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

鳥居 洸希 †　　梶間 大地 †　　菊池 浩明 †

明治大学総合数理学部 先端メディアサイエンス学科 †

1　はじめに

近年のプライバシ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．

2　半自動クローラシステム

図 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．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．

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．

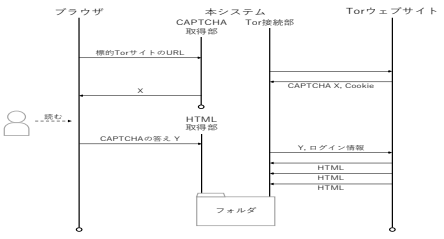

図 1　システム構成図

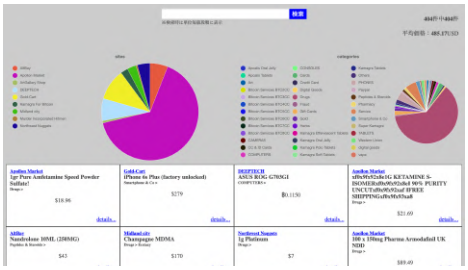

図 2　サイト表示例

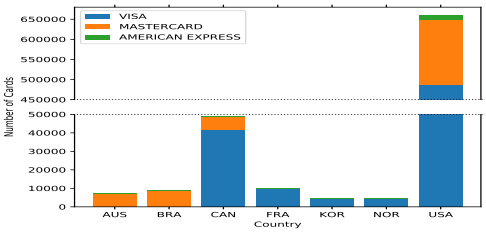

図 3　国別の販売クレジットカード割合

次に，ユーザは CAPTCHA X を解く．X 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にアクセスする．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．HTML をサーバ内に保存する．

3　違法商品販売の調査

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) に加えて，手動によるデータ取得も含め，どのようなサイトがあるのか調査する．

次に，カード販売を専門としているウェブサイト Tormarket[*3] に着目する．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．

3.1　調査結果 (1) サイトとカテゴリ

表 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 2 にカテゴリごとの情報を示す．表 1 と表 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．

Development of Tor crawler system (1) – survey of illegal products in sale in Tor market

†Hiroki Torii, Daichi Kajima, and Hiroaki Kikuchi, School of Interdisciplinary Mathematical Science, Meiji University.

[*3] http://tt2mopgckifmberr.onion/

表 1　Tor サイトと主要カテゴリ

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
	AltBay	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 2　カテゴリと平均価格

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　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

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 2 に本サイト*2の実行画面を示す．

3.2　調査結果 (2) クレジットカード

図 3 に調査の結果を示す．カード所有者の国について，イシュア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 3 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 (80%) である．次に多かったのは CAN のカードで 48,895 枚 (6%) であった．イシュア毎では 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．AMEX は USA のカードしかなかった．

3.3　処理時間の評価 (3)

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 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．

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低かった．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．例えば，Cards に属する商品は 1372 個で平均価格が 227.86 USD であるのに対し，Arts は 3 個で 3733.33 USD，Jewelry は 41 個で 1554.39 USD だった．

偽造されたカードが販売されていると仮定すると，イシュアや国の分布は幅広くなるはずである．しかし本調査の結果では，VISA カードの販売数が圧倒的に多く，AMEX は USA のものしかなかった．また，国は USA と CAN に偏りがあった．つまり，Tormarket で販売されているカードが偽造されている可能性は低いと考えられる．

4　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．

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．

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] 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．

*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/