複数ホストの通信視覚化による比較解析

清野 祥之 †　　小池 英樹 †

† 電気通信大学 大学院情報システム学研究科
182-8585 東京都調布市調布ヶ丘 1-5-1

seino@vogue.is.uec.ac.jp , koike@is.uec.ac.jp

あらまし 本研究では，ネットワーク管理者がマルウェアの感染を早期発見するため，複数ホストやネットワークの通信を視覚化する解析ツールを開発した．解析ツールでは，通信ログから IP アドレスを取り出し，通信したホストやネットワークごとに色分けをして，通信の変化をアニメーションで視覚化する．視覚化には，2 次元マップに IP アドレスの第 3 第 4 オクテットを割り当ててプロットする機能と，4 つのフレームに各オクテットの値変化を座標で示す機能を実装した．ハニーネットで収集した攻撃通信データを解析ツールで解析した所，2 つのホストから 3 種のスキャンが同時に行われている様子を容易に発見し，解析ツールの効果を確認した．

Visualization and Analysis of Multi-Host Traffic

Yoshiyuki SEINO†　　Hideki KOIKE†

†Graduate School of Information Systems, The University of Electro-Communications
1-5-1 Chofugaoka Chofu-shi Tokyo 182-8585

seino@vogue.is.uec.ac.jp , koike@is.uec.ac.jp

Abstract To find the malware infections, we developed an analysis tool that visualizes hosts and network traffic. This tool displays an animation of traffic patterns which are color-coded according to hosts and network. The system is composed of two modules. One plots third and fourth octets on two-dimensional maps. The other displays octets change on four frames. Then we analyzed a honeynet log, we could easily find three malware scans came from two hosts in same time.

1　はじめに

近年，マルウェアの感染はインターネットだけでなく，私たちの社会を脅かす危険な存在として認識されている．ネットワーク管理者がマルウェアの感染や不正な通信を発見する手段として，通信ログの解析が挙げられるが，通信ログには大量の情報が含まれており，手作業での解析は困難である．この問題を解決するために，ログ情報を視覚化するという手法がある [1][2][3].

情報を視覚化して解析する手法には，井上らの Nicter[1]，大野らの IP Matrix[2]，向坂らの内部ネットワーク監視のための視覚化システム [3] などがある．Nicter はリアルタイムで通信を地理的，論理的に視覚化することで，国内外への通信を監視し，不正な通信を発見，解析することができる．大野らの IP Matrix では，Snort のログから IP アドレスを取り出し，第 3 オク

テットを縦軸に，第 4 オクテットを横軸に割り当てて 2 次元マップ上の座標に示すことで，どのホストへと攻撃が行われていたかを視覚化している．向坂らの手法では，内部ネットワークの IP アドレスと対応した建物の図を利用し，通信が行われたホストの実際の位置を視覚化することで，内部ネットワークの通信を監視可能にしている．しかし，これらの手法では通信先の IP アドレスの値の変化が詳細に視覚化されておらず，マルウェアのスキャンが行われている様子を表現できない．また，ホストの位置などの環境がネットワークによって異なるため，それらを考慮した視覚化が必要である．

マルウェアのスキャンは時間をかけて様々なホストへと行われる．そのため，ログの解析には時間に着目した解析と，空間に着目した解析が必要となる．そこで我々は解析ツールとして，時間視覚化を行う WIV[4] に，空間視覚化の機能を加えて実装した．

空間視覚化の機能としては，大野らの IP Matrix の手法を参考に，Layered IP Matrix を開発した．Layered IP Matrix では，ホストやネットワークごとの通信を表示した 2 次元マップを透過的に重ねることによって，マルウェアのスキャンが行われた通信先を空間的に比較して解析可能にした．

時間視覚化の機能を担う WIV には，通信先 IP アドレスの各オクテットの値変化をホストやネットワークに応じて色分けする手法を実装した．これにより，マルウェアのスキャンの通信先が時間に応じてどのように変化しているかと，そのスキャンの発生場所を確認可能にした．

最後に，開発した解析ツールを用いて CCC DATAset 2010[5] の攻撃通信データを解析した結果として，複数のホストにマルウェアが感染し，特徴的な通信が複数行われていた様子を示す．

2　解析ツールの概要

開発した解析ツールは，パケットキャプチャを行ってリアルタイムで通信を視覚化する機能と，ログ情報から通信内容を視覚化する機能を持つ．

使用可能な通信ログは，pcap ファイルと，tcpdump を使用して pcap ファイルから出力したログの 2 種類である．以下は，tcpdump によってログを出力する際に使用したコマンドである．

```
$ tcpdump -n -r data.pcap > data.log
```

読み込んだログ情報は，我々が実装した Layered IP Matrix，WIV，その他の機能である LogViewer と TrafficViewer の合計 4 つの視覚化機能によって動画で視覚化される（図 1）．動画での表示にはログの時間を実時間と対応付けて視覚化を行うことで，実際にどのようなタイミングで通信が行われていたかを視覚的に認識できる．ログ情報の時間と実時間との対応付けには任意の伸縮が可能となっており，実際の 2 倍の速さで動画を表示することや，実際よりも遅く動画で表示することも出来る．さらに，実時間との対応付けを無視して，読み込んだログ情報を次々と視覚化し，動画をなるべく速く表示することも可能である．これらの速度調節の機能を使用することで，ログ情報の量に応じて速度調整が可能になり，より解析しやすくなるものと考える．また，通信ログには通信を全く行っていない時間が存在するため，通信の変化の時間相関を解析する時以外にはこの無通信の時間を早送りして，解析時間の短縮をはかる．

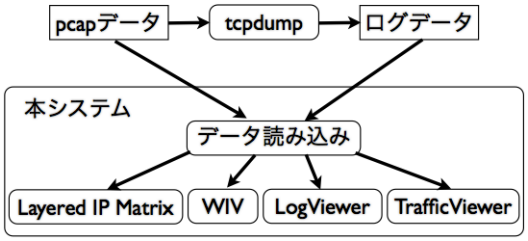

図 1: システム概要図.

2.1　Layered IP Matrix

IP アドレスの 2 つのオクテットをそれぞれ 2 次元マップ上の Y 座標と X 座標に割り当てて視覚化する手法として，大野らの IP Matrix[2] がある．

IP Matrix を用いることで，IP アドレスの論理的な近接関係を表現でき，かつディスプレイ表示域に IP アドレスを経済的に視覚化することが可能となる．例えば，外部のネットワークへ行われた通信の下位オクテットを視覚化することで，内部から不正に行われたスキャンを視覚化することができる．

しかし，IP Matrix では全てのホストの通信を一つの 2 次元マップに視覚化してしまうため，特定のホストやネットワークの通信に焦点を当てることができず，それぞれの通信がどのホストから行われたものか判断することができない．

今回我々は IP Matrix を利用して，複数のホストの通信をそれぞれ 2 次元マップ上に色分けして視覚化し，それらのマップを透過的に重ねあわせることで，複数のホストの通信を比較できる視覚化手法，Layered IP Matrix を実装した（図 2）．

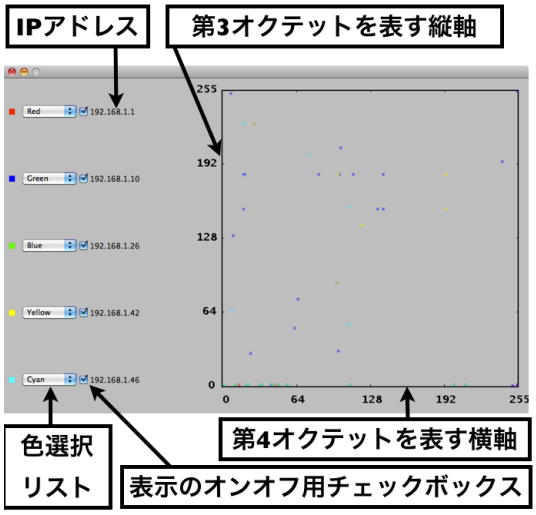

図 2: Layered IP Matrix 概要図．ウィンドウ右側に IP Matrix のマップを描画し，左側のコントロールでそれぞれのマップの色や表示のオンオフを設定する．

2.2 WIV

すでに我々は通信の変化を視覚化する機能として，WIV（Worm Infection Visualizer）を開発した [4]．IP アドレスを動画で視覚化する手法，通信をポート別に視覚化する手法に加えて，今回，ホストやネットワーク別に視覚化する手法，通信回数の偏りを視覚化する手法の 2 つを付加した WIV を開発した．

IP アドレスを視覚化する際，IP アドレスは 2^{32} 個のユニークな値を表現できるため，その 1 つ 1 つをディスプレイ上の点に対応付けて表現するだけでも，現在の一般的なディスプレイの表示総画素数をはるかに超えた巨大なディスプレイを必要としてしまう．そのため，IP アドレスの各オクテットの値を，それぞれ対応した枠の X 座標位置に縦線で表現し，合計 4 本の縦線で IP アドレス 1 つを表現した（図 3）．

IP アドレスを動画で表示する際，視覚化した IP アドレスを表示し続けておくと，後から視覚化した IP アドレスがどの位置に表示されたのかわからなくなってしまう．そこで，表示した IP アドレスの縦線を時間の経過と共に減色させ，2 秒程度で完全に見えなくなるようにした（図 4）．

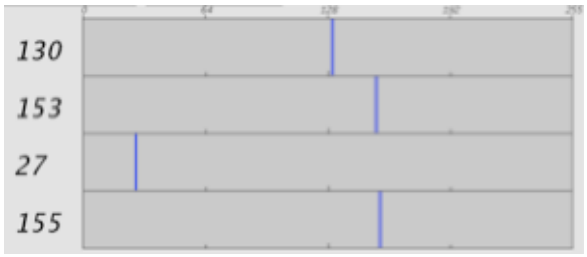

図 3: WIV による IP アドレス視覚化の様子．

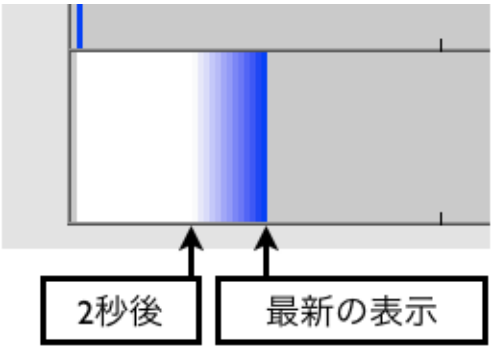

図 4: 時間経過による減色の様子．

マルウェアは特定のサービスへと攻撃を行う

ため，特定のポートへの通信を分けて視覚化することで，マルウェアの特徴的な通信を観測できるようになる．そこで，いくつかの指定されたポートごとに分けて視覚化した．指定されたポートへの通信があった場合，IP アドレスを表現するそれぞれの縦線の表示位置の高さを変えることで，指定されたどのポートへの通信かを表現する．例えば，HTTP や FTP といった普段から利用されているサービスへの通信，上記のサービスを除いた well-known ポートへの通信，ポート番号 1024 以上のポートへの通信，といった 3 つの通信別に分割して視覚化することができる（図 5）．

マルウェアは感染を行うと感染したホストのローカルネットワークに対して攻撃を行うことがある．どのネットワークに感染が行われたか，どのホストに感染したかをネットワーク管理者が素早く確認することで，マルウェアの感染拡大の防止に繋がると考えられる．そこで，特定のホストやネットワークからの通信を色分けし，IP アドレスを表現する各縦線の色に対応づけて表現した（図 6）．

特定のネットワークやホストに対して頻繁に通信を行っていないか調査できるよう，各オクテットの値に対しての通信回数を視覚化した．通信回数の視覚化には，IP アドレスを視覚化した縦線の表示のあとに，通信回数に応じた長さの白い縦線を表示することで実現した．白い縦線の長さは，各オクテットごとに縦線で表示した値ごとの回数を記録しておき，最も回数の多かった値の位置の縦線の長さを 1 として正規化して決まる（図 7）．

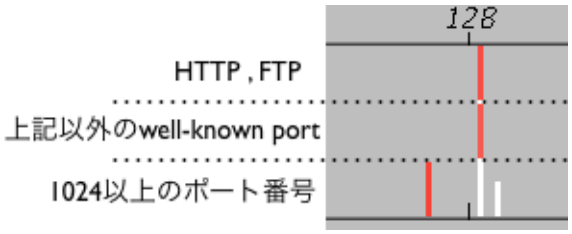

図 5: WIV による指定ポートの通信視覚化．各オクテットの縦線の表示域が指定数で分割されている．

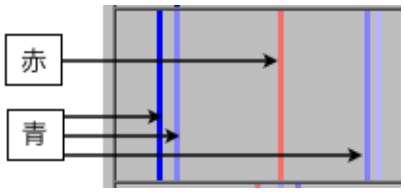

図 6: ホストやネットワークによる色分け．

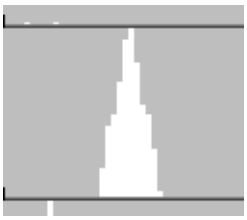

図 7: 正規化された通信回数．

2.3 TrafficViewer

通信回数を折れ線グラフで時系列表示した（図 8）．定期的なスキャンや感染が行われれば，時系列のグラフに特徴があらわれる．特に，マルウェアは感染を拡大させるために定期的なスキャンを行う傾向があるため，この視覚化手法は有効だと考えられる．

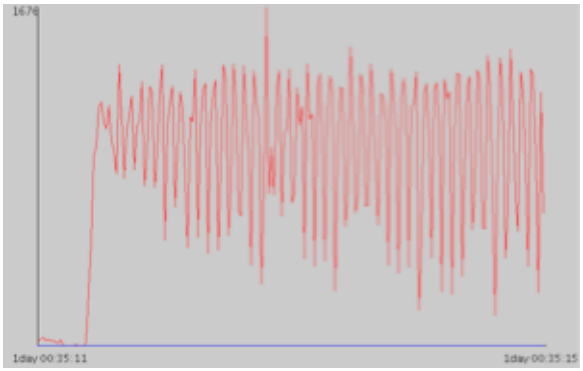

図 8: 通信回数を縦軸に，時間を横軸に割り当てて描画したグラフ．

2.4 LogViewer

具体的な通信内容としてのポート番号や時間を確認できるように，視覚化中のログ情報を表示することとした．ログ情報として表示する内

容は，マイクロ秒単位までの時間情報，送信元 IP アドレス，送信元ポート番号，送信先 IP アドレス，送信先ポート番号の 5 つである．

3　結果

今回我々が開発したツールを利用することで，どのような結果を得ることが可能かを示すために，実際に CCC DATAset 2010 の通信ログを解析した．CCC DATAset 2010 の通信ログは，プライベート IP アドレスを割り振られた 2 台のホストの通信が記録されているため，この 2 台のホストの通信を比較した．

各ログを解析した結果，最も特徴的な通信として，2 台のホストから 3 種類のスキャンが同時に行われている様子を容易に発見することができた（図 9 参照）．

青色で視覚化したホストの通信は，2 つのポートに対して，通信先の IP アドレスの第 1 オクテットを固定し，第 2 第 3 第 4 オクテットの値をランダムに変えながらスキャンを行っていた．

赤色で視覚化したホストの通信では，2 種類のスキャンが行われていた．一方のスキャンは，1 つのポートに対して，通信先の IP アドレスの第 1 オクテットを固定し，第 2 第 3 第 4 オクテットの値をランダムに変えながら行われていた．もう一方のスキャンは，第 1 第 2 第 3 オクテットを固定し，第 4 オクテットの値を 1 ずつ増やしてシーケンシャルに行われていた．複数種類のスキャンが同時に行われている際に，WIV の IP アドレス表示の縦線だけではシーケンシャルなスキャンの行われているネットワーク部の値がわかりづらいが，各オクテットにおける通信回数の変化を示した白い縦線によって，どの値に通信を行っているのかが確認できた．

また，Layered IP Matrix では，特定ネットワークに通信を行った様子と，そのネットワークの多くのホストに通信を行った結果が見えた（図 10）．2 つのホストの通信の表示を切り替えることで，双方とも同じネットワークを狙って，IP アドレスのホスト部に広くスキャンをしていたことがわかった．

その他，3 種類のスキャン時には通信回数のグラフに 2 つのホストの定期的な通信が見て取れた（図 11）．ログ情報の表示を見ることで，通信一つ一つのポート番号や時間情報を確認することができた．

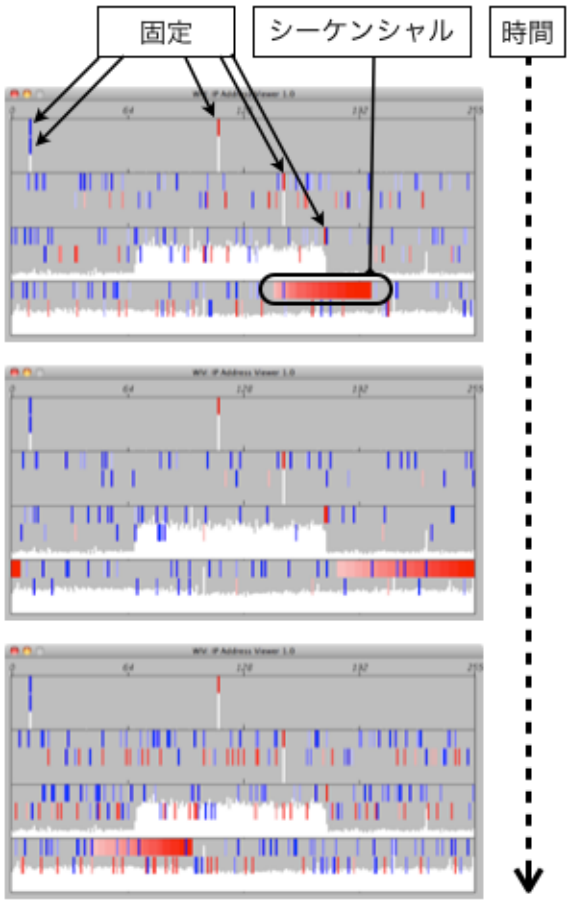

図 9: 2 つのホストからの 3 種類のスキャン．

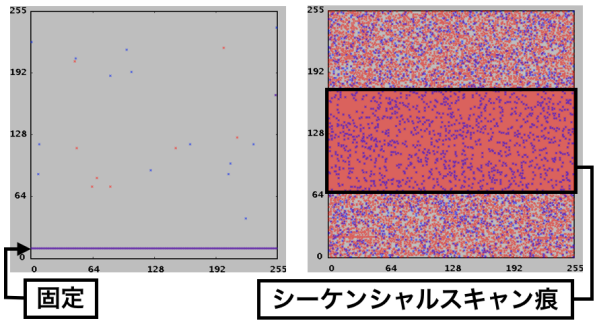

図 10: 2 つのホストの送信先 IP アドレスを Layered IP Matrix で視覚化した図．左がネットワーク部を，右がホスト部を視覚化したマップ．

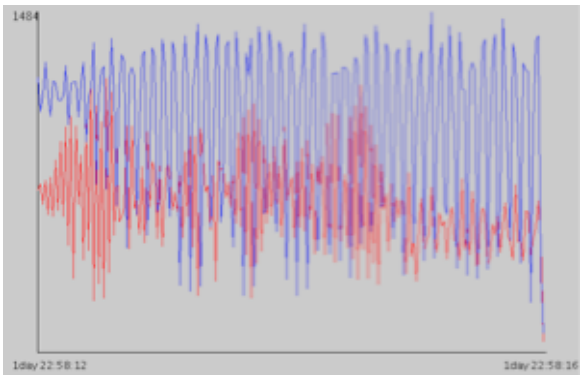

図 11: 3 つのスキャンが同時に行われている際の，2 つのホストの通信回数のグラフの変化．

4　おわりに

本論文では，マルウェアのスキャンを発見するための解析ツールを開発し，解析ツールで CCC DATAset 2010 を解析した結果を示した．解析ツールでは指定したホスト，ネットワークをそれぞれ色分けして視覚化することで，注目すべき複数のホストやネットワークに焦点を当てた解析を可能にした．

解析ツールを利用して CCC DATAset 2010 を解析した結果では，2 台のホストから 3 種類のスキャンが同時に行われていたことを容易に発見することが出来た．これは，WIV によって複数スキャンの通信先の変化を，Layered IP Matrix でそれぞれのスキャンにおける通信先の空間的広がりを，視覚的に表現できたからである．

今回我々は解析ツールをハニーネットのログ情報の解析に利用したが，解析ツールは企業などのネットワークにも利用可能である．各ネットワークやホストで色分けを行うことで，感染確認を素早く行える．また，Layered IP Matrix には，nmap のログの結果をマップに割り当てて表示する機能がある．この機能を利用して，管理しているネットワーク内のスキャン結果を表示し，ネットワーク内部への通信を視覚化することで，どのホストに通信が行われていたか調査することも可能である．多数のホストを管理しているネットワークでこの機能を利用することで，より解析しやすくなるものと考える．

解析ツールの今後の展望として，表現力の強化や特定の通信パターンへの対応が挙げられる．特定のホストやネットワークを色だけで区別して視覚化するには限界があるため，色だけでなく記号などを用いる手法が考えられる．また，マルウェアの通信に似たパターンを検知させ，視覚的，または聴覚的に知らせることで，より解析が容易になるものと考える．

謝辞

今回の論文執筆または研究にあたり，御助言くださった株式会社ラック金子博一氏に感謝の意を表します．

参考文献

[1] Inoue, D., Eto, M., Yoshioka, K., Baba, S., Suzuki, K., Nakazato, J., Ohtaka, K., Nakao, K., Nicter: An Incident Analysis System toward Binding Network Monitoring with Malware Analysis. WOMBAT Workshop on Information Security Threats Data Collection and Sharing (WISTDCS 2008), pp. 58-66, 2008.

[2] Koike, H., Ohno, K., Koizumi, K., Visualizing Cyber Attacks using IP Matrix, Proc. of Workshop on Visualization for Computer Security (VizSEC 2005).

[3] Mukosaka, S., Koike, H., Integrated Visualization System for Security Monitoring in Large-scale Local Area Network, Proc. of Asia-Pacific Symposium on Visualization (APVIS2007), pp.41-44, 2007.

[4] 清野祥之, 小池英樹, マルウェア解析のための通信視覚化, コンピュータセキュリティシンポジウム 2009, 情報処理学会, pp. 265-270, 2009.

[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とワークショップを通じた研究成果の共有，MWS2009，2009 年 10 月．