CCC DATAset 2009 によるマルウェア配布元の可視化

松木 隆宏 †　　新井 悠 †

† 株式会社ラック サイバーリスク総合研究所
105-0001 東京都港区虎ノ門 4-1-17　神谷町プライムプレイス 3F

あらまし　サイバークリーンセンターによる注意喚起の継続によって国内のボット感染端末は減少傾向にある．しかし，未知のマルウェアの配布元については大半が海外に存在すると報告されている．本論文では，CCC DATAset 2009 の攻撃元データを用いて，未知のマルウェアの配布元の地理的分布を可視化し，実態を明確にする．ダウンローダ型のマルウェアの増加によって複数のマルウェアに感染することも増えている．ダウンローダによる感染拡大を防止する方法の 1 つとして，攻撃通信データを用いたマルウェア配布元 URL リストの作成を検討する．

Visualization of the Malware distribution by CCC DATAset 2009

Takahiro Matsuki†　　Yuu Arai†

†Risk Research Institute of Cyber Space, Little eArth Corporation Co., Ltd.
4-1-17 Toranomon Minato-Ku Tokyo 105-0001 Japan

takahiro.matsuki@lac.co.jp, y.arai@lac.co.jp

Abstract　The domestic BOTs infection decreases by continuation of the attention by the Cyber Clean Center. However, most undetectable malware are distributed from foreign countries. In this paper, promote the grasp of the actual situation by making the geographical distribution of the malware distribution visible with CCC DATAset 2009. Also downloader and infection to plural malware are increasing recently. Because make a URL list of the distribution of the malware and examine a method to prevent infection expansion with the downloader.

1　はじめに

最近 Web の閲覧を介して感染するマルウェア（以下，Web ベースマルウェア）や USB メモリなどの記憶媒体を介して感染するマルウェアなど新たな経路で感染を広げるマルウェアが流行しているが，サイバークリーンセンターによるとボットについては注意喚起の継続によって国内の感染端末が減少傾向にあり，ハニーポットによるマルウェア検体の収集数も減少している．しかし，未知のマルウェアの配布元については大半が海外に存在すると報告されており，今後も未知のマルウェアの配布は継続されると思われる．

全体的にダウンローダ型のマルウェアが増加し，それによって多数の未知のマルウェアへの感染，複数のマルウェアが連動する例もある．著者は，マルウェア同士の関連性に着目し，連鎖感染を可視化する研究を MWS 2008 において発表した [1]．しかし，連鎖感染のパターンを増加させるファイル感染型検体の存在や長い分析対象時間によってグラフのノードの増加し，結果として作成した連鎖感染マップは非常に複

雑になった．

本論文では，可視化の対象を検体から検体の配布元に変え，CCC DATAset 2009 の攻撃元データから未知検体の配布元の地理的分布を一見して把握できるようにする．また，攻撃通信データを用いてダウンローダによる感染拡大を防止する URL ブラックリストの作成を検討する．

2　未知検体の初期配布元の分布

サイバークリーンセンター活動実績では，新たな未知検体や注目すべき検体の配布元の国や地域が記載されているが，本稿では，一般利用者やボット感染者がより理解しやすいように未知のマルウェアの配布元である国や地域を可視化を行なう．これによって未知のマルウェアが世界各地から発生していることが改めて認識できる．

CCC DATAset 2009 の攻撃元データの IP アドレスについて Geolocation データベースである GeoLite City [3] を用いて緯度経度を取得し，Google Maps API [4] を用いて世界地図上にマッピングを行なった[1]．マルウェアの感染端末やスパムメールの送信元の地理情報の可視化は，F-Secure なども行っている [6].

2.1　未知検体を多数配布した IP アドレス

攻撃元データによると 2008 年 11 月から 2009 年 4 月の間で合計 1,494 種類のファイルハッシュが異なる未知検体が収集されている．本稿では，検体の種類はファイルハッシュによって識別した．各未知検体について最初に配布を行った IP アドレスを調査した結果，未知検体を最初に配布した IP アドレス（未知検体の初期配布元）はハニーポット自身を除いて 357 件であった．これらの IP アドレスについて配布した未知検体の種類の数の分布を調査した（表 1）．

表 1: 未知検体の配布数と IP アドレス

未知検体の配布数	IP アドレスの数
1	274
2〜5	45
6〜10	17
11〜25	11
26〜50	5
51〜100	2
101〜200	3

全体の 76.8 %にあたる 274 の IP アドレスは 1 種の未知検体のみ配布しており，10 種以下を配布した IP アドレスまでで 94% を占めている．これらの IP アドレスはボットに感染し，ボットネットによって検体の拡散に利用されたものと推測できる．一方，11 種以上の未知検体を配布していた IP アドレスは 21 件存在した．さらに 51 種類以上の未知検体を配布していた IP アドレスは 5 件まで絞り込むことができ，これらは通常のボットに感染した IP アドレスではない可能性が考えられる．未知検体を最初に配布した IP アドレス の位置情報を取得し，世界地図にマッピングしたものが図 1 である．配布した未知検体の種類の数によってマーカーの色を変更した．1 種のみ配布した IP アドレスは黄色，11〜50 種は赤色，51 種以上配布した IP アドレスは紫色のマーカーとした．

可視化によって未知検体の最初の配布元は世界中に分散していることと，特に多種類の未知検体を配布している IP アドレスが 福岡（日本），北京（中国），リガ（ラトビア），メイデンヘッド（イギリス）とヒューストン（アメリカ）に存在することが確認できる．この 5 つの IP アドレスから最初に配布された未知検体は合計で 577 種あり，全体の 38.6 % である．

2.2　多数の配布元から配布された検体

次に攻撃元データに含まれる 1,494 種の未知検体について，配布元の IP アドレスの数を調査した（表 2）．全体の 79.3 % は，1 つの IP アドレスから配布されていた．

[1]商標等に関する表示
本論文に記載している商品，サービス等の名称は，それぞれの所有者の商標または登録商標です.

図 1: 未知検体の初期配布元の分布

表 2: 未知検体の配布数と IP アドレス

配布元 IP アドレスの数	未知検体の種類
1	1,185
2〜50	268
51〜100	21
101〜500	13
501〜1,000	3
1,001〜5,000	4

多数の IP アドレスから配布された検体ついて詳しく調査を行なった結果，TSPY_KOLABC.CH という名称がつけられた検体が特徴的であった．TSPY_KOLABC.CH はファイルハッシュの異なる検体が 11 存在し，未知検体の中で最も多数の 2,084 の IP アドレスから配布されていた．これを A とする．923 の IP アドレスから配布されていたものもあった．こちらを B とする．2 つの TSPY_KOLABC.CH の配布 IP アドレスの推移を図 2 に示す．A，B ともに未知検体として最初に収集されたのは，2008 年 12 月 29 日であり，2009 年 1 月 7 日のログから TSPY_KOLABC.CH という名称がつけられている．A は 3 月上旬に配布元が消滅し，最後に配布されたのは，2009 年 3 月 4 日であったが，B は 2009 年 4 月 30 日（データセットの最終日）にも配布されていた．

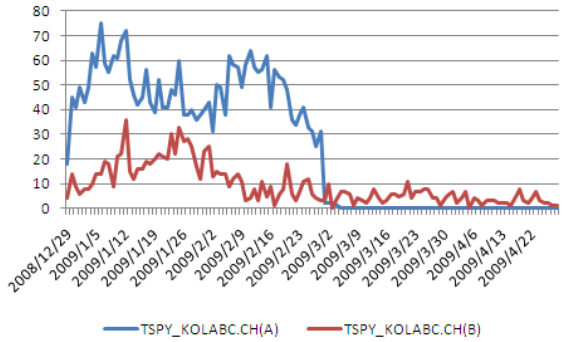

図 2: TSPY_KOLABC.CH の配布元の推移

トレンドマイクロによると，TSPY_KOLABC.CH は複数の検体と関連しており，BKDR_POEBOT.GN，WORM_SWTYMLAI.CD という検体を作成し，感染させるという．さらに Web サイトから別の検体をダウンロードし，感染させるという [7]．

なお，攻撃元データでは，BKDR_POEBOT.GN は 569，WORM_SWTYMLAI.CD は 1 つの IP アドレスからのみ配布されていた，

TSPY_KOLABC.CH と 関連する検体の配布元の位置情報，時刻情報から KML データを作成し，Google Earth [5] を用いてこれらの分布と時間的変化を可視化し，関連性を確認した．

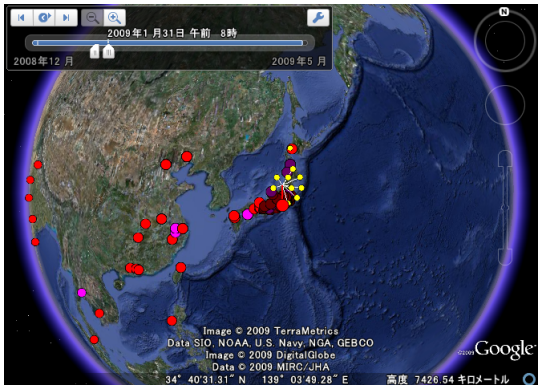

図 3: TSPY_KOLABC.CH 関連検体の配布元の分布と時間的変化

2.3 Web ベースマルウェアとの比較

多くの未知検体がポート 80 を用いて配布されていることから，流行している Web ベースマルウェアの配布元との関連性を調査を行なった．Web ベースマルウェアを調査するために開発しているクライアント型ハニーポットシステムを用いて 2009 年 8 月中に調査した約 600 件のマルウェア感染源サイトの IP アドレスの位置の可視化を行なった（図 4）．これまでの調査では，Web ベースマルウェアの配布元は中国に最も多く発見している．

CCC DATAset 2009 による未知検体の配布元と比較すると，多種の未知検体を配布している地域的には大きくずれていないが，CCC DATAset 2009 と Web ベースマルウェアの配布元に共通する IP アドレスは存在しなかった．

3 検体配布に使われるポート

CCC DATAset 2009 の攻撃元データ全件について検体の配布に使われたポートを調査した結果，ポート 80 を用いた配布は，全 894,517 件中 372,165 件（41.6 %）であった．CCC DATAset 2008 の攻撃元データでは，全 2,942,221 件中 1,157,101 件（39.3 %）で，ほぼ同じ割合であり，著しい変化はないと考えられる．

著者は CCC DATAset 2008 を用いた連鎖感染の可視化によって「連鎖感染の半数以上はポート 80 が用いられ，特に未知検体との連鎖はポート 80 が用いられる割合が高い」という傾向を確認した．図 1 に可視化した CCC DATAset 2009 の未知検体の配布で用いられたポートについて調査し，同じ傾向を再確認した．

攻撃元データに記録された未知検体 1,494 種類のうちポート 80 を用いて配布された検体は 1,066 件であった．これは未知検体全体の 71.4 % である．

図 5 は未知検体の配布ポートの分布を示したグラフであり，X 軸は未知検体が収集された順，Y 軸はポートである．6ヶ月の間 ポート 80 が継続して多用されているが，200 番目付近（日時では 2008 年 12 月 6 日〜 9 日）では，図 1 で示した多種の未知検体を配布した IP アドレスの 1 つでポート 8889 による未知検体配布が増加していた．2008 年 12 月度のサイバークリーンセンター活動実績で述べられている BKDR_PROTUX.AHB と考えられる [2]．また，600〜1000 番目の間にも 80 以外の特定ポートを用いて配布される未知検体が存在している．

未知検体の配布はポート 80 を用いて行われる割合が高く，ファイアウォールでのブロックされない可能性が高い．ポート 80 を用いた通信のプロトコルが HTTP であるかどうかは攻撃元データから判断することができないが，HTTP であった場合は通常の通信との区別が容易でないと考えられる．

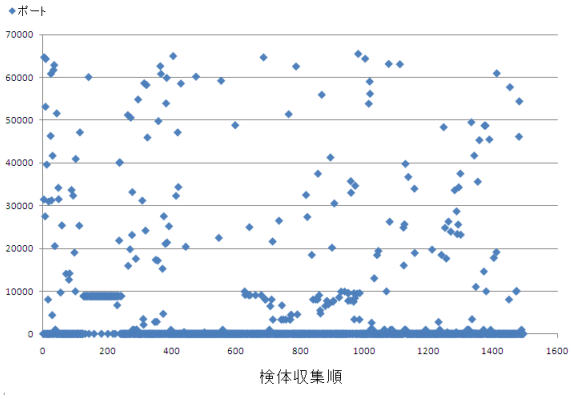

図 5: 未知検体の配布に使われたポート

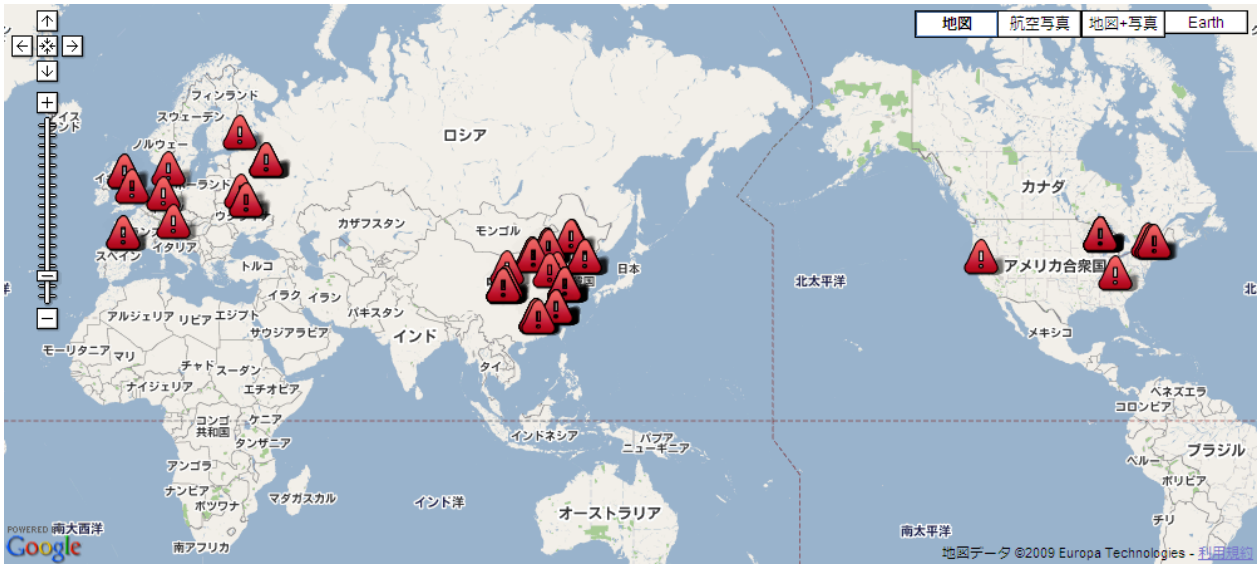

図 4: Web ベースマルウェアの配布元サイトの位置

4 URL ブラックリストの作成

検体の配布に使われるポート 80 の通信について，CCC DATAset 2009 の攻撃通信データで利用されているプロトコルの分析を行なった．プロトコルが HTTP の場合は，検体配布元 URL のブラックリストを作成し，それを既存の URL フィルタリング装置などに導入することでマルウェア感染の予防に利用可能と考える．攻撃通信データの解析によって検体配布元 URL のブラックリストの作成を試みた．

3 月 13 日と 3 月 14 日それぞれの攻撃通信データを解析した結果，3 月 13 日は，宛先ポート 80 の通信 560 件中 238 件（42.5 %），同 14 日は 464 件中 154 件（33.2 %）が HTTP であると確認できた．セッションから GET リクエストとユニークな URL を抽出した結果を表 3 に示す．

解析の結果判明した特徴的な URL として，2.2 節で述べた TSPY_KOLABC.CH にダウンロードされる検体の URL が存在した．その他に ICQ，AIM，Firefox などの正規ソフトウェアの URL に偽装したマルウェアのダウンロードが存在した（表 4）．調査の結果，PE_BOBAX.AF-O という検体によるものであった [8]．

各ハニーポットが検体取得のためにアクセス

表 3: 攻撃通信データに含まれる HTTP 通信

日付	ハニーポット	GET	URL
3/13	honeypot1	102	10
	honeypot2	130	13
3/14	honeypot1	64	8
	honeypot2	82	10

した URL は 1 日に 10 前後であり，2 日間のデータから抽出できたユニークな URL は 22 件であった．分析対象のハニーポットを増加したとしても URL ブラックリストの作成とフィルタリングが実現可能な数だと考える．

4.1 セキュリティ情報 DB との照合

Web ベースマルウェアや Web に関するセキュリティ脅威の増加に伴い，セキュリティ組織やウイルス対策ベンダ各社は，脅威の発信元に関する情報をあらゆる情報源から収集，蓄積したデータベースを構築し，対策に利用している．具体的な形として，IP アドレスや URL のブラックリスト，サイトの安全性をチェックするサービスやブラウザアドオンなどがある．

複数のブラックリストへの登録状況を確認することのできるサービスやいくつかのベンダの

表 4: 正規ソフトウェアの URL に偽装したマルウェア

URL
http://205.188.226.xx/aim/win95/Install_AIM.exe
http://209.170.96.xx/pub/ICQ_Win95_98_NT4/ICQ_4/Lite_Edition/icq4_setup.exe
http://193.74.22.xxx/pub/mozilla.org/firefox/releases/1.0/win32/en-US/Firefox%20Setup%201.0.exe

安全性チェックツールを用いて各種セキュリティ情報データベースに今回データセットから抽出した TSPY_KOLABC.CH にダウンロードされる検体の URL の情報が存在するか確認した．その結果，Stopbadware.org，surbl.org，Norton Safe Web，Trend Micro Smart Protection Network に登録されていた．

5 まとめ

本稿では，CCC DATAset 2009 の攻撃元データを用いて，多くの未知検体を配布している配布元，多数の配布元から配布された未知検体について地理的位置と時間的変化を可視化し，検体拡散の実態を把握した．

また，攻撃通信データを用いて検体配布元の URL リストを作成し，ダウンローダによる感染拡大を防止する方法を検討した．

最後に今後も継続して研究用にデータセットが提供されること，より最新の情報がデータセットとして提供されることを期待したい．

謝辞

本研究は，情報通信研究機構（NICT）「インシデント分析の広域化・高速化技術に関する研究開発」 の支援を受け実施しています．また，財団法人 日本データ通信協会 Telecom-ISAC Japan ならびにサイバークリーンセンターの支援を受け実施しています．本研究を進めるにあたり，有益な助言と協力を頂いた関係者各位に深く感謝致します．

参考文献

[1] 松木 隆宏：時系列分析による連鎖感染の可視化と検体種別の推測 (2008). 情報処理学会シンポジウムシリーズ Vol.2008. No.8

[2] 2008 年 12 月度　サイバークリーンセンター活動実績
https://www.ccc.go.jp/report/200812/0812monthly.html

[3] MaxMind - GeoLite City
http://www.maxmind.com/app/geolitecity

[4] Google Maps API
http://code.google.com/intl/ja/apis/maps

[5] Google Earth API - Google Code
http://code.google.com/intl/ja/apis/earth

[6] F-Secure Weblog : News from the Lab
http://www.f-secure.com/weblog/archives/00001606.html

[7] Trend Micto TSPY_KOLABC.CH
http://www.trendmicro.co.jp/vinfo/grayware/ve_graywareDetails.asp?GNAME=TSPY%5FKOLABC%2ECH&VSect=Td

[8] Avira Worm/Bobic.K.3
http://www.avira.com/jp/threats/section/fulldetails/id_vir/1189/worm_bobic.k.3.html

[9] StopBadware.org
http://www.stopbadware.org

[10] surbl.org
http://www.surbl.org

[11] Norton Safe Web
http://safeweb.norton.com