ダークネット観測パケットへの位相的データ解析に関する一検討

柏倉 潤平[†1] 成田 匡輝[†1] 猪股 俊光[†1] 杉野 栄二[†1] 今井 信太郎[†1]

概要：ダークネットと呼ばれる未使用の IP アドレス空間には，不正な通信によるパケットが到着する．それら到着パケットを観測することで，インターネット上で発生している悪性な活動の動向を把握することが可能である．近年，観測される不正なパケットは増加傾向にあり，ビッグデータ解析と同様，大量の不正パケットから攻撃動向を効率的に把握する技術が必要となっている．本稿では，位相的データ解析というトポロジーの考えに基づくデータ解析により，ダークネットへの到着パケットの可視化を試みる．高次元のデータに位相的データ解析による可視化を行うことで，攻撃動向の変化の検出が容易になることを示す．

キーワード：ダークネット，位相的データ解析，可視化技術，異常検知，ネットワークセキュリティ

1. はじめに

近年，ビッグデータが注目を浴びている．これは，インターネットを通して様々なサービスを受けられるようになったためである．特に，IoT（Internet of Things）の普及の影響が大きい. IoT の普及により, デバイスの増加が目立ち，結果的にインターネット上のパケットデータのやり取りも急激に増加している．そのため，インターネット上には大量のパケットデータが溢れている．

また，パケットデータの増加に伴い，その裏で不正なパケットデータのやり取りも目立っている．こうした背景から，ビッグデータ時代におけるネットワークセキュリティの重要性はより高まったといえる．

ネットワークセキュリティに関する情報を収集するため，ダークネット観測により，インターネット上の悪性な活動の動向を把握しようという試みがある. ダークネットとは，インターネット上で到達可能かつ特定のホストコンピュータに割り当てられていない IP アドレス空間のことを指す．ダークネットの IP アドレスには, 使用されていないにもかかわらず，実際には相当数のパケットが到達している．ここには主に, ランダムに IP アドレスを生成し行うポートスキャンや, DDoS 攻撃による不正なパケット等が到着する．そうした不正パケットが到達する理由は，攻撃者が，ある特定のアドレス空間に対し総当たりでパケットを送り，その宛先の IP アドレスの中にホストが割り当てられていないものも存在するためである．つまり，ダークネット観測により届いた不正なパケットを分析することで，悪性な活動，すなわちインターネット上の攻撃の動向を調査することが出来る．

ビッグデータ時代の到来により，ダークネットに到達するパケットデータは今後も急激に増加していくと考えられる．パケットデータが増加することで，ダークネット観測により悪性活動の動向を把握するにも，様々な工夫が必要になると考えられる．1 つは，不正パケット数の増加への対処が挙げられる．不正パケット数が増加することで，解析に要するコストも増大すると考えられる．例えば，多次元のパケットデータを 2 次元への次元削減を行おうと考えた場合，圧縮のための処理が求められる．

もう 1 つは，不正パケットの効率的な検知手法が挙げられる．ダークネット観測における不正パケット数増加に伴い，セキュリティ管理者には，より効率的な悪性活動の検知手法が必要と考えられる．より簡単で分かりやすい検知結果を得ることで，増加した不正パケットへの早急な対応を可能にする必要がある．

以上のことから，ビッグデータ時代に対応したダークネット観測は重要である．

本稿では，ビッグデータ時代のダークネットパケットの解析手法に焦点を当てる．その中で，より簡単で分かりやすい攻撃動向の可視化と，ダークネット観測における攻撃動向の変化を検出することを目的とし，この 2 つの実現手法を提案する．

†1 岩手県立大学大学院ソフトウェア情報学研究科

2. 関連研究

畑中ら[1]の研究では，ダークネット観測によるスキャンパケットから特徴ベクトルを作成し，2 次元への次元削減を行い平面に可視化している．非線形な次元削減手法として，t 分布型確率的近傍埋め込み法（以下 t-SNE 法）を用いている．低次元なデータに対し有効である Isomap や Diffusion maps とは異なり，多次元なデータに有効な手法として用いられる次元削減手法が t-SNE 法である．畑中らは，17 次元のデータに対して適用している．ただし，この t-SNE 法は計算コストが高いという欠点があり，当該研究のような多次元データには必ずしも推奨されていない．

また，梅田[2]は新しいデータ分析手法として，位相的データ解析（Topological Data Analysis，以下 TDA）を調査している．位相幾何学（トポロジー）は，データの形を捉える分析手法であり，数学の考え方を導入した手法である．位相幾何学の元となる幾何学は，高次元を含めた図形を数値や数式を用いて把握する学問である．これまでデータ分析手法として主流となってきた統計学的手法とは異なるアプローチといえる．

インターネット上のデータが大量かつ複雑になる中，従来の手法では詳細な分析が難しくなってきている．そこで着目されたのが，この TDA である．ビッグデータを焦点とし，データの形を捉えることで新たな知見を取り出そうという新しい技術である．特に，ここでは位相幾何学の分野におけるモース理論を元にした TDA における Mapper 技術（以下 TDA Mapper）を用いる．

モース理論とは，図形の特徴を捉える理論である．図形の中で特性が変わる部分を臨界点と呼び，この臨界点を見ることで，図形の特徴を捉えようと試みる．TDA Mapper[3]は，データの臨界点付近のデータをまとめて 1 つのノードとし，連続したデータのあるノード間をエッジで繋ぐことで，データの集合をグラフに変換する技術である．TDA Mapper は，たとえ元のデータの次元が大きい場合でも 2 次元で可視化が可能となっている．

実際に梅田は，3 次元空間上に 3 つの混合ガウス分布に従って発生させたデータに対し，統計学的手法による分析結果を示している．主成分分析，カーネル主成分分析，Isomap，Autoencoder の 4 手法において，2 次元への次元圧縮が示されている．混合ガウスモデルでは，適切なガウス分布の数を決定するために膨大な計算時間を必要とすることがある．また，複雑な確率密度関数の場合，ガウス混合モデルで表現することが難しい場合がある．実際の結果として，どれを取ってみても，3 つの混合ガウス分布から構成される 2 次元データであると判別することは難しい結果となっている．

ここで，TDA Mapper による分析を試みる．TDA Mapper では，3 つの特徴的なノードが表れている．

Marc ら[4]の研究では，実際にダークネット観測パケット解析に TDA Mapper の適用を行っている．これまで TDA をダークネット観測の可視化，ひいてはパケットデータへ応用している例はなく，当該研究が初となっている．ここでは，6 次元データからの簡易的な可視化を実行している．ダークネット観測により得られたパケットデータを加工せずにグラフ化すると，非常に乱雑で，特徴を読み取ることが困難なグラフが生成されてしまう．しかし，TDA Mapper を用いることで，データの特徴的な部分のみを残したグラフを生成し，データの特徴を読み取ることが可能となっている．これにより，パケットデータへの TDA Mapper の有用性が示されている．

しかしながら，Marc らの研究ではパケットデータへの TDA Mapper 適用が有用である可能性を示すのみとなっており，従来手法との性能評価などは行われていない．また，時間の流れを考慮していないグラフが示されただけとなっている．

3. 研究概要

本研究では，ビッグデータ時代に対応したダークネットパケットの解析において，TDA Mapper の適用を提案する．これにより，不正パケットの動向をデータの形として分かりやすく捉え，これからのパケット分析手法の 1 つとして有用であることを示す．

まず，ダークネット観測により得られたパケットデータを TDA Mapper に適用する．その解析結果として得られた描画データから，特徴的な形を取り出し，不正パケットによる悪性な活動の抽出を試みる．

TDA Mapper は，入力されたベクトルデータをフィルタにかけクラスタリングすることで解析を行い，2 次元のデータとして描画する．入力データには独自に設計したベクトルデータを用意し，フィルタ，クラスタリングのパラメータを操作することで，よりパケットデータ分析に適した TDA Mapper の運用を目指す．TDA Mapper の処理フローを図 1 に示す．

図 1　TDA Mapper の処理フロー

4. 提案手法

Marc らの手法と同様に，TDA Mapper によるダークネット観測のパケットデータの可視化を試みる．本研究では，独自に設計したダークネット観測パケットのベクトルデータを TDA Mapper に適用する．これにより，新たなダークネット観測の解析手法を提案する．

表 1　本手法で使用したベクトルデータの例

	攻撃者IPアドレス(10進数)	送信総パケット数	宛先IPアドレス数(MAX:4096)	宛先IPアドレス数毎のパケット数の分散	TCP:送信元ポート番号の総数	TCP:送信元パケット分散	TCP:送信先ポート番号の総数	TCP:送信先パケット分散	UDP:送信元ポート番号の総数	UDP:送信元パケット分散	UDP:送信先ポート番号の総数	UDP:送信先パケット分散
1	1571708296	135973	4096	4.378611028	2	3277729203	0	0	39	269503.0703	0	0
2	3575646025	108873	4096	23.73085302	27666	3.602944297	0	0	1	0	0	0
3	2745977282	106478	4096	0.004375219	1	0	0	0	26	0.982248521	0	0
4	1539697418	81411	4096	3.165954053	3	292737708.7	0	0	23	384060.8469	0	0
5	1541375240	74046	4096	1.463699102	1	0	0	0	20	168942.91	0	0
6	2513054034	62526	39	31533290.74	1	0	0	0	34	35741837.06	0	0
7	3395435301	39429	28	1036851.004	1	0	0	0	2	2493398903	0	0
8	1541375245	39110	4096	6.18906951	1	0	0	0	61	365592.7815	0	0
9	3109973761	36727	4096	0.146586359	7140	0.152801042	18048	0.665414833	2	0	8	1573940.359
10	3496221412	36724	4096	7.101859093	2	245674276	0	0	55	555172.9336	0	0
●												
●												
●												
990	2379512453	478	394	0.167744595	475	0.0062759	0	0	2	8281	0	0
991	2379512456	478	386	0.181535075	474	0.008367605	0	0	2	8281	0	0
992	2379512457	478	388	0.178153895	476	0.004184027	0	0	2	9025	0	0
993	3638218332	478	451	0.056282909	473	0.010459082	0	0	10	994.96	0	0
994	1492744908	477	477	0	1	0	0	0	1	0	0	0
995	2379512454	477	388	0.176765597	475	0.004192798	0	0	2	8372.25	0	0
996	1191630906	476	476	0	4	2296.5	0	0	4	2296.5	0	0
997	1191630903	475	475	0	4	2338.1875	0	0	4	2338.1875	0	0
998	2086225768	475	471	0.008420445	7	17521.26531	0	0	3	28140.22222	0	0
999	758580833	474	474	0	469	0.010547324	0	0	1	0	0	0
1000	3247902089	474	15	260.5066667	1	0	0	0	2	34596	0	0

4.1 ベクトルデータ

ベクトルデータの元となるパケットデータは，NICTER が NONSTOP 上で提供するダークネット観測のデータセット[5]を用いる．このデータセットを加工することで，ベクトルデータの作成を行う．また，ダークネット観測パケットのベクトルデータ作成には，2017 年 6 月 1 日から 2017 年 6 月 30 日までに観測された 1 日毎の合計 30 日分のパケットデータを用いた．

ベクトルデータ作成には，ダークネット観測パケットから 12 個の特徴を抽出した．以下にその特徴を示す．

- 攻撃者 IP アドレス
- 攻撃者の送信総パケット数
- 宛先 IP アドレス数（MAX 値：4096）
- 宛先 IP アドレス数毎のパケット数の分散
- TCP：送信元ポート番号の総数
- TCP：送信元ポート番号毎のパケット分散
- TCP：送信先ポート番号の総数
- TCP：送信先ポート番号毎のパケット分散
- UDP：送信元ポート番号の総数
- UDP：送信元ポート番号毎のパケット分散
- UDP：送信先ポート番号の総数
- UDP：送信先ポート番号毎のパケット分散

パケットデータの中でも，特に攻撃において意義のある特徴を抽出した．本手法では特に攻撃者に注目し，攻撃者の IP アドレスを重要視している．

以上の 12 次元の特徴を抽出し，攻撃者の送信総パケット数が多い順に上位 1000 件までのデータを取得する．12 次元 1000 行のベクトルデータの例を表 1 に示す．

4.2 TDA Mapper による処理

TDA Mapper による処理として，フィルタとクラスタリングのパラメータ操作を行う．TDA Mapper のフィルタとクラスタリングには，様々なパラメータの組み合わせが考えられる．このパラメータを操作し，様々な組み合わせを試すことで，よりダークネットパケット解析に適した解析結果を検討する．

4.3 描画結果の評価方法

TDA Mapper による解析で得られた描画結果から，主に相対的な比較を行うことで，描画結果の評価を試みる．ただし，現状では主観的な比較に留まるが，今後の課題として，目的関数を定めて定量的な評価方法を検討する．

描画結果の評価にあたって，4 つの比較方法を挙げる．以下にその比較方法を示す．

1. 類似した描画結果同士の比較
2. 類似した描画結果から逸脱した描画との比較
3. 逸脱した描画結果同士の比較
4. 統計学的手法による解析結果との比較

ここでは，TDA Mapper により得られた描画結果の比較により，より簡易的な評価を行う．類似した描画結果同士からは，同類のパケットデータの特徴を有していると考えられる．また，そういった類似性から逸脱した描画結果となったパケットデータには，なんらかの異常，もしくはそれに近い特徴を持つと考える．その 2 つの異なる特徴を持つパケットデータを解析することで，攻撃動向の変化もしくは攻撃の特徴を捉えることを試みる．

5. 実験

ここでは実際に，ダークネット観測により得られたパケットデータをベクトルデータに変換し，そのベクトルデータを TDA Mapper へと適用する実験を行う．

5.1 評価項目

実験の評価項目として，次の 3 つを挙げる．

1. パケットデータの動きを描画可能か
2. パケットデータの異常な動きを，通常の動きと比較して，識別可能か
3. 既存の統計学的手法では識別が困難であった特徴を取得可能か（今後の課題）

5.2 実験内容

ダークネット観測パケットをベクトルデータに変換し，TDA Mapper に適用する．そして，TDA Mapper による解析の結果として描画されたグラフの比較を行う．

使用するベクトルデータは，2017 年 6 月 1 日から 2017 年 6 月 30 日まで 1 日毎に分け，それぞれ 30 パターンとして TDA Mapper に適用する．

次に，TDA Mapper の処理として，フィルタとクラスタリングのパラメータ操作を行う．フィルタには Eccentricity，クラスタリングには Complete を適用する．TDA Mapper の処理内容を図 2 に示す．

最後に，TDA Mapper の解析により得られた描画結果を評価する．

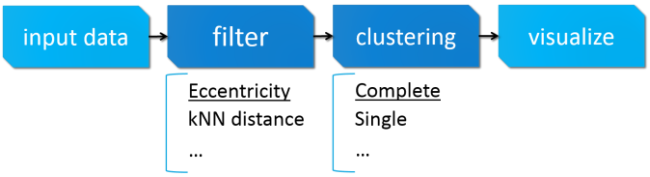

図 2　TDA Mapper の処理内容

5.3 実験結果

ダークネット観測パケットに TDA Mapper を適用した解析により得られた描画結果の評価を行う．

まずは，2017 年 6 月 1 日から 2017 年 6 月 30 日までの 30 パターンの TDA Mapper 解析結果から，特徴的な解析結果を取り上げる．2017 年 6 月 1 日観測パケットの解析結果を図 3 に，2017 年 6 月 12 日観測パケットの解析結果を図 4 に，2017 年 6 月 16 日観測パケットの解析結果を図 5 にそれぞれ示す．

ここで取り上げた 3 つの解析結果は，主観的ではあるが，それぞれ類似した特徴を持っていることが分かる．多くの解析結果は，これら 3 つの解析結果と同様に，類似した特徴を持った 2 次元データで表されている．ダークネット観測パケットのそれぞれの日付の元データと比較してみても，

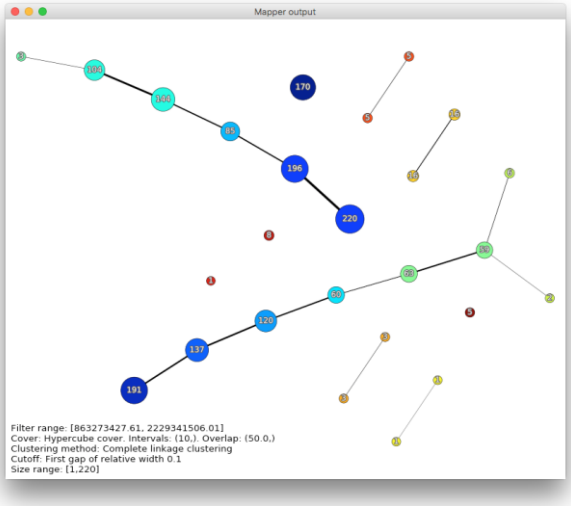

図 3　2017 年 6 月 1 日観測パケットの解析結果

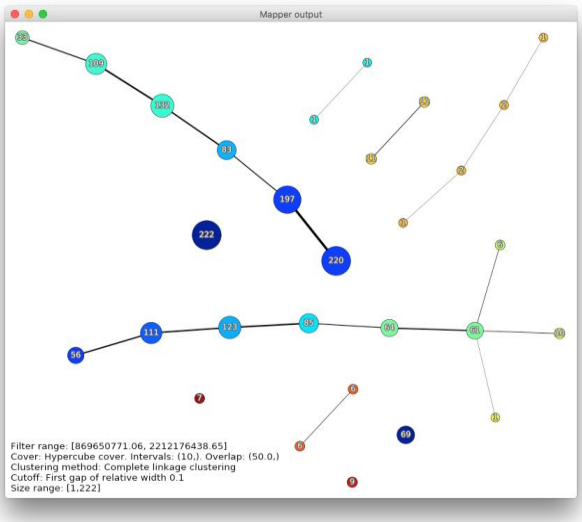

図 4　2017 年 6 月 12 日観測パケットの解析結果

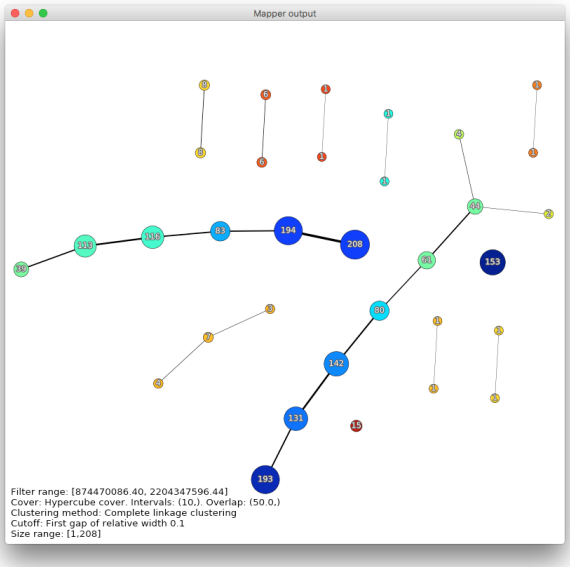

図 5　2017 年 6 月 16 日観測パケットの解析結果

大きな差異を見つけることは困難であった．

一方，2017 年 6 月 14 日観測パケットの解析結果に関しては，唯一他の解析結果とは明確に異なる結果となっていた．2017 年 6 月 14 日観測パケットの解析結果を図 6 に示す．

図 3 から図 5 までに挙げたダークネット観測パケットの解析結果のグラフと比較してみても，とりわけ異質な解析結果として描画されていることが分かる．これを別日のダークネット観測パケットの元データと比較してみると，2017 年 6 月 14 日観測パケットは，攻撃者の送信総パケット数が他と比べて多く集中していることが明らかとなった．しかし，現時点では，これが明確に異なる解析結果の要因とまでは断定するには不十分である．

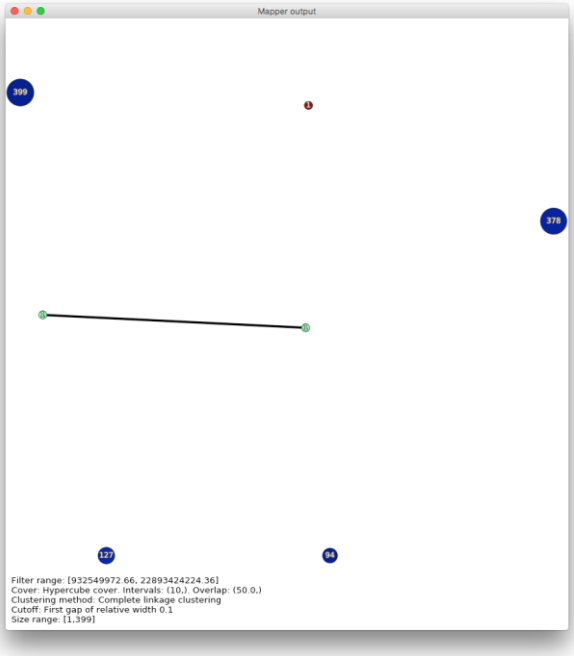

図 6　2017 年 6 月 14 日観測パケットの解析結果

6. 考察

実験結果は，ダークネット観測のパケットデータの動きを，TDA Mapper の適用で描画することができたと考える．つまり，ダークネット観測パケットの解析手法として TDA Mapper の適用の可能性を示すことができた．

類似した結果を示したダークネット観測パケットの解析結果は，それぞれ類似した特徴を有しているためと考えられる．そうでなければ，同様な描画として膨大なパケットデータの解析結果が，偶然に重なるとは考えにくい．また，明確に異なる特徴を示したダークネット観測パケットには，パケットデータの動きの特異点となる特徴を持つ可能性が考えられる．ただし，要因は不明確なため，今後の課題として，評価方法の再検討が必要である．

7. おわりに

本稿では，ダークネット観測パケットに対して，位相的データ解析（TDA）における TDA Mapper といわれる解析手法を適用し，ダークネット観測パケットへの TDA Mapper 解析手法の有用性を提案した．提案手法では，ダークネット観測パケットから独自のベクトルデータを生成し，TDA Mapper の処理を通して適用を試みた．

実験結果では，ダークネット観測パケットの動きを，TDA Mapper により 2 次元データに描画することに成功した．これはダークネット観測パケットの，より簡単で分かりやすい新たな解析手法となりうる可能性を示すことができたと考える．また，解析結果により特徴的なデータの形を捉えたことで，今後のパケットデータにおける TDA Mapper 適用の重要な先駆的事例になると考える．

今後の課題として，より簡単で分かりやすい解析結果を目指し，TDA Mapper のパラメータ操作を試行する必要がある．そのためには，ダークネット観測パケットのベクトルデータ設計の再検討も挙げられる．また，解析結果の根拠や評価基準といった，定量的な判断方法の検討も必要と考えられる．

謝辞

貴重なデータセットを提供してくださった，MWS2017 実行委員会，ならびに NICTER 運営関係者に感謝する．

参考文献

[1] 畑中拓哉 他，”ダークネットトラフィックの可視化とオンライン更新によるモニタリング”，Computer Security Symposium 2016，pp.11−13，Oct. 2016.

[2] 梅田裕平((株)富士通研究所)，”データの形が教えてくれること ートポロジカル・データ・アナリシスとその応用ー”，情報処理 Vol.57，No.11，Nov. 2016.

[3] G. Singh, F. Memoli, and G. Carlsson, “Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition”, in Eurographics Symposium on Point-Based Graphics. The Eurographics Association, 2007.

[4] Marc Coudriau, Abdelkader Lahmadi, Jerome Francois, “Topological Analysis and Visualisa-tion of Network Monitoring Data: Darknet case study” 8th IEEE International Workshop on Information Forensics and Security - WIFS 2016, Abu Dhabi, United Arab Emirates. IEEE, 2016, Information Forensics and Security, Dec 2016.