地理的可視化を用いたマルウェアの統合解析

金子 博一 †

† ラックホールディングス株式会社 サイバーセキュリティ研究所
102-0093 東京都千代田区平河町 2-16-1 平河町森タワー
hirokazu.kaneko@lachd.co.jp

あらまし サイバー攻撃は多岐に渡って行われており, 社会インフラを脅かしている. 特にマルウェアは日々洗練され, 被害者に気づかれることなく価値のある情報を盗み出す事が攻撃者の主目的になっているといえる. そのためマルウェアによる攻撃は, どのような種類のマルウェアがどのような通信活動を行っているか直感的に判断しづらい傾向にある.

本研究では CCC DATAset 2011 の内, 攻撃通信データと攻撃元データを用いて, 各マルウェアの通信地点の可視化を行った. この際, 各マルウェアの通信元の情報から攻撃目的に着目し, マルウェアの統合的解析の為の支援システムを実装した.

Integrate analysis malware from geographic visualization

Hirokazu Kaneko†

†LAC Holdings, Inc. Cyber Security Laboratory
Hirakawacho Mori Tower, 2-16-1 Hirakawacho, Chiyoda-ku, Tokyo 102-0093, JAPAN
hirokazu.kaneko@lachd.co.jp

Abstract Cyber attacks use several methods and threaten social inflastracture. Especially, malware is highly sophisticated to steal more valuable information, and the victimes are not aware of it's infection. Thus, it is difficult to detect and distinct malware infections by physical sence.

In this paper, we geographical visualize malware's attack point with CCC DATAset 2011's Attack-Connection Data and Attack-Source Data, and creates support system for integrate analysis malware to obvious attacker's purpose.

1 背景

サイバー攻撃は益々多様化しており, 社会インフラを脅かす重大な問題となってきている. 特にマルウェアによる攻撃は目的に応じて様々な手法を用いて攻撃が行われている. ウラン濃縮設備の制御システムを狙ったと考えられる Stuxnet や, Android 端末の情報を狙った Droid Dream などが例に上げられる. マルウェアはこれまで攻撃対象としていなかった SCADA システムやスマートフォンといった新たなプラットフォームを対象としている.

また, 一般のコンピュータ利用者はマルウェア検知・認識がしづらいといえる. マルウェアが既存のプロセスになりすまして感染活動を行い, コンピュータ利用者に気づかれないように情報を盗み出すなど秘密裏に攻撃を行う手法がとられてきた. そのため各種ウイルス対策ソフトベンダー等が対応に追われているが, 膨大なデータに対して適切なルールを適応しなければ

異常を検出できない状態となっている.

1.1 可視化

サイバー攻撃は一般のコンピュータ利用社に認識しづらく, 攻撃されている事を実感できない場合があるため, 特に認識が難しいものについては様々な可視化が行われてきた. ログ解析支援システムとしてはログデータを統計的に解析して時間毎に可視化を行った「見えログ」[1], 地図上に攻撃元地点を描画する「WASABI」「Gumblar の地理的可視化」[2]「DShield」[4] などが行われてきた.

1.1.1 地理的可視化

インターネット上のサイバー攻撃を可視化する手法の一つとして, 地理的可視化手法が用いられている. 地図は人間にとってより直感的に地点を表現する表現方法の一つであり, 例えば情報通信機構 Nict のシステムである「Nicter」[3] が挙げられる.

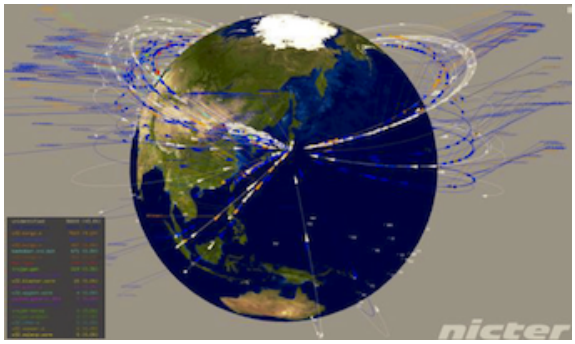

図 1: Nicter の例

図 1 は Nicter の例である. このシステムは主にダークネットと呼ばれる, 未使用の IP アドレスから日本に対する攻撃を可視化している. 感染していると思われるマルウェアの名前を色分けして表示しており, 図 1 の例では, 特定のマルウェアの活動が活発であることがわかる. また利用しているポート番号が大きい番号であればあるほど高い軌道を描くため, 同じマルウェアでも挙動が違う場合は認識しやすい.

1.1.2 IP アドレスと地理情報

サイバー攻撃の攻撃元情報の特定には, IP アドレスと地理情報の紐付けが必須である. そこで, 地域認識技術 (Geolocation Service) を用いて IP アドレスから地理情報へ変換する. 日本国内ではサイバーエリアリサーチ株式会社の SurfPoint が存在し, 海外では MaxMind 社の GeoIP[5] といったサービスが提供されている.

しかし SurfPoint は日本国内における精度は高いが, 海外の地域認識はできない. そのため, 本研究では全世界の地域認識が可能な GeoIP を利用した.

1.1.3 ポイントマップ

図 2 は地理的可視化の一つとして挙げられるポイントマップの例である. ポイントマップは目印の大きさ, 色や形で他のポイントと違いを表現することができるため, 詳細な位置情報が必要な場合に効果を発揮する可視化手法であるといえる.

図 2: ポイントマップの例

1.1.4 プリズムマップ

別の地理的可視化手法の一つとしてプリズムマップが挙げられる. プリズムマップは一定の基準で境界を区切り, 該当部分を 3D ポリゴンで表現する手法である.

図 3 はプリズムマップの例であり, この場合は乳幼児の死亡率を表現している. 国等の地域

単位で概要を掴む場合に有効な表現手法であり，主に色と高さで特徴を表現している．

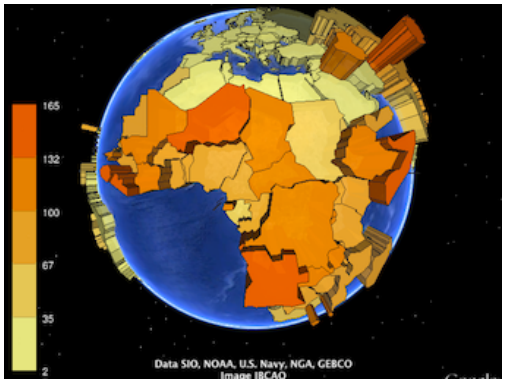

図 3: プリズムマップの例 - 乳幼児の死亡率

1.1.5　既存の研究の問題点

既存の研究ではマルウェアの特徴を複数または単一の通信データから可視化を行っている．しかし明示的に活用したログの差を設け，且つ同時に可視化しているものは少ない．

2　目的

本研究では各マルウェアによる通信とその他の通信を同時に解析するための統合解析支援システムを開発し，各マルウェアや攻撃の通信の特徴を得る．特にマルウェアの特徴的な挙動を人間が認識できる点について着目し，システムによる有効性を示す．

3　システム概要

主に可視化技術をベースに，マルウェアの統合解析を行うシステムを構築する．より様々な特徴を掴むため，複数の可視化手法を同時に行う．

3.1　対象となるデータ

使用するデータは CCC DATAset 2011 の内，攻撃通信データと攻撃元データとしている．ここではそれぞれのデータの特徴と実際に利用するデータについて述べる．

攻撃元データは下記データが収集されており，主にマルウェアの感染活動に関するデータを収集することができる．

(a) 攻撃通信が発生した時刻

(b) 通信元 IP アドレス／ポート番号

(c) 通信先 IP アドレス／ポート番号

(d) 通信プロトコル (TCP,UDP)

(e) SHA1 ハッシュ値

(f) マルウェア検知名

(g) 検知する通信を行ったプロセス名

攻撃通信データは pcap ファイルであり，実際の通信を取得している．しかし，通信はパケット単位よりもセッション単位で判断するため，このままではデータが膨大である． そのため今回はネットワークパケットを解析するツール Wireshark に含まれる,Tshark で通信概要を取得した．取得した概要情報は下記で構成される．

(I) パケット ID

(II) 取得日の 00 時 00 分 00 秒からの経過時間 (秒)

(III) 通信元 IP アドレス

(IV) 通信先 IP アドレス

(V) 通信プロトコル (HTTP,FTP 等アプリケーションプロトコル含む)

(VI) パケットサイズ

(VII) ペイロード

攻撃通信データはマルウェアが行った通信全般のデータを取得できる．しかし，攻撃通信データと攻撃元データはそれぞれ期間や観測対象が異なっている．

図 4 は CCC DATAset の特徴を示している．これにより，期間・観測対象は共に攻撃通信データの方が少ないことがいえる．また,HoneypotID の表記に差があるなど相違点は多い．

そのため本研究では主に攻撃通信データに合わせ，2010 年 8 月 18 日から 31 日と 2011 年 1

データ	期間	観測対象
攻撃通信データ	2010年8月18日 ～ 2010年8月31日	Honeypot001 Honeypot002
	2011年1月18日 ～ 2011年1月31日	
攻撃元データ	2010年5月1日 ～ 2011年1月31日	Honeypot001 ～ Honeypot078

図 4: MWS2011 データセットの特徴

月 18 日から 31 日のデータのみを対象とした. 観測対象も攻撃通信データに会わせて Honeypot001,Honeypot002 のみとし, 相違点は攻撃元データに合わせるよう調整した.

4 システム構成

システムの概要を図 5 に示す.

図 5: システム概要図

全てのデータを MySQL データベースに格納した. その後必要に応じて SQL で必要なデータの抽出し, 各々のデータに対して可視化を行った. 攻撃元データと攻撃通信データの時間情報を結びつけることで, マルウェアによる感染活動の通信とマルウェアが行った通信内容を判断することができる. そのため通信を行っている全ての IP アドレスをポイントマップで表現した.

5 実装と運用

可視化手法はプリズムマップとポイントマップを同時に描画させる手法を用いている. 今回はマルウェアに着目し, どの Honeypot へ攻撃を行っているかに着目した.

図 6 はその例であり, 互いに色や大きさによって攻撃の特徴を表示している.

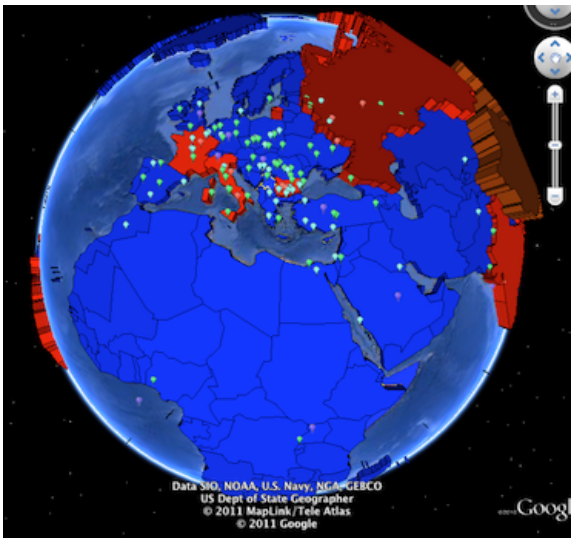

図 6: プリズムマップとポイントマップの統合

プリズムマップは攻撃量の多いマルウェアの順に赤, 橙, 黄, 黄緑, 緑に 5 段階で塗り分けた. この時 5 位以下は緑とし, 全く攻撃がない国に関しては青とした. また, 攻撃量に応じてポリゴンを高く表示するようにした.

そしてポイントマップには攻撃先 Honeypot を色で判別できるようにしており, マルウェアによる攻撃を行っている IP アドレスを含んでいる場合は赤色とした. それ以外の攻撃地点の色は,Honeypot001 への攻撃地点が水色, Honeypot002 への攻撃拠点が緑色, 両方の Honeypot への攻撃地点が紫色とした.

しかし対象データは 2010 年 8 月 18 日から 31 日,2011 年 1 月 18 日から 31 日と時期的に開きがあるため, それぞれを独立したデータとして扱う.

5.1 特定国の攻撃傾向

今回は主に 2011 年 1 月 18 日から 31 日の攻撃元 IP アドレスに着目して可視化を行い, 特定の国からの攻撃傾向に着目する.

ポリゴンマップを見ると全体的に攻撃が全くないか, マルウェアの感染活動が行われている国に分けることができる. 全世界で攻撃量第一位のマルウェア WORM_ DOWNAD 系による攻撃が世界各国で行われており, 猛威を振るっている.

図 7 は 2011 年 1 月 29 日のアジアの可視化例である. 中国のポリゴンマップは他の国と比較するとマルウェアによる攻撃量が突出しており,Mal_DLDER による感染活動が行われている. 図 8 は WORM_DOWNAD と Mal_DLDER に着目し, 感染通信量を表にしたものである. 特にマルウェアによる攻撃通信は常時同じ地点から行われており, 特定 IP から何日にも渡って感染活動が行われていた. 該当のマルウェアは, 実害のあるマルウェアを他サイトからダウンロードさせるダウンローダであるため, 対象の IP アドレスへの通信を止めるなどの対策が考えられる.

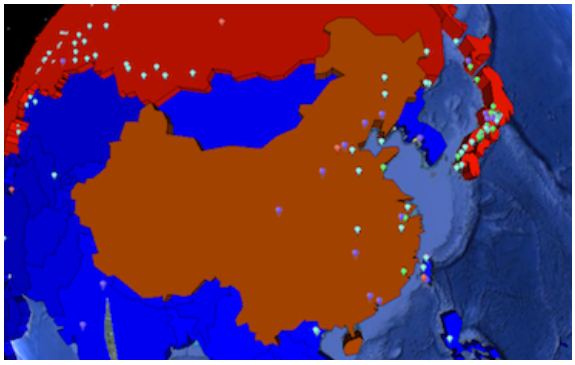

図 7: 中国のマルウェアの可視化

また, 通常よりも複数の Honeypot へ攻撃通信を行っている事が多く, 日付を変更した場合でもその傾向に変動はなかった. 2010 年 8 月 18 日から 31 日のデータではマルウェアによる感染活動は少ないが, 同等の傾向がみられた.

通信元の IP アドレスを調査した結果, モバイル端末からネットワークへ通信を行う LF-dacheng-ZTE-device-manager(以下 ZTE) からの接続が

	WORM_DOWNAD	Mal_DLDER
日本	261	0
アメリカ	818	0
中国	89	10309
ロシア	1178	0

図 8: 2011 年 1 月 18 日から 31 日の特定マルウェア感染通信量

多い. ZTE は主にモバイル端末関連の事業を手がけている企業であり, 一般家庭のマシンから攻撃が行われていると推察される. また Honeypot001, Honeypot002 の両方にほぼ同時期に攻撃を行っていることが多く, 広範囲の IP アドレスへ攻撃を行っていると思われる.

5.2 特定 IP 帯からの攻撃状況

可視化システムを用いて日毎にマルウェアの状況と Honeypot への攻撃元 IP アドレスを描画した場合, 特定の地域に限り多量の攻撃元地点が出現する場合がある. 図 9 は 2010 年 8 月 18 日の攻撃通信データから攻撃元 IP アドレスを可視化したものである. 可視化された地図を見ると特定の地域から攻撃を行っており, アメリカの特定の IP アドレス帯から行われた通信であることがわかる. また, 前後数日間のデータを確認するとアメリカの別の地域から同じ Honeypot へ同様の攻撃が行われており, アメリカの特定 IP アドレス帯を掌握したボットネット等による攻撃が行われていると推測できる. 通信内容は主に Echo Request であり, フランスを中心にヨーロッパからも同様の攻撃が行われている.

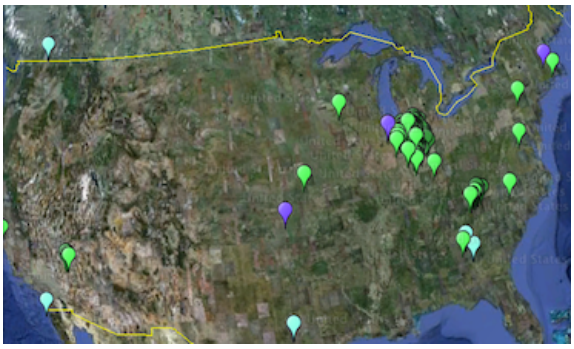

図 9: 密集した攻撃地点

諸国のマルウェアによる直接的な攻撃と, 特定の IP アドレス帯からの攻撃状況を比較すると, 同一の地点から攻撃が行われている事は少なかった. 主に日本や台湾といった比較的面積が小さい国では行われていたが, 誤差範囲と考えられる.

図 10 は 2010 年 8 月 31 日のアメリカの一部に同様の可視化したものであり, 主にテキサス州とルイジアナ州を表示している. このように攻撃地点が特定の州に多々見られる事があった. 全体的にアメリカでは大陸東側の州から調査のためと思われる攻撃が行われており, 海に面する州からマルウェアによる攻撃が行われている傾向にあった. また他国に関しても地域による傾向が存在してあり, 少なくともフランスでは同様の攻撃が確認されている.

図 10: 攻撃地域とマルウェアによる攻撃

6 考察

可視化により, 特定の国から行われている攻撃の傾向と, マルウェア感染活動とその後の通信について簡単に気づきを得る事が出来たとえいる. 特に同じ IP アドレス帯の通信は地理的に似た攻撃元 IP アドレスから多量の通信があるため, 視覚的に気づきを得やすい結果となった.

また, 特定の IP アドレス帯から調査目的で通信を行っていると思われる攻撃と, 実際にマルウェアによる感染活動を行っている IP アドレスとは地理的な差があることがわかった. この結果はボットにされやすいマシンが多い地域や, 攻撃拠点とされやすいマシンが多い地域を判断できると思われる.

7 まとめ

複数のログを可視化することにより, 特定国に感染するマルウェアの傾向を掴む事ができた. またマルウェアの感染活動と通信を同時に可視化することにより, 攻撃対象の Honeypot へ対する調査の動向や感染活動を行う IP アドレスの特徴に気づく事ができた. 今回は攻撃通信ログが短期間で少数の Honeypot を対象としていたため, 今後は長い期間で多数の Honeypot のデータを対象としてシステムの有用性を評価したい.

8 謝辞

研究に伴い CCC DATAset を提供して頂いた NTT コミュニケーション様や関係機関の皆様に深く感謝いたします. また研究においてご指導を頂いた伊東寛氏, 新井悠氏, 松木隆宏氏に御礼を申し上げます.

参考文献

[1] 見えログ:情報視覚化とテキストマイニングを用いたログ情報解析支援システム
http://www.ipa.go.jp/NBP/12nendo/12mito/mdata/4-10h/4-10h.pdf 高田哲司（電気通信大学）

[2] 通信トラフィックの分析による Gumblar 感染 PC の可視化
http://ci.nii.ac.jp/naid/110007889774
金子博一，松木隆宏，新井悠（株式会社ラック）

[3] Nicter
http://www.nict.go.jp/glossary/4otfsk00000l6sm.html Nict

[4] DShield
http://www.dshield.org/reports.html
Internet Storm Center

[5] GeoIP
http://www.maxmind.com/app/ip-location
MaxMind