Onion サイトのリンクによって形成されるグラフを作成し，トップページのテキストからカテゴリ分類を行った．グラフをコンテンツの属性ごとに着色することで，ハイパーリンク関係とコンテンツ内容に関連性があることが可視化された．

新井ら [9], [10] は，Onion サイトをクロールし，レスポンスヘッダを用いて違法物品取扱サイトか否かを判定する学習器を構築した．新井らは 2019 年 7 月に Onion サイトの URL をクロールし，全 Onion Domain の 94%にあたる 4,340 件の Onion Domain を取得したとしている．しかし Tor Metrics [5] によると，2019 年 7 月に存在した Onion Domain は約 72,000 件であるとされている．そのため新井らが取得した Onion Domain は全体の 6%にすぎない可能性がある．

Pastor-Galindo ら [7] は，2021 年までに Onion サイトのドメイン収集を行った研究をまとめた．しかし，Onion サイトのドメイン収集数が最大でも 8 万件程度止まりである点や，規約で禁止されている Hidden Service Directry の脆弱性を用いた収集手法を用いている点，またいずれの研究においても，旧バージョンである V2 Onion のドメインを収集している点などが課題としてあげられる．

Shodan [11] は，インターネット上の機器の情報を収集し，検索可能にするサービスである．Shodan は，インターネット上の機器の情報を収集するため，インターネット上の機器に対してスキャンを行い，レスポンスヘッダフィールドおよび Web サイトのタイトルを記録する．しかし，Shodan はサーフェスウェブのみを対象としているため，Onion サイトを対象としていない．

我々 [4], [12] は，2023 年 1 月に Onion サイトをクロールし，39,118 件の Onion Domain を取得した．さらに収集した Onion Domain にリクエストを行い，メタデータを比較することでサーフェスウェブのサイトと Onion サイトを紐づけることに成功した．しかし Tor Metrics によると，当該期間には約 760,000 件の Onion Domain が存在したとされている．よって我々が収集した Onion Domain は全体の 5%にすぎない可能性がある．そこで本研究では，先行研究の Onion Domain 収集数を上回る Onion Domain を収集する手法を述べ，それらの Onion Domain で運用されている Web サイトを対象とした分析およびその結果について示す．

4. 調査手法

本研究では，Onion Domain の収集と，収集した Onion Domain から得られる HTTP レスポンスの傾向調査に分けて調査を行う．

4.1 Onion Domain の収集

本節では，我々の以前の研究 [4] と同様の手法を用いている．

Onion Domain を収集するために，Onion サイトの URL のクロールを行う．クロールには初期巡回リストを作成し，初期巡回リストから幅優先探索で各 URL にアクセスを行い，HTTP レスポンスボディを取得する．そのうえで，レスポンスボディより “.onion” を含む URL を収集する．また，収集した URL から “.onion” を含む Onion Domain を収集し，記録する．収集した URL についてアクセスを行い，以上の動作を繰り返しクロールを行う．クロールを行う順序を図 **2** に示す．

本調査では高速化のために，特にリクエストを最大で 500 スレッドに並列化し，我々の以前の研究からスレッド数を増加させてクロールを実施する．

4.2 HTTP レスポンスの傾向調査

本研究では，収集した Onion Domain について，Tor を経由した HTTP/1.1 [13] リクエストを行い，得られた HTTP レスポンスの傾向調査を行う．

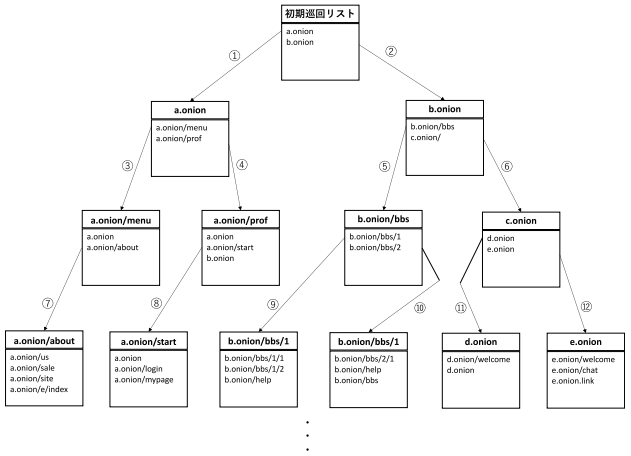


図 2 Onion サイトに対する幅優先探索の巡回順
Fig. 2 Circuit order of breadth-first search for Onion sites.

レスポンスヘッダフィールドに着目すると，レスポンスヘッダの行数や，レスポンスヘッダの種類などが特徴としてあげられる．また，Server ヘッダの値には，サーバの詳細な情報が含まれる場合がある．そこで，本研究では，レスポンスヘッダの行数や，レスポンスヘッダの種類，Server ヘッダの値に着目し，Onion サイトの特徴を調査する．

レスポンスボディに着目すると，Onion サイトのコンテンツの特徴を調査することができる．本研究では特に，HTML の `<title>` タグの内容に着目し，Onion サイトの特徴を調査する．`<title>` 以外のコンテンツは自動で生成されたり，時間によって変化したりするため，ハッシュ値をとって一致を確認したり，類似度を求めることは困難である一方，`<title>` タグの内容は，Web ページの内容を表す要素であるため，変化が少なく内容比較に適していると考えられる．

リクエストに工夫をすることで，通常とは異なるレスポンスを得ることができる場合がある．具体的には，HTTP/1.0 [14] でリクエストを行うと，Host 名を指定できないため，バーチャルホスト機能が有効になっているサーバではデフォルトのコンテンツを返す場合がある．そこで，本研究では，HTTP/1.0 でリクエストを行った結果と，HTTP/1.1 でリクエストを行った結果の違いに着目し，Onion サイトの特徴を調査する．

5. 調査結果

5.1 Onion Domain の収集

4.1 節の手法に基づいて，収集を行った結果を**表 2** に示す．ただし，表 2 でいう発見した URL，調査した URL とはそれぞれ，HTTP レスポンスボディから発見された URL と，クローラが実際にアクセスを行った URL を指す．

表 2 に示すとおり，9 日間で約 20 万件の Onion Domain を収集した．なお，収集では時間およびシステムリソースの制約があり，発見した URL すべてを調査することは現実的でないため，発見した URL のうち，未調査の Onion Domain がなくなった段階で収集を終了した．Tor Project の報告 [5] によると，収集を行った 2023 年 12 月 9 日から 2023 年 12 月 17 日では，Onion Domain は平均で 812,340 件存在したとされている．そのため本研究の収集は，当該期間における全 Onion Domain の 25.13%の Onion Domain を収集したといえる．

表 2 Onion Domain の収集結果概要
Table 2 Summary of Onion Domain acquisition results.

調査期間	2023/12/09-2023/12/17（9 日間）
発見した URL	18,122,686 件
調査した URL	1,332,347 件
発見した Onion Domain	204,173 件

表 3 利用率が 1%以上のレスポンスヘッダ
Table 3 Response headers with a usage rate of 1% or more.

ヘッダ名	件数	利用率
Date	197,508	99.99%
Content-Type*	197,275	99.87%
Server	193,579	98.00%
Connection	184,242	93.27%
Vary	182,715	92.50%
Etag*	9,078	4.60%
Set-Cookie	8,488	4.30%
Last-Modified*	6,192	3.13%
Content-Length*	5,422	2.74%
Cache-Control	5,104	2.58%
X-Frame-Options	3,405	1.72%
Expires*	3,206	1.62%
X-Content-Type-Options	3,076	1.56%
Pragma	2,872	1.45%
Accept-Ranges	2,458	1.24%
Referrer-Policy	2,246	1.14%

5.2 HTTP レスポンスの傾向調査

5.2.1 レスポンスヘッダ名と利用率

実験の結果，調査対象とした Onion サイトより得られた全レスポンス中から，合計 289 種類のレスポンスヘッダを発見した．それらのうち，調査対象 Onion Domain における利用率が 1%以上のレスポンスヘッダを，**表 3** に示す．ただし，ヘッダ名の末尾に “*” が付与されているものは，レスポンスエンティティヘッダである．

また，発見したレスポンスヘッダのうち，101 種類が 1 つの Onion サイトでしか発見されなかった．すなわち，約 35%のレスポンスヘッダが 1 つの Onion サイトでしか発見されなかったといえる．

5.2.2 レスポンスヘッダの行数

各レスポンスヘッダフィールドに含まれるレスポンスヘッダの行数の分布を，**図 3** に示す．

図 3 より，レスポンスヘッダフィールドに含まれるレス

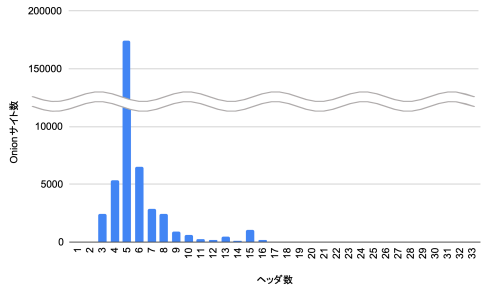


図 3 各レスポンスヘッダフィールドに含まれるレスポンスヘッダの行数の分布

Fig. 3 Distribution of the number of response headers included in each response header field.

表 4 5 件のレスポンスヘッダの組合せ

Table 4 Combination of 5 response headers.

レスポンスヘッダの組合せ	件数
Date Content-Type Server Connection Vary	171,638
Date Content-Type Server Set-Cookie X-Powered-By	878
Date Content-Type Connection Vary Keep-Alive	546
Date Content-Type Server Vary X-Xss-Protection	128
Date Content-Type Server Connection Set-Cookie	114

ポンスヘッダの行数は，5 件が最多であり，続いて 6 件，4 件，7 件，3 件，8 件が多かった．またレスポンスヘッダフィールドに含まれるレスポンスヘッダの最小行数は 1 件，最大行数は 33 件であった．

レスポンスヘッダフィールドに含まれるレスポンスヘッダの行数が 5 件であるレスポンスについて，含まれるレスポンスヘッダの組合せのうち，上位 5 パターンを**表 4** に示す．ただし，レスポンスヘッダの順序については考慮していない．

表 4 より，レスポンスヘッダフィールドに含まれるレスポンスヘッダの行数が 5 件であるレスポンス 173,969 件のうち，約 98.7%にあたる 171,638 件のレスポンスは，`Date`，`Content-Type`，`Server`，`Connection`，`Vary` の 5 つのレスポンスヘッダの組合せであることが分かる．これは，表 3 に示したレスポンスヘッダの利用率を調査した実験の結果と矛盾しない．

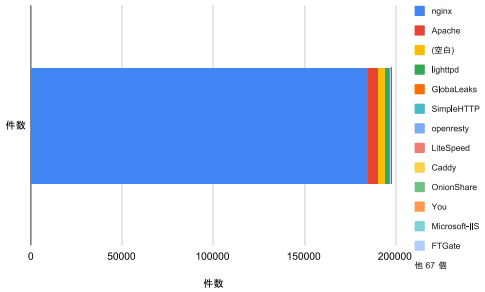


図 4 サーバソフトウェア名の分布

Fig. 4 Distribution of server software names.

表 5 サーバソフトウェア名上位 10 件

Table 5 Top 10 server software names.

サーバソフトウェア名	件数	割合
nginx	185,066	93.68%
Apache	5,409	2.74%
（空白）	3,970	2.01%
lighttpd	1,564	0.79%
GlobaLeaks	1,092	0.55%
SimpleHTTP	60	0.03%
openresty	47	0.02%
LiteSpeed	36	0.02%
Caddy	30	0.02%
OnionShare	26	0.01%

5.2.3 Server ヘッダの調査

表 3 であげたとおり，5 種類のレスポンスヘッダが調査対象 Onion Domain のうち 90%以上の利用率を示した．

この 5 種類のうち本項では特に，`Server` ヘッダについて，その値の分布を調査した．サーバソフトウェア名について，その分布を**図 4** に示す．

また，多かったサーバソフトウェア名上位 10 件とその件数を**表 5** に示す．

図 4 および表 5 より，サーバソフトウェア名が “nginx” であるレスポンスが全体の 9 割以上と最も多く，続いて “Apache”，“lighttpd” であった．

次に，`Server` ヘッダに含まれる OS 名について，その分布を**図 5** に示す．また，件数について，**表 6** に示す．

OS 名とみられる文字列が抽出できたのは，合計で 2,581 件のレスポンスであった．図 5 および表 6 より，OS 名が “Win64” であるレスポンスが最も多く，続いて “Ubuntu”，“Debian” であった．

5.2.4 レスポンスボディ

レスポンスボディについて，`<title>` タグの内容を抽出

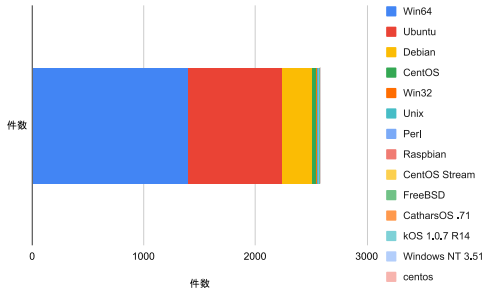


図 5 OS の分布
Fig. 5 Distribution of OS.

表 6 OS 名の分布
Table 6 Distribution of OS names.

OS 名	件数
Win64	1,395
Ubuntu	840
Debian	276
CentOS	30
Win32	11
Unix	10
Perl	6
Raspbian	5
CentOS Stream	2
FreeBSD	2
CatharsOS .71	1
kOS 1.0.7 R14	1
Windows NT 3.51	1
centos	1

表 7 多かったタイトル上位 10 件とその件数
Table 7 Top 10 titles and their counts.

タイトル	件数
Best Onion C****d Porn Collection	6,322
Raped Bitch - Real Rape Material	6,066
Paypal Account	4,001
GC King -_- GiftCard Shop	3,500
Buy Real Money — The Way	3,016
Real Rape	2,968
BlackMarket Cc - REAL SELLER CARDS — WESTERN UNION — PAYPAL	2,891
QF Market - Fast Transfers	2,890
All BTC .:. Everything you needed	2,872
CCPPShop - Paypal accounts and cloned cards for sale	2,672

し，その重複件数を調査した結果，件数が多いタイトルとその件数を**表 7** に示す．

合計で 196,459 件の Onion サイトから，9,377 種類のタイトルを収集した．2 件以上の Onion サイトで重複していたタイトルが 3,180 件，1 件のみの Onion サイトで発見されたタイトルが 6,197 件存在した．すなわち，今回の調査対象においては，7 割程度の Onion サイトは，他の Onion サイトと重複したタイトルを用いているといえる．

表 8 HTTP/1.0 によって得られたレスポンスボディの多かったタイトル上位 10 件とその件数
Table 8 Top 10 titles and their counts obtained by HTTP/1.0.

タイトル	件数
HTTP Status 400 – Bad Request	171,348
Error	3,975
Welcome to nginx!	1,139
403 Forbidden	958
LoliPorn	838
301 Moved Permanently	717
（空白）	498
Bitcoin Wallet - Free and Secure Bitcoin Wallet	262
You or your scanner suck [NO HTTP-HOST FIELD]	251
Fresh Onions	233

表 9 サーフェスウェブにおけるレスポンスヘッダの利用率 [11]
Table 9 Usage rate of response headers on the surface web [11].

レスポンスヘッダ	件数	利用率
Date	162,892,279	94.94%
Content-Type	164,640,601	95.96%
Server	158,741,768	92.52%
Connection	152,906,873	89.12%
Vary	7,460,567	4.35%

5.3 HTTP/1.0 によるリクエスト

HTTP/1.0 によるリクエストによって得られたレスポンスボディについて，`<title>` タグの内容を抽出し，その重複件数を調査した結果，多かったタイトルとその件数を**表 8** に示す．

HTTP/1.1 によるリクエストでタイトルを収集できた 196,459 件の Onion サイトに HTTP/1.0 でリクエストをした結果，合計で 191,223 件の Onion サイトから，3,911 種類のタイトルを収集した．2 件以上の Onion サイトで重複していたタイトルが 903 件，1 件のみの Onion サイトで発見されたタイトルが 3,008 件存在した．

6. 分析

調査結果をもとに，分析とその検証を示す．

6.1 調査結果の分析

6.1.1 レスポンスヘッダ

表 3 に示すとおり，`Date`，`Content-Type`，`Server`，`Connection`，`Vary` の 5 種のレスポンスヘッダが 90%以上の利用率を示した．これらのレスポンスヘッダについて，サーフェスウェブ上での利用率を**表 9** に示す．ただし，サーフェスウェブ上での利用率は，以降では Shodan [11] による 80 番ポートでの利用率の調査結果を用いる．

表 9 より，`Vary` タグの利用率について，サーフェスウェブ上での利用率が 4.35%であるのに対し，Tor 上での利用率が 92.50%と，大きな差があることが分かる．ここで，

表 10 サーフェスウェブにおけるサーバソフトウェアの利用率 [11]
Table 10 Usage rate of server software on the surface web [11].

サーバソフトウェア名	件数	割合
CloudFront	107,733,513	67.87%
nginx	12,572,560	7.92%
AkamaiGHost	8,853,399	5.58%
Apache	8,090,036	5.10%
awselb	3,628,516	2.29%
Microsoft-IIS	2,119,805	1.34%
Microsoft-HTTPAPI	764,395	0.48%
micro_httpd	760,841	0.48%
lighttpd	638,615	0.40%
openresty	506,398	0.32%

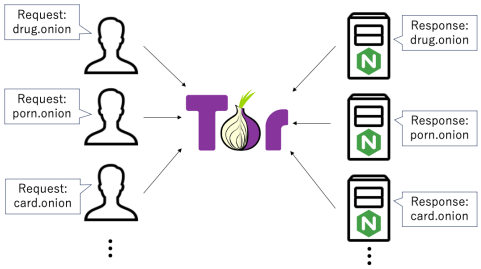


図 6 仮説 1
Fig. 6 Hypothesis 1.

サーフェスウェブにおけるサーバソフトウェアの利用率について調査した結果を**表 10** に示す．

表 10 より，Tor での利用率と異なる点として，`CloudFront` が 67.87%で最も多く利用されていることがあげられる．`CloudFront` は Amazon Web Services（AWS）のコンテンツデリバリネットワーク（CDN）であり，デフォルトの設定では `Vary` ヘッダを含まない．そのため，サーフェスウェブと Tor 上での `Vary` ヘッダの利用率に大きな差がある要因は，用いられているサーバソフトウェアの割合の差に起因していると考えられる．

`CloudFront` の特性上ダークウェブでは利用されにくいことを考慮して，`CloudFront` を除いて分析を行っても，Tor での状況と異なりサーフェスウェブでは `nginx` が支配的なシェアを有しているとはいえない．そのため，Tor において `nginx` が支配的なシェアを有していることは，サーフェスウェブと異なる特徴であるといえ，何らかの要因があると考えることが自然である．以下に，2 種類の仮説を示す．またそれぞれの仮説を**図 6** および**図 7** に示す．

仮説 1 多くの運用元が同種のサーバソフトウェアを採用している（図 6）．

仮説 2 少数の運用元が同種のサーバソフトウェアを用いてそれぞれが大量の Onion サイトをホストしている（図 7）．

2 つの仮説のうちどちらの可能性が高いかについて，次項でレスポンスボディとバーチャルホストの分析結果を踏まえて議論する．

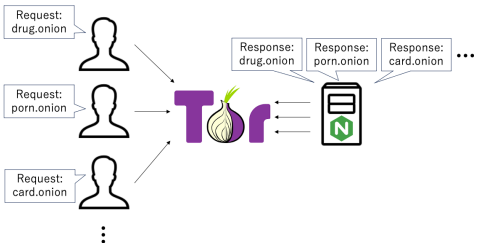


図 7 仮説 2
Fig. 7 Hypothesis 2.

表 11 HTTP/1.0 によってエラー無く得られたレスポンスボディの多かったタイトル上位 10 件とその件数
Table 11 Top 10 titles obtained without errors by HTTP/1.0.

タイトル	件数	件数（HTTP/1.1）
Welcome to nginx!	1,139	25
LoliPorn	838	1,000
（空白）	498	863
Bitcoin Wallet - Free and Secure Bitcoin Wallet	262	262
Fresh Onions	233	236
+27500 Brutal teen porn videos, watch or download.	133	7
DARK LIST	127	40
Как вы здесь оказались?	126	0
shop John The Ripper Private Hacking Tools For Sale — Hacker tolls Online in Dark web - Password Hacking Service Online	119	6
CARDING CASHOUT SHOP	115	5

6.1.2 レスポンスボディとバーチャルホスト

HTTP/1.0 によるリクエストで得られたレスポンスのうち，エラーメッセージを含まないタイトルが得られたものについて，タイトルと件数を**表 11** に示す．ただし，HTTP/1.1 によるリクエストで得られた件数も併記する．

表 11 より，HTTP/1.1 によるリクエスト時と HTTP/1.0 によるリクエスト時で，大きく発見数が異なるタイトルが存在することが分かる．特に，“shop John The Ripper Private Hacking Tools For Sale — Hacker tolls Online in Dark web - Password Hacking Service Online” のように，HTTP/1.0 によるリクエスト時のほうが多く発見されたタイトルについては，バーチャルホスト機能におけるデフォルト設定である可能性があるといえる．またこの集計結果から，少数の運用元が，それぞれ 1 つのサーバで大量の Onion サイトをホストしている可能性があるといえ，Onion Domain の数と比べて運用元の数は著しく少ない可能性があるといえる．このことから，前項で示した仮説については，“仮説 2” の可能性が高いといえる．

表 12 検証対象の Onion サイト
Table 12 Onion sites for verification.

	title
Onion サイト 1	Cash Closet
Onion サイト 2	Omega Flippers - Serious Middlemen Only!
Onion サイト 3	HOME
Onion サイト 4	404 Not Found
Onion サイト 5	Old New Money
Onion サイト 6	g6x...6yd.onion — Coming Soon
Onion サイト 7	Verifo Financial Services - Amazon Gift Cards & Western Money Order
Onion サイト 8	Million Multiplier — The Right BTC Multiplier
Onion サイト 9	Deep Web Escrow
Onion サイト 10	VMB - Amazon and Itunes Gift Cards

6.2 検証

6.2.1 検証手法

6.1 節で示した “仮説 2” について，バーチャルホスト機能の特性を用いて検証を行う．しかし，2.3 節で述べたとおり，Tor 上の Onion サイトに対しては，通常の DNS クエリを用いてドメイン名から IP アドレスを取得することができない．そのため，IP アドレスと名前ベースのバーチャルホスト機能の特性を用いて検証を行うことは困難である．そこで，複数のホストが同一のサーバで運用されているかを調査するため，HTTP/1.1 リクエストにおける Host ヘッダを書き換えてリクエストを行う．具体的には，あるドメイン A とドメイン B が名前ベースのバーチャルホスト機能を用いて同じサーバで運用されてることは，ドメイン A に対して，ドメイン B を Host ヘッダに指定してリクエストした際に，ドメイン B のコンテンツをレスポンスボディとして得ることで検証することができる，という特性を用いる．

着目するドメインの組合せは，同一のレスポンスヘッダである Onion Domain とする．ただし，すべての Onion Domain について調査を行うことは難しい．そこで本研究では，複数の異なる内容を扱う Onion サイトが，異なる Onion Domain を用いて同一のサーバで運用されていると考えられる結果を示す．また本研究は，個別の Onion サイトの運用元が同じであることを示すことが目的ではない．そのため研究倫理の観点から同一サーバで運用されているといえる具体的な Onion Domain は示さない．

6.2.2 検証結果

HTTP/1.0 によるリクエスト時に “Omega Flippers - Serious Middlemen Only!” というタイトルのコンテンツを返した 10 件の Onion サイトに着目する．着目する Onion サイトについて，HTTP/1.1 でリクエストした際のレスポンスボディに含まれる HTML の `<title>` タグの内容を**表 12** に示す．

表 13 検証結果
Table 13 Verification results.

	title
Onion サイト 1	VMB - Amazon and Itunes Gift Cards
Onion サイト 2	VMB - Amazon and Itunes Gift Cards
Onion サイト 3	VMB - Amazon and Itunes Gift Cards
Onion サイト 4	VMB - Amazon and Itunes Gift Cards
Onion サイト 5	VMB - Amazon and Itunes Gift Cards
Onion サイト 6	VMB - Amazon and Itunes Gift Cards
Onion サイト 7	VMB - Amazon and Itunes Gift Cards
Onion サイト 8	VMB - Amazon and Itunes Gift Cards
Onion サイト 9	VMB - Amazon and Itunes Gift Cards

表 12 に示すとおり，HTTP/1.1 でリクエストした際のレスポンスボディに含まれる HTML の `<title>` タグの内容は，10 件ともそれぞれ異なる．ここで，6.2.1 項で示した手法を用いて，Onion サイト 1〜9 に対して，Onion サイト 10 のドメインを Host ヘッダに指定してリクエストを行った結果を**表 13** に示す．

表 13 に示すとおり，Onion サイト 1〜9 に対して，Onion サイト 10 の Onion Domain を Host ヘッダに指定してリクエストを行った結果，すべての Onion サイトが Onion サイト 10 と同一のコンテンツを返した．6.2.1 項で示したとおり，Onion サイト 10 と Onion サイト 1〜9 が同一のサーバで運用されていない限り，このような結果は得られない．したがって，Onion サイト 1〜9 はそれぞれ Onion サイト 10 と同一のサーバで運用されているといえる．また，Onion サイト 1〜10 を Host ヘッダに指定してリクエストを行った結果，すべての Onion サイトが実際にリクエストした Onion サイトのコンテンツにかかわらず，Host ヘッダに指定した Onion Domain のサイトのコンテンツを返した．また，Onion サイト 1〜10 以外の Onion Domain を Host ヘッダに指定してリクエストを送ると，エラーを返した．これらの結果から，Onion サイト 1〜10 はプロキシではなく，同一のサーバで運用されているといえる．

この検証結果から，同一のサーバで大量の Onion サイトを運用している場合があり，“仮説 2” が成立しうることが示唆された．また，表 11 で示すタイトルを持つ他の Onion サイトでも同一の事例が見られることから，多くの Onion サイトについて，少数の運用者が大量の Onion サイトを運用している傾向があるといえる．

7. 考察

5 章で示した調査結果および 6 章で示した分析結果から，サーフェスウェブと Tor では，メタデータについて異なる分布を示すことが分かった．また，図 7 および 6.2 節で示したように，少数の運用元がバーチャルホスト機能を用いて大量の Onion サイトをホストしている場合があることが示された．一般に，バーチャルホスト機能を用いて多

くのサイトをホストする理由として，業としてホスティングサービスを行っていることが考えられる．しかし，バーチャルホスト機能におけるデフォルト設定のコンテンツが，6.1.2 項で示したような違法な可能性がある Web サイトである場合があることを考えると，当該運用元が業としてホスティングサービスを行っているとは考えにくい．そのため，運用元は自身の Web サイトとして複数の Onion サイトを運用している可能性が高いといえる．たとえば，マーケットプレイスと決済サイトを同一運用元が別個の Onion サイトとして運営すれば，ユーザの経済活動を掌握でき，ユーザの囲い込みが可能となる．実際に，表 12 で示したサイト群については，ギフトカードを扱うマーケットプレイスのほか決済サイトや取引仲介サイト，ミキシングサイトなどが含まれており，6.2.2 項で示したとおり，これらのサイトが同一の運用元によって運営されているといえるため，ユーザの取引開始から完了までをすべて掌握できる．また，複数の違法コンテンツ掲載サイトを別個の Onion サイトとして運用すれば，アクセス数を増やすことができる可能性がある．さらに，1 つのサービスに対して複数の Onion Domain を割り当てることで，1 つの Onion Domain が有名になりすぎることを防げるといえる．

本研究により，同一の運用者である可能性がある Onion サイトを検証することができるようになり，今後の調査において，ダークウェブサイトの経済圏の把握や，類似違法サイトを運用している運用者の摘発とそれにともなう Onion サイトの一斉閉鎖などに活用できるといえる．また，これまで少ない数のダークウェブサイトを分析して得られてきた知見について，再検討が必要である可能性が示唆され，同時に，これらの知見の正当性を検証するための新たなアプローチが可能となった．たとえば，少数の運用者が大量の Onion サイトを運用しているために，従来のダークウェブのコンテンツの流行などに関する研究は，運用者数の偏りによって結果の一部が歪んでいる可能性があるといえる．これについて，本研究のデータおよび手法を用いることで，多くの運用者が同一のコピーサイトを運営しているサイトを除外して，流行情勢を分析できる．さらに，ダークウェブにおいて決済サービスを用いる場合など，外部サイトに誘導することで第三者性を確保しているように主張されている場合においても，当該サービスの第三者性の検証が不可欠であることも示唆された．

一方で，運用元が多くの Onion サイトをホストするモチベーションは必ずしも明らかではない．そのため，運用元がバーチャルホスト機能を用いて，同一または異なるコンテンツの Onion サイトを運用するモチベーションについての調査が必要である．

8. まとめ

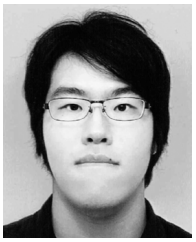
本研究では，クローラによって Onion Domain を約 20 万件収集し，その Onion Domain に対して HTTP/1.1 によるリクエストを送信し，そのレスポンスヘッダとレスポンスボディを収集した．また，HTTP/1.0 によるリクエストによって得られたレスポンスボディについても収集し，HTTP/1.1 によるリクエストによって得られたレスポンスボディと比較した．その結果，Tor におけるサーバソフトウェアの利用状況がサーフェスウェブと異なることが分かった．さらに，Tor とサーフェスウェブではメタデータの特徴に異なる傾向がある場合が示され，Tor ではバーチャルホスト機能によって，少数のサーバが大量の Onion サイトをホストしている可能性があることが示唆された．さらに，実際にバーチャルホスト機能を用いて検証を行い，同一のサーバで運用されている複数の Onion サイトが存在することを示した．

課題として，今回の調査対象とした Onion Domain の数が全 Onion Domain の 4 分の 1 程度であることがあげられる．杉生らの研究 [8] や我々の以前の研究 [15] では，ハイパーリンク上の距離が近い Onion サイトは類似のコンテンツである場合が多く，同一の運用元である場合があるとしており，クローラによる Onion Domain の収集は，メタデータの分布に偏りがある可能性がある．そのため，メタデータに偏りがないといえる Onion Domain の数を算出するか，100%に近い件数の Onion Domain の収集を目指す必要があるといえる．

参考文献

[1] 総務省：情報通信白書令和 5 年版 (2023).
[2] TorProject: Tor Project – Anonymity Online, available from ⟨https://www.torproject.org/⟩ (accessed 2023-01-14).
[3] TorProject: Tor Project – Onion Services, available from ⟨https://community.torproject.org/onion-services/⟩ (accessed 2023-02-05).
[4] Kimura, Y., Akiyama, S., Inomata, A. and Uehara, T.: On Collecting Onion Server Fingerprints and Identification of Their Operators, *IEEE International Conference on Software Quality, Reliability and Security Companion* (*QRS-C*), pp.530–538 (online), DOI: 10.1109/QRS-C60940.2023.00045 (2023).
[5] TorProject: Onion Services – Tor Metrics, available from ⟨https://metrics.torproject.org/hidserv-dir-onions-seen.html⟩ (accessed 2024-01-26).
[6] 小野諒人，神薗雅紀，笠間貴弘，上原哲太郎：HSDir の Snooping と秘匿サービスへのスキャンを組み合わせたダークウェブ分析システム，Vol.117, No.481, 電子情報通信学会，pp.103–108 (2018)（オンライン），入手先 ⟨https://cir.nii.ac.jp/crid/1520290883139979264⟩.
[7] Pastor-Galindo, J., Gómez Mármol, F. and Martínez Pérez, G.: On the gathering of Tor onion addresses, *Future Gener. Comput. Syst.*, Vol.145, No.C, pp.12–26 (online), DOI: 10.1016/j.future.2023.02.024 (2023).
[8] 杉生貴成，猪俣敦夫：Dark Web のコンテンツ分析とつながりの解明，研究報告インターネットと運用技術（IOT），Vol.2018, No.10, pp.1–6 (2018).
[9] 新井 悠，吉岡克成，松本 勉：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，コンピュータセキュリティシンポジウム 2019 論文集，Vol.2019, pp.482–487 (2019)（オンライン），入手先 ⟨https://cir.nii.ac.jp/crid/1050855522099615488⟩.
[10] 新井 悠，吉岡克成，松本 勉：ダークウェブ内の違法物品取扱サイトの HTTP ヘッダ情報を特徴量にした同サイトの自動検出，情報処理学会論文誌，Vol.61, No.9, pp.1388–1396（オンライン），DOI: 10.20729/00206786 (2020).
[11] Shodan: Shodan Search Engine, available from ⟨https://www.shodan.io/⟩ (accessed 2024-01-26).
[12] 木村悠生，穐山空道，猪俣敦夫，上原哲太郎：Onion Service における Server Fingerprint の収集及び運用元特定可能性の検討，信学技報，ICSS2022-75, Vol.122, No.422, 沖縄，pp.163–168 (2023).
[13] Nielsen, H., Mogul, J., Masinter, L.M., Fielding, R.T., Gettys, J., Leach, P.J. and Berners-Lee, T.: Hypertext Transfer Protocol – HTTP/1.1, RFC 2616 (1999).
[14] Nielsen, H., Fielding, R.T. and Berners-Lee, T.: Hypertext Transfer Protocol – HTTP/1.0, RFC 1945 (1996).
[15] 木村悠生, 石川琉聖, 穐山空道, 猪俣敦夫, 上原哲太郎：Tor 内のリンクネットワークグラフ可視化と Server ヘッダによる Onion Service 運用元特定手法の検討，コンピュータセキュリティシンポジウム 2023 論文集，pp.261–268 (2023).



木村 悠生 （学生会員）

2022 年立命館大学情報理工学部卒業．2024 年立命館大学大学院情報理工学研究科博士課程前期課程修了．現在，同大学院情報理工学研究科博士課程後期課程在学中．総務省サイバーセキュリティ統括官付参事官付総務技官．ダークウェブ，印章，サイバーセキュリティ全般に興味を持つ．

穐山 空道 （正会員）

2010 年京都大学工学部情報学科卒業．2015 年東京大学大学院情報理工学系研究科創造情報学専攻修了．博士（情報理工学）．日本電信電話株式会社，産業技術総合研究所，東京大学を経て，2022 年 4 月より立命館大学情報理工学部セキュリティ・ネットワークコース准教授．メモリシステム，性能分析，仮想化技術等の研究に従事．

猪俣 敦夫 （正会員）

2002 年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了，博士（情報科学）．通信キャリア研究所を経て，2008 年奈良先端科学技術大学院大学准教授，2016 年東京電機大学教授，2019 年大阪大学教授，CISO，現在に至る．一般社団法人 JPCERT コーディネーションセンター理事，一般社団法人公衆無線 LAN 認証管理機構代表理事，大阪府警察・奈良県警察サイバーセキュリティアドバイザー．暗号理論，ネットワークセキュリティに関する教育，研究開発に従事．電子情報通信学会会員．著書に『サイバーセキュリティ入門』（共立出版）等．

上原 哲太郎 （正会員）

1995 年京都大学大学院工学研究科博士後期課程研究指導認定退学．同大学院工学研究科助手，和歌山大学システム情報学センター講師，京都大学大学院工学研究科附属情報センター助教授，同大学学術情報メディアセンター准教授，総務省情報通信戦略局通信規格課標準化推進官を経て，2013 年より立命館大学情報理工学部教授．京都大学博士（工学）．デジタル・フォレンジック，システムセキュリティ等の研究に従事．共著に『基礎から学ぶデジタル・フォレンジック』（日科技連出版）等．