通信ログを基にしたマルウェア配布オペレーションの抽出と可視化の試み

森久 和昭 †　神宮 真人 †　神田 慎也 †　Gregory Blanc†　門林 雄基 †

† 奈良先端科学技術大学院大学情報科学研究科
630-0192 奈良県生駒市高山町 8916 番地の 5

{kazuaki-m, masato-j, shinya-ka, gregory, youki-k}@is.naist.jp

あらまし マルウェアは USB メモリといった物理デバイスだけでなく，インターネットを通じてダウンロードされる．本研究では CCC DATAset2010 攻撃通信データの通信宛先 IP アドレスに着目する．複数のホストの通信ログにおいて，同一の IP アドレスが観測された場合，ダウンロードサイト間には関係があると推測できる．そこで我々はこの関係を明らかにするために，同一 IP アドレスの抽出と可視化を試みた．IP アドレスが所属する AS 番号の情報を組み合わせてグループ化をおこない，その結果を用いてマルウェアダウンロードホストの関係について考察をおこなった．

Towards Extracting and Visualizing Malware Distribution Operations Based on Network Traffic Logs

Kazuaki Morihisa†　Masato Jingu †　Shinya Kanda†　Gregory Blanc†
Youki Kadobayashi†

†Graduate School of Information Scinece, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-city, Nara 630-0192 JAPAN
{kazuaki-m, masato-j, shinya-ka, gregory, youki-k}@is.naist.jp

Abstract A malware has been downloaded from not only physical device like USB memory but the Internet. In this research, we focus on the IP addresses contained in the attack communications of the CCC DATAset2010. Through several host traffic logs, we were able to follow a same given IP address and guess it is related to malware download sites. In order to verify such hypothesis, we attempted extracting and visualizing such IP address relationships. By combining IP addresses with the information of ASes they belong to, we were able to cluster the malware download hosts and made considerations about their relationships.

1　はじめに

マルウェアは USB メモリなどの物理デバイスだけではなく，ネットワークを介して感染する事例が多発している．これはネットワーク上の脆弱性のあるホストや Web サーバを攻撃することによって，広範囲に感染活動をおこなえることが原因である．また攻撃者はボットネットを使用したり，複数のホストを踏み台として経由したりすることで攻撃者自身が特定されることを防ぎながら，より手軽にマルウェアの配布をおこなうことが可能である．そのためマルウェア対策者としては広範囲に拡散しているマルウェアの情報から攻撃者の特定と，その感染対策をおこなう必要がある．ここでは同一マルウェアの配布をおこなったり，あるマルウェアが通信をおこなうホストをまとめて「マルウェア配布オペレーション」と呼ぶ．

マルウェアは先に述べたようにネットワークを介して感染する．現在のインターネットではホストの

一意性を示す IP アドレスを持たなければ通信をおこなうことができない．したがってマルウェアを配布する場合の通信や，マルウェアがおこなう通信にはこの IP アドレスが必ず含まれる．そこで我々はこの IP アドレスに注目した．マルウェアに感染したホストがおこなう通信ログを基に IP アドレスの関係を抽出し，その IP アドレスが所属する自律システム (AS: Autonomous System) 番号でグループ化をおこなう．このグループ化をおこなった結果を分析することによって，マルウェア配布に関する特徴を得ることが目的である．また特徴を分類するために可視化をおこなった．現在，マルウェアの活動に関する研究はいくつかおこなわれている [6],[7]．しかし AS 番号に基づいた分析と可視化に関する研究はおこなわれていない．そこで我々は IP アドレスに着目し，それが依存する AS 番号に基づく情報を用いるアプローチをおこなった．なお使用する通信ログは CCC DATAset 2010 を用いた．以下，第 2 章では関連研究について述べる．第 3 章ではマルウェア配布オペレーションについて定義とその抽出アルゴリズムおよび可視化について説明する．第 5 章では抽出結果に基づいて考察をおこなう．最後に第 6 章ではまとめと今後の課題を述べる．

2　関連研究

これまでにマルウェアに関する研究がいくつかおこなわれている [6], [7]．文献 [6] ではマルウェア配布元の可視化をおこなっている．著者らはマルウェア配布ホストを地理的分布に基づいて世界地図にマッピングしている．その結果，マルウェアの配布は世界各地からおこなわれているということと，Web ベースのマルウェア配布サイトは中国に集中しているということが分かっている．しかしこの可視化のアプローチは IP アドレスを実世界の地理情報に当てはめているため，ネットワークトポロジとしてのまとまりを把握することが困難である．そこでマルウェアに感染したノードがおこなう通信から発見した IP アドレスと AS の関係を組み合わせて可視化し，視覚的に把握しやすくなるようなビューワーを作成する．

文献 [7] ではマルウェアの配布元とその通信ホストの関係と，マルウェアの活動の特徴について分析をおこなっている．その結果，ダウンロードホストの活動期間が 3 つに分類できること，ダウンロードの回数とマルウェアの種類に関係があることを明らかにしている．しかしこの分析についてもネットワークにおけるマルウェア配布サイトの関係は調査されておらず，同一マルウェア配布サイト同士の関係が明らかになっていない．そこで我々はこの同一マルウェア配布サイト同士の関係について分析をおこなう．

3　マルウェア配布オペレーションの定義

この章ではマルウェア配布オペレーションについて定義する．まずマルウェアとは悪意のある第三者が他人のコンピュータから情報を盗み出したり，破壊したりするためのプログラムのことである．このマルウェアは Web サイトを通じてダウンロードされることが多発している．またその種類は多く改変された亜種も存在する．これらのマルウェアを配布しているホストには共通点があると考えられる．たとえば同一のマルウェアを配布している場合，一人の攻撃者がホストを攻撃した結果，マルウェアの配布活動をおこなうようになったと想定できる．そこでマルウェア配布をおこなうホスト同士をグループ化し，これらの活動をマルウェア配布オペレーションと呼ぶ．このような状況においてマルウェア配布オペレーションを抽出することは，マルウェアの感染活動範囲を特定したり被害状況を確認したり，攻撃者を特定するための手がかりにしたりすることができるようになる．次にこのマルウェア配布オペレーションの具体例を示す．

図 1 では，ハニーポットが異なるマルウェア配布サイトから同一のマルウェアをダウンロードするモデル図である．この場合，マルウェア配布サイト A とマルウェア配布サイト B には，同一のマルウェアを配布したという関係がある．そのため両者はマルウェア配布オペレーションにグループ化できる．

またマルウェアがおこなう通信においてペイロード部分に IP アドレスが含まれている場合にもホストに関係があると仮定する．図 2 では，マルウェアに感染したハニーポットがマルウェア配布サイト A(以下，ホスト A) と通信しているモデル図である．ハニーポットとホスト A は直接通信をしている (実線で表し，直接ノードと呼ぶ) が，ホスト B とは直接通信していない．しかし，ホスト A との通信内容 (ペイロード部分) においてホスト B の IP アドレス

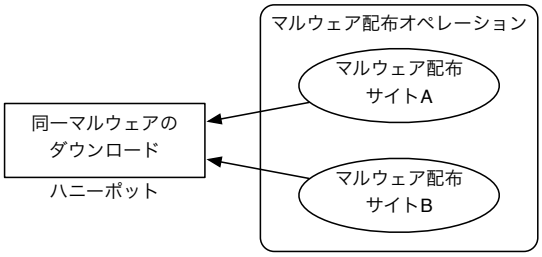

図 1: マルウェア配布オペレーション例 1

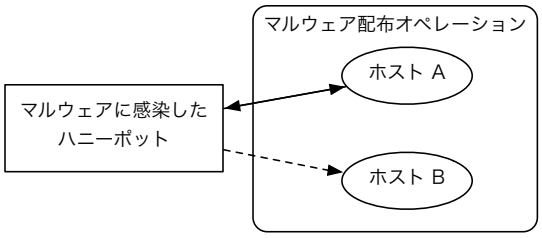

図 2: マルウェア配布オペレーション例 2

が含まれていたり，マルウェア内部にホスト B の IP アドレスが含まれていたりした場合，ホスト A とホスト B には関係があると考えられる (点線で表し，間接ノードと呼ぶ)．そのため両者はマルウェア配布オペレーションにグループ化できる．

今回はグループ化の条件として IP アドレスに着目しているが，マルウェア配布オペレーションの調査にはこの IP アドレスの他に AS 番号を利用している．AS とは大きくインターネットを分割する単位である．AS 同士は自己管理下のネットワークの経路情報を交換する．このとき，経路情報を提供する AS をプロバイダー AS と呼ぶ．一方，経路情報を受け取る側の AS をカスタマー AS と呼ぶ (参考文献 [5])．この 2 つの組み合わせによって，経路情報を流すプロバイダーと受け取るカスタマーの P2C 関係と，お互いが経路情報を流す P2P の組み合わせが発生する．AS が互いに経路情報を交換することによって，より大きなネットワークを構成することができる．AS が管理するネットワークには IP アドレスが割り当てられ，必ず 1 つの IP アドレスに対して，1 つの対応する AS がある．また AS にはインターネット上において重複しない唯一の AS 番号が割り当てらている．そのため AS 番号と IP アドレスは一対多の関係になっている．つまり異なる IP アドレスであったとしても，AS 番号が同じであれば，同一 AS のネットワークに存在しているということが断定できる．この性質を利用してマルウェア配布サイトの IP アドレスと AS 番号の関係を調査することによって，マルウェア配布オペレーションの特徴を抽出することが可能である．

4 マルウェア配布オペレーションの抽出

マルウェア配布オペレーションを抽出するために，データセットからデータを取り出し，AS 番号との対応付け処理をおこなう必要がある．そこでこの章では，データセットから抽出した IP アドレスを AS 番号にマッピングするアルゴリズムを示す．

4.1 アルゴリズム

本節では，マルウェア配布オペレーションの抽出アルゴリズムについて詳述する．抽出アルゴリズムは下記の 3 つのフェーズで成り立っている．

phase1 通信ログから IP アドレスを抽出する．

phase2 ある期間の AS 番号を抽出する．

phase3 IP アドレスと AS 番号をマッピングする．

phase1 では，IP アドレスの正規表現を利用し，通信ログのペイロードに IP アドレスが含まれているパケットを特定する．通信ログは CCC DATAset 2010 の攻撃通信データを用いた．phase2 では，通信ログを取得した時期の AS 番号を取得する．AS 番号の情報は CAIDA[1] のデータを利用した．phase3 では，phase1 で特定したパケットの送信元 IP アドレスと宛先 IP アドレス，及びペイロードに含まれる IP アドレスに対して，通信当時に当該 IP アドレスが存在していた AS 番号をマッピングする．次にアルゴリズムの疑似コードを示す．

以上の処理をおこなうことによって，攻撃通信データから IP アドレスの抽出と AS 番号のマッピングをおこなうことができる．このマッピングデータを用いてマルウェア配布オペレーションの調査と可視化をおこなう．

次に攻撃通信データに出現した IP アドレスと AS 番号を表 1 にまとめる．この表 1 より 1 日ごとに接続数の増減があることが読み取れる．

Algorithm 1 AS 番号マッピングの方法

```
pcap ← 通信ログ
for all packet in pcap do
    packet の送信元 IP アドレスと送信先 IP アドレスを保存する
    if packet のペイロードに IP アドレスが含まれている then
        ペイロードに含まれている IP アドレスを保存する
    end if
end for
for all IP in 保存した IP アドレス do
    パケット送信日時の AS データを用意する
    IP に該当する AS をさがす
    探した結果の AS 番号と IP をマッピングして保存する
end for
```

表 1: IP アドレスと AS 番号の出現数

日付	IP アドレス数	AS 数
3 月 5 日	200	101
3 月 6 日	298	145
3 月 7 日	266	126
3 月 8 日	210	102
3 月 9 日	271	136
3 月 10 日	227	108
3 月 11 日	293	148

4.2 攻撃通信データの可視化

攻撃通信データと抽出したデータはテキストデータであるため直感的にどのような構造になっているのか把握することが難しい．そこで今回は 2 つの方法で可視化をおこなった．

攻撃通信データの可視化：AS 番号のマッピング情報は用いず，マルウェアに感染したホスト (ハニーポット:H) とその通信相手 (ノード:N) のデータだけを用いて可視化をおこなった．これはハニーポットとノードの基本的な関係を明らかにすることが可能である．

可視化した結果を基にモデル化した図を図 3 に示す．なお可視化には Graphviz[2] を用いた．図 3 中の実線は直接通信がおこなわれたことを示し，点線は通信パケットのペイロード部分に IP アドレスが含まれていたことを示す．図 3 の 1-A はハニーポットとノードが 1 対 1 で通信をおこなうモデルである．これに対し 1-B はハニーポットからノードへ HTTP 通信をおこなう際に観測できる．なぜなら URL が IP アドレスで指定されており，これがペイロード部分に含まれていたためである．そこで 1-A および 1-B は同じ通信モデルである．次に 2 は 1 つのノードに対して複数のハニーポットが接続しているモデルである．3 はノードからハニーポットへのパケットの中に第三者のノードの IP アドレスが含まれているモデルである．4 はハニーポットとノードの通信において両方向とも，パケットの中にノードの IP アドレスが含まれているものである．これはノードが FTP サーバであり，マルウェアをダウンロードさせるための命令をハニーポットへ送信しているためこのようなモデルになる．5 は 1 台のハニーポットが複数のノードと通信をおこなっている．さらに特徴として一方のノードへ送信したパケット中に他方のノードの IP アドレスが含まれていることである．これはノード同士が連携して攻撃をおこなっていたり，ボットネットである可能性がある．以上の 5 種類がマルウェアオペレーションのモデルとして明らかになった．この結果より，異なるハニーポットであっても共通のノードが存在していることと，共通の間接ノードが存在することが読み取れる．したがってこれらのノードをグループ化することでマルウェア配布オペレーションを抽出可能である．

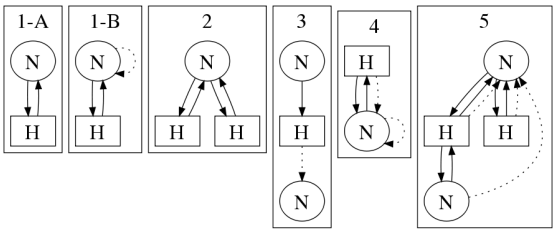

図 3: マルウェア配布サイトの構造

AS 番号をマッピングしたデータの可視化:AS 番号とマッピングしたデータを可視化した結果を図 4 に示す．なお可視化には Java と，グラフィック描画ライブラリとして Prefuse[3] を用いた．図 4 の円形の部分はハニーポットである．このハニーポットから矢印で通信先の AS を示している．この図からも，異なるハニーポットから共通して使用している AS があることが読み取れる．したがってマルウェア配布サイトの一部はハニーポットとは別の，AS のネッ

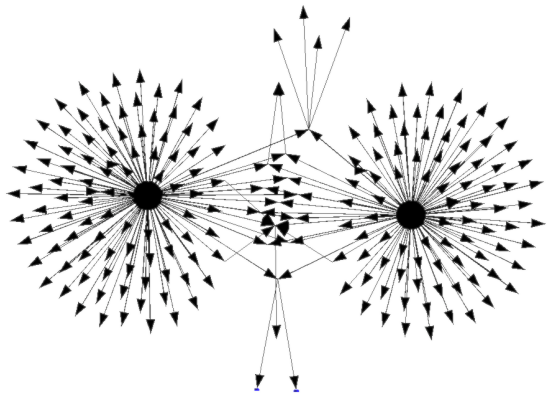

図 4: マッピングデータの可視化

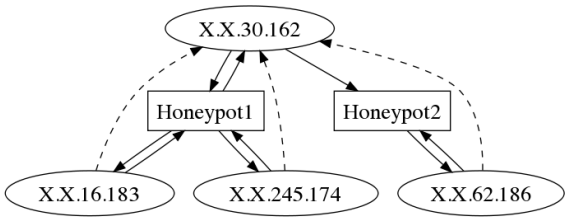

図 5: オペレーションの構造例

トワーク内にあることが考えられる．

5　抽出結果の一例と考察

5.1　マルウェア配布オペレーションの一例

本節では手動で解析したデータの中から発見したオペレーションを例として挙げる．注目点としてハニーポットを中心とし，そのハニーポットが感染しているマルウェアの活動を観測し他のノードや AS とどのような関係があるかを調査する．

ハニーポットの通信ログを可視化した結果，先に述べたように図 3 のようなモデル構造が得られた．この構造に実際に用いられた IP アドレスを当てはめると図 5 になる．ここで用いられた IP アドレスは実際の攻撃ノードであるため一部を伏せている．この図 5 からは攻撃ノードがどのように連携しているのか判断することはできない．そこで IP アドレスを AS 番号でマッピングした情報を用いると図 5 は 2 つの AS が関係していることがわかった．すなわち，X.X.30.162 が所属する AS と，それ以外が所属する AS である．さらに通信ログを調査した結果，X.X.30.162 はマルウェアを実際に配布するノードであった．下段のノード (X.X.16.183, X.X.245.174, X.X.62.186) は間接ノードであり，上段の攻撃ノード (マルウェアの配布と，FTP で接続を促すノード) と関係があることが推測される．この攻撃ノードと間接ノード間の通信ログは得られないため，実際に通信がおこなわれたか，また通信がおこなわれている場合には，どのような内容であったかを確認することはできない．ただし仮説として， X.X.30.162 が情報収集を目的とするような悪意のあるノードであった場合，下段ノードはユーザ情報や，コンピュータの情報を送信していることが想定される．以上より，マルウェア配布ノードの通信パケットをマイニングしたことにより，攻撃ノードの発見とそれと連携しているノードを推測することができた．しかし，攻撃元を隠すためにマルウェア配布オペレーションの中にさらに，役割を分割している可能性がある．この別々の役割のノードが連携した場合，組織的な攻撃が可能となる．この場合はオペレーションを詳細に把握することによって対策が立てやすくなる．例えばそのオペレーションが属する AS からの経路情報を受け取らないことや接続関係を無くすことが挙げられる．

5.2　AS の関係を用いた調査結果の考察

5.1 節で示したマルウェア配布オペレーションについて考察をおこなう．図 5 の構造において便宜的に攻撃ノードである X.X.30.162 が所属する AS を AS_X とする．またハニーポットと下段のノード (X.X.16.183, X.X.245.174, X.X.62.186) が所属する AS を AS_Y とする．この両者の AS について，Route Views[4] が公開している経路広告情報を用いて分析・可視化すると図 6 の構造であった．AS_X は $AS1$ のカスタマー AS であり，AS_Y は $AS2$ のカスタマー AS である．また $AS1$ と $AS2$ は P2P の関係にあることがわかった．つまり AS_X および AS_Y のプロバイダー AS 同士が P2P 関係であるため，AS_X と AS_Y は AS 階層構造の同じ階層に位置しているということがいえる．ここで AS_X と AS_Y が所属しているノードの役割に注目すると，AS_X が攻撃をおこなう側のノードが存在しており，AS_Y は攻撃を受ける側のノードが存在している．したがって，AS 階層構造において同じ階層に位置する AS 同士でも攻撃がおこなわれていることが確認できた．以上より IP アドレス情報から AS 情報とマッピングする

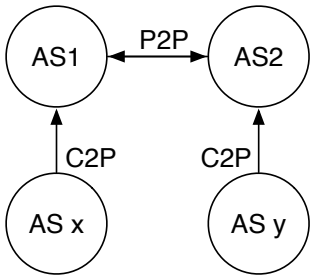

図 6: AS_X と AS_Y の関係図

ことで，攻撃ノードから攻撃を受けるノードまでのネットワークトポロジの作成が可能であることを示した．

以上の結果を用いると，AS の信頼性を評価することができるのではないかと考えられる．つまりハニーポットで得られた攻撃通信データを AS 情報にマッピングすることによって，攻撃ノードが存在する数量を AS 単位で集計することができるためである．現状のネットワークでは，階層構造の下位に位置する AS は個人が運営していたり，ポリシーが明確でないまま運営していたりするため，悪意のあるホストを設置しやすいといえる．そこで攻撃ノードから攻撃を受けるノードまでのネットワークトポロジを調査することによって，どの AS に攻撃ノードが多いのかという情報を得ることが可能である．その情報を基に AS の信頼性を客観的に判断することができるようになる．

次に AS 情報をマッピングすることによる利点について考察する． AS 情報と経路広告情報は公開情報であるため誰もが過去の情報を取得することが可能である．また通信ログを取得するときに時間情報を合わせて保存することによって，後で AS 情報とマッピングできる．つまり 1ヶ所に設置したセンサーで通信ログを取得するだけで，後からマルウェア配布ホストがおこなった攻撃のネットワークトポロジを明らかにすることができる．そのため少数のセンサーで攻撃に使われた AS を明らかにすることができ，攻撃範囲の特定およびマルウェアの拡散状態について把握することができるのではないかと考えられる．

6 まとめ

本研究ではマルウェア配布オペレーションの抽出と可視化をおこなった．オペレーションの抽出は CCC DATAset 2010 の攻撃通信データを用いて IP アドレスに着目しておこなった．またその IP アドレスが所属する AS 番号とマッピングすることによって被害者 (ハニーポット) と攻撃者の関係を明らかにすることを試みた．その結果，5 種類のマルウェア配布構造が存在していることが明らかになり，一部のマルウェア配布オペレーションの AS の関係性を手動で解析することができた．その結果， AS の信頼性評価をおこなうことが可能ではないかということと，少数の通信を記録するセンサーを用いることで，攻撃範囲の特定およびマルウェアの拡散状態について把握できるのではないかということがわかった．

今後の課題として，インターネット上すべての AS と攻撃に使われた AS の関係をマッピングし，より大きなマルウェア配布オペレーションの発見をおこなうことが考えられる．また可視化の方法も改善する余地があり，より詳細な情報を確認できる機能や異なる通信ログから同一マルウェアオペレーションを抽出し比較するというアプローチが考えられる．

謝辞 本研究をおこなうにあたり，AS 情報の提供および助言をいただいた，奈良先端科学技術大学院大学の櫨山寛章氏に感謝の意を示す．

参考文献

[1] Caida. http://www.caida.org/home/.

[2] Graphviz. http://www.graphviz.org/.

[3] Prefuse. http://prefuse.org/.

[4] Route views. http://routeview.org/.

[5] Xenofontas Dimitropoulos, et al. AS Relationships: Inference and Validation. *ACM SIGCOMM CCR*, 2007.

[6] 松木隆宏, 新井悠. CCC DATAset 2009 によるマルウェア配布元の可視化. *MWS2009*, 2009.

[7] 石井宏樹, 佐藤和哉, 田端利宏. ダウンロードホストに着目したマルウェアの活動傾向分析. *MWS2008*, 2008.