2015年度 修士論文

スロースキャンの偵察行為の特徴分析

提出日：2016年2月1日

指導：後藤滋樹教授

早稲田大学大学院 基幹理工学研究科 情報理工・情報通信専攻
学籍番号：5114F008-4

池西 大起

目次

図一覧

表一覧

第 1 章

序論

1.1 研究の背景

インターネットが通信のインフラとして普及し、様々な恩恵を現代社会にもたらすようになった。その反面、インターネットの普及に伴いネットワーク上の弱点を突いた悪意のある通信が増加している。安定したインターネットのネットワーク攻撃の大半は、攻撃する対象を目的とした偵察行為から始まる。したがって、偵察行為を早期に発見することがホストに対しての攻撃を防ぐための鍵となる。また攻撃が始まっている場合は、その攻撃に付随する偵察を突き詰めることが攻撃をつなぎ合わせる最初の作業である。

このようなスキャン攻撃は、ネットワーク攻撃における第一段階であり、「偵察行為」と呼ばれる。攻撃を防御する側では、この偵察行為の段階の攻撃を検出できれば、攻撃が差し迫っているという警告を早期に把握できるという重要な役割を果たす。特に悪用可能なサービスは、ほんの数秒もあれば攻撃できる。という弱点がある。

多くの偵察行為に対しては，IDS (Intrusion Detection System) に代表される侵入検知システムを用いて検出することが可能である。しかし、IDS ではスロースキャンと呼ばれるパケット数や帯域を制限したスキャンは正常なトラヒックとの区別が困難である。つまり、IDS による検出が難しい。そのため、スロースキャンの検出が課題となっている。

1.2　研究の目的

節 1.1 で、偵察行為の段階での攻撃の検出が重要であることと、IDS によるスロースキャンの検出が難しいということを述べた。通常のスキャンを IDS により検知し、さらにスロースキャンを検知することができれば、より安全なネットワーク環境が実現できる。本研究では、既存研究 [1] の手法により、実ネットワーク上を流れるトラフィック上からスロースキャンと思われる IP アドレスを抽出し、その IP アドレスの特徴分析を行い、スロースキャンの行動把握を行うことを目的とする。

1.3　論文の構成

本論文は以下の章により構成される。

第 1 章 序論

本論文の概要を述べる。

第 2 章 偵察行為

本研究の対象としている、偵察行為の概要および方法について解説する。

第 3 章 提案手法

本研究の提案手法について説明する。

第 4 章 スロースキャンの抽出実験

提案手法を用いて、スロースキャンを行っているパケットを抽出し、その結果を提示する。

第 5 章 スロースキャンの特徴分析

第 4 章のスロースキャン抽出実験で、スロースキャンと判定されたパケットを対象として特徴分析を行う。

第 6 章 結論

本論文のまとめおよび今後に残された課題について述べる。

第 2 章

偵察行為

2.1 偵察行為の概要

攻撃者は何をターゲットとして攻撃するのか、また、対象となるホストに対してどのような攻撃を仕掛けるのかを決定するためにあらかじめ下調べを行う。この行為を、偵察行為をいう。クラッキングを行う攻撃者は、少ないアクセス数によって多くの情報を得ることを目的として、以下の 3 段階を順次実行する。

- 偵察行為段階
  悪意を持つ攻撃者は、ターゲットとなるホストやネットワークに対して、ターゲットのホストやネットワークに関する情報を収集する。この情報を元にして、ターゲットのネットワークやホストの情報をマッピングする。偵察行為は、少ないアクセス数でより多くの情報を得ることを目的とするため、何度も同じ情報を調査しないという特徴がある。

- 権限取得段階
  攻撃者は偵察行為の段階で目的となるホストを絞り、ホストにアクセスして特権アカウント [13] の取得を試みる。ここで権限取得に成功した攻撃者は、次の不正攻撃の段階へ進む。また、バックドアを設置することにより次回以降も容易にアクセスできるようにしたり、不正アクセスに気付かれた場合の抜け道を作ったりする。特権アカウントが取得できず一般ユーザアカウントしか取得ができなかった場合、攻撃者はローカルの脆弱性を利用して特権アカウントの取得を試みる。

- 不正攻撃段階

  目的となるホストの権限を取得した攻撃者は、そのホストを占拠して踏み台として Dos 攻撃やウイルス、攻撃の痕跡消去、情報リソースの盗用、システムの破壊など、様々な攻撃を行う。

2.2　偵察行為の種類

偵察行為の段階で行われる偵察行為の種類を以下の表 2.1 に示す。

表 2.1: スキャンの種類

スキャンの種類	プロトコル名	スキャン名
ネットワークスキャン	ICMP	ping スキャン
ポートスキャン	UDP	UDP ポートスキャン
	TCP	TCP 接続スキャン
		TCP SYN スキャン
		TCP FIN スキャン

2.2.1　ネットワークスキャン

ネットワークスキャンとは、攻撃者が行うスキャン攻撃のうち、対象となるノードが存在するかどうかを調査する。代表例として、ICMP を用いた ping スキャンがある。ping に対して応答があった場合はそのホストの存在が確認されたことになる。

2.2.2　ポートスキャン

ポートスキャンとは、攻撃者が行うスキャン攻撃のうち、攻撃対象となるノード上で特定のサービスが動作しているかどうかどうかを調査する目的のものを指す。ポートスキャンでは、対象となるサービスの有無の調査を行うものだけでなく、ノードが返す情報から、対象となるサービスの種類やバージョン情報を取得できる場合がある。ネットワーク上で動作するサービスが使用している通信プロトコルには、TCP と UDP が存在する。したがって、そのサービスを調査するスキャンが使用するプロトコルも、UDP ポートスキャンと TCP ポートスキャンとに分類される。

- UDP ポートスキャン
  UDP ポートスキャンは、UDP パケットを送信して、エラーメッセージの応答の有無によりポートの状態を推測するものである。この手法では、ポートの開閉の判定を「エラーメッセージの応答がない場合にはポートが開いている」という推測に基づいているため、UDP プロトコルの性質を考慮すると疑わしい結果になる場合がある。UDP はコネクションレス型のプロトコルであるため、UDP ポートスキャンは相手先ノードのログには残らない。

- TCP 接続スキャン
  TCP 接続スキャンは、3 ウェイハンドシェイクを用いたスキャンのことである。3 ウェイハンドシェイクとは、TCP コネクションを確立する際の手順のことである。この手順が成功したかどうかで対象となるサービスの動作の有無を判断する。なお、TCP 接続スキャンは相手先とのコネクションを確立するため、成功した場合は相手先ノードのログに残ることとなる。

- TCP SYN スキャン
  TCP SYN スキャンは、3 ウェイハンドシェイクの途中までを行うことでサービスの有無を調査するスキャンである。3 ウェイハンドシェイクを完全に実行せずにコネクションを確立しないため、TCP ハーフスキャンやステルススキャンとも呼ばれる。まず、スキャン実行者は対象のノードへ TCP の SYN パケットを送信する。次に相手先ノード上で対象のサービスが動作していれば、3 ウェイハンドシェイクに従い TCP の SYN+ACK パケットが返信される。しかし対象のサービスが動作していない場合、相手先ノードはコネクションをリセットするために RST+ACK パケットを返信する。TCP SYN スキャン

はこの返信の違いを利用して、サービスの有無を調査する。なお TCP SYN スキャンは相手先とコネクションを確立しないため、相手先ノードのログに残らない。

- TCP FIN スキャン
  TCP FIN スキャンは、相手先ノードへ TCP の FIN パケットを送信することで、サービスの有無を調査するスキャンのことである。したがってスキャン実行者は、TCP の RST+ACK パケットが返ってきた場合は対象のサービスが動作していない状態、何も返信がない場合には対象のサービスが動作している状態であると判断する。また TCP FIN スキャンは、ネットワークの途中でスキャンパケットや応答パケットが消失した場合でも、サービスが動作していると判断してしまうため, 信頼性の低いスキャン方法である。

2.3　偵察段階における攻撃検出の意義

クラッキングにおける第 2 段階のアクセス段階および第 3 段階である不正攻撃段階においては、攻撃に用いられるパケットが IP Spoofing [10] をされていることが多い。IP Spoofing とは、攻撃者を特定されにくくする攻撃手法のことで、自分の IP アドレスを相手の IP アドレスに偽装して攻撃を仕掛けたり、Firewall を突破したりする。つまり、アクセス段階や不正攻撃段階に入った場合は攻撃者の特定が難しい。一方、攻撃の第 1 段階である偵察行為段階においては、偵察行為を行っているパケットが IP Spoofing されている可能性は低い。なぜならば、攻撃者は、目的となるホストからの応答を期待して偵察を行っているためである。したがって、クラッキング対策においては、早い段階での攻撃を検出することで、標的ホストの監視を徹底することに専念できる。つまり偵察段階での攻撃の検出が重要である。

2.4　スロースキャン

スロースキャンとは、IDS によって検知されないように、緩慢な時間間隔で行われる低速のスキャンのことである。通常のスキャン攻撃は、対象とするネットワークに対して、連続してスキャンパケットを送信するため、単位時間当たりのパケットの量が急増する。IDS はこのパケット量の急速な増加を判断して、スキャンが行われていることを検知できる。スロースキャンは、IDS による検知から逃れるために、スキャンとの時間間隔を開けて行う。IDS によっては、少ないパケット量でも検知するように設定を変更することにより、スロースキャンを検知することが可能となる。しかし、スロースキャンは一般ユーザのパケットとの差異がつきにくいため、正常な通信を誤って異常な通信ととらえてしまう false positive の割合が増加してしまい、IDS を利用してスロースキャンを実用的に検出することは難しい。

第 3 章

提案手法

3.1 提案手法の概要

本研究では、時間間隔をあけてパケットを送信して標的となるホストの情報を収集するスロースキャンの特徴分析をすることを目標とする。まず、ネットワーク上を流れているトラヒック上の中からスロースキャンと思われるパケットを抽出する。そして、抽出したスロースキャン IP アドレスに対して、地理情報を調査し、送信元 IP アドレスの可視化を行うことで、スロースキャン IP アドレスの特徴を把握する。

3.2 既存手法の概要

本節では、本研究で使用するスロースキャンの検出手法について説明する。本研究では、既存研究 [1] で考案された手法を活用して、スロースキャンの抽出を行う。既存手法では、与えられたパケットデータに対して、固有アクセス率という尺度を用いることで、IP スキャンおよび Port スキャンの検出を行った。srcIP (以下、送信元 IP のことを srcIP とする) ごとに、固有アクセス率に関する情報を収集し、その情報を用いてスロースキャンの検出を行う。なお、本研究では、srcIP ごとにデータを集計しており、異なる宛先ホスト数、異なる宛先ポート数、総パケット数を以下のように定義した。

- 異なる宛先ホスト数 = #(dstIP)
- 異なる宛先ポート数 = #(dstport)
- 総パケット数 = #(pkts)

3.3　スロースキャンの検出手法

本研究では、固有アクセス率という尺度を用いて、スロースキャンの検出を行う。検出における閾値を決定する際に、固有アクセス率の累積度数分布 (CDF : Cumulative Distribution Function) [11] を使用する。

3.3.1　固有アクセス率

ここで、固有アクセス率を以下の式で定義する。

- 固有アクセス率 = (固有宛先ホスト数/総パケット数) (=#(dstIP)/#(pkts))

この式の分母の総パケット数は、観測対象となる総パケット数である。分子の固有宛先ホスト数は、観測対象となるパケットに出現するユニークなホスト数である。つまり、重複して出現する宛先ホスト数についてカウントする。固有アクセス率は、スキャンによる情報収集の効率を表す指標である。

3.3.2　ホストおよびポートに関する固有アクセス率

正常な通信データと悪性な通信データの、ホストの固有アクセス率の累積度数分布を比較すると、分布上の差異が認められた。この性質を利用することでホストの固有アクセス率に基づくスロースキャンの検出を行う。具体的には TCP、UDP、ICMP に関するホストの固有アクセス率からスキャンの判断を行う。一方、正常な通信データと悪性な通信データの、ポートの固有アクセス率の累積度数分布を比較した場合、ポート固有の固有アクセス率の累積度数分布を比較してもほとんど差が見られなかった [1]。つまり、ポートを順番に偵察するポートスキャンは、スロースキャンの場合にはほとんど行われていないことが分かる。したがって、ポートの固有アクセス率は、スロースキャンの検出には利用しない。

3.3.3　スロースキャン検出の閾値

以上をふまえて、ホストの固有アクセス率に着目したときのスロースキャン検出における閾値を以下のように決定した。

スロースキャン検出における閾値は、[1] の手法を用いて表 3.1 に設定した値を使用する。また、高須ら [7] の研究にならって、設定の閾値を超えた IP アドレスでも、1 つの送信元 IP アドレスから 5 パケット以下の通信はフィルタリングで除外した。

表 3.1: スロースキャン検出の閾値

プロトコル	#(dstIP)/#(pkts)
ICMP	96.1%
TCP	85.2%
UDP	98.0%

第 4 章

提案手法の評価

4.1 評価実験の環境

本研究では、toumon (早稲田大学のネットワークと学外のネットワークとを接続するゲートウェイ) において、3 日間のパケットキャプチャデータを利用した。キャプチャの様子を図 4.1 に示す。

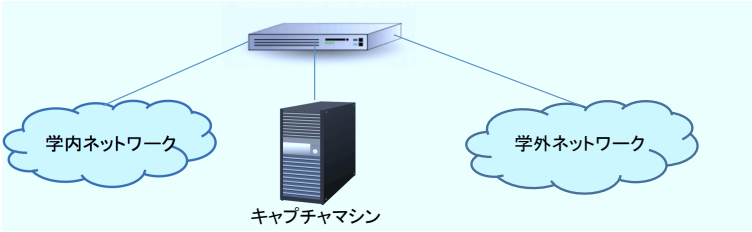

図 4.1: toumon でのパケットキャプチャの様子

本研究では収集したパケットの全部を対象とするわけではない。以下のように観測対象を定めた。

- TCP、UDP、ICMP プロトコルを用いたパケットのみを観測対象とする
  偵察行為段階で使用されるプロトコルには、TCP、UDP、ICMP があり、この 3 つのプロトコルを観測対象とした。

- Web 通信プロトコルを除外する
  Web 通信プロトコルを用いた偵察行為はほとんど存在しないため、Web 通信プロトコルをキャプチャ対象から除外した。

このように観測対象を定めた結果、3 日間のパケットキャプチャデータの概要を、表 4.1 に示す。

表 4.1: 3 日間の toumon 観測データ

キャプチャ期間	2016/01/21 – 2016/01/23
IP アドレス総数	2,892,702

4.2　集計する情報

ホストの固有アクセス率の関する情報を、送信元IP アドレスごとに集計する。TCP、UDP、ICMP ごとのプロトコルに対して、パケット数と固有宛先 IP アドレス数に関する情報を表 4.2 のように集計する。

表 4.2: 集計する情報

偵察行為の種類	プロトコル名	集計データ
IP Scan	ICMP	ICMP パケット数
		ICMP の固有宛先 IP アドレス数
Port Scan	UDP	UDP パケット数
		UDP の固有宛先 IP アドレス数
	TCP	TCP パケット数
		TCP の固有宛先 IP アドレス数

第 5 章

スロースキャンの特徴分析

5.1 分析の対象となるデータの概要

スロースキャンの特徴分析に用いるデータについて、概要を説明する。本研究では、本論文の第 4 章でキャプチャしたデータの中から、スロースキャンを行っていると思われる IP アドレスを抽出したものを使用する。表 4.1 で得られた IP アドレスのうち、本手法によってスロースキャン IP と判定された IP アドレス数を、表 5.1 に示す。

表 5.1: 各プロトコルごとのスロースキャン IP と判定されたパケット数

プロトコル	IP アドレス数
ICMP	6,175
TCP	1,994
UDP	2,246
IP アドレス総数	10,395

IP アドレス総数が ICMP、TCP、UDP を単純に合計した値と一致しないのは、ICMP、TCP、UDP の中で重複した IP アドレスが存在するためである。

5.2　GeoIP を用いた地理情報

節5.1で説明したデータを、GeoIP データベース [15] を用いて、地理情報の取得を行う。GeoIP データベースを使用したときの国別コード表を表5.2 に示し、地理情報を取得した結果を図5.1 に示す。

表 5.2: 国別コード

国コード	国名	国コード	国名
AR	アルゼンチン	KZ	カザフスタン
AU	オーストラリア	MX	メキシコ
BA	ボスニア・ヘルツェゴビナ	MY	マレーシア
BR	ブラジル	NL	オランダ
CA	カナダ	NZ	ニュージーランド
CN	中国	None	不明
CZ	チェコ	PK	パキスタン
DE	ドイツ	PL	ポーランド
ES	スペイン	RO	ルーマニア
FR	フランス	RU	ロシア
GB	イギリス	SG	シンガポール
HK	香港	TH	タイ
ID	インドネシア	TR	トルコ
IE	アイルランド	TW	台湾
IN	インド	UA	ウクライナ
IT	イタリア	US	アメリカ
JP	日本	VE	ベネズエラ
KR	韓国	VN	ベトナム

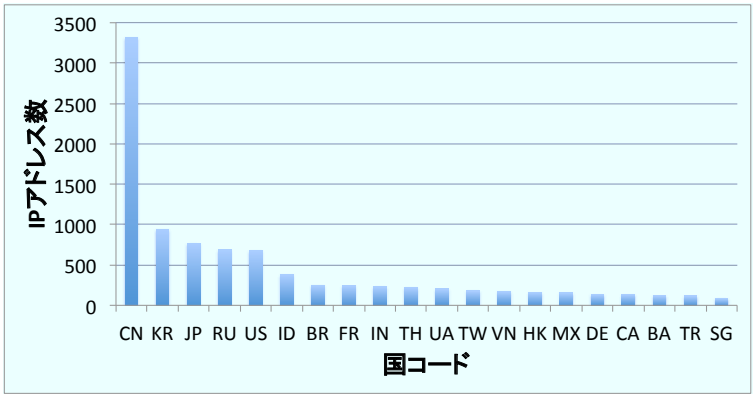

図 5.1: スロースキャン IP アドレスの地理情報 (全体)

図 5.1 から、スロースキャンを行っている国は中国が 1 番多いという結果が得られた。、他に多い国としては、韓国、日本、ロシア、アメリカ、などがある。次に、IMCP、TCP、UDP の各プロトコルごとの地理情報の結果を、ICMP は図 5.2 に、TCP は図 5.3 に、UDP は図 5.4 に示す。

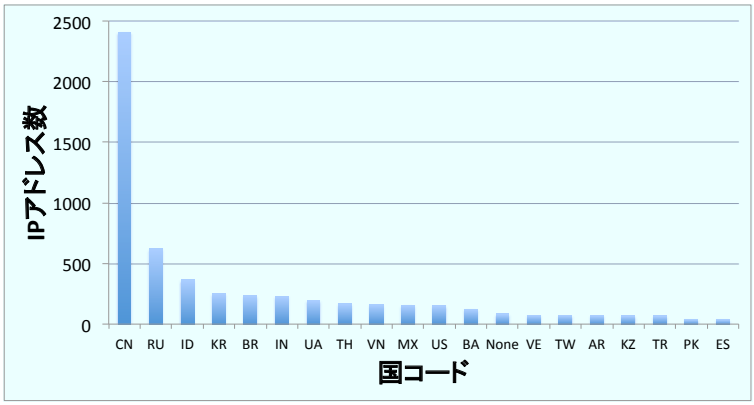

図 5.2: スロースキャン IP アドレスの地理情報 (ICMP)

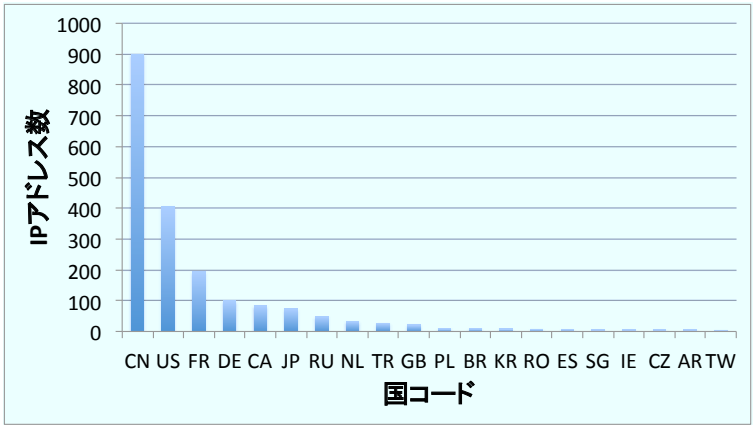

図 5.3: スロースキャン IP アドレスの地理情報 (TCP)

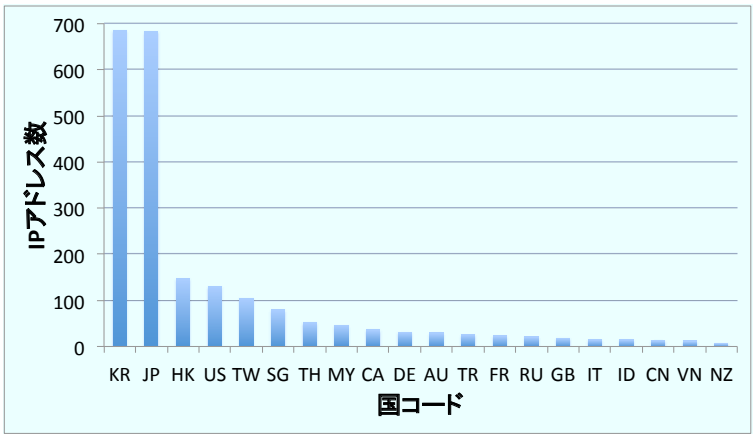

図 5.4: スロースキャン IP アドレスの地理情報 (UDP)

各プロトコルごとの送信元アドレスの地理情報を見ると、ICMP と TCP に関しては、偵察行為を行っているパケットの多くが中国からのものであるという結果になった。UDP を用いた偵察行為は、日本と韓国のが IP アドレス多いという結果になった。千葉らの研究 [6] は、IP 攻撃通信データを送信元 IP アドレスごとに国別に集計した論文である。この論文では、中国とアメリカからの IP 攻撃が多いという結果が得られているが、本研究では、それに加えて韓国やロシアなどのアジア圏からのスキャン行為を観測できている。全体的に見ると文献 [6] と同様の結果が得られていることが分かる。

5.3　ヒルベルト曲線を用いた送信元の可視化

5.3.1　ヒルベルト曲線

ヒルベルト曲線は空間充填曲線の一つであり、多次元空間の情報を 1 次元空間に写像する手法として用いられる。カタカナのコの字の 4 つの基本図形を LUR (Left-Up-Right) , ULD (Up-Left-Down) , DRU (Down-Right-Up) , RDL (Right-Down-Left) としたとき、式 (5.1) の 4 つのルールに従って、$n = 0$ になるまで再帰的に呼び出すことでヒルベルト曲線が生成される。

$$\begin{cases} \mathbf{ULD}(n) = \mathbf{LUR}(n-1), \mathrm{Up}, \mathbf{ULD}(n-1), \mathrm{Left}, \mathbf{ULD}(n-1), \ \mathrm{Down}, \mathbf{RDL}(n-1) \\ \mathbf{DRU}(n) = \mathbf{RDL}(n-1), \mathrm{Down}, \mathbf{DRU}(n-1), \mathrm{Right}, \mathbf{DRU}(n-1), \ \mathrm{Up}, \mathbf{LUR}(n-1) \\ \mathbf{RDL}(n) = \mathbf{DRU}(n-1), \mathrm{Right}, \mathbf{RDL}(n-1), \mathrm{Down}, \mathbf{RDL}(n-1), \ \mathrm{Left}, \mathbf{ULD}(n-1) \\ \mathbf{LUR}(n) = \mathbf{ULD}(n-1), \mathrm{Left}, \mathbf{LUR}(\mathbf{n-1}), \mathrm{Up}, \mathbf{LUR}(n-1), \ \mathrm{Right}, \mathbf{DRU}(n-1) \end{cases} \tag{5.1}$$

式 (5.1) に従ってヒルベルト曲線を描く。

5.3.2　送信元 IP の可視化

提案手法を用いて抽出したスロースキャンの送信元の IP アドレスをヒルベルト曲線を用いて可視化した結果を図 5.5 に示す。各ブロックは IP アドレスの X.0.0.0/24 に属する送信元の個数を示しており、各ブロックの色が濃いほど多くのスロースキャンが行われていると予測できる。この図より、送信元は一様に分散しているのではなく、ある箇所にまとまっており、空間的な局所性が確認できる。

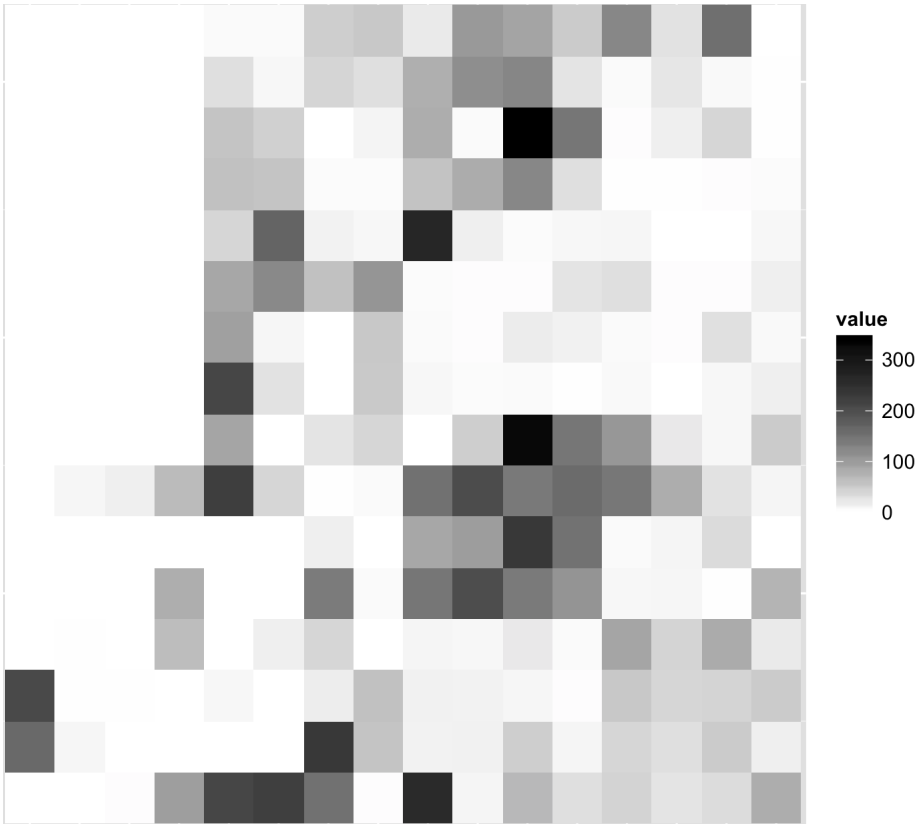

図 5.5: ヒルベルト曲線を用いたスロースキャンの送信元の分布の可視化

第 6 章

結論

6.1 まとめ

本論文では、既存研究 [1] の手法を用いて、スロースキャンの検出が可能であることを示し、GeoIP を利用した攻撃元の地理情報の取得を行った。ネットワークの利用が拡大する中で、ネットワークを介したクラッキングやワーム等のサイバー攻撃による被害が拡大している。これらの攻撃は、攻撃の前段階に偵察行為を行うことが多く、偵察行為段階での検出が重要である。本研究では、スロースキャンの攻撃を検知するため、スロースキャンの特徴分析を行い、スキャンの精度を高めることを目的とした。

6.2 今後の課題

本研究では、スロースキャンを行う IP アドレスの特徴分析にとどまっているが、新たなスロースキャンの検出手法および検出の精度を高めることが今後の課題となる。

謝辞

本修士論文の作成にあたり、日ごろよりご指導をいただいた早稲田大学基幹理工学研究科の後藤滋樹教授に深く感謝いたします。本研究活動を進めるにあたり、多くのご助言をいただいた後藤研究室 OB 西宇基氏に心より感謝いたします。最後に、ともに研究をおこなった後藤滋樹研究室の諸氏に感謝いたします。特に、篠宮一真氏、青木一樹氏、志村正樹氏、高橋一基氏、武部嵩礼氏には大変お世話になりました。

参考文献

[1] 西宇 基, 固有アクセス率に基づく Slow Scan 検出法, 早稲田大学大学院修士論文, February 2013.

[2] 西宇 基, 多段階フィルタリングによる Slow Scanning 検出法, 早稲田大学卒業論文, February 2011.

[3] 渡邊 雄二, スロースキャン IP の特徴分析, 早稲田大学卒業論文, February 2014.

[4] 篠宮 一真, 二段階の機械学習を用いた高精度な不正通信検知法, 早稲田大学卒業論文, February 2015

[5] 古岡 達也, ダークネットのトラヒック分析に基づくスロースキャン検知法, 早稲田大学修士論文, February 2015.

[6] 千葉 大紀, IP アドレス構造を用いた攻撃通信の判別法, 早稲田大学卒業論文, February 2011.

[7] 高須 雄一, ダークネット観測に基づく攻撃の時間変化の可視化, 早稲田大学修士論文, February 2014

[8] 井上 貴裕, 石田 賢治, 小畑 博靖, 舟阪 淳一, 悪意の度合いに基づくネットワーク偵察行為の順位付け, 電子情報通信学会技術研究報告. ICSEC2002-18, pp.37-42, 2002.

[9] 竹下 隆文, 村山 公保, 荒井 透, 苅田 幸雄, マスタリング TCP/IP 入門編第 4 版, オーム社, 2007.

[10] IP Spoofing, https://www.ipa.go.jp/security/fy14/contents/soho/html/chap1/spoof.html

[11] 累積度数, http://next1.msi.sk.shibaura-it.ac.jp/MULTIMEDIA/statistics/node3.html

[12] TCP Slice,
http://www.tcpdump.org/manpages/tcpslice.1.txt

[13] 特権 ID 管理 SHieldWARE のご紹介,
http://www.ssl.fujitsu.com/products/network/netproducts/shieldware/shieldware-presen.pdf

[14] Xcart, DDoS 攻撃による通信障害について,
http://www.xcart.jp/

[15] GeoIP,
http://readit.l8r.in/post-337/

[16] Dale Dougherty, Arnold Robbins, sed&awk プログラミング, 株式会社オライリージャパン, 2011.

[17] Bruce Blin, 入門 UNIX シェルプログラミング, ソフトバンククリエイティブ株式会社, 2006.