支配的なトラフィックの変化に着目したダークネット通信分析

金井 登威† 角田 裕† キニ グレン マンスフィールド‡

†東北工業大学
982-8577 宮城県仙台市太白区香澄町 35-1
m142802@st.tohtech.ac.jp tsuno@m.ieice.org

‡株式会社サイバー・ソリューションズ
989-3204 宮城県仙台市青葉区南吉成 6-6-3 ICR ビル 3F
glenn@cysols.com

あらまし 本研究では支配的なトラフィックに着目してダークネットの分析を進めている．インターネットのある地点を流れるトラフィックは概ね一定数の種類のものにより支配的であるという考え方があり，これはダークネットでも例外ではないと思われる．そこで，NICTER Darknet データセットのトラフィックに対して同様の考え方が成り立つのかを調査した．本研究では，支配的なトラフィックを，特定のフィールドの値が一致しているパケットをグループ化した際，全パケット数に対して一定割合以上のパケット数を持つグループとし，そのグループ数をTop$N$と定義した．4年間のデータに対し5つのフィールド毎でTop$N$を日毎に算出し変化を調べた結果，複数のフィールドで連動してTop$N$が増減している期間をいくつか発見した．本発表では，その調査結果に加えTop$N$に着目することで得られた情報について報告する．

Darknet Traffic Analysis by Focusing on Variations in Dominant Traffic

Toui Kanai† Hiroshi Tsunoda† Glenn Mansfield Keeni‡

†Tohoku Institute of Technology
35-1, Yagiyama Kasumi-cho, Taihaku-ku, Sendai, Miyagi, 982-8577, JAPAN
m142802@st.tohtech.ac.jp tsuno@m.ieice.org

‡Cyber Solutions Inc.
ICR Bldg. 6-6-3, Minami Yoshinari, Aobaku, Sendai, Miyagi, 989-3204, JAPAN
glenn@cysols.com

Abstract In this research, we analyze the dominant components in darknet traffic. In the Internet, it is known that there are dominant traffic components, and these components characterize the traffic. We apply this principle to darknet traffic analysis. In our analysis, a dominant component is one for which the corresponding packet group has a traffic-occupancy larger than a pre-defined ratio. We analyze darknet traffic for a span of 4 years, provided by NICTER. This is the Darknet dataset 2015. We found several significant variations in the number of packet groups constituting the dominant components. By investigating the reason of those changes, we show that the number of packet groups in the dominant component provides useful information for darknet traffic analysis.

1 はじめに

ダークネットとは，到達可能かつ特定のホストが割り当てられていない IP アドレス空間を指す．通常，ダークネットに通信が届くことは考えづらいが，実際には日々大量の通信が観測される．それらは，マルウェアによるスキャンや送信元 IP アドレスが詐称された DoS 攻撃の跳ね返り（以下，バックスキャッタ）などの不正な通信である．そのため，ダークネットで観測されたトラフィックを分析することでインターネットでの不正活動の傾向を把握する研究が広く行われている[1][2][3][4].

我々は支配的なトラフィックに着目してダークネットトラフィックの分析をしている[5]．インターネットのある地点を流れるトラフィックは通常時には概ね一定数の種類のものにより支配的であるという考え方があり，それをダークネットトラフィックに適用し分析を行っている．ダークネットで観測されるパケットの多くは基本的に悪意があるものとされるが，常に同じ攻撃により発生したパケットがダークネットに届くのなら，ここで観測されるトラフィックでも一定数の種類のものにより支配的になるはずである．そして，新種のマルウェアからの通信といった，今までは見られなかった種類のパケットが観測されることでダークネットの支配的なトラフィックの傾向は変わると予想する．

本研究では，支配的なトラフィックを全体に対して一定割合以上のパケット数を持つグループと定義し，支配的なトラフィックを構成するグループ数を示すTop$N$の変化に着目してダークネットトラフィックを分析する．支配的なトラフィックとは，言うなればその時点で最も用いられているパケットの種類のため，その数を表すTop$N$からダークネットに届くパケットの傾向の変化を知ることが出来る．

以下，2 章ではダークネット分析の関連研究について述べ，3 章では支配的なトラフィックとTop$N$の考え方を述べる．4 章ではダークネットトラフィックに対し提案手法を適用して得られた結果からの分析を報告し，5 章でまとめとする．

2 関連研究

本章では運用されているダークネット観測システム及びダークネット分析の関連研究について述べる．

2.1 ダークネット観測システム

ダークネット観測システムは国内外で様々な組織が運用している．

情報通信研究機構（以下，NICT）[6]では，インターネット上におけるセキュリティインシデントの早期検知・分析・対策の確立を目的としたNICTER (Network Incident Analysis Center for Tactical Emergency Response)の研究開発を行っており，約 24 万の IPv4 アドレスを持つダークネットの観測と分析を行っている．

JPCERT コーディネーションセンターのTSUBAME システム[7]では，ダークネットセンサを国内外の多数の地点に分散配置している．そして，収集したトラフィックを共有・分析することで地域毎でのマルウェアの活動やスキャンの動向を把握し，国内に脆弱性関連情報や影響範囲の広い深刻な情報セキュリティ上の脅威の注意喚起の発信を行っている．

総務省の PRACTICE プロジェクト[3][8]は，プロジェクトに参加国のダークネットにセンサを設置しトラフィックを収集している．そのトラフィックは各国で共有され国際連携によるサイバー攻撃の対策のため分析が行われている．

米国では，CAIDA (Center for Applied Internet Data Analysis) の The UCSD Network Telescope プロジェクト[3][9]が/8 の大規模なダークネットを観測し，一部のデータを公開している．

2.2 ダークネット分析の関連研究

文献[10]では，ヘッダ情報を利用して通信源ホストの分類をし，ペイロードを含まないダークネットトラフィックの傾向を分析している．

ダークネットで観測されたバックスキャッタを

対象とした分析も行われおり，文献[1]では宛先IPアドレスや送信先ポート番号の特徴に着目してバックスキャッタを分類し，発生原因となったDoS 攻撃の特徴を考察している．文献[4]でも同様にバックスキャッタに含まれるパラメータから特徴を抽出し，使用されたDoSツールの特定を行っている．

異なるダークネットのトラフィックを利用してネットワークの傾向を把握する研究も行われている．文献[11]では，日本と海外のダークネットトラフィックの相関分析を行っており，文献[12]では，複数国のダークネットトラフィックのポートなどの情報を用いて分析をしている．どちらの研究でも，異なるダークネットトラフィックの分析結果を突合することで，多くの国で共通して観測される広域な不正通信の存在や，一部の国でしか見られない局地的な不正通信の特徴を発見している．

また，特定の地域からのトラフィックがダークネットで観測されるかに着目することで，災害時のインターネット死活監視にダークネットを活用する研究も行われている[13]．

3 支配的なトラフィックの変化に着目した分析手法

本章では，本研究で提案する支配的なトラフィックに着目したダークネット分析の手法を述べる．

支配的なトラフィックとは，全パケットに対して大部分を占めているパケットグループである．従来では，同じ種類のパケットをまとめ，個数が多い上位 X 種類（TopX）のグループを支配的なトラフィックとする方法が一般的である．これを，ダークネットで観測されたパケットの宛先ポートに適用することで攻撃者が頻繁に狙うサービスという特徴が分かる[11]．しかし，着目するポートの種類数を固定にすると次のような問題が発生する．仮に，単純にTop10として固定してしまうと，11 種類以上のポートに対してスキャンが行われれば必要な情報を取りこぼしてしまう．逆に 1 種類のポートに集中した DDoS の場合には，残り 9 種類のポートには本来着目する必要はないはずである．そこで，固定ではなくパケットの種類毎の割合により支配的なトラフィックを動的に決める．

本研究で支配的なトラフィックとは，全パケットに対して一定以上の割合を占めているパケットのグループとし，これを構成するパケットグループ数をTop$N$と表す．そして，支配的なトラフィックを構成するTop$N$個のパケットグループをTop$N$グループと記す．図 1 に宛先ポートのフィールドを対象とした場合のTop$N$の算出方法を示す．

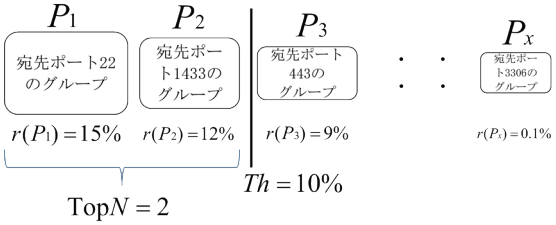

図 1 支配的なトラフィックとTop$N$の算出方法

図 1 のP_iはパケットのグループを表し，そのグループのパケット数を$\mathrm{n}(P_i)$と表す．ただし，$\mathrm{n}(P_i) > n(P_{i+1})$とする．このとき，グループ$P_i$の支配率$r(P_i)$を式(1)で算出し，それが閾値$Th$以上になるグループ数を式(2)によって求めTop$N$とする．

$$\boldsymbol{r(P_i)} = \frac{\boldsymbol{n(P_i)}}{\sum \boldsymbol{n(P_x)}} \times \ \boldsymbol{100}\ (\%) \qquad (\boldsymbol{1})$$

$$\mathbf{Top}\boldsymbol{N} = \max \boldsymbol{i} \text{ where } \boldsymbol{r(P_i)}$$

2 つの式から，図 1 の場合ではTop$N$は 2 となり，Top$N$グループは 22 と 1433 となる．

Top$N$により注目するべきパケットの種類数をスロット毎で動的に変更することができ，種類数を固定にした場合と比べスロット毎に大量に届く種類のパケットを過不足なく抽出できると考えられる．また，支配的なトラフィックとは言わばその時点で最も用いられているパケットの種類のため，その数を表すTop$N$が増加した場合には頻繁に使用されるようになった送信元ポートが出現した，減少した場合には特定のホストか

らのパケットが大量に届いたといった傾向の分析ができると思われる.

4 章では, Top*N*の変化を基にダークネットトラフィックを分析した結果を報告する.

4 Top*N*の推移による分析

本研究では NICTER Darknet データセット [14] の 2011 年～2014 年のトラフィックから以下の 2 つの種類のパケットそれぞれについて Top*N*の変化を調査した.

① TCP/UDP パケット
② バックスキャッタ

なお, バックスキャッタは SYN Flood 攻撃の結果として発生する TCP の SYN-ACK パケットとした.

これらのパケットに対し, 以下の 5 つのフィールドの値を基にパケットをグループ化し, それぞれのTop*N*を日毎に算出した.

- 送信元アドレス
- 宛先アドレス
- 送信元ポート
- 宛先ポート
- TTL

なお, Top*N*を求める閾値*Th*は 1.0%, 0.75%, 0.5%のそれぞれについて調査した. 以降では, Top*N* の変化が最も顕著であった閾値 *Th*=0.5%の場合について述べる.

4.1 観測期間全体のTop*N*の推移

2011 年～2014 年におけるそれぞれのパケットのTop*N*の 30 日毎の移動平均を図 2 と図 3 に示す.

図 2と図 3から, Top*N*はフィールドにより大きく違うことが見て取れる. これは, フィールド毎に対象とする値の最大値が違うためである. 送信元及び宛先ポートは 65,536 種類であり, TTL では 256 種類である. 宛先アドレスはネットワークの環境により異なり, NICTER Darknet データセットにおいては 4,096 種類である. また, 送信元アドレスは全てのアドレス空間が対象となるため, IPv4 では最大で約 43 億種類である. このように, フィールドにより対象とする値の種類数が大きく異なるためTop*N*の大きさも異なる. それでも, 図 2と図 3 でTop*N*は最大でも 60 以内に収まっており, それぞれのフィールドにおいて注目すべき種類数の尺度を合わせることができる. これにより, 種類数が異なるフィールド毎の比較が直感的に行えるようになるというメリットがある.

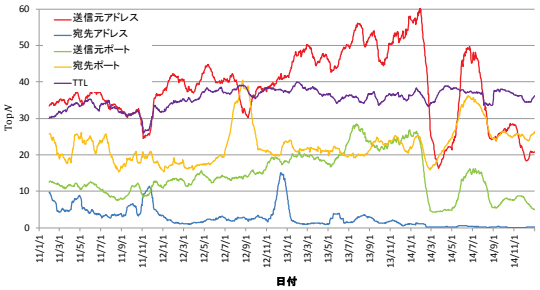

図 2 TCP/UDP パケットの 5 つのフィールドのTop*N*の推移

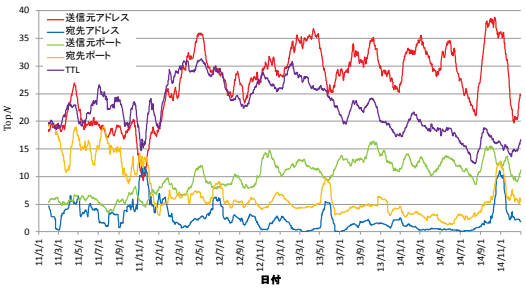

図 3 バックスキャッタの 5 つのフィールドのTop*N*の推移

図 2 から TCP/UDP パケットに関する送信元アドレスと送信元ポートのTop*N*は時間と共に増加傾向にあることがわかる. これは, 大量のパケットを送るホストが年々増加しているためだと考えられる. 宛先ポートのTop*N*は対象期間の始まりと終わりで大きな差はないため, パケットが頻繁に届きやすいポートの種類数は変わっていないことが伺える. 特徴的な傾向として, 2014 年初頭にこれらの 3 つのフィールドのTop*N*が連動して大きく減少したことが挙げられる. この原因については 4.2 節で考察する.

図 3 からバックスキャッタのパケットは, 送信

元アドレスと TTL のTop*N*が増加傾向にある．このことから年々DoS 攻撃の規模が大きくなっている，あるいは狙われるホスト数が増加していることが伺える． 特徴的な傾向として，2011 年 11 月始め頃にバックスキャッタの送信元アドレスと TTL に関するTop*N*が大きく減少していた．この原因については 4.3 節で考察する．

4.2 TCP/UDP パケットのTop*N*が急激に減少した期間に発生していた攻撃

4.1 節で述べた通り，2014 年 1 月末にTCP/UDPパケットに関して複数のフィールドのTop*N*が連動して大きく減少していた． 図 2 から送信元アドレス，送信元及び宛先ポートの2013 年 11 月～2014 年 4 月におけるTop*N*を抽出したグラフを図 4 に示す．

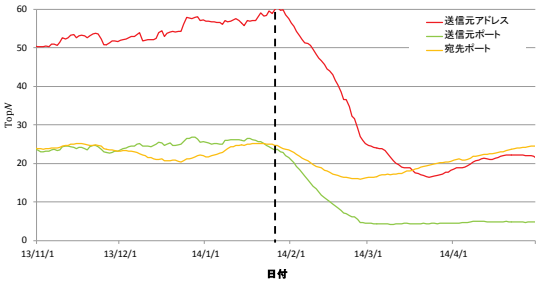

図 4 2014 年初頭の 3 つのフィールドのTop*N*の推移

同時期における TCP/UDP の日毎のパケット数の推移を図 5 に，フィールド毎のユニークな値の数及び全パケットに対するTop*N*グループの総パケット数の支配率を図 6 と図 7 に示す．なお．図 6と図 7 は 30 日毎の移動平均で表している．

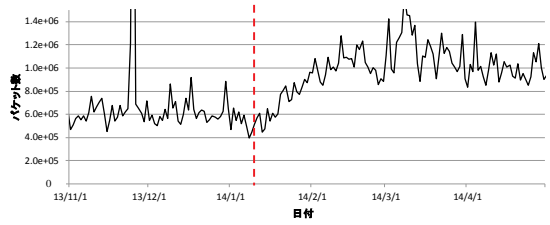

図 5 TCP/UDP パケット数の推移

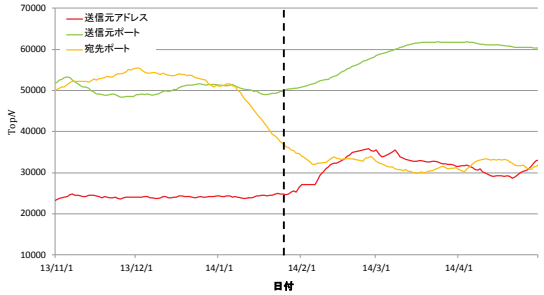

図 6 フィールド毎のユニークな値の数

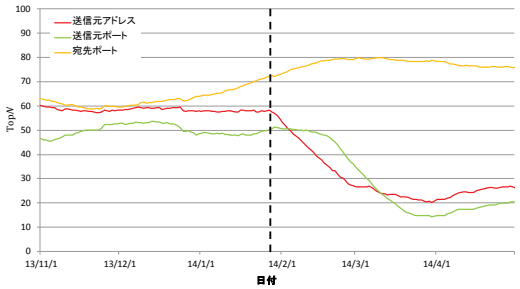

図 7 フィールド毎の全パケットに対してのTop*N*グループの支配率

図 4～図 7 から 2014 年初頭を境に以下のような変化が起きていることがわかる．

- 総パケット数が増加(図 5)
- 各フィールドの値のユニーク数が変化(図 6)
  - ◆ 宛先ポートは増加
  - ◆ 送信元アドレスとポートは減少
- Top*N*は減少(図 4)
- Top*N*グループの総パケット数は減少(図 7)

以上のことから，この期間に同じポートを使用して特定の宛先ポートに大量にパケットを送るホストの通信がTop*N*グループに加わったと考えられる．そのため，他のパケットの支配率が相対的に小さくなり，Top*N*やTop*N*グループの総パケット数が減少したと見られる．ただ，宛先ポートのTop*N*グループに含まれる総パケット数は増加していた．これは，Top*N*の減少が他の 2 つのフィールドと比べ小さいことから，常にTop*N*グループに含まれている宛先ポートが攻撃のターゲットとなったことが原因であると考えている．それにより，他のパケットの支配率は

相対的に小さくなったが，Top*N*グループに加わったパケット数に対して外れるパケット数が少なかったためTop*N*は減少しTop*N*に含まれるパケット数は増加したと思われる．

この考察の裏付けのため，Top*N*グループの送信元及び宛先ポート番号を調査した．図 8 と図 9 は 2013 年 11 月～2014 年 5 月について，各日におけるTop*N*グループの送信元・宛先ポート番号を月毎に分けプロットしたものである．

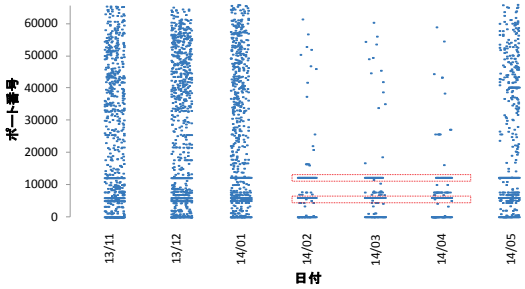

図 8　Top*N*グループの送信元ポート番号

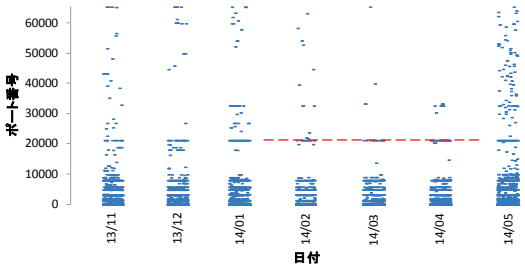

図 9　Top*N*グループの宛先ポート番号

図 8 に破線で示す通り，2014 年初頭には送信元ポート 6000 番と 12200 番から大量にパケットが送られていた．また，図 9 から，2014 年初頭では赤線で示す 21320 番より大きいポート番号のほとんどがTop*N*グループから外れており，10000 番以下のポートにパケットが集中していることがわかる．よって，6000 番と 12200 番ポートから 10000 番以下の宛先ポートに集中してパケットが送られたことで，21320 番より大きい宛先ポート番号のグループがTop*N*グループからまとめて外れたため，Top*N*が大きく減少したと思われる．

MWS Cup 2014 でもこの期間に同様の特徴を発見したチームがあり，6000 番ポートは 1433（MSSQL），4899（Radmin），3306（MySQL），22（ssh）の探索を，12200 番ポートは 1080（SOCKS プロキシ），8080，21320、3128，1998 のプロキシを探索していたとの報告がある[15]．使用されている送信元及び宛先ポート番号から同じ事象だと見られる．

これらのことから，Top*N*では同じ種類のパケットが送られるといった事象が反映されていると考えることができ，スキャンを効果的に発見できるといえる．

4.3　バックスキャッタのTop*N*が大きく増減した期間に発生していた事象

4.1 節で述べたとおり，2011 年 10 月 31 日～11 月 3 日にバックスキャッタに関して 2 つのフィールドのTop*N*が大きく減少していた．2011 年における 1 日あたりのバックスキャッタのパケット数の平均値は約 15.6 万件であるが，11 月 1 日には最大で約 72 万件のパケットが観測されていた．このことから，大規模な攻撃が発生したため 2 つのフィールドのTop*N*が大きく変化したと見られる．

2011 年 10 月～11 月の全バックスキャッタのパケットに対して送信元アドレスと TTL のTop*N*グループの総パケット数を図 10 に示す．

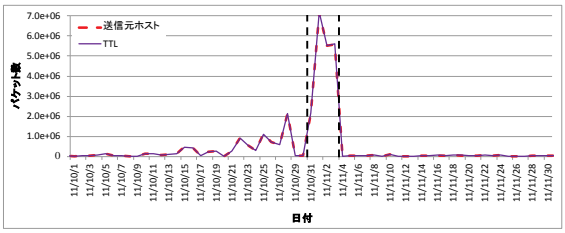

図 10 Top*N*グループの総パケット数

図 10 から，2011 年 10 月 31 日～11 月 3 日においては，送信元アドレスのTop*N*グループの総パケット数が最も多い 11 月 1 日には約 74 万件であり，次に多かった 3 日で約 58.7 万件ものパケットが届いていた．加えて，TTL のTop*N*グループの総パケット数は 1 日が約 72

万件，3 日は 56.5 万件であり送信元アドレスのパケット数とほぼ同数であるが分かる．このことから，大量に届いたバックスキャッタは同じホストから送られたと考えることができる．つまり，特定のホストに大規模な DoS 攻撃が行われたことにより発生したバックスキャッタが原因で Top*N*が増減したと考えられるため，パケットを大量に送っていたホストを調査した．

図 11 は 2011 年 10 月 29 日～11 月 6 日の 9 日間についてTop*N*グループの送信元アドレスを日毎にプロットした散布図となる．なお，縦軸は送信元アドレスを 10 進数で表している．

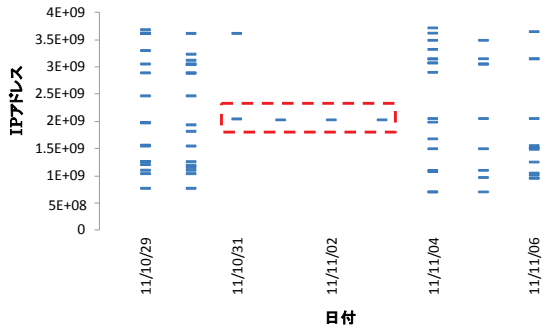

図 11 Top*N*グループである送信元アドレス

図 11 から，2011 年 10 月 31 日～11 月 3 日の 4 日間では破線に囲まれたアドレスから大量のバックスキャッタが送られていたと考えられる．調査したところ，このアドレスは中国のホストに割り当てられたものであった．関連研究[1]でも 2011 年 11 月 1, 2 日に中国のある 1 台のホストから 7000 万件を超えるバックスキャッタを観測したと述べられており，この期間に大規模な DoS 攻撃が発生したと思われる．

今回発見した事象の影響により，送信元アドレス，宛先アドレス，TTL と複数のフィールドの Top*N* が大きく変化していた．このことから，Top*N*の変化が大きいフィールドや変化の度合いから発生した攻撃や規模がある程度予測できると考えている．

4.4 Top*N*グループから得られる情報

本節では 2013 年 1 月 26 日のトラフィックを例に，支配率に基づいて求めたTop*N*とTop10のそれぞれで分析した場合の結果の違いを述べる．図 12 に 2013 年 1 月 26 日におけるTop*N*グループの支配率を示す．なお，この日の TCP/UDP パケットに関する宛先ポートの Top*N*は 22 であった．

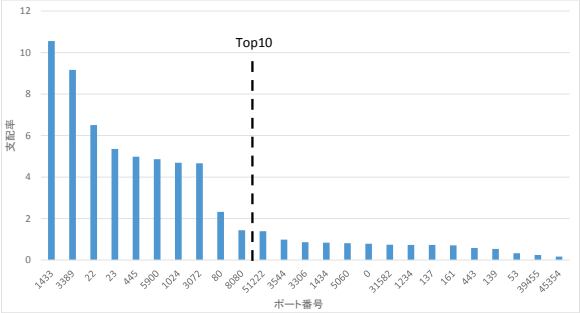

図 12 2013 年 1 月 26 日の Top*N*グループの支配率

図 12 において上位 10 番目である 8080 番ポートの支配率は約 1.44%だが，上位 11 番目である 51222 番のそれは 1.39%と大きな違いはない．しかし，Top10までに着目すると，この 51222 番に関するパケットは支配率に大きな差がないにも関わらず，調査対象から外れてしまう．

この 51222 番について調査したところ，午前 3 時～4 時の 1 時間にわたり，882 台のホストから 1 つのセンサに向けて 8,149 個ものパケットが送られていた．また，前後の日付である 2013 年 1 月 20 日～30 日の 11 日間でこのポートにパケットが届いているか調べたところ，26 日を除く全ての日で数個あるいはまったく観測されていなかった．一般には公開されていないダークネットセンサに対して大量のパケットが送られたのは，ランダムで生成した IP アドレスを狙った DDoS の可能性がある．

このことから，見るべきパケットの種類数を固定にしてしまうと，それ以降にある注目すべき情報も取りこぼしてしまう恐れがある．しかし，Top*N*により種類数を動的に変更することで，それらの事象に気づくことができる．

5　まとめと今後の課題

本稿では，NICTER Darknet データセットに対して，支配的なトラフィックの変化に着目して分析を行った．

分析の結果，複数のフィールドのTop*N*が大きく変化している期間にはスキャンや大量のバックスキャッタが届いていたことが判明した．またTop*N*を用いることで，見るべきパケットの種類数を固定にした場合では取りこぼしてしまう情報も観測できることを示した．

今後の課題としては，Top*N*を日毎以外で算出して分析をすることや現在の閾値の妥当性を評価することが挙げられる．

謝辞

本研究で用いたデータセットを提供して頂いた NICTER の関係者の皆様に深く感謝します．

参考文献

[1] 井上大介,中里純二,衛藤将史,中尾康二," DoS 攻撃:3.3 DoS/DDoS 攻撃対策（3）～ダークネット観測網を用いたバックスキャッタ分析～", 情報処理, Vol. 54, No. 5, May 2013

[2] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha, "Practical Darknet Measurement" in Proceedings of the 40th Annual Conference on Information Sciences and Systems (CISS '06) , pp. 1496-1501, Princeton, New Jersey, USA, March 2006

[3] 鈴木将吾，小出駿，牧田大祐，村上洸介，笠間貴弘，島村隼平，衛藤将史，吉岡克成，松本勉，井上大介，"複数国ダークネット観測による攻撃の局地性分析", Computer Security Symposium 2014, 22-24 October 2014

[4] 中里純二，島村隼平，衛藤将史，井上大介，中尾康二，"nicter によるネットワーク観測および分析レポート ~DDoS 攻撃によるバックスキャッタの推移と分類~", 信学技報, IA2012-7, ICSS2012-7 (2012-06)

[5] IPA 情報セキュリティ対策研究開発評価等事業　高トラヒック観測・分析法に関する技術調査 http://www.ipa.go.jp/security/fy15/reports/traffic_mon/documents/traffic_mon.pdf

[6] 国立研究開発法人 情報通信研究機構 NICT Cyber security Laboratory http://nict.go.jp/nsri/cyber/research.html

[7] TSUBAME（インターネット定点観測システム）, http://www.jpcert.or.jp/tsubame/

[8] PRACTICE Dataset 2013 http://www.iwsec.org/mws/2013/files/20130612_PRACTICE-Dataset-2013.pdf

[9] CAIDA (Center for Applied Internet Data Analysis) PROJECTS - The UCSD Network Telescope http://www.caida.org/projects/network_telescope/

[10] 笹生憲，森達哉，後藤滋樹，“通信源ホストの分類を利用したダークネット通信解析”, Computer Security Symposium 2013, 21-23 October 2013

[11] 深澤成孝，佐藤直，"ダークネットトラフィックの相関分析", Vol. 2015-DPS-162 No. 20 ,Vol. 2015-CSEC-68 No. 20

[12] 村上洸介，蒲谷武正，千賀渉，鈴木将吾，小出駿，島村隼平，牧田大祐，笠間貴弘，衛藤将史，吉岡克成，井上大介，中尾康二，“複数のダークネット観測拠点で同時期に急増する攻撃を検知する手法の提案”, Computer Security Symposium 2014, 22-24 October 2014

[13] 井上大介，中里純二，島村隼平，衛藤将史，中尾康二，“災害時における大規模ダークネット観測網の活用に関する検討“，信学技報，IA2011-5, ICSS2011-5 (2011-06)

[14] NICTER Darknet 2014 http://www.iwsec.org/mws/2014/files/NICTER_Darknet_Dataset_2014.pdf

[15] MWS Cup 2014, 事前課題 4「Darknet Traffic Analysis」解答例, http://www.iwsec.org/mws/2014/files/mws_cup_2014_pre4.pdf