明治大学総合数理学部

2018 年度

卒　業　研　究

ビットコインネットワーク探索パケットのダークネット観測調査

学位請求者　先端メディアサイエンス学科

半澤映拓

目次

第 1 章

はじめに

近年，匿名性から違法な売買でも用いられている暗号通貨 Bitcoin に注目が集まっている．Bitcoin はその匿名性から違法な売買でも用いられているが，Bitcoin ネットワークの複雑さのため送信元の解析は困難であるとされている．

そこで，本研究ではビットコインネットワークの観測を行い，その拡大状況等を明らかにすることを目的とする．そのため，未使用ネットワークアドレス空間であるダークネットに到達した，ビットコインのピア探索に用いられる 8333/tcp ポート宛パケットのヘッダー情報に注目し，ビットコイン P2P ネットワークのトラフィックや特徴について調査，分析を試みる．

本研究の目的は次の 2 つである．

(1) NICTER Darknet へ到達した 8333/tcp ポート宛のパケット本体，ヘッダー情報の収集を行うシステムの構築
(2) 収集したヘッダー情報の送信元に関する調査

第 2 章

要素技術

2.1 NICTER Darknet，NONSTOP について

NICTER Darknet は国立研究法人 情報通信研究機構が開発しているインシデント分析システム NICTER（Network Incident analysis for Tactical Emergency Response）[2] プロジェクトで観測を行っている/20 の連続したダークネットである．NONSTOP（NICTER Open Network Security Test-Out Platform）は NICTER の保有するサイバーセキュリティ情報を外部から利用するための分析基盤である．

NICTER Darknet に到達したパケットは PCAP サーバに日毎にダンプファイルとして保存される．パケットの IP ヘッダ，TCP ヘッダ，UDP ヘッダ，ICMP ヘッダそれら以外のプロトコルを使用するパケットのヘッダ，ペイロード情報，オプション情報を DB サーバに保存する．

2.2 Bitcoin ネットワーク

図 2.1 に Bitcoin ネットワークに接続する際の動作である“Peer Discovery”と“Connecting to Peers”について示す．

“Peer Discovery”は Bitcoin ネットワークに接続する際に隣接するピアを探索する動作である．初回は DNS シードと呼ばれるハードコーディングされたドメイン一覧を照会する．60 秒以内に応答がなかった場合はハードコーディングされた IP アドレスを照会する．以降接続する際には，前回までに接続が成功しているピアを照会する．

“Connecting to Peers”は“Peer Discovery”が成功した後に PeerP_1 が Bitcoin ネットワークへの接続を維持する動作である．P_1 は P_2 の宛先ポート 8333 に対して TCP 接続を試みる．接続が成功すると通信プロトコルのバージョン番号等を含む version メッセージを送信する．接続先の P_2 は verack メッセージと version メッセージで応答し，それに対して verack メッセージを送ることで接続が確立される．接続が確立された後は接続先の P_2 に P_1 の IP アドレスを含む addr メッセージと他のピアの IP アドレスリストを要求する getaddr メッセージを送信する．また，接続しているピアの内，30 分以上メッセージを送信していないピアに対しては，ping メッセージを送信し起動状態を確認する．ピアの起動状態が 90 分以上確認されなかった場合，そのピアを接続リストから除外する．

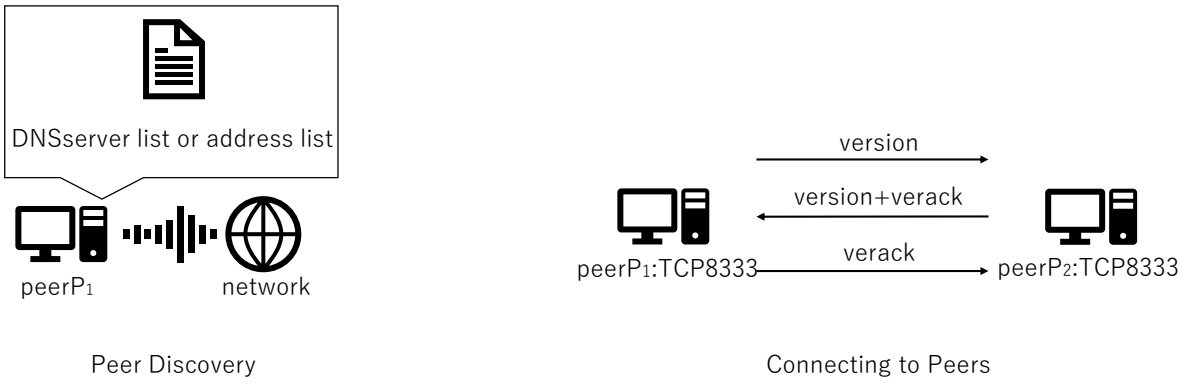

図 2.1　Bitcoin ネットワーク接続の仕組み

2.3　先行研究

今村ら [1] が複雑な Bitcoin ネットワークの分析を行う手法として，ダークネットに到達するパケットを用いた分析を提案している．その中で本来ダークネット上で観測されることが想定されない 8333/tcp 宛のパケットが 2012 年ごろから確認されており，ハッキング被害を受けた Mt.Gox が閉鎖された 2014 年 2 月 24 日前後でダークネット上で 1 日に観測される平均パケット数が大幅に増加していることを明らかにしている．

本研究ではこれらの点に注目，ダークネット上での当該パケット観測を行い，ビットコインネットワークの傾向を探る．

第 3 章

提案システム

3.1　システム

本調査では，NONSTOP を利用してシステムを構築した．図 3.1 に構成図を示す．DB サーバには，IP ヘッダ，TCP・UDP ヘッダの情報を，PCAP サーバには，pcap ファイルや spam メールを移動する．本研究で提案したシステムでは DB サーバから TCP8333 宛パケットの到着日時，送信元アドレス，送信元ポート，国コードを 8333header.csv に出力する．

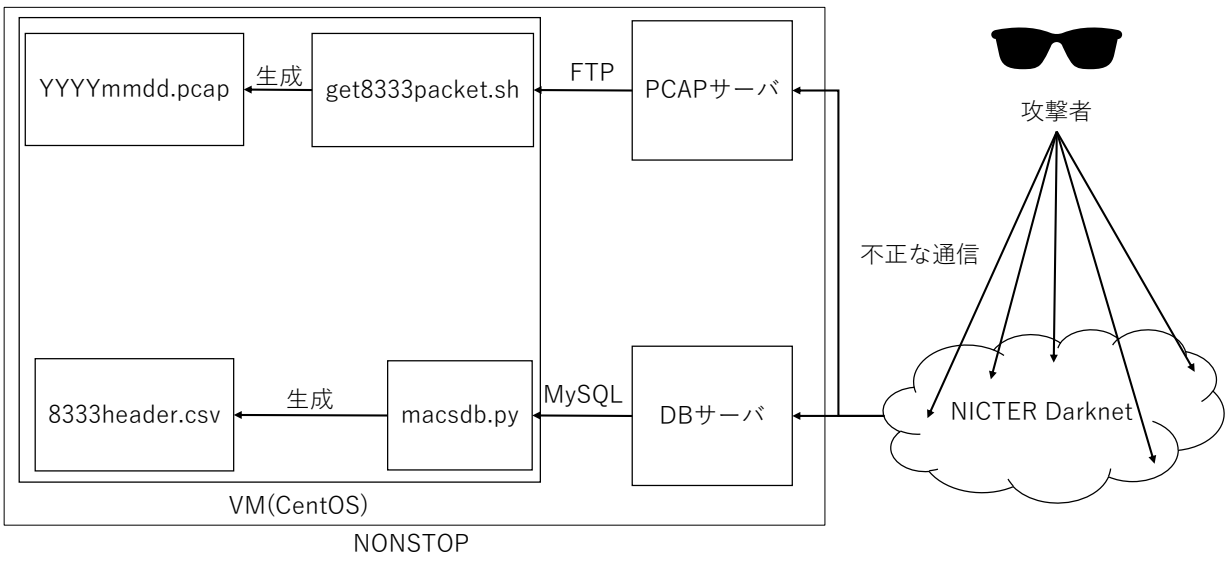

図 3.1　システム構成図

3.2　観測結果

本研究では，2018 年 1 月 1 日から 11 月 12 日の 316 日間にかけて 3.1 節で提案したシステムを用いて観測を行い，8333/tcp ポート宛パケットの TCP ヘッダ情報のパケット受信日時，送信元アドレス，送信元ポート，送信元国コードを収集した．

図 3.2 にダークネット上の 8333/tcp ポートに対するパケット数の推移を示す．最もパケット数の多かった日は 2 月 5 日で 9509 パケット，総パケット数は 674,304 であった．

受信時刻と送信元アドレス，送信元ポートについての散布図を図 3.3 と 3.4 に示す．

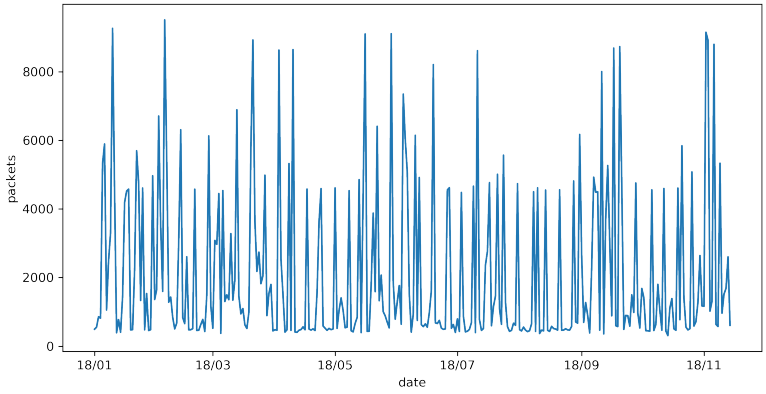

図 3.2　パケット数の推移 (2018 年)

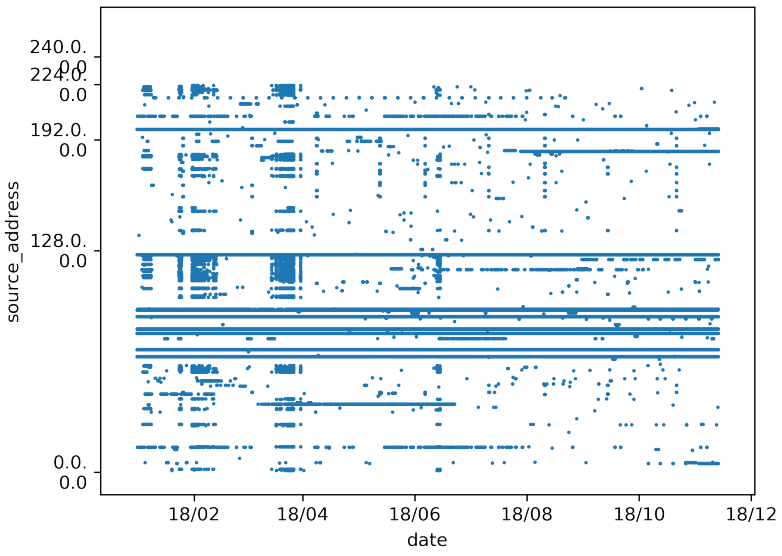

図 3.3　受信日時についての送信元アドレス

送信元ポートの種類数は 53,110 だった．8 月 31 日以前のポート番号 16000 から 33000 の間，ポート番号 33000 以上，8 月 31 日以降に特徴的な送信パターンがあることが観測できる．そこでこれらをグループ A, B, C に分け，この正体について分析を行う．

表 3.1 にグループごとのパケット送信数上位の国を示す．グループ A は 210,722 レコードで全体の 31.3 %，グループ B は 116,515 レコードで全体の 17 %，グループ C は 174,178 レコードで全体の 26 %を占める．3 グループすべての上位に米国（US）と中国（CN）が入っているが，グループ B の CN は他の 2 グループでの

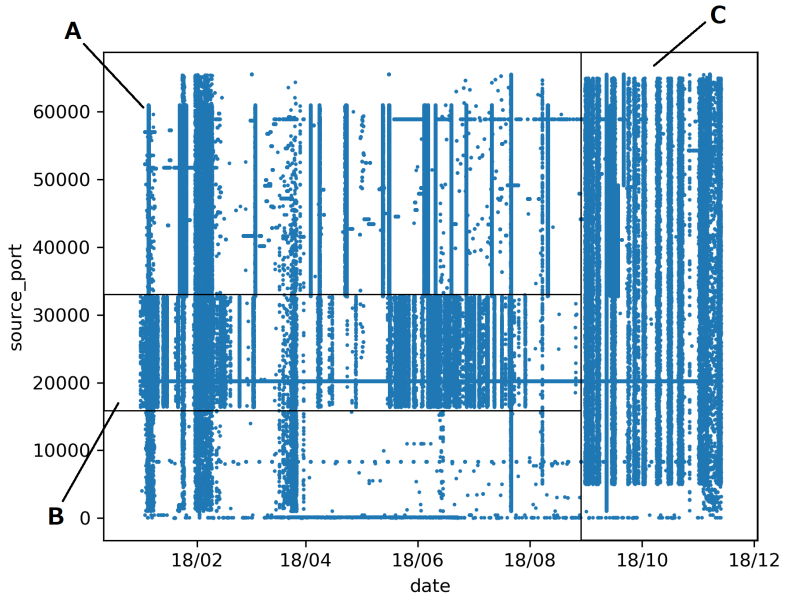

図 3.4　受信日時についての送信元ポート

表 3.1　グループごとのパケット送信件数上位国

グループ A	送信件数		グループ B	送信件数		グループ C	送信件数	
CN[*1]	48735	23.1%	US	51957	40.1%	US	73351	42.2%
US[*2]	45105	21.4%	NL	41725	32.2%	CN	47280	27.2%
CL[*3]	33778	16.0%	VN[*4]	13249	10.2%	NL	16201	9.3%
RU[*5]	29746	14.1%	CN	12870	9.9%	GB[*6]	13105	7.5%
NL[*7]	17933	8.5%	IS[*8]	5590	4.3%	RU	12604	7.3%
計	210722		計	116515		計	174178	

[*1] 中国
[*2] アメリカ
[*3] チリ
[*4] ベトナム
[*5] ロシア
[*6] イギリス
[*7] オランダ
[*8] アイスランド

CN の送信数に比べて少ない．NL が A，B，C で，RU が A，C で上位に入っている

3.3　送信元国について

表 3.2 に送信元国コードから送信パケット数上位の国と観測された国数，並びに送信元 IP アドレスを MaxMind の提供する GeoIP2 database サービスで参照した送信パケット数上位の国と観測された国の数を示す．

これらの結果から US と CN は NICTER NONSTOP と GeoIP2 で共に順位は変わらず，パケット数の差も US は 3860 件，CN は 2 件なので大きな差がないことがわかる．一方，NL は 27430 件減少し，RU と CL は上位 5 カ国から外れ，代わりに SC と GB が上位 5 カ国に入るなど，NONSTOP と GeoIP2 database サービスで送信元国が異なるレコードも存在することがわかる．NONSTOP と GeoIP サービスで送信元国が異なるレコードについての集計を表 3.3 に示す．

表 3.2 と表 3.3 より GeoIP2 で送信元国が SC と判定されたパケット 57860 件のうち，57859 件は NICTER NONSTOP で NL から送信されたパケットであると判定されていることがわかる．

3.4　送信元ポートについて

表 3.4 にパケット送信数上位の送信元ポートとそれらの送信元ポートからパケットを送信した IP アドレス数を示す．

表 3.4 の結果から送信元ポート番号 20217 からの通信は 137862 件あるにも関わらず，33 のホストからし

表 3.2　送信元国コード別パケット数と GeoIP2 を参照した際の送信元国別パケット送信数

送信元国コード	送信件数	Geoip database	送信件数
US	320460	US	316600
CN	115552	CN	115550
NL	75868	SC[*9]	57860
RU	42380	NL	48438
CL	33778	GB	41481
観測された国数	45	観測された国数	45

表 3.3　NICTER NONSTOP と Geoip database で送信元国が異なるパケット送信件数

NICTER NONSTOP	Geoip database	送信件数
NL	SC	57859
CL	NL	23473
RU	GB	17181
CL	US	10304
RU	BG[*10]	8508
US	NL	8165

[*9] セイシェル
[*10] ブルガリア

か送信されていないことがわかる．そこで最も送信件数の多い 8333/tcp からのパケットと 20217/tcp からのパケットのレコードについて詳細な分析と比較を行う．8333/tcp から送信されたパケットで送信件数の多いホスト上位 8 つを表 3.5，送信元国を上位から表 3.6，受信時刻と送信元アドレスについての散布図を図 3.5，20217/tcp から送信されたパケットで送信件数の多いホスト上位 8 つを表 3.7，送信元国を上位から表 3.8，受信時刻と送信元アドレスについての散布図を図 3.6 に示す．

8333/tcp から送信されたパケットはユニークアドレス数が 361 であるが，送信元国は 2 カ国のみでその送信元アドレスは第 1 オクテットから第 3 オクテットが全て 5.63.151, 71.6.233, 88.202.190, 216.98.153 のどれかと一致していた．また US のアドレスは 2018 年の初頭からパケットを送信しているのに対し，GB のアドレスは 2018 年の途中から観測されていた．

一方，20217/tcp から送信されたパケットは US と NL がその通信の大半を占め，上位 2 アドレスが 3 位のアドレスの 2 倍近い送信件数であった．また全体の送信数では 2 番目に多かった CN がこのポートからの通信に関しては 2 件のみにとどまっている。そこで US，NL，CN の送信元ポートに関する散布図を図 3.7,3.8,3.9 に示し比較を行う．

図 3.7 の US と図 3.8 の NL の散布図においてポートの 20000 から少し上の部分に年間を通して見られる通

表 3.4　送信元ポート別パケット送信件数とアドレスのユニーク数

送信元ポート番号	送信件数	送信元アドレスユニーク数
8333	192507	361
20217	137862	33
58914	13781	4294
49139	4107	8
52842	4107	7
43404	4103	7

表 3.5　送信元ポート 8333 からのパケット送信件数が多いホストとその国コード

送信元アドレス	送信元国コード	送信件数
71.6.216.45	US	2627
216.98.153.234	US	2624
216.98.153.247	US	2620
216.98.153.254	US	2613
71.6.216.46	US	2610
216.98.153.250	US	2610
216.98.153.227	US	2610
216.98.153.251	US	2608
216.98.153.248	US	2605
71.6.216.36	US	2603

表 3.6　送信元ポート 8333 の送信元国とパケット送信件数

送信元国コード	送信件数
US	180202
GB	12305

*11 ルーマニア

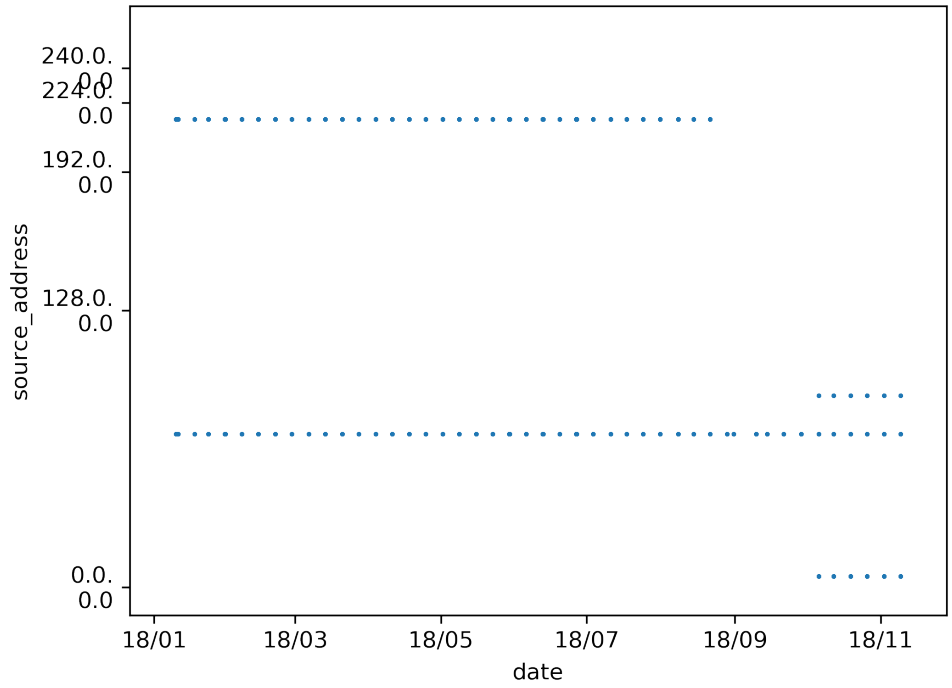

図 3.5　送信元ポート 8333 からのパケットの受信日時についての送信元アドレス

信がある．一方，図 3.9 の CN の散布図には US と NL で見られる部分が存在しない．これら 3 つの散布図からも国ごとの送信元ポートの特徴の違いが見られる．

表 3.7　送信元ポート 20217 からのパケット送信件数が多いホストとその国コード

送信元アドレス	送信元国コード	送信件数
80.82.77.139	NL	16509
80.82.77.33	NL	16263
125.212.217.215	VN	8562
125.212.217.214	VN	8527
71.6.146.185	US	7371
71.6.158.166	US	6766
89.248.167.131	NL	6536
93.174.95.106	NL	6331
71.6.146.186	US	5716
71.6.167.142	US	5484

表 3.8　送信元ポート 20217 の送信元国とパケット送信件数

送信元国コード	送信件数
US	57369
NL	55843
VN	17089
IS	7540
RO[*11]	19
CN	2

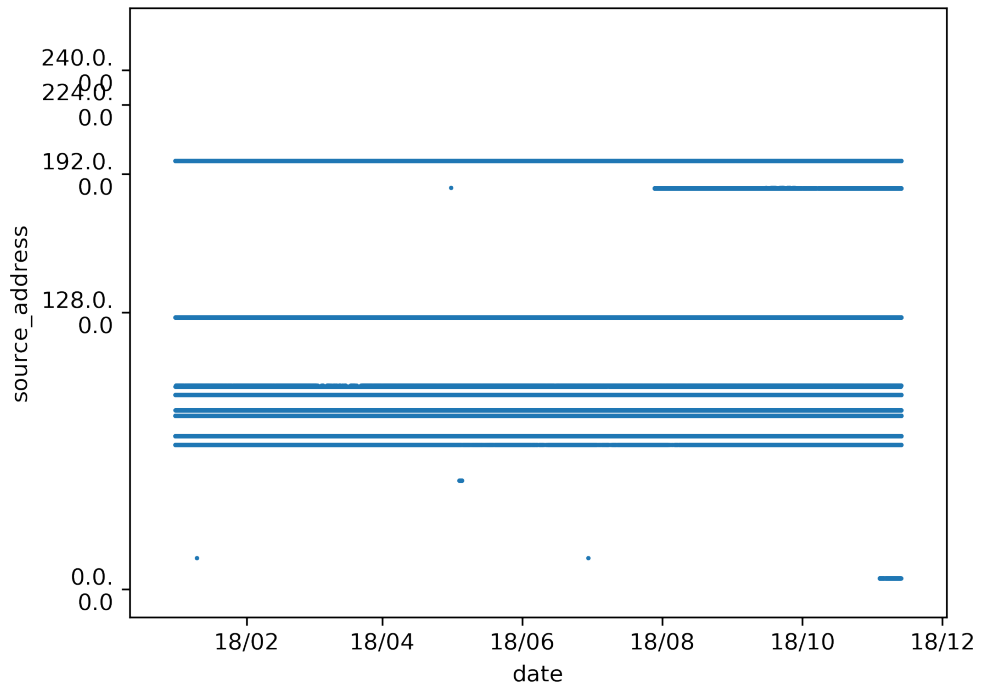

図 3.6　送信元ポート 20217 からのパケットの受信日時についての送信元アドレス

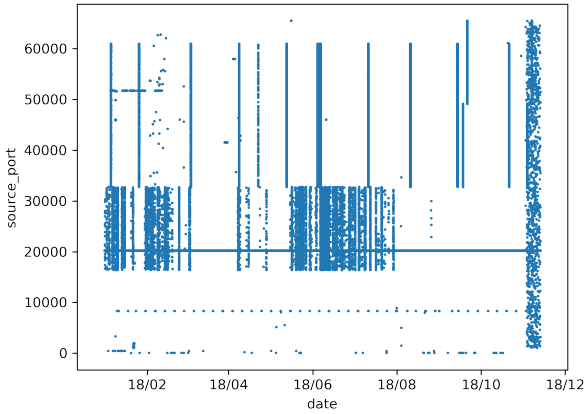

図 3.7　US の受信日時についての送信元アドレス

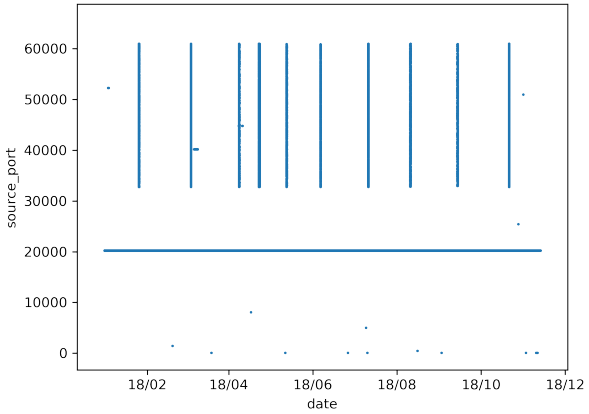

図 3.8　NL の受信日時についての送信元アドレス

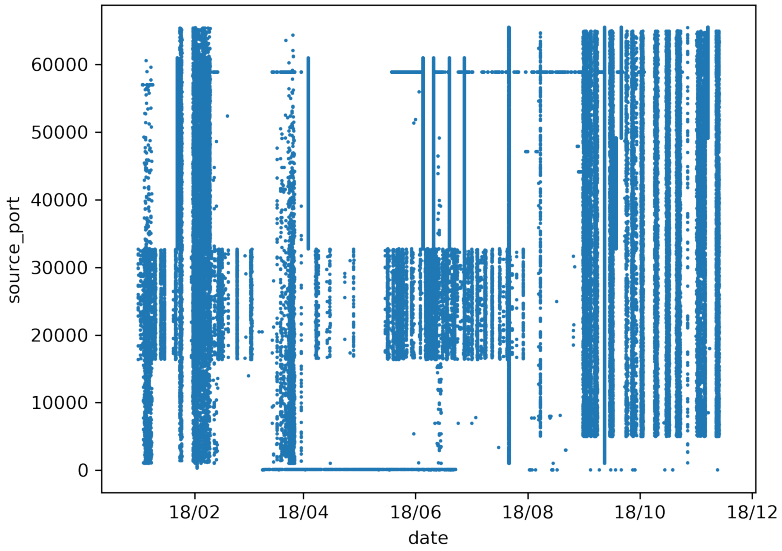

図 3.9　CN の受信日時についての送信元アドレス

第 4 章

考察

3.2 節では受信日時と送信元ポートの散布図から特徴的な部分についてグループ分けを行い，グループごとに上位を占める国の構成に差異があることを示した．したがって，国によって 8333/tcp にパケットを送信する際に使用するポートの傾向が異なると考えられる．

3.3 節では NICTER NONSTOP で割り振られた送信元国コードと GeoIP2 database サービスを参照し IP アドレスから割り出した送信元国の差について分析を行った．GeoIP2 で SC の IP アドレスであると判断されたレコードのほぼ全てが NICTER NONSTOP では NL からの送信者であると判定されていることがわかった．NICTER は送信元国の判定を行う際，送信元アドレスではない情報から決定を行っていると考えられる．

3.4 節では送信元ポート番号 8333 と 20217 のパケットに注目して送信元の国，アドレスについて分析を行った．8333/tcp から送信されたパケットは送信元アドレス数は多いもののその殆どは第 1 オクテットから第 3 オクテットまでの値が 4 種類に分別でき，送信元国も 2 カ国だけであることがわかった．20217/tcp から送信されたパケットは US と NL がそのつ通信の大半を占め，上位 2 アドレスが 3 位のアドレスの 2 倍近い送信件数であることがわかった．一方で，全体では送信数が 2 番目に多かった中国が当該ポートからの通信は 2 件のみの結果となっていた．これらのことから送信元ポートや送信元国によって送信者の特徴が異なる可能性を示せたと考える．さらに，表 3.5 の送信件数上位のホストの中で NL から送信されたものに対して IP アドレスから送信元の国を調査したところ、全て SC からの送信されたものとなっており 3.3 節の分析と合わせて，NL からの通信と SC の関係についてさらに調査が必要であると考える．

第 5 章

おわりに

本研究では，ダークネット上に到達した 8333/tcp ポート宛のパケットの送信元について調査を行い，送信元国，送信元ポートに基づく解析を行った．特定の期間，送信元ポート区間において送信元国に差異が見られた．特定のポートからパケットの送信を行った送信者の送信元アドレス，送信元国に関する特徴を明らかにした．今後は到達パケットのペイロードについて調査し，ダークネットに当該ポート宛パケットが到達する要因について分析を行うことを課題とする．

参考文献

[1] 今村光良，面和成，　“ダークネット観測情報を用いたビットコインネットワークの分析”，暗号と情報セキュリティシンポジウム (SCIS) 2018，pp.1-8，2018

[2] D Inoue, et al.， nicter: An incident analysis system toward binding network monitoring with malware analysis. In Information Security Threats Data Collection and Sharing, 2008. WISTDCS'08. WOMBAT Workshop on, pp. 58-66. IEEE, 2008.

謝辞

本研究を進めるに当たり，指導教官の菊池浩明教授からは多大な助言を賜りました．深く感謝を申し上げます．また NONSTOP の利用に際してお世話になった NICT の笠間貴弘様、ニッシンの畑太一様にも厚く御礼を申し上げます．最後に協力してくださった菊地研究室の皆様に深く感謝の意を表するとともに、謝辞にかえさせていただきます．

付録 A

SSH ハニーポットによる攻撃の観測

A.1　はじめに

2017 年 3 月には日本郵便が不正アクセスの被害を受け，約 3 万人分の個人情報が流出，同 6 月には国土交通省不動産取引システムが不正アクセスの被害を受けるなど，日本の重要機関を狙った不正アクセスが発生している．これらの脅威に対抗するために攻撃者がどのような方法で不正アクセスを行うのか，不正アクセスに成功した後どのような行動を取るのかを調査し対策を講じる必要がある．

佐藤らはハニーポット“Kippo”を用いて攻撃者が SSH サーバーにアクセスしログインするまでの動向について調査と解析を行い，狙われやすいユーザー名，パスワードを明らかにした [1].

A.2　SSH ハニーポット“Cowrie”

A.2.1　概要

ハニーポットは攻撃者の動向の観察や仕掛けられたマルウェア，ワームの分析を目的とした脆弱性のある振りをするサーバーである．“Cowrie”は攻撃ごとに観測した情報を内容ごとに connect，version，closed，success，failed，command の 6 つのテーブルに分けてデータベースに格納する．表 A.1 にデータベースと情報を示す．

connect テーブルには接続時の IP アドレス・ポート，version テーブルには通信の際の暗号化アルゴリズムの種類，closed テーブルには接続遮断時の接続時間，success テーブルにはログインに成功した時の入力情報，failed テーブルにはログインに失敗した時のユーザー名・パスワード，command テーブルには攻撃者がログイン成功後にターミナルで入力したコマンドが記録される．

表 A.1　データベースの内容

テーブル名	格納される情報の種類
connect	eventid,timestamp,session,message,src_port,system,isError,src_ip,dst_port,dst_ip,sensor
version	eventid,macCS,timestamp,session,kexAlgs,keyAlgs,,message,system,isError,src_ip,version,compCS,sensor,encCS
closed	eventid,timestamp,message,message,system,isError,src_ip,duration,session,sensor
success	eventid,username,timestamp,message,system,isError,src_ip,session,password,sensor
failed	eventid,username,timestamp,message,system,isError,src_ip,session,password,sensor
command	eventid,timestamp,message,system,isError,src_ip,session,input,sensor

A.2.2　実験内容

本研究では SSH サーバーにアクセスした攻撃者がどのようなユーザー名，パスワードを使用するのか，あるいは不正に侵入した際どのような行動を取るのかを調査するために，“Cowrie”を用いて 2017 年 7 月 20 日から 85 日間さくらの VPS サーバー上で観測した．

A.3　実験結果

A.3.1　アクセス件数

図 A.1 にアクセス件数の変化を示す．最もアクセス件数の多かった日は 8 月 18 日で，4266 件のアクセスがあった．9 月 30 日から 10 月 2 日の 3 日間にかけて 2000 件以上のアクセスが観測された．総アクセス件数は 81,070 件だった．

表 A.2 に接続回数，平均接続時間，パスワード入力回数，パスワード入力種類数の上位 5 つの IP アドレスを示す．各 IP アドレスの第 3，4 オクテットはアルファベットに置き換えている．

接続回数とパスワード入力数の項目を比べてみると一致する IP アドレスが存在しないことから，接続の多い攻撃者が必ずしもログイン試行を行っているわけではないということがわかる．

表 A.3 に入力回数の上位のパスワードを示す．“password”，“12345”，“admin”などのパスワードを設定しているサーバーが狙われやすい．また，2 位の“usuario”はスペイン語で“ユーザ”の意味であり英語圏以

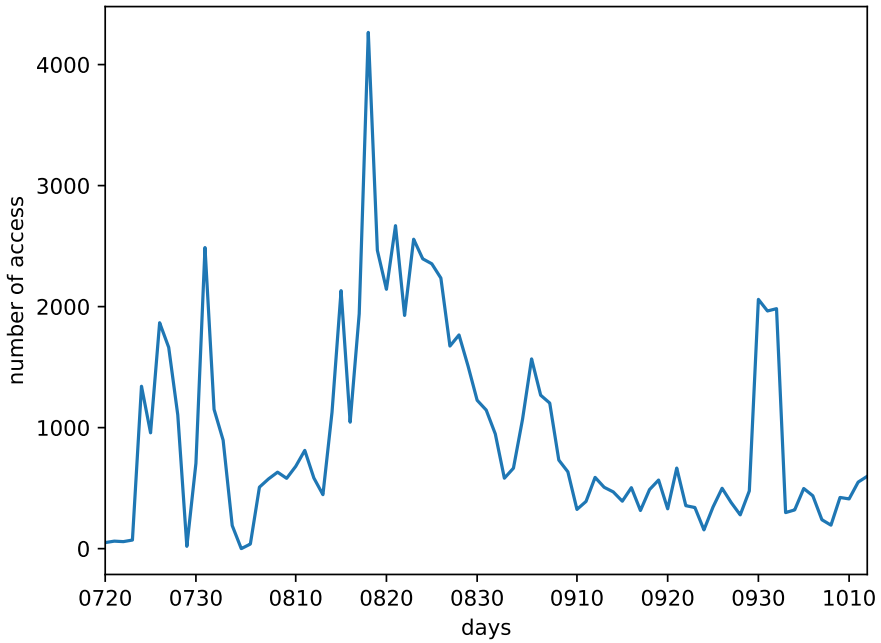

図 A.1　日ごとのアクセス件数

表 A.2　各項目の上位 IP アドレス

接続回数	平均接続時間	パスワード入力回数	パスワード入力種類数
211.23.A	171.231.F	211.110.K	211.110.K
116.196.B	103.27.G	121.194.L	121.194.L
91.47.C	108.172.H	180.168.M	213.179.O
118.70.D	5.249.I	181.118.N	180.168.M
123.16.E	88.99.J	213.179.O	50.97.P

表 A.3　入力されたパスワード上位 10 種

password	974
usuario	871
ubnt	867
12345	815
support	764
default	712
service	619
admin	600
123456	580
admin1234	438

表 A.4　入力されたコマンド上位 10 種

mkdir /tmp/.xs/	29052
cat >/tmp/.xs/daemon.i686.mod	5812
/tmp/.xs/daemon.i686.mod	5811
cat >/tmp/.xs/daemon.armv41.mod	5811
cat >/tmp/.xs/daemon.mips.mod	5811
cat >/tmp/.xs/daemon.mipsel.mod	5811
chmod 777 /tmp/.xs/daemon.i686.mod	5811
/tmp/.xs/daemon.armv41.mod	5809
chmod 777 /tmp/.xs/daemon.armv41.mod	5809
/tmp/.xs/daemon.mips.mod	5809

外からの攻撃も観測されていることがわかる．

表 A.4 に攻撃者がログイン後に入力したコマンドの上位 10 件を示す．2 位から 10 位まではほとんど差がなく，またどれも 1 位のコマンドで作成したディレクトリ，またはその下位のディレクトリでの操作を行っている．

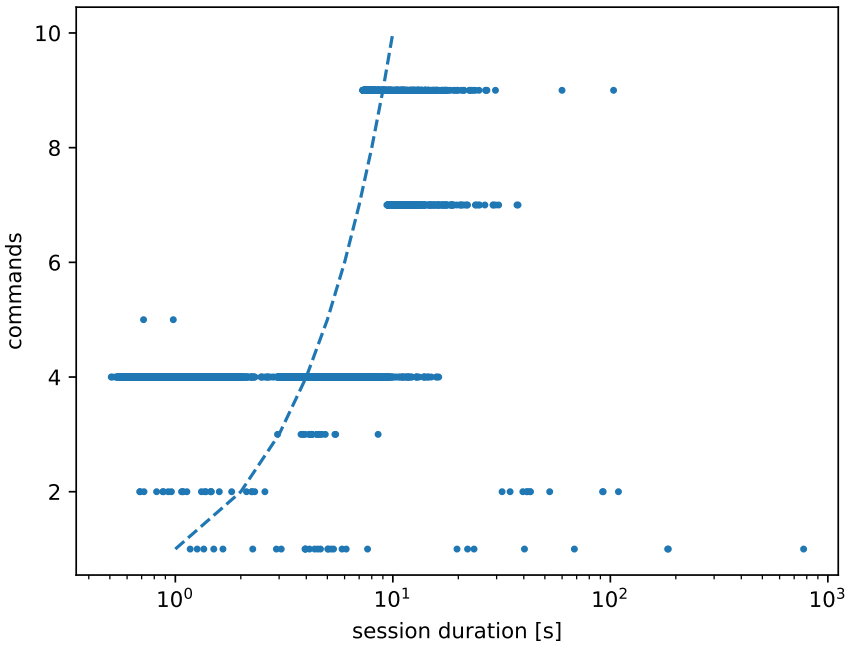

図 A.2　接続時間とコマンド入力数

A.3.2　攻撃者の判別

セッションごとの接続時間とコマンド入力数から，攻撃が人間の入力によるものか，あるいはスクリプト等の人間以外の入力によるものかの推定を行った．接続時間とその間入力されたコマンド数の散布図を図 A.2 に示す．

接続時間をコマンド入力数で割った平均入力時間の長さによって判別を行った．特に平均入力時間が 1 秒を下回るセッションは機械によるアクセスと入力がなされたセッションであると推定する．

A.3.3　NICTER Darknet との照合

ハニーポット上では 4179 個の IP アドレスが観測された．それらのアドレスが同時期に NICTER Darknet 上で観測されているのかを検証した．結果として 1674 個のアドレスがダークネット上でも観測されていた．

表 A.5 に国別の双方で観測された IP アドレスの数を示す．

表 A.5　国別の NICTER とハニーポット両者で観測されたアドレス数

国コード	アドレス数
CN	422
AR[*12]	174
US	159
RU	122
EC[*13]	82
KR[*14]	72
BR[*15]	70
FR[*16]	51

A.4　おわりに

本研究では SSH ハニーポット“Cowrie”を用いて SSH サーバーへの攻撃の観測を行った．突破される可能性の高いパスワードやログイン後の攻撃者の動向を明らかにした．また，1 セッションあたりの接続時間とコマンド入力数から攻撃が人間の手によるものか，あるいは機械によるものかの推定を行った．

[*12] アルゼンチン
[*13] エクアドル
[*14] 韓国
[*15] ブラジル
[*16] フランス

参考文献

[1] 佐藤聡，小川智也，新城靖，吉田健一　“筑波大学におけるハニーポットを用いた不適切な SSH アクセスの収集とその解析”，情報処理学会，pp.1-6，2014