2023 年 3 月 31 日

一般財団法人　日本産業科学研究所
理事長　宮地 尚　様

下記の通り，一般財団法人日本産業科学研究所研究助成金による研究実績を報告致します．

岩手県立大学　ソフトウェア情報学部
ソフトウェア情報学科
准教授　成田 匡輝

令和 4 年度　研究助成実績報告書

研究課題名：ビッグデータ化した攻撃パケットデータを視覚化する情報システムの開発

・研究背景と目的

セキュリティ対策が不十分な情報サービスがサイバー攻撃の標的となり，大量の個人情報が外部に流出するなどの被害が連日発生している．ダークネット観測システム（図 1）は，インターネット上で未割り当ての IP アドレス空間に配備した観測点に到着する攻撃パケットを捕捉し，善良なインターネット利用者に注意喚起・セキュリティ対策情報を迅速に提供するためのシステムである．

インターネット上を流れる攻撃パケットは増加し，もはやビッグデータとなった．ビッグデータ化した攻撃パケットに埋もれている未知のサイバー攻撃を検出し，攻撃対策に活かす需要は確実に高まるであろう．ビッグデータが活用される現代，このままではビッグデータに埋もれた重要なサイバー攻撃に関する知見が見逃される可能性がある．

そこで本研究は，攻撃パケットデータに位相的データ解析を適用し，膨大な攻撃パケットデータに含まれるサイバー攻撃を視覚的に把握・分析できる，データの形に着目する情報提供システムの開発を目的とする．位相的データ解析とは，柔らかい幾何学ともよばれる．この解析手法は，位相幾何学の考え方を応用し，膨大なデータの特徴を直感的に理解しやすい視覚的な形として表現する手法である．

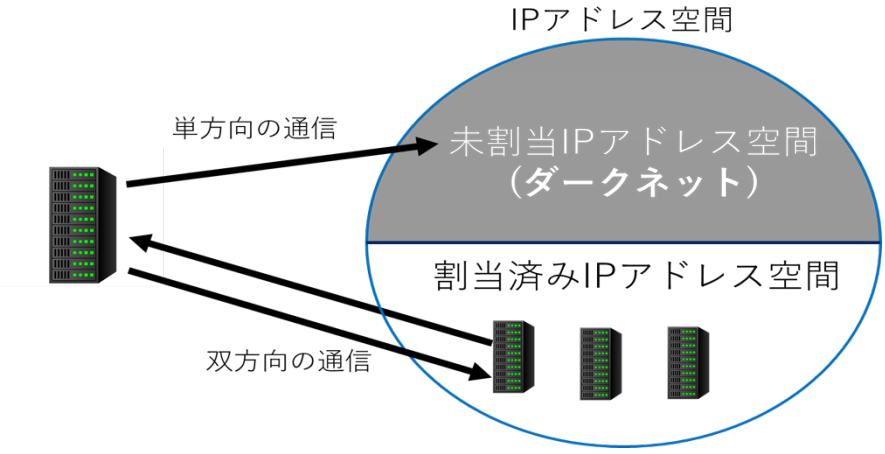

図 1 ダークネット観測システムの概要

・研究成果 1　攻撃パケットの傾向変化の分析

インターネットに接続する機器の増加に伴い，悪意を持ってインターネット上に送出される攻撃パケットも増加の一途を辿っている．一般家庭にもインターネット接続されたカメラやスマートスピーカーなど，いわゆる IoT デバイスが普及してきた．

攻撃者は，攻撃対象の機器に攻撃を成功させるため，1 つのサービスポートを標的とするのではなく，複数のサービスポートを同時に攻撃対象とする戦略（マルチベクタ攻撃）を取ってきている可能性がある．そこで仮説ではあったが，以前よりも近年のほうがその傾向が強くなってきていると考えた．そこで，本研究ではまず，インターネット上に送出されている攻撃パケットの動向がここ 10 年程度でどのように変化しているか，複数のサービスポートを標的とした攻撃の割合はどれくらいであるかといった，調査を行なった．

本研究室では，国立研究開発法人情報通信研究機構(NICT)の協力のもと，前述の組織が運用しているダークネット観測システム(nicter)で観測された一部のパケットデータを研究利用可能である．本研究では，このパケットデータを調査対象とした．

まず，年別マルチベクタ攻撃の割合 (図 2) の変化であるが，2012 年という 10 年以上前に比べ，2017 年，2022 年では，明らかに複数のサービスポートを標的とした攻撃が猛威を振るっていることが分かる．一方，そういう状況であっても図 3 で示したように，5 種類程度のサービスポートを監視すれば，全体の攻撃の約半分をカバーできることが分かる．すなわち，図 2，図 3 が示すところは，今後攻撃パケットのビッグデータを解析対象とする場合，複数のサービスポートが攻撃対象とされているということを念頭に置く必要があることを示唆している．

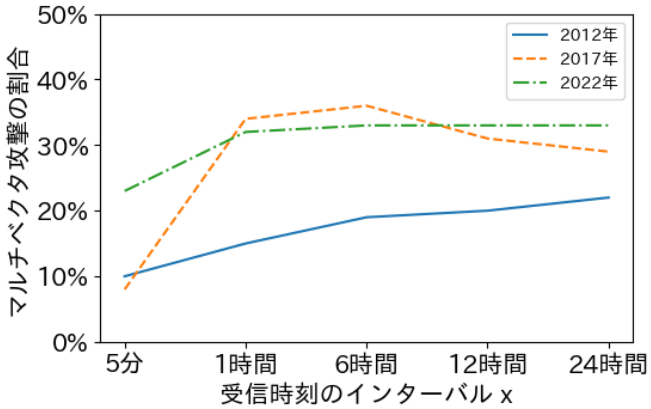

図 2 年別マルチベクタ攻撃の割合の変化

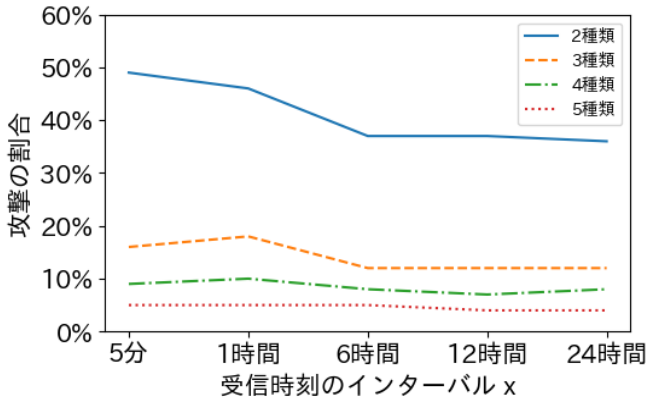

図 3 標的にされるサービスポートの数の割合

なお，表 1 に示したのは，具体的にどのような複数のサービスポートが標的とされているかを上位 5 件まで示したものである．どの時刻を切り取っても 80 番，443 番ポートといった，通常ウェブページ公開に利用するためのポート番号(http，https)の組み合わせが最も多く，この傾向は，一貫して変化していない．民間企業や政府機関をはじめとして，各々のホームページへの妨害が最も多いという結果である．

表 1 標的にされるサービスポートの組み合わせ Top5

	受信時刻のインターバル x				
	5分	1時間	6時間	12時間	24時間
1	80, 443 (0.39%)	80, 443 (1.02%)	80, 443 (2.41%)	80, 443 (2.87%)	80, 443 (3.15%)
2	9916, 9918 (0.27%)	9916, 9918 (0.56%)	9916, 9918 (1.20%)	9916, 9918 (1.51%)	9916, 9918 (1.71%)
3	22, 80 (0.11%)	22, 80 (0.32%)	22, 80 (0.84%)	22, 80 (1.01%)	22, 80 (1.10%)
4	22, 443 (0.06%)	22, 443 (0.14%)	22, 443 (0.31%)	22, 443 (0.40%)	22, 443 (0.45%)
5	9917, 9918 (0.06%)	22, 53 (0.13%)	22, 53 (0.25%)	9917, 9918 (0.30%)	9917, 9918 (0.32%)

また，解析対象とするビッグデータのデータセットを作成するため，他に着目すべき（抽出すべき要素）があるかを検討した結果，パケットの送信元に着目するという考えに至った．図 4 は，国別に前述のマルチベクタ攻撃を散布図で表現したものである．

パケットの収集時期にも結果は依存するが，図 4 は 2022 年の結果である．左下に偏ったデータとなってはいるが，全体に広がる興味深い分類結果も見られ，パケットが送信されてきた国別に視覚化を行うことも一定の成果に繋がると考えられる．

収集したパケットの中には，ウクライナを標的とした攻撃パケットも含まれており，このような戦時下におけるサイバー攻撃の情報を収集できたことは，パケットが発出された国の情報もラベルとしてデータセットに組み込む価値があるという結論となった．

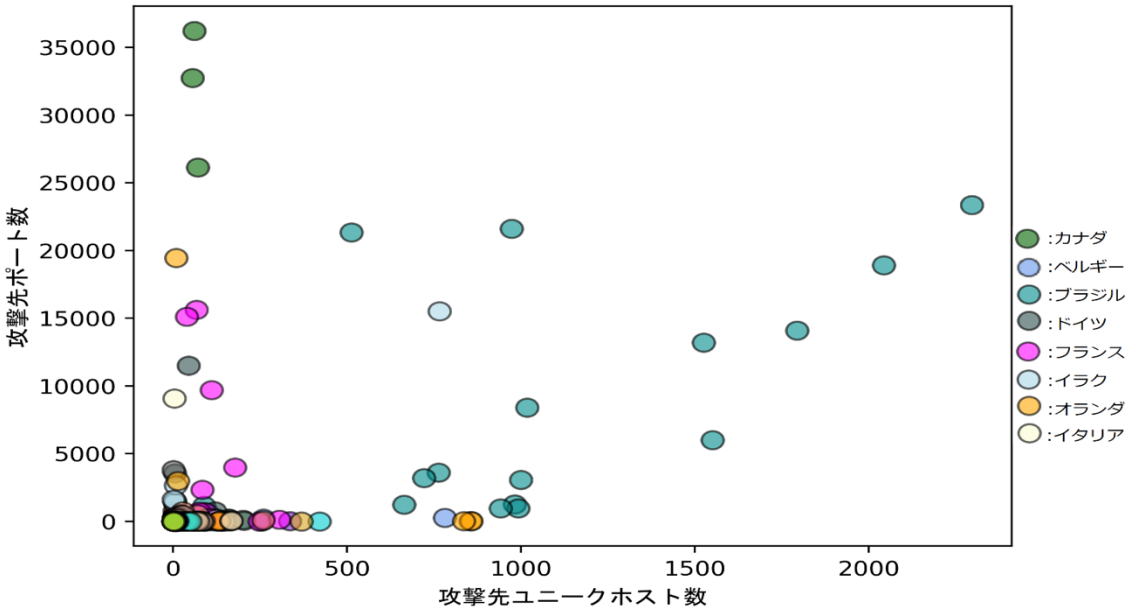

図 4 マルチベクタ攻撃を国別で分類した散布図

・研究成果２　ビッグデータ化した攻撃パケットデータセットの作成

次に，研究成果 1 で得られた知見を基に，攻撃パケットデータセットの作成を行なった．攻撃パケットの収集環境は，国立研究開発法人情報通信研究機構(NICT)の協力のもと，前述の組織が運用しているダークネット観測システム(nicter)を遠隔利用するシステム NONSTOP である．本研究の前段階の検証では，下記の属性を攻撃パケットの収集対象としていた．

- タイムスタンプ（UNIX 時間）
- 送信元 IP アドレス（32 ビットの整数値）
- 送信元ポート番号（16 ビットの整数値）
- 宛先ポート番号（16 ビットの整数値）
- プロトコル番号（8 ビットの整数値）

これらの属性でも一定の成果は得られていたが，今後の攻撃の多様化を想定する必要があると考えた．具体的には，研究成果 1 の成果からこれに加えて送信元の国別コードや，ICMP パケットのタイプやコードといった，可能な限り多くの情報を含めた攻撃パケットデータセットを作成した．すなわち，パケットの送信元のポート番号，送信先のポート番号，国別コードをより確実に収集することで，マルチベクタ攻撃の想定や，国という観点からリアルタイムな攻撃動向の把握が可能になると考えられる．また，他の国で発生した攻撃が時間差で我が国に波及するといったインシデントにも対応できるであろう．

表 2 は，本研究で作成した攻撃パケットのデータセットのラベル構成である．TCP，UDP，ICMP といったインターネット上で主要なプロトコルに対し，6 次元のデータとした．このデータセットを位相的データ解析による視覚化を行う情報システムへの入力とすることで，ビッグデータ化した攻撃パケットを分析する．

表 2 作成した攻撃パケットデータセットのラベル構成

プロトコル	ラベル 1	ラベル 2	ラベル 3	ラベル 4	ラベル 5	ラベル 6
TCP	受信時刻	送信元 IP	宛先 IP	送信元国	送信元ポート	宛先ポート
UDP	受信時刻	送信元 IP	宛先 IP	送信元国	送信元ポート	宛先ポート
ICMP	受信時刻	送信元 IP	宛先 IP	送信元国	TYPE	CODE

また本研究では，インターネット上のダークネットで観測された攻撃パケットの他に，ローカルにインターネットから隔離した仮想環境を構築し，マルウェアの持つ特徴や動作ログなども収集して簡易データベースとした．こちらは，今回実際に観測された攻撃パケットとマルウェアとを照合するといった利用が考えられる．

・研究成果 3　位相的データ解析による攻撃パケットデータの視覚化

本研究 3 つ目の研究成果は，研究成果 2 で作成したデータセットの開発中のシステムへの入力である．下記の図 5 は，研究成果 2 で定義したラベル構成の攻撃パケットデータセットを開発中のシステムに入力・結果を出力した例である．2013 年の出力結果は，最新の 2022 年の出力結果と比較すると，構造が非常に単純である．これは，入力としたデータセット全体のパケット数が少なく，攻撃自体が少ないことによるものである．そして，2017 年，2022 年の出力結果を順に見ていくと，そのグラフ構造が次第に複雑化していくことがわかる．これは，攻撃パケット数の増加に伴い，攻撃数だけでなく，攻撃の種類や手法が多様化していくためである． これは現代のサイバー攻撃の増加・多様化を視覚的に反映しており，本研究が目的とする情報システムの開発方針の妥当性を補強する根拠となる．

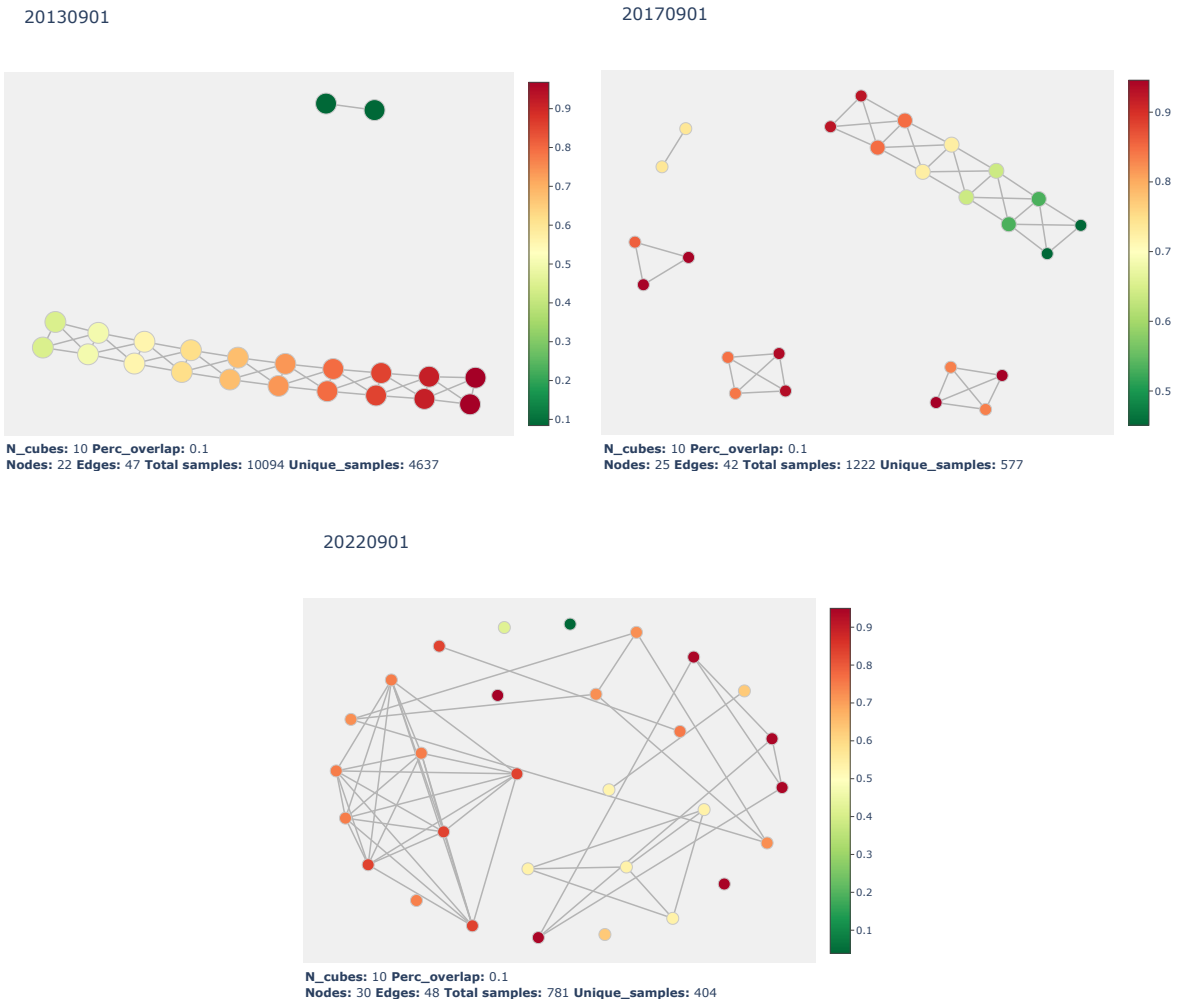

図 5 新たなラベル構成で攻撃パケットデータを開発中のシステムに入力した場合の出力例（左上は 2013 年，右上は 2017 年，下は 2022 年の 9 月 1 日のデータセットを入力とした）

現在は出力結果の詳細な分析を行なっている途上であり，順次進めていく予定である．一方，新たな問題も明らかとなってきた．近年の解析対象とする攻撃パケットデータ量を扱うにあたり，開発に使用している計算機性能に限界があり．時間効率のよい分析が出来ていない．より高性能な計

算機を調達する必要があると考えている．また，本システムが実際にサービスを提供する場合は，そうした高性能な計算機が位相的データ解析を行う役割を担い，そこで得られた出力結果を別のウェブサーバーに転送・公開するといった運用形態が望ましいであろう．

・達成と今後の展望

ビッグデータ化した攻撃パケットデータを視覚化する情報システムの開発という目的を達成するため，インターネット上に現在も送出され続けている攻撃パケットについての過去から現在までの攻撃傾向の変化を把握することができた．インターネット上に送出される攻撃パケット数は，年々増加傾向にあり，ビッグデータ化し続けていることも実際に確認した．攻撃者も複数のサービスポートを同時に攻撃するなど，その攻撃範囲を拡大させている．

こうした状況を受けて，開発中のシステムへの入力データには，新たなラベルとしてパケットの送信元国の情報を付与した．これにより，時間差で我が国に波及しうるインシデントにも対応できるようになるであろう．

一方，現状の開発環境では，入力となる攻撃データセット全体を一度に読み込めず，解析の時間効率が低下している．より高性能な計算機の導入の必要性，処理対象のビッグデータのサイズに応じて読み込むデータ量を変動させる工夫など，システム開発における新たな課題も明らかとなった．

・謝辞

本研究プロジェクトを遂行するにあたり，このような貴重な機会を与えていただいた一般財団法人日本産業科学研究所の皆様，審査に携わった方々に心より感謝申し上げます．