ネットワーク上の距離に着目したマルウエアの配置に関する一考察

東角　芳樹†　　寺田剛陽†　　鳥居　悟†

†株式会社富士通研究所 ソフトウエア＆ソリューション研究所
211-8588　川崎市中原区上小田中４－１－１

あらまし ボットによる被害を軽減する対策を策定するうえで，攻撃者からの制御を行う制御通信及びマルウエアをダウンロードするサイトへの各ボットからの距離を把握する事は，非常に重要であると考えられる．本論文では，CCC DATAset2010の攻撃通信データの分析において，ネットワーク上での距離に関するパラメータ（TTL値，RTT）に着目し，各マルウエアのボットネットでの役割を考慮したネットワーク上の配置に関する考察を行った．

A study of an arrangement of Malware that consider of a distance on network

Yoshiki Higashikado†　Takeaki Terada†　Satoru Torii†

†Software and Solution Laboratories, Fujitsu Laboratories LTD.

Abstract Measures to reduce damage by Bots are thought that it is very important to understand the distance to site's where the control communication and malware that controls by a high decision by the attacker are downloaded from each Bots. In this thesis, when the attack communication data of CCC DATAset2010 was analyzed, it paid attention to the parameter （TTL, RTT）concerning the distance on the network, and it considered it concerning arrangement on the network where the role on the botnet of each malware had been considered.

1　はじめに

近年，マルウエアの活動目的が金銭的な搾取に特化してきた結果，その活動は非常に目立たないものとなりつつあり，また，手口は非常に巧妙化しつつある．そして，ボットにおいてもこの傾向は顕著になっており，感染させる環境に応じて，さまざまな挙動を示すことから，ボットの検知を非常に困難なものとしている．ボットの基本的な通信は大きく分けると，ボットに感染させるための準備としてのプログラムの脆弱性を狙った感染通信，ボット本体のダウンロード，ボットに指示を与えるための制御通信，DDoS等の攻撃通信，搾取したデータのアップロード通信といった様に，非常に様々な通信が存在する．本論文では，CCC DATAset2010[1]の攻撃通信データの解析による，ネットワーク上の距離から得られた情報から，ボットの通信先までの距離について考察する．本稿では，第２章でネットワーク上での距離の概念について述べ，第３章でボットの挙動と各エージェントの位置関係について述べ，第４章で本研究のア

プローチについて述べる. 第 5 章で通信データの解析結果を示す. 第 6 章で考察, 第 7 章でまとめる.

2　ネットワーク上での距離ついて

ネットワーク上での距離は, ネットワークトポロジの一部の特性として扱われる. 一般的にネットワーク上の距離の代用としてよく使われる指標としては, 通信負荷等の変動の影響を受けるが, 通信速度が用いられることが多い. しかし, TCP/IP 通信においては, ネットワーク上の距離として用いることができるものとして IP ヘッダの TTL 値, および, TCP 通信の RTT を用いることが出来るものと考えられる. IP ヘッダの TTL(Time To Live)値は, その初期値として, 各ノードの OS に依存した一定の値をとる. そして, エンドノードから1つのルータを IP データグラムが通過する場合に, TTL 値を1つ減じ, その値が 0 となった場合当該パケットをルータが破棄するというプロトコルとなっている. このため, IP ヘッダの TTL 値は, パケットの通過してきた通信経路上のルータや L3 スイッチの数を表すこととなり, 距離的な指標として用いることができるものと考えられる. 次に, RTT であるが, 図1に示すように, TCP の通信パケットが往復する時間を示しており, TCP プロトコル制御(パケットの送出間隔の調整等)でこの RTT を用いている. パケットが通過してくる経路の帯域幅や負荷の状況に依存するが, RTT もネットワークの距離的な指標として用いることが出来るものと考えられる.

本稿では, これらの値(IP-TTL, TCP－RTT)を用いることで, ネットワーク上でのボットの各通信先までの距離を計測するものとする.

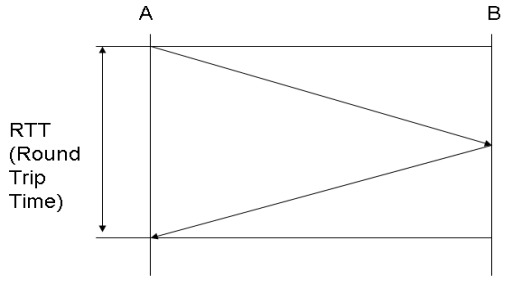

図1. RTT の概念図

2.1　関連研究

ネットワーク特性の測定として距離を測定するものとして, Internet Distance Map Service(IDMaps)[2]がある. IDMaps は, RTT や帯域幅といった距離情報を収集・提供する基盤環境である. Tracer と呼ばれる特別な測定エージェントを置き, 他の Tracer との間で相互に距離情報の測定を行うものである. また, このほかにネットワーク上の距離をあらわすために, 座標系を用いたネットワーク情報の推定も行われており Global Network Positioning(GNP)[3]ではユーグリッド空間において RTT をネットワーク上の距離に対応させる研究が行われている.

本研究では, この RTT と TTL 値を用いることにより, ボットとその通信先である, 感染元システム, ダウンロードサーバ, 制御通信サーバまでの距離について評価する.

3　ボットの挙動

ここでは, ボットの挙動を整理すると共に, 各フェーズで発生する通信に対して想定される特徴を整理する.

3.1　ボットの挙動

一般に, インターネット上のサーバからプログラム等をダウンロードする「ダウンローダ」を介して, ボットに感染するといわれている. 感染後, ハーダからの制御通信に基づき様々な挙動を示す. この動作を図2に示す.

感染に至る流れは以下の4つの手順に分類できる.

① 攻撃/侵入:主にポート番号135 のRPC(Remote Procedure Call)を使ってダウンローダが送り込まれる.

②不正プログラムのダウンロード:TFTP やHTTP を使って不正プログラムをダウンロードする.

③ 攻撃者の制御通信:IRC メッセージのやり取りが行なわれ, 動作を制御される.

使用されるポート番号はIANAのIRC の割り当て番号とは異なる場合がある.

④ 踏み台の拡大/攻撃:SPAM 送信やDoS 攻撃, 他PC の探索などが行なわれる.

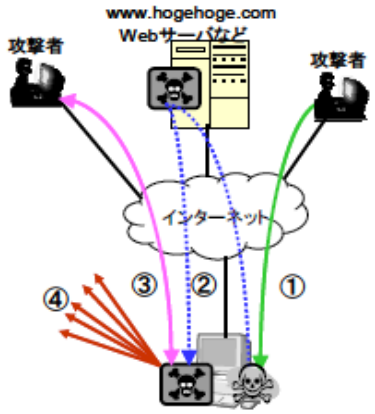

図2. ボットの動作

本論文では，①～③の通信について研究の対象とする.

4 アプローチ

4.1 着眼点

ボットの各通信相手先（感染通信元，制御通信サーバ，ダウンロードサーバ）とのネットワーク上の距離の特徴を特定するにあたり，ボットのそれぞれの通信フェーズにおける TCP セッションの IP ヘッダの TTL 値と TCP のセッションの RTT に着目する.

攻撃者の制御通信はその性格上，指示が確実に伝わるTCP プロトコルが用いられるケースが多いと考えられる. また，攻撃通信DATAの解析結果から，攻撃元との通信，制御通信サーバとの通信，および，データのダウンロードにおいて使用されているプロトコルは，その大部分がTCP プロトコルによるものであった.

これらのことから，ネットワーク上での距離を推定するに当たり，TCP セッションに着目することは妥当であると考える.

4.2 セッションの再構築および RTT の計測

攻撃通信データは，pcap形式のデータであり，パケットキャプチャしたデータである. ここから，IP アドレスやセッション情報などIP ヘッダの情報を基に，同一セッションを割出しグループ化する.

この際，RTTの測定にパケットの消失や再送などの考慮を必要とするために，セッションごとに分離したpcapデータをtcptrace[4]にかけて解析を行った.

5 通信データの解析

5.1 ハニーポットの送受信データ

ハニーポットの送受信データを感染通信，ダウンロード，制御通信に分類する事は、昨年，TCP セッションの特徴に基づくボット制御通信の検知方式の検討[5]で説明した，3つのフェーズに分けて，TTL，RTTを解析した. なお，RFC791でデフォルトのTTL値は64とされているが，実際には，各システムのOSによるIPデータグラム送出時のデフォルトTTLの値は，以下の通りとなっている. なお，ボット等のアプリケーションにより変更されていないものとする.

Windows[1]: 128
Linux[2]: 64
BSDなどのUNIX[3]: 255

5.2 感染通信の解析

ボットの感染通信は，今回の感染システムがWindowsであることから，主に135（DCE-RPC）のポートに対して感染元サーバからの感染通信を受信しているセッションについて，そのパケットのTTLを横軸に，縦軸に同じTTL値の個数を集計したものを図3に示す.

---

1 Windows is a registered trademark of Microsoft Corporation in the United States and other countries.
2 Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries.
3 UNIX is a registered trademark of The Open Group

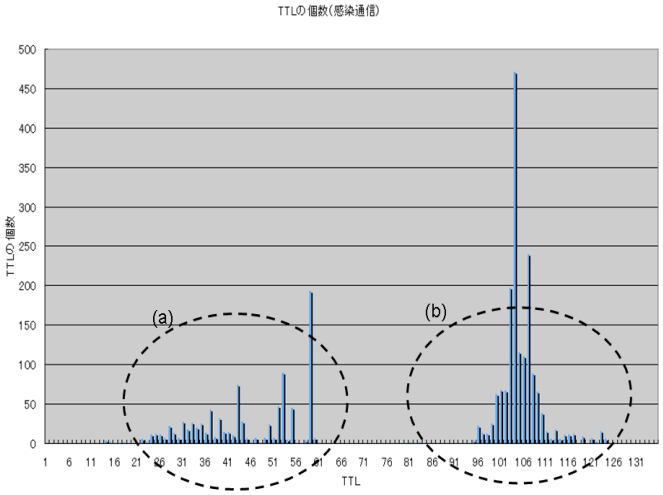

図3.感染通信のTTLの集計

図3では，TTL値が40付近を中心とするグループ(a)と，100付近を中心とするグループ(b)の大きく2つのグループに分類できることがわかる．

これは，TTLの値の差は，感染源のサーバからの距離（ホップ数）が大きく異なる，あるいは，パケット送出時のTTL初期値が異なるかのいずれかの原因が考えられるが，後者である可能性が高いものと思われる（詳細は考察で検討する）．パケット送出時のTTL初期値が原因であるとすると，感染源からのホップ数は24程度と考えられグループ(a)は，Linux系のシステムが感染元のサーバ，グループ(b)は，Windows系のシステムが感染元のサーバと考えられる．また，感染通信の送信元システムの数は，Linuxに比較して圧倒的に，Windowsが多いことが図3から見て取れる．

5.3　ダウンロード通信の解析

ボットのダウンロード通信は，主にHTTPあるいは，FTP相当のプロトコルを用いて行われることがわかっている．そこで，これらのダウンロード通信を行うパケットのTTL値を横軸に，縦軸に同じTTL値の個数を集計したものを図4に示す．

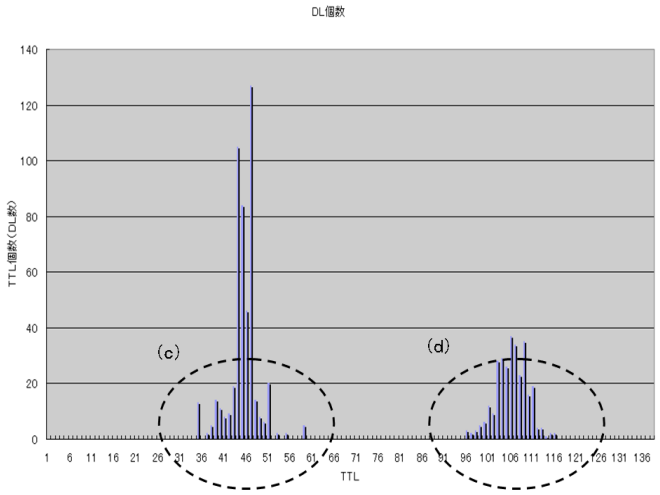

図4．ダウンロード通信のTTLの集計

図4では，TTL値が40付近を中心とするグループ(c)と，100付近を中心とするグループ(d)の大きく2つのグループに分類できることがわかる．この2つのグループに分類できる原因は5.2の感染通信の場合と同じと考えられる．そこで,ダウンロードサイトからのホップ数は22程度と考えられ,ここでもグループ(c)は，Linux系のシステムがダウンロードサイトで用いられ，グループ(d)は，Windows系のシステムがダウンロードサイトで用いられているシステムであることが考えられる．また，ダウンロード通信では，送信元システムは，Linuxの方がWindowsより多い．これは，FTPやHTTPサーバとして、Linuxの方が用いやすい事が関連しているものと推測できる．

5.4　制御通信の解析

ボットの制御通信は,テキストによるIRCをベースとした制御コマンド体系が用いられ,使用されるポートは,IRCのポート6667以外にもさまざまなポート番号が使用されていることが多い．そこで，これらの制御通信を行うパケットのTTL値を横軸に，縦軸に同じTTL値の個数を集計したものを図5に示す．

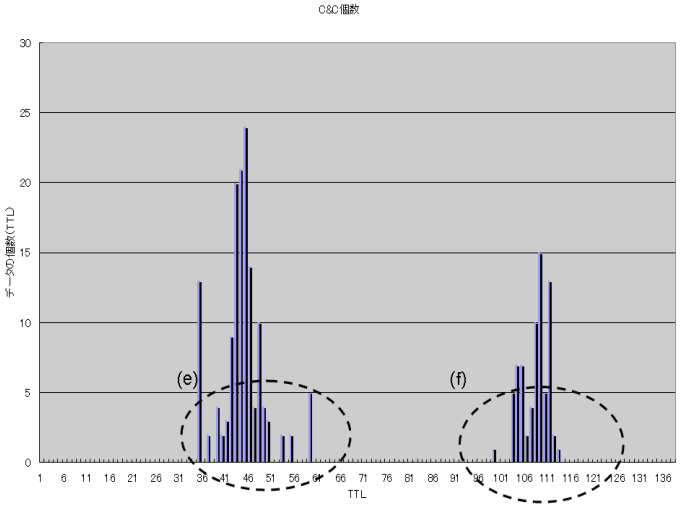

図5. 制御通信のTTLの集計

図5では，TTL値が40付近を中心とするグループ(e)と，100付近を中心とするグループ(f)の大きく2つのグループに分類できることがわかる．このグループに分類できる原因は，5.2の感染通信の場合と同じと考えられるので,制御通信サイトからのホップ数は20程度と考えられる．このため,グループ(e)は，Linux系のシステムが制御通信サーバ，グループ(f)は，Windows系のシステムが制御通信サーバとして用いられているシステムであることが考えられる．また，制御通信サーバで使用されているシステムがLinuxが多い理由は，制御通信で用いられているプロトコルが，主にIRCであることから，IRCサーバを動作させやすい，LinuxやUNIXサーバが主に用いられているものと推測する事が出来る．

6　考察

6.1　各通信における2グループ化の理由

感染通信についてTTLの値が40付近を中心とするグループと，100付近を中心とするグループの2つに大きく分かれている．この理由は，実際のホップ数が異なるということで考えると60以上のホップ数の差がある送信元から感染通信を受けている事になる．しかし，RFCでのIPのTTL値の推奨値が64であることや，実際Linux等のIPヘッダのTTL初期値が64であることから考えると，実際のネットワーク上で60以上のホップ数の違いがある場合，様々なケースでIPデータグラムがTTL値が0になることにより破棄されるということになる．しかし，実際は，ネットワーク上でのループなどの特殊な場合を除けば，データグラムが破棄されたといった話を聞かないことから，このTTLの値の差は，ホップ数の違いからくるものではないものと推測できる．そこで，この値の差が送信元のシステムのOSのTTLのデフォルト値からくるものと考えた場合，ボットに感染しているシステムにWindowsが多い実情から考えると，Windowsが感染通信を行っているとの推測は，整合性がとれているものと思われる．これらのことから，感染通信，ダウンロード通信および，制御通信のTTLの値が2つのグループに分類される理由は，送信元のシステムOSのTTLの初期値に違いがあると考えることが妥当であると思われる．

6.2　感染通信の送信システム

一般的にWindowsに感染するマルウエアは，Windowsから感染するものといわれている．しかし，6.1.の考察結果から考えると，必ずしもWindowsシステムからの感染だけではなく，Linuxのシステムから，マルウエアに感染することがあることを示している．

6.3　各サーバとの TTL 値と RTT 値について

ダウンロード通信の送信元サーバからのTTL値とSYN-SYN/ACK時のRTTをプロットすると図6のようになる．

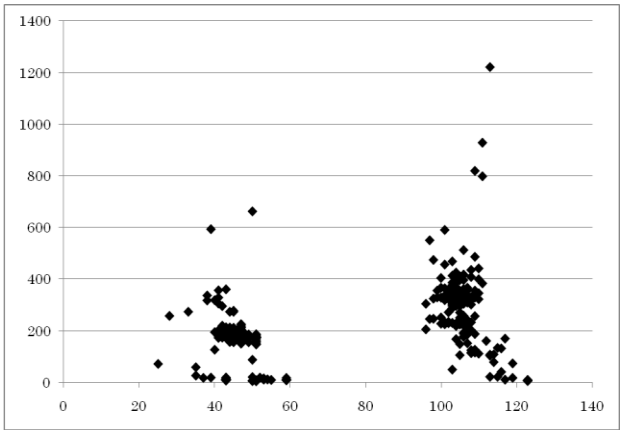

図6. RTTとTTLの関係（ダウンロード通信）

次に、制御通信の送信元サーバからのTTL値とSYN-SYN/ACK時のRTTをプロットすると図7のようになる。

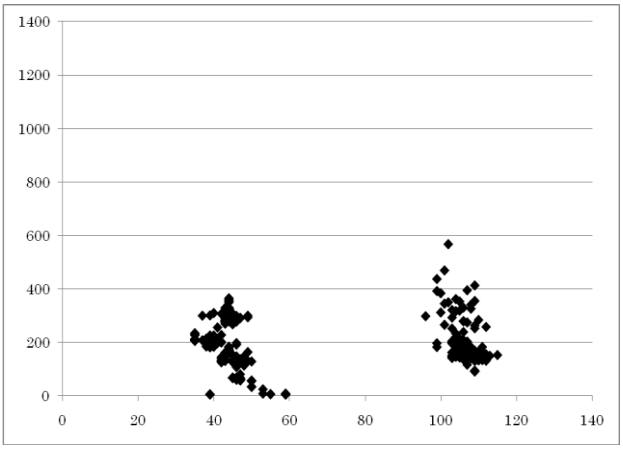

図7. RTTとTTLの関係（制御通信）

図6,図7より,サーバからのホップ数は，ダウンロードサーバ，制御通信サーバ共に，TTL値のみをプロットし場合と同じくTTL値が40付近のグループと100付近の2つのグループに分類されるが，この理由は6.1で述べたとおりである．また，ダウンロード元のサーバとボット感染ホストとのネットワーク距離と制御サーバとボット感染ホストとのネットワーク距離に関しては若干ダウンロード通信の送信元の方が，RTTにばらつきが大きい点を除けば，さほど大きな差は認められなかった．このため現状では，攻撃者はまだ，それほどネットワーク的な距離を考慮してボット感染ホストを配置していないものと思われる．なお，感染元サーバとボットとのネットワーク距離に関しては，SYN-SYN/ACKの3WAYハンドシェイクの方向が逆向きであるため，6.3の方法では計測できないため，まだ評価していない．

7　まとめ

(1)ボットの感染通信の送信元として従来考えられてきたWindowsシステムのみでなく，Linux等のUNIX系システムが存在する事を示した．
(2)ボットのダウンロードサイトおよび制御通信サイトには，Linux等のUNIX系ホストが比較的多く使われていることを示した．
(3)ボットのネットワーク上の位置関係に着目し，ボットとダウンロードサーバ，および，制御サーバとのネットワーク距離については，大きな差は無く，攻撃者は，まだネットワーク的な距離を考慮した上でボット感染ホストを配置していないものと思われる．

8　今後の方向性

・感染元ホストとボット感染ホストとの間では，SYN-SYN/ACKの方向が異なるため6.3の方法で評価が行えない．このため，通信セッション全体でのRTTの平均値を求める方法を用いて，RTTとTTL値の評価を行いう予定である．そこで，攻撃者の何らかの意図が働いているか否かについて考察する予定である．

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 ～MWS 2010 Datasets～，MWS2010（2010 年 10 月）
[2]P.Francis et.al, IDMaps: A Global Internet Host Distance Estimation Service., IEEE/ACM Transaction on Networking, Oct, 2001.
[3]T.S.Eugent Na and Hui Zhang., Prediction Internet Network Distance with Coordinates-Based Approaches., IEEE INFOCOM, pp.170-179, 2002
[4]O.Shawn, The tcptrace project, http://www.tcptrace.org
[5]東角芳樹，他：TCP セッションの特徴に基づくボット制御通信の検知方式の検討，MWS2009(2009年10月)