

## ハニーポットを用いたアドレスハーベスタと spam 送信者の spam 活動の調査

金高一<sup>†1</sup> 松井一乃<sup>†1</sup> 加来麻友美<sup>†2</sup> 池部実<sup>†2</sup> 吉田和幸<sup>†3</sup>

spam 送信者は、ボットネットを用いて不特定多数のあて先に対して大量の spam を送信する。spam のうち、81.2%はボットに感染したエンドユーザコンピュータから送信される。spam 送信者は、spam のターゲットとなるあて先メールアドレスを入手すること、spam の送信手段であるボットネットを維持・管理することが重要となる。本研究は、アドレスハーベスティングと spam 送信に着目し、ボットネットの spam 活動を調査する。アドレスハーベスタ、spam 送信者がボットネットであるかの分類、ボットネットの活動拠点を調べるため IP アドレスの分布、どのような頻度で活動をするかアクセス回数の3つの観点から調査し spam 活動を考察した。ボットネットの分布に関しては、ハニーポットへアクセスしてきた IP アドレスをヒルベルト曲線にて2次元マトリックスへマッピングした。ヒルベルト曲線は、類似した IP アドレスを近傍に表示できる。ボットネットは集団感染をするという性質から、IP アドレスが近傍に表示された際、同じボットネットに所属している可能性がある。今回の観測から、ボットネットのアクセスのひとつの特徴として、連続する IP アドレスから同時に、短期間で大量のアクセスを試みることを確認した。

### A Study for Spam Activities of E-mail Address Harvesters and Spammers Using HoneyPots

HAJIME KANETAKA<sup>†1</sup> KAZUNO MATSUI<sup>†1</sup> MAYUMI KAKU<sup>†2</sup>  
MINORU IKEBE<sup>†2</sup> KAZUYUKI YOSHIDA<sup>†3</sup>

Spammers send a large amount of spam to a large number of the e-mail addresses using botnet. End-user computers infected with a bot sent 81.2% of all spam. The following activities are important in spammers. (1)To obtain e-mail addresses for spam targets. (2)To manage the botnets for spam sending. Our research purpose is to study the activities of e-mail address harvesters and spam bots from access log of honeypots. In this paper, we investigated the botnets activity focused on spam activity with e-mail address harvesting and spam sending. We investigated botnet activities from the activity base in IPv4 address space of spammers and e-mail address harvesters, classification into botnets, and access frequency to the honeypots. We represent a mapping of the first and second octets of IPv4 addresses of botnet onto the Hilbert curve. The Hilbert curve visualizes the IPv4 address space in 2-dimensions. This curve keeps adjacent address physically near each other. If IP addresses appear in the vicinity, the IP addresses might belong to the same botnet. As a result, we verified one of features of botnets that amount of access from consecutive IP addresses have visited in short term.

#### 1. はじめに

インターネットの急速な発展と普及に伴い、メールを始めとするネットワークを介したコミュニケーションは必要不可欠となっている。メールと通常の郵便を比べると、送信者は容易かつ安価に大量のメッセージを送信できる。これに伴い spam が大きな社会問題となっている。spam は受信者の意図を無視して無差別かつ大量に一括して送信されるメールを指し、UCE (Unsolicited Commercial E-mail)、UBE (Unsolicited Bulk E-mail)とも呼ばれる。現在、インターネットを流れるメールの約64%が spam である[1]。Symantec社の調査報告によると spam のうち、ボットに感染したエンドユ

ーザコンピュータから送信される割合は81.2%である[2]。メールが我々の生活にとって不可欠なサービスである現在において、添付によるウイルス感染、大量の spam に紛れた通常メールの紛失による受信者の精神的ストレスなどの spam による被害は無視できない問題である。

アドレスハーベスティングは、spam 送信の前段階として、ターゲットとなる受信者のメールアドレスを収集する行為である。一般的に spam は無差別かつ大量に送信されるため、多くのメールアドレスを収集する必要がある。spam 送信とアドレスハーベスティングには密接な関係がある。

本研究は、ボットネットを用いたアドレスハーベスティングと spam 送信それぞれの IP アドレスの近接関係、ハニーポットに対するアクセスの種類、アクセス回数を調査し、ボットネットによる被害を防ぐため、spam 活動の特徴からボットネットの活動を考察する。

本論文の構成は以下の通りである。第2章で研究目的について述べ、第3章では関連研究について述べる。そして第4章で本研究の調査方法について述べ、第5章で調査結

†1 大分大学大学院工学研究科知能情報システム工学専攻  
Course of Computer Science and Intelligent Systems, Graduate School of Engineering, Oita University

†2 大分大学工学部知能情報システム工学科  
Department of Computer Science and Intelligent Systems, Faculty of Engineering, Oita University

†3 大分大学学術情報拠点情報基盤センター  
Center for Academic Information and Library Services, Oita University

果について述べ、考察する。最後に第6章でまとめと今後の課題について述べる。

## 2. 調査対象の spam 活動

### 2.1 spam 活動とボットネット

インターネットを流れるメールの約64%がspamであり、そのうち81.2%が、ボットに感染したエンドユーザコンピュータから送信される。ボットとは、悪意を持った第三者がインターネットを通じてコンピュータを外部から遠隔操作することを目的とした悪性プログラムのことである。アンチウイルスソフトによるウイルス対策がなされていないコンピュータが多数存在するネットワークにおいてボットが1台でも感染すると、感染したコンピュータが属するネットワークのコンピュータがボットに集団感染することがある。インターネット上には複数のボットネットが存在しており、ボットネットはメールアドレスやソフトウェアのシリアルナンバーなどの重要な情報の収集などにも用いられる[3]。

本研究ではボットネットの活動を調べるために、ボットネットを使った活動の中でも、特にボットネット使用率が高いとされるspam活動に着目した。spam活動には、spamを送信するだけでなく、spamを送信する前段階であるアドレスハーベスティングも含まれる。ボットネットを使用したspam活動として、アドレスハーベスティング、spam送信について調査した。

### 2.2 アドレスハーベスティング

spam活動のうちspam送信は、ターゲットとなる受信者の意図を無視して無差別かつ大量に一括して送信する行為である。本節では、spam活動の残りのアドレスハーベスティングについて説明する。

アドレスハーベスティングは、spamのターゲットとなるメールアドレスを収集する行為を意味し、以下の3つの方法に分類される[4]。

#### (a) 公開情報(Web ページ)からのメールアドレス収集

Webサイトを巡回し、Webページからメールアドレスを収集する。

#### (b) 非公開情報からのメールアドレス収集

社外などへ公開していない情報システムに不正侵入し、顧客名簿などからメールアドレスを入手する。

#### (c) プログラムによるメールアドレスの生成

辞書ファイルを用いてローカルパート(@マークよりも前の部分)をプログラムにより生成し、あて先ドメインと組み合わせて、メールアドレスを生成する。有効なメールアドレスの情報を取得する目的で大量のメールを送りつけるディレクトリハーベスティング攻撃(Directory Harvesting Attack; DHA)はこれに分類さ

れる。

アドレスハーベスタがメールアドレスを収集するうえで効率が良い方法は、ボットネットやプログラムを用いて自動でメールアドレスを収集する方法である。(b)の方法は確実に使用されているメールアドレスを入手することができるが、他の方法と比べてメールアドレス入手のためにかかる手間が大きい。(b)に比べ、(a)、(c)はプログラムを実行するだけで人の手をかけずに、メールアドレスを入手できる。さらに、それらのプログラムをボットネット上で実行すれば大量のメールアドレスを短期間で大量に入手することができる。アドレスハーベスティングはspam活動の準備段階として非常に重要である。

### 2.3 ボットの分類

ボットは、悪意を持った第三者がインターネットを通じて感染したコンピュータを外部から遠隔操作して、spam送信や、Webページからの情報を収集するための悪性プログラムである。

本論文では、spam活動のうちメールアドレスを収集するボットを”アドレスハーベスタ”、spamを送信するボットを”spamボット”と定義する。

また、検索エンジンのインデックス作成のために、Webページの内容を収集する良性のボットが存在する(例: Googlebot, Baiduspider)。本論文では、これらのボットを”クローラ”と呼ぶ。

## 3. 関連研究

本研究ではボットネットによるアドレスハーベスティングとspam送信のspam活動を調査する。そこで、関連研究としてハニーボットを用いてspam活動を調査した研究について述べる。

Oliverら[5]は、Web上に存在するメールアドレスを収集するハーベスティングボットの収集プロセスとそれら进行操作するハーベスタを3年にわたり調査した。その結果、様々なspamの原因はハーベスティング活動を行う特定のわずかな人物・組織であること、またWeb上でメールアドレスを公開するにあたり、JavaScriptを用いた難読化技法がメールアドレスを収集されることを防ぎ、spamを受信しないために有効であることなどを示した。

Wilfriedら[6]は、運用開始時期やmailto掲載数の異なる複数のハニーボットを用いてspamを8ヶ月間収集し、どのハニーボットにspamが送られてきたかを調査し、spamトラフィックとしていくつかの種類に分類した。その結果、メールアドレスを公開して初めてspamを受信するまでの時間が最速で13日、最遅で64日であり、メールアドレス公開からspam受信まで最低でも10日以上の間隔があること。また、多くのspamは公開しているメールアドレス宛

てに送信してくるが、一部の spam は存在しないアカウントへ送信してくることを観測し、greylisting などの spam 対策の排除率を検証した。

本研究では、アドレスハーベスタ、spam 送信者のハニーポットへのアクセスログから、アクセス元を S25R や whois によって分類することにより、ボットネットの発見、spam 活動を調査する。

#### 4. アドレスハーベスタと spam ボットの調査

本研究は、アドレスハーベスティングと spam 送信に着目し、ボットネットの spam 活動を調査する。アドレスハーベスタ、spam 送信者がボットネットであるかの分類、ボットネットの活動拠点を調べるために IP アドレスの分布、どのような頻度で活動をするかアクセス回数の3つの観点から調査し spam 活動を考察した。

##### 4.1 調査環境

本研究における調査環境の構成を図1に示す。アドレスハーベスタと spam 送信者の情報を収集する手順を以下に示す。

- (1) Web サーバは、CGI(Common Gateway Interface)によってページ要求ごとに異なるメールアドレスを含む Web ページを返す。メールアドレスは、クライアントがページを要求したアクセス時刻をもとにして、自動生成する。メールアドレスにアクセス時刻を含むことにより、メールを受信した際、どのタイミングで収集されたかを判別できる。
- (2) メールサーバは、Web サーバで公開したメールアドレスの他に、自身のドメイン宛で存在しないアカウントに送信されたメール、オープンリレーによる第三者宛へのメールを含むすべてのメールを spam として受信する。

本研究では(1)のアクセスログと(2)のメールログを調査することにより、アドレスハーベスタと spam 送信者の活動状況を調査する。

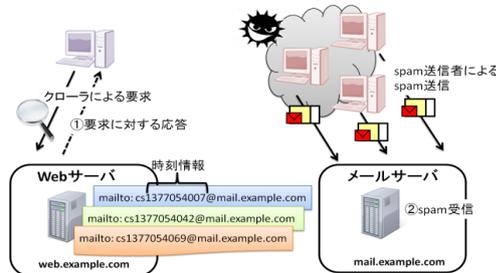


図1 アクセス情報の収集

##### 4.2 IP アドレス空間の表現方法

多次元空間の情報を一次元に写像する手法として、空間充填曲線が用いられる。空間充填曲線の代表的なものとし

て、ヒルベルト曲線[7][8]、ルベーク曲線(Z-Ordering)、シェルピンスキー曲線、ゴスパー曲線などがある。二次元空間充填曲線は平面上において近隣のノードが一次元線上でも近くに配置されるように空間を充填するため、平面上のノードの近接性と変換後の一次元線上での近接性に相関をもたせることができる[9][10]。

空間充填曲線を用いる利点は、図2に示す空間充填曲線を用いない場合と比べ、連続する IP アドレスを1つの塊として描写できるため、アクセス元の IP アドレスの近接関係の把握が容易になる。またアドレスハーベスタと spam 送信者に関する分布を比較する際に、IP アドレスがどの IP アドレスブロックに存在しているか分布の差分を調べる際に直感的に違いが分かる点にある。

図3に示すルベーク曲線は7-8間、15-16間、31-32間、63-64間、127-128間では IP アドレスブロックが大きく離れている部分がある。シェルピンスキー曲線など他の曲線も基本図形の性質から同様に IP アドレスブロックが大きく離れる部分ができる。これに対してヒルベルト曲線は、IP アドレスブロックが大きく離れることなく近接性を保持したまま描写できる。このことから、本研究では他の空間充填曲線と比較して IP アドレスの近接関係の描写に優れているヒルベルト曲線を用いる。

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

図2 空間充填曲線を用いない0~255の分布の描写の一例

0	7	16	27	40	55	72	91	112	135	160	187	216	247
2	9	18	29	42	59	78	99	122	147	174	203	234	267
5	14	25	38	52	67	84	103	126	151	178	207	238	271
10	21	34	49	64	81	100	121	144	169	196	225	256	289
17	30	45	62	81	101	122	145	170	197	224	253	284	317
26	41	58	77	98	120	143	168	195	222	251	282	313	346
37	54	73	94	117	141	166	193	220	249	278	309	340	373
50	71	94	119	145	172	201	230	260	291	322	353	384	417
65	90	117	146	174	203	232	262	293	324	355	386	417	450
82	109	138	169	200	231	262	293	324	355	386	417	448	481
101	132	165	200	231	262	293	324	355	386	417	448	479	512
122	157	194	231	262	293	324	355	386	417	448	479	510	543
145	184	223	262	293	324	355	386	417	448	479	510	541	574
170	213	254	293	324	355	386	417	448	479	510	541	572	605
197	244	287	324	355	386	417	448	479	510	541	572	603	636
226	277	318	355	386	417	448	479	510	541	572	603	634	667
257	300	343	384	417	448	479	510	541	572	603	634	665	698
290	333	378	417	448	479	510	541	572	603	634	665	696	729
325	368	407	448	479	510	541	572	603	634	665	696	727	760
362	405	446	479	510	541	572	603	634	665	696	727	755	788
401	440	477	510	541	572	603	634	665	696	727	755	783	816
442	481	518	541	572	603	634	665	696	727	755	783	811	844
485	524	561	584	603	634	665	696	727	755	783	811	839	872
530	569	606	629	648	667	696	727	755	783	811	839	867	900
577	616	653	676	695	724	753	782	810	838	866	894	922	955
626	665	702	725	744	773	802	831	859	887	915	943	971	1004
677	716	753	776	795	824	853	882	910	938	966	994	1022	1055
730	769	806	829	848	877	906	935	963	991	1019	1047	1075	1108
785	824	861	884	903	932	961	990	1018	1046	1074	1102	1130	1163
842	881	918	941	960	989	1017	1045	1073	1101	1129	1157	1185	1218
901	940	977	1000	1019	1048	1076	1104	1132	1160	1188	1216	1244	1277
962	1001	1038	1061	1080	1109	1137	1165	1193	1221	1249	1277	1305	1338
1025	1064	1101	1124	1143	1172	1200	1228	1256	1284	1312	1340	1368	1401
1090	1129	1166	1189	1208	1237	1265	1293	1321	1349	1377	1405	1433	1466
1157	1196	1233	1256	1275	1304	1332	1360	1388	1416	1444	1472	1500	1533
1226	1265	1302	1325	1344	1373	1401	1429	1457	1485	1513	1541	1569	1602
1297	1336	1373	1396	1415	1444	1472	1500	1528	1556	1584	1612	1640	1673
1370	1409	1446	1469	1488	1517	1545	1573	1601	1629	1657	1685	1713	1746
1445	1484	1521	1544	1563	1592	1620	1648	1676	1704	1732	1760	1788	1821
1522	1561	1598	1621	1640	1669	1697	1725	1753	1781	1809	1837	1865	1898
1601	1640	1677	1699	1718	1747	1775	1803	1831	1859	1887	1915	1943	1976
1682	1721	1758	1781	1800	1829	1857	1885	1913	1941	1969	1997	2025	2058
1765	1804	1841	1864	1883	1912	1940	1968	1996	2024	2052	2080	2108	2141
1850	1889	1926	1949	1968	1997	2025	2053	2081	2109	2137	2165	2193	2226
1937	1976	2013	2036	2055	2084	2112	2140	2168	2196	2224	2252	2280	2313
2026	2065	2102	2125	2144	2173	2201	2229	2257	2285	2313	2341	2369	2402
2117	2156	2193	2216	2235	2264	2292	2320	2348	2376	2404	2432	2460	2493
2210	2249	2286	2309	2328	2357	2385	2413	2441	2469	2497	2525	2553	2586
2305	2344	2381	2404	2423	2452	2480	2508	2536	2564	2592	2620	2648	2681
2402	2441	2478	2501	2520	2549	2577	2605	2633	2661	2689	2717	2745	2778
2501	2540	2577	2600	2619	2648	2676	2704	2732	2760	2788	2816	2844	2877
2602	2641	2678	2701	2720	2749	2777	2805	2833	2861	2889	2917	2945	2978
2705	2744	2781	2804	2823	2852	2880	2908	2936	2964	2992	3020	3048	3081
2810	2849	2886	2909	2928	2957	2985	3013	3041	3069	3097	3125	3153	3186
2917	2956	2993	3016	3035	3064	3092	3120	3148	3176	3204	3232	3260	3293
3026	3065	3102	3125	3144	3173	3201	3229	3257	3285	3313	3341	3369	3402
3137	3176	3213	3236	3255	3284	3312	3340	3368	3396	3424	3452	3480	3513
3250	3289	3326	3349	3368	3397	3425	3453	3481	3509	3537	3565	3593	3626
3365	3404	3441	3464	3483	3512	3540	3568	3596	3624	3652	3680	3708	3741
3482	3521	3558	3581	3600	3629	3657	3685	3713	3741	3769	3797	3825	3858
3601	3640	3677	3700	3719	3748	3776	3804	3832	3860	3888	3916	3944	3977
3718	3757	3794	3817	3836	3865	3893	3921	3949	3977	4005	4033	4061	4094
3837	3876	3913	3936	3955	3984	4012	4040	4068	4096	4124	4152	4180	4213
3958	3997	4034	4057	4076	4105	4133	4161	4189	4217	4245	4273	4301	4334
4075	4114	4151	4174	4193	4222	4250	4278	4306	4334	4362	4390	4418	4451
4194	4233	4270	4293	4312	4341	4369	4397	4425	4453	4481	4509	4537	4570
4313	4352	4389	4412	4431	4460	4488	4516	4544	4572	4600	4628	4656	4689
4432	4471	4508	4531	4550	4579	4607	4635	4663	4691	4719	4747	4775	4808
4551	4590	4627	4650	4669	4698	4726	4754	4782	4810	4838	4866	4894	4927

ヒルベルト曲線には図 5 に示す 4 つの基本図形がある。基本図形を ULD, DRU, RDL, LUR とする。この基本図形を 4 つのルールに従って再帰的に呼び出すことでヒルベルト曲線ができる。4 つのルールを以下に示す。↑, ↓, →, ← は各方向への描画を示す。

ルール 1 を例にルールを説明する。まず, DRU(n) が呼び出されると, RDL(n-1) を呼び出す。RDL(n-1) の描画が終了すると, RDL(n-1) の描画が終了した場所から ↓ 方向に線を描画し, DRU(n-1) を呼び出す。以後同様に DRU(n-1), →, DRU(n-1), ↑, LUR(n-1) の順で呼び出し, n=0 になるまで繰り返す。

• ルール 1 (Down-Right-Up)

DRU(n)=RDL(n-1), ↓, DRU(n-1), →, DRU(n-1), ↑, LUR(n-1)

• ルール 2 (Up—Left-Down)

ULD(n)=LUR(n-1), ↑, ULD(n-1), ←, ULD(n-1), ↓, RLD(n-1)

• ルール 3 (Right-Down-Left)

RDL(n)=DRU(n-1), →, RDL(n-1), ↓, RDL(n-1), ←, ULD(n-1)

• ルール 4 (Left-Up-Right)

LUR(n)=ULD(n-1), ←, LUR(n-1), ↑, LUR(n-1), →, DRU(n-1)

IP アドレスの第 1 オクテットをヒルベルト曲線で表現すると, クラス A~E のクラスフル IP アドレスは図 6 となる。

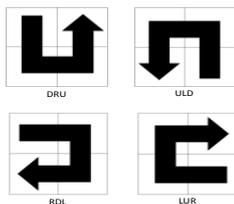


図 5 ヒルベルト曲線の基本図形

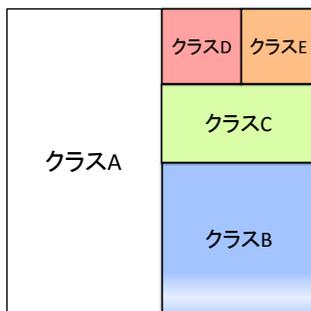


図 6 ヒルベルト曲線を用いたクラス A~E のクラスフル IP アドレスの表現

本論文ではハニーポットへアクセスしてきた IP アドレスの上位 16 ビット(第 1 オクテット, 第 2 オクテット)を 1 ブロックとして扱い, ヒルベルト曲線を用いてアドレスハーベスタ, spam 送信者の分布を表現する。その結果から, ボットネットとして使用されているアドレスハーベスタと spam 送信者の IP アドレス空間における活動拠点を知らるために, 分布と近接関係を調査する。

### 4.3 エンドユーザコンピュータの推定

Symantec 社の調査報告[2]によると, spam の 81.2% はボットに感染したエンドユーザコンピュータから送信されている。ボットに感染したエンドユーザコンピュータからの spam を排除する対策に S25R(Selective SMTP Rejection)がある[11]。S25R はメール送信元の IP アドレスを逆引きして得た FQDN(Fully Qualified Domain Name)を規則と照合し, エンドユーザコンピュータを推定し, SMTP(Simple Mail Transfer Protocol)アクセスを拒否する。企業や学術機関, ISP が管理するメール送信サーバの多くは FQDN を設定している。しかし, エンドユーザコンピュータの多くは FQDN に IP アドレスの数字を含む場合や, FQDN を設定していないことが多い[12]。S25R で検知できる FQDN の例を図 7 に示す。

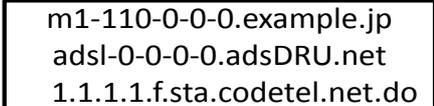


図 7 S25R で検知できる FQDN の例

本研究ではボットネットの spam 活動を調査するため, ハニーポットで収集したアドレスハーベスタと spam 送信者をエンドユーザコンピュータとその他に分類した。以下に分類の手順を示す。

① S25R による分類

IP アドレスを逆引きして得た FQDN を S25R の規則と照合し, 一致したらエンドユーザコンピュータに一致しなければその他に分類する。

② RBL(Real-time Blocking List)[13]による分類

S25R でその他に分類したホストが RBL に登録されていたならばエンドユーザコンピュータに, 登録されていなければその他に分類する。

エンドユーザコンピュータからのアクセスであれば, ボットに感染している可能性が考えられる。また本研究におけるエンドユーザコンピュータは, 一般家庭で使用されているコンピュータだけでなく企業や組織でも使用されているコンピュータを含む。

## 5. 調査結果

### 5.1 Web サーバ, メールサーバへのアクセス分類

調査期間は 2013 年 8 月 11 日から 2013 年 10 月 28 日までの約 3 カ月間である。ハニーポットへのアクセスには, ポートスキャンが含まれている。そこで, 我々が開発・運用している不正通信検知システム[14]を用いて, アクセスしてきた IP アドレスをスキャン攻撃かどうか判定した。ただし HTTP(HyperText Transfer Protocol)と SMTP(Simple Mail Transfer Protocol)は, 誤検知が多いため不正通信検知システムの検知対象ポートから除外している。表 1 にハニーポッ

トで観測したアクセスのうち不正通信検知システムで検知した IP アドレスの数を示す。

表 1 ハニーポットへのアクセスのうち不正通信検知システムで検知された IP アドレスの数

ハニーポット	IP アドレスの総数	不正通信検知システムで検知した IP アドレス数
Web サーバ	414	48
メールサーバ	61	4

(集計期間：2013/08/11~2013/10/28)

不正通信検知システムでは、ハニーポットだけでなく他のホストや 80 番、25 番のポート以外に対して水平スキャンや垂直スキャンを仕掛けていた IP アドレスを攻撃者として検出している。今回の調査で収集したハニーポットのアクセスログと検知システムのログを照合した結果、ハニーポットと不正通信検知システムに同時刻にアクセスがあったものをポートスキャンと判断した。

ポートスキャンを除外したハニーポットのアクセスのうち、検索エンジンのインデックス作成のためのクローラからのアクセスをユーザエージェントから判断した。本論文では、“Googlebot”、“bingbot”、“Baiduspider”などをクローラとした。これらが正しいユーザエージェントか確かめるために S25R、RBL の結果を用いて判別した。

正規の Googlebot であれば、crawl-A.B.C.D-googlebot.com などからアクセスする(A.B.C.D は数字)。これらは S25R の規則から除外する。ユーザエージェントを Googlebot と偽装した IP アドレスからのアクセスの場合、FQDN は S25R のルールセットに該当すると推測できる。

これらのホストが RBL に登録されているか調査した結果、RBL では該当する IP アドレスはなかった。表 2 に分類したクローラ、アドレスハーベスタ、spam 送信者の IP アドレスを逆引きし、エンドユーザコンピュータかどうか判別した結果を示す。

表 2 クローラ・アドレスハーベスタ・spam 送信者の S25R によるホスト分類

	エンドユーザコンピュータ	その他	総数
クローラ	53	96	149
アドレスハーベスタ	169	48	217
spam 送信者	46	11	57

(集計期間：2013/08/11~2013/10/28)

表 2 の結果より、アドレスハーベスタや spam 送信者に関しては、エンドユーザコンピュータと思われるアクセスが 8 割ほどを占めており、ボットに感染したコンピュータからのアクセスの疑いがある。また表 2 に示すクローラに関して、エンドユーザコンピュータとして分類した 53 件は、ユーザエージェントを偽装したアドレスハーベスタの疑いがあるため、これ以降はアドレスハーベスタとして扱う。

表 3 Web サーバへのアクセス回数

	アクセス回数
クローラ	764 回
アドレスハーベスタ	768 回
アクセス総数	1532 回

(集計期間：2013/08/11~2013/10/28)

表 2、表 3 より観測したアドレスハーベスタの IP アドレスの数がクローラと比べて多いにもかかわらず、アクセス回数にあまり差がみられなかった。この理由として、クローラは新しい Web ページや、更新された Web ページを発見し、収集結果をデータベースに反映させるためにアクセスを繰り返す。それに対して、アドレスハーベスタは、メールアドレスの収集が目的のため、1 度アクセスしてメールアドレスを収集した Web ページに再訪することは少ないためであると推定される。

## 5.2 IP アドレスの分布

これ以降、Web サーバへのアクセスはクローラを除外し、アドレスハーベスタのみを調査対象とする。

図 8、図 9 はハニーポットへアクセスしてきたアドレスハーベスタと spam 送信者の分布をヒルベルト曲線にて表現し、IP アドレスブロックの色でアクセス数を表現している。論文中の図では、1 ドットを IP アドレス上位 8 ビットで表現している。実際には、1 ドットを IP アドレス上位 16 ビットとして描画し、調査した。1 ブロックを 8 ビットと 16 ビットで描写した場合では、32 ビットで表現した場合と比べて IP アドレスの近接性が曖昧になるが、アクセス数を色で表現することで近接性の問題点を補完する。

図 8 (1) は、Web サーバへの単一の IP アドレスからの大量のアクセスがあったブロックである。これらの IP アドレスは全て逆引きが設定されていなかった。図 8 (2) は ISP に所属する 47 の IP アドレスからのアクセスであり、逆引きが設定されていなかった。このアクセスは、ISP を介して身元を隠し、様々な情報を収集あるいは spam を発信するアクセスが多く、アドレスハーベスタからのアクセスの可能性はある。

図 9 (3) は、1 つの IP アドレスから大量にアクセスがあったブロックである。アクセスは 1 秒間隔で 510 回であった。

この IP アドレスは逆引きが設定されており、PC 販売業者ということがわかった。図 9(4)は 13 個の連続した IP アドレスからのアクセスであり、1 つの IP アドレスから 2 秒間隔で 181 回のアクセスがあった。これらはフィッシングサイトとしてブラックリストに報告されている IP アドレスからのアクセスであり、すべて逆引きが設定されていなかった。

ハニーポットへアクセスしてきた IP アドレスは重複していなかった。今回の調査期間中では、アドレスハーベスタ、spam 送信者の関係性を確認できなかった。

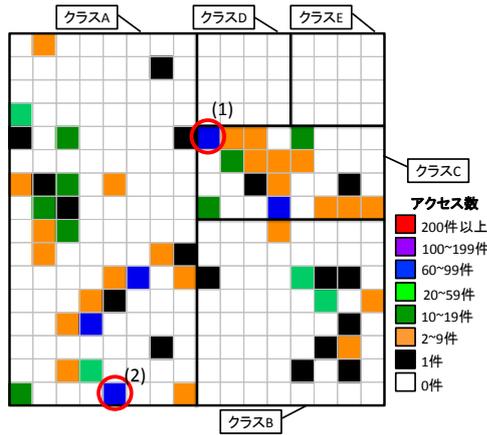


図 8 Web サーバへのアクセス数と IP アドレスの分布

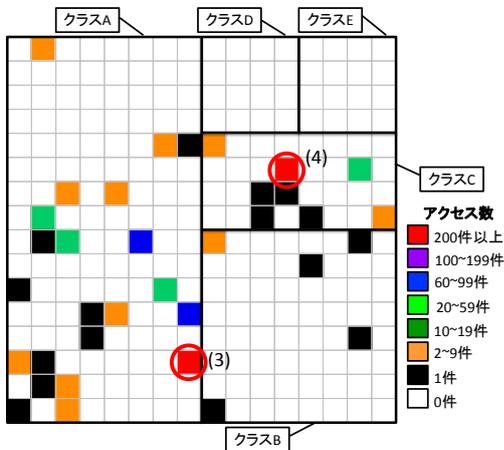


図 9 メールサーバへのアクセス数と IP アドレスの分布

### 5.3 IP アドレス別リクエスト分類

本調査では、アドレスハーベスタの Web ページへのアクセスを追跡するため、同じ内容の Web ページを複数用意した。用意したページは、以下のとおりである。

- index.html
- index-english.html
- kadai[1-20].cgi
- eng[1-20].cgi
- kadai[11-20].html
- eng[11-20].html

本調査では、index.html、index-english.html は、それぞれ日本語と英語で記述したトップページ、index.html からリンクしたメールアドレスを含む kadai[1-20].cgi と

kadai[11-20].html、index-english.html からリンクしたメールアドレスを含む eng[1-20].cgi と eng[11-20].html の計 62 の Web ページを用いた。kadai[11-20]と eng[11-20]について同じ内容を CGI と HTML で別々に提供している。そのため今回の集計期間には kadai[11-20]と eng[11-20]について CGI のページと HTML のページ 2 つが混在している。図 10 にそれぞれのページへのアクセス頻度、表 4 にアクセスをページ毎に集計した結果を示す。表 4 では、アクセス回数の上位 5 件の IP アドレスを示している。

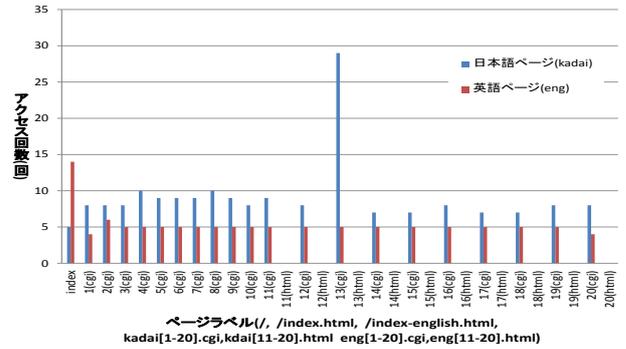


図 10 各ページへのアクセス頻度  
(集計期間：2013/08/11~2013/10/28)

表 4 Web サーバへのアクセス上位 5 件

IP アドレス	アクセス回数	アクセスしたページ
A	69 回	phpMyAdmin
B	43 回	すべてのページ
C	40 回	index-english.html, kadai[1 2 7 8 9 10 12 14 15 17].cgi
D	38 回	kadai[3 4 5 6 11 13 16 18 19 20].cgi
E	23 回	index.html, index-english.html, eng[1-19].cgi

(集計期間：2013/08/11~2013/10/28)

図 10 より、調査期間内に html ファイルのページへのアクセスはみられなかった。しかしアドレスハーベスタとして分類したホストの多くは偏りなく CGI ファイルのページへアクセスしていることが分かる。図 10 の kadai13.cgi だけが他のページと比べアクセスが突出していた。このときのアクセス傾向は、同じネットワークセグメントに所属する複数のホストからそれぞれ 1 回ずつのアクセスを確認した。観測した 2 つのネットワークセグメントからのアクセスのうち、1 つは逆引きを設定してあるホスティングサービスが運用しているネットワークからのアクセス、残りの 1 つは逆引きを設定していない ISP からのアクセスである。同じネットワーク上に存在し、同じ Web サーバの異なるページをそれぞれ参照していることから、これらはポットに

感染したエンドユーザコンピュータの可能性があると考えられる。アクセス元を調査したが、なぜこのページだけが増えているのか明確な理由は不明である。

表 4 に示す最もアクセスを試みた IP アドレス A は、phpMyAdmin に対するページ要求であり、脆弱性を狙った攻撃であると考えられる。IP アドレス C, D についてはクラス C に属する近隣のホストからのアクセスであり、互いに Web ページのファイル名に対して、重複がないようにアクセスしてきた。この 2 つのホストには逆引きが設定されていなかった。IP アドレス A 以外のホストについては、トップページから始まりほぼすべてのページを順に参照していたことから、ページに埋め込まれたメールアドレスなどの目的の情報を探す、ボットネットだと推察した。

#### 5.4 spam メールに関する分類

メール送信者のメールサーバへ対する挙動を目的別に集計した結果を表 5 に示す。メール送信者の挙動は「自ドメイン宛へのメール」、「オープンリレーを目的としたメール」、「SMTP セッションを張るのみで何も行わない」の 3 つに分類した。

表 5 目的別にみた spam 送信者の分類

目的	サーバ数
自ドメイン宛	0
オープンリレー	9
セッションのみ	48

(集計期間：2013/08/11~2013/10/28)

表 5 より、spam 送信者は、メールアドレスを入手したターゲットへ直接 spam を送るのではなく、オープンリレーを許可している MTA を介して身元がばれないように spam を送る傾向がある。次に表 5 の「セッションのみ」においてどのタイミングでセッションを切断したかの割合を表 6 に示す。1 つの IP アドレスから複数のセッションを受ける場合もあるため、ここでは割合として示す。全セッションの数は 3401 回である。

表 6 SMTP コマンド内訳

SMTP コマンド	回数	割合	サーバ数
lost connection after CONNECT from ****	5 回	0.15%	3
possible SMTP attack	3 回	0.09%	2
did not issue MAIL/EXPN/VRIFY/ETRN during connection to MTA	3393 回	99.7%	47

(集計期間：2013/08/11~2013/10/28)

表 6 に示すコマンド内訳のうち、セッションを張るのみで何もせずにタイムアウトになったことを意味する” did

not issue MAIL/EXPN/VRIFY/ETRN during connection to MTA” が 99% を占めていた。このログは、外部ネットワークからメールサーバを使用してメールを送信できるシステムの有無を確認しようとした場合に出され、不正中継を試みるためのサーバチェックの可能性がある。そのため、メールを送信することが目的ではなく MTA のサービスの生存確認、25 番ポートを狙った不正通信検知システムで検知されなかったホストからのポートスキャンの可能性はある[15]。

表 7 メールサーバへのアクセス上位 5 件

IP アドレス	アクセス時刻	アクセス間隔	アクセス回数
F	10/8 07:47-7:55	1 秒	510 回
G	9/20 14:57-15:10	2 秒	465 回
H	9/8 13:18-13:18, 9/16 2:55-2:58, 10/19 15:45-15:47	2 秒	181 回
I	10/9 11:11-11:11	1 秒間に 10 回以上	65 回
J	9/18 2:35-2:35	1 秒間に 10 回以上	65 回

(集計期間：2013/08/11~2013/10/28)

表 7 は Web サーバへ最もアクセスを試みた上位 5 件を示す。IP アドレス F~J のアクセスの内容はすべて” did not issue MAIL/EXPN/VRIFY/ETRN during connection to MTA”であった。それぞれの IP アドレスを学内ファイアウォールの外側のスイッチで取得したミラーパケットで確認した結果、IP アドレス I は、学内ホストの 25 番ポートに対して水平スキャンをしており、スキャン攻撃後にメールサーバの稼働を確認して大量にアクセスをしてきたと推測する。IP アドレス G, H についてはメールサーバの 25 番ポートを狙ってアクセスしてきていた。また、IP アドレス H については、H の他に同じネットワークに所属する連続する 11 のホストから同じく 181 回のアクセスを観測しており、更にこれらのホストはフィッシング詐欺サイトとして報告されていた。IP アドレス I, J についてもメールサーバの所属するサブドメイン宛にアクセスをしていた。また、IP アドレス I 以外には逆引きが設定されていなかった。

一般ユーザが 1 秒おきに何度もアクセスを繰り返すということは、ほとんど無いと考えられる。そのため、今回のメールサーバへのアクセスは、自前の MTA をもつ spam 送信業者か、ボットに感染したエンドユーザコンピュータを使ったプログラムからのアクセスだと推測できる。

## 5.5 考察

ハニーポットを用いてアドレスハーベスティングと spam 送信について、spam 活動を調査した。その結果、ハニーポットへのアクセスにおいてボットネットの疑いのあるアクセスを発見した。Web サーバに関しては Web ページを探索してメールアドレスを収集する動きがみられたが、メールサーバに関しては SMTP システムの悪用を目的としたサービス確認のためのアクセスが大半であり、spam を受信できなかった。今回 spam 送信者に関しては十分なデータを集められなかった。Wilfried ら[6]の調査によると、メールアドレスを公開してから最遅でも 64 日で spam を受信したが、今回の 3 ヶ月間の調査では spam を 1 通も受信することはなかった。しかし、これから調査期間を増やし、時間が経つにつれてメールアドレスやドメインがアドレスハーベスタや spam 送信者に拡散し spam 送信や DHA のターゲットにされ、データを収集できると考えられる。

## 6. まとめ

ボットネットの活動をアドレスハーベスタと spam ボットの spam 活動に注目して調査した。ハニーポットとして仕掛けた Web サーバとメールサーバのアクセスログから、アドレスハーベスタや spam ボットがどのような IP アドレスブロックを拠点に活動しているか、また、それらのアクセス時刻やアクセスの規則性について調査し、特徴を発見した。今回の観測からひとつの特徴として、ボットネットは短期間に大量のアクセスを試み、連続する複数の IP アドレスからのアクセスである場合、Web ページへのアクセスは、互いに Web ページのファイル名に対して、重複がないようにアクセスをする。メールサーバへのアクセスに関しては、アクセス間隔やアクセス回数が等しくなることを確認した。

また、大学や研究所などの歴史的 PI(Provider Independent) アドレスを所有するクラス B よりも、CIDR(Classless Inter-Domain Routing)を用いて企業や ISP に IP アドレスを割り振られているクラス A、クラス C を拠点としたボットネットからのアクセスを多く観測したことから、組織や ISP などによる対策により、ボットネットの spam 活動を抑制する試みがなされているが、依然として greylisting や S25R などによるボットネットを排除するための spam 対策は有効であると考察する。

今回の調査でアドレスハーベスタと思われるホストを発見できたが、メールに関しては spam を送信してきたホストはなかった。

今後の課題として、観測期間を数カ月単位ではなく、数年単位で継続して調査し、アドレスハーベスタと spam 送信者の spam 活動の調査する。

## 参考文献

- [1] シマンテックインテリジェンス月次レポート 2013 年 1 月号, 入手先  
<[http://www.symantec.com/content/ja/jp/enterprise/white\\_papers/sr\\_wp\\_spam\\_report\\_1301.pdf](http://www.symantec.com/content/ja/jp/enterprise/white_papers/sr_wp_spam_report_1301.pdf)>(参照 2013-07-13)
- [2] シマンテック インターネットセキュリティ脅威レポート第 17 号, 2012 年 4 月, 入手先  
<[http://www.symantec.com/content/ja/jp/enterprise/white\\_papers/istr17\\_wp\\_201207.pdf](http://www.symantec.com/content/ja/jp/enterprise/white_papers/istr17_wp_201207.pdf)>(参照 2013-09-12)
- [3] D.Plohmann, E.Gerhards-Padilla and F.Leder, "Botnets:Measurement, Detection, Disinfection and Defence",European Union Agency for Network and Information Security, CIIP and Resilience, Jul.2011,入手先  
<[http://www.enisa.europa.eu/activities/Resilience-and-CIIP/critical-applications/botnets/botnets-measurement-detection-disinfection-and-defence/at\\_download/fullReport](http://www.enisa.europa.eu/activities/Resilience-and-CIIP/critical-applications/botnets/botnets-measurement-detection-disinfection-and-defence/at_download/fullReport)>(参照 2013-11-05)
- [4] 渡部稜太, 愛甲健二, "スパムメールの教科書", データハウス, 2006 年
- [5] O.Hohlfeld, T.Graf and F.Ciucu, "Longtime Behavior of Harvesting Spam Bots", proceedings of the 2012 ACM conference (ICM'12), pp.453-460,Nov.2012
- [6] W.N.Gansterer and M.Ilger,"Analyzing UCE/UBE Traffic",proceedings of the ninth international conference on Electronic commerce(ICEC'07), pp195-204,Aug.2007
- [7] B.Irwin, and N.Pilkington, "High Level Internet Scale Traffic Visualization Using Hilbert Curve Mapping", proceedings of the VizSEC 2007, Mathematics and Visualization, Springer Berlin Heidelberg, pp147-158,Oct.2007
- [8] R.Munroe, Map of the Internet,入手先,  
<<http://www.xkcd.com/195/>>(参照 2013-11-05)
- [9] 木谷友哉,中村嘉隆,"遅延状況を考慮した構造型 P2P オーバレイネットワーク構築法",情報処理学会研究報告,マルチメディア通信と分散処理,2009-DPS-139,Vol3,pp.1-6(2009 年 6 月)
- [10] C.Muelder and KL.Ma,"Rapid Graph Layout Using Space Filling Curves",Proceedings of the IEEE Transactions Visualization and Computer Graphics 2008,Vol.14,No.6,pp.1301-1308,Nov.2008
- [11] 阻止率 99%のスパム対策方式の研究報告,入手先  
<<http://www.gabacho-net.jp/anti-spam/anti-spam-system.html>>(参照 2013-09-10)
- [12] 松井一乃,小刀称知哉,金高一,池部実,吉田和幸,"大分大学宛のメール送信サーバの分布", 情報処理学会 第 61 回 CSEC・第 21 回 IOT 合同研究発表会, Vol.2013-IOT-21,No19,pp.1-7, May.2013
- [13] RBL.JP,入手先<<http://www.rbl.jp/>>(参照 2013-11-18)
- [14] 小刀称知哉, 天本大地, 池部実, 吉田和幸,"scan 攻撃検知システムを用いた被検知ホストの挙動についての調査", 電気関係学会九州支部第 65 回連合大会, pp.278-278, 2012 年 9 月
- [15] 長谷川明生,鈴木常彦,山口栄作,"たかが spam されど spam:spam を呼び込んだら",情報処理学会研究報告,分散システム/インターネット運用技術,2005-DMS37,pp75-78,May.2005