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

永尾　禎啓†　　鈴木　博志†　　齋藤　衛†

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

nagao@iij.ad.jp, hiroshi-suzuki@iij.ad.jp, msaito@iij.ad.jp

あらまし　広範な観測網による研究用データセット CCC DATAset 2008 の攻撃元データと、自社観測網による局所的な観測データを比較し、その差異について検討し考察を加える。

A Comparison Between Malware Observations from Honeypot Networks of Different Sizes

Tadaaki Nagao†　　Hiroshi Suzuki†　　Mamoru Saito†

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

nagao@iij.ad.jp, hiroshi-suzuki@iij.ad.jp, msaito@iij.ad.jp

Abstract　In this paper, we compare two observational data sets of malware infection activity, one of which is CCC DATAset 2008 Attack Source Data from wide honeypot network and another data set from IIJ's locally installed honeypot network. We study and discuss differences observed between them.

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いる [3]。

研究用データセット CCC DATAset 2008 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 2008 の攻撃元データ (以降、CCC2008 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、その差異を検討および考察する。

2　攻撃元データの比較

今回、CCC2008 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
利用ポート番号／プロトコル
観測点 IP アドレス
ハッシュ値

表 1: 比較対象とする MITF 攻撃元データの情報項目

なお、本稿では、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 個と数える。

2.1　共通するマルウェア

CCC2008 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 588 個あった。本稿ではこれらを共通マルウェアと呼ぶことにする。この個数は、MITF 攻撃元データに現れるマルウェア全体の 26% 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 63% であった。CCC2008 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 1% に過ぎないものの、それらの取得件数合計は全マルウェア取得件数の 47% に及んでいる。このことから、共通マルウェアは、CCC の広範な観測網と MITF の局所的かつ密な観測網のどちらから見ても非常に活発に感染活動を行っているマルウェアを多く含んでいると考えられる。

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 CCC2008 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)。

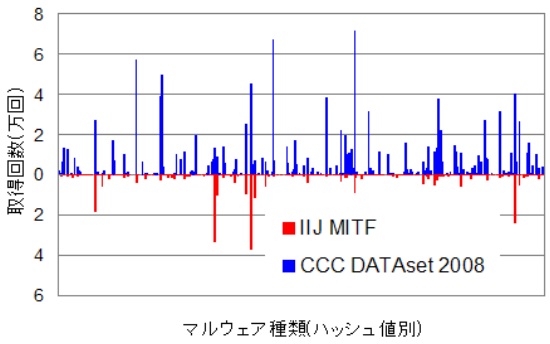

図 1: 共通マルウェア群のハッシュ値ごとの取得件数

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 37% は、CCC2008 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、CCC2008 攻撃元データにおけるマルウェア総取得件数のうち 53% が、MITF 攻撃元データには含まれていない、CCC2008 攻撃元データに固有のマルウェアである。

まず、CCC2008 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 アドレス範囲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。そして、CCC2008 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 2 に示す。

次に、MITF 攻撃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、各取得時点での実際の観測点 IP アドレスと取得元 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 3 に示す。

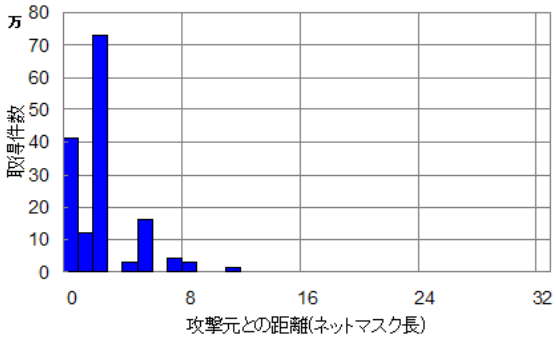

図 2: CCC2008 攻撃元データに固有のマルウェアにおける取得元から MITF 観測点範囲までの距離と取得件数の関係

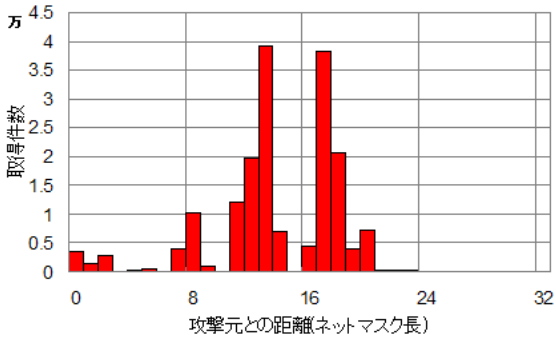

図 3: MITF 攻撃元データに固有のマルウェアにおける取得元から観測点までの距離と取得件数の関係

3 考察

節 2.1 で示した図 1 からは、CCC2008 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、MITF 攻撃元データでは活発とは言えないものの存在が目立つ。このように、CCC2008 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。

節 2.2 では、まず図 2 で CCC2008 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、CCC2008 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 3 からは、MITF の観測点から比較的近い 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。さらに両者の分布を並べて比較すれば、IP アドレス空間上で遠い位置でのマルウェア感染活動は観測しにくく、反対に近い位置での感染活動はよく観測されると言える。

以上のように、いずれもマルウェアの感染活動には局所性があることを示す結果となった。

謝辞

研究用データセット CCC DATAset 2008 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 2008 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] “「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”, 日経 BP ITpro, http://itpro.nikkeibp.co.jp/article/NEWS/20071115/287291/