L-004

複数のダークネットに対するトラフィックデータ解析とそこからの情報漏洩について

Traffic Data Analysis Result for Multiple Darknet and Information Leakage from there

村井 健祥 †　古本 啓祐 †　村上 洸介 ‡　中尾 康二 ‡, ††　森井 昌克 †
Kensho Murai　Furumoto Keisuke　Kosuke Murakami　Koji Nakao　Masakatu Morii

1　はじめに

近年，世界中にインターネットは拡大し，身近な端末を用いて国境を越えた通信を行うことができるようになった．しかし，それに伴い世界中でサイバー攻撃が増加し，複数国をまたいだ大規模なものも観測されている．そこで各国が連携して情報共有を行うことによるサイバー攻撃の検知，対応を目的とした研究が行われている．情報共有の具体的な手段として，連携する各国の所有するダークネットのトラフィックデータを収集し大規模な解析を行った後に，得られた統計データを各国にフィードバックするということが考えられる．フィードバックされた統計データを見ることで連携する各国はサイバー攻撃の現状を共有でき，早期対処につなげることが可能となる．

ダークネットとはインターネット上の未使用の IP アドレス空間であり，通常の利用目的においてダークネットにパケットが送信される可能性は低い．そのためダークネットを観測することで不正な通信を効率良く見つけることができる．しかし反対にダークネットの情報を悪意ある者に把握されてしまうとそれ以降アクセスが無くなる他，攻撃に利用される可能性もある．したがってダークネットの位置や規模などの情報は不用意に外部に公開してはならない．サイバー攻撃の情報共有の為に用いられる各国のダークネット統計データの公開も例外ではなく，公開した情報から各国のダークネットに関する情報の漏洩を防ぐ必要がある．そこでダークネット統計データの公開における匿名化処理の方法についても研究 [1] が行われている．

本稿では，時系列分析の一種である自己相関と多変量解析手法の一種である多次元尺度法（MDS: Multi-Dimensional-Scaling）を利用した複数のダークネットに対するパケット解析を行う．ダークネットにパケットを送信する IP アドレス毎の自己相関結果に多次元尺度法を適用することにより，攻撃パターンの類似する送信元 IP アドレスの判別を行う．

2　サイバー攻撃とダークネット観測

サイバー攻撃の概要とダークネットを利用したサイバー攻撃の観測法について本章で解説する．

2.1　サイバー攻撃

サイバー攻撃はネットワークを介したクラッキング行為全般を指し，ネットワークそのものやコンピュータを含むネットワークシステムを対象に行われる．具体的には標的のコンピュータやネットワークに不正に侵入してデータの詐取や破壊、改ざんなどを行ったり標的のシステムを機能不全に陥らせるといった攻撃である．世界中，特に先進国においてはコンピュータネットワークによって提供される様々なサービスに依存しており，サイバー攻撃を受けてシステムが不能になった際の被害は甚大である．したがって，サイバー攻撃に関する情報を可能な限り早く収集し他者への注意喚起や適切な対策を施すことが必要となる．また，国をまたいだサイバー攻撃の多くは必ずしも全世界一斉に行われるわけではなく，国や地域ごとに時間を変えて狙われるケースも存在する．そのようなケースにおいて国際連携による情報共有を行い，被害を低減することは重要である．サイバー攻撃に対するネットワーク解析手法として，未使用の IP アドレス空間を利用したダークネット観測手法の他に対話性を備えたハニーポットを利用した手法がある．ハニーポットは OS やアプリケーションにおいて故意に脆弱性を持たせたシステムの総称で，実際に攻撃を受けることで攻撃者の情報を詳細に得ることができる．しかしハニーポットとしたコンピュータが乗っ取られて攻撃の踏み台にされる可能性もあり，運用管理面でのコストは大きくなる．そのためダークネットとハニーポットは得られる情報量とコストにおいて対をなす存在と言える．本稿で利用するダークネット観測について次節で解説する．

2.2　ダークネット観測

ダークネットとは特定のホストコンピュータが割り当てられていない未使用の IP アドレス空間のことである．正常なネットワーク活動において，未使用 IP アドレスであるダークネットにパケットが到着する可能性は低い．したがって，ダークネットに到着するパケットはマルウェアやボットによるスキャン行為や感染活動といった，不正な通信に起因する場合が多い．ダークネット観測システムは，ダークネットの IP アドレスブロック宛に到着したパケットを観測用のホストにルーティングすることで構成される．ダークネット観測の利点として低い管理運用コストで広域なネットワークの観測ができることが挙げられる．しかし，到着パケットに対して応答しないためアクセス元 IP アドレスの詳細な情報を得られないという欠点もある．実際に使用したダークネットトラフィックデータにおけるパケットに含まれる情報を図 1 に示す．なお解析に扱うデータの都合上，送信元アドレス (SourceIP) 及び送信先アドレス (DestinationIP) は一部伏せている．次節以降，ダークネットへパケットを送信する行為を“攻撃”，またパケットを送信する IP アドレスを“攻撃元 IP アドレス”と呼ぶこととする．ダークネットへの到着パケットは攻撃が目的でないものも含まれている可能性があるが，解説の都合上“攻撃”と表記する．

† 神戸大学大学院工学研究科, Graduate School of Engineering, Kobe University
‡ KDDI 株式会社, KDDI Corpotation
†† 国立研究開発法人情報通信研究機構, National Institute of Information and Communications Technology

Time	Source	Destination	src port	dst port	Protocol	Length
0	*.*.29.185	*.*.*.131	80	51195	TCP	60
3.10161	*.*.14.154	*.*.*.227	53	15707	DNS	88
4.668908	*.*.79.43	*.*.*.137	59614	23	TCP	60
5.360658	*.*.224.130	*.*.*.225	42586	8009	TCP	60
6.103231	*.*.249.185	*.*.*.135	54404	23	TCP	74
8.200241	*.*.26.42	*.*.*.131	80	51195	TCP	60
8.34233	*.*.204.11	*.*.*.252	80	54608	TCP	60
9.104628	*.*.249.185	*.*.*.135	54404	23	TCP	74
9.659819	*.*.224.129	*.*.*.172	60744	1158	TCP	60
15.10454	*.*.224.130	*.*.*.182	34206	8009	TCP	60

図 1　ダークネット到着パケット

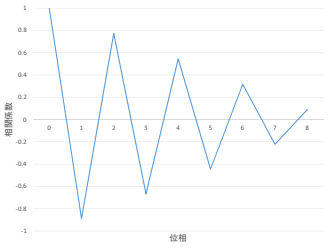

図 2　自己相関サンプル

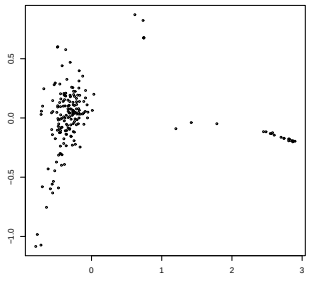

図 3　多次元尺度法サンプル

3　複数のダークネットに対するパケット解析手法

解析対象のパケットはダークネットの規模（レンジ）によって差はあるものの量が膨大であり，かつ 1 つのパケットに含まれる情報量も多い．そこで本稿ではデータ中のパケットの到着パターンに着目し，時系列分析及び多変量解析の手法を適用することを考える．自己相関と多次元尺度法を用いた手法について本章で簡潔に解説し，実際のデータを用いた解析結果については次章で述べる．本章の手法を利用することで，パケット送信元 IP アドレスの攻撃傾向を得ることが可能となる．解析にはフリーの統計ソフトである”R”[3] を用いた．得られた攻撃傾向は 5 章で述べる他国のダークネット情報の推定に利用することができ，その推定結果から統計データを公開する際の対策を検討することが可能となる．

3.1　特定のトラフィックデータに対する時系列分析

3.1.1　自己相関

時系列分析における自己相関は対象の時系列データを入力として，データが過去の履歴に対してどのくらい影響を受けているかを求めることが可能となる．また自己相関の結果が周期性を持った場合，元の時系列データも周期性を持つため，データの周期性を判定することができる．時系列データを $x_i\{i = 1, 2, \cdots\}$，x_iの相加平均を $\bar{x}$，τ を位相差とすると自己相関係数 R_{xx}は以下の式で定義される．

$$R_{xx} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(x_{i-\tau} - \bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(x_{i-\tau} - \bar{x})^2}} \quad (1)$$

3.1.2　トラフィックデータへの自己相関の適用

サイバー攻撃は人の手によるものに加え，マルウェアと呼ばれる悪意を持ったソフトウェアによっても行われる．マルウェアによる攻撃はプログラムによって自動化されており，マルウェアの検体によって攻撃に同一の特徴が見られる場合がある．またその特徴に基づいたマルウェアの分類に関する研究 [2] もなされている．ダークネットにパケットを送信する IP アドレスが多数存在する場合，攻撃の特徴の分かりやすい項目として攻撃パターンの一致がある．その理由として感染したホスト群は C & C サーバーと呼ばれる 1 つのホストから命令を受けて一斉に活動するケースが多いことが挙げられる．そこでダークネットで観測される各 IP アドレスに対して自己相関を用いて攻撃の周期性すなわち攻撃パターンの判定を行い，比較することで IP アドレスの分類を行うことを考える．まずダークネットで観測される IP アドレスに対して一定期間毎の送信パケット数を計算する．次に計算結果を時系列データとして自己相関関数の入力とする．得られた相関係数の推移から IP アドレスの攻撃パターンを判定する．自己相関のグラフサンプルを図 2 に示す．位相差のスケールはサンプルデータにおける期間である．送信パケット数によって攻撃元 IP アドレス毎に相関係数の推移が異なり，これを IP アドレス X の攻撃パターンとすることができる．また，相関係数をグラフ化することで IP アドレス間での比較を行うことができ，類似度を直感的に判定できるメリットがある．

3.2　トラフィックデータにおける多変量解析

3.2.1　多次元尺度法

多次元尺度法とは各個体の多次元データを、2 次元あるいは 3 次元空間に配置する方法である．データ間において類似したものを近く、そうでないものを遠くに配置することにより，データの構造を考察することが可能となる．多次元尺度法は計量多次元尺度法と非計量多次元尺度法に大別される．実際の距離データを低次元に配置する計量多次元尺度法に対し，非計量多次元尺度法は順序尺度のデータの類似度あるいは距離に変換可能な親近性データを低次元に配置する方法である．R における非計量多次元尺度法の手順を簡潔に述べる．まずデータにおける個体間のユークリッド距離を求める．次にストレスと呼ばれる統計量を最小にするように座標を定め，最後に 2〜3 次元上に個体を配置し散布図を作成する．多次元尺度法のグラフサンプルを図 3 に示す．図 3 より，各個体の散布状況によって類似度を判別し，また点の集合によって各個体の特徴による分類が可能となる．

3.2.2　自己相関係数への適用

各攻撃元 IP アドレスの自己相関係数の推移に対して非計量多次元尺度法を適用することで，攻撃元 IP アドレスの分類を行うことを考える．特定の攻撃元 IP アドレス間において自己相関係数の値及びその推移が似ている場合，多次元尺度法によってそれらの IP アドレス間の距離は近くなる．逆に自己相関係数が似ていない IP アドレス間の距離は遠くなる．

本稿では 2 種類の散布図を作成する．まず，複数の国に攻撃を行う IP アドレスに対しては各国別に自己相関係数の推移を導出し，それらをまとめて多次元尺度法によって 2 次元の散布図で表現する．これにより各国に対

表 1　攻撃元 IP アドレスリストの抜粋

順位	IP アドレス	項目 1（日）	項目 2（ヶ国）	項目 3（パケット）
1	*.*.144.64	30	7	218761
2	*.*.144.65	30	7	213692
3	*.*.144.67	30	7	212431
4	*.*.144.66	30	7	197696
5	*.*.224.129	30	7	94671
500	*.*.50.189	8	1.8	2492
501	*.*.142.244	8	1.8	1993
502	*.*.34.235	8	1.7	3816
998	*.*.154.214	4	4.2	2699
999	*.*.249.210	4	4.2	2244
1000	*.*.156.16	4	4.2	1567

する攻撃パターンの類似する IP アドレスの判別を行う．次に，複数の国に攻撃を行う IP アドレスに対して各位相における各国の自己相関係数の標準偏差を求める．標準偏差とは要素間のばらつきを表す指標であり，この場合特定の攻撃元 IP アドレスから各国への攻撃のばらつきを表すものとする．得られた各攻撃元 IP アドレスの標準偏差の推移を同様に 2 次元の散布図で表現する．

4　実際のダークネットのトラフィックデータに対する解析結果

7 ヶ国がそれぞれ所有するダークネットのトラフィックデータを用いて，観測パケットの解析を行った．国名は解析に扱うデータの都合上，A 国，B 国... といった表記で扱う．データの利用期間を 2015 年 4 月 1 日から 2015 年 4 月 30 日までの 1 ヶ月間として，自己相関と多次元尺度法による解析を行った．解析対象の攻撃元 IP アドレスを選出するにあたり，1 ヶ月間において攻撃が観測された日数が多いものから順に攻撃元 IP アドレスを順位付けした．実際に順位付けしたリストの抜粋を表 1 に示す．また表中の項目内容を以下に示す．観測日数が同じ場合は項目 2 及び項目 3 に着目して順位を判定した．

項目 1： 攻撃が観測された日数
項目 2： 1 日に攻撃が観測された国数の平均
項目 3： 1 ヶ月の合計パケット数

4.1　各国別の攻撃による分析

順位付けした攻撃元 IP アドレスのうち上位 1000 件に対して各国別に自己相関を適用した．ここで自己相関関数の入力とする時系列データの単位を 1 日の合計パケット数とした．さらに導出した自己相関係数の推移に対して非計量多次元尺度法を適用した．得られた散布図を図 4 に示す．図 4 において，各攻撃元 IP アドレスから各国への攻撃が点で表されている．図中の円 a～d に含まれる点が表す攻撃についてそれぞれ隣接する 2 点を抜き出し，自己相関係数をグラフ化したものを図 5～図 8 に示す．4 つの図に示した 8 本の自己相関グラフより，攻撃パターンの類似するものは距離が近く，似ていないものは距離が遠く配置されている．この結果より，複数の IP アドレスから各国に対する同一な攻撃パターンが存在することがわかる．また，円 a に含まれる攻撃はグラフの波形がなめらかであるのに対し，円 c や d に含まれるものははっ

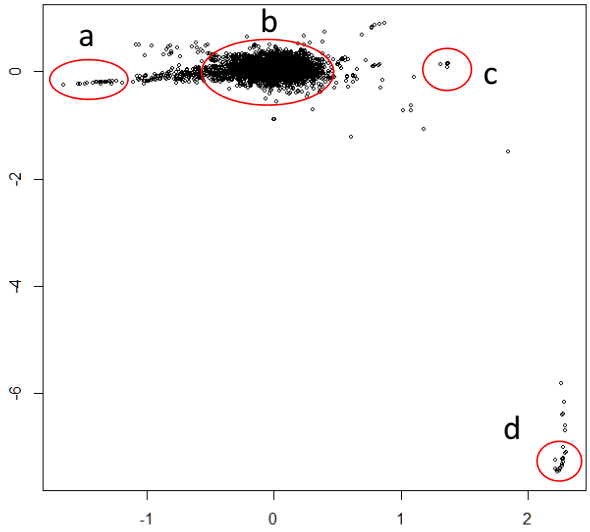

図 4　各国別の攻撃による散布図

きりと周期的に相関係数の増減を繰り返す波形であった．円 b に含まれる攻撃は相関係数の増減は見られるものの，規則性の不明瞭なものが多く見られた．

円 a 及び円 d に含まれる隣接する点が表す攻撃について，図 5 及び図 8 に示すように攻撃パターンが一致している．それぞれの図における両 IP アドレスについて DNS によってホスト名を検索した結果は，どちらも不明であった．また，攻撃元 IP アドレスのレンジは異なっていた．このことから同じマルウェアに感染したホストであるか，同じスキャンツールを使用する攻撃元であると考えられる．さらに円 b に含まれる点が表す攻撃について両 IP アドレスについて DNS によってホスト名を検索した結果，census6.shodan.io 並びに census12.shodan.io であった．どちらも shodan[4] と呼ばれるサーチエンジンによるスキャンであることがわかり，同様のシステムを利用しているため攻撃のパターンも類似していると考えられる．以上より，散布図における点の位置関係によって攻撃元 IP アドレスの攻撃パターンを特徴毎に分類することが可能であるといえる．

4.2　標準偏差による分析

数の国に攻撃を行う IP アドレスに対して各位相における各国の自己相関係数の標準偏差を求め，さらに導出した標準偏差の推移に対して非計量多次元尺度法を適用した．得られた散布図を図 9 に示す．図 9 において，各攻撃元 IP アドレスの攻撃が点で表されている．図中の円 a 及び円 b に含まれる隣接する 2 点について標準偏差の推移をグラフ化したものを図 10 及び図 11 に示す．各図より標準偏差の推移が類似するものは距離が近く，似ていないものは距離が遠く配置されている．図 11 より，両者の標準偏差は小さな値をとっている．図 11 において攻撃元 IP アドレスから各国への攻撃の自己相関グラフを図 12 及び図 13 に示す．図 12 及び図 13 より，各国への攻撃パターンがそれぞれ一致していることからも，図 11 に示した攻撃は各国への攻撃パターンに差がないことがわかる．また，図 10 において攻撃元 IP アドレスから各国への攻撃

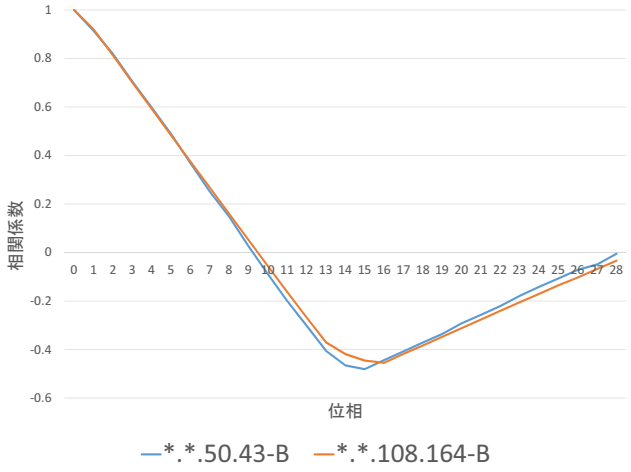

図 5　a の円に含まれる 2 点の攻撃

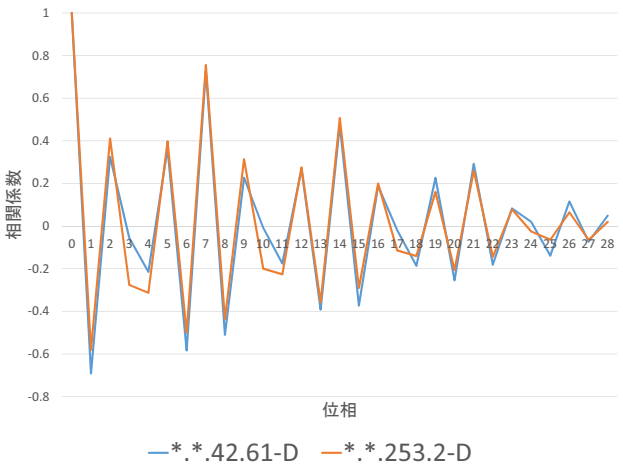

図 7　c の円に含まれる 2 点の攻撃

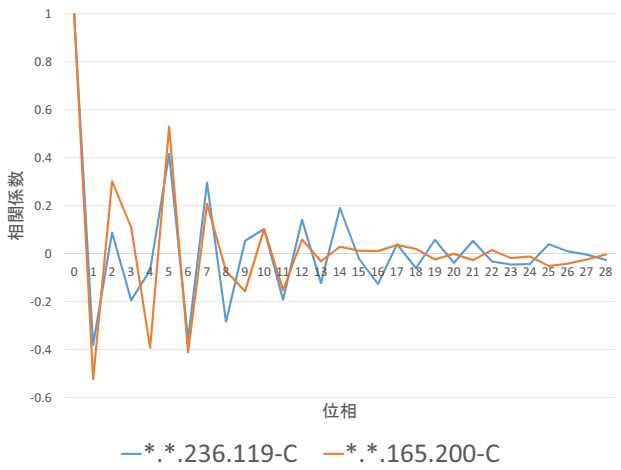

図 6　b の円に含まれる 2 点の攻撃

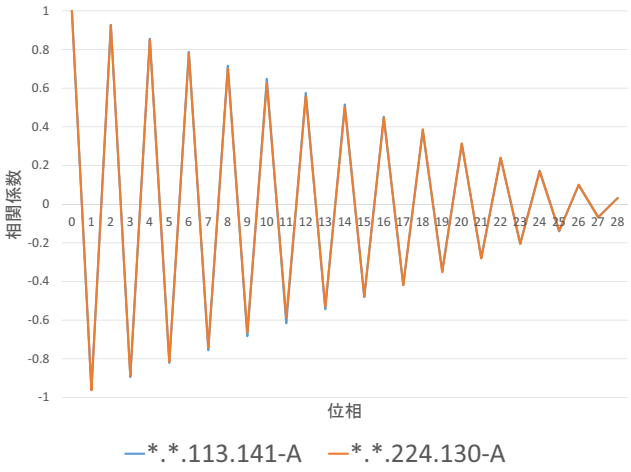

図 8　d の円に含まれる 2 点の攻撃

の自己相関グラフを図 14 及び図 15 に示す．図 14 及び図 15 より，両者の各国への攻撃のばらつきは同程度である．両 IP アドレスについて DNS によってホスト名を検索した結果, census3.shodan.io 並びに atlantic.census.shodan.io であった．このことから，サーチエンジン等による大規模なスキャンを行う IP アドレスは各国への攻撃のばらつきの傾向が類似していると考えられる．以上より，複数国へ攻撃を行う IP アドレスの攻撃のばらつき傾向の特徴を利用して攻撃元 IP アドレス分類をすることができる．

5　ダークネットに関する情報の漏洩

本章では前章の結果も踏まえて，他国のダークネットの規模などの情報の推定に関して述べる．図 16 にダークネットの情報共有における統計データの扱いのモデルを示す．図 16 の流れについて簡潔に解説する．まず自国を含む参加国は所有するダークネットのデータを解析機関へ提供する．次に解析機関は参加国から収集したダークネットのデータをまとめて解析し，統計データを作成する．最後に解析機関は作成した統計データを参加国に公開する．参加国は自国以外のダークネットに関する情報として，公開された統計データのみを参照することができる．また，それぞれ自国のダークネットに関する情報を全て参照できる．図 17 に統計データの例を示す．

第 4 章の解析結果を検証すると，複数国に定常的または周期的に攻撃を行う攻撃元 IP アドレスが実際に存在することがわかる．複数国への攻撃の相関関係が明確な攻撃元 IP アドレスの攻撃パターンが既知のものである状況を仮定する．このとき他国のダークネットの規模の情報に関して，特定の IP の攻撃傾向や他国の特定のポートにおける攻撃割合を示す寄与率を利用して，公開する統計データから推定を行う方式 [5] が提案されている．この方式では多くの攻撃元 IP アドレスの攻撃先のポート番号の傾向は一定であるという傾向を利用している．公開す

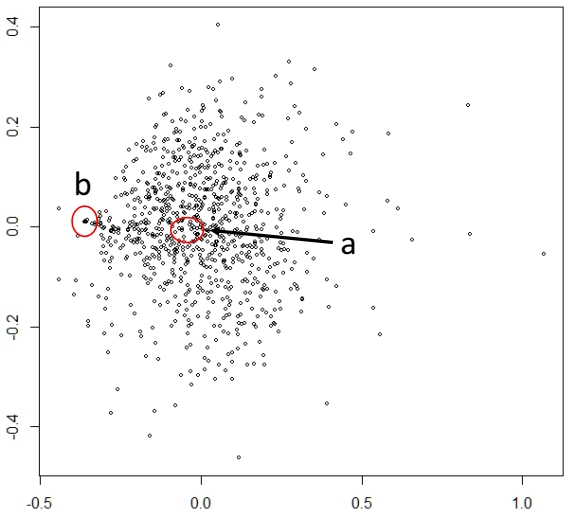

図 9　標準偏差による散布図

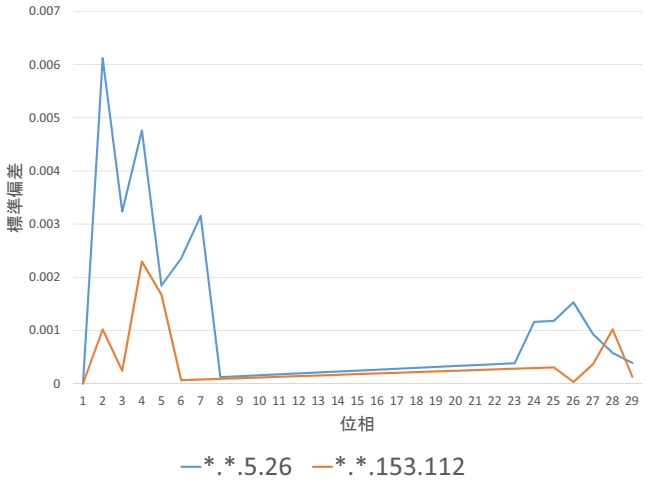

図 11　b の円に含まれる 2 点の攻撃

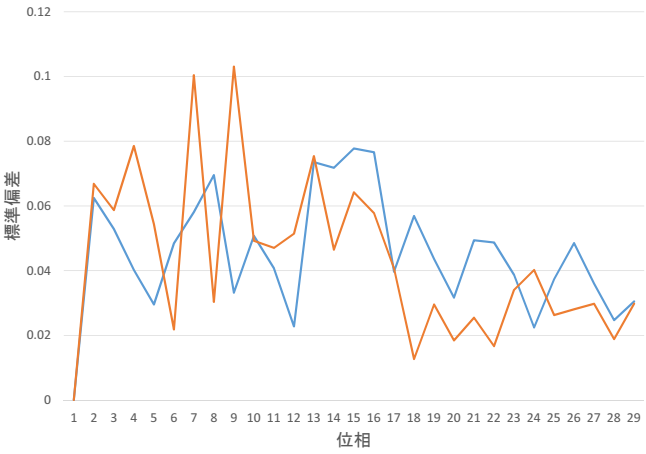

図 10　a の円に含まれる 2 点の攻撃

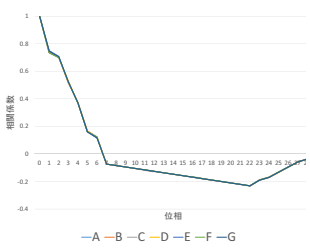

図 12　*.*.5.26 から各国への攻撃の自己相関

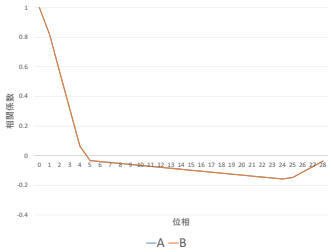

図 13　*.*.153.112 から各国への攻撃の自己相関

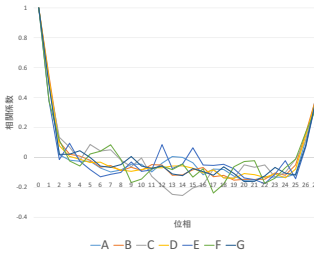

図 14　*.*.70.114 から各国への攻撃の自己相関

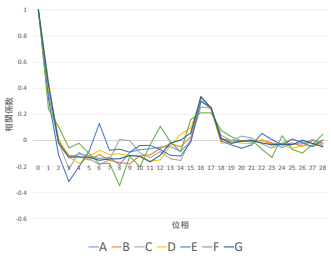

図 15　*.*.9.50 から各国への攻撃の自己相関

る統計データにおける特定のポート番号に対して送られてきたパケット数に着目して，ある攻撃傾向を有している攻撃元 IP のそのポートに占める寄与率を利用して推定するという方式である．ただし，この方式ではある特定の一つの IP アドレスの情報を利用するのみであった．ここで，本稿で述べてきた多次元尺度構成法を利用した解析結果より，ある攻撃傾向を有している特定の IP アドレスを複数クラス分けすることが可能である．これらの同様の攻撃傾向を有する複数の攻撃元 IP アドレスの攻撃パターンを利用することにより，より精度の高い推定を行うことが可能であると考えられる．

6　まとめ

本稿では複数のダークネットのトラフィックデータに対する解析手法の提案と実際の解析結果からの考察を行った．4 章で述べたように自己相関解析や多次元尺度法，標準偏差を利用した解析により，複数のダークネットからなる膨大なトラフィックデータに対しても有益な攻撃傾向を把握することが可能である．また，5 章で述べたように，4 章の解析結果を利用することにより，他国のダークネットの情報の推定にも発展させていくことも可能である．今後は解析における攻撃元 IP アドレスの分類精度の向上を行う．そのためにはダークネットトラフィックデータに含まれる宛先ポート番号などのパラメータを取り入れることが挙げられる．また，向上させた分類法を元により詳細なダークネット情報の推定にも取り組む．

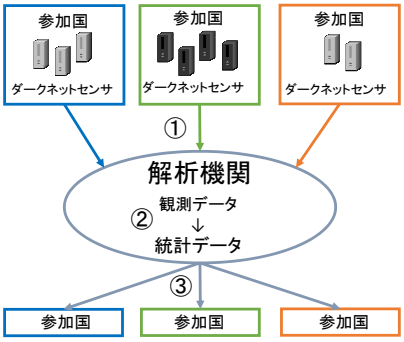

図 16　ダークネット情報共有のモデル

ポート	1日	2日	3日	4日	5日
3128	16363	23685	1328	1399	1303
3306	10852	13184	9846	12750	12669
23	12098	12724	12598	11663	10721
1433	5513	11689	6650	7625	10767
22	5255	5215	5709	6348	5806
8080	3563	4064	2618	3026	3574
3389	3093	3039	2689	3618	3629
9200	1173	169	548	1643	2043
80	1605	2131	1462	3287	2218
445	977	915	1348	957	1418
443	861	1076	1440	1251	904
28537	5	2	4	0	386
53	903	625	1192	937	944
8888	262	12	178	22	278
1900	960	1781	770	1086	1066
137	827	616	892	895	836
123	470	797	514	418	653
62915	0	0	0	0	0
5900	1179	2010	1302	966	1153

28日	29日	30日	31日
59617	43096	38746	7903
4830	4660	3937	4441
12058	9148	7879	8337
10123	18663	11830	7441
5962	8069	8025	8978
2416	1736	2425	2071
3112	1796	1939	2303
1812	2712	2971	2816
2547	1526	3163	2175
1634	1549	1733	2040
1761	859	2047	1397
3199	1381	3575	1837
1227	1149	1110	670
148	270	1485	10389
1695	1217	993	994
993	1269	1271	1175
1353	1046	904	915
7511	4887	5991	1788
401	588	734	775

図 17　公開する統計データ

参考文献

[1] 総務省｜「国際連携によるサイバー攻撃予知・即応プロジェクト『PRACTICE』」, http://www.soumu.go.jp/menu_news/snews/01ryutsu03_02000039.html
[2] 柏井祐樹, 森井昌克, 井上大介, 中尾康二 “NONSTOP データを用いたマルウェアの時系列分析,” *CSS 2013*, 3A3-2, 2013.
[3] R: The R Project for Statistical Computing, www.r-project.org/
[4] SHODAN - Computer Search Engine, www.shodanhq.com/
[5] 村井健祥, 古本啓祐, 村上洸介, 中尾康二, 森井昌克 “複数のダークネットに対するトラフィックデータ解析とその応用,” 電子情報通信学会技術研究報告, 情報通信システムセキュリティ (ICSS) 研究会, Vol.115, No.81, ICSS2015-7, pp.33-38, 2015 年 6 月.