Vol.2017-CSEC-78 No.7
Vol.2017-SPT-24 No.7
2017/7/14

MDSを用いた結託型走査グループの活動状況の分析手法

梶川 慶太[1]　中村 康弘[1,a)]

概要：不特定多数のネットワークアドレスやポートに対する走査活動は，サイバー攻撃の準備段階として行われる可能性があるため，日頃からこれを観測・分析しておく必要がある．とくに走査活動検知を回避するための低速走査や分散型走査は，走査パケットの頻度に基づく検知を困難にする．そこでこの研究では，不特定多数のアドレスやポートに向けた低速走査や分散型走査の検知および傾向の分析を目的として，TCP のコネクション要求に擬似応答を返すことにより初期ペイロードを取得し，そのハッシュ値の同一性に基づいて低速かつ分散化された一連の走査活動を行うアドレスグループを検知する手法を提案する．さらに，これまでは 1 日単位の検知であったが，1 年分の検知結果を元に MDS を用いて複数のアドレスグループの類似性を検証した．異なるペイロードを送付するアドレスグループであっても，挙動が類似するアドレスグループは何らかの関連性を持つものと推定できる．実データに対して提案手法を適用し，初期ペイロードが異なる複数のアドレスグループ間の挙動の類似性について検証した結果を示す．

キーワード：ダークネット観測，ハニーポット，ネットワーク走査活動，多次元尺度法 (MDS)

An analysis method of collaborative scanning group activity using MDS

Keita Kajikawa[1]　Yasuhiro Nakamura[1,a)]

Abstract: Scanning activities for unspecified large numbers of network addresses and ports may be happen as preparation stages of cyber attacks, it is necessary to observe and analyze this on a daily basis. In particular, low-speed scanning and distributed scanning to avoid detection of scanning activity makes detection based on the frequency of scanning packets difficult. Therefore, in this research, in order to detect and analyze trends for low-speed scanning and distributed scanning for unspecified large numbers of addresses and ports, we obtains the initial payload by returning the pseudo response to the TCP connection request, and we propose a method to detect address groups which perform a series of low-speed scanning and distributed scanning based on the identity of the hash value. Furthermore, the research so far is detection on a per-day basis, we verified similarity of multiple address groups using MDS based on detection results for one year. Even in the address group that sends different payloads, we can be presumed that address groups with similar behavior have some relevance. We show the results of verifying the similarity of behaviors among multiple address groups with different initial payload by applying the proposed method to real data.

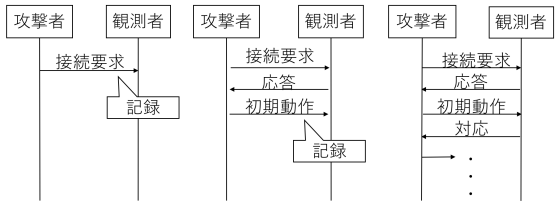
Keywords: Darknet Monitoring, Honeypot, Network Scanning , Multi-Dimensional Scaling (MDS)

1. はじめに

一般に，サイバー攻撃の準備段階には，攻撃対象のネットワークに対して走査活動が行われる [1]．攻撃者は走査活動を行うことで，機器が接続されている IP アドレス，各ポートの開閉状況，各ポートで待ち受けているサービスのバージョンや脆弱性等を調べることができる．

特定の IP アドレスから多数のポートスキャンが行われた場合，そのアクセス回数の頻度の変化を観測することにより，走査活動を検知することが可能である．しかしながら，近年では検知を回避するためにスロースキャンや分散型スキャンが行われる場合がある．スロースキャンは単位時間あたりのアクセス数の増大を低減させるため，個々の

[1] 防衛大学校 理工学研究科 サイバーセキュリティ工学
Yokosuka, Kanagawa 239–8686, Japan
[a)] yas@nda.ac.jp



(1) パッシブ観測 (2) セミアクティブ観測 (3) アクティブ観測

図 1 観測手法

走査時間間隔を意図的に長くする．分散型スキャンは送信元アドレス，宛先アドレス，宛先ポート番号を分散させることにより，アドレスあたりの頻度やポート番号あたりの頻度を低減させる．いずれも，走査活動が行われていることを検知するのは難しい．

本研究では，未使用アドレスへの接続要求に対して疑似応答を返すことで初期ペイロードを取得し，同一時期に同一のペイロードを送付してくる複数の送信元アドレスのグループを分散化された結託型走査グループと想定して検知する．これまでの研究で，1 日単位のグループの抽出とそのアドレス範囲や活動状況の類似性を確認してきたが，複数日あるいは長期にわたる活動状況の分析は不十分であった [2]．そこで，個々のアドレスグループの活動状況に対し，多変量解析手法の一種である多次元尺度法 (MDS: Multi-Dimensional Scaling) を適用し，各グループ間の活動状況の類似性を定量的に評価する手法について検討した．グループ間の類似性が適切に得られれば，複数日に分割されたグループ活動や長期にわたって休止していたグループの再活動などを検知できる可能性が高くなるものと期待できる．

以下，第 2 章では走査活動とその観測，検知に関する既存研究について述べ，第 3 章で擬似応答を返すことで得られる初期ペイロードを用いた走査グループの検出法を提案し，さらに得られた個々のグループに MDS を適用し，各グループの活動状況の定量的な類似性を調査する手法を提案する．第 4 章では，実観測データに対し提案手法を適用し，個々のアドレスグループの活動状況の類似性および異なるグループ間の類似性の評価結果について述べる．

2. 走査活動に対する観測手法

ネットワーク走査活動の検知，観測手法は，大きく 3 種類に分けられる (図 1)．

2.1 パッシブ観測

NICTER が行うダークネット観測では，ダークネットへ着信するパケットに対し応答を返さないパッシブな観測であり，ネットワークインシデントを早期に発見するとともに，大局的な攻撃傾向の把握が可能となっている [3]．また，長期に渡って少量のパケットしか送信されないスロースキャンを検知するとともに，同じような特徴をもった攻撃を抽出する研究がある [4]．

しかし，ダークネット観測においては，パケットを受信するのみであり，分析に使用できる情報は少ない．参考文献 [5] では，ダークネット送信元ごとの通信パターンや OS フィンガープリントを利用して，送信元を分類することで，より多くの情報を得ることが出来ている．

2.2 セミアクティブ観測

着信するパケットに対し，何らかの擬似応答を返し，初期ペイロード等を観測，分析することで，パッシブ観測に比べて情報を多く得ることが出来る．またダークネット観測においては，80 番ポートに対しパケットが増加した場合，HTTP リクエストが増加していると判断するしかないが，これまでの研究 [2] で, 必ずしも HTTP リクエストだけではなく，プロキシ探索を行っているホストが発見されており, パッシブ観測だけでは得られない情報が得られる観測手法である．

本研究では，分散化された走査活動を行う攻撃者が使用するプログラムは, 同一または類似したものであると考え，送信される初期ペイロードが同一になるという予測に基づき，その特徴を利用してグルーピングを行う．

2.3 アクティブ観測

着信するパケットに対し，Honeyd 等 [6] のハニーポットを用いてサービスをエミュレートすることで，接続元とアクティブに通信を行い，一連の攻撃活動を観測する．しかしながら，ハニーポットは，意図的に脆弱性を設定し，マルウェアや不正アクセスを行わせるよう設置するので，攻撃を実際に受けた際の対策や，運用面からも大きなコストが掛かり，大規模な観測には不向きである．

3. 提案手法

3.1 走査活動観測手法

セミアクティブ観測を行うために，以下の設定をルータに行いパケットを取得する．

(1) 使用していない IP アドレス帯に対し，接続要求である SYN パケットが着信した場合 SYN+ACK 応答を返す．
(2) 設定ミスによるパケットの接続確立の防止，初期ペイロード以外の観測の防止のため，SYN+ACK 応答後は RST を送信する．
(3) SYN+ACK パケット応答後に接続元から PSH+ACK が着信した場合，パケットを取得する．
(4) 取得したパケットから着信日時，初期ペイロードのハッシュ値, 送信元 IP アドレス, 宛先 IP アドレス, 送信元ポート番号，宛先ポート番号を抽出しデータベースに格納する．

3.2 送信元アドレスのグルーピング

(1) 時間間隔 T ごとに，ペイロードのハッシュ値が同じであるパケットをグループとし，そのパケット数 N 及び宛先ポートの種類数 P_N を求める．
(2) (1) で求めた N に対し，閾値 N_T より少ないものは，誤送信と攻撃を区別するため除外する．
(3) (1) で求めた P_N に対し，閾値 P_T より多いものは，他手法で検知可能なので除外する．
(4) 更に送信元アドレスが 1 個であった場合，分散化されていないので除外する．

3.3 パラメータの設定

観測期間中におけるグループの遷移及び活動状況を分析するため，パケットがどのように到達，分散化されているかを表す，次のパラメータを設定する．

(1) 同一のハッシュ値 H かつ同一の送信元 S であるパケットの着信時刻 T_{HS}.
(2) 着信時刻 T_{HS} を基にした，各パケットの到達時間間隔 D_{HS}，総数 N_H，最大時間 T_{Hmax}，最小時間 T_{Hmin}，平均時間 T_{Havg}，時間の標準偏差 T_{Hsd}，時間の最頻値 T_{Hmo}.
(3) 同一のハッシュ値が最初に現れた時刻 T_{H1}，最後に現れた時刻 T_{H2}，最も現れた時刻 T_{H3}.

3.4 MDS を用いたハッシュ間の類似性の推定

MDS は分類対象間の関係を多次元空間において表現し，似ているものは近く，異なるものは遠くに配置する手法である．MDS には計量的多次元尺度法と非計量的多次元尺度法があるが，本研究では使用するデータが順序尺度の水準以上であれば良い非計量的多次元尺度法を使用する．

ハッシュ値 H に対応する各パラメータに対し，MDS を用いて計算することで，各ハッシュ値を 2 次元座標上に配置し，類似性のあるものは近くに配置されることから，異なるペイロードをもつグループであっても，近くに配置された場合にはその挙動に類似性が見られる．

4. 観測データに対する適用結果

4.1 実環境における観測データ

2014 年 1 月 1 日から 2014 年 12 月 31 日の間，約 1500 個の未使用 IP アドレスに対し，3.1 の方法で観測を行った際のパケットキャプチャデータを用いる．

4.2 送信元アドレスのグルーピング結果

観測したパケットキャプチャデータから，3.2 に基づいて T=1(日), 閾値 $N_T = 5$, $P_T = 9$ によりグルーピングを行った結果，2,256 個の送信元アドレスのグループが得られた．得られたグループの一部を表 1 に示す．

4.3 パラメータの設定結果

ユニークなハッシュ値は 331 個が得られており，それぞれのハッシュ値ごとに 3.3 に基づいて尺度値を求めた．結果の一例を表 2 に示す．

3 種類のハッシュ値のそれぞれについて，送信元アドレスグループの年間の活動状況を図 2 に示す．横軸は，送信元 IP アドレスを数値に変換したものであり，縦軸はパケットが到達した月日である．図中のプロット点は当該アドレスからペイロードが送付された月日を表す．縦線は同一の送信元アドレスが複数日にわたって同一のペイロードを送付していたとき，それらの日を結んだものである．

図 2(a) では，同一のペイロードが年間を通して多数のアドレスから送られているだけではなく，同一送信元アドレスが多数の日に出現していることがわかる．図 2(b) では，年始めに多数のパケットを送信をした後，10 月頃に同一の送信元アドレスが再び現れたことがわかる．図 2(c) では，ほとんどの送信元アドレスが多数回現れており，(b) と同様に複数の送信元がほぼ同一の時期に活動していることがわかる．

このように，初期ペイロードのハッシュ値を使用することで，結託型走査グループが存在することが分かるとともに，グループの活動状況を得ることができる．

4.4 MDS を用いた異なるハッシュ値間の挙動の類似性

4.3 で設定したパラメータを用いて，MDS により，ハッシュ値が異なる送信元グループを 2 次元座標にプロットした結果を図 3 に示す．

図 3 において，それぞれ近接した 3 つの領域 A, B, C に着目する．それぞれの領域に属する代表的な 3 つの送信元グループの年間活動状況を図 4，5，6 に示す．図 4 では，領域 A に属する 3 つの送信元グループを比較すると，それぞれ 11 月から 12 月末の期間にのみ活動しており，またその出現タイミングが類似していることがわかる．図 5 では，領域 B に属する 3 つの送信元グループは，すべて 2 月前後にのみ活動しており，その時期が類似していることがわかる．同様に，図 6 の領域 C では，2 月頃に初期の活動があり，間をおいて 10 月から 11 月にかけて一斉に活動していることがわかる．さらに，それぞれの領域に属する各送信元グループは，IP アドレス範囲が近接しており，年間の出現パターンも極めて類似している．

5. まとめ

本研究では，未使用アドレスへの接続要求に擬似応答を返すことで得られた初期ペイロードのハッシュ値を利用して，同一時期に同一のペイロードを送信するアドレスグループを結託型走査活動グループとして検知し，年間の活動状況をプロットすることによりその活動状況の類似性を検証した．また，各送信元アドレスグループの特徴量を用

表 1 2014 年におけるグループ一覧

日付	ハッシュ値	ポート番号	IP アドレス数	IP アドレス
2014/01/01	48303e3e839ed923f7638bbbc7404625	3389	27	5.135.***.133, 91.197.***.33, 109.235.***.219, etc
2014/01/01	7669afc57b6742d1abb7777bf6c8fad9	445	9	190.145.***.227, 221.172.***.115, 112.105.***.27,etc
2014/01/01	682bd3057a6ab9a5b191bd53a4b60179	80, 4899, 4971	12	95.105.***.209, 120.234.***.172, 109.207.***.187,etc
⋮	⋮	⋮	⋮	⋮
2014/12/31	6d1c7b1e800863342aef86ae7fc69228a	80, 4899, 4971	12	119.9.***.14, 119.9.***.64, 162.13.***.234,etc

表 2 生成したパラメータの一例

H	N_H	T_{Hmax}	T_{Hmin}	T_{Havg}	T_{Hsd}	T_{Hmo}	T_{H1}	T_{H2}	T_{H3}
022903f2d29c8e5b5b7aac2c74296d13	12	12	5	7.69	12.56	5	128	360	343
078c78295ffaad2d780d3bf4a6d99080	7	11	1	4.75	26.78	1	308	364	364
0a8ead81bef975756c87a48a83a26d6c	1137	28	1	3.88	50.00	1	265	301	265
0a986294c226383a13010e69e3182bcb	5	31	31	31	0.00	31	313	344	313
0ef9ab407b747d585d47e21dbed51791	113	81	1	9.29	294.49	1	2	233	64
10ae05884a5b74d795729dd8f1f74198	1	15	15	15	0	15	236	251	236
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

い，MDS により各グループを 2 次元平面上にプロットした結果，異なるハッシュ値のアドレスグループが相互に近接している場合には，その年間活動状況が類似していることが確認できた．活動時期やタイミングが極めて類似しているアドレスグループは，送信するペイロードが異なっていても，それらのグループ間には何らかの関係性があるものと考えられる．今後は，特徴が近接するアドレスグループとそのペイロードを個別に分析することにより，アドレスグループのクラスタリングを行い，分散化されたネットワーク走査活動の全容を明らかにする手法について検討する．

参考文献

[1] Eric M. Hutchins, Michael J. Clopperty, Rohan M. Amin, “Intelligence–Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains”, Lockheed Martin Corporation, http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf, last accessed: 13th June 2017.

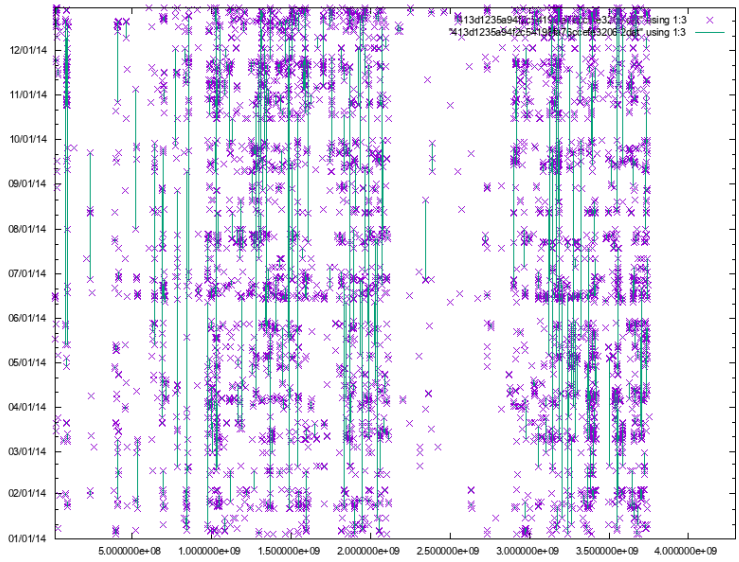
[2] 中村康弘, “初期ペイロードに着目したネットワーク走査活動の分析”, 情報処理学会第 79 回全国大会, 5D-02, Vol.3, pp.523–524, 2017.

[3] 国立研究開発法人 情報通信研究機構 サイバーセキュリティ研究所 サイバーセキュリティ研究室, “NICTER 観測レポート 2016”, http://www.nict.go.jp/cyber/report/NICTER_report_2016.pdf

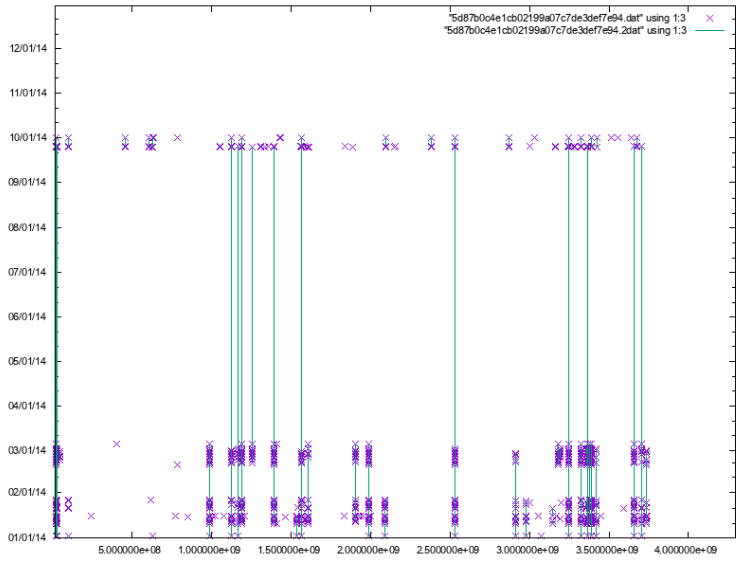
[4] 福島 祥郎, 堀 良彰, 櫻井 幸一,“ダークネット観測データに基づく攻撃挙動の特徴抽出に関する考察” ICSS2009, pp.37–42, 2009.

[5] 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”, IPSJ2013, Vol.4, pp.729–736, 2013.

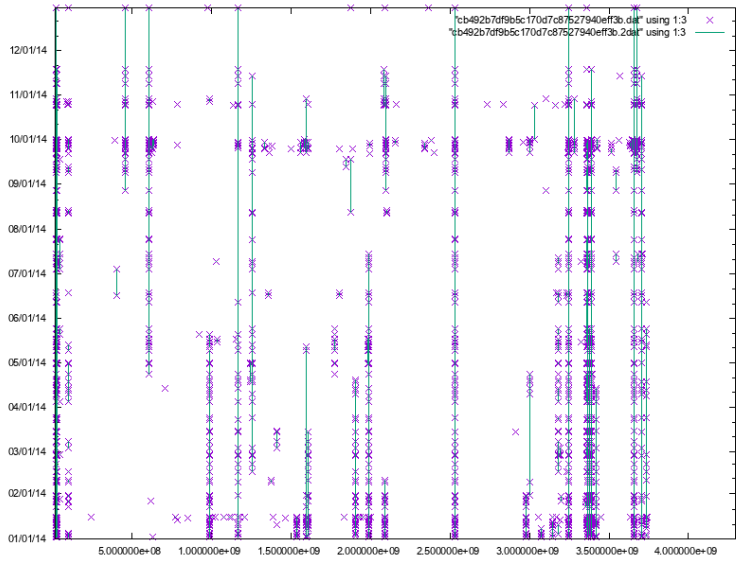
[6] 池部 実, 宮崎 桐果, 吉田 和幸, “ハニーポットによる大分大学におけるダークネット宛通信の分析”, 情報処理学会研究報告, Vol.2015-CSEC-69, No.17, 2015.



(a) ハッシュ値 = 413d1235a94f2c54198fa76ccefe3206



(b) ハッシュ値 = 5d87b0c4e1cb02199a07c7de3def7e94



(c) ハッシュ値 = cb492b7df9b5c170d7c87527940eff3b

図 2 ハッシュ値別アドレスグループの年間活動状況

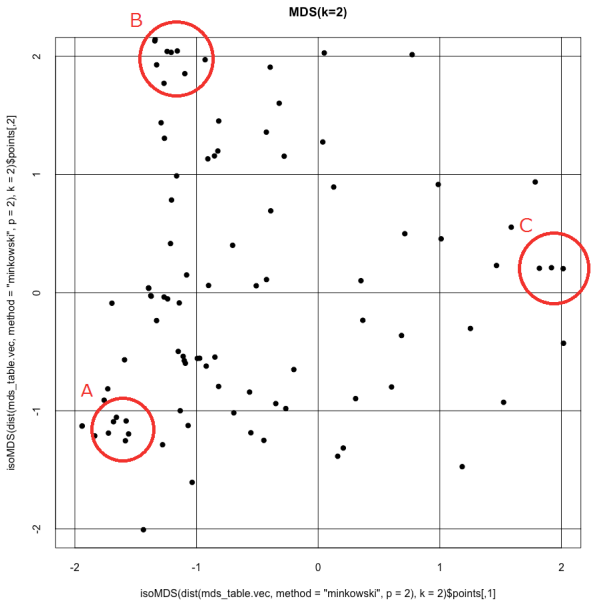


図 3　MDS の結果と注目する 3 つの領域

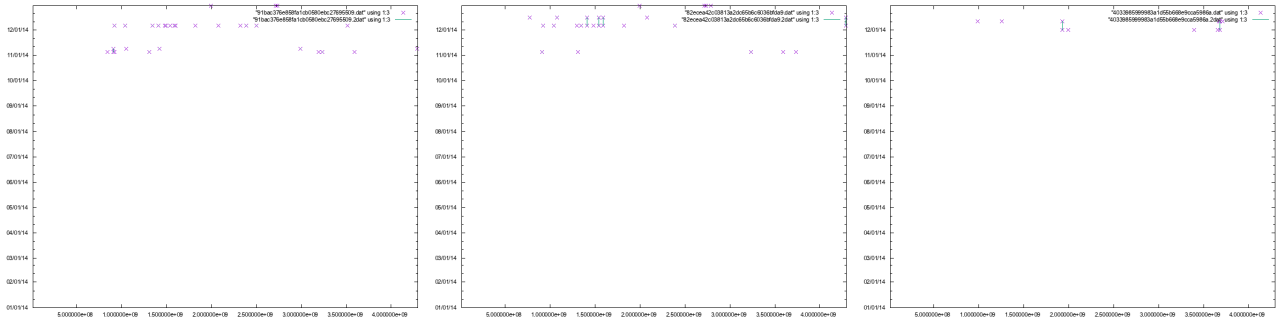


図 4　領域 A の送信元アドレスグループの年間活動状況

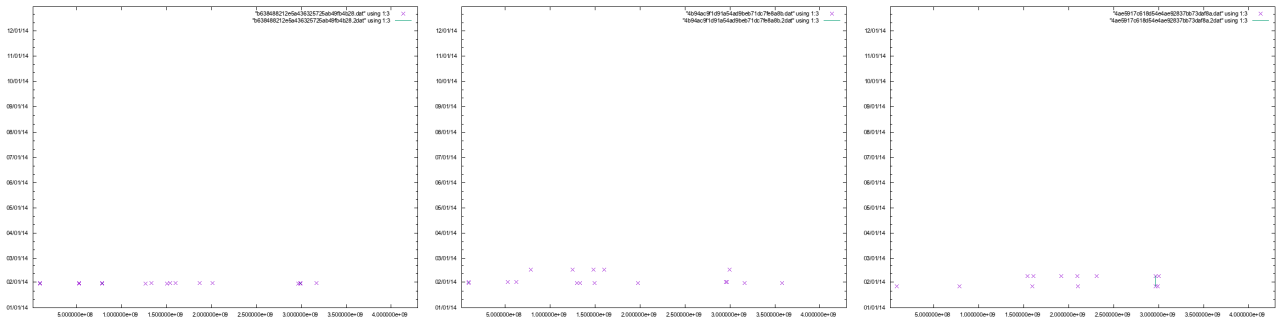


図 5　領域 B の送信元アドレスグループの年間活動状況

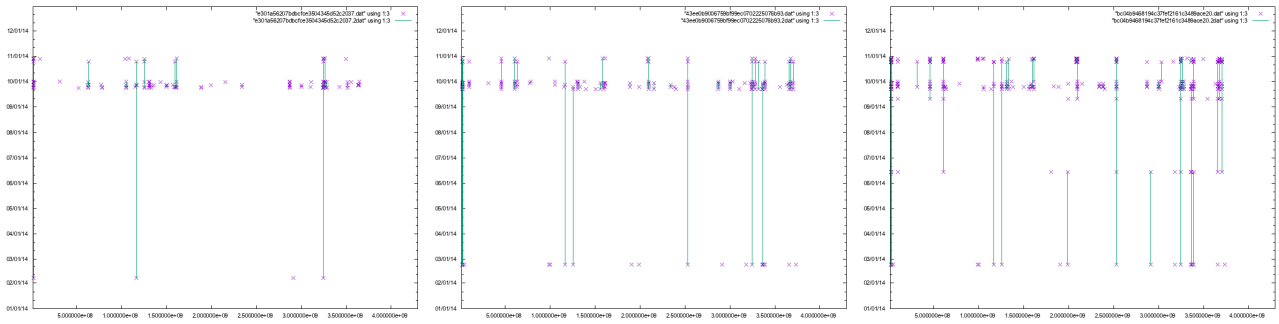


図 6　領域 C の送信元アドレスグループの年間活動状況