大小2つの観測網による結果から見たマルウェアの挙動と対策に関する一考察

永尾　禎啓†　　鈴木　博志†　　加藤　雅彦†　　齋藤　衛†

†株式会社インターネットイニシアティブ
サービス本部 セキュリティ情報統括室
101-0051 東京都千代田区神田神保町 1-105

{nagao,hiroshi-suzuki,masa,msaito}@iij.ad.jp

あらまし　CCC DATAset 2011 の広範な観測網によるマルウェア感染活動観測と、自社観測網による局所的な観測からマルウェアの挙動を分析し、ネットワーク上で活動するマルウェアへの対策について考察する。

A study of malware countermeasures and an analysis of malware behavior based on observations from two honeypot networks of different sizes

Tadaaki Nagao†　　Hiroshi Suzuki†　　Masahiko Katoh†　　Mamoru Saito†

†Office of Emergency Response and Clearinghouse for Security Information
Service Division
Internet Initiative Japan Inc.
Jinbocho Mitsui Bldg., 1-105 Kanda Jinbo-cho, Chiyoda-ku, Tokyo 101-0051, Japan

{nagao,hiroshi-suzuki,masa,msaito}@iij.ad.jp

Abstract　In this paper, we compare two observational data sets of malware infection activities, one of which is CCC DATAset 2011 Attack Source Data from Cyber Clean Center's wide honeypot network and another from IIJ's locally installed honeypot network. We study and discuss differences observed between them, and moreover, we also discuss countermeasures against infection activities.

1　はじめに

インターネットにおけるマルウェア活動の脅威が大きく問題になる中、その状況を把握し対策につなげるために、サイバークリーンセンター [1](CCC) をはじめとして、The Honeynet Project [2] など各所でマルウェア活動の観測が実施されている。インターネットイニシアティブ (IIJ) でも、2007 年 4 月より、マルウェア捕獲、解析、対策プロジェクト Malware Investigation Task Force (MITF) を開始し、その一環として、サービス利用者のネットワーク上に観測点を設置してマルウェア活動の観測を行っており [3]、四半期毎発行の Internet Infrastracture Review (IIR)[4] などでその結果を報告している。

研究用データセット CCC DATAset 2011[5] は国内インターネット上の広範にわたる観測点を持つサイバークリーンセンターの観測網を使用して得られたデータであるのに対し、MITF では IIJ ネットワーク内のみに密に設置した観測点で構成する観測網を使用している。

本稿では、研究用データセット CCC DATAset 2011 の攻撃元データ (以降、CCC2011 攻撃元データ) を用いて、これを MITF で得られた攻撃元データ (以降、MITF 攻撃元データ) と比較し、検討し

考察を加える。ここでは、MWS2008、MWS2009、MWS2010 における筆者らによる調査 [6][7][8] と同様の手法を用いる。また、ネットワーク上で活動するマルウェアへの対策について検討する。

2　攻撃元データの比較

CCC2011 攻撃元データの比較対象として、MITF で取得しているデータから、同期間 2010 年 5 月 1 日から 2011 年 1 月 31 日までの、表 1 に示す攻撃元関連情報を抽出して MITF 攻撃元データとした。

項目
時刻
マルウェア取得元 IP アドレス
マルウェア取得元ポート番号
観測点 IP アドレス
観測点ポート番号
プロトコル
マルウェア検体のハッシュ値

表 1: MITF 攻撃元データの情報項目

なお本稿では、以下、マルウェアをハッシュ値で同定して数えることにする。すなわち、同一のハッシュ値を持つマルウェアは 1 種類と数える。

2.1　共通するマルウェア

まず、CCC2011 攻撃元データと MITF 攻撃元データとで共通するマルウェアハッシュ値は 204 種類あった。本稿ではこれらを共通マルウェアと呼ぶことにする。この種類数は、MITF 攻撃元データに現れるマルウェア全体の約 4.8% であり、MITF 攻撃元データにおける共通マルウェアの取得件数合計は全マルウェア取得件数の約 78% であった。

CCC2011 攻撃元データにおいては、共通マルウェアは観測されたマルウェア全体の約 1.6% であり、それらの取得件数合計は全マルウェア取得件数の約 17% であった。

このことから、共通マルウェアという少数のマルウェア群には、CCC の広範な観測網に MITF の観測網のどちらから見ても、比較的活発に感染活動を行っているマルウェアが多く含まれていると考えられる。

これら共通マルウェアをより詳細に見るために、共通マルウェアのみに着目して CCC2011 攻撃元データと MITF 攻撃元データそれぞれでの各マルウェアの取得件数を求めた (図 1)。

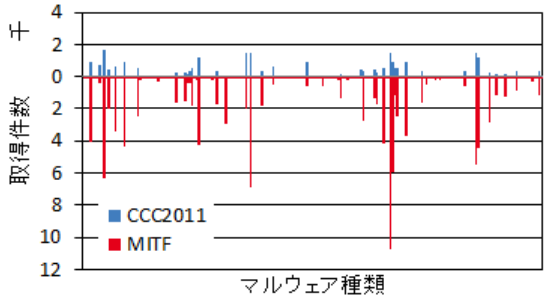

図 1: 共通マルウェアのハッシュ値ごとの取得件数

また、共通マルウェアについて、CCC2011 攻撃元データと MITF 攻撃元データにおける初出日時の差を求めた。CCC2011 攻撃元データで先に観測されたマルウェアについて、その時間の差と種類の数の関係を図 2 に示す。

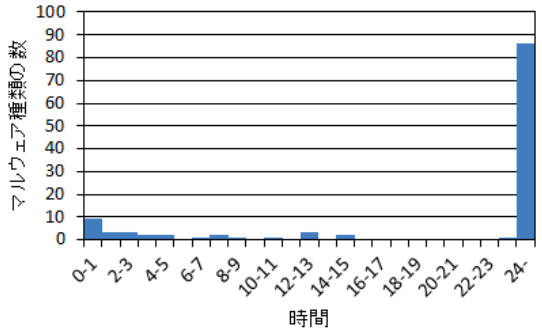

図 2: CCC で先に観測された共通マルウェアの初出日時の差 (単位: 時間)

同様にして、MITF 攻撃元データで先に観測されたマルウェアについて図 3 に示す。

CCC2011 攻撃元データで先行して観測されたマルウェアは 116 種類で、時間差の平均は約 27.0 日、最大は約 235 日であった。これらのうち、CCC2011 攻撃元データに現れてから 1 時間未満で MITF 攻撃元データにも現れたマルウェアは 9 種類 (共通マルウェア 204 種類の約 4.4%) あった。その一方で、24 時間以上経過してから現れたマルウェアは 86 種類 (約 42%) あった。

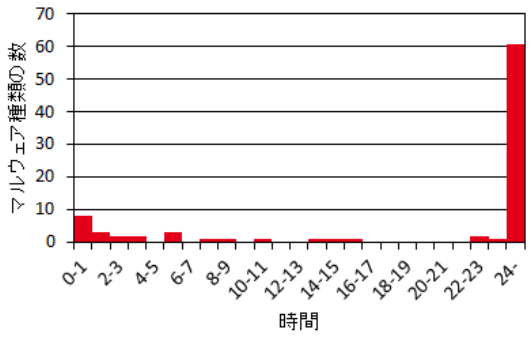

図 3: MITF で先に観測された共通マルウェアの初出日時の差 (単位: 時間)

そして、MITF 攻撃元データで先行して観測されたマルウェアは 88 種類で、時間差の平均は約 19.0 日、最大は約 222 日であった。これらのうち、MITF 攻撃元データに現れてから 1 時間未満で CCC2011 攻撃元データにも現れたマルウェアは 8 種類 (約 3.9%) あり、24 時間以上経過してから現れたマルウェアは 61 種類 (約 30%) あった。

2.2 一方の観測に固有のマルウェア

前節前半部で示した共通マルウェアの割合を言い換えれば、MITF 攻撃元データにおけるマルウェア総取得件数のうち約 22% は、CCC2011 攻撃元データには含まれていない、MITF 攻撃元データに固有のマルウェアによるものだということになる。

同様に、CCC2011 攻撃元データにおけるマルウェア総取得件数のうち約 83% が、MITF 攻撃元データには含まれていない、CCC2011 攻撃元データに固有のマルウェアである。

まず、CCC2011 攻撃元データに固有のマルウェアについて、MITF の観測点から IP アドレス空間上でどれほど離れたところで活動しているかを明らかにする。そのために、各取得時点での取得元 IP アドレスと MITF の観測点が存在する IP アドレスとの間で、アドレス共通部分を示すネットマスク長を求め、これを両者間の距離を表す指標として使うことにした。両者の IP アドレスが数値として近ければ、ネットマスク長は大きくなる (仮に両者が一致した場合には最大値の 32 となる)。MITF の観測点は複数あるが、それぞれの取得元 IP アドレスについて、最も近い観測点 IP アドレスを採用した。そして、CCC2011 攻撃元データ中の固有マルウェアの取得件数をネットマスク長ごとに分類した。このようにして取得元の分布を表したものを図 4 に示す。

次に、MITF 攻撃元データに固有のマルウェアについても、IP アドレス空間上で観測点からどれほど離れたところで活動しているかを明らかにする。先程と同様に計算すればよいが、MITF 攻撃元データでは各取得時点での実際の観測点 IP アドレスがわかっているため、これと取得元 IP アドレスの共通部分を示すネットマスク長を求め、MITF 攻撃元データ中の固有マルウェアの取得件数をネットマスク長ごとに分類した。このようにして取得元の分布を表したものを図 5 に示す。

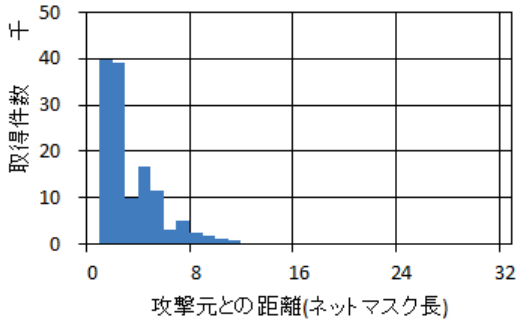

図 4: CCC2011 攻撃元データに固有のマルウェアにおける取得元から MITF 観測点範囲までの距離と取得件数の関係

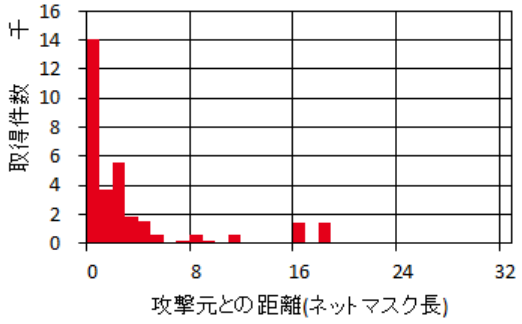

図 5: MITF 攻撃元データに固有のマルウェアにおける取得元から観測点までの距離と取得件数の関係

3 考察

3.1 過去の観測結果との比較

今回の CCC2011 攻撃元データの全マルウェア取得件数は、畑田らによる報告 [5] にもあるとおり昨年と比べて約 86% 減少しており、また昨年も、一昨年と比べて約 53% の減少であった。一方、MITF 攻撃元データは昨年と比べて約 80% 減少し、昨年も一昨年と比べて約 77% の減少であった。今回のデータ収集期間は 9 か月間、昨年と一昨年は 1 年間というように期間の違いはあるが、それを考慮してもなお、観測網の大小によらず大幅減少の傾向が継続していることがわかる。この傾向の理由としては、畑田らの指摘する理由のほか、マルウェアの感染活動の主流がネットワーク経由の直接攻撃から離れ、Web ブラウザ経由の Drive-by-download 攻撃等に移行していることが考えられる。

3.2 マルウェア感染活動の局所性

節 2.1 で示した図 1 からは、MITF 攻撃元データで活発な感染活動が見られるマルウェアであっても、CCC2011 攻撃元データでは活発とは言えないものの多く存在する。このように、CCC2011 攻撃元データと MITF 攻撃元データとで活発な感染活動が観測されるマルウェアには差異が見られた。

また、CCC2011 攻撃元データと MITF 攻撃元データにおける共通マルウェアの初出日時比較 (図 2 および図 3) からは、一方での観測から相当程度遅れてもう一方で観測されるマルウェアが多く存在することがわかった。一度に広範囲のネットワークを対象として感染活動を行うマルウェアでは、このような時間差は小さいはずであるから、多くのマルウェアでは一回あたりの活動範囲が局所的であるものと考えられる。

節 2.2 では、まず図 4 で CCC2011 攻撃元データに固有のマルウェアの取得元分布を示したが、ここからは、CCC2011 攻撃元データに固有のマルウェアについて、その取得元のほとんどは MITF 観測点から遠く離れていたことがわかる。次の MITF 攻撃元データに固有のマルウェアの取得元分布を示した図 5 からは、MITF の観測点から遠い IP アドレスからのマルウェア取得が大半を占めていたことがわかるが、この点は過去の調査で比較的近いアドレスが大半を占めていた状況と大きく異なっている。

謝辞

研究用データセット CCC DATAset 2011 を提供下さり、本考察の機会を与えて下さったサイバークリーンセンターの皆様およびマルウェア対策研究人材育成ワークショップ 2011 実行委員会の皆様に感謝致します。

参考文献

[1] サイバークリーンセンター,
https://www.ccc.go.jp/

[2] The Honeynet Project,
http://www.honeynet.org/

[3] ITpro, 「マルウエアを専用装置で捕獲、挙動を解析」――IIJ が新システム,
http://itpro.nikkeibp.co.jp/article/NEWS/20071115/287291/

[4] IIJ Internet Infrastructure Review,
http://www.iij.ad.jp/development/iir/

[5] 畑田充弘, 他:
マルウェア対策のための研究用データセット 〜MWS 2011 Datasets〜,
MWS2011(2011 年 10 月)

[6] 永尾禎啓, 他:
観測網の大小に基づく結果の比較とその差異に関する一考察,
MWS2008, pp.93-95, 2008

[7] 永尾禎啓, 他:
観測網の大小に基づく結果の比較とマルウェア対策に関する一考察,
MWS2009, http://www.iwsec.org/mws/2009/paper/A2-2.pdf, 2009

[8] 永尾禎啓, 他:
大小 2 つの観測網による結果の時間変化とマルウェア対策に関する一考察,
MWS2010, http://www.iwsec.org/mws/2010/manuscript/1A2-1.pdf, 2010