Vol.2010-CSEC-51 No.12
2010/12/10

ボットネットの国別マルウェア活動時間 なぜインドからの攻撃は日本時間で行われるか？

松尾 峻治[†1] 菊池 浩明[†1]
寺田 真敏[†2] 藤原 将志[†2]

ボットはマルウェアダウンロードサーバを国外に設置していることがあり，マルウェアのダウンロードはその国の時刻に応じてが行われている．本研究では，CCC Dataset の攻撃元データを解析し，各々の感染ホストの活動時間とダウンロード活動の相関を取り，各国での現地時間と日本で観測された時間との位相のずれがあることを見つけ，位相のずれのなぞを解いた．

Time-Country Analysis in downloading botnet malware Why are attacks from India synchronized with Japanese local time?

Shunji Matsuo,[†1] Hiroaki Kikuchi,[†1]
Masato Terada[†2] and Masashi Fujiwara[†2]

The botnet controls malware downloading servers distributed in to make abroad, download of malware is performed in the time of the target country. In the study, I analyze source addresses of attack data of CCC Dataset and found the correlation between active time for downloading and the time zone in which source address belong to. We propose a method to estimate The phase shift between the time observed and local time in countries.

†1 東海大学大学院工学研究科情報理工学専攻
Tokai University, Gradulate school of Engineering
259-1292 神奈川県平塚市北金目四丁目１番１号
ozuma,kikn@cs.dm.u-tokai.ac.jp
†2 日立製作所 Hitachi Incident Response Team(HIRT), Hitachi Ltd.
212-8567 神奈川県川崎市幸区鹿島田 890　日立システムプラザ新川崎
masato.terada.rd,masashi.fujiwara.zz@hitachi.com

1. はじめに

現在は 2 億を超える Web サイトが存在し，16 億人のネットユーザ人口がいると言われている．それに伴い，ユーザ層が多様化し，技術力の乏しいユーザが不正アクセスやスパムメール送信などの温床となっている．中でも，多量の一般的なネットユーザの PC に感染し，外部から操つって，不正アクセスや多量のスパムメール送信を行うネットワーク群は，ボットネットと呼ばれて新たな脅威となっている．

これらのボットネットの振る舞いには不明な点が多く，いくつかの解析が試みられている．2) ではボット C&C サーバのについて調査しており，3,600 台を超える C&C サーバ数の推移と位置を定期的に報告している．Stone-Gross らは，Torpig という実際のボットネットの C&C サーバを乗っ取って，能動的にボットネットの通信を観察している[4]．彼らによると，ボットネットはマルウェアをダウンロードした際にボット ID を一緒に送り，その ID によってどのボットネットに所属するか決定する．ボット ID と観測された IP 数の推移に一定の周期性があることを報告している．9) ではマルウェアダウンロード数とインターネット利用者のトラフィック量が密接な関係があると述べられているが，国地域の時差などを考慮した国にいてはロシア 1 件数のみであり，十分な解析が残されていた．

人の活動時間とマルウェアが活動している時間に関連性があり，国ごとにマルウェア活動時間が異なっていることに着目する．この差を活用することで，攻撃ホストの国の特定できるの可能性がある．本論文では，ハニーポッドデータセット CCC Dataset 2010[7] の国ごとのダウンロード活動とボットネットの関連性を基に，各国の活動時間とボットネットの関連から攻撃ホストが多い国を抽出することを目的とする．

2. ボットネット活動時間について

2.1 各国のマルウェア活動時間の観測

インプレスジャパン[5] のアンケート調査によるとインターネット利用時間帯は夜 22 時がピークであり，夜 20 時から 23 時の間が活発である．図 1 に，その結果を示す．

ボットネットの活動は感染しているホストのユーザの活動と深く関係してるはずである．そこで，CCC Dataset 2010 の攻撃元データにおける，ダウンロード数の推移を図 1 に重ねて示す．CCC Dataset 2010 の攻撃元データ[7] はダウンロード元が日本のダウンロード回数である．アンケート調査では百分率で推移を示しているのに対し，CCC Dataset 2010 ではダウンロード数で推移しているため，ダウンロード数は正規化して表示している．正規

Vol.2010-CSEC-51 No.12
2010/12/10

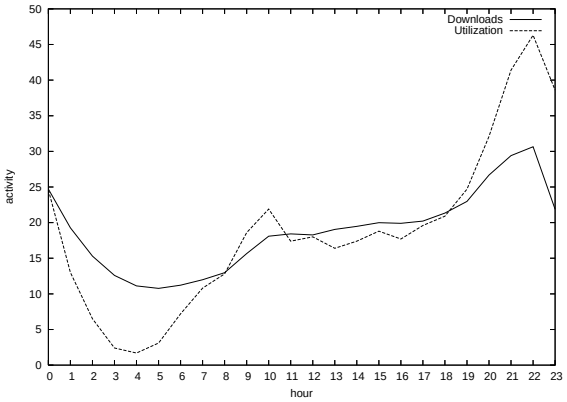

図 1 インターネット利用率時間帯[5] とマルウェアダウンロード時間帯

化した図 1 に示されるように，マルウェア活動時間と利用率の推移が酷似しており，インターネット利用時間帯にマルウェアも動機して活動していることが分かる．

2.2 国別のマルウェア活動時間

そこで，CCC Dataset におけるダウンロード行っているホストの IP アドレスを国別に分類し，その活動時間を調べる．ここで国の判別は，商用の GeoIP サービス GeoLite City[8] を用いた．2009 年 5 月からの一年間におけるマルウェアダウンロード数の統計情報を表 1 に示す．マルウェアをダウンロードしているホストは大きく 2 種類に分けられた．図 2 に示す日本とほぼ同じ時間で活動している J(Japan Time) 群と，図 3 の現地時間で活動していると思われる L(Local Time) 群である．

国 i における時刻 t から一時間の間に観測された総ダウンロード数を $d_i(t)$ とする．24 時間の総数を D_i，平均を $\bar{d}_i$ である．すなわち，$D_i = \sum_{i=0}^{n} d_i(t)$，$\bar{d}_i = D_i/24$ となる．

各国でのマルウェアダウンロード数の差が激しいため，次の式のように正規化した．

$$d'_i(t) = \frac{d_i(t)}{D_i} \tag{1}$$

図 2，3 は，こうして求めた時刻 t における正規化ダウンロード数 $d'_i(t)$ である．国 i への時差 p をシフトして現地時間へ変換したダウンロード数 $d'_i(t-p)$ を図 4 に示す．時差を考慮した場合，図 4 の様になり，日本とほぼ同じ活動が行われいることがわかる．

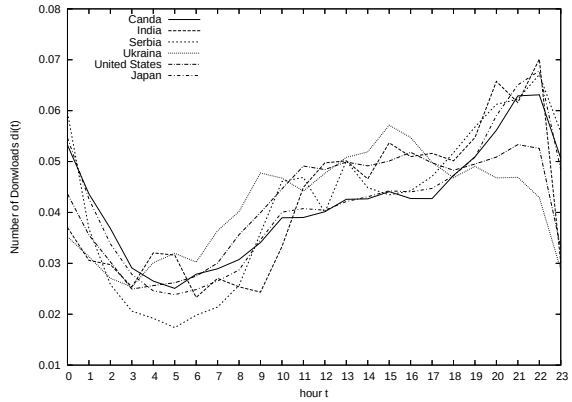

図 2 時間当りのダウンロード数の変化 (日本と活動時間が同一の国=J 群)

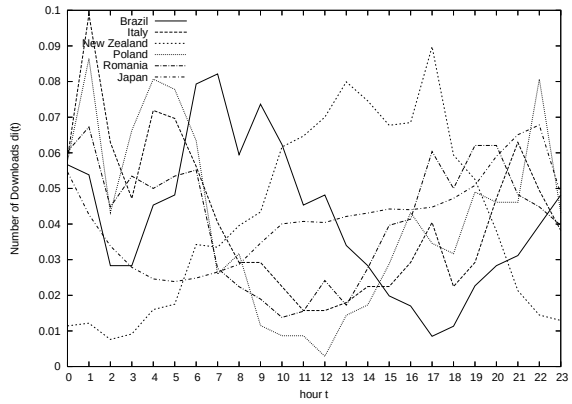

図 3 時間当りのダウンロード数の変化 (t，日本と活動時間が異なる国=L 群)

2.3 ダウンロード活動の位相

マルウェアダウンロード活動時間に時差があることをより正確に示すため，次の位相同定方式を考える．

国 i と国 j のダウンロード数の類似の度合いを次の相関係数 $S_{i,j}$ で定める．

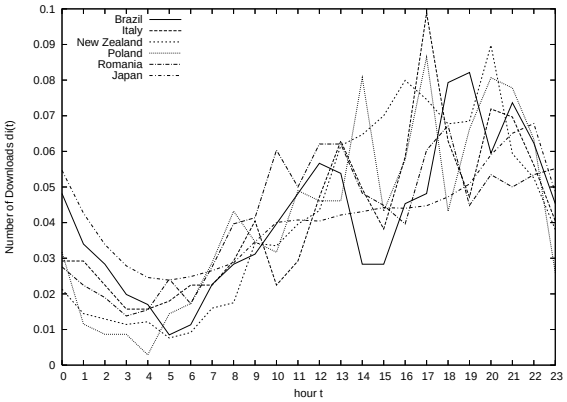

図 4 時間当りのダウンロード数の変化 (L 群，時差を考慮して，現地時間について集計)

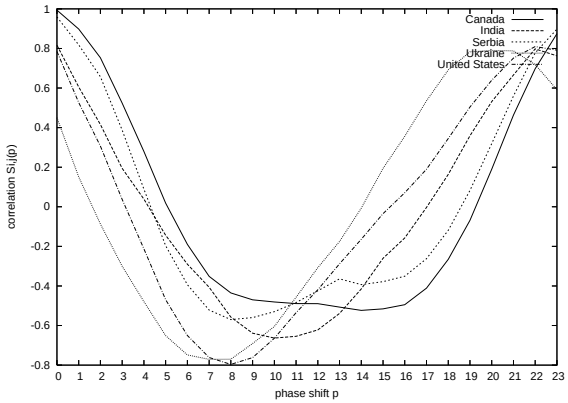

図 5 時間位相 p について相関係数の推移 (日本時間 J 群)

表 1 国別マルウェアダウンロード数の統計情報

群	国名 i	総ユニーク IP 数	DownLoad 回数	日本との時差
	カナダ	256	293693	15
J	アメリカ	1145	127052	16
	ウクライナ	180	51560	7
群	セルビア	7	4948	8
	インド	1212	4806	3
	ブラジル	345	353	12
L	イタリア	433	445	8
	ニュージーランド	282	1314	-3
群	ポーランド	338	347	8
	ルーマニア	537	580	7

$$S_{i,j} = \frac{\sum_{t=0}^{23}(d_i(t)-\bar{d_i})(d_j(t)-\bar{d_j})}{\sqrt{\sum_{t=0}^{23}(d_i(t)-\bar{d_i})^2}\sqrt{\sum_{t=0}^{23}(d_j(t)-\bar{d_j})^2}} \quad (2)$$

結果を表 2 に示す．表 2 より，活動時間が現地時間とずれている国と活動時間が現地時間である国との差がはっきりと別れた．

この位相 p を機械的に求めよう．時間差 p をずらした相関係数

$$S_{i,j}(p) = \frac{\sum_{t=0}^{23}(d_i(t-p)-\bar{d_i})(d_j(t-p)-\bar{d_j})}{\sqrt{\sum_{t=0}^{23}(d_i(t-p)-\bar{d_i})^2}\sqrt{\sum_{t=0}^{23}(d_j(t-p)-\bar{d_j})^2}} \quad (3)$$

を定める．$p = 0, \ldots, 23$ について求めた相関係数 $S_{i,j}(p)$ の変化を図 5(J 群)，6(L 群) に示す．

明らかに，相関係数に周期があり，その極大値 $S_{i,j}(p^*)$ を取る p^* を定めることができる．時差を考慮した時の $S_{i,j}(p')$，極値 $S_{i,j}(p^*)$ とその時の p^* それぞれの結果を表 3 に示す．表 3 では L 群は日本の活動時間に対して正の相関があり，J 群は時差 p' 遅らせて高い相関係数を示している．L 群のほとんどの国の p^* は，その国 i の時差との差がに収まっている．なお，表 3 の時差は 12) による．

3. なぜ，日本でダウンロードが起きたか？

3.1 仮　　説

図 4 の L 群の国は各国のローカル時間で活動していたのに対し，図 2 の J 群では日本時間にが近かった．マルウェアがダウンロードされるにはいくつかのプロセスが必要である．図 7 にて，攻撃ホスト A とマルウェアダウンロードサーバ B，ハニーポット C3 台から成る感染するまでのプロセスを示す．A は C に脆弱性を狙った攻撃を行い，マルウェアを B からダウンロードさせ，ハニーポットに感染させる．攻撃の起点となるのは攻撃ホストであ

Vol.2010-CSEC-51 No.12
2010/12/10

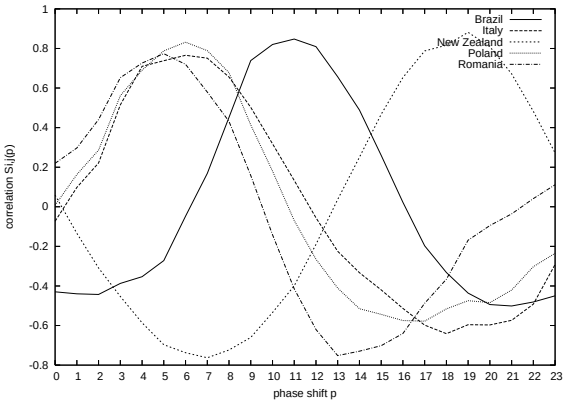

図 6 時間位相 p について相関係数の推移 (現地時間 L 群)

表 2 国 i と日本 j の活動時間の相関係数

	国名	相関係数 $S_{i,j}$
	カナダ	-0.49
J	インド	0.038
	セルビア	-0.570
群	ウクライナ	-0.771
	アメリカ	0.190
	ブラジル	0.809
L	イタリア	0.655
	ニュージーランド	0.671
群	ポーランド	0.677
	ルーマニア	0.579

り，マルウェア本体が実行されるのはダウンロードサーバに依存している．したがって，A，B，C，が日本，J 群，L 群の 3 通りの場合があり，表 4 で示す 7 パターンが考えられる．

今攻撃元データの観測より，1 と 2a のパターンが日本時間に属することが分かっている．しかし，その原因が A によるのか B によるのか分からない．すなわち，次の二つの仮説が考えられる．

仮説 1(攻撃ホスト説) A が現地時間に合わせて活動しているために，ダウンロードが現地時間になる (3c)．

仮説 2(ダウンロードサーバ説) B が現地時間で動いているので，B が停止中に A からの

表 3 時差と最適な位相の相関係数について

	国名	日本との時差 p'	最適な位相 p^*	$S_{i,j}(t)$	$S_{i,j}(t-p')$	$S_{i,j}(t-p^*)$
	カナダ	15	0	0.993	-0.494	0.993
J	アメリカ	16	0	0.816	0.038	0.816
	ウクライナ	7	0	0.959	-0.57	0.959
群	セルビア	8	20	0.450	-0.771	0.789
	インド	3	22	0.785	0.19	0.809
	ブラジル	12	11	-0.429	0.809	0.847
L	イタリア	8	6	-0.0731	0.655	0.765
	ニュージーランド	-3	19	0.058	0.671	0.881
群	ポーランド	8	6	0.008	0.677	0.831
	ルーマニア	7	5	0.219	0.579	0.773

表 4 攻撃パターン

	攻撃ホスト A	ハニーポット C	DL サーバ B	活動時間
1	*	日本	日本	日本時間
2a	日本	日本	J 群	
2b	J 群	日本	J 群	日本時間
2c	L 群	日本	J 群	
3a	日本	日本	L 群	
3b	J 群	日本	L 群	現地時間
3c	L 群	日本	L 群	
CCC Dataset	攻撃通信データ	-	攻撃元データ	

ダウンロードコマンドが失敗して，結局観測時間帯が現地時間となる．

まず，前提として，C は 24 時間休みなく運用されているとする．従って，C が位相の原因とは考えにくい．次に，仮説 2 を仮定するならば，A からダウンロードを命じられた C が通信を試みて，TCP のハンドシェークに失敗 (Syn はあるが，Ack がない) をするはずである．一方，仮説 1 が正しいならば，A は現地時間に合わせてオンラインになる．そこで，次の実験を行い，これら仮説の検証を行った．

3.2 実験目的

時刻に差が出たのは，攻撃ホストの国の時間に影響しているのか (仮説 1)，マルウェアダウンロードサーバが置いてある国の時間に影響しているのか (仮説 2) のどちらかであるか明らかにすることを目的とする．

3.3 実験方法

CCC Dataset 2010 の攻撃通信データ[7] を用いて検証する．このデータはハニーポットで収集したパケットキャプチャデータであり，定期的にシステムリブートしており，これ

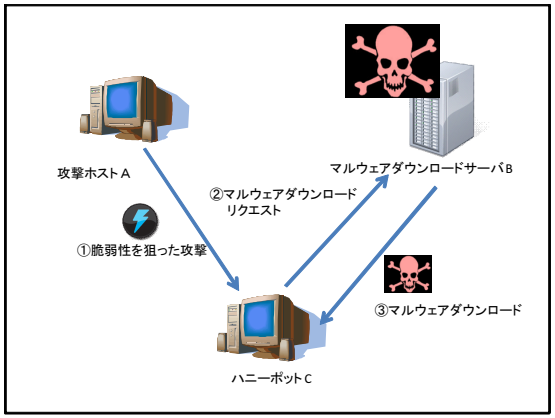

図 7 マルウェア感染フロー

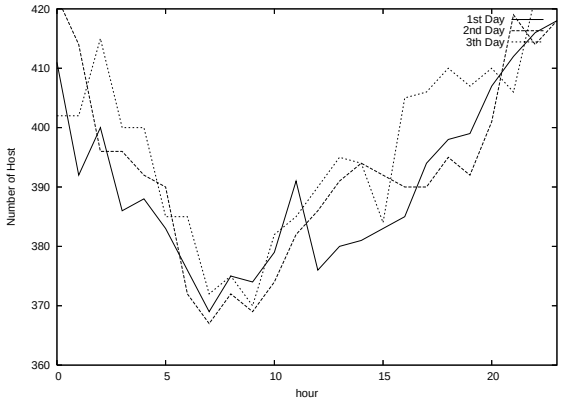

図 8 オンラインの総攻撃ホスト数

を 1 つのマルウェア感染の通信データとする．

攻撃ホストは，TCP のハンドシェークを引き起こすホストとする*1．TCP のハンドシェークは，次の 3 つからなる．

(1) syn
(2) ack
(3) syn-ack

この内，syn と ack パケットは送信されたのに，syn/ack が送信されないものを syn-ack の失敗と呼ぶ．syn-ack が失敗したホストは攻撃が失敗したものとし，マルウェアがダウンロードされない．

実験 1 L,J の両群について syn-ack の通信が失敗している割合を調べる．

実験 2 2010 年 9 月 1 日〜4 日かけて，各攻撃ホストに毎時 ping(ICMP Echo パケット) を行い，オンラインの攻撃ホストの割合を調べる．

この二つを検証する．

3.4 実験結果

実験 1：syn-ack の通信が失敗している割合

攻撃通信データから攻撃ホストは 1164 件あり，その中で，syn-ack 通信が失敗している IP アドレスは 39 件あった．syn-ack 通信が失敗しているデータの一部を表 5 に示す．表 5 の FQDN を見ると，ほとんどがプロバイダーがつけた名前となっており，独自のドメインではないこと示している．このことから，攻撃ホストは一般ホストに混じっていることが分かる．ここでスロットとは観測単位の番号を示す．攻撃ホストで攻撃に失敗しているスロットは 45 件あった．

実験 2：ping によるオンライン率

全攻撃ホスト 1164 件に対し，ping を行い求めたオンラインの攻撃ホスト割合の結果を図 8 に示す．L,J 群の代表的な国についてのオンライン攻撃ホスト数の変化を図 9，10，11，12 に示す．L 群のイタリアとブラジルでは活動時間が変動していたのに対し，J 群のアメリカは一定であった．

3.5 考　　察

図 9，12 を見ると，日本，アメリカは 24 時間動いていることが分かる．それに対し，ブラジル，イタリアは図 10，11 に示されるように，停止時間がある．このことから，アメリカの攻撃ホストはほぼ全て 24 時間稼働しているのではないかと考えられる．つまり，日本の攻撃ホストがダウンロードを引き起こしている (表 4 の 2a)．

一方，イタリアやブラジルでは現地時間がマルウェア活動時間に一致しており，攻撃の起点となるホストが現地にある (表 4 の 3c)．

以上より，マルウェアの活動は攻撃ホストに依存している，すなわち，仮説 1 が正しいと

*1 push の時のダウンロードサーバも含まれてしまうが無視する，

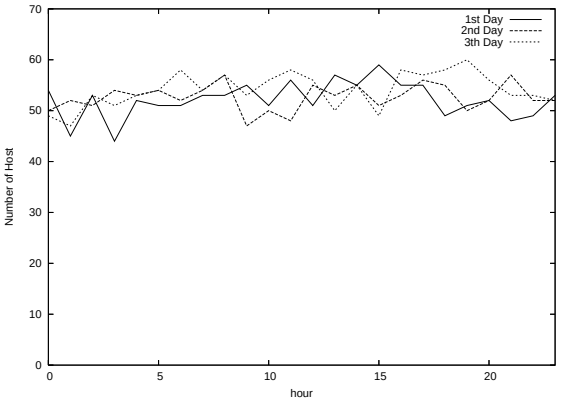

図 9 オンラインの日本攻撃ホスト数 (日本)

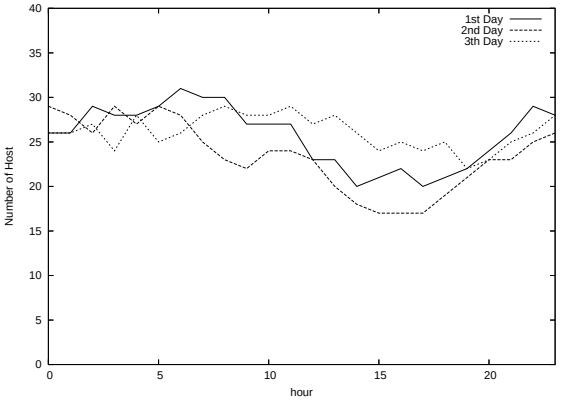

図 10 オンラインのブラジル攻撃ホスト数 (ブラジル)

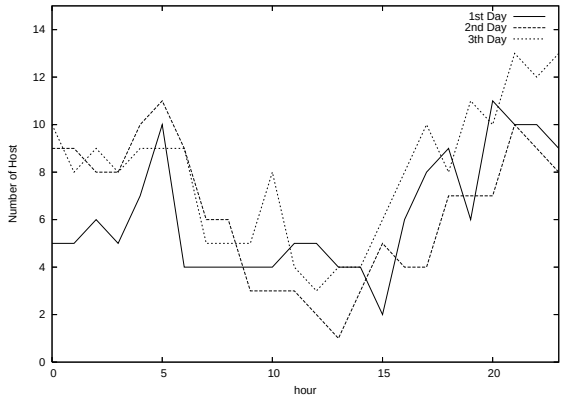

図 11 オンラインのイタリア攻撃ホスト数 (イタリア)

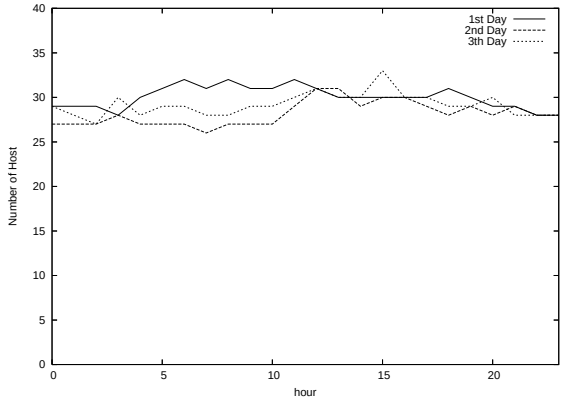

図 12 オンラインのアメリカ攻撃ホスト数 (アメリカ)

結論づける．

4. ま と め

CCC Dataset 2010[7] を用いて，マルウェアの活動時間について調査を行った．マルウェアの活動は一般ユーザの活動時間に深いかかわりがあり，攻撃ホストの属するタイムゾーンに依存している．攻撃ホストが多いほど，現地時刻に近くなるので，各国活動時間との相関を取ることにより，観測された不正ホストの中で攻撃ホストの割合多いかどうかを判定できる．今後，攻撃ホストとダウンロードサーバとの活動時間について調べ，不正ホストが攻撃ホストとダウンロードサーバどちらかなのかを特定できるよう検討していきたいと考えている．

表 5 syn-ack が失敗した IP の詳細の一部

スロット	IP	FQDN	Country-Code	国名
1	67.43.236.68	Nothing	LB	レバノン
3	123.205.232.146	xxx.dynamic.seed.net.tw.	TW	台湾
6	69.64.147.243	ash.parking.local.	US	アメリカ
7	69.64.147.243	ash.parking.local.	US	アメリカ
10	122.18.181.213	xxx.tokyo.ocn.ne.jp.	JP	日本
11	41.97.253.199	Nothing	DZ	アルジェリア
16	122.18.181.213	xxx.tokyo.ocn.ne.jp.	JP	日本
27	189.84.197.171	xxx.projesom.com.br.	BR	ブラジル
27	5.160.60.120	Nothing	?	?
29	130.22.1.10	Nothing	US	アメリカ
57	124.86.121.64	xxx.kanagawa.ocn.ne.jp.	JP	日本
58	124.86.121.64	xxx.kanagawa.ocn.ne.jp.	JP	日本
80	218.232.43.140	Nothing	KR	韓国
89	66.2.3.7	xxx.algx.net.	US	アメリカ
89	77.28.192.12	Nothing	MK	マセドニア
89	89.106.98.77	xxx.optilinkbg.com.	BG	ブルガリア
118	39.99.169.152	Nothing	US	アメリカ

表 6 各群における上位ユニーク IP 数

	攻撃ホスト A			ダウンロードサーバ B		
	順位	国	ユニーク IP 数	順位	国	ユニーク IP 数
	1	JP	228	1	JP	49
J	4	US	89	3	US	9
	8	IN	34	9	UA	3
群	17	CA	11	13	IN	2
	22	UA	9	19	CA	1
	6	BR	51	6	RO	6
L	9	RO	29	10	PL	3
群	10	IT	28	13	BR	2
	12	PL	20	19	IT	1

参 考 文 献

1) ITmedia, “米 Yahoo!が設立 15 周年”, http://www.itmedia.co.jp/news/articles/1003/03/news016.html, Mar 2010.
2) Shadowserver, “Botnet Charts”, http://www.shadowserver.org/wiki/pmwiki.php/Stats/BotnetCharts, Mar 2010.
3) 野津誠, “ボットネットに関与した ISP の接続停止で世界のスパムが 38 %減少”, http://internet.watch.impress.co.jp/docs/news/20090903_312659.html, 2009.
4) Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Chris Kruegel, and Giovanni Vigna, “Your Botnet is My Botnet: Analysis of a Botnet Takeover” in Proceedings of the ACM CCS, Chicago, IL, November 2009.
5) 柴谷 大輔, “実態調査でみる個人のインターネット利用動向”, “インターネット白書 2010”, pp.180-193. インプレスジャパン, 2010.
6) 大類他, “分散ハニーポット観測からのダウンロードサーバ間のアソシエーションルール抽出”, 情報処理学会, コンピュータセキュリティシンポジウム, CSS2009, 2009.
7) 畑田 充弘, 中津 留勇, 秋山 満昭, 三輪 信介, “マルウェア対策のための研究用データセット ～MWS 2010 Datasets～”, MWS 2010, 2010
8) MaxMind. GeoIP,http://www.maxmind.com/app/ip-location, 2008.
9) 金井 瑛, “マルウェアの転送ログを利用した地域毎のボット活動分析”, CSS 2010, 2010
10) 水谷 正慶, 武田 圭史, 村井 純, ”Web 感染型悪性プログラムの分析と検知手法の提案 “, 電子情報通信学会論文誌. B, pp.1631-1642, 2009.
11) 竹森 敬祐, 三宅 優, 田中 俊昭. “ネットワーク間プロファイル比較による攻撃異常検知”, 情報処理学会研究報告, マルチメディア通信と分散処理研究会報告, pp.87-92, 2005
12) “世界各国の時差一覧 時間の計算の仕方”, http://www.travelerscafe.jpn.org/world_time.html.