マルウェア解析のための通信視覚化

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 近年，インターネット上でマルウェアの活動による被害は拡大しており，その問題は深刻である．マルウェア対策には解析が必要であり，マルウェアの特徴的なトラフィックを観測することは解析の一助となる．本論文では，Honeynet により収集された大量の攻撃通信データから，あて先 IP アドレスを視覚化し，通信の変化の様子を動画として表示する解析ツールを開発した．解析ツールを使用することで，大量のログを効率よく動画で表示することと，マルウェアの特徴的な通信を発見することができた．

Traffic Visualization for Malware Analysis

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 In recent years, serious damages by the malware's activities on Internet are increasing. The malware analysis is required to prevent from the malware propagation. It is a set for the malware analysis to observe the malware's characteristic traffic. We developed an analysis tool that visualizes destination IP Address. This tool can show the transmission status as an animation from the Honeynet data. By using this tool, we efficiently display a lot of connections on movie, and easily discover the malware's characteristic traffic.

1　はじめに

近年，インターネット上でマルウェアの活動による被害は拡大しており，ハニーポットなどを利用したマルウェア対策のための研究が活発に行われている．マルウェアはメールによって感染活動を行う種類と，IP アドレスを基に感染活動を行う種類に大別されるが，今回我々は IP アドレスを基に感染活動を行う種類のマルウェアを対象に研究を行った．マルウェアが多数のコンピュータへと感染を拡大させるために行う攻撃は，コンピュータの通常の使用では現れにくい特徴的なトラフィックを示す．マルウェア対策にはマルウェアの解析が必要であり，マルウェアの特徴的なトラフィックを観測することは，マルウェア解析の一助となる．Honeynet により収集された攻撃通信データ CCC DATAset 2009[1] には様々なマルウェアの通信が記録されているが，テキストベースで大量の通信ログからマルウェアの通信を解析するのは，非常に時間のかかる困難な作業である．

大量の通信ログを効率よく見る手法は以前から研究されており，テキストログを解析しやすい形に視覚化する MieLog[2] や，異常な通信を

検出しやすくするために動画で視覚化する Visual Firewall[3] など，様々な方法が考案されてきた．

マルウェアの解析には，マルウェアのスキャンパターンやその通信速度などの情報が有益となるため，それらを考慮した視覚化手法が必要である．本研究では，IP アドレスを独自の手法で視覚化し，通信速度を考慮して動画で表示する解析ツールを開発した (図 1 参照)．実際にこの解析ツールを使用した結果，通信の変化の様子を動画で直感的に理解することが可能となり，攻撃通信データからマルウェアの特徴的な通信を簡単に発見することができた．

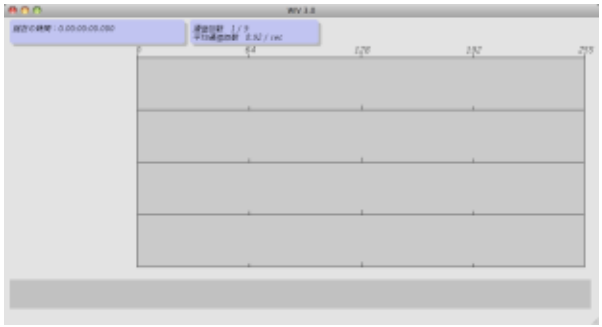

図 1: 解析ツールの概要図．

2　解析ツール

本解析ツールはテキストで記録された通信ログを読み込み，あて先 IP アドレスを視覚化し，通信の変化の様子を動画で表示する．今回我々は CCC DATAset 2009[1] の攻撃通信データを解析するにあたり，攻撃通信データを tcpdump で読み込んで出力したテキストログを使用した．以下はテキストログの出力に使用したコマンドである．

$ tcpdump -n -r 20090313.pcap > 090313.log

2.1　IP アドレスの視覚化

通信ログの IP アドレスを視覚化する際，IP アドレスは 2^{32} 個まで表現できるため，IP アドレスと画面の点 1 つを対応させて表示しようとすると，現在使用されている画面の解像度をはるかに超えた巨大な解像度が必要になってしまう．

そこで今回我々は IP アドレスを視覚化するために，表示する IP アドレスをオクテットごとに分解し，分解した 4 つの値をそれぞれに対応した枠の X 座標位置に縦線で示し，合計 4 本の縦線で 1 つの IP アドレスを表現することとした (図 2 参照)．

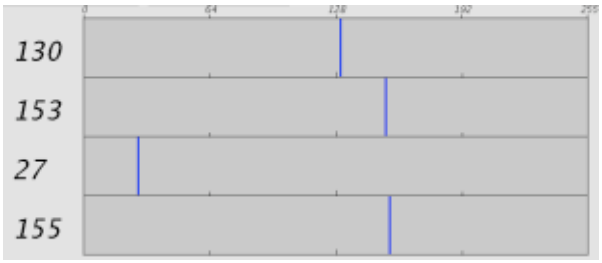

図 2: IP アドレスを視覚化している様子．IP アドレスの各オクテットを対応する枠の X 座標に縦線で示す．

2.2　特定ポートの視覚化

マルウェアは感染を拡大させるために特定のサービスへと攻撃を行うので，特定のポートへの通信を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，解析ツールにいくつかのポートを指定することにより，それらのポートを分けて視覚化することとした．指定されたポートへの通信があった場合，指定されたどのポートへの通信だったかを，IP アドレスを表現する縦線の Y 座標位置に赤色で表現することとした (図 3 参照)．

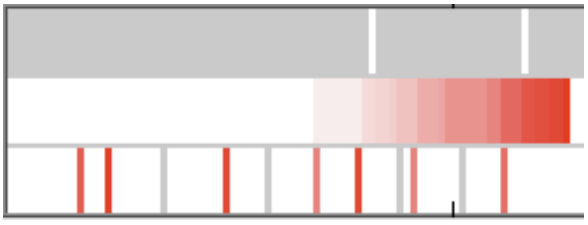

図 3: 3 つのポートを指定して視覚化した様子の拡大図．枠内を 3 つに分割して視覚化を行う．

例えば，視覚化するポートに 80，135，445 の 3 つを選んだ場合，ポート 80 への通信は各枠の上部に赤色の短い縦線で示し，ポート 135 への通信は各枠の中程に赤色の短い縦線で示し，ポート 445 への通信は各枠の下部に赤色の短い縦線で示し，その他のポートへの通信は各枠に青色の縦線で表示する．

2.3 動画表示

通信速度を考慮して動画で表示を行うために，解析ツールが通信ログから通信を行った時間を抜き出して利用し，通信の表示タイミングを調整した．その際，IP アドレスの縦線を表示し続けると，後から表示される IP アドレスがどのように表示されているのかわからなくなってしまう．そこで，表示した IP アドレスの縦線は時間の経過と共に減色させ，最終的には白い跡をアクセス痕として残すことにした．

マルウェアの攻撃が非常に速いため，動画表示の時間を現実の時間に合わせて表示を行うと，膨大な数の通信が同時に表示されてしまう．そのため，特に指定がない限り，動画表示は実際の通信速度の 100 倍の時間でゆっくり表示することとした．この動画表示の速度を指定することも可能であり，通常の通信速度と同じ速度やそれ以上の速度など，望んだ速度で動画表示することができる．

通信ログには全く通信を行わない空白の時間が存在することがあるが，マルウェアの通信の時間相関を解析する場合を除き，このような待ち時間を見つけた場合には解析ツール内部の時間を進め，次の通信をすぐに表示することとした．

2.4 補助機能

送信元 IP アドレスを指定することで，対象の IP アドレスから送られた通信のみを視覚化することを可能にした．今回使用した Honeynet の通信ログには複数の Honeypot の通信が記録されているが，この機能を使用することで特定の Honeypot からの通信のみを表示することができる．

通信先のポートを指定することで，特定のポートへの通信のみを表示可能とした．マルウェアの解析を進め，特定のマルウェアの攻撃以外の通信を表示したくない場合などに用いる．

2.5 補助表示

通信活動のその他の情報は，ウィンドウ上で補助的に表示することとした．動画の表示時間は実際の時間と異なるため，動画のウィンドウ左上部に通信ログ上の時間を常に表示することとした．また，通信が全く無い時間を自動で進めた場合にも，時間を進めたことをウィンドウ左上部に表示することとした (図 4 参照)．

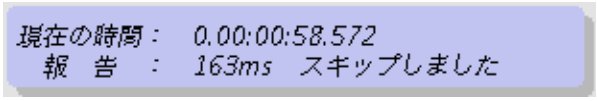

図 4: 現在のログ上での時間と，自動で時間を進めたことを表示する．

通信を表示する際に，通信が時系列に沿ってどのような頻度で行われているかをウィンドウ下部に表示することとした．ウィンドウ下部に濃い灰色で四角く塗りつぶしたフレームを準備し，通信が行われた時に縦線をフレーム内右端に描写する (図 5 参照)．縦線は時間に応じて左へと移動し，時系列表示用フレームの左端に達すると描写されなくなる．

図 5: 通信を時系列で表示している様子．

通信速度がどのようになっているかを理解するために，秒間の平均通信回数と，現在までに視覚化を行った通信回数をウィンドウ上部に表示した (図 6 参照)．

動画表示した IP アドレスの値を，数値またはログ形式で表示可能とした．数値表示では各オクテットの値をそれぞれ対応する枠の左に表

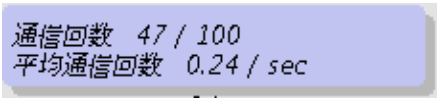

図 6: 通信回数と秒間の平均通信回数．通信回数の左側の数は表示済みの通信数で，右側の数はログから読み込み済みの通信数である．

示し，ログ表示ではウィンドウ左にログ表示用の枠を設けて枠内に記述した (図 7 参照)．

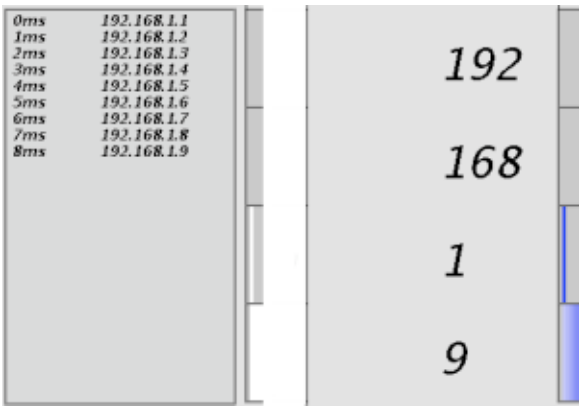

図 7: ログ表示部分と数値表示部分．

3　マルウェアのスキャンパターン

マルウェアは攻撃の最初の段階として様々なコンピュータへとスキャンを行う．このスキャンのパターンには特徴があり，スキャンパターンはマルウェアが通信先の IP アドレスをどのように選んでいくかによって 2 つに大別される [4]．

3.1　ローカルスキャン

ローカルスキャンは，IP アドレスの上位オクテットを固定し，下位オクテットを変化させてスキャンを行っていく手法である．下位オクテットの変化には，ランダムに変化させるものと，1 から 254 までを順次に変化させるものがある．この手法は，マルウェアが特定のネットワークを攻撃する時などに利用するものと考えられる．この手法において特筆すべき点として，マルウェアが感染したコンピュータの所属しているネットワークのネットワークアドレス部を利用してスキャンを行うと，感染したコンピュータと類似した環境を持つ脆弱なコンピュータに攻撃できる可能性が高く，効率よく感染を拡大させることが可能である点が挙げられる．

3.2　ランダムスキャン

ランダムスキャンは，IP アドレスをランダムに選択してスキャンを行っていく手法である．様々な IP アドレスに攻撃を行うことで，物理的に様々な場所のコンピュータへと感染を拡大させることが可能である．

4　結果

実際に解析ツールを用いて攻撃通信データを解析した結果，マルウェアのスキャンと考えられる通信を即座に発見することができた．

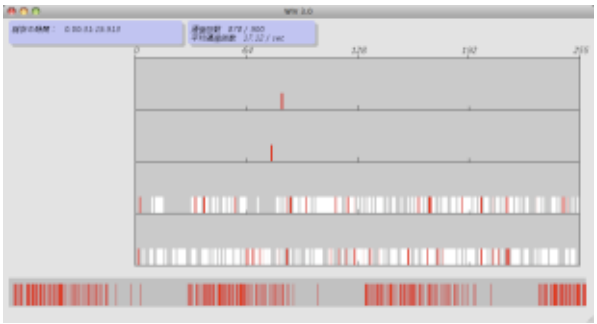

図 8: マルウェアのスキャンの様子 1.

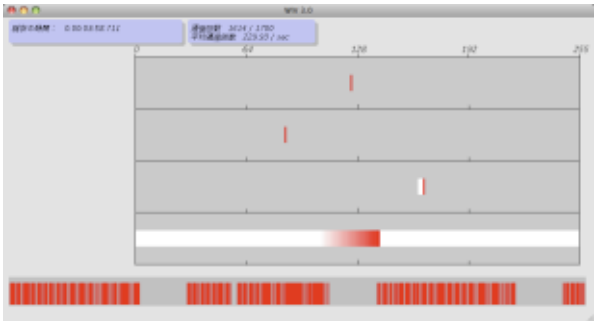

図 9: マルウェアのスキャンの様子 2.

図 8 は上位 2 オクテットを固定し下位 2 オクテットの値をランダムにスキャンしている様子

であり，図 9 は上位 2 オクテットを固定しながら下位 2 オクテットの値を 1 ずつ増加させてスキャンしている様子である．図 8 のスキャンの平均通信回数は秒間およそ 17 回であったのに対し，図 9 のスキャンの平均通信回数は非常に速く秒間およそ 220 回であった．また，図 8 と図 9 のスキャンは，それぞれ別のポートに対して違ったスキャンパターンを用いていた．上記の速度や通信先のポート，スキャンパターンの違いから，これらは別の種類のマルウェアが行ったものであると考えられる．

2 種類のスキャンが同時に行われている様子を発見することもできた (図 10，図 11，図 12，図 13 参照)．それぞれのスキャンは，上位 2 オクテットをそれぞれ違った値で固定されながら，別々のポートに対して行われていた．一方のスキャンは，短時間に第 4 オクテットの値を 0 から 255 まで 1 ずつ増やしながら行われていた．このスキャンは第 4 オクテットの値を 255 まで増やすと，第 3 オクテットの値を 1 増やし，再度第 4 オクテットの値を 0 から 1 ずつ増やすという方法で行われていた．もう一方のスキャンは第 3 第 4 オクテットの値をランダムに変えながら，比較的ゆっくりとした速度で行われていた．ポートを分けて動画で視覚化することにより，2 種類のスキャンがそれぞれ違う IP アドレスの別のポートに対して異なった速さで行われていたことを容易に判別できた．また，2 種類のスキャンがそれぞれ別の手法を用いて行われていることから，これらのスキャンは 2 種類のマルウェアがそれぞれ行ったものだと考えられる．

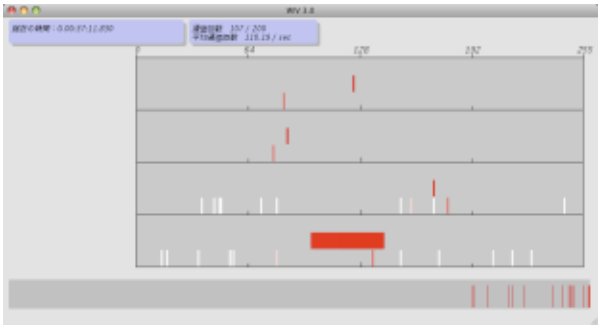

図 10: 2 種類のマルウェアが同時にスキャンしていたと考えられる様子 1.

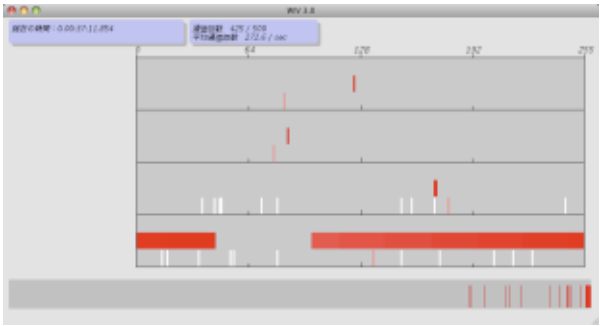

図 11: 2 種類のマルウェアが同時にスキャンしていたと考えられる様子 2.

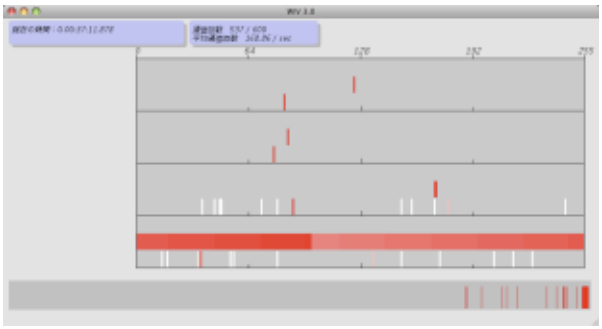

図 12: 2 種類のマルウェアが同時にスキャンしていたと考えられる様子 3.

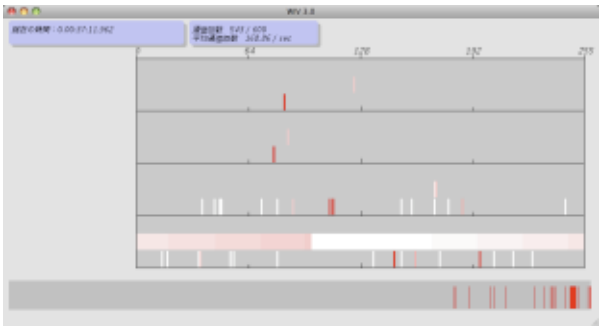

図 13: 2 種類のマルウェアが同時にスキャンしていたと考えられる様子 4.

5　おわりに

本研究では IP アドレスを視覚化して動画で表示する解析ツールを開発した．解析ツールを使用して大量の通信ログを動画表示で解析することで，通信の変化を効率よく見ることが可能となり，CCC DATAset 2009[1] の攻撃通信データの解析を行った結果では，マルウェアの特徴的な通信を簡単に発見することができた．また，いくつかのポートを別々に視覚化することで，複数のマルウェアのスキャンパターンを観測することもできた．本解析ツールは Honeynet の通信ログ以外の通信ログを解析することもできるため，企業や ISP における通信の解析においてもマルウェアの早期発見に寄与することが可能である．

6　謝辞

今回の論文執筆または研究にあたり，御助言くださった産業技術総合研究所高田哲司氏，また，本研究に協力してくださった小池研究室セキュリティ班の皆様に感謝の意を表します．

参考文献

[1] 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．

[2] Tetsuji TAKADA, Hideki KOIKE, MieLog: A Highly Interactive Visual Log Browser Using Information Visualization and Statistical Analysis, Proceedings of LISA XVI Sixteenth Systems Administration Conference, The USENIX Association, pp.133-144, November, 2002.

[3] C P Lee, J Trost, N Gibbs, Beyah Raheem, and J A Copeland, Visual firewall: real-time network security monitor, In: VizSEC: Proceedings of the IEEE Workshop on Visualization for Computer Security, pp. 129-136.

[4] 小泉 芳，小池英樹，安村通晃，低対話型ハニーポットログを活用したウイルス分布の解析と生態学的解釈，情報処理学会 コンピュータセキュリティシンポジウム 2004，pp.415-420，2004.