ダークネットトラフィックデータの解析による サブネットの脆弱性判定に関する研究

西風 宗典† 班 涛‡ 小澤 誠一†

†神戸大学大学院工学研究科
〒 657-8501 兵庫県神戸市灘区六甲台町 1-1
137t252t@stu.kobe-u.ac.jp ozawasei@kobe-u.ac.jp

‡情報通信研究機構ネットワークセキュリティ研究所
〒 184-8795 東京都小金井市貫井北町 4-2-1
bantao@nict.go.jp

あらまし 本研究では，NICTER darknet Dataset 2013 を用いて，マルウェアの活動を観測する，いわゆる動的解析手法の開発を行う．ダークネットとは実在しない IP アドレスのことであり，ダークネットを宛先として設定されているパケットデータは，マルウェアによるスキャン行為，感染行為，DDoS 攻撃のバックスキャッタ，設定ミスなどが考えられる．特定 IP 領域（サブネット）から送出されるパケットデータをポート別に解析し，マルウェア活動を特定するとともに，その感染タイプの分析を行う．これにより，サブネットの脆弱性を判定するシステムの開発を行う．

A Study on Vulnerability Inspection of Internet Subnets by Darknet Traffic Data Analysis

Hironori Nishikaze† Tao Ban‡ Seiichi Ozawa†

†Guraduate School of Engineering, Kobe University
1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, JAPAN
137t252t@stu.kobe-u.ac.jp ozawasei@kobe-u.ac.jp

‡Network Security Research Institute, National Institute of Information and Communications Technology
4-2-1 Nukuikita-machi, Koganei, Tokyo 184-8795, JAPAN
bantao@nict.go.jp

Abstract In the research, we develop a dynamical analysis method of malware activities using NICTER darknet Dataset 2013, which includes unsolicited packet data generated by malware scan activities, infection activities, backscatter by DDos attacks, human setup errors, etc. We analyze packets sent out from specific IP regions (subnetworks) at all the destination ports, and identify the types of malwares. Then, we develop a vulnerability inspection system of subnetworks based on the darknet traffic data analysis.

1　はじめに

近年，コンピュータが急速に普及し，それに伴いネットワークの高速化・大容量化が進んでいる．これにより多くの人が様々な恩恵をインターネットで享受するようになったが，一方で情報の漏洩などを行う不正なプログラム（マルウェア）による信用不安も広がっている．また，パッカーという実行ファイル圧縮ソフトなどの登場によりマルウェア本体の発見が難しい場合もある．よって，プログラムの振る舞いによりマルウェアかどうかを検知する動的解析に注目が集まっている．

マルウェアのいくつかは他のコンピュータの脆弱性の探索（ネットワークスキャン）などの感染行為を行うとき，特定のサーバを攻撃するときに実在しないIPアドレス（ダークネット）を宛先としたパケットを短時間に発生することがある [1]．また，特定のサーバに集中的にパケットを送信し，機能を停止させる攻撃をDDos攻撃というが，この攻撃では虚偽の送信元を記載した，接続の確立を要求するパケットを送信する．よって，これを受信したサーバからの返答が，ダークネットに届くことがある．以上よりダークネットに届くパケットはマルウェアの活動と関連性があると考えられ，その解析によりマルウェアによる感染の程度や拡大の様子を知ることが可能である．

世界中には非常に多くのコンピュータが存在しており，個々の感染を把握するのは容易ではない．そこで，感染しているコンピュータの分布や感染の程度を調査する単位を，サブネットとすることもよく行われる．サブネットの通信トラフィックは，そのサブネットが含むIPアドレスを運用する組織・団体の方針やセキュリティレベルに影響されると考えられる．注視すべき通信トラフィックは，あるサブネットに属するいくつのホストがダークネットにパケットを送信したか，また，どのポートを標的としているか，いくつのIPアドレスを標的としているかといった情報である．サブネットを脆弱性の程度によりグループ化できると，危険なサブネットをあらかじめ知ることができるため，そのサブネットと通信をしている組織に警告を行うことができる．また，危険なサブネットに所属しているユーザにも警告を発し，マルウェアの感染拡大阻止に効果があると考えられる．

本研究では，サブネットの脆弱性を判定することを目標とし，まず，サブネットの通信トラフィックにより，サブネットのクラスタリングを行う．さらに，クラスタリングされたサブネットの解析を通して，クラスタリングの有効性を検証する．

2　実験手法

2.1　サブネットへの分割

ネットワーク感染状況を観測，分析，可視化するため，IPアドレス空間をサブネットに分割する．サブネットの分割については，32ビットのIPv4アドレス空間を2^{16}，すなわち，65536のサブネットに分割する．このとき，各サブネットが含むIPアドレスは65536個となる．これはクラスBのサブネットであり，一般に，大企業や国の機関などが所有している規模である．複数の団体や国にまたがっている場合があるため，これだけで意味があるネットワーク監視ができるとは限らないが，本研究の目的はその仕組みを構築することにあるため，より詳細なネットワーク監視への対応は今後の課題とする．

2.2　特徴ベクトル生成

サブネットごとに，その振る舞いを特徴ベクトルとして表現する．マルウェアによって利用されるポートは決まっていることが多いことから，ポート別に攻撃される頻度を解析をする．各サブネットが送信したパケットを，宛先ポートごとに分類して特徴ベクトルを定義する．各宛先ポートごとにパケットの送信先IPアドレスを取得し，ポートpに対して，サブネットsが何種類の送信先IPアドレスに対してパケットを送信しているかカウントした情報をP_{sp}とする．これによりポート数長のベクトルが生成される．ここで，サブネットのトラフィック分布を比較するため，ベクトルの各成分を宛先ポートと送信先IPアドレスの組合せ総数N_sで除算

して正規化する．さらに各要素に $\log(N_s{+}1)$ を乗算し，N_s による重みづけを行う．N_s は宛先ポートと送信先 IP アドレスの組合せ総数であるため，広範囲のポートや IP アドレスにパケットを送信しているサブネットに対して大きな値をとる．つまり，広範囲のポートや IP アドレスをスキャンするマルウェアの活動は危険度が高いと考えて，そのようなトラフィックをもつサブネットを区別するために $\log(N_s{+}1)$ を乗算する．以上より，特徴ベクトル $\boldsymbol{x}_s = \{x_{sp}\}_{p=1}^{65536}$ の各要素 x_{sp} は，

$$x_{sp} = \frac{P_{sp}}{N_s}\log(N_s + 1) \tag{1}$$

と表す．

2.3 クラスタリング

類似した通信トラフィックをもつサブネットをグループ化するため，クラスタリングを行う．クラスタリングは，階層的手法である最長距離法 [2] を用いる．特徴ベクトル総数 M とし，クラスタ C_i に含まれる特徴ベクトルを $\boldsymbol{x}_{ia}$ とする．最初に，クラスタを M 個だけ設け，各特徴ベクトル $\boldsymbol{x}_{ia}$ をクラスタ C_i とする．クラスタ間の距離 $\mathrm{d}_1(C_i, C_j)$ を，式 (2) を用いて計算し，この距離が最も近い 2 つのクラスタを併合する．ここで，$\mathrm{d}_2(\boldsymbol{x}_{ia}, \boldsymbol{x}_{jb})$ はベクトル同士のユークリッド距離を表す．この操作を繰り返すことによりクラスタを形成する．

$$\mathrm{d}_1(C_i, C_j) = \max_{a,b}(\mathrm{d}_2(\boldsymbol{x}_{ia}, \boldsymbol{x}_{jb})) \tag{2}$$

$$\boldsymbol{x}_{ia} \in C_i, \boldsymbol{x}_{jb} \in C_j \tag{3}$$

2.4 次元削減による可視化

クラスタリング結果が妥当であるかを確認するため，可視化を行う．各サブネットのトラフィックを表す特徴ベクトルはポート数（65536）次元のベクトルであるため，可視化するために次元削減を行う．次元削減には，多様体上のデータ構造を学習する Isomap[3] を用いる．Isomap において，多様体上の距離（測地距離）情報から低次元空間への点の配置は MDS[3] という手法を用いている．MDS のアルゴリズムを Agorithm 1 に，Isomap のアルゴリズムを Algorithm 2 に示す．

Algorithm 1 Multi-Dimensional Scaling (MDS)

1: Input $\boldsymbol{D} \in \boldsymbol{R}^{n\times n}, D_{ii} = 0, D_{ij} \geq 0$.
2: Set $\boldsymbol{B} = -\frac{1}{2}\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}$, where $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^t$ is the centering matrix.
3: Compute the spectral decomposition of $\boldsymbol{B} : \boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^t$.
4: Form $\boldsymbol{\Lambda}_m$ by setting $[\boldsymbol{\Lambda}_m]_{ij} = \max(\Lambda_{ij}, 0)$.
5: Set $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}_m^{\frac{1}{2}}$.
6: Return $[\boldsymbol{X}]_{n\times d}$.

Algorithm 2 Isomap

1: Input $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, k$.
2: Form a k-nearest neighbor graph with edge weighs $W_{ij} = ||\boldsymbol{x}_i - \boldsymbol{x}_j||$ for neighboring points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.
3: Compute the shortest path distances between all pairs of points using Dijkstra's or Floyd's algorithm.
4: Store the squared distances in $\boldsymbol{D}$.
5: Return $\boldsymbol{Y} = MDS(\boldsymbol{D})$.

3 実験と考察

3.1 クラスタリングの有効性

本実験では，NICT が提供している NICTER darknet Dataset 2013[4] という，NICT のダークネットに送られたパケットデータ（pcap 形式）を用いた．2011 年 1 月から 2013 年 8 月の中旬までの約 31 か月分のパケットの情報である．このデータの 2011 年 1 月分（約 686 万パケット）を用いて実験を行った．活動が活発なサブネットを選ぶため，4 種類以上のホストが NICT のダークネットにパケットを送信したサブネットを解析対象とした．解析対象とした 1296 サブネット分のパケットデータを，2.1 節の通りサ

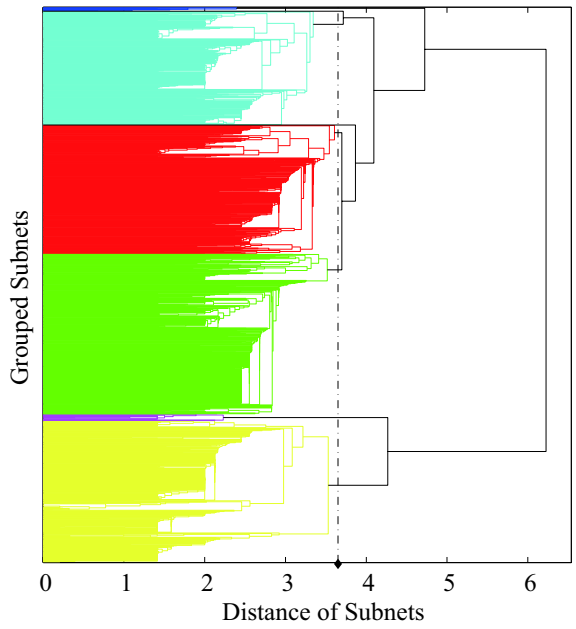

図 1: 階層構造のデンドログラム

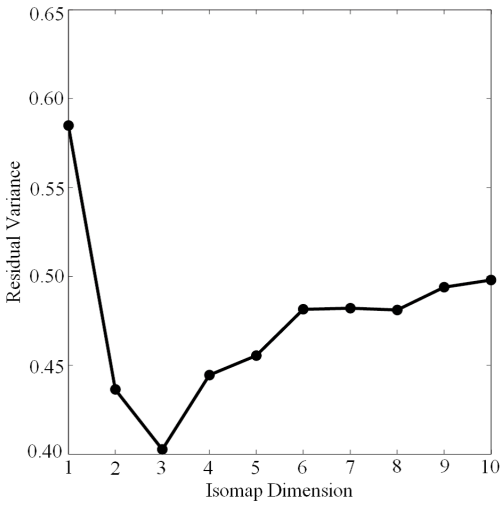

図 2: Isomap の誤差と次元の関係

ブネットにより分割し，それぞれのサブネットにおいて 2.2 節の通り特徴ベクトルを生成した．また，2.3 節に従い階層クラスタリングを行い，8 クラスタを生成した．その結果を，横軸を距離とした 1 次元のデンドログラムとして図 1 に示す．図 1 において，クラスタごとに色を分けて示している．これより，特徴ベクトル同士のユークリッド距離によりクラスタリングができていることが確認できる．

図 1 ではポート数（65536）個の要素がある特徴ベクトル同士の距離を用いて 1 次元で表示している．この次元圧縮による結果が正しいものであるか検証するため，2.4 節の手法を用いて次元削減を行った．Isomap を用いて高次元から L 次元に変換した際，L の値によりどの程度誤差が発生しているかを図 2 に示す．図 2 より，本実験では 3 次元に圧縮すると元データとの分布の誤差が小さいことがわかる．よって，3 次元空間に変換した．クラスタごとに異なる記号でプロットを行った結果を図 3 に示す．

図 3 より，同じクラスタにクラスタリングされた特徴ベクトルは次元圧縮した空間においても近い位置に存在しており，本実験で行ったクラスタリングが妥当であることが確認できた．

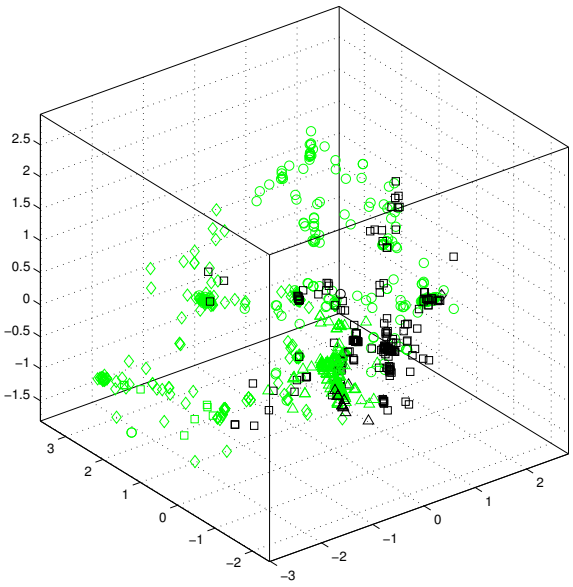

図 3: Isomap によるクラスタリング結果の可視化

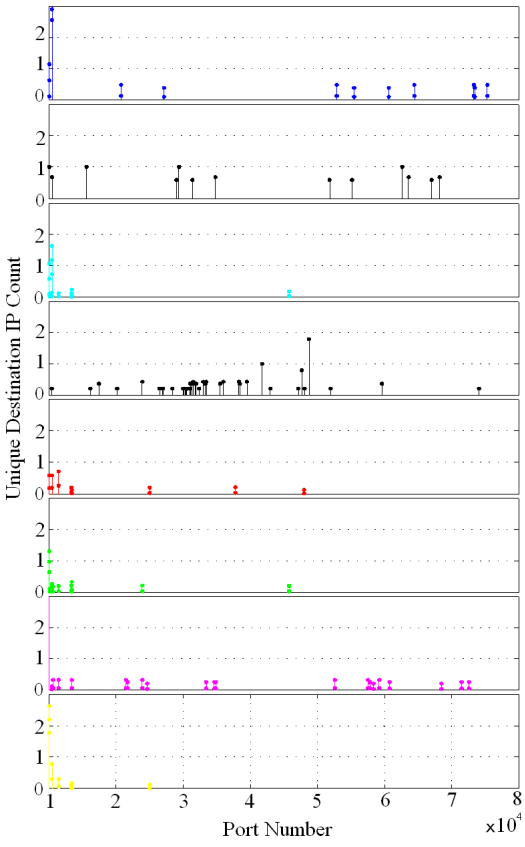

図 4: 各クラスタに属する特徴ベクトルの傾向

3.2 クラスタのトラフィック特徴

生成した 8 クラスタに対し，宛先ポート特徴量の平均と標準偏差をローソク足表示したものを図 4 に示す．なお，図 4 の色は図 1 と対応させており，平均値が 0.01 以下の場合は表示していない．図 4 より，クラスタごとに異なるトラフィック特徴をもっていることがわかる．

4 まとめ

サブネットの脆弱性判定を目標とし，サブネットの通信トラフィックの特徴による分類を行い，その結果を可視化した．NICTER darknet Dataset 2013 の通信データを用いた実験において，IP アドレスをサブネット単位に分割し，そのサブネットの通信トラフィックを解析することで，サブネットのクラスタリングを行うことができた．特徴ベクトル同士の距離を基準に，クラスタリングを行い階層構造で表示し，次元削減を行って得た結果と比較することで，クラスタリングの有効性を確認した．また，クラスタリング結果において，クラスタごとにトラフィックの特徴を図示することで，クラスタごとに異なるトラフィック特徴をもっていることが確認できた．

サブネットの脆弱性判定や危険度推定を行うには，各クラスタが，どの程度危険であるかランク付けする必要がある．これについてはあらかじめ収集が可能なマルウェアの情報と関連付ける必要があり，今後の研究課題としたい．また，本実験では 1 か月分のデータのみを用いて実験を行ったが，他の期間のデータに対しても本手法が有効であるか，引き続き検討していく．

参考文献

[1] 中尾康二, 井上大介, インシデント分析センター nicter の研究開発概要, 情報通信研究機構季報, Vol. 57, pp. 3 - 16, 2011.

[2] 神蔦敏弘, データマイニング分野のクラスタリング手法 (1), 人工知能学会誌, Vol. 18, pp. 59 - 65, 2003.

[3] L. Cayton, “Algorithms for Manifold Learning,” *UCSD Tech Report CS2008 - 0923,* 2008.

[4] 神薗雅紀, “マルウェア対策のための研究用データセット (MWS Datasets 2013),” 2013.

[5] K. Nakano, K. Yoshida, D. Inoue, M. Eto , “A Novel Concept of Network Incident Analysis based on Multi-layer Observation of Malware Behavior,” *The 2nd Joint Workshop on Information Security,* pp. 267 - 279, 2007.

[6] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, S. Sinha, “Practical Darknet Measurement,” *Information Sciences and Systems,* pp. 1496 - 1501, 2006.

[7] D. Inoue, K. Yoshida, M. Eto, M. Yamagata, E. Nishino, J. Takeuchi, K. Ohkuchi, K. Nakano, “An Incident Analysis System NICTER and Its Analysis Engines Based on Data Mining Techniques,” *Proc.*

International Conference on Advances in Neuro-Information, pp. 579 - 586, 2008.

[8] K. Nakano, D. Inoue, M. Eto, K. Yoshida, "Practical Correlation Analysis between Scan and Malware Profiles against Zero-Day Attacks Based on Darknet Monitoring," *IEICE Transactions on Information and Systems,* Vol. 92, No.5, pp. 787 - 798, 2009.

[9] N. Provos, "A Virtual Honeypot Framework," *Proc. USENIX Security Symposium,* pp. 1 - 14, 2004.