

明治大学総合数理学部

2019 年度

卒 業 研 究

Tor ネットワークのクローラシステム (1)
違法商品販売の調査

学位請求者 先端メディアサイエンス学科

鳥居洸希

目次

第 1 章	はじめに	2
1.1	背景	2
1.2	Tor の匿名性	2
第 2 章	半自動クローラシステム	4
2.1	概要	4
2.2	システムの流れ	5
第 3 章	違法商品販売	6
3.1	概要	6
3.2	調査結果	6
3.3	処理時間の評価	8
3.4	考察	8
第 4 章	カード専門の販売サイトの調査	9
4.1	概要	9
4.2	調査結果	9
第 5 章	Tor 内の全ウェブサイトの動向調査	11
5.1	概要	11
5.2	調査結果	11
5.3	考察	12
第 6 章	おわりに	13
	参考文献	15

第 1 章

はじめに

1.1 背景

近年のプライバシーに対する意識の高まりから、匿名通信の重要性は益々高まっている。中でも、匿名通信システム Tor(The Onion Router) は、発信元の情報を隠したままの通信が可能であり、内部告発などに広く用いられている。本来、匿名通信はプライバシー保護の目的に設計されているが、ランサムウェアによる身代金送付などの不正な目的で利用されるケースも多い。

そこで、ダークウェブ上の違法物品販売サイト上で扱われている商品の種別やサイト運営期間等に関する調査に大きな注目が集まっている。しかし、違法商品の売買の実態をエージェントによる機械的に調査されることを避ける為に、ほとんどのサイトでは CAPTCHA を用いたセキュリティ対策を施している。この問題に対して、我々は、(1)OCR による CAPTCHA の自動解答と (2) ログイン時の CAPTCHA を解く操作のみは人間が行う半自動クローラの 2 つで解決を試みる。本稿は、主に (2) の試みと、システム全体について述べ、(1) については [1] で報告する。ただし、CAPTCHA の種類はテキストベースのものに限る。

1.2 Tor の匿名性

Tor ネットワーク内のウェブサイトは、特有の.onion という TLD(Top Level Domain) を持つ。この TLD のサイトには、通常のブラウザではアクセスすることが出来ず、Tor ブラウザが必要になる。

図 1.1 に Tor の原理図を示す。匿名性を確保するために、いくつかの中継サーバを経由し、目的のウェブサイトへアクセスする。これらの中継サーバ(リレー)は、クライアントに近い方から Guard, Middle, Exit と呼ばれている。図 1.1 の例では、 n_1 , n_8 , n_2 の順に多重に暗号化がなされている。これらのリレーは一定時間ごとに接続先が変わり、これによって匿名性を高めている。

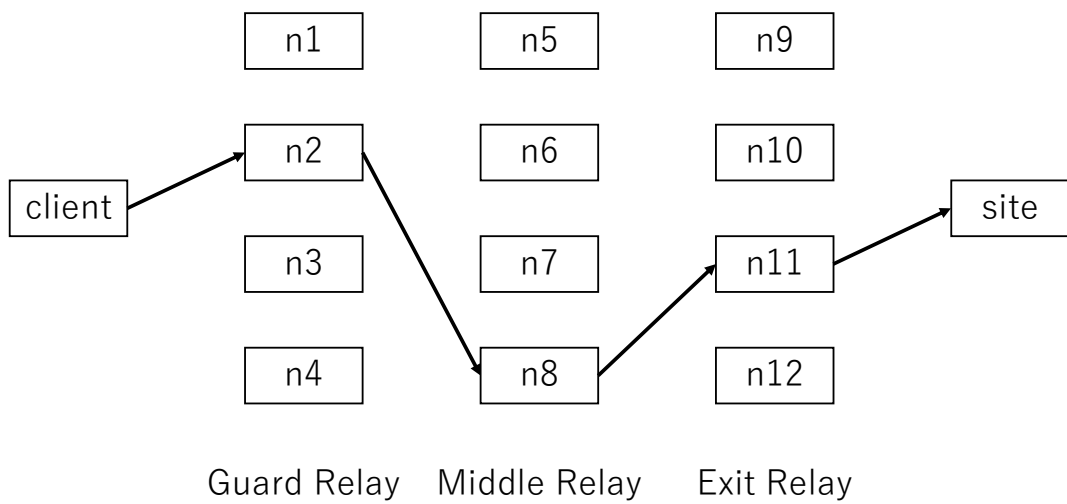


图 1.1 Tor 原理图

第 2 章

半自動クローラシステム

2.1 概要

図 2.1 に本研究で開発したクローラシステムの構成図を示す。図 2.2 に実行例を示す。本システムは、CAPTCHA 取得部、HTML 取得部、Tor 接続部、ブラウザへのインターフェースの 4 点から成る。本システムは、CAPTCHA を解くところのみ人間が操作し、以降はシステムが機械的にクローリングして動作する。

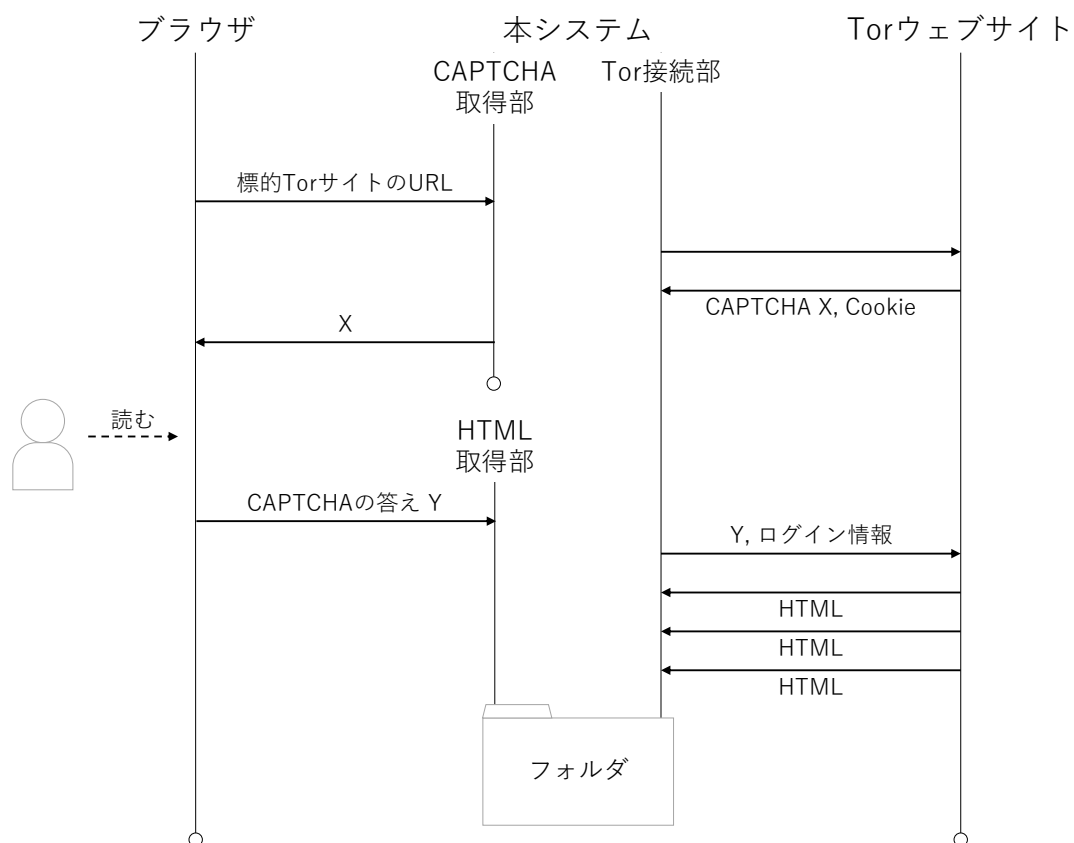


図 2.1 システム構成図


<p>サイト名</p>	<p>* Apollon Market ▼</p>
<p>CAPTCHA</p>	
<p>送信</p>	

図 2.2 本サイトの CAPTCHA 解答の実行例

2.2 システムの流れ

まず、ユーザがブラウザを用いて専用のウェブサイトへアクセスする。ユーザは、クローラの対象である URL から一つを選択する。CAPTCHA 取得部では、Tor 接続部を経由して標的ウェブサイトから CAPTCHA 画像と Cookie を取得する。Cookie により、後にアクセスした時にも CAPTCHA なしでアクセスが可能になる。

次に、ユーザは CAPTCHA を解く。CAPTCHA の答え Y が HTML 取得部に渡されると、先程の Cookie を用いて再び Tor 接続部を経由してウェブサイトへアクセスする。この時、 Y の他にユーザ名とパスワードなどのログイン情報も POST 送信する。一旦ログインに成功すると、その HTML をサーバ内に保存する。

第3章

違法商品販売

3.1 概要

Tor ネットワークのサイトは多様性があり、標的 Tor サイト毎に対処が必要である。そこで、(1)、(2) のアプローチに加えて、Tor ブラウザを使用し手動によるデータ取得も含めて調査する。

3.2 調査結果

表 3.1 は本研究で調査したサイトの一覧とデータである。手動での取得は 2019 年 9 月上旬、本システムを用いた半自動での取得は 2019 年 10 月 21 日から 11 月 19 日である。表 3.2 にカテゴリごとの情報を示す。表 3.1 と表 3.2 の平均価格は、1 EUR = 1.12 USD, 1 BTC = 7223.83 USD を用いて USD に換算した。

表 3.1 調査したサイトの情報

調査方法	サイト名	主要カテゴリ	商品数	平均価格 [USD]	通貨
手動	Northwest Nuggets	Drugs	17	111	USD
	Midland City	Drugs	12	90	USD
	DEEPTECH	Digital Goods	30	370	BTC
	Alhtbpay	Drugs	24	254	USD
	ArtGallery Shop	Arts	3	3691	EUR
	Gold-Cart	Cards	42	1146	USD
	Kamagra For Bitcoin	Drugs	10	23	BTC
	Murder Incorporated Hitmen	Service	2	12500	USD
半自動	Apollon Market	Drugs	264	368	USD,BTC
	Tor Market	Drugs	129	*1	NZD,GBP,EUR,TAB,USD
自動	Tenebra marketplace	Cards	2643	830	USD
	UnderMarket	Cards	134	520	BTC,LTC,ETH

表 3.3 に実際に販売されている商品を示す。また、データを DB に入れ、カテゴリや商品の検索をウェブサイトを用いて提供している。図 3.1 に本サイト^{*2}の実行画面を示す。

*1集計が困難なため割愛している。

*2<https://windy.mind.meiji.ac.jp/~tori/2018/summer/work/>

表 3.2 カテゴリ

カテゴリ名	商品数	サイト数	平均価格 [USD]
Drugs	790	8	*1
Cards	1372	6	227.86
Digital Goods	472	6	801.75
Service	238	5	179.77
Counterfeit	226	2	1391.47
Jewelry	41	2	1554.39
Arts	3	1	3733.33
Others	644	3	*1

表 3.3 商品例

商品名	カテゴリ	価格
Champagne MDMA	Drugs	\$170
IPHONE XS MAX 512GB	Digital Goods	\$340
Basic Hitman	Service	\$5,000

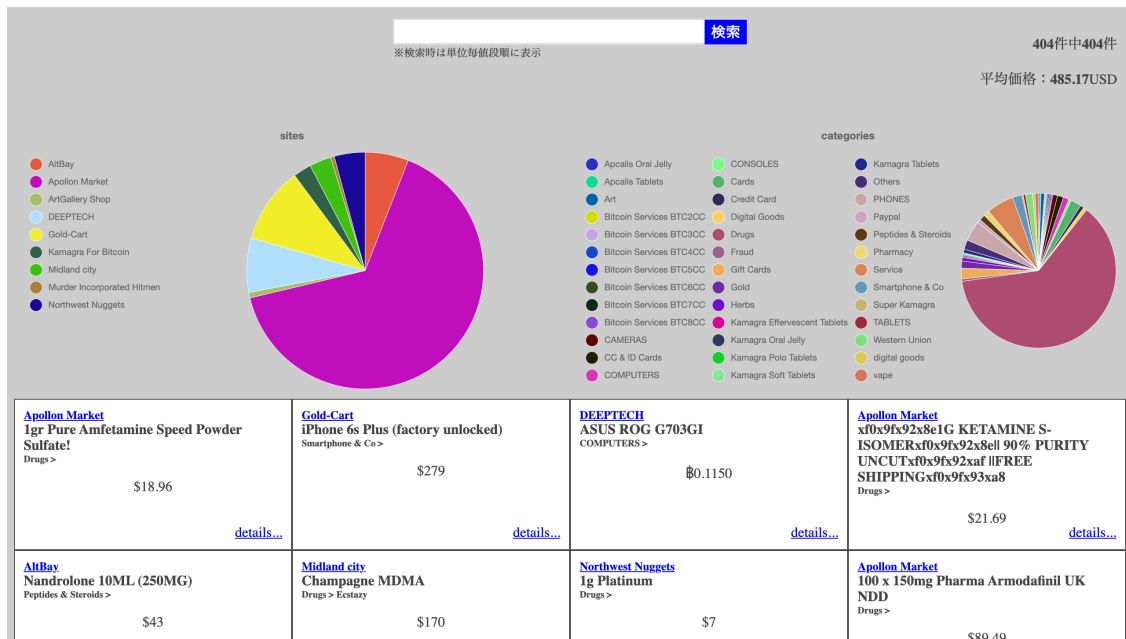


図 3.1 サイト表示例

3.3 処理時間の評価

半自動による処理時間を調査する。本システムを用いて指定したサイトを5回計測し、平均時間を求める。表3.4に計測の結果を示す。手動での調査は、半自動での調査より1.7倍以上の時間がかかる。

表 3.4 HTML 取得にかかった時間

取得方法	平均 [s]	標準偏差
手動	41.93	3.50
半自動	24.08	2.04

3.4 考察

本研究で調査した12個のサイトのうち、6個のサイトはDrugsに属する商品が一番多かった。

通貨で一番使われていたのはUSDであった。一方で、仮想通貨のBTCはEURやNZDよりも多かった。

Drugsを主要カテゴリとするサイトで販売されている商品の平均価格は比較的良かった。また、商品数が少ないものほど平均価格が高くなっている。

第 4 章

カード専門の販売サイトの調査

4.1 概要

Tormarket^{*3}はカード販売を専門としているウェブサイトである。Tormarket に売られているカードは、CVV 有りと CVV 無しの 2 つに分けられる。本研究では CVV 有りのカードを対象として調査を行った。カード自体は窃盗されたものか偽造されたものか定かではないが、データにはカード発行者 (issuer) やカードホルダーの名義と国などの情報が含まれている。

4.2 調査結果

図 4.1 に調査の結果を示す。カード所有者の国について、イシュー毎のカードの数とした。データの中には国の情報が欠落しているものもあり、図 4.1 中には取得したデータ全ては入っていない。

国毎に一番多かったのは USA のカードで 659,092 枚あった。これは全体の 80% 以上である。次に多かったのは CAN のカードで 48,895 枚あった。これは全体の 6% にも満たなかった。

イシュー毎では VISA が一番多く、545,935 枚で全体の 66.3% であった。次に多かったのは MASTERCARD で 186,580 枚あり、全体の 22.7% であった。AMERICAN EXPRESS は USA のカードしかなかった。

^{*3}<http://t2mopgckifmberr.onion/>

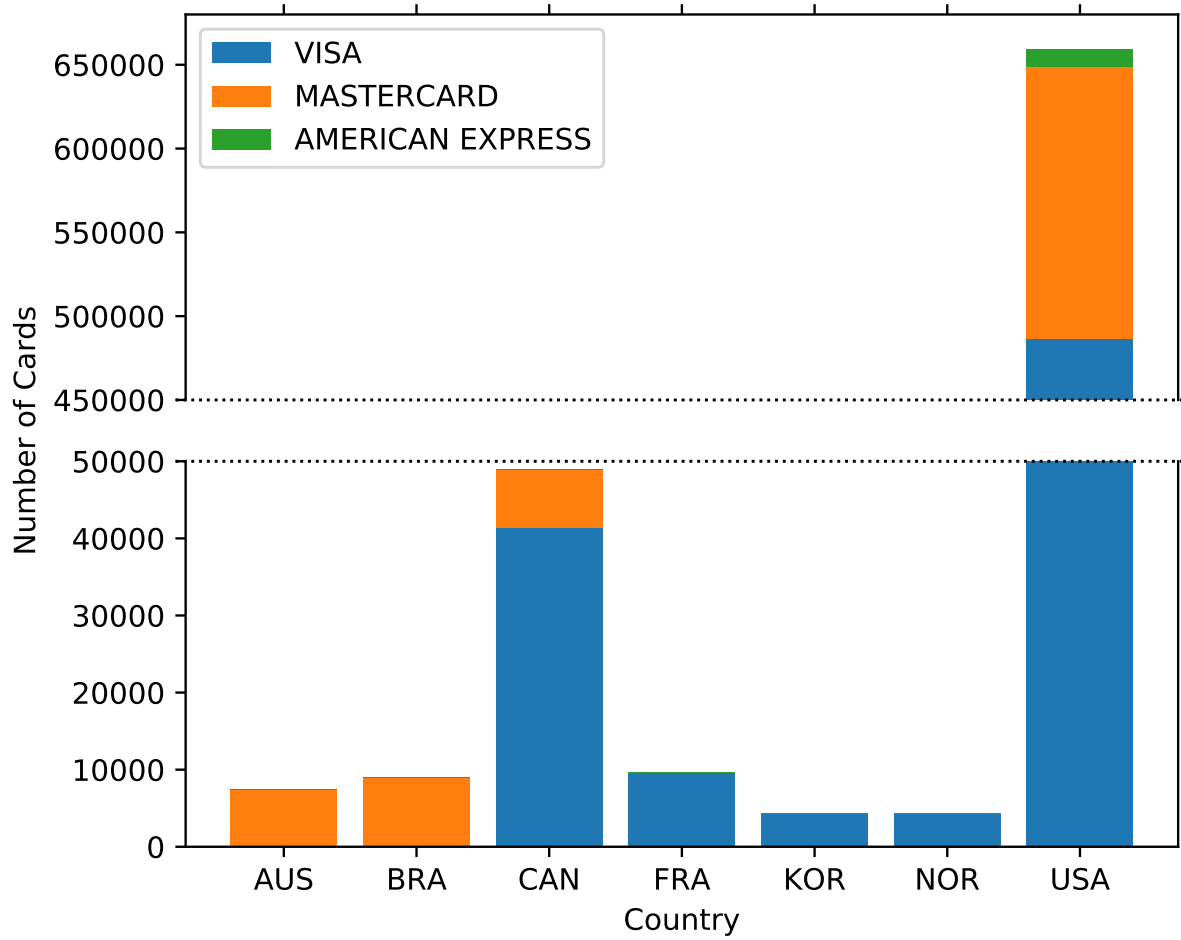


図 4.1 クレジットカードの情報

第 5 章

Tor 内の全ウェブサイトの動向調査

5.1 概要

2019 年に Tor ネットワーク内で観測されたことのあるウェブサイトの URL を取得した。URL の数は 13,698 件である。2 章の Tor 接続部を流用したプログラムと Linux の crontab を用いて定期的にアクセスすることでサイトの運用期間を調査する。運用しているか閉鎖されているかは、ステータスコードで判断する。調査は 2019 年 9 月 23 日から 12 月 7 日までに行った。

5.2 調査結果

図 5.1 に調査の結果を示す。横軸を調査を行った日付とし、縦軸をステータスコード毎のサイトの数とした。

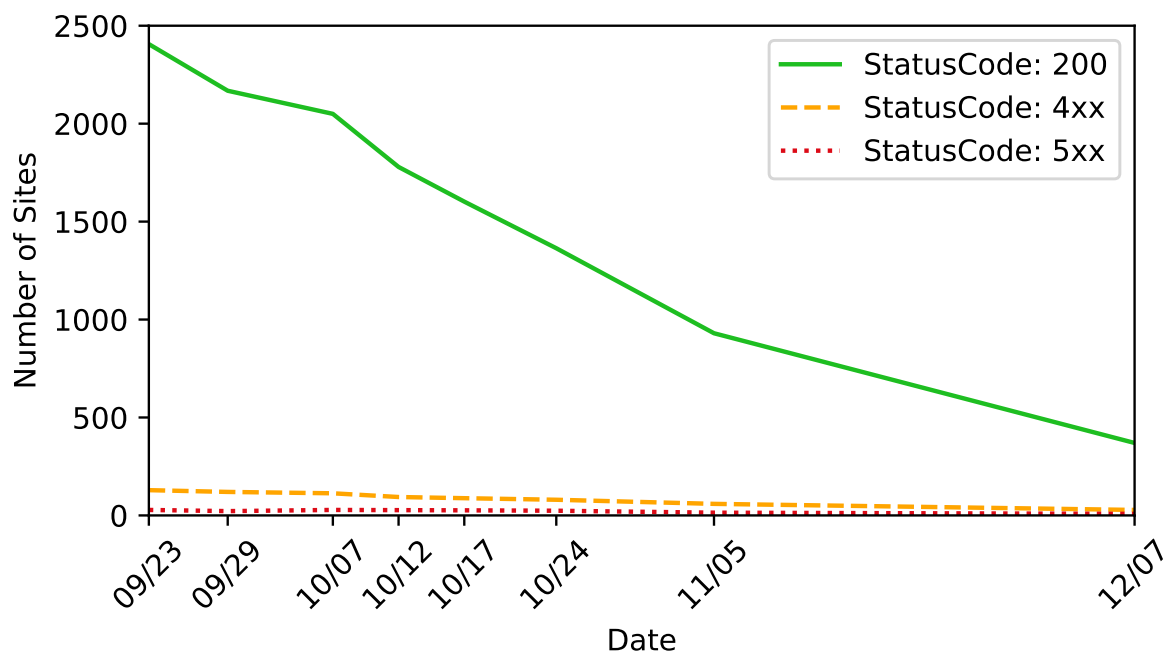


図 5.1 サイトの数の推移

13,698 件の URL のリストのうち，調査開始日にステータスコードが 200(OK) を返してきたウェブサイトは 2,406 件であった．1 日あたり数十件ずつ減少していき，観測を開始してから約 1 ヶ月の間には 1,000 件以上，観測終了日までの約 2 ヶ月半の間には 2,000 件のウェブサイトが閉鎖され，その時に残っていたサイトは 370 件であった．

1 日あたりに何件のサイトが閉鎖されているのかより詳しく知るために，観測開始日を 0 とした相対日付 x とステータスコード 200 を返すサイトの数 y について線形回帰を行い， $y = -28.035x + 2326.6$ を得た．従って，Tor サイトは 1 日あたり平均 28 件減少している．

5.3 考察

調査開始日に確認された 2,406 件のサイトは，1 日あたり約 28 件ずつ減少している．これは観測開始時に得ていたサイトの約 1% である．Tor ネットワーク全体でも同様に，1 日あたりに全体の 1% のサイトが閉鎖されていると考えれば，3 ヶ月で Tor ネットワーク内のウェブサイトは全滅すると思われる．そうなっていないことから，ウェブサイトは減少と同時に増加していることがわかる．増加量は減少量と同等若くはそれ以上と考える．なぜならば，Tor プロトコルを用いたトラフィックが，ここ数年で増加しているからだ [3]．

第 6 章

おわりに

本研究では 533 件の商品情報を取得し、それらの種別と値段について調査をした。調査の結果、主要カテゴリは Drugs であることが明らかとなった。

サイト数の遷移の調査では、約 28 件/日のサイトが閉鎖されていることが明らかになった。本研究では閉鎖までの期間を調べたが、閉鎖された URL を観測し続け、再びサイトが立ち上がれば興味深い。

また、本システムを使用することで CAPTCHA があるサイトのクローラが可能となった。手動での調査と時間比較をし、有用性があることが示された。今後は、クローラシステムの汎用性を高めるとともに、長期的な観測をして販売商品の傾向の遷移を調査することを課題とする。

謝辞

本論文は筆者が明治大学総合数理学部先端メディアサイエンス学科学士課程に在籍中の研究成果をまとめたものである。本研究の完成は多くの方々からの御指導と御援助がなければ成しえなかった。ここに感謝の意を表す。特に、学部3年時から2年間お世話になった明治大学総合数理学部先端メディアサイエンス学科教授、菊池浩明先生には指導教官として本研究の実施の機会を与えて戴き、その遂行にあたって終始、ご指導を戴いた。さらに、2年間苦楽を共にした明治大学菊池研究室の同期には、本研究のデータ取得の御協力に加え、研究に対する有益な意見を戴いた。最後に、ここまで育ててくれた両親には、ここ明治大学で学ぶ機会を頂いた。本研究だけでなく、著者の学生生活は皆様の支えなくしては成り立たなかった。この場を借りて、改めて深謝の意を表す。

参考文献

- [1] 梶間大地, 鳥居洸希, 菊池浩明, “Tor ネットワークのクローラシステム (2) CAPTCHA 自動解析”, 情報処理学会第 82 回全国大会, 発表予定.
- [2] 鳥居洸希, 菊池浩明, “Tor ネットワークのクローラシステムの開発と違法商品販売サイトの調査”, 情報処理学会第 81 回全国大会, pp.3.435-3.436, 2019.
- [3] Traffic - Tor Metrics, <https://metrics.torproject.org/bandwidth-flags.html?start=2010-01-01&end=2019-12-31>, 2020 年 1 月 5 日参照.
- [4] 大中彩香, “Tor ネットワーク内の違法商品販売サイトの調査”, 明治大学 2017 年度卒業論文, 2017.
- [5] I Gede Surya Rahayuda, Ni Putu Linda Santiari, “Crawling and Cluster Hidden Web Using Crawler Framework and Fuzzy-KNN”, 5th International Conference on Cyber and IT Service Management, 2017.
- [6] Jonghyeon Park, Hyunsu Mun, Youngseok Lee, “Improving Tor Hidden Service Crawler Performance”, IEEE Conference on Dependable and Secure Computing, 2018.
- [7] NHK ニュース, <https://www3.nhk.or.jp/news/html/20191218/k10012219611000.html>, 2019 年 12 月 28 日参照.