1ZD-01

ダークネットにおけるアドレス空間規模の変化に伴う可視性の変化

水谷　剛大†　　小谷　大祐‡　　岡部　寿男‡
京都大学‡

1. はじめに

ダークネットから得られる知見をもとに新たなサイバー攻撃への対策が提案されているが、研究機関等の持つダークネットのアドレス空間は異なり、かつ大規模なものが多い[1]。IPv4 アドレスは需要が高い一方で、新規割り当てのための IP アドレス空間が枯渇しており、大規模なダークネットを維持していくことが今後困難になる可能性がある。ダークネットのサイバー攻撃に関する可視性を維持し、そのアドレス空間の大きさを最適化することが望まれる。そのためにはダークネットの規模の変化に伴う可視性の変化について調査する必要がある。先行研究[1]では、異なるダークネットを比較してそれらの類似度を計算することにより、ダークネットの規模とその可視性との関連について報告している。本研究では、異なる規模のダークネットの通信トラフィックに関して、ポートスキャンの傾向の変化点に着目し、それらの可視性の変化について評価した。

2. ダークネットとポートスキャン

2.1 ダークネットとは

未使用の IP アドレス空間のことである。未使用の IP アドレス空間であるダークネットには、本来通信が来ることはないと考えられるが、サイバー攻撃の初期段階として行われるスキャン活動のような、不特定多数を対象とする通信が観測される。つまり、ダークネットではない、一般的な通信に利用されているアドレス空間の場合は、観測された通信から不特定多数に対するサイバー攻撃の通信と、それ以外の通信とを分ける必要があるが、ダークネットの場合は分ける必要がなく、効率的にサイバー攻撃に関するデータを収集することができるため、サイバー攻撃に関する研究で広く利用されている。

本研究で分析に用いるデータは京都大学のダークネット観測網によって収集されているパケットキャプチャデータである。このデータから分析に使用するポート番号やパケット数などの特徴量に関する時系列データを抽出している。

2.2 ポートスキャン

ダークネットで観測される通信トラフィックは主に 3 つに分類される。1 つは、本研究で焦点を当てているサイバー攻撃の初期段階として行われるスキャントラフィック。2 つ目は DDoS 攻撃などによる攻撃者の送信元 IP アドレスのなりすましの影響による跳ね返りトラフィックである。3 つ目は、単純にユーザの入力ミスやルータなど機器の設定ミスによるエラートラフィックである。本研究では TCP SYN スキャンによるポートスキャンに焦点を当て、TCP 通信における SYN パケットを抽出して分析を行う。

3. 評価手法

3.1 評価手法の概要

観測期間 2020 年 3 月 1 日～2020 年 6 月 21 日における、ダークネットのアドレス空間規模の変化による、ポートスキャンの傾向変化に関する可視性を評価するために、異なる規模のダークネットを準備した。まずは、/21 のダークネットでこれを基準とする。この/21 のダークネットを分割する形で、/22（2 分割）、/23（4 分割）、/24（8 分割）のダークネットを定義する。

これらのダークネットを用いた場合に検出されるポートスキャンの傾向の変化点のタイミングがどの程度近いのか、変化点として検出された日の集合をベクトルの形で抽出し、/21 から抽出されたベクトルに対して、/22、/23、/24 のそれぞれから抽出されたベクトルとのコサイン類似度を計算することによって評価する。

3.2 変化点検出に使用する特徴量

ポートスキャンの傾向が変化したことを検出するために使用する特徴量は、どのような脆弱性を標的としているのかに関する情報としてポート番号（/21 のデータにおいてパケット数の絶対量の変化が大きかった 7 つのポート 22, 23, 445, 1433, 2323, 5555, 52869 に焦点を当てる）とそれぞれのポート番号に対応するスキャンパケットの数と送信元 IP アドレスの数の 3 つである。これらの時系列データは 1 時間ごとに集計し、変化点検出の判定も 1 時間単位で行う。

[1] 例：UCSD Network Telescope のダークネット規模は/8
https://www.caida.org/catalog/datasets/telescope-darknet-scanners_dataset/

Change in Visibility due to changing of IP address range using Darknet
† Kodai MIZUTANI, Daisuke KOTANI, Yasuo OKABE・Kyoto University

3.3 変化点検出方法

ダークネットを用いたスキャンの傾向変化検出方法として、定量的にポートスキャンの傾向の変化点を検出するために、[2]の手法と同様にChangeFinder とボリンジャーバンドを組み合わせることによって傾向の変化を捉える。

3.3.1 ChangeFinder とは

時系列モデルの 2 段階学習によって時系列データの変化点のスコアを出力することができる手法である。時系列データの変化が大きいほど出力されるスコアも大きくなる。ChangeFinder のパラメータは、[2]を参考に忘却率 0.02、平滑化区間 48h、AR モデルの次元 1 とした。

3.3.2 ボリンジャーバンドとは

移動平均と標準偏差を用いて、時系列データの異常な値を検出することができる統計学における古典的な手法である。バンド幅を移動平均区間 x と標準偏差の y 倍として定義し、バンド幅に収まらない値を異常な値として検出する。データが正規分布に従うとした時、バンド幅に収まる確率が y=2 の時は約 95.4%となり y=2 とした。移動平均は、[2]と今回のデータの観測期間を考慮し 72h とした。

3.4 ベクトルの作成方法

変化点として検出された日（[2]との対応を取るためにそれに倣って検出の基準を日とした）の集合をもとに、時系列の文脈を含むベクトルにするため、全観測期間 113 日間のそれぞれの日にベクトルの次元が一つずつ対応する 113 次元の零ベクトルを定義し、変化点として検出された日に対応する次元の値のみを 1 とするベクトルを作成した。例えば初日である 3 月 1 日が変化点として検出されている場合はベクトルの 1 次元目が、10 日目である 3 月 10 日が変化点として検出されている場合は、10 次元目が 1 となる。

3.5 コサイン類似度

コサイン類似度は、0～1 の値をとる。コサイン類似度を計算するために用いるベクトルの各次元を観測期間の各日と対応させることで、変化点として検出された日の数だけではなく、その時系列的な近さを考慮できるようにコサイン類似度を用いる。

4. 評価結果

/21 のダークネットに対して、/22,/23,/24 のダークネットのコサイン類似度がどのように変化しているのか図 1 に示す。

ポート番号ごとに類似度の変化の仕方は大きく異なる結果となった。またダークネットの規模が/21 から/22 のように半分になった場合でも、

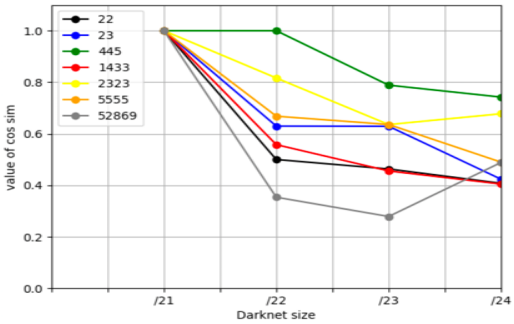

図 1 コサイン類似度（各ポートにおける異なる規模のダークネットの変化点検出に関する）

約半分のポートで類似度は 0.6 を下回った。

5. 考察

本研究のようにポートスキャンの変化点検出に関して、アドレス規模が半減するだけでも類似度が簡単に低下しているということで、ポートスキャンの傾向の変化に関する可視性が、ダークネットの規模変化の影響を受けやすいものであると考えることができ、慎重に扱う必要があると言うことが言える。

また 2323 番ポートや 52869 番ポートのように、ダークネット規模が/24 へ小さくなっている場合でも、/21 とのコサイン類似度が上昇しているケースもある。これは、大きいアドレス空間において偏りが大きかったアドレス範囲が、ダークネットの分割とともにうまく分散したことが原因だと考えることができる。このことから、ダークネットの分割の仕方が、可視性の変化に大きく影響すると考えられる。

6. おわりに

本研究では、変化点検出の可視性の変化を、コサイン類似度を用いて比較することで評価したが、時系列的な類似度についてより適切な評価方法があるのではないかと考えている。次元を大きくしすぎることで、コサイン類似度の解釈がより難しくなってしまうこともあり、より適切な評価方法を活用することが望ましい。

7. 参考文献

[1] F.Soro,et al.;Are Darknets All The Sam-e? On Darknet Visibility for Security Moni-toring. 2019 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN),2019,pp.1-6

[2] 今永大遥、大谷誠、堀良彰;ダークネットを用いた新たなサイバー攻撃傾向の変化検出、情報処理学会火の国情報シンポジウム 2019 C2-2