Computer Security Symposium 2018
22 - 25 October 2018

宛先変化数を用いた協調型走査活動を行なうホスト群抽出手法の提案

梶川 慶太[1]　中村 康弘[2]　芦野 佑樹[3]　鮫島 礼佳[3]　須尭 一志[3]　矢野 由紀子[3]

概要：近年，ボットネット等が行っている走査活動には分散型と呼ばれるものがあり，送信元と宛先をそれぞれ分散化させ，従来手法では検出することを困難にしている．本研究では，同一プログラムによる走査活動は挙動が類似するという推定のもと，一定期間内に宛先を変更した回数を宛先変化数と定義し，宛先の変化のパターンが類似する協調型走査活動を検出する．また，実際に連続したアドレスを持つ観測環境において本手法を適用し，検出された結果について示す．

キーワード：ダークネット, ネットワークスキャン, セキュリティ, 宛先変化

An extraction method for collaborative scanning host groups using number of destination changes

Keita Kajikawa[1]　Yasuhiro Nakamura[2]　Yuki Ashino[3]　Reika Samejima[3]　K Sugyo[3]
Yukiko Yaho[3]

Abstract: In recent years, there is a scanning activity what is called a distributed type. It has distributed sources and destinations respectively, making it difficult to detect by conventional methods. In this study, we estimated that the scanning activity by the same program has similar behavior. Then, the number of changing the destination within a certain period of time is defined as the destination change number, and I will detect collaborative scanning activity in which the changed pattern to the destination is similar. Also, we apply this method in an observation environment with actual continuous addresses and show the detected results.

Keywords: Darknet, Network Scan, Security, Destination Change

1. はじめに

近年，インターネットを介したサイバー攻撃が活発になっており，それらサイバー攻撃を検知し，対策を行うには，ネットワークを観測し，その脅威を把握することがますます重要になってきている [1,2].

脅威を把握する手段の一つとして攻撃者が行う走査活動（スキャン）を分析し，走査の発信元，走査の規模，目的や種類を分析する走査活動分析がある [3,4].

例えば，Mirai ボットネットの規模を把握しようとする場合には，受信したパケットの TCP の初期シーケンス番号と宛先 IP アドレスが同じになっているものを Mirai ボットネットとして扱うことができる [5]．Hajime ボットネットの場合には，TCP の初期シーケンス番号の上 16bit，もしくは下 16bit の値が 0 であるものを Hajime マルウェアファミリーによる走査活動として扱う事ができる [6].

しかしながら実際の通信には，そういった特徴ある通信は少数である．そこで，まず走査活動を検知する手段とし

[1] 防衛大学校理工学研究科
Graduate School of Science and Engineering, National Defense Academy
[2] 防衛大学校情報工学科
Computer Science, National Defense Academy
[3] 日本電気株式会社ナショナルセキュリティ・ソリューション事業部サイバーセキュリティ・ファクトリ
Cyber Security Factory, National Security Solution Division, NEC Corporation

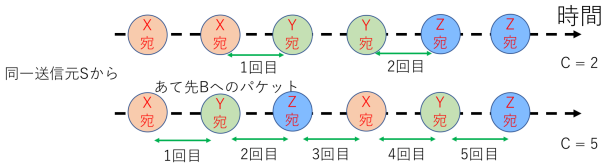

図 1　宛先変化数 C

Fig. 1　C: Number of destination changes

て，パケットの送信頻度に対し閾値を用いた検知があり，パケットを連続して対象のホストに送信すると，パケットの時間当たりの頻度が極端に高くなることから，走査活動であると検知する手法になる．

しかし，近年ではパケット間の送信間隔を非常に長期間にし，頻度によるスキャン検知を回避するスロースキャンが行われている．

また，複数のスキャン先を複数の送信元に分担し，それぞれ独立して走査を行う分散型走査活動（分散型スキャン）も行われており，その検知は困難である．

走査活動を実際に観測すると，数ヶ月から年間に渡って行われているものもあり，その実態を調査するために，そういった活動を行うホスト群を抽出することが必要である．

そこで，本研究では 1 ～ 2 ヶ月程度活動しているホスト群を対象とし，長期間に渡って行われる分散型走査活動を協調型走査と定義し，ダークネットアドレスへ到達するパケットに対し，協調型走査を行うホスト群を抽出する手法について提案する．

2. 関連研究

土性ら [7] は，協調的な動作をするホスト群を人の目で把握できるように，そういったホスト群を散布図に示す手法，及びそれらを抽出する手法を提案している．新規 IP アドレス群の出現時期と終了時期を用いて，協調的なホスト群をボットとみなしているが，同じボットネットによる通信であっても，異なる時期に現れるホスト群は別グループとして区分されてしまう可能性がある．

福島ら [8] は，長時間にわたって少量のパケットしか投げられないような，気づかれにくい攻撃の検知と，その検知された攻撃パターンからの特徴抽出を提案している．

笹生ら [9] は，ダークネットトラフィックから，ホスト毎の通信パターン及び OS フィンガープリントを利用することで，ホストの分類を行い，実際に 2 年間収集したデータに適用し，セキュリティ対策に有用な新規情報を抽出することが出来ている．

3. 提案手法

3.1 方針

本研究では，一定期間内に宛先を変更した回数を宛先変化数 C と定義する．**図 1** は，宛先変化数 C の例である．

表 1　クラスタリングで使用する特徴量

Table 1　Attribute.

パラメータ名	意味
宛先種類数 C	ユニークな宛先ホストの数
宛先変化数 U	観測期間中に宛先が変化した回数
宛先変化率 R	宛先変化数 C/宛先種類数 U

一つの丸はある宛先に対するパケットを示しており，6 個のパケットが時間の経過とともに，どの宛先へ送信されているかを表している．この例では，6 個のパケットを 2 種類のパターンで送信している．上段では連続したパケット間で宛先が変化した数は緑矢印で示す 2 回であり，下段は 5 回である．上段と下段ではパケットの宛先のパターンが異なっているが，各 X，Y，Z 宛のパケットはそれぞれ 2 個ずつと等しい．この場合，通常の統計的な指標では区別することが出来ないが，宛先変化数 C 使用することで，区別することができる．

また，同一プログラムによる走査活動は挙動が類似するという推定を行い，宛先変化数 C を用いて，宛先の変化のパターンが類似するものは，協調して走査活動を行っている協調型走査活動として抽出する手法を提案する．

3.2 観測環境

宛先の変化を特徴量として取得するには複数アドレスを同時に観測できる環境が必要である．また，本提案手法はパケットの前後間の宛先変化を特徴量として必要とするため，連続したグローバルアドレスが観測できる環境において適用できると考える．

3.3 パケットのフィルタリング

まず，本研究では対象とするホストアドレスを，1 ～ 2 ヶ月程度の活動を行なうホスト群とするため，観測期間中，その送信元が最初に現れた時刻と，最後に現れた時刻との差を活動期間と定義し，活動期間が 2592000 秒 (30 日)～5184000 秒 (60 日) であるホストのみを抽出するフィルタリングを行う．

3.4 宛先変化

観測期間全体において，同一送信元から送信されたパケットの前後間の宛先が異なっていた回数である宛先変化数 C を計算する．また，各送信元からユニークな宛先ホストの数を示す宛先種類数 U，及び宛先変化数 C から宛先種類数 U を割った値，宛先変化率 R を計算し，これら**表 1** に示す特徴量を，宛先変化を表す特徴量として使用する．

3.5 協調型走査活動の抽出

最後に，表 1 の特徴量を用いて，クラスタリングを行なうことで，類似した挙動を行なう協調型走査活動を抽出

表 2　観測環境
Table 2　Observated environment.

アドレス数	約 1500 ダークネットアドレス
観測期間	2014 年 1 月 1 日〜2014 年 12 月 31 日
パケットサイズ	377GB
パケット数	44，604，863 パケット
観測送信元アドレス	10，821，497 アドレス

表 3　計算結果の例
Table 3　Result example.

送信元アドレス	宛先変化数 C	宛先種類数 U	宛先変化率 R
1.0.100.*	7	8	0.875000
1.93.29.*	77	75	1.026667
1.228.145.*	64	62	1.032258
81.26.178.*	1763	2	881.5
⋮	⋮	⋮	⋮

する．

4. 実験結果

4.1 観測環境

提案手法を**表 2** の観測環境において得たパケットに対し適用した．

4.2 フィルタリング結果

フィルタリングした結果，165761 アドレスが得られた．また，得られた送信元アドレスの活動期間と宛先種類数 U の散布図を**図 2** に示した．横軸は活動期間であり，2592000 秒 (30 日)〜5184000 秒 (60 日) の範囲である．縦軸は宛先種類数 U を示している．観測範囲全体に対し送信する A 群，/22 の範囲に対し送信する B，C 群，/24 の範囲に対し送信する D 群が存在することが図 2 から読み取れる．

4.3 宛先変化

観測環境において得られた各ホストの宛先変化数 C，宛先種類数 U，宛先変化率 R を計算した例を**表 3** に示す．宛先変化の挙動が類似する 1.93.29.*と 1.228.145.*のそれぞれの宛先変化数 C と宛先種類数 U は異なる値ではあるが，宛先変化率 R では近い値となっている．

4.4 クラスタリング

今回クラスタリングには，K-means 法を用いた．K-means 法は非階層クラスタリング手法の 1 つであり，クラスタ数を設定することでクラスタの中心とクラスタ対象との距離を求め，分類する手法である．距離計算には，宛先種類数 U，宛先変化数 C，宛先変化率 R を使用し，各特徴量は標準化して行った．

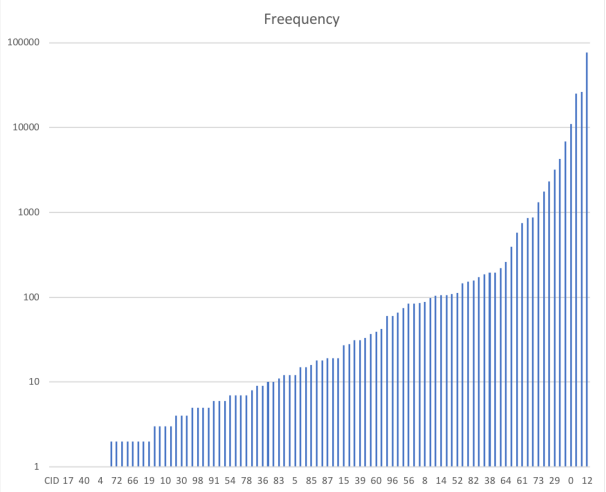

図 3　クラスタリング結果
Fig. 3　Clustering result.

また，クラスタ数は 100 を設定し，その結果について**図 3** に示す．横軸はクラスタ番号，縦軸はそのクラスタ番号の度数を示している．クラスタの約 7 割は 100 アドレス以下で構成されていることが分かる．

5. 検証と考察

クラスタリングされたアドレスがそれぞれどのような特徴になっているか確認し，提案手法によって類似した挙動を行なうホスト群が抽出できているか検証を行った．

クラスタ番号 16 は 11 アドレスで構成されており，約 2 秒の間に約 1200 アドレスに対し走査を行っており，送信元が最初に SYN を送信し，再送要求を送信してくるまでの間隔が約 12 時間前後というホスト群であった．各アドレスの詳細については，**表 4** に示す．平均間隔時間をクラスタリングする際に用いていないが，非常に近い値となっていることが分かる．

また，クラスタ番号 32 は 6 アドレスで構成されており，ポート 5900 に対し，ランダムに宛先アドレスへ走査を行っているホスト群である．各アドレスの詳細については，**表 5** に示す．また，すべてのホストがスリーウェイハンドシェイクによるセッションが確立後即座に FIN+ACK を持ってセッションを切断する挙動を行っていた．

6. まとめと今後の課題

本研究では，1 〜 2 ヶ月程度活動しているホスト群を対象とし，ダークネットアドレスへ到達するパケットに対し，協調型走査を行うホスト群を抽出する手法について提案，検証を行った．結果として，同じような挙動を行なうホスト群を抽出し，その特徴について分析することができた．

今後は，クラスタリング手法を別の手法で行なうとともに，クラスタリングされたホストが本当に協調して行っているのかを明らかにする．また，走査活動の活動パターン

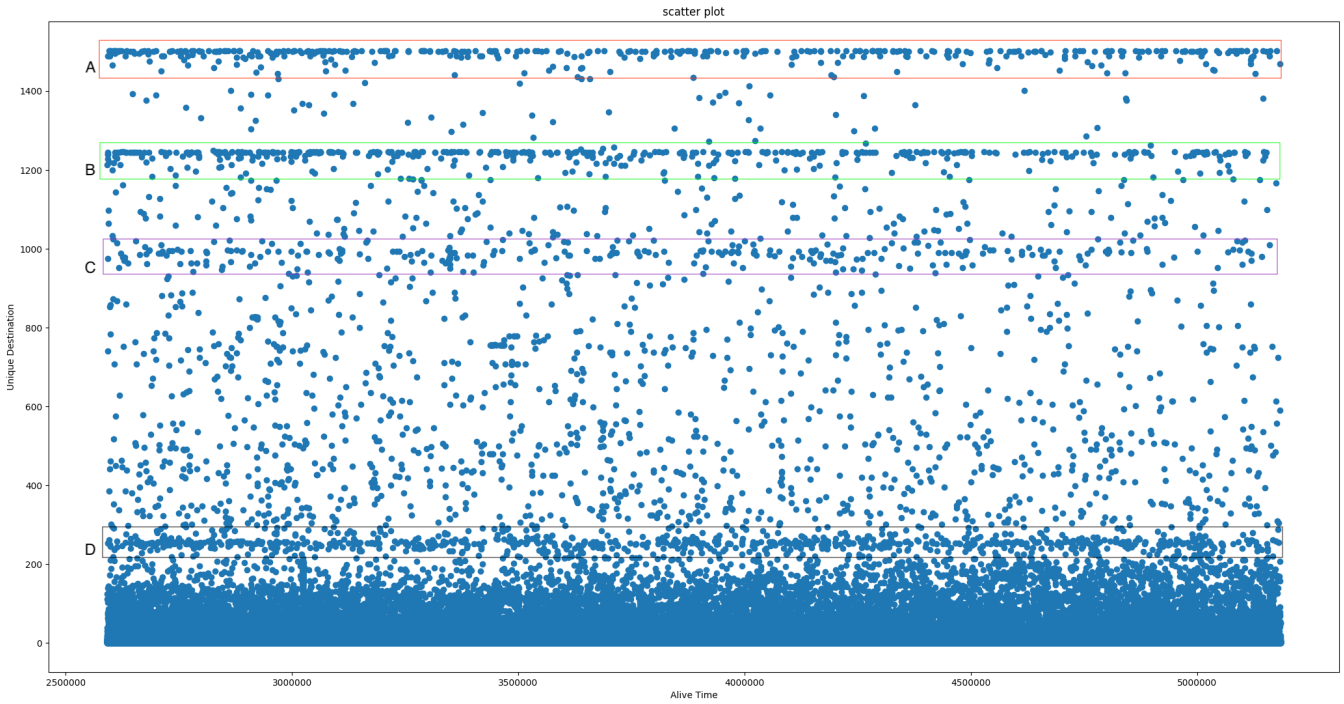

図 2　活動期間と宛先種類数 U の散布図

Fig. 2　Scatterplot of Alive Time and Number of Unique Destination

表 4　クラスタ番号 16 に含まれるアドレスの詳細

Table 4　No.16 Cluster Details.

送信元アドレス	パケット数	平均間隔時間	宛先変化数 C	宛先種類数 U	宛先変化率 R	生存期間
59.53.67.*	125962	38.0	113765	1246	91.304173	4861181
61.147.103.*	103566	29.0	100561	1245	80.771888	3004348
61.160.215.*	108164	28.0	101186	1245	81.273896	3040075
122.226.160.*	126572	38.0	97236	1245	78.101205	4890841
180.225.197.*	131281	39.0	111965	1246	89.859551	5147444
200.70.40.*	165914	30.0	110172	1240	88.848387	5015748
211.143.243.*	129901	28.0	87643	1232	71.138799	3718248
218.2.22.*	122493	21.0	114591	1245	92.040964	2614835
218.17.156.*	90278	42.0	88535	1231	71.921202	3865483
222.186.52.*	103021	25.0	89847	1219	73.705496	2605946
222.186.62.*	123034	24.0	111658	1245	89.685141	3010819

をより正確に定義できるような特徴量についても検討する．

参考文献

[1] https://www.fireeye.jp/current-threats/stopping-todays-cyber-attacks.html, (2018/8/6).

[2] 最近のサイバー攻撃の実情, https://www.jst.go.jp/pr/img/sjsympo2011/presentation_nawa.pdf, (2018/8/6).

[3] インターネット定点観測の結果と攻撃の技術的手法, https://www.npa.go.jp/cyberpolice/material/pdf/20130709_teiten.pdf, (2018/8/6).

[4] NICTER 観測レポート 2017, https://www.nict.go.jp/cyber/report/NICTER_report_2017.pdf, (2018/8/6)

[5] Internet Infrastructure Review (IIR) Vol.33, https://www.iij.ad.jp/dev/report/iir/033/01_04.html, (2016/12/15).

[6] Rapidity Networks, Hajime: Analysis of a decentralized internet worm for IoT devices, https://security.rapiditynetworks.com/publications/2016-10-16/hajime.pdf, (2017/9/1).

[7] ダークネット観測データを用いたボットネット抽出手法の提案，土性文哉，杉生貴成，笠間貴弘，佐々木良一, DICOMO2016,pp.826-831.

[8] ダークネット観測データに基づく攻撃挙動の特徴抽出に関する考察，福島祥郎，堀良彰，櫻井 幸一, ICSS2009(60).

[9] 通信源ホストの分類を利用したダークネット通信解析, 笹生憲，森達哉，後藤 滋樹, CSS2013,pp.729-736.

表 5 クラスタ番号 16 に含まれるアドレスの詳細

Table 5 No.32 Cluster Details.

送信元アドレス	パケット数	平均間隔時間	宛先変化数 C	宛先種類数 U	宛先変化率 R	生存期間
80.152.143.*	263263	12.0	128872	118	1092.135593	3251534
87.228.206.*	382085	7.0	84867	80	1060.837500	3012409
174.61.113.*	355228	8.0	136108	116	1173.344828	3009825
180.43.49.*	308812	10.0	124079	100	1240.790000	3251097
217.128.182.*	190282	15.0	103423	97	1066.216495	2910976
220.132.230.*	561052	5.0	114105	90	1267.833333	3010952