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

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

あらまし　CCC DATAset 2009 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を比較し、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検討し考察を加える。

A study of malware countermeasures based on a comparison between observations from honeypot networks of different sizes

Tadaaki Nagao†　　Hiroshi Suzuki†　　Masahiko Katoh†　　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,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 2009 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, 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いる [3]。

研究用データセット CCC DATAset 2009[4] 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 2009 の攻撃元データ (以降、CCC2009 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、昨年の MWS2008 における筆者らによる研究 [5] 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

今回、CCC2009 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 2008 年 5 月 1 日から 2009 年 4 月 30 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
利用ポート番号／プロトコル
観測点 IP アドレス
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、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

まず、CCC2009 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 1,956 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約 7% 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 61% であった。

CCC2009 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 3% 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約 43% 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 CCC2009 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)。

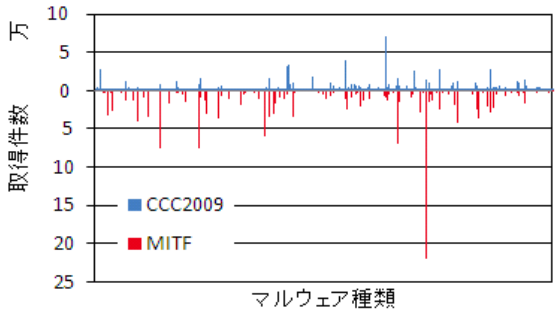

図 1: 共通マルウェアのハッシュ値ごとの取得件数

また、共通マルウェアについて、CCC2009 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。CCC2009 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 (1 時間単位に切り捨て) と個数の関係を図 2 に示す。

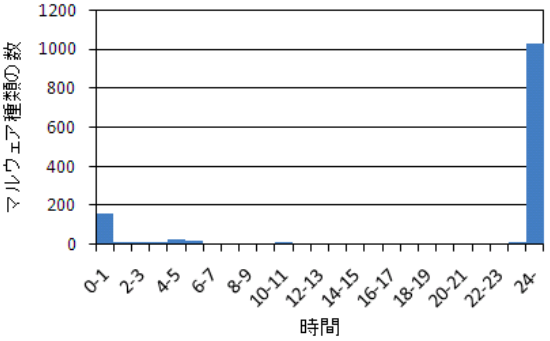

図 2: CCC で先に観測された共通マルウェアの初出日時の差 (単位: 時間)

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。

CCC2009 攻撃元データで先行して観測されたマルウェアは 1359 個で、時間差の平均は約 37.2 日、最大は約 353 日であった。これらのうち、CCC2009 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

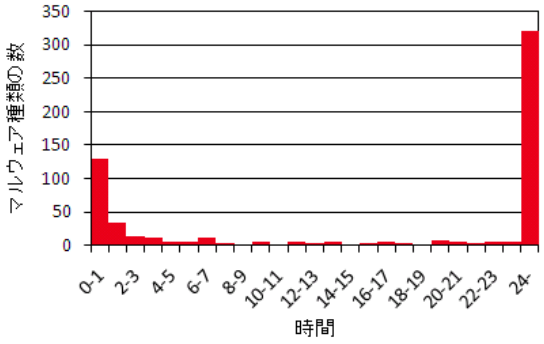

図 3: MITF で先に観測された共通マルウェアの初出日時の差 (単位: 時間)

ウェアは 155 個あった。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 1028 個あった。

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 597 個で、時間差の平均は約 9.0 日、最大は約 165 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 CCC2009 攻撃元データにも現れたマルウェアは 130 個あり、24 時間以上経過してから現れたマルウェアは 320 個あった。

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 39% は、CCC2009 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、CCC2009 攻撃元データにおけるマルウェア総取得件数のうち 57% が、MITF 攻撃元データには含まれていない、CCC2009 攻撃元データに固有のマルウェアである。

まず、CCC2009 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、CCC2009 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 5 に示す。

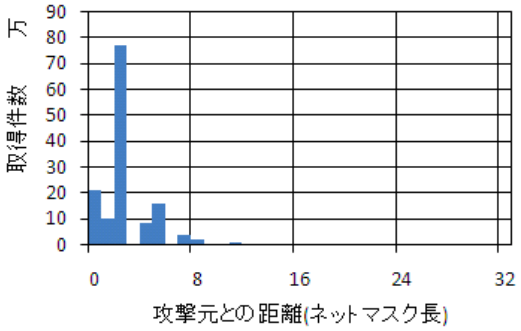

図 4: CCC2009 攻撃元データに固有のマルウェアにおける取得元から MITF 観測点範囲までの距離と取得件数の関係

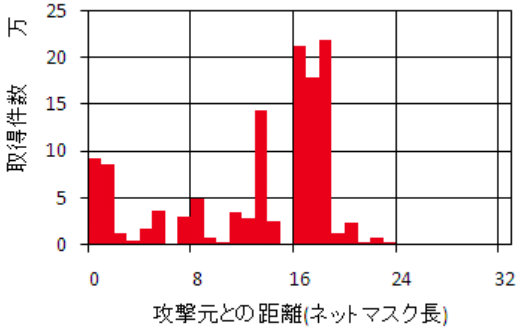

図 5: MITF 攻撃元データに固有のマルウェアにおける取得元から観測点までの距離と取得件数の関係

3 考察

3.1 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、CCC2009 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、CCC2009 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、CCC2009 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。このような時間差が生じる理由の一つとして、両者の観測網の密度や広さの違いが考えられる。

節 2.2 では、まず図 4 で CCC2009 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、CCC2009 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。

3.2 マルウェア感染活動への対策

ここでは、対策の例として、マルウェアのネットワーク感染活動を阻害するために、CCC でマルウェアの感染活動を観測したら 24 時間以内に、その活動のネットワーク的特性 (使用プロトコル、ポート番号、IP アドレス等) の情報を IIJ ネットワークに展開し、フィルタリング等の対策を施すことを考える。この対策手法には技術面だけでなく、法制度面などで検討すべき課題が多いが、以下では、仮にこの対策を実施できたとして検討する。節 2.1 に示した初出日時の比較では、CCC2009 攻撃元データに現れたのち 24 時間以上経過してから MITF 攻撃元データに現れたマルウェアは、種類にして共通マルウェアの約 52% に上っていた。これらのマルウェアの感染活動を MITF 観測データにおける取得件数で数えれば、全取得件数の約 33%に上っている。言いえれば、仮に CCC から日々情報提供を受けて上述のような対策を実施できたとしたら、IIJ ネットワークで観測されたマルウェア感染活動全体のうち、約 33% を未然に防止できていたということになり、非常に有益な対策であるといえる。

謝辞

研究用データセット CCC DATAset 2009 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 2009 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] 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とワークショップを通じた研究成果の共有, MWS2009(2009 年 10 月)

[5] 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