SCIS 2007 The 2007 Symposium on Cryptography and Information Security
Sasebo, Japan, Jan. 23-26, 2007
The Institute of Electronics, Information and Communication Engineers

定点観測によるボットネットの挙動観測とログ情報の視覚化
A Development of the BOTNET monitoring system and data visualization

堀合 啓一*† 大橋 洋一* 朝長 秀誠† 田中 英彦†
Keiichi Horiai Yoichi Ohashi Shusei Tomonaga Hidehiko Tanaka

あらまし 近年，ボットネットによるスパムメールの大量送信やDDoS攻撃，情報の奪取などの不正行為が問題となっている．ボットネットは，従来のワームやウィルスのように自動的に感染を拡大せず，Herderと呼ばれる攻撃者からの指令を受けて活動するため，その実態の把握が難しいといわれている．本研究は，定点観測のセンサーとしてハニーポットを利用したシステムを構築し，定点観測で得られたログ情報を視覚化することでボットネットの実態把握を試みたものである．観測データを視覚化することによって，観測サイトと攻撃元ＩＰアドレスの関連性，攻撃ペイロードの変化などを直感的に把握可能となった．また捕獲したMalwareを模擬環境で実行した際の挙動を解析し，ボットネットの影響を緩和するための対策案を提案する．

キーワード 定点観測，ボットネット，ハニーポット，視覚化

1 はじめに

近年，ボットネットによるスパムメールの大量送信やDDoS攻撃，情報の奪取などの不正行為が問題となっている．ボットネットとは，一種のバックドアを埋め込まれた多数のPCで構成されるネットワークの総称であり,現在では,多くの場合IRC (Internet Relay Chat)の仕組みを利用して指令を受け制御されている．「IPアドレスの2%～2.5%程度がボットに感染している」との調査結果[1]もあり，ボットネットが大規模なDDoS攻撃に利用された場合には，インターネットの利用に対して甚大な被害が発生する可能性がある．一方，ボットネットは，従来のワームやウィルスのように自動的に感染を拡大せず,Harderと呼ばれる攻撃者からの指令を受けて活動するため，その実態の把握が難しいといわれている．

本研究は，定点観測のセンサーとしてハニーポットを利用したシステムを構築し，定点観測で得られたログ情報を視覚化することでボットネットの実態把握と基本的な対策案の導出を目指している．ボットネットの挙動を把握するためには，固有のIPアドレスを付与した，できるだけ多数のセンサーをインターネットへ接続する必要がある．このため本研究では，仮想OSの一種であるVMware[2]を利用することにより，1台のPC上に複数個のゲストOSを稼動させ，それぞれのゲストOS上でハニーポット用のソフトウェアを定点観測のためのセンサーとして利用した．また，収集したログは，正規化(Normalize)，集約化(Aggregate)のステップを経て視覚化(Visualize)の処理を行うことにより，時系列表示，マトリクス表示，表形式表示など用途に応じて多様な表示が可能となった．

2 システムの概要

本研究では，将来的に多数の観測サイトからデータを集めて処理することにより，ボットネットの状況を常続的に把握することを一つの狙いとしている．ただし，当面利用可能なネットワーク等の制約から，今回は3つの拠点を観測サイトとして稼動させた．以後，

*防衛庁技術研究本部電子装備研究所 〒154-8511, 東京都世田谷区池尻 1-2-24, ESRC TRDI JDA, 1-2-24 Ikejiri Setagaya Tokyo
†情報セキュリティ大学院大学情報セキュリティ研究科,〒221-0835, 神奈川県横浜市神奈川区鶴屋町 2-14-1, INSTITUTE of INFORMATION SECURITY, 2-14-1 Tsuruya-cho Kanagawa-ku Yokohama Japan

3 個の観測サイトは A,B,C のラベルで区別する.

観測サイトはすべて固定 IP アドレスであり，観測サイト A,B は 1 個，観測サイト C は 2 個ブロックでそれぞれ 4 個及び 5 個の IP アドレスにセンサーを設置した．定点観測システムの構成を図 1 に示す．Whois 及び DNS サーバは，必要に応じて攻撃元 IP アドレスに関する情報の問合わせに利用した.

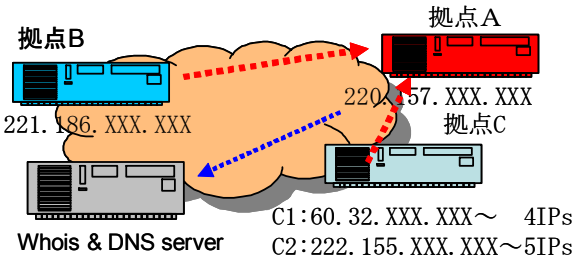


図 1　定点観測システムの構成

2.1　センサーの構成

センサー及びログの処理には，入手が容易な一般的な PC を利用し，OS として Linux を使用した．複数の IP アドレスを利用できる環境では，VMware を利用して 1 台の物理マシンへ複数のゲスト OS をインストールし，最大 5 個のセンサーを 1 台で実現した．センサーとしては，ミディアム・インタラクション型のハニーポットの一種である Nepenthes[3]と IDS として広く利用されている Snort[4]を使用した．Nepenthes は Windows の脆弱性を模擬し，攻撃を受けると受け側の応答をエミュレートして攻撃のパケットを記録する．この攻撃が，あらかじめ用意されている応答手順に含まれる場合には，さらに Malware の本体（ペイロード）をダウンロードする．Nepenthes では作動状況を示すログの記録に加えて，ダウンロードしたペイロードがバイナリ・データとして得られる．複数個設置した観測サイトから，センサーが記録したログ情報等を観測サイト A へ転送・蓄積し，ここで一括してログの処理を行った.

これらのセンサーはインターネットへ接続していることから，攻撃を受けて攻略されるリスクが存在し，特にハニーポットの場合には，このリスクは少なくない[8]．システムの安全性確保には十分な対策が必要でありこのため，観測に不要なポートの閉塞やセキュリティ・パッチの適用を確実に実施し，安全性の確保に配慮した．センサーの構成例を図 2 に示す.

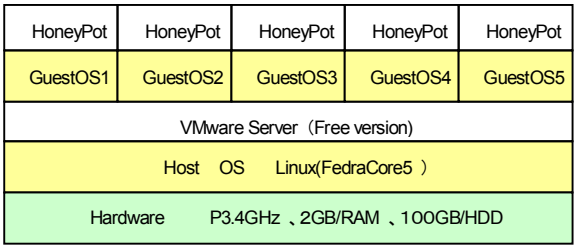


図 2　センサーの構成例

2.2　Malware の実行環境

捕獲した Malware を WindowsXP 上で実行させてネットワークアクセスの状況を調査した．Malware の実行環境としては，ネットワークの模擬と DNS,IRC サーバ等の模擬を行う Truman[5]の一部の機能を利用し，WindowsXP は VMware を利用した仮想 OS 環境上で実行した．実行した際の通信の状況を tcpdump でキャプチャし，あて先 IP アドレス，ポート番号，IRC サーバへのログイン名などを抜き出した．Malware の実行の都度，WindowsXP をクリーンな環境に戻す必要があるが，これには VMware の Snapshot 機能を利用した.

2.3　ログの収集(Collect)

複数の観測サイトから，ログを転送する手段として，UNIX の標準的なツールである RSYNC 及び SSH を利用した．ログの転送は，一定間隔で自動的に実行する必要があり，このためには人手を介さない自動ログインが必要となる．そこで，あらかじめ観測サイト A の公開鍵を観測サイト B,C へ配布し，SSH の公開鍵認証機能を利用して自動ログインを実現した．Snort のログについては，テキスト情報の alert ログを転送した.

2.4　ログ情報の正規化(Normalize)

ログ情報は，出力する OS やアプリケーションによって各種各様であり，生のログ情報を直接処理すると，ログの種類ごとに視覚化などのソフトウェアを開発する必要があり，非効率である．このため，ログ情報の中からキーとなる項目を取捨選択し，必要に応じてデータの様式を変換する必要がある．ここで，選択した項目をログの「要素」と呼ぶ．正規化ログの基本要素として｛タイムスタンプ，発信元 IP，あて先 IP（センサー名），プロトコル，発信元ポート，あて先ポート｝を基本とした．また，これらの要素はハニーポッ

ト，IDS を問わずほぼ共通しているが，対象機器固有の情報の中にも，要素として必要な項目が存在する．例えば，ハニーポットの例では，ファイルをスキャンして判明した Malware の種類や IDS の場合には，Signature 等がこれに該当する．これらの情報は，先に述べた基本要素に任意の個数を付加できるように工夫した．

2.5 ログ情報の集約処理(Aggregate)

集約処理は，2.4 で示した正規化処理に続いて実行するもので，複数サイトから集めた複数種類のログの情報からトップ 10 の算出及び各種の関連付けを行う．例えば IP アドレスと国別コードの関連付け，ダウンロードした Malware の hash 値とウィルススキャンした結果の関連付けなどの処理を行う．本プロトタイプでは UNIX 系 OS で標準的に用意されているツールを利用し，観測サイト A にて一括して処理を行った．

2.5.1. 出現回数のトップ 10 算出

正規化したログ情報の要素の中の注目すべき項目毎に，一定期間内の出現回数の算出を行った．例えば，要素の 1 つである発信元 IP アドレスに注目すると，特定の期間（例えば 7 日間）内にログに記録されたすべてのユニークな IP アドレス毎に，それぞれ何件記録されているかを算出する．次に，算出した結果から，出現回数の多いものから順にトップ N を算出する．N の値は，図表化した際の見やすさの等の観点から，用途に応じて適当な整数（例えば 5, 10 等）を用いるが，表記上は代表してトップ 10 としている．この処理は，必要に応じて正規化ログの各要素に適用し，例えば検出した Malware の種類トップ 10，発信元 IP アドレスの国別トップ 10 等を算出した．算出した結果は，表形式表示や時系列グラフ表示等の基本データとなる．

2.5.2. Malware の検出

ハニーポットで捕獲した Malware をオープンソースのウィルス・スキャン・ソフトウエア ClamAV[6]を利用してスキャンし，ファイル名と Malware 名の対応表を生成した．スキャンの結果，Malware が検出されなかった場合は，新種または亜種である可能性が高いことから，*unknown*と表示した．

2.5.1～2.5.2 の処理は，観測サイト毎にそれぞれ実施するが，これらの結果を総合した観測サイト全体のデータを生成した．

2.6 ログ情報の視覚化(Visualize)

ログ情報の視覚化処理は，閲覧のために専用のソフトウェアの配布が不要な Web ブラウザを利用することを前提として開発した．このため開発言語としては HTML との親和性が高い，PHP を利用し，グラフィックスの描画には，PHP のライブラリである JpGraph[7]を利用した．視覚化処理では，要素毎のトップ 10 の表形式表示，時系列グラフ表示，攻撃元 IP アドレスの分布を示すマトリクス表示[9]及び攻撃元 IP アドレスと観測サイトの関係を示すアニメーション，Malware のアップデート履歴を作成した．

2.6.1. トップ 10 データの表形式表示, 時系列表示

2.5.1 で生成したデータから，表及び時系列のグラフを含む HTML ファイルを生成する．この処理の際に，HTML ファイルの中に必要な識別タグを埋め込んで，各要素のトップ 10 が表示されている状態から，対話形式で検索ができるよう工夫した．

このグラフによって，時系列的な全般的傾向の把握が可能である．捕獲した Malware の種類のトップ 10 の時系列表示例を図 3 に示す．攻撃元 IP アドレス及び Malware の種類トップ 10 の表形式表示例を図 4 示す．

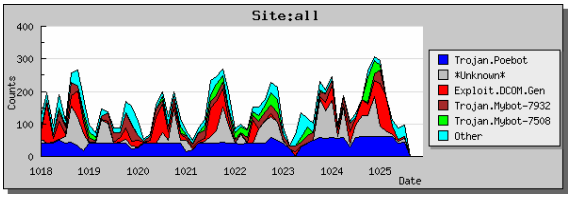


図 3 捕獲した Malware の種類トップ 10 の表示例

Source IP			Malware Name		Source IP Today			Malware Today	
60	60.32.28.xxx	JP	574	Exploit.DCOM.Gen	20	60.32.51.xxx	JP	46	Exploit.DCOM.Gen
41	60.32.24.xxx	JP	72	Trojan.Sdbot-3424	15	60.32.44.xxx	JP	13	Trojan.IRCBot-778
40	60.32.53.xxx	JP	59	Trojan.IRCBot-778	14	221.188.150.xxx	JP	11	Trojan.IRCBot-776
33	60.32.62.xxx	JP	57	Trojan.IRCBot-776	10	60.32.38.xxx	JP	5	Trojan.Sdbot-3424
33	60.32.53.xxx	JP	38	Trojan.Mybot-5073	6	60.32.37.xxx	JP	3	*Unknown*
32	60.32.52.xxx	JP	38	Trojan.Mybot-7508	4	60.32.23.xxx	JP		
27	60.32.25.xxx	JP	35	*Unknown*	3	60.32.56.xxx	JP		
24	60.32.25.xxx	JP	34	Trojan.Mybot-7936	1	60.28.9.xxx	CN		
23	60.32.39.xxx	JP	30	Trojan.Spybot-170	1	60.32.60.xxx	JP		
23	60.32.59.xxx	JP	13	Trojan.Mybot-7899	1	60.50.165.xxx	MY		

Trojan.Sdbot-3424 送信 2006 11 26 4 W

図 4 攻撃元 IP アドレス及び Malware のトップ 10

2.6.2. 攻撃元 IP アドレスのマトリクス表示

観測した攻撃元の IP アドレスが，アドレス空間上のどの位置に分布するかを把握するために，IP アドレスのマトリクス形式による表示を採用した．表示例を図 5 示す．

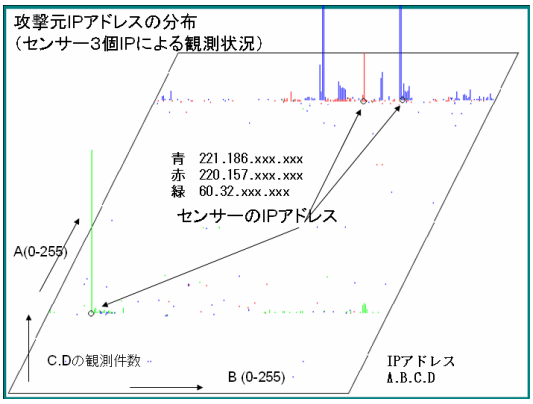


図 5　攻撃元 IP アドレスの分布表示例

IP アドレスは，一般に A.B.C.D のように，8 ビット毎 に区切って表現されるが，発信元 IP アドレスについて，A を斜め上方向に，B を横方向とし，上位 2 オクテットが A.B となるユニークな IP アドレスの個数を上方向としてプロットした．このグラフからセンサーの IP アドレスと上位 2 オクテットが等しい IP アドレスを持つホストからの攻撃が多いことが読み取れる．また，攻撃元 IP アドレスの分布が，横方向に並んでいることから，IP アドレスの 1st オクテットがセンサーと等しいホストからの攻撃が多いことがわかる．

2.6.3. アニメーション表示

2.6.2 の表示を拡張し，攻撃元 IP アドレス，センサーの IP アドレス，捕捉した Malware の種類及び観測日時の関係を，動画形式で表示するソフトウェアを開発した．この表示では，攻撃元 IP アドレスの上位 2 オクテットが，各センサーの IP アドレスと等しいものを表示の対象としている．このソフトウェアによって，複数の攻撃元からの攻撃が，連続した IP アドレスを持つセンサーで，どのような順序で観測されるかについてアニメーション表示することが可能である．また，センサーと攻撃元 IP アドレスの位置関係，Malware の変化の様子，複数のセンサーで観測した順序関係などを直感的に把握できる．表示例を図 6 示す．

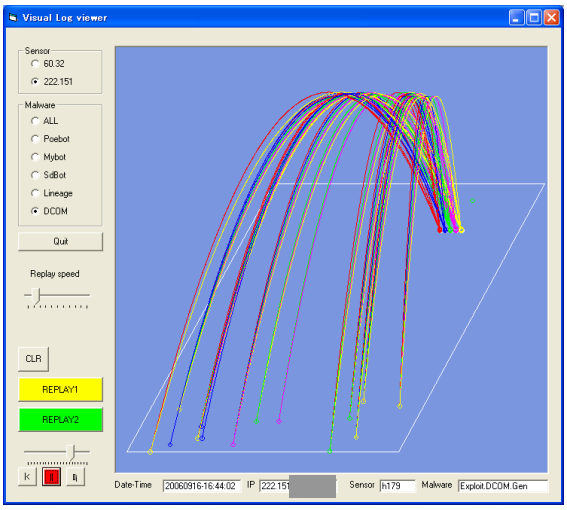


図 6　アニメーション表示例

3　観測結果及び考察

2.2 に示した構成の定点観測システムを使って，2006 年 5 月下旬から 11 月下旬までの間に 1700 種類以上の Malware を収集することができた．観測した攻撃元の IP アドレス数の総数は約 15,000 であり，1 個のセンサーが観測した 1 日毎のユニークな攻撃元 IP アドレス数の平均数は 11.6 で，同様に捕獲した 1 日毎の Malware の種類の平均数は，6.1 個であった．

3.1　捕獲した Malware の種類と感染ポート

収集した Malware を ClamAV でウィルススキャンした結果，表 1 に示すように，ほとんどがボット関連の Malware として分類され，全体の約 20%が未知の Malware と判定された．また，併設した snort で観測した結果，感染を広げるための攻撃に利用されたポートは表 2 のとおり 135, 139 及び 445 の 3 種類で大半を占めた．これらのポートをインターネット上で意図的に利用する可能性は低いと思われることから，ボットネットの感染を拡大させないためには，これらのポートを制御することが有効な対策の一つと考えられる．

表1 Malwareの分類

Name	割合
Trojan.Mybot	28.8%
Unknown	18.8%
Trojan.Poebo	13.1%
Trojan.Sdbot	19.3%
Other	20.0%

表2 snortのAlert

Port	割合
445	37.4%
139	31.1%
135	18.7%
1433	2.3%
Other	10.5%

3.2　攻撃元 IP アドレスの特徴

観測した攻撃元の IP アドレスの総数は約 1 万 5 千を超えたが，日を違えて繰り返し観測される IP アドレスが，どの程度の割合かを調査した結果を図 7 示す．

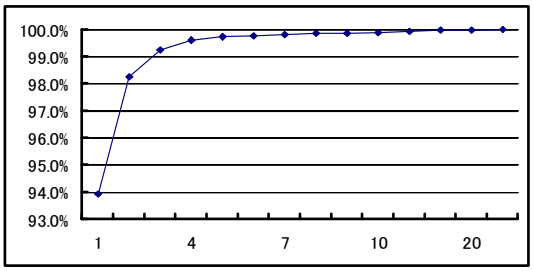


図 7 特定のアドレスを繰り返し観測する累積確率

横軸が延べ観測日数，縦軸が累積確率を示す．この図から，1 日だけ観測される IP アドレスが約 94%であり，約 99%が延べ 3 日間以内の観測となっている．これは，ボットに感染して数日以内に Malware の駆除等の対策を完了しているとも考えられるが，インターネットへ接続する際に，DHCP サーバから動的に IP アドレスが付与される方式の利用者が大半を占めるために，接続の都度 IP アドレスが変化していることを反映しているものと推定される．

センサー毎に，ＩＰアドレスの 1st オクテットが等しい攻撃元の集合を CLASS-A，1st オクテットと 2nd オクテットの両者が等しい攻撃元の集合を CLASS-B とした場合の割合を表 3 に示す．

表3 観測IPアドレスの内訳

Sensor	A	B	C1	C2
Class-A	92.5%	86.8%	86.8%	88.3%
Class-B	0.4%	33.1%	60.1%	50.4%

この表から，どのセンサーについても CLASS-A の範囲に属する IP アドレスからの攻撃が 90%前後を占めていることが分かる．Class-B の範囲の割合はセンサー毎に大きく異なっている．

この状況を 2.6.2 で示したマトリクス表示を利用して視覚化した例を紹介する．センサーの IP アドレスを T，攻撃元 IP アドレスを S と表記する．また，IP アドレスを 8 ビット毎に区切った要素を上位から a, b, c, d の記号を付記して表現する．そこで{Sa=Ta}と {Sa=Ta and Sb=Tb}の場合のマトリクス表示を行うことにより，攻撃元とセンサーの IP アドレスの位置関係を段階的に詳細化して視覚化できる．一例として，{Sa=Ta and Sb=Tb}， Sabcd = 60.32.XXX.XXX の場合について図 8 に示す．図において Y 軸が IP アドレスの 3rd オクテットを，X 軸が 4th オクテットを示し，マトリクス内のドットは，攻撃元の IP アドレスの位置を示したプロットである．図中の○の点はセンサーの IP アドレスを示している．

この図から，攻撃元 IP アドレスの分布に大きな偏りの存在が読み取れる．センサーの周辺からの攻撃は極めて少なく，攻撃元が図の下部周辺に集中している．このような偏りの発生の原因を特定するため，センサー周辺の IP アドレスに関する情報を whois と nslookup コマンドを利用して調査した．その結果，攻撃の少ない IP アドレスの領域は，ほぼ法人向けの割当または利用されていない IP アドレスであり，攻撃の多い領域は個人利用者向けの割当て範囲であることが判明した．このように，定点観測で得られる感染 PC の IP アドレス等の情報を基にして，ISP（Internet Service Provider）を通じて感染者へ連絡し，駆除等を行う対処フローの確立が望まれる．

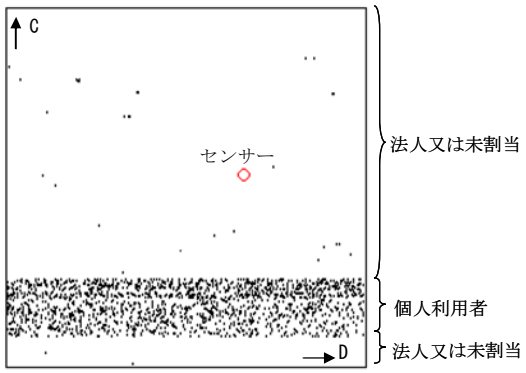
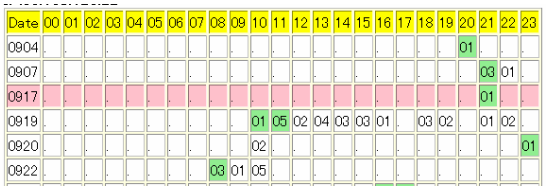


図 8　攻撃元 IP アドレスの分布

3.3　Malware アップデート履歴

ボットネットに使用されるプログラムは頻繁に更新されているとの報告がある[1]．観測したデータの中から，日を違えて繰り返し観測される IP アドレスの数は，3.2 で述べたように多くはないが，繰り返し観測される IP アドレスの記録を自動的に抽出し，日々の 1 時間毎の攻撃回数を表形式で表示した例を図 9 に示す．横方向は 1 時間毎の時間帯を示し，縦方向は月日を示す．表中の数値は，当該時間帯に観測された攻撃回数を示す．数値の背景が緑色の箇所は，ダウンロ

ードした Malware のハッシュ値が変化したことを示している．この例では何回もハッシュ値が変化していることから，ボットのアップデートが頻繁に行われていることがわかる．



Date	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0904																					01			
0907																						03	01	
0917																						01		
0919											01	05	02	04	03	03	01		03	02		01	02	
0920											02													01
0922									03	01	05													

図 9　Malware のアップデート履歴の表示例

3.4　Malware 実行に伴うトラフィックの発生

2.2 で述べた模擬環境で Malware を実行した際に発生したトラフィックを観測し，そのポート番号を調査した結果を表 4 に示す．

表4 Malware実行時の通信ポート番号

Port	65520	80	8585	135	6667	other
割合	23.6%	23.5%	16.6%	15.3%	14.7%	6.3%

ここでポート 65520, 8585, 6667 は irc サーバとの通信であり，80, 135 は Malware の更新や感染拡大のための通信と考えられる．これは文献[1]の調査結果とはやや異なったものとなった．原因としては，捕獲の環境（ハニーポットの種類，センサーの IP アドレス）の違いなどの影響と，特に観測した期間が違うことで，インターネット上で実際に活動しているボットの種類が変化していることを反映しているものと推定される．ボットネットに既に感染した数多くの PC は，これらのポートを利用して harder からの指令を受けることから，65520，8585，6667 などのポートを制御することで，ボットネットの影響を緩和できるものと考えられる．ただし C&C サーバとの通信が IRC だけでなく P2P ネットワークを利用する種類の存在も報告[1]されており，ボットの挙動の変化にも注意が必要である．

4　おわりに

ハニーポットを利用した定点観測によって，ボットネットを構成する Malware を収集し，その挙動の解析を支援するシステムを構築した．正規化したログ情報から，時系列表示，IP マトリクス表示，アニメーション表示及びアップデート履歴等を生成し，捕獲したボットの種類や拡散に利用されるポート番号などの各種統計値，ボットの IP アドレスの分布，ボットのアップデートの状況などのボットネットの特徴的な挙動を把握できることを示した．

また，捕獲した Malware の実行結果のログからボットの感染活動や C&C サーバとの通信に利用するポートを明らかにし，これらのポートを制御することで，ボットネットの影響を緩和できる可能性を示した．ただし，今後ボットネットの対策が促進されると，ボットネットのメカニズムも環境に適応して変化することが予想される．したがって，ボットネットの挙動を把握し，影響を緩和するための対策案を得るには，観測サイトの増強と継続的な観測が必要であり，Malware の解析についてもさらに自動化できるようなシステムの検討と構築が今後の課題である．

参考文献

[1] 高橋　正和, 村上　純一, 須藤　年章, 平原　伸昭, 佐々木良一　「フィールド調査によるボットネットの挙動解析」　情報処理学会論文誌 Vol47.N0.8 Aug,2006

[2]VMware http://www.vmware.com/

[3]Nepenthes http://nepenthes.mwcollect.org/

[4]snort http://www.snort.org/

[5]TJoe Stewar, truman - The Reusable Unknown Malware Analysis Net http://www.lurhq.com/truman/

[6]ClamAV AntiVirus http://www.clamav.net/

[7]JpGraph http://www.aditus.nu/jpgraph

[8]大平　健司,　宋　中錫,　高倉　弘喜,　岡部　寿男　「未知のコードを安全に収集するための定点観測装置の構築手法」　信学技報 IEICE Technical Report IA2006-1(2006-05)

[9]大野一広，小池英樹，小泉　芳　「IP Matrix:広域ネットワーク監視のための視覚化手法」　情報処理学会論文誌 Vol.47 No.4 Apr.2006

[10]伊藤　貴之,　高倉　弘喜,　沢田　篤志,　小山田　耕二　「平安京ビューによる IDS データの視覚化～第２報」

[11]Yinru Chen, Kegang Diao, Istvan Orci, Mats Astrom　“Normalizing Security Audit Data in XML-Format” (2004,Rechnical Report)

[12]Robert Ball, Glenn A. Fink Chris North “Home-Centric Visualization of Network Traffic for Security Administration” (2004 ACM)

[13]江端真行，小池英樹　「不正侵入調査を目的とした複数ログの時系列視覚化システム」　情報処理学会論文誌 Vol.47 No.4 Apr.2006