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

東角　芳樹、鳥居　悟
株式会社富士通研究所　ソフトウェア＆ソリューション研究所

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，攻撃者からの指示を与える制御通信を見つけることは，非常に重要であると考えられる．本論文では，我々が CCC DATAset 2009 の攻撃通信データを解析して得られた，制御通信の検出手法について報告する．本手法を評価したところ，様々なボットに適用可能であり，ほぼ 100%という高い確率で制御通信を特定できるという結果が得られた．
キーワード：ボット，ハニーポット，ログ解析，制御通信検知

Study for Detecting C&C Communication Based on Behavior of TCP/IP Session

Yoshiki Higashikado, Satoru Torii
Software and Solution Laboratories, Fujitsu Laboratories LTD.

Abstract : The command and control communication(C&C communication) is used for attacker to manage the behavior of their bots. To identify these communications is very important for examining the method of decreasing the harmful bots damage. In this paper, the detection method of the C&C communication is described. The method was obtained by analyzing the attack communication data of CCC DATAset 2009. It was possible to apply to various bots, and it was able to specify the C&C communication by short odds of almost 100%.
Keyword : Bot, honeypot, log analysis, C&C communication detection

1. はじめに

近年のマルウェア特にボットは，環境に応じて動作を変更するなど巧妙になってきている．特に，攻撃者から随時指示を受けて自分自身の挙動を変更するといわれており，検知に必要な挙動を規定することが困難となっている．このような指示を与える制御通信を見つけることは，感染検知のみならず攻撃者の特定など，ボットによる被害を軽減する対策を策定するうえで，非常に重要であると考えられる．

本論文では，我々が CCC DATAset 2009[1]の攻撃通信データを解析して得られた，制御通信の検出手法について報告する．

本手法を導くにあたり，我々は，攻撃通信データから TCP セッションを再構築し，全てのセッションの特徴を分類/整理した．その結果，幾つかのグループに分類できることが明らかになった．そこで，得られた特徴の違いを元に不正プログラムのダウンロード通信と攻撃者からの制御通信に切り分ける手法を検討した．

この手法を CCC DATAset 2009 をもちいて評価したところ，様々なボットに適用可能で

あり，ほぼ 100%という高い確率で制御通信を特定できた．

以降では，攻撃者の制御通信を特定する研究動向を紹介し，ボットの挙動とその関連通信の特徴を整理する．本研究における手法のアプローチについて述べ，攻撃通信データを用いた評価実験とその結果について述べる．最後に，実験結果に関する考察を行う．

2. 関連研究

ボットに関する研究は，挙動分析，感染検出，攻撃手法，感染規模の把握など，様々な観点で行われている[2][3][4]．

特に，攻撃者の制御通信を特定する研究に関しては，主に IRC プロトコルが使用されていることか利用しそこで使用されるキーワードを用いた検出[5]や，ボットネットの挙動モデルに基づく検出[6]など，各通信パケットやセッションに着目した手法の研究が行われている．

一方，本研究では，制御通信全般に見られる特徴の有無に着目している．すなわち，ある期間の通信ログを集計し，それら全体の傾向を分析することで，検出に利用できる特徴を導くことを目的としている．

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，攻撃者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 1 に，その時の通信状態を図 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 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．

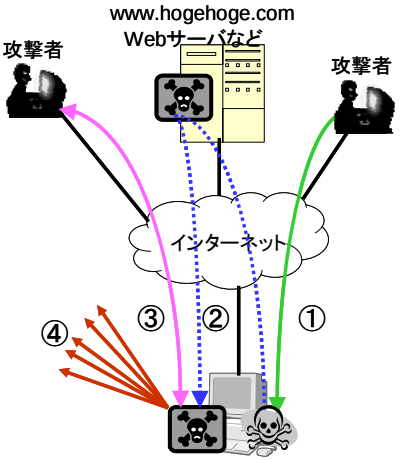

図 1:ボットの動作

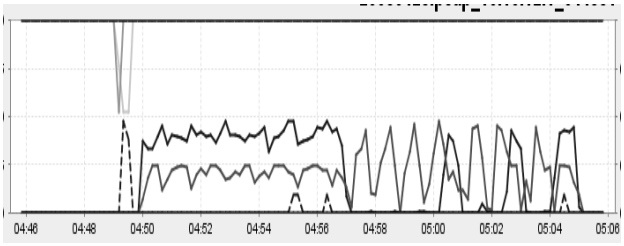

図 2:特徴的な挙動を示すセッション例

3.2. ボット関連通信の特徴

今回の攻撃通信データでは，各通信はおおむね以下のような特徴を備えるものと考えられる．特徴を一覧表として表 1 に示す．

表 1:ボット関連通信の特徴

通信種別	コネクション	データ量
攻撃侵入	外部から張る	inbound ＞ outbound
不正プログラムのダウンロード	内部から張る	inbound ≫ outbound
攻撃者の制御通信	内部から張る	inbound ≒ outbound
踏み台の拡大/攻撃	n/a	n/a

(1)攻撃侵入
通信の契機は，外部から行われる．データ量は，流入量のほうが多いと考えられる．
(2)不正プログラムのダウンロード
通信の契機は，主にハニーポットから行われると考えられる．データ量は，流入量のほうがかなり多いと考えられる．
(3)攻撃者の制御通信
通信の契機は，主にハニーポットから行われると考えられる．データ量は，流入量と流出量との差はほとんどないと考えられる．
(4)踏み台の拡大/攻撃
今回の攻撃通信データでは，外部への攻撃は成功していない．すなわち，ハニーポットの仕組みにより，このようなセッション確立やデータ送信がブロックされているものと考えられる．

4. アプローチ

4.1. 着眼点

攻撃者の制御通信を特定するにあたり，攻撃通信データにおける TCP セッションに着目する．

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，攻撃通信データに該当する攻撃元データにおいては，UDP プロトコルによる検体取得はほとんどみられず，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，制御通信を特定するにあたり，TCP セッションに着目することは妥当であると考える．

すなわち，攻撃通信データから TCP セッションを再構築し，再構築したセッションに対して適用することで制御通信の特定が可能であると考える．

4.2. セッションの再構築

攻撃通信データはパケットキャプチャしたデータであり，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を行う．パケットの消失や再送などの考慮はしていない．

なお，セッションとして割出すにあたっては，何らかのデータ(1byte 以上)が流れたものを対象とする．すなわち，3-way ハンドシェイクが成功せずセッションが確立できなかったものなどはセッションとしてみなさない．

4.3. ハニーポットの送受信データ

先に示したボット関連通信の特徴が妥当なものであるか，攻撃通信データの一部を用いて得られた TCP セッションに対して評価を行った．

図 3 は，ハニーポットから接続する通信の送受信データ量の分布を示したものである．ここで，ハニーポットは WindowsXP[i]であり，各 TCP セッションを対象としている．横軸に送信データ量を，縦軸を受信データ量を，それぞれ対数目盛りで示している．

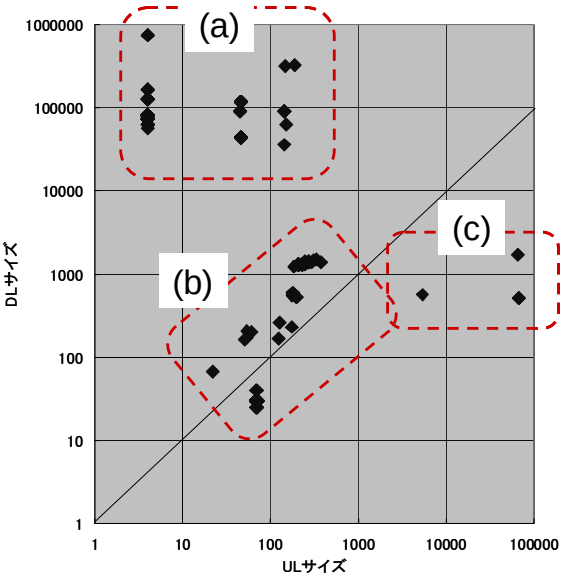

図 3:送受信データ量の分布

ここから，各セッションは，いくつかのグループに分類できることがわかる．すなわち，図中(a)は不正プログラムのダウンロードの通信，図中(b)は攻撃者の制御通信ではないかと類推できる．先に示したボット関連通信の特徴が妥当なものであるとの感触がこの分布結果から得られた．

図中(c)は何かをアップロードする通信として，これまでには見られない新たな挙動ではないかと類推される．当該通信セッショ

ンは，以降の集計対象から外している．当該通信については『6.4 何かをアップロードする通信』にて詳述する．

4.4. 検知手法の確立に向けて

攻撃者の制御通信を特定するにあたり，図 3 における(a)と(b)を区別するための何らかの指標が必要である．そこで，それぞれが区別可能となるような特徴がないか調査・検討した．

図 4 は，ハニーポットが WindowsXP の場合における，各 TCP セッションのセッション継続時間と平均パケット長による分布を示したものである．ここで，横軸はセッション継続時間を，縦軸は平均パケット長を示している．

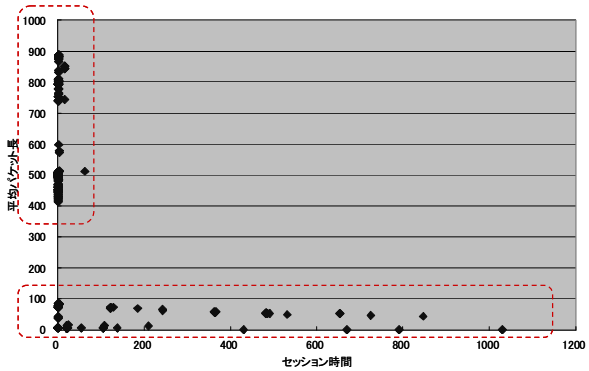

図 4:セッション時間とパケット長の分布

この図からわかるように，平均パケット長が 0byte から 100byte のグループと，それ以上のパケット長のグループとに分類が可能であることが見とれる．

この平均パケット長が，攻撃者の制御通信を特定するための指標として有効であると考えられる．

5. 評価実験

5.1. 前提

ハニーポットから接続を行う TCP セッションの特徴に着目し，攻撃者の制御通信が特定可能であるか評価する．特定するための指標として，セッションの平均パケット長を使用する．

評価にあたって，対象は CCC DATAset2009 の攻撃通信データにある全セッションとする．制御通信であるかどうかは，攻撃元データと照合されない通信かどうかで判断する．すなわち，照合されたセッションは不正プログラムのダウンロード通信であり，照合されなかったセッションは攻撃者の制御通信とみなす．

5.2. 攻撃元データとの照合手順

攻撃元データのログとの照合は，再構築したセッション単位に行った．

具体的には，検体取得日時刻，IP アドレス(pull は DstIP，push は SrcIP)，SrcPort/DstPort，TCP/UDP，通信方向（pull/push）が一致する通信セッションを抽出した．ここで，攻撃元データの時刻が秒単位の時刻であり，攻撃通信データの単位と異なるので，照合には前後 1 秒のマージンをとった．照合結果を表 2 に示す．

表 2:照合結果

		WindowsXP	Windows 2000
TCP/pull			
	セッション照合数	190	212
	攻撃元データ数	193	213
TCP/push			
	セッション照合数	1	1
	攻撃元データ数	1	1
UDP/pull			
	セッション照合数	6	6
	攻撃元データ数	6	6

照合の結果，一致しなかった攻撃元データが 4 個あった．これらの検体取得日時刻は，攻撃通信データの境目，すなわち，異なる日付ファイルにまたがる場合であった．そこで評価にあたっては，照合できたセッション(WindowsXP:190, Windows2000:212)を対象とすることとした．

なお，攻撃通信データに格納されているダウンロード通信は，その大部分がハニーポットから通信リクエストを発行する TCP/pull 型であったことがわかる．

5.3. 検出結果

特定するための指標である平均パケット長の値を 100byte/パケットとした．便宜上，

この値より大きいセッションを「長パケットセッション」，小さいセッションを「短パケットセッション」と呼ぶ．すなわち，長パケットセッションが不正プログラムのダウンロード通信であり，短パケットセッションが攻撃者の制御通信とみなせる．

表 3 に，再構築したセッション数，長短パケットセッションの数，攻撃元データとの照合結果を示す．

表 3:検出結果

			セッション数	攻撃元データと照合	
CCC Dataset 2009					
	Windows XP				
		長パケット	213	190	89.2%
		短パケット	117	0	100.0%
	Windows 2000				
		長パケット	224	212	94.6%
		短パケット	160	0	100.0%

平均パケット長(100byte/パケット)を指標とした場合，単純な手法であるものの，その検出性能は高いといえる．また，制御通信と検出してはいけないものを抽出した割合(false positive)は 0%であり，検出すべきものを見逃した割合(false negative)は 5.4%から 10.8%であった．

6. 考察

6.1. 検出を見逃した通信

攻撃元データと照合されていない通信が攻撃者の制御通信として検出されていないケースが 35 個ある．これらの該当するセッションに対して，通信の詳細を分析してその原因を調査した．

まず，再構築して得られた通信セッションの中から類似セッションがあるかどうかを調査した．ここで類似セッションとは，当該セッションと同じ内容のセッション(DST アドレス，DST ポート，DL 量，UL 量)のものをいう．

その結果，類似セッションを持つものが 30 個あり，それらの類似セッションは攻撃元データと照合されていた．このうちのひとつは，途中でセッションが切れていたようである．おそらく定期的なリブートによるものと考えられる．残りの 29 個は，入手するファイル名が異なっているなど，明確な相違点が見受けられず，マルウェアがダウンロードされていたのではないかと考えられる．

残りの 5 個に関しては，類似セッションが見つからず，実行ファイルのようなものがダウンロードされているが，詳細は不明である．

これらのことから，攻撃元データと照合されていない通信のほとんどは，ダウンロード通信であり，本手法による検出は正しく行われたとみなすことができる．

6.2. 特定した各通信セッションの特徴

CCC DATAset 2009 攻撃通信データを対象に，特定した各通信セッションの特徴を整理する．

(1) 不正プログラムのダウンロード

長パケットセッションとして特定されたものである．本通信セッションでは，その全てにおいて，実行ファイルのようなものがハニーポット宛に転送されていた．

平均パケット長が長く，おおむね 500Byte と 800Byte のところに集中していた．セッション時間は短く，4 秒未満か 16 秒前後であった．宛先ポート番号は，80 番もしくは 1 万番台（5 桁）に集中しており，かなり大きな数値であった．

(2) 攻撃者の制御通信

短パケットセッションとして特定されたものである．本通信セッションは，実行ファイルのようなものが転送された形跡がまったくなかった．

平均パケット長が短く，100byte 以下であった．セッション時間は長く，最長で 1,000 秒間継続しているものもあった．宛先ポート番号は，80 番もしくは 1000 番台（4 桁）に集中していた．

6.3. CCC DATAset 2008 での評価

同様に，CCC DATAset 2008 攻撃通信データを対象とした評価実験を行った．

その結果，Windows2000 をハニーポットとするデータに対しては、特定するための指標である平均パケット長の値が 200byte/パ

ケットと値が異なるものの，おおむね同様の特定が可能であった．

一方，WindowsXP をハニーポットとするデータに対しては，特定するための指標を求めることができなかった．但し，特定のあて先ポート番号(80，443)を除く通信セッションに対しては特定が可能であった．

6.4. 何かをアップロードする通信

図 3(c)に相当する通信では，どのようなやり取りされているか調査した．該当セッションは 8 個あり，その全てが基本的には同じ処理手順であった．

ハニーポットから大量(64Kbyte)のデータを POST メソッドの引数として送信している．成功(200)の場合は，サーバから実行ファイルのようなものがダウンロードされる．失敗(403)の場合は，サーバからのダウンロードは行われていない．実行ファイルのようなものがダウンロードされた場合であっても，攻撃元データには該当するメッセージはなかった．

さらに，当該セッションが発信される状況を元に，ウィルス関連情報を調査したところ，『ダウンロードされるファイルは「nonmalicious」』とのことである．そのため攻撃元データには検出したとのメッセージがないものと考えられる．

データ転送量から判断すると，本通信では何かをアップロードしているように見えるが，結果的には無害のファイルをダウンロードする通信であると考えられる．

7. まとめ

本論文では，ボットに対する攻撃者の制御通信の検出手法について述べた．本手法は通信セッションの平均パケット長を指標とするものであり，非常に単純な仕組みで検出することが可能である．

CCC DATAset 2009 の攻撃通信データを元に評価した結果，様々なボットに適用可能であり，ほぼ 100%という高い確率で制御通信を特定できるという結果が得られた．すなわち，本手法により，ボットの挙動における制御通信，少なくとも実行ファイルをダウンロードするものではないものを特定することができると考える．

今後は，一般的な環境で本手法がどの程度有効であるか，すなわち，攻撃者の制御通信と類似する特徴を持つ通信の発生頻度，区別するための適切な指標の値などに関しての精査を行っていく．

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] Z. Zhu, G. Lu, Y. Chen, Z. J. Fu, P. Roberts, and K. Han. *Botnet research survey.* In 2008 32nd Annual IEEE International Computer Software and Applications Conference (COMPSAC'08), Turku, Finland, July 2008.

[3] Michael D. Bailey, Evan Cooke, Farnam Jahanian, Yunjing Xu, and Manish Karir. *A Survey of Botnet Technology and Defenses.* In Proceedings of the Cybersecurity Applications & Technology Conference For Homeland Security (CATCH '09), Washington, District of Columbia, USA, March 2009.

[4] 東角芳樹，他：*DNS 通信の挙動からみたボット感染検知方式の検討,* MWS2008（2008 年 10 月）

[5] J. Goebel and T. Holz. Rishi: *Identify bot contaminated hosts by irc nickname evaluation.* In First Workshop on Hot Topics in Understanding Botnets (HotBots'07), Cambridge,MA, April 2007.

[6] G. Gu, R. Perdisci, junjie Zhang, and W. Lee. *BotMiner: Clustering analysis of network traffic for protocol- and structure-independent botnet detection.* In Proceedings of the 17th USENIX Security Symposium (Security'08), San Jose, CA, July 2008.

---

[i] Windows は米国 Microsoft Corporation の米国およびその他の国における登録商標です。