Vol.2012-CSEC-58 No.38
Vol.2012-SPT-4 No.38
2012/7/20

長期間のマルウェア動的解析を支援する通信可視化手法とユーザインタフェースの提案

森博志† 吉岡克成† 松本勉†

我々は，同一の検体に対してマルウェア動的解析を数週間から数か月という期間，継続的に行うことで，マルウェアを遠隔から操作する攻撃者の動向を調査することを検討している．このように長期間に渡り，解析を行う場合，観測される通信は膨大となるため，解析者の負担が大きくなる．そこで，このような長期間の観測データを短時間で把握するための通信可視化手法とユーザインタフェースを提案する．

Traffic Visualization and User Interface for Supporting Long-term Malware Sandbox Analysis

HIROSHI MORI† KATSUNARI YOSHIOKA†
TSUTOMU MATSUMOTO†

In order to investigate the behavior of attackers who remotely control malware-infected hosts, we are developing a malware analysis environment where malware sample can be run and monitored for a long period such as several weeks or months. When we let a sample run for that long period, the traffic it creates becomes huge and diverse. Thus, we propose a new user interface and traffic visualization method that can help the human analyst with the burden of analyzing the huge traffic.

1. はじめに

近年のサイバー攻撃は，ボットネットや標的型攻撃に代表されるようにマルウェアに感染したホスト群を攻撃者が遠隔から操作することで行われる場合が多い．我々は，同一のマルウェア検体を長期間解析環境内で動作させ，攻撃者との通信を観測することで，攻撃者の振る舞いの観測を行う技術の研究開発を進めている．

本稿では，このような長期動的解析で得られる大量の通信の内容を解析者が効率的に把握するためのユーザインタフェースを提案する．長期動的解析により得られる通信データは短時間解析に比べて膨大である．一例として我々が Spybot [1]を 1 週間動作させた際には，リモートエクスプロイト攻撃のように短時間で大量のセッションを確立する通信は観測されなかったにも関わらず，通信データサイズは 1GB 以上となっている．このように膨大な通信データを解析し，目的に応じて必要な情報を得るのは容易ではない．

提案手法では，解析の起点となる通信データのサマリをグラフビューおよび世界地図ビューにより簡略表示する．グラフビューでは宛先ポート毎の通信量の時間推移がグラフ表示され各ポートへの通信の時間推移を概観することができる．一方，世界地図ビューでは，IP アドレスから得られる地理情報に基づき宛先を表示することで，当該マルウェアのアクセス先を瞬時に把握できる．これら 2 つのビューは連動しており，一方に対する操作が他方にも反映されるようになっている．例えば，グラフビュー上で特定のポートを選択表示したりグラフ中の特定の期間をマウスでドラックすることで当該期間にフォーカスすることが可能であるが，この操作に対応して指定されたポートおよび期間の通信だけが世界地図ビュー上で表示される．逆に世界地図ビュー上で特定のホストを指定した場合には，対応する通信がグラフ上で強調表示されるようにした．また操作のリアルタイム性を確保するため，各ビューの表示に必要な情報のうち事前計算可能なものは予め計算しておくことで操作時に生じる計算処理によるオーバーヘッドを削減した．本稿の構成は次のとおりである．2 章では，まず通信データを可視化する関連研究を挙げ，3 章では提案手法について述べる．4 章では提案手法を実装した通信可視化ユーザインタフェースについて説明し，5 章では 4 章で説明したインタフェースによる解析例を示す．最後に，6 章でまとめとする

2. 関連研究

マルウェアの通信を可視化する先行研究には文献[2]や nicter[3]がある．文献[2]では複数の場所に設置されたハニーポットによる観測データを用いてサイバー攻撃の可視化を行う手法が提案されている．

図 1 は文献[2]の提案手法によってマルウェアの攻撃を可視化したものである．本手法ではポイントマップと呼ばれる，点による可視化手法とプリズムマップと呼ばれる，境界（この例では国の境界）ごとに 3D で表示する可視化手法によりどの国からどのくらいの量の攻撃が来ているのか，

† 横浜国立大学
Yokohama National University

ⓒ2012 Information Processing Society of Japan

またどの地点のハニーポットが攻撃の対象となっているのかを知ることができる．

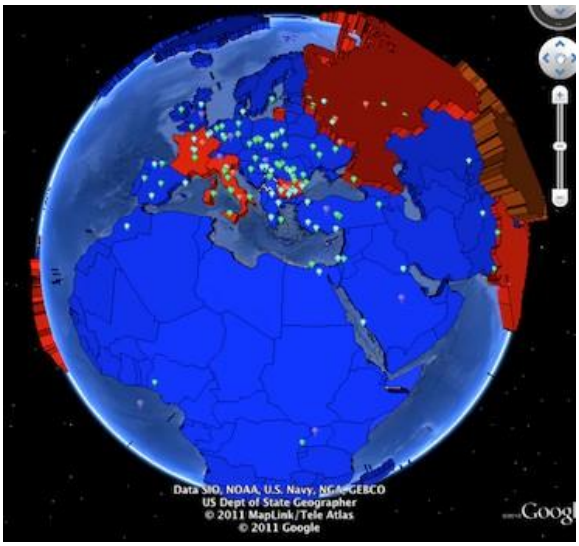

図 1　文献[2]の提案手法によるマルウェアが行う攻撃の可視化の例

nicter は情報通信研究機構で研究開発されているシステムでダークネットと呼ばれる未使用の IP アドレス空間に届く通信を観測することでネットワーク攻撃の観測を行う．図 2 は Atlas と呼ばれる nicter の可視化エンジンのひとつである．Atlas では，世界中から日本各地にあるダークネットに対して送信される通信を可視化することでマルウェアの解析を行う．図 2 で赤や青や黄色で表現されている物体が通信データで，仰角が宛先ポートを，色が通信の種類を表している．

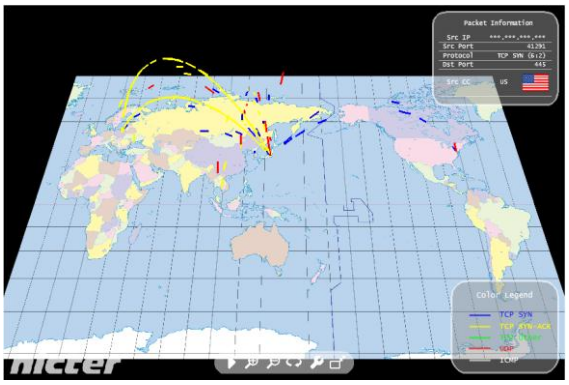

図 2　nicter の Atlas による通信の可視化

本稿で提案する手法と以上に述べた先行研究は，

- マルウェアが行う通信を可視化することを目標としている
- 位置情報を利用して世界地図上に可視化を行なう

という 2 点において似ているが，上記の 2 件の研究はマルウェアから送られてくる通信データの解析を行なっているのに対し，本稿で提案する手法は解析環境下でマルウェアが送信する通信データの解析を目的とする．また，我々の提案手法では世界地図による表示だけでなく，マルウェアが行う通信の時間推移を表示するグラフビューを連動させて操作することでより直感的に解析作業を進めることができる．

3.　長期間の通信を可視化する手法の提案

提案手法ではマルウェアが行う通信の時間推移を宛先ポート毎に示すグラフビューとマルウェアの通信先などの情報を示す世界地図ビューにより解析者を支援する．

提案手法は以下のような特徴を持つ．

- 通信ポート番号毎に通信量の時間推移をグラフで表示する（グラフビュー）
- 通信先ホストを位置情報で分類して地図上に表示する（世界地図ビュー）
- 表示する情報の内容が異なる 2 つのビューを同時に表示し，それらのビューを連携して操作できる．

図 3 は提案手法を実装したユーザインタフェースの画面である．

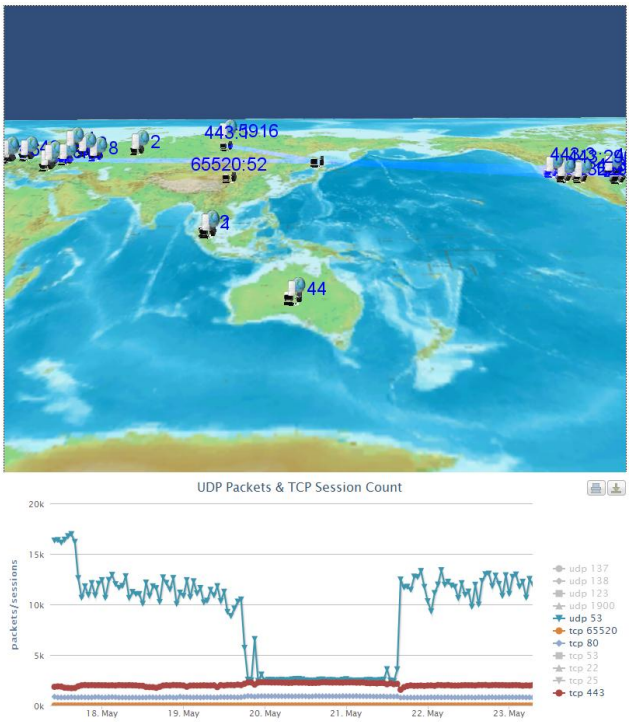

図 3　提案ユーザインタフェースの表示例

3.1　動的解析環境

本節では次節で述べる提案インタフェースにおいて可視化の対象となるマルウェアの通信データを観測するための動的解析環境について説明する．これまで文献[4]など，様々なマルウェア動的解析手法が検討されているが，提案インタフェースでは図 4 のような構成の動的解析システム

によって得られる通信データを可視化の対象とする．

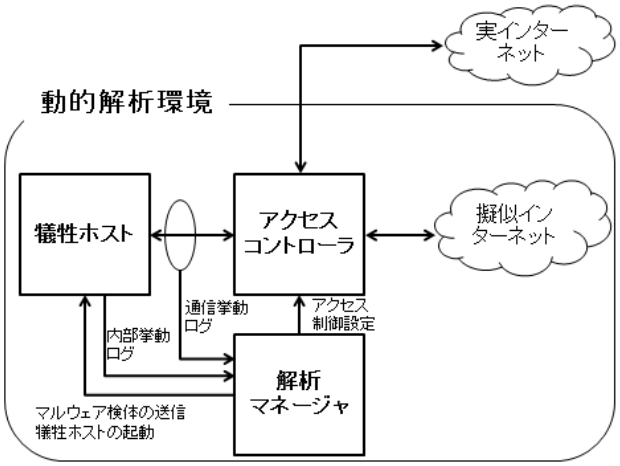

図 4　動的解析システムの構成

図 4 の動的解析システムでは，解析のためにマルウェアを実行する犠牲ホストの通信をアクセスコントローラによって制御する．アクセスコントローラは解析マネージャによって設定されたルールに従い，犠牲ホストが行う通信のうち安全なものは実インターネットに通すが，リモートエクスプロイトなどの攻撃性のある通信は実インターネットには通さずに代わりに実インターネットを模擬した擬似インターネットに転送する．擬似インターネットは実インターネット上に存在するサーバやホストを模擬したダミーサーバによって構成されており，攻撃性のある通信を実インターネットに通すこと無く解析ができるようになっている．

提案手法では以上に述べた動的解析環境下で犠牲ホストを観測し，その際に得られる通信挙動ログ（pcap ファイル）を可視化の対象とする．

3.2 提案手法

3.2.1 グラフビュー

観測した通信データから犠牲ホストが行った通信のうち TCP/UDP 通信に関して宛先ポート別に分類を行いそれぞれ単位時間当たりの通信量を TCP 通信に関してはセッション数で，UDP 通信に関してはパケット数でカウントし，宛先ポート毎の通信量の時間推移を折れ線グラフとして描画する．また，外部からセッションが確立された通信に関しては犠牲ホスト側のポート番号で分類を行い，犠牲ホストがセッションを確立した通信と区別がつくように表示する．

グラフビューの一例としてある通信データに対する可視化結果を図 5 に示す．図 5 上段のグラフはグラフビュー表示の初期状態である．この状態から特定期間に絞り込んでグラフを表示したい場合は，当該期間をグラフ上でドラッグにより指定することで，1 回のマウス操作で簡単にドリルダウンできるようにしている．図 5 の中段のグラフは，上記の方法により上段グラフの水色の部分を絞り込んで表示したものである．またグラフ右側に配置された凡例のポート番号をクリックすることで当該ポートのグラフの表示・非表示を切り替えることができる．（図 5,下段）の下段のグラフは中段のグラフを見る際に邪魔になるポートのグラフを非表示にした例である．

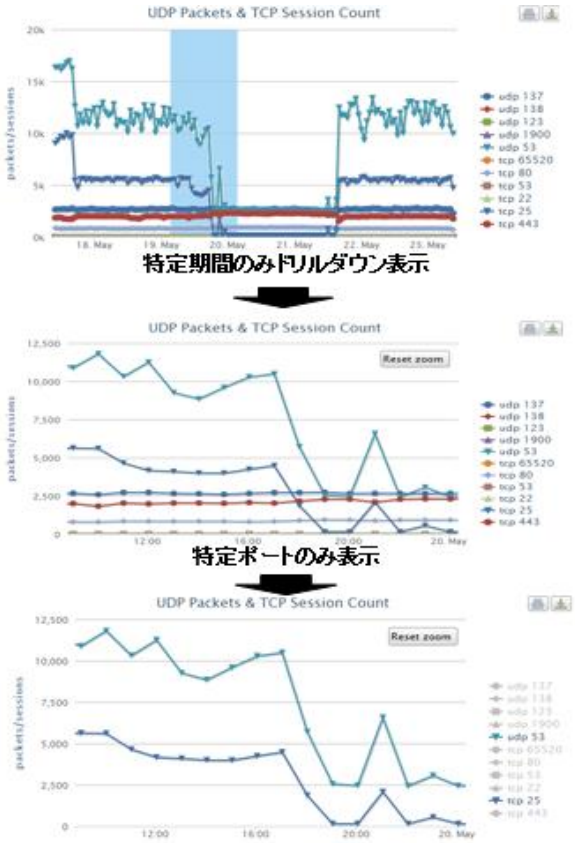

図 5　グラフビューによる表示の例

3.2.2 世界地図ビュー

世界地図ビューでは観測した通信データと MaxMind 社の GeoIP と呼ばれる Group データベース（IP アドレスとそれに対応するホストの位置情報のデータベース）を元に，犠牲ホストとその通信先ホストのシンボルを世界地図上に描画する．ローカルネットワークのホストは犠牲ホストのシンボルの上空に表示する．また 3.1 節で述べたとおり，マルウェア動的解析環境では，マルウェアが動作している犠牲ホストと実際のインターネット上の実ホストとの通信を許可する場合と，実ホストとの通信を許可せず，ダミーサーバへ通信を転送する場合があるが，これを視覚的に区別するためインターネット上の実ホストへの通信は世界地図上の地表付近に色つきのホストシンボルを表示し，一方，ダミーサーバとの通信の場合は，当該地点の上空に黒塗りのホストシンボルを表示することとした（図 6）．

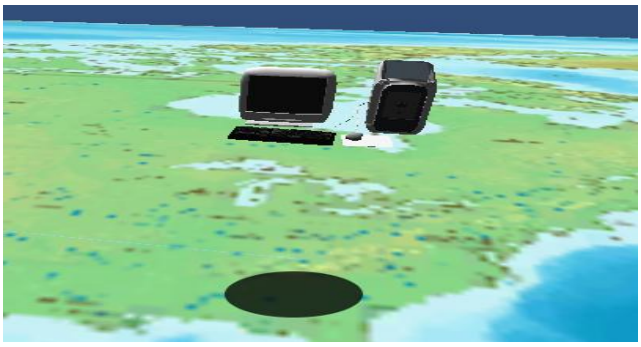

図 6　ダミーサーバの表示例

また犠牲ホストと通信先ホストのシンボル間は両ホスト間でやり取りされたパケット数に比例する太さの線で結ぶことで通信量の把握を容易にする．

さらに，マルウェアの通信先ホストのシンボル付近には通信内容を示すシンボルとその通信量（TCP 通信であればセッション数，UDP 通信であればパケット数）を同時に表示する．図 7,8 は世界地図ビューの一例である． HTTP 通信は，www をイメージした地のアイコン，DNS 通信は単純に”DNS”と記載したアイコン， SMTP 通信をメールアイコンで表示している．

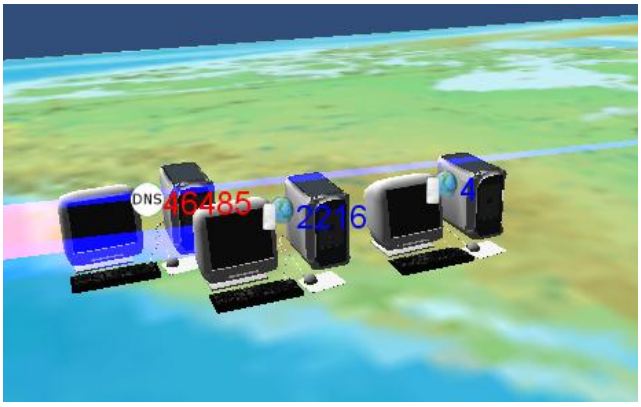

図 7　世界地図ビューによる表示の例 1

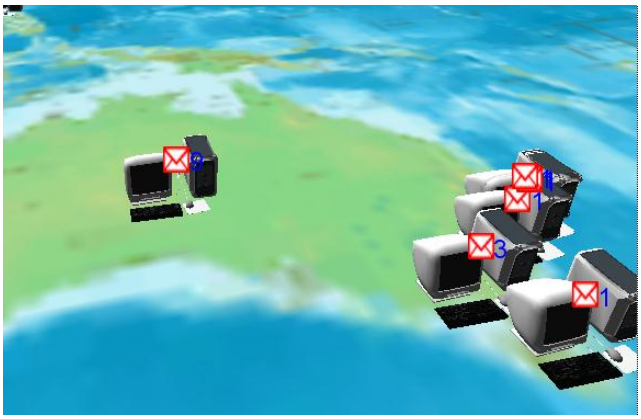

図 8　世界地図ビューによる表示の例 2

3.2.3 ビュー間の連携

上記で述べたビューは互いに協調して動作する

(1)　グラフビューの操作による世界地図ビューへの連携

3.2.1 項で述べた通り，グラフビューをドラッグすることで世界地図ビューの表示内容をドラッグした期間に絞ることができる．また，グラフビューで非表示にしたポート番号に関する情報は世界地図ビューでも非表示になる．

(2)　世界地図ビューの操作によるグラフビューへの連携

世界地図ビューでは，表示されるマルウェアの通信先ホストのシンボルをクリックすると対応するホストの IP アドレスを指定することができる．世界地図ビューで IP アドレスの指定が行われると，グラフビューにその情報が反映され，マルウェアが行う当該 IP アドレスへの通信量が追加表示される．

4.　提案手法を実装したユーザインタフェース

提案手法を導入したユーザインタフェースを Web アプリケーションとして実装した．Web アプリケーションとすることで，グラフビュー，世界地図ビューの表示に必要な通信データの処理をすべて Web アプリケーションサーバに任せることができ，解析者は Web ブラウザと 3D 描画に利用するプラグインさえインストールしていればすぐに解析を始めることができる．また，通信データを一度サーバにアップロードしておけば pcap ファイルが手元に無くても解析を行うことができる．図 9 は提案ユーザインタフェースを含むマルウェア動的解析システム(以下，単に本システムと呼ぶ)の全体図である．

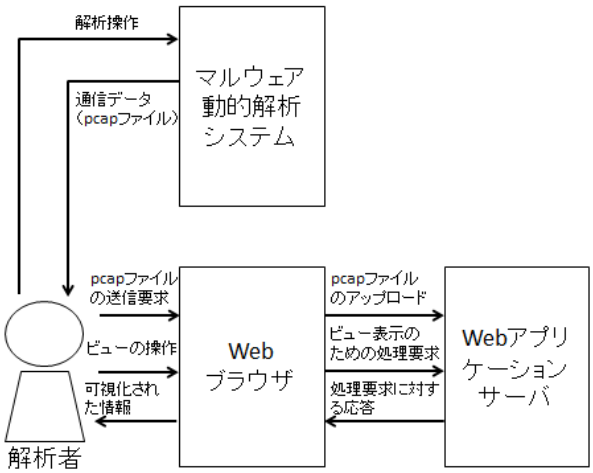

図 9　提案ユーザインタフェースを含むマルウェア動的解析システムの全体図

Web アプリケーションとして提案手法を実装するとユーザインタフェースの利用が容易になる反面，可視化処理を行なう際に pcap ファイルの解析に加え，通信処理も発生するためユーザインタフェースのリアルタイム応答性低下の要因になる．また，ユーザインタフェースの利用可能な資

ⓒ2012 Information Processing Society of Japan

源が Web ブラウザに制限されるデメリットもある．

4.1 Web アプリケーションサーバの構成

図 10 は本システムの Web アプリケーションとサーバの構成である．解析者が操作する Web ブラウザからのリクエストを Web フレームワークが受け，Web ブラウザへの応答に必要な処理を行う．

図 10　本システムの Web アプリケーションサーバの構成

4.2 提案ユーザインタフェースのリアルタイム応答性

本システムでは，特定期間の通信状況の表示などで，サーバに保存されている pcap ファイルに対する処理が必要になった場合，サーバ側で pcap ファイルに対する処理を行いその結果を解析者が操作する Web ブラウザに送信する必要がある．このためリアルタイム応答性の低下が懸念される．そこで通信データを可能な限り事前に解析しておくことで可視化の際に必要な処理を減らす．

4.2.1 事前処理

本システムでは pcap ファイルがサーバにアップロードされた段階でサーバが①～④の項目の処理を行う．

① pcap ファイルを 1 時間ごとに分割する
② TCP 通信の情報をセッションごとにまとめる
③ 宛先ポート番号別の通信量の時間推移の計算を行う
④ 1 時間ごとの宛先ポート番号別の通信量の計算と通信先ホストの IP アドレスに対する位置情報の取得

①では長期間の動的解析で得られる膨大な通信データを効率良く処理できるように tcpslice を使用して通信データを 1 時間ごとに分割する．
②では①によって分割されたそれぞれの pcap ファイルに対して tcpflow を実行し，TCP 通信の情報をセション単位でまとめ，report.xml というファイル名で保存する．また tcpflow はセッションごとにペイロードを再構築するためセッション単位でのペイロードの抽出が容易になる．
③ではグラフビューの表示に必要な宛先ポート番号別の通信量の時間推移を計算する．
④では①で分割した通信データに対してそれぞれ宛先ポート番号別の通信量と通信先ホストの IP アドレスに対する位置情報の計算を行う．
図 11 では処理①，②の流れを示している．

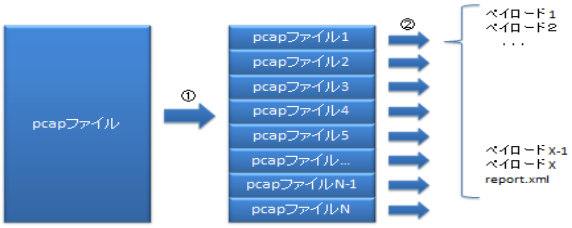

図 11　通信データの処理の流れ

4.2.2 事前処理したデータを利用した応答

本項では前項で述べた事前処理を利用して解析者の操作に対してどのように応答を行うのかを述べる．

(1) グラフビューの表示

グラフビューの表示には前項で挙げた処理③の結果をそのまま利用することで，オンデマンドでの計算を一切することなく Web ブラウザに情報を送信する．そのためグラフビューの表示に関するリアルタイム応答性は高いと言える．

(2) 世界地図ビューの表示

提案手法では解析者がグラフビューから指定した任意の期間におけるマルウェアの通信の様子を世界地図ビューに表示することで解析を支援する．

世界地図ビューの表示には指定された期間にマルウェアが行う通信の宛先ポート番号別の通信量と通信先ホストの位置情報の取得が必要となる．解析者によって指定される期間は任意なので世界地図ビューの表示に必要となる情報を全て事前計算しておくことは難しいが，前項で述べた処理④の結果を利用することである程度処理を削減できる．次に，処理④の結果を利用した世界地図ビューの表示方法を述べる．

解析者の期間の指定は大きく分けて 2 つに分類できる．

（ア）1 時間未満の期間の指定

解析者によって指定された期間が 1 時間未満だった場合，処理④による結果を利用することは難しい．そのため指定された期間の情報をオンデマンドで処理する必要がある．しかし期間が 1 時間未満であるため処理する情報量はそれほど多くなくリアルタイム応答性にあまり影響しない．

（イ）1 時間以上の期間の指定

期間の指定が 1 時間以上であった場合は処理④の結果が利用できることがある．例えば解析者によって図 12 のように期間が指定された場合，図中の赤色の期間に関してはオンデマンドで情報を処理する必要があるが，緑色の期間に

関しては処理④の結果を利用することができるので計算する必要がない．このため（イ）のケースにおいてオンデマンドで情報を処理する必要のある期間は 2 時間未満である．

上記で述べたことから世界地図ビューを表示する際に必要となるオンデマンドでの計算処理は 2 時間未満の期間に対する処理に抑えられると言える．

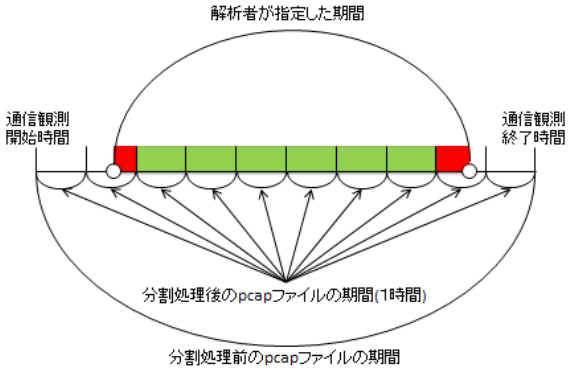

図 12　期間指定の例

5. 提案ユーザインタフェースを用いた解析例

本章では提案手法を実装したユーザインタフェースを用いて通信を解析した実例を述べることで提案手法の有用性を示す．

解析に用いる通信データは Spybot の亜種をを 5 月 17 日 9 時から 5 月 23 日 9 時まで継続的に実行して得られた通信で，総パケット数は 7,609,502 パケット，データサイズは 1.1GB である．

(1) 通信データのアップロードと事前処理

提案インタフェースを用いて解析を行うにはまず Web ページを通して通信データをアップロードする必要がある．通信データがアップロードされるとサーバは可視化を効率良く行うために当該データを解析し事前計算を行う．通信データをアップロードして事前計算を完了するまでにある程度時間がかかるため，本インタフェースを利用する場合，通信データは逐次サーバにアップロードし，予めこれらの処理を済ませておく．

(2) グラフビューによる通信の把握

まず，グラフビューによる解析期間全体の通信の概要を示す(図 13).

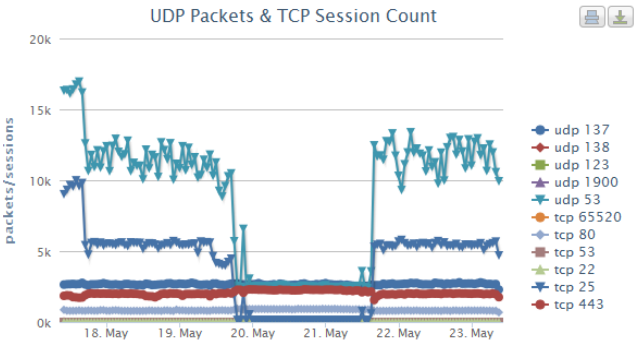

図 13　解析期間全体の通信の概要

なお，縦軸は TCP については 1 時間当たりのセッション数，UDP については 1 時間当たりのパケット数である．グラフビューから，Spybot に感染したホストは少なくとも解析した期間内ではたかだか 11 種類の宛先ポート番号に対してしか通信を行わなかったことが分かる．

次にこれら 11 種類のポートについて通信量の時間推移をみると TCP65520 番ポートが最も特徴的であることがわかる（図 14）．すなわち，当該検体の実行後 1 時間に 1 回の定常的な通信を続けているが，5 月 19 日 12 時頃突如通信を行わなくなっている．なお，世界地図ビューとの連携により，当該通信は中国のサーバと行っていることがわかる．さらに通信内容を調べると明らかに IRC であり，このサーバが C&C サーバである可能性が高いことが分かる（図 15）．

図 14　TCP65520 番ポートへの通信の概要

```
6.545267 IP 114.112.255.81.65520 > 192.168.228.240.1041: P 1:233(232
 0x0000:  4500 0110 6d0e 4000 2f06 c67e 7270 ff51  E...m.@./..~rp.Q
 0x0010:  c0a8 e4f0 fff0 0411 afad 0de0 8ed5 ac39  ...............9
 0x0020:  5018 16d0 4bf3 0000 3a75 2e20 5052 4956  P...K...:u..PRIV
 0x0030:  4d53 4720 6d67 6862 7a75 7661 203a 2167  MSG.mghbzuva.:!g
 0x0040:  6574 2068 7474 703a 2f2f 7370 2e67 6875  et.http://sp.ghu
 0x0050:  7261 2e70 6c2f 7275 732e 7068 700d 0a3a  ra.pl/rus.php..:
 0x0060:  752e 2050 5249 564d 5347 206d 6768 627a  u..PRIVMSG.mghbz
 0x0070:  7576 6120 3a21 6765 7420 6874 7470 3a2f  uva.:!get.http:/
 0x0080:  2f6b 6572 6b65 7274 2e63 6f6d 2f30 3038  /kerkert.com/008
 0x0090:  6463 2e74 7874 0d0a 3a75 2e20 5052 4956  dc.txt..:u..PRIV
 0x00a0:  4d53 4720 6d67 6862 7a75 7661 203a 2167  MSG.mghbzuva.:!g
 0x00b0:  6574 2068 7474 703a 2f2f 6d61 736d 6573  et.http://masmes
 0x00c0:  642e 636f 6d2f 7465 6d70 2f73 7364 642e  d.com/temp/ssdd.
 0x00d0:  6578 650d 0a3a 752e 2050 5249 564d 5347  exe..:u..PRIVMSG
 0x00e0:  206d 6768 627a 7576 6120 3a21 6765 7420  .mghbzuva.:!get.
 0x00f0:  6874 7470 3a2f 2f39 312e 3232 302e 3335  http://91.220.35
 0x0100:  2e32 372f 7465 6d70 2f31 2e65 7865 0d0a  .27/temp/1.exe..
```

図 15　65520 番ポートの通信の内容の例

また，TCP25 番ポートに注目すると実行後，一時的に通信量が急増し，その後，5,000 セッション/時程度で定常的に通信を行っていることが分かる(図 16).

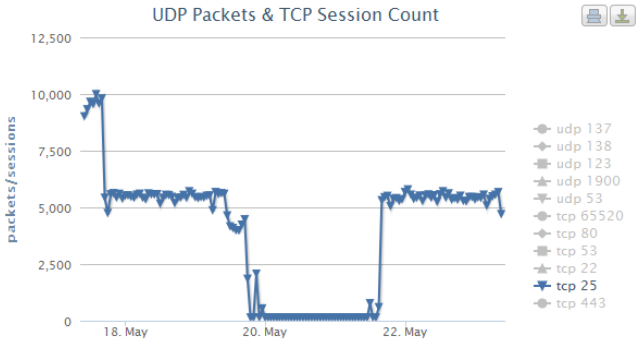

図 16　TCP25 番ポートへの通信の概要

世界地図ビューにより，宛先は世界中に広がっていることから，当該通信はスパムメールである可能性が高いことが分かる(図 17).

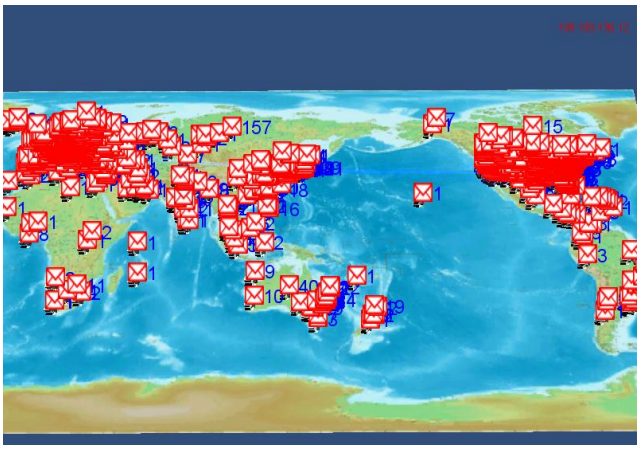

図 17　TCP25 番ポートのみを表示した世界地図ビュー

図 18 に TCP25 番ポートにおける，あるセッションの通信内容を例として示す．実際に SMTP によりメールが送信されていることを確認できる．

```
220 bgag500556.tk.mesh.ad.jp ESMTP Sendmail shby/4714030211; Fri, 18 May 2012 15:34:19
+0900
HELO FL1-110-233-151-254.kng.mesh.ad.jp
250 bgag500556.tk.mesh.ad.jp Hello FL1-110-233-151-254.kng.mesh.ad.jp
[110.233.151.254], pleased to meet you
MAIL FROM: <[redacted]@mtlo.co.jp>
250 2.1.0 <[redacted]@mtlo.co.jp>... Sender ok
RCPT TO: <[redacted]@mtlo.co.jp>
250 2.1.5 <[redacted]@mtlo.co.jp>... Recipient ok
DATA
354 Enter mail, end with "." on a line by itself
To: <[redacted]@mtlo.co.jp>
Subject: [redacted]@mtlo.co.jp . BREITLING Inc . Discount-87037
From: <[redacted]@mtlo.co.jp>
MIME-Version: 1.0
Content-Type: text/html;
      charset="UTF-8"
Content-Transfer-Encoding: quoted-printable

<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN" "http://www.w=
3.org/TR/REC-html40/loose.dtd">
<html>
  <head>
```

図 18　TCP25 番ポートの通信内容の例

また，UDP53 番ポートの通信に注目すると，その時間推移は TCP25 番ポートへの通信とほぼ同期しており，DNS 通信の大半は上記のスパム送信のための宛先アドレスのドメイン名前解決のために行われていることがわかる（図 19）.

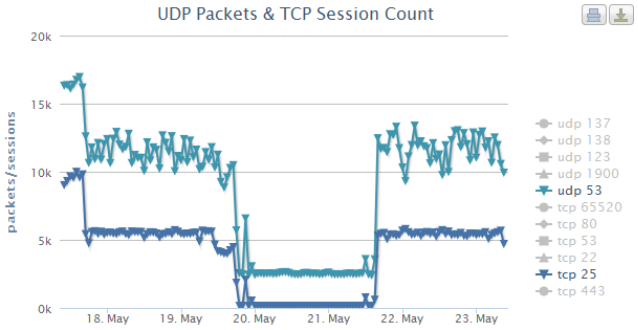

図 19　UDP53 番と TCP25 番ポートへの通信の概要

また，スパム送信は上記 C&C サーバとの通信が行われなくなった 5 月 19 日の午後から一時的に止まっているが，その後 5 月 22 日頃に再開し，5000 セッション/時程度でスパム送信を続けていることが分かる．

さらに，TCP 443 番ポート(HTTPS)や TCP 80 番ポート(HTTP)に注目すると，ほぼ定常的に通信を行っているものの，上記のスパム送信の停止時期に微増し，再開の時期に微減していることが分かる（図 20）.

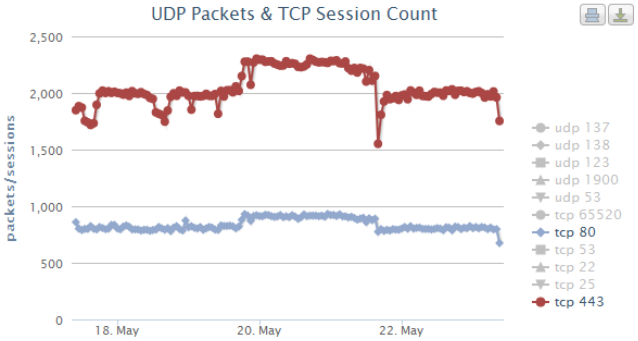

図 20　TCP80,443 番ポートへの通信の概要

上記から，この 2 つの時期に攻撃者からの何らかの命令が届き，それに従いマルウェアの挙動が変化したことはほぼ間違いないといえる．このことから解析者はこの 2 つの時

期に注目して攻撃者から命令を探索するといった解析に進むことができる．

以上の解析例では，本ユーザインタフェースを利用することで Spybot に感染したホストが行う大量かつ多様な通信の概要と比較的容易に把握し，攻撃者からの命令が行われた可能性のある時間帯を絞り込む作業のサポートができたといえる．

6. まとめと今後の課題

本稿では長期間に渡るマルウェアの動的解析で得られた通信を解析者が効率的に把握することができる通信可視化手法について提案し，提案手法の実装例を示した．また，実装したユーザインタフェースによる解析例を説明した．Spybot の数週間における動的解析では，これまで我々が行ってきた数分間程度の短い実行時間での動的解析とは比較にならないほど多くの挙動が観測できた．一方，解析者が解析すべき情報が膨大になるため，本稿で提案したユーザインタフェースによる支援は，解析者にとって有用な解析支援ツールに成り得ると考える．

今後は，他のマルウェア検体を含め解析の事例を増やし，必要な機能の拡充を行う予定である．特に，長期的なマルウェア動的解析では，1 つの検体に解析環境が長期間占有されるため，解析環境を並列化し，同時に多数の検体の解析を行うことが想定される．このような場合に，多数の検体の挙動を迅速に把握するためのインタフェース構築も今後の課題である．

謝辞　本研究の一部は，総務省情報通信分野における研究開発委託／国際連携によるサイバー攻撃の予知技術の研究開発／サイバー攻撃情報とマルウェア実体の突合分析技術類似判定に関する研究開発により行われた．

参考文献

1) Symantec W32.Spybot.Worm, http://www.symantec.com/ja/jp/security_response/writeup.jsp?docid=2003-053013-5943-99&tabid=2
2) 金子博一: 地理的可視化を用いたマルウェアの統合解析,CSS2011(2011).
3) 中尾康二,松本文子,井上大介,馬場俊輔,鈴木和也,衛藤将史,吉岡克成,力武健次,堀良彰: インシデント分析センタ nicter の可視化技術,ISEC,Vol.106,No.176,pp.83-89(2006).
4) KATSUNARI YOSHIOKA, TSUTOMU MATSUMOTO: Multi-Pass Malware Sandbox Analysis with Controlled Internet Connection, IEICE Trans, E93A, No.1, pp.210-218, (2010).

ⓒ2012 Information Processing Society of Japan