日本ソフトウェア科学会第 27 回大会 (2010 年度) 講演論文集

Darknet に到着する UDP パケットの特徴解析

大田 昌幸　福田 健介　廣津 登志夫　菅原 俊治

インターネット上には異常トラフィックが常に大量に流れており，それらの異常トラフィックに対する検知・防御が求められている．異常トラフィックを検知・防御するためには，異常トラフィックを解析し，その特徴をつかむ必要がある．そのため，Darknet に到着するパケットをプロットし散布図として確認するとともに，異常トラフィックの特徴を把握してきた．これまでは，パケット数が多いことから TCP パケットのうち syn フラグのみが 1 のパケットを解析し，1 つのソースアドレスからであっても IP ヘッダの情報や OS 情報を活用して，いくつかの系列に分けられることを示してきた．本稿では，Darknet に到着する異常トラフィックのうち，これまで解析の対象外となっていた UDP パケットを解析し，その系列を分類することを試みる．

1　序論

インターネットと情報機器の急速な発展により，インフラストラクチャとしてのインターネットの重要性は飛躍的に高まってきた．インターネットを利用した商取引や娯楽などのサービス提供は現在の社会活動において必要不可欠である．しかし，インターネット上のあらゆるサービスは (D)DoS, Virus, Worm などの異常トラフィックとそれを原因としたサービス停止問題に直面しており，これらの異常トラフィックに対する検知・防御が求められている．そのため，インターネット定点観測システム ISDAS [4]，Honeypot/Honeynet/Honeyfarm [3], [7], [5]，Darknet を利用した異常パケット収集 [1], [2], [6] などネットワークへの攻撃を検知・防御するシステムの研究が盛んである．特に Darknet は，経路広報はされているが未使用のアドレスであり，そこに到着するパケットのほとんどは異常トラフィックである．したがって Darknet を観測することでネットワーク上の異常トラフィックのパケットを効率的に収集でき，その特徴を掴むことができる．

そのため，これまで我々は Darknet に到着するパケットのうち TCP パケットで syn フラグのみ 1 のパケット (以下 tcp-syn と呼ぶ) を横軸を時間，縦軸を宛先 IP アドレスとする散布図にプロットし，パケットが到着する様子 (パケット到着パターン) を示し，その特徴を視覚的に把握してきた [11]．ここでは攻撃パターンを使い観測情報から隣接するアドレスへの攻撃を予測するシステム [9] を実現するために，特に特徴を把握しやすいアドレススキャンと見られるパケットに注目し，解析を行った．一方で，散布図でも一見は無秩序な雑音のように見える異常トラフィックをパケットのヘッダ情報を利用して解析し，その中にも特定の規則性や共通的特徴があることも報告してきた [15]．

本稿では，これまで解析対象としていなかった UDP パケットについて，特に [15] で使用した手法を用いて，その特徴を調査した結果を報告する．UDP パケットは特定の宛先 IP アドレスに対してパケットを送信するパケットの割合が多く，また宛先ポート番号もほぼ変化しない特徴がある [14]．そのため，UDP パケッ

Characteristic analysis of UDP packets arrived in Darknet

Masayuki Ohta, Toshiharu Sugawara, 早稲田大学大学院基幹理工学研究科情報理工学専攻, Dept. of Computer Science and Engineering, Waseda University.

Kensuke Fukuda, 国立情報学研究所/科学技術振興機構, National Institute of Informatics / PRESTO JST.

Toshio Hirotsu, 法政大学情報科学部, Faculty of Computer and Information Sciences, Hosei University.

トの特徴を解析することで，特に狙われるポート番号や宛先 IP アドレスを調査できると考えられる．特徴を調査するため，本稿ではまずパケット到着パターンを送信元 IP アドレス，宛先ポート番号を元に分類する．さらに，IP ヘッダの TTL，identification フィールドを用いて到着パケットを解析し，無秩序に見えるいくつかの攻撃パケットの群について，送信元ホストの数やそれぞれのホストがどのようにネットワークに属するかについて推定する．

2 解析手法

先の研究 [15] で，異常トラフィックには，アドレス空間へのパケット到着パターンを確認するだけでは捉えきれない隠れた特徴があることを報告した．本稿では，UDP ヘッダの持つ情報を解析フィールドとして，パケット到着パターンの色分けを行い，全体的な特徴を観測する．また，IP ヘッダの持つ情報を元に，パケットを送信するホスト数の推定や，同一送信元 IP アドレスに対するホストの (非) 同一性の推定を試みる．本論文で使用した解析情報は以下の通りである．

1. ヘッダ情報による分類
   本稿では，UDP ヘッダの宛先ポート番号や，IP ヘッダの送信元 IP アドレスごとにパケット到着パターンを色分けする．
2. 到着パケット数の推移
   到着したパケットの数の推移を調べることで，異常トラフィックが特に多く到着する時間帯を確認する．
3. ヘッダ情報によるパケット到着パターンの解析
   各パケット到着パターンに対して，IP ヘッダの TTL(Time to Live)，IP ヘッダの Identification フィールド (以下 IPID フィールド)，到着パケット数を組み合わせて分類を試みる．

3 解析結果

3.1 送信元 IP アドレス毎のパケット到着パターン

我々は，2006 年 10 月 1 日から 2009 年 4 月 7 日までを解析対象としたが，本稿では特に 2008 年 12 月 20 日，2009 年 1 月 16 日の 2 日分のデータのうち UDP パケットの解析結果を報告する．それぞれのデータのパケット到着パターンを散布図として示し，送信元 IP アドレス毎に色分けした．なお，パケット到着数の多い送信元 IP アドレスから到着するパケットの特徴に注目するため，到着パケット数の多い上位 10 件の送信元 IP アドレス分のみを出力した．また，送信元 IP アドレスは到着数が多い順に，Src A〜Src J として表示している．

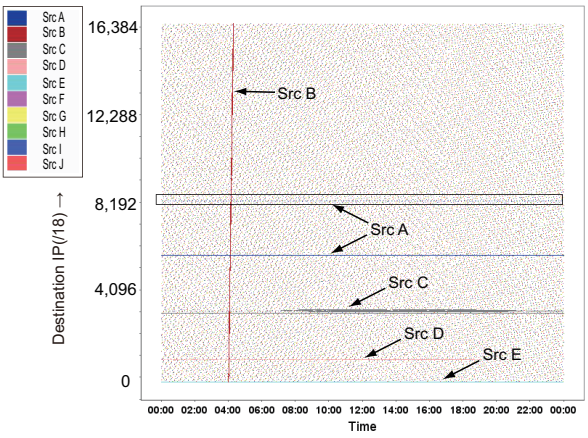

図 1 送信元 IP アドレスによるパケット到着パターンの色分け結果-upd-2008 年 12 月 20 日

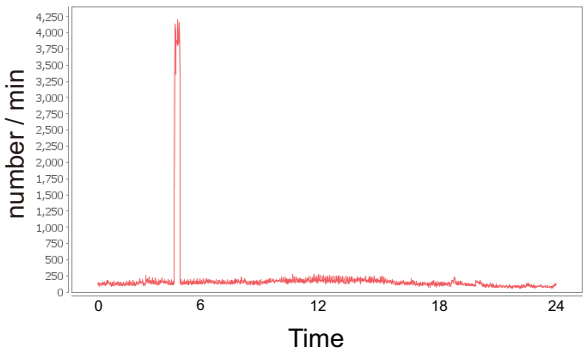

図 2 パケット到着数-upd-2008 年 12 月 20 日

図 1 は，2008 年 12 月 20 日の色分けされたパケット到着パターンである．表 1 に，上位 10 件の送信元 IP アドレスごとのパケット到着数と割合も示す．最も到着パケット数の多い Src A は，1 日中 2 つの特定のアドレスに対してパケットを送信していた．

図 2 は，到着パケット数の遷移を示している．これ

表 1　送信元 IP アドレスごとのパケット到着数と割合 -udp- 2008 年 12 月 20 日

順位	送信元 IP アドレス	パケット数	割合 (%)
1	Src A	100,474	37.0
2	Src B	65,487	24.1
3	Src C	29,809	11.0
4	Src D	16,478	6.1
5	Src E	12,954	4.8
6	Src F	9,556	3.5
7	Src G	9,477	3.5
8	Src H	9,227	3.4
9	Src I	9,191	3.4
10	Src J	9,123	3.4

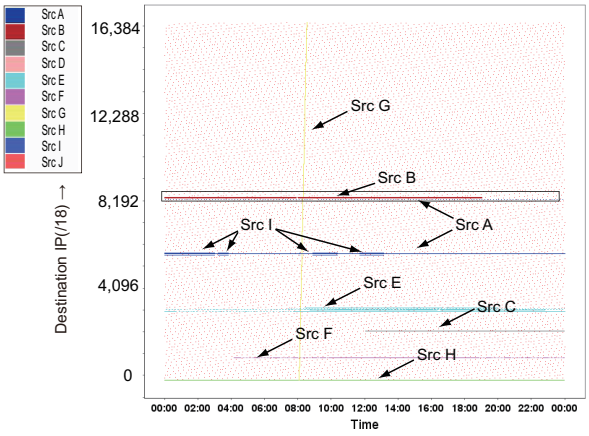

図 3　送信元 IP アドレスによるパケット到着パターンの色分け結果-upd-2009 年 1 月 16 日

を図 1 と比較すると，Src B のアドレススキャンと見られる攻撃が到着した際にパケット到着数が大幅に増えたことが分かる．

図 3 は，2009 年 1 月 16 日の色分けされたパケット到着パターンである．表 2 に，この日の上位 10 件の送信元 IP アドレスごとのパケット到着数と割合を示す．複数の送信元 IP アドレス (Src A, Src B, Src C, Src E, Src F, Src H, SrcI) がそれぞれ特定の 1 つないし複数のアドレスに対してパケットを送信し続けていた．Src A のように 1 日にわたってパケットを送信し続ける送信元 IP アドレスもあれば，Src I のように 1 時間程度パケットを送信し続ける送信元 IP アドレスもあった．また，複数の宛先 IP アドレスに対してパケットを送信する送信元もあった．例えば，

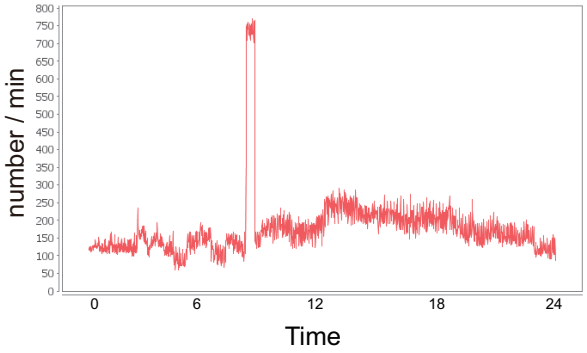

図 4　パケット到着数-upd-2009 年 1 月 16 日

表 2　送信元 IP アドレスごとのパケット到着数と割合 -udp- 2009 年 1 月 16 日

順位	送信元 IP アドレス	パケット数	割合 (%)
1	Src A	106,975	42.3
2	Src B	21,103	8.3
3	Src C	20,598	8.2
4	Src D	18,887	7.5
5	Src E	18,549	7.3
6	Src F	17,673	7.0
7	Src G	16,384	6.5
8	Src H	13,009	5.1
9	Src I	9,859	3.9
10	Src J	9,852	3.9

SrcB は 2 つのアドレスに，SrcE は 7 つの宛先 IP アドレスに対してパケットを送信していた．

この日の到着パケット数の遷移を図 4 に示す．図 3 と図 4 を比較すると，Src G のアドレススキャンと見られる攻撃が到着した際にパケット到着数が大幅に増加したことが分かる．また，特定の宛先ポートを狙った攻撃 (水平に見える直線) の数の増減で，パケット到着数も増減することも分かる．

3.2　宛先ポート番号毎のパケット到着パターン

送信元 IP アドレスと同様に，解析対象としたそれぞれの日について宛先ポート番号でパケット到着パターンを色分けした．なお，パケット到着数の多い宛先ポート番号宛に到着するパケットの特徴に注目するため，上位 10 件のポート番号のパケットのみをグラフに示した．この 10 件に含まれるパケットは，前説で示した上位 10 件の送信元 IP アドレスに含まれる

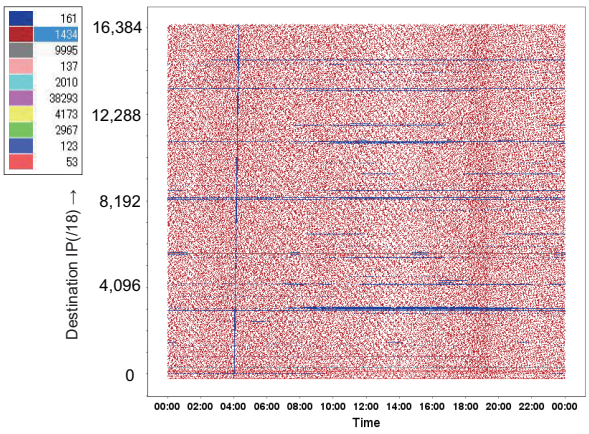

図 5　宛先ポート番号によるパケット到着パターンの色分け結果-upd-2008 年 12 月 20 日

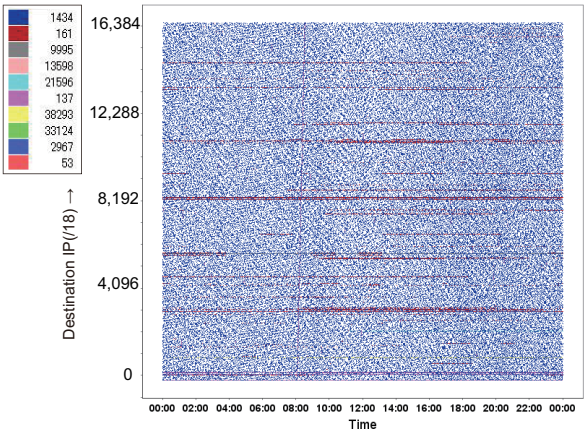

図 6　宛先ポート番号によるパケット到着パターンの色分け結果-upd-2009 年 1 月 16 日

表 3　宛先ポート番号ごとのパケット到着数と割合 -udp-2008 年 12 月 20 日

順位	宛先ポート番号	パケット数	割合 (%)
1	161	164,340	32.0
2	1434	132,495	25.9
3	9995	100,330	19.6
4	137	47,747	9.3
5	2010	30,871	6.0
6	38293	9,935	1.9
7	4173	8,014	1.6
8	2967	7,164	1.4
9	123	6,764	1.3
10	53	4,706	0.9

表 4　宛先ポート番号ごとのパケット到着数と割合 -udp-2009 年 1 月 16 日

順位	宛先ポート番号	パケット数	割合 (%)
1	1434	143,478	22.9
2	161	132,740	21.2
3	9995	106,833	17.0
4	13598	86,555	13.8
5	21596	53,397	8.5
6	137	52,548	8.3
7	38293	18,475	2.9
8	33124	12,680	2.0
9	2967	10,859	1.7
10	53	10,086	1.6

パケットとは異なる．

図 5 に，2008 年 12 月 20 日の色分けされたパケット到着パターンを示した．また，上位 10 件の宛先ポート番号宛に送信されたパケットの数と割合を表 3 にまとめた．まず，昇順の攻撃や特定の宛先 IP アドレスへの攻撃がポート 161 番 (色は青) で到着パケット数全体の 32%を占めていた．次に，1 日にわたってパケット到着パターンの画面を埋め尽くす攻撃がポート 1434 番 (色は赤) で，到着パケット数全体の 26%ほどを占めていた．上位 3 件目のポート番号 9995 番 (色は灰) のパケットは，1 日にわたって特定の宛先 IP アドレスに対して到着していた．

これら上位 3 件のポート番号のパケットだけで全体のパケットの約 77%を占めていた．なお，宛先ポート番号の種類は全体で 540 件あった．このことから，ネットワーク上で狙われるポート番号には偏りがあり，ネットワークの防御のためには，特に狙われるポート番号を把握することが重要であると考えられる．

図 6 に，2009 年 1 月 16 日の色分けされたパケット到着パターンを示す．また，この日の上位 10 件の宛先ポート番号に向けて到着したパケットの数と割合を表 4 にまとめた．上位 4 件の宛先ポート番号宛のパケットだけで全体のパケットの 74%を占めていた．なお，宛先ポート番号の種類は全体で 511 件あった．また，順番は異なるが 2008 年 12 月 20 日と上位 3 件のポート番号は同じだった．今回は，全体の 23%を占めた 1434 番宛 (色は青) のパケットは 2008 年 12 月 20 日と同様に Darknet 全体に 1 日にわたっ

て到着していた．また，全体の 21%を占める 161 番宛のパケット (色は赤) は複数の宛先 IP アドレス宛に継続的に到着しており，2008 年 12 月 20 日と非常に似通ったパケット到着パターンを示した．また全体の 17%を占める 9995 番宛のパケット (色は灰) は，2008 年 12 月 20 日と同様に特定の宛先 IP アドレスに継続的に到着しており，狙われている宛先 IP アドレスも全く同じであった．2008 年 12 月 16 日のデータでは上位 10 位にすら入らなかった 13598 番宛のパケットが今回大幅に増加しており，全体の 13.8%を占めていた．これらのパケットは 18 時から 24 時の間に特定の宛先 IP アドレス向けに到着していた．

ちなみに 1434 番宛のパケットは IPA の TALOT2 [13] におけるアクセス解析においても，UDP パケットのポート番号別到着数で 2 位であった．このことから，観測するアドレス空間が異なっても，ネットワーク上で共通して特に狙われるポート番号が把握できることが分かった．Darknet の解析から特に狙われるポート番号の情報を取得し，ファイヤーウォールの設定などに応用できると考えられる．

4 IP ヘッダ情報を利用したパケット到着パターンの解析

前節でそれぞれの送信元 IP アドレスごとに攻撃の特徴が異なることが分かった．ある送信元 IP アドレスから送信されてくるパケットは，一見すると何の特徴もないが，さらに詳しく解析を行うことで特徴を抽出できると考えられる．そこで，特定の送信元 IP アドレスに注目し解析を行う．解析は TTL, IPID フィールドを用いて行った．本稿では 2008 年 12 月 20 日と 2009 年 1 月 16 日のそれぞれの Src A(到着したパケット数が最も多かった送信元 IP アドレス) から到着したパケットについて述べる．

4.1 2008 年 12 月 20 日の解析結果

図 7 は，2008 年 12 月 20 日に Src A(到着パケット数が最も多かった送信元 IP アドレス) から到着したパケットの IPID をプロットしたものである．図 7 から同時に 2 つの昇順の系列を確認できる．通常 1 つのホストから到着するパケットの IPID は 0〜65535

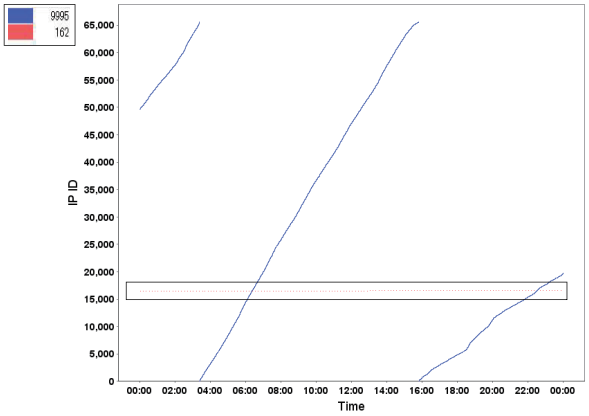

図 7 Src A のみの IPID を宛先ポート番号で色分けした結果-upd-2008 年 12 月 20 日

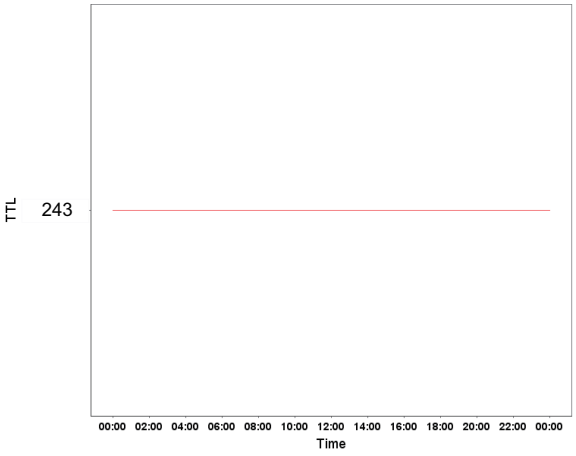

図 8 Src A のみの TTL-upd-2008 年 12 月 20 日

の間で増加していき上限に達したところで 0 に戻る．つまり，1 つのホストからのパケットの到着であれば，IPID フィールドの値は 1 つの昇順の系列として表される．

しかし今回，独立に IPID が上昇する 2 つの系列を確認できた (赤色の系列は確認しにくいため，四角の枠で囲んだ．この部分も右上がりになっている．)．

2 つの系列が確認できる理由として 3 つの可能性が考えられる．第 1 は，一つの送信元 IP アドレスの内部に 2 つのホストがある可能性である．第 2 は，2 つ

の NIC を持つホストが存在し，パケットを送信してきた可能性である．第 3 は，送信者が送信元の IP アドレスを偽装し，偶然同一の送信元 IP アドレスから 2 つの系列のパケットが到着する可能性である．

IPID の値は OS によって自動的に割り振られ，通常は 1 つの OS からパケットが到着する場合には 1 つの系列が観測される．そのため，全く同じ観測期間に IPID の系列が 2 つ存在することを考えると，それぞれ異なるホストからパケットが送信されてきた可能性が高い．従って，上記の 3 つの可能性のうち，第 2 の「NIC を持つホストが存在する」の可能性は低いと考えられる．

攻撃元ホストについてさらに詳しく調査するために，Src A の TTL に関する調査も行った．図 8 にこの日の Src A の TTL を示す．図 8 から，TTL の値は 1 つであることが確認できる．

TTL の値が常に一定であることを考慮すると，異なるグローバルネットワークからパケットが送信されてき可能性は低いと考えられる．なぜなら，異なるネットワークから送信されたパケットはルータを経由する回数が異り，TTL の値も異なる可能性が高いからである．そのため，第 3 の「送信元ホストを偽装し，偶然同一の送信元 IP アドレスから 2 つの系列のパケットが計測された」という可能性も低い．従って，Src A には 1 つのグローバルネットワーク内に 2 つのホストがあると考えられ，また，TTL が常に一定であることを考慮すると，第 1 の「2 つのホストは NAT 内の同一のローカルネットワークに接続されている」の可能性が高くなる．

4.2 2009 年 1 月 16 日の解析結果

2009 年 1 月 16 日のデータに対しても同様の調査を行った．図 9 はこの日の IPID を示している．この日も 2008 年 12 月 20 日の全く同じ送信元 IP アドレスが Src A であった．図 9 から，この日も Src A の IPID に複数の系列を確認した．また，Src A の IPID の宛先ポート番号は 2008 年 12 月 20 日も 2009 年 1 月 16 日も共通して 9995 番と 162 番であった．割合に関しても 9995 番が Src A の到着パケット数全体のうちの 99.8%を占める点で共通していた．この日の

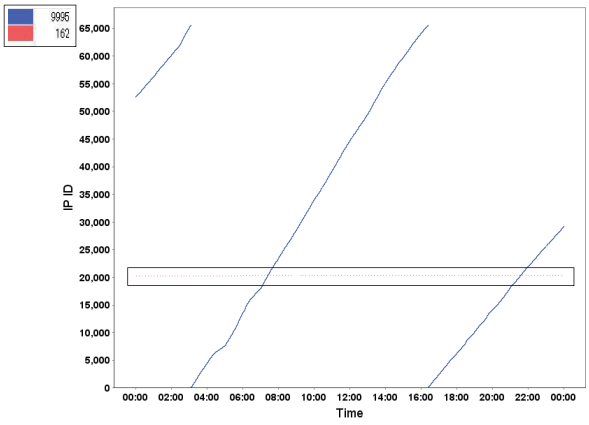

図 9 Src A のみの IPID を宛先ポート番号で色分けした結果-upd-2009 年 1 月 16 日

TTL の値についても確認を行ったところ，TTL の値に関しても 2008 年 12 月 20 日と同様であった．

IPID の系列の傾きは異なるが，系列が 2 つ存在する点や，TTL の値が常に一定であることを考慮すると，2008 年 12 月 20 日と同様，Src A には NAT 内の同一のローカルネットワークに 2 つのホストが存在し，その 2 つのホストが Darknet にパケットを送信したと考えられる．

解析結果から，全く同一の送信元 IP アドレスが同じ傾向でパケットを送信したことが分かり，異常トラフィックを大量に送信してくる特定の送信元 IP アドレスの存在を把握できた．また，異常トラフィックを送信してくる際にも送信者ごとの特徴があることが分かった．異常トラフィックを大量に送信するホストの危険性を考慮すると，ファイヤーウォールや IDS 内での特定の送信元 IP アドレスに対する設定を，送信元毎の特徴に合わせて行う必要があると考えられる．

5 結論

本稿では，Darknet に到着する異常トラフィックの内 UDP パケットを解析し，その特徴解析を行った．まず，パケット到着パターンを送信元 IP アドレス，宛先ポート番号で色分けし，到着するパケットの特徴を調査した．さらに，一つの送信元 IP アドレスからの到着パケットに着目し，詳しい特徴を抽出するため

に TTL，IP ヘッダの ID フィールドを用いて解析を行った．

本稿の調査結果から，パケット到着パターンを送信元 IP アドレスで色分けすることで，各送信元 IP アドレス毎からの到着パケットの特徴を把握できた．また，パケット到着パターンを宛先ポート番号で色分けすることで，特に狙われるポート番号を把握できた．宛先ポート番号ごとの色分けは，その時期に流行するウイルスや，特に狙われるポート番号自体の把握に役立つと考えられる．

本稿の調査結果は，1 つの送信元 IP アドレスからの到着パケットに着目し，TTL や IP ヘッダの ID フィールドを調査すると，パケットを送信するホスト数やそのホストが接続されたネットワークの構成を推測できることが分かった．加えて，異常トラフィックを大量に送信してくる特定の送信元 IP アドレスが存在すること，またその送信方法にも特徴があることが分かった．本稿での解析結果を通して，異常トラフィックの特徴の一端を観測できた．

今後は IP ヘッダの ID フィールドに対して統計学的な手法を用いて解析を行うことで，異常トラフィックを判別する方法を調査する予定である．

参 考 文 献

[1] Michael Bailey, Evan Cooke, Farnam Jahanian, Andrew Myrick, and Sushant Sinha, “Practical darknet measurement, ”Proceedings of Conference on Information Sciences and Systems 2006 (CISS ’06), 2006.

[2] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, and David Watson, “The Internet Motion Sensor: A distributed blackhole monitoring system, ”Proceedings of Network and Distributed System Security Symposium Conference 2005 (NDSS ’05), 2005.

[3] Xuxian Jiang, Dongyan Xu, and Yi-Min Wang, “Collapsar: a vm-based honeyfarm and reverse honeyfarm architecture for network attack capture and detention, ” J. Parallel Distrib. Comput., Vol. 66, No. 9, pp. 1165-1180, 2006.

[4] JPCERT/CC. “インターネット定点観測システム.”http://www.jpcert.or.jp/isdas/.

[5] The Honeynet Project. http://www.honeynet.org/.

[6] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson, “Characteristics of internet background radiation, ” IMC ’04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pp.27-40, New York, NY, USA, 2004.ACM.

[7] Fabien Pouget, Marc Dacier, and Herv´ Debar, “Attack processes found on the Internet, ” NATO Research and technology symposium IST-041 “Adaptive Defence in Unclassified Networks, ” 19 April 2004, Toulouse, France, Apr 2004.

[8] Xuxian Jiang, Dongyan Xu, and Yi-Min Wang, “Collapsar: a vm-based honeyfarm and reverse honeyfarm architecture for network attack capture and detention, ” J. Parallel Dis- trib.p=4 Correliation Coefficient Comput., Vol. 66, No. 9, pp. 1165-1180, 2006.

[9] 廣津登志夫, 福田健介, 栗原聡, 明石修, 菅原俊治, “断片アドレスを用いた分散協調インターネット監視に関する一考察, ”情報処理学会 OS 研究会研究報告 (83), pp. 39-45, 2007.

[10] Evan Cooke, Michael Bailey, Z.Morley Mao, Danny Mcpherson, David Watson and Farnam Jahanian, “Toward understanding distributed blackhole placement, ” Proceedings of the 2004 ACM Workshop on Rapid Malcode (WORM-04), pp.54-64, 2004. ACM Press.

[11] 杉本 周, 福田 健介, 廣津 登志夫, 明石 修, 菅原 俊治, “特定のアドレス空間を基準とした遅延相関解析によるインターネット上の攻撃予測の可能性, ”日本ソフトウェア科学会インターネットテクノロジ研究会, 2009.

[12] David Moore, Colleen Shannon, Geoffrey M. Voelker and Stefan Savage. “Network telescopes: Observing small or distant security events. ”Invited Pre- sentation at the 11th USENIX Security Symposium, 2002.

[13] TALOT2(IPA), http://www.ipa.go.jp/security/ciadr/txt/list.html

[14] Kensuke Fukuda and Romain Fontugne, “Estimating Speed of Scanning Activities with a Hough Transform, ”IEEE International conference on communications, CapeTown, South Africa, 2010.

[15] 大田 昌幸, 薄田 広志, 福田 健介, 廣津 登志夫, 菅原俊治, “Darknet に到着する背景雑音異常トラフィックの特徴解析, ”日本ソフトウェア科学会インターネットテクノロジ研究会, 2010.