Vol.2020-DPS-182 No.52
Vol.2020-CSEC-88 No.52
2020/3/13

スロースキャン攻撃検知のための特徴量の提案

山下 智也[1,a)]　宮本 大輔[1,b)]　関谷 勇司[1,c)]　中村 宏[1,d)]

概要：ネットワーク上の攻撃者は，本格的な攻撃を行う前にスキャン攻撃を行い，ネットワーク上のホストに存在する脆弱性についての情報を収集する．スキャン攻撃の検知は，さらなる本格的な攻撃を未然に防ぐための重要な課題といえる．スキャン攻撃に対して侵入検知システム（IDS）が提案されているが，長期間にわたる低速なスキャンによるスロースキャン攻撃は検知が難しい．そこで本稿では，スキャン攻撃を行うホストと正常な通信を行うホストの通信挙動の違いを捉えることのできる特徴量を提案する．そして，提案する特徴量を用いてホストの分類を行い，スロースキャン攻撃検知における有効性を検証する．

1. 背景

近年のコンピュータネットワークの発展と普及により，情報ネットワークは一層人々にとって身近な存在となった．現在ではメーリングリストやホームページ，ブログや SNS など様々なサービスが現れ，人々は簡単に情報交換を行うことが可能である．しかし，コンピュータネットワークの普及とともに，ネットワークを悪用して情報を盗み取る，他者のホストへの不正アクセスを行う，などといったネットワーク上の攻撃も年々増加している [1]．ネットワーク上の攻撃者は，本格的な攻撃の準備としてまずスキャン攻撃を行い，ネットワーク上のホストに存在する脆弱性についての情報収集を試みる．

スキャンそのものが違法な活動であるというわけではない．たとえば，スキャンは OSINT という諜報活動における合法の検索エンジン，Shodan や Censys などにおいて利用される．Shodan や Censys とは，定期的にネットワーク上のアクセス可能なホストに対してスキャンによる情報収集を行い，収集した情報の提供を行う検索エンジンである [2,3]．しかし，攻撃者がスキャン攻撃を行う際には，スキャン攻撃によって得られる情報を利用した本格的な攻撃が後に続く可能性が高い．本格的な攻撃としては，ホストの機密情報を盗み出す，ホストのデータやシステムを破壊する，ホストを乗っ取りボット化するなどが考えられる．したがって，スキャン攻撃の検知は，セキュリティ分野における重要な課題のひとつといえる．

代表的なスキャン攻撃に，水平ポートスキャン攻撃，垂直ポートスキャン攻撃，ホストスキャン攻撃の 3 つがある [4]．水平ポートスキャン攻撃とは 1 台の標的ホストに対して，複数ポートに信号を送ることで，利用可能なポートを探索する攻撃である．一方，垂直ポートスキャン攻撃とは複数の標的ホストに対して，単一ポートに信号を送ることで，ネットワーク上で特定のポートを開いているホストを探索する攻撃である．そしてホストスキャン攻撃とは，Ping などのコマンドを用いて，ICMP エコー要求をネットワーク上の複数 IP アドレスに送ることで，通信可能なホストを探索する攻撃である．

ネットワーク上のスキャン攻撃を検知しネットワーク管理者に通知するシステムに，侵入検知システム（IDS）が存在する．IDS はシステムやネットワーク上で発生するイベントを監視，分析することでネットワーク上の攻撃検知を行う．既存の IDS では，シグネチャと呼ばれるルールに通信データを照らし合わせて攻撃通信を検知する手法，もしくは，通信データの統計値に対して閾値を定めて攻撃通信を検知する手法が主流である．しかし，通信の時間間隔を大きくしてスキャンを行うスロースキャン攻撃のシグネチャを網羅的に用意したり，スロースキャン攻撃に対応する通信データの適切な閾値を定めることは難しい．そのため，既存の IDS ではスロースキャン攻撃を検知することは困難とされる．

そこで本稿では，シグネチャや通信データの閾値に頼った検知ではなく，通信挙動にもとづいて検知を行うアノマリ検知に着目し，スロースキャン攻撃の検知を考える．アノマリ検知においては，スキャン攻撃と正常な通信の通信挙動の違いを捉えることのできる特徴量を利用することが

[1] 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
[a)] yamashita@hal.ipc.i.u-tokyo.ac.jp
[b)] daisu-mi@nc.u-tokyo.ac.jp
[c)] sekiya@nc.u-tokyo.ac.jp
[d)] nakamura@hal.ipc.i.u-tokyo.ac.jp

重要である．本稿では，スキャン攻撃と正常な通信の通信挙動の違いを捉えることのできる 2 つの特徴量を提案する．そして，提案する特徴量を用いることで，スロースキャン攻撃を行うホストを検知できることを実験によって示す．以下に本稿の構成について述べる．

本稿では，まず 2 章で関連研究について述べ，3 章では提案する 2 つの特徴量について説明する．4 章では，スロースキャン攻撃検知の評価実験について述べ，5 章で実験結果を述べる．6 章で考察をまとめ，7 章で結論と今後の課題について述べる．

2. 関連研究

2.1 スロースキャン攻撃

攻撃者は，スキャン攻撃を行うことで，ネットワーク上のホストに存在する脆弱性に関する情報を収集する．一般的な IDS では，主に時間あたりのトラフィック量に着目して，スキャン攻撃の検知を試みる．したがって，長い時間をかけ低速に実行されたスキャン攻撃（スロースキャン攻撃）を検知することは難しい．過去には，1 時間に数回程度の頻度で行われたスロースキャン攻撃が観測されている [5,6]．

2.2 スロースキャン攻撃検知に関する既存研究

スロースキャン攻撃を検知するための既存研究に，Fuzzy Rule を利用した検知に関する研究がある [7–9]．Fuzzy Rule とはスキャン攻撃検知に利用する特徴量の値に対し，攻撃通信か否かという結論を対応づけたシンプルなルールである．セキュリティのエキスパートが前もって Fuzzy Rule を定めて検知を行う．Fuzzy Rule を用いることで攻撃通信か否かだけでなく，個々の通信の攻撃の度合いを特徴量の値に応じて決めてしまうことも可能である．[8,9] における，Fuzzy Rule を用いた検知に関する研究では，各ホストの単位時間あたりのパケットの送受信数に着目した特徴量が用いられている．この特徴量に対し Fuzzy Rule を用いた検知実験を行った結果，一般的な IDS（Snort）では検知が困難であるスロースキャン攻撃に対して，検知性能が向上したと述べられている．ただし，検知実験では，実際の通信データではなく，擬似的に構築したテスト用のネットワーク環境において検知性能の評価が行われており，実際の通信に対する検知性能についての言及はされていなかった．

また，TCP の 3 ウェイハンドシェイクの情報を用いた検知手法に関する既存研究がある [10]．この研究では，TCP の 3 ウェイハンドシェイクの情報をもとに，ホストを攻撃ホスト，怪しいホスト，正常ホストの 3 つに分類する．そして，怪しいホストに対しては，一定時間だけ判断を先延ばしにして改めて判断を行うという方法をとっている．怪しいホストに対する判断を先延ばしにすることにより，正常ホストに対する誤検知を低減することができると考えられる．検知実験においては，通信間隔の異なるスキャンを行う攻撃ホストにスキャン攻撃を実行させて，検知性能の評価を行っている．検知性能の評価の結果，用意した攻撃ホストによるすべてのスキャン攻撃を検知できたと述べられている．ただし，怪しいホストに対する判断の先延ばし時間はあらかじめ定めておく値であり，この先延ばし時間を超えた時間間隔で，スキャンを行うスロースキャン攻撃に対する有効性は確認されていない．

さらに，固有アクセス率に着目したスロースキャン攻撃検知に関する既存研究がある [11]．固有アクセス率とは 1 パケットあたりの宛先 IP アドレスの種類数，もしくは 1 パケットあたりの宛先ポートの種類数であり，スキャン攻撃による情報収集の効率を表す指標となっている．[11] ではホストごとに固有アクセス率を計算し，閾値を定めて，スキャン攻撃を行うホストの検知を行っている．早稲田大学と学外のネットワークをつなぐゲートウェイで取得した通信データを用いて検知実験を行った結果，既存の IDS に比べて低い誤検知率を達成し，スロースキャン攻撃と思われる通信を検知できるということが確認されている．ただし，検知に用いる閾値は，特徴量の値に対するホストの累積度数分布を観察することで決定されており，スキャン攻撃を行うホストが正常な通信を行うホストに比べて極端に少ない場合に，閾値を適切に決定することができるか否かについての確認はされていない．

3. 提案手法

3.1 既存研究が持つ課題と提案手法の概要

Fuzzy Rule を用いた検知に関する既存研究 [7–9] における課題として，時間依存性を持つ特徴量を用いて検知を試みているということが挙げられる．本稿において，時間依存性を持つ特徴量とは単位時間あたりの情報を捉える特徴量と定義する．このような時間依存性を持つ特徴量を用いた検知では，既存研究で想定されているスロースキャン攻撃よりも，さらに低速なスロースキャン攻撃への対応はしばしば困難になると考えられる．また，[10] における先延ばし時間にも，同様の課題があると考えられる．すなわち，怪しいホストに対する判断の先延ばし時間を超えた時間間隔でスキャンを行うようなスロースキャン攻撃に対しては，適切な検知は困難であると考えられる．

また [7–9] では仮想的なネットワーク環境で生成した通信データを用いて検知実験が行われており，実際の通信に対してどれほど検知効果があるのかについて検証がされていないという課題もある．

固有アクセス率に着目したスキャン検知に関する研究 [11] では，検知の閾値を累積度数分布を観察することで決定していた．しかし，累積度数分布をみて検知閾値を決定する場合，スキャン攻撃のホストの数が正常ホストの数に比べ

表 1　通信フローの例

開始時刻	srcIP	srcPort	dstIP	dstPort
08:14:59	198.51.100.30	52676	198.51.100.199	443
08:15:00	198.51.100.45	52913	192.0.2.40	443
08:15:00	192.0.2.39	52668	192.0.2.14	443
08:15:01	192.0.2.23	80	198.51.100.14	443
08:15:02	198.51.100.23	52668	192.0.2.16	80

て極端に少ない場合に，閾値を適切に決定できなくなる可能性があると考えられる．

上記の課題を踏まえ，本稿では，スキャン攻撃検知のための時間依存性を持たない特徴量を提案する．これにより，低速にスキャンを行うスロースキャン攻撃の検知にも対応できると考えられる．また，提案した特徴量を用いて通信データ中に含まれる全てのホストの分類を行い，スキャン攻撃を行うホストと正常な通信を行うホストの分類が可能であることを確認する．全てのホストの分類を行うことで，スキャン攻撃を行うホストが正常な通信を行うホストに比べて極端に少ない場合にも，スキャン攻撃の検知が可能になると考えられる．

3.2　検知の方針

本稿では通信フローから計算できる，スロースキャン攻撃の検知において有効な特徴量を提案する．通信フローは，任意の送信元と宛先との間の，双方向のパケットのストリームと定義する．送信元と宛先は，ネットワーク層の IP アドレスと，トランスポート層の送信元および宛先のポート番号によってそれぞれ定義する．通信フローは，F=（srcIP，srcPort，dstIP，dstPort，t_s）の形式で表される．srcIP は送信元 IP アドレス，srcPort は送信元ポート番号，dstIP は宛先 IP アドレス，dstPort は宛先ポート番号，t_s は通信フローの開始時刻である．定義より通信フローは，どこから，どこに，いつ，どのような通信が行われていたのかという情報を持つデータとなっていることがわかる．通信フローを攻撃検知に用いるメリットとしては，パケット毎に解析を行うのに比べて，保持するデータが少量で済む，軽量な検知が可能などが挙げられる．表 1 に通信フローの例を示す．

本稿では，スキャン攻撃と正常な通信の通信挙動における 2 つの違いに着目する．1 つは，通信の際に利用する宛先ポートの数と，通信を行う宛先 IP アドレスの数の比である．もう 1 つは，src として行う通信フローの数と，dst として行う通信フローの数の比である．

3.3　提案する特徴量

3.3.1　宛先ポートの数と宛先 IP アドレスの数の比（PPI）

攻撃者がスキャン攻撃を行う際には，標的とするホスト，もしくは標的とするネットワークの情報を収集することを目的としている．水平ポートスキャン攻撃を行う攻撃者は，標的とする少数のホストに対して，多くの宛先ポートを用いた通信を行うと考えられる．一方，垂直ポートスキャン攻撃，もしくはホストスキャン攻撃を行う攻撃者は標的とするネットワーク上の多くのホストに対して，特定の少数の宛先ポートを用いた通信を行うと考えられる．この通信挙動は正常な通信を行うホストにはめったに見られないものである．したがって，この通信挙動の違いにもとづいた特徴量として，Ports Per IPs（PPI）を提案する．

$$\mathrm{PPI} = \log \frac{(\text{ホストが通信に用いる dstPort の種類数})}{(\text{ホストが通信を行う dstIP の種類数})}$$

水平ポートスキャン攻撃を行う攻撃者は，少数のホストに対して多くの宛先ポートを用いて通信を行うため，特徴量 PPI の値は大きなものになると考えられる．一方，垂直ポートスキャン攻撃，ホストスキャン攻撃を行う攻撃者は，多くのホストに対して少数の宛先ポートを利用して通信を行うため，特徴量 PPI の値は小さなものになると考えられる．そして正常な通信を行うホストの特徴量 PPI の値は，ゼロに近いものになると考えられる．したがって，特徴量 PPI を用いることで，スキャン攻撃を行う攻撃者と，正常な通信を行うホストを区別することができると考えられる．なお，特徴量 PPI において対数を取っているのは，分子に対して分母が大きい場合に，特徴量 PPI がゼロに収束することを回避するためである．

ただし，特徴量 PPI の問題点として，正常な通信を行う Web サーバや DNS サーバの特徴量 PPI の値が大きくなる可能性があることが考えられる．これは Web サーバや DNS サーバに対して，少数のクライアントホストが複数の接続を確立することがしばしば起こるからである．したがって，特徴量 PPI のみを用いてスキャン攻撃検知を行った場合，正常な通信を行う Web サーバや DNS サーバを誤って検知してしまう可能性があると考えられる．このような誤検知を回避するために，スキャン攻撃を行う攻撃者と，正常な通信を行う Web サーバや DNS サーバの通信挙動の違いを捉える特徴量を提案する．

3.3.2　src のフロー数と dst のフロー数の比（SPD）

正常な通信を行う Web サーバや DNS サーバは，自らが送信元として始まる通信と同程度かそれ以上に，自らが宛先として始まる通信を行うと考えられる．一方，スキャン攻撃を行う攻撃者は，自らが送信元として始まる通信が主であり，自らが宛先として始まる通信は少ないと考えられる．この通信挙動の違いにもとづいた特徴量として，以下の特徴量 Src Per Dst（SPD）を提案する．

$$\mathrm{SPD} = \log \frac{(\text{src として行う通信フロー数})}{(\text{dst として行う通信フロー数})}$$

正常な通信を行う Web サーバや DNS サーバは，src として行う通信フロー数と dst として行う通信フロー数が同程度となると考えられるため，特徴量 SPD の値は小さな

ものになると考えられる．一方，スキャン攻撃を行う攻撃者は，src として行う通信フロー数が，dst として行う通信フロー数を大きく上回ると考えられ，特徴量 SPD の値は大きなものになると考えられる．したがって，特徴量 SPD を用いることで，スキャン攻撃を行う攻撃者と正常な通信を行う Web サーバや DNS サーバを区別することができると考えられる．

まとめると，各ホストに対して特徴量 PPI と特徴量 SPD を計算すると，水平ポートスキャン攻撃を行うホストの特徴量 PPI，SPD はともに大きな値をとると考えられる．また，垂直ポートスキャン攻撃，もしくはホストスキャン攻撃を行うホストについては，特徴量 PPI が小さく，特徴量 SPD が大きな値をとると考えられる．そして，正常な通信を行うホストの特徴量 PPI と特徴量 SPD は，スキャン攻撃を行うホストとは大きく異なる値を取ると考えられる．したがって，特徴量 PPI と特徴量 SPD を用いることで，正常な通信を行うホストとスキャン攻撃を行うホストの分類が可能になると考えられる．

4. 評価実験

特徴量 PPI と特徴量 SPD を用いることで，スキャン攻撃を行うホストと，正常な通信を行うホストを分類できることを実験 1 により確認する．そして，特徴量 PPI と特徴量 SPD が，スロースキャン攻撃の検知においても有効であることを実験 2 により確認する．特徴量の有効性を確認するための評価実験には，4.1 節で説明する通信データを利用する．

4.1 利用するデータセット

特徴量 PPI と特徴量 SPD の評価実験のためのデータセットとして，東京大学の一部のネットワークにおいて取得された通信データを利用する．取得日時は 2020 年 2 月 1 日の 8 時 14 分から 14 時 14 分までの 6 時間である．本稿では通信フローの情報を用いた特徴量を考えるために，フロー生成ツール Yet Another Flowmeter（yaf）[12] を用いて，利用する通信データから通信フローを抽出し，それ以外のデータは破棄している．また，本稿では IP アドレスを RFC5737 にしたがって変換することで，データの匿名性を担保する [13]．なお，今回用いたデータセットには 521500 個の IP アドレスが存在する．

4.2 実験 1：特徴量 PPI，SPD によるホストの分類

データセット 1 時間分（8 時 14 分から 9 時 14 分のデータセット）に含まれる全てのホストに対し，特徴量 PPI と特徴量 SPD を計算する．そして，各ホストを特徴量 PPI と特徴量 SPD を用いてマッピングする．マッピング結果に対して，いくつかの領域のホストの通信データを観察し，それぞれの領域にどのような通信挙動を持つホストが存在するのか確認する．観察する領域は特徴量 PPI，SPD がともに大きな値をとる領域，特徴量 PPI が小さな値をとり，特徴量 SPD が大きな値をとる領域，また特徴量 PPI が大きな値をとり，特徴量 SPD が小さな値をとる領域の 3 つとする．

表 2　利用するデータセットの時間と挿入する攻撃のフロー数

時間幅	1 時間	6 時間
利用するデータの時間	8:14〜9:14	8:14〜14:14
挿入する攻撃のフロー数	30 フロー	180 フロー

4.3 実験 2：特徴量を用いたスロースキャン攻撃検知

つぎに，データセットに対し，スロースキャン攻撃を模した擬似的な攻撃通信を行うホストの通信データを挿入する．そして，特徴量 PPI と特徴量 SPD をデータセットに含まれる全ホストに対して計算し，各ホストのマッピングを行う．マッピングされた結果において，挿入された擬似的な攻撃通信を行うホストが，どのようにマッピングされるかを確認する．実験結果としてデータセット 1 時間分（8 時 14 分から 9 時 14 分のデータセット）を用いたマッピングと，データセット 6 時間分（8 時 14 分から 14 時 14 分のデータセット）を用いたマッピングの 2 つを示すこととする．マッピングに利用するデータセットの時間が長くなるにつれて，挿入される擬似的なスキャン攻撃のフロー数は増加することとなる．挿入する擬似的なスロースキャン攻撃について 4.4 節で説明する．

4.4 挿入するスロースキャン攻撃

本稿において挿入する擬似的なスロースキャン攻撃を行うホストは，2 分あたり 1 スキャン程度の頻度で水平ポートスキャン攻撃を行うホストとする．一般的なポートスキャンツールである nmap は，デフォルトで 0.2 秒あたり 1 スキャンを実行する [14]．また，一般的な IDS である Snort を用いた検知では，2 分あたり 1 スキャン程度の頻度で実行される水平ポートスキャン攻撃を検知できないということを確認した．したがって，本稿で挿入する水平ポートスキャン攻撃は，十分に低速なスロースキャン攻撃であるといえる．なお，スキャン攻撃は 8 時 14 分から 14 時 14 分の 6 時間，利用するデータセット全時間にわたって挿入するものとする．表 2 に，利用するデータセットの時間と挿入するスキャン攻撃のフロー数を示す．

5. 実験結果

5.1 実験 1：特徴量 PPI，SPD によるホストの分類

1 時間分のデータセット（8 時 14 分から 9 時 14 分のデータセット）に対する，特徴量 PPI と特徴量 SPD によるマッピング結果を図 1 に示す．横軸に特徴量 PPI，縦軸に特徴量 SPD の値をとっている．観察する 3 つの領域は，図 1

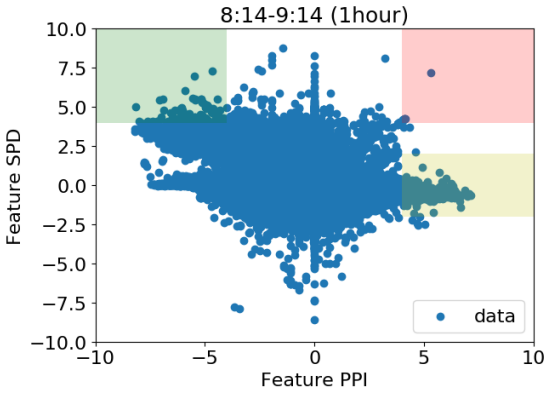

図 1　通信データを観察する領域

表 3　赤色の領域に存在するホストの通信データ

srcIP	srcPort	dstIP	dstPort
198.51.100.1	55164	192.0.2.24	18842
198.51.100.1	51470	203.0.113.15	790
198.51.100.1	36002	203.0.113.13	59999
198.51.100.1	44486	192.0.2.24	59999
198.51.100.1	46742	192.0.2.24	7376
198.51.100.1	47021	203.0.113.13	7376
198.51.100.1	60209	203.0.113.15	7376
...	...	...	...
198.51.100.1	42806	203.0.113.15	49244
198.51.100.1	59257	192.0.2.24	49244
198.51.100.1	45600	203.0.113.13	49244

における赤色の領域（特徴量 PPI，SPD がともに大きい領域），緑色の領域（特徴量 PPI が小さく，SPD が大きい領域），黄色の領域（特徴量 PPI が大きく，SPD が小さい領域）とする．それぞれの領域は，赤色の領域（PPI ≥ 4.0，SPD ≥ 4.0），緑色の領域（PPI ≤ -4.0，SPD ≤ 4.0），黄色の領域（PPI ≥ 4.0，$-2.0 \leq$ SPD ≤ 2.0）とする．

5.1.1　赤色の領域のホスト

マッピングの結果，赤色の領域には 7 個の IP アドレスが存在することがわかった．赤色の領域に存在するホストのひとつについて，表 3 に通信データを示す．

表 3 の通信データを見ると，ホスト 198.51.100.1 は 3 つのホスト（192.0.2.24，203.0.113.15，203.0.113.13）に対して多くの宛先ポートを利用して通信を行っており，これは水平ポートスキャン攻撃の挙動と考えられる．赤色の領域に存在するホストの多くは，表 3 のように少数のホストに対して，多くの宛先ポート番号を利用して通信を行うホストであることがわかった．したがって，この領域には水平ポートスキャン攻撃を行うホストが多く存在していると考えられる．

5.1.2　緑色の領域のホスト

マッピングの結果，緑色の領域には 190 個の IP アドレスが存在していることがわかった．緑色の領域のホストのひとつについて通信データを表 4 に示す．

表 4　緑色の領域に存在するホストの通信データ

srcIP	srcPort	dstIP	dstPort
198.51.100.23	51347	203.0.113.16	445
198.51.100.23	51349	203.0.113.23	445
198.51.100.23	51350	203.0.113.25	445
...	...	...	...
198.51.100.23	51356	130.69.243.14	445
198.51.100.23	51358	130.69.243.21	445

表 5　黄色の領域に存在するホストの通信データ

srcIP	srcPort	dstIP	dstPort
198.51.100.35	443	203.0.113.212	43922
198.51.100.35	443	203.0.113.212	43922
198.51.100.35	443	203.0.113.212	44062
...	...	...	...
198.51.100.35	443	203.0.113.212	33650
198.51.100.35	443	203.0.113.212	33710

表 4 の通信データを見ると，ホスト 198.51.100.23 は多くのホストに対して，宛先ポート番号 445 を利用して通信しており，これは垂直ポートスキャン攻撃の挙動と考えられる．緑色の領域に存在するホストの多くは表 4 のように多数のホストに対して，少数の宛先ポートを利用して通信を行うホストであることがわかった．したがって，この領域には垂直ポートスキャン攻撃，もしくはホストスキャン攻撃を行うホストが多く存在していると考えられる．

5.1.3　黄色の領域のホスト

マッピングの結果，黄色の領域には 502 個の IP アドレスが存在していることがわかった．黄色の領域のホストのひとつについて通信データを表 5 に示す．

黄色の領域に含まれるホストの通信データを見ると，少数のホストに対して，多くの宛先ポートを利用して通信を行っており，水平ポートスキャン攻撃を行っているように見える．しかし，送信元ポート番号が 443 番，すなわち HTTPS ポートが利用されているということと，宛先ポートの一部が重複していることから，水平ポートスキャン攻撃の挙動であるとは考えにくい．このホストは正常な通信を行う HTTPS サーバであると考えられる．黄色の領域に存在するホストの多くは表 5 のように正常な通信を行う HTTP(S) サーバ，もしくは DNS サーバと考えられるホストであることがわかった．

特徴量 PPI と特徴量 SPD を用いたホストのマッピングに対し，3 つの領域の通信データを観察した結果，正常な通信を行うホストと，水平ポートスキャン攻撃を行うホストと，垂直ポートスキャン，もしくはホストスキャン攻撃を行うホストを分類できることが確認できた．また，マッピング結果の観察により，水平ポートスキャン攻撃を行っていると考えられるホストの数は，垂直ポートスキャン攻撃やホストスキャン攻撃を行っていると考えられるホストの数に比べて少ないということもわかった．これは，多くの

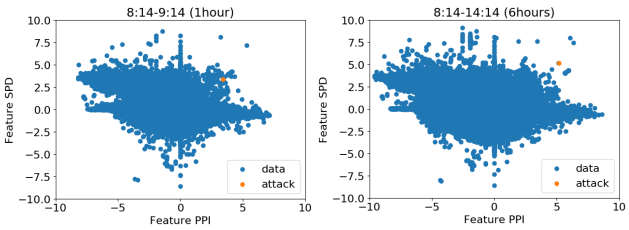

図 2　特徴量 PPI，SPD によるマッピング

攻撃者の目的が，ターゲットとなるホストをしぼって開けているポートを探索することではなく，特定のポートを開けているホストを探索することにあるためと考えられる．

5.2　実験 2：特徴量を用いたスロースキャン攻撃検知

つぎに，擬似的なスロースキャン攻撃を含むデータセットに対する，特徴量 PPI，SPD による各ホストのマッピング結果を図 2 に示す．横軸に特徴量 PPI，縦軸に特徴量 SPD の値を取っている．また青色の点がデータセットに含まれていたホストのマッピング結果であり，オレンジ色の点が擬似攻撃を行うホストのマッピング結果である．

マッピング結果を見ると，オレンジのプロット，すなわち挿入した擬似攻撃のプロットが，マッピングに利用するデータセットの時間が長くなるにつれて，右上に移動していることがわかる．これは，時間が長くなるにつれて水平ポートスキャン攻撃が進行し，特徴量 PPI と特徴量 SPD の値が徐々に大きくなっていくからであると考えられる．そして，6 時間のデータセットを用いた場合のマッピングにおいては，オレンジのプロットが特徴量 PPI と特徴量 SPD において，目立った値を取っていることが確認できる．

垂直ポートスキャン攻撃，もしくは低速なホストスキャン攻撃が行われた場合にも同様に，スキャン攻撃が進行するにつれてプロットは左上に移動，すなわち特徴量 PPI が小さく，特徴量 SPD が大きくなっていくと考えられる．

したがって，特徴量 PPI と特徴量 SPD を用いることでスロースキャン攻撃を行うホストを検知できる可能性があることを確認した．

6.　考察

特徴量 PPI と特徴量 SPD に対する評価実験によって以下のことを確認した．

まず，特徴量 PPI と特徴量 SPD を用いたホストの分類の結果，正常な通信を行うホストと，水平ポートスキャン攻撃を行うホストと，垂直ポートスキャン攻撃，もしくはホストスキャン攻撃を行うホストを分類できることを確認した．これは特徴量 PPI と特徴量 SPD がスキャン攻撃の通信挙動と正常な通信の通信挙動を捉えることができているためと考えられる．また，データセットに含まれる全てのホストに対して，特徴量 PPI と特徴量 SPD によるマッピングを行うことで，水平ポートスキャン攻撃のように数が少ないホストに対しても検知できる可能性があることを確認した．

さらに，スロースキャン攻撃に対しても，特徴量 PPI と特徴量 SPD が検知に有効であることを確認した．これは，特徴量 PPI と特徴量 SPD が時間依存性を持たない特徴量となっており，特徴量 PPI と特徴量 SPD を計算する際に，利用するデータセットの時間を長くとることで，スロースキャン攻撃の進行を捉えることができるためと考えられる．

7.　結論

本稿では，スキャン攻撃を行うホストと，正常な通信を行うホストの通信挙動の違いを捉えることのできる 2 つの特徴量，特徴量 PPI と特徴量 SPD を提案した．そして，特徴量 PPI と特徴量 SPD を用いることで，正常な通信を行うホストと，スキャン攻撃を行うホストの分類が可能であることを確認した．さらに，既存の IDS では検知が困難であったスロースキャン攻撃に対し，特徴量 PPI と特徴量 SPD による検知が有効であることを確認した．

今後の課題としては，特徴量 PPI と特徴量 SPD によるホストの分類において，スキャン攻撃を行っていると判断する閾値の決定方法を考える．また，そのうえで特徴量 PPI と特徴量 SPD によるスキャン攻撃検知の性能評価を行うなどが挙げられる．

謝辞

本研究の一部は，JST CREST JPMJCR1783 の支援を受けたものである．

参考文献

[1] 国立研究開発法人情報通信研究機構 サイバーセキュリティ研究所サイバーセキュリティ研究室. Nicter 観測レポート 2018. https://www.nict.go.jp/cyber/report/NICTER_report_2018.pdf.
[2] Shodan. https://www.shodan.io/.
[3] Censys. https://censys.io.
[4] 王サン, フォンヤオカイ, 川本淳平, 堀良彰, 櫻井幸一. 挙動に基づくポートスキャン検知の自動化に向けた学習アルゴリズムの提案とその性能評価. 情報処理学会論文誌, Vol. 56, No. 9, pp. 1770–1781, sep 2015.
[5] 武仲正彦, 鳥居悟, 清水聡. ランダムで低速なポートスキャンの検知についての検討. コンピュータセキュリティシンポジウム 2012 論文集, 第 2012 巻, pp. 736–741, oct 2012.
[6] 嶌田一郎, 津田侑, 衛藤将史, 井上大介. ライブネットにおける低速スキャン検知手法. コンピュータセキュリティシンポジウム 2014 論文集, 第 2014 巻, pp. 458–465, oct 2014.
[7] Jaekwang Kim. A slow port scan attack detection mechanism based on fuzzy logic and a stepwise policy. *IET Conference Proceedings*, pp. 25–25(1), January 2008.
[8] W. El-Hajj, F. Aloul, Z. Trabelsi, and N. Zaki. On detecting port scanning using fuzzy based intrusion detection system. In *2008 International Wireless Communications and Mobile Computing Conference*, pp. 105–110, Aug 2008.

[9] M. Almseidin, M. Al-Kasassbeh, and S. Kovacs. Detecting slow port scan using fuzzy rule interpolation. In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, pp. 1–6, Oct 2019.
[10] M. Dabbagh, A. J. Ghandour, K. Fawaz, W. E. Hajj, and H. Hajj. Slow port scanning detection. In *2011 7th International Conference on Information Assurance and Security (IAS)*, pp. 228–233, Dec 2011.
[11] 西宇基. 固有アクセス率に基づく slow scan 検出法, feb 2013.
[12] Yet another flowmeter. `https://linux.die.net/man/1/yaf`.
[13] RFC 5737 ipv4 address blocks reserved for documentation. `https://tools.ietf.org/html/rfc5737`.
[14] Nmap. `https://nmap.org/man/ja/man-performance.html`.