Snort ルールの組合せによるボット通信検知方式の確立と 改ざんサイト自動検知システム DICE の機能拡張

田中 達哉† 佐々木 良一†

†東京電機大学
〒101-8457 東京都千代田区神田錦町 2-2
tanaka@isl.im.dendai.ac.jp

あらまし ボットネットは PC にマルウェアを感染させることで範囲を広げ，被害を増加させている．マルウェアを感染させる手段として，WEB ページからの感染と，不正侵入からの感染がある．不正侵入を検知する手段として，IDS が存在するが IDS のルールは膨大であり，設置環境に合わせたルールのチューニングが必要となる．そこで，本論文では CCC DATAset 2011 を IDS の一種である Snort に適応することで，検知に最適なルールを抽出し，このルールの中から最適なルールの組み合わせを検証する．そして，この組み合わせを先に開発した改ざんサイト自動検知システム DICE の機能として追加し，WEB ページ，不正侵入，双方からの感染を防止する手法を提案する．

Development of Bot Communication Detection Method Using Snort Rules and Its Use to Extend DICE (Defaced Sites Automatic Detection System)

Tatsuya TANAKA† Ryoichi SASAKI†

†Tokyo Denki University
2-2, Kanda-Nishiki-cho, Chiyoda-ku, Tokyo, 101-847 JAPAN
tanaka@isl.im.dendai.ac.jp

Abstract Botnets are expanding the scale by infecting the malware on PC and spreading damage. As a way to infect the malware, there exist attacks from Web page and unauthorized intrusion. In order to use IDS for detecting unauthorized intrusion, big amounts of Rule set of IDS must be tuned to be suitable to the required environment. In this paper, we analyze CCC DATAset 2011 using Snort and extract best rules of Snort and validate the best rules among them. In addition, we proposed a method to prevent infection caused by not only web pages attack but unauthorized intrusion by using the rules to expand the function of DICE (Defaced Sites Automatic Detection System).

1. はじめに

近年，ボットネットを利用した様々なサイバー犯罪が横行している．このボットネットは，ハーダーと呼ばれる攻撃者が C&C サーバと呼ばれる制御サーバへ命令を送り，ボットに感染したユーザ PC に不正な動作をさせるネットワークである．命令には，指定されたサーバへの DDoS 攻撃，スパムメールの送信，フィッシングサイトの構築，マルウェアの拡散活動など多数存在し，命令の数は 100 種類以上存在すると言われている.[1]そのため，ネットワーク管理者は自身のネットワーク内でこれらの命令通信を発見するために，侵入検知システム(IDS)を用いて通信内容を監視し，命令通信を発見している．この IDS は通信内容から固有の通信パターンを定義したシグネチャ(ルール)を用いて命令通信を判別している．

しかし，先に述べたように命令の数は 100 種類以上存在し，かつボット等に代表されるマルウェアは短期間で亜種が発生するため，ルールが更新されない限り，新たなボットに対応できない問題がある．また，ルールは数が膨大であることから，検知した危険な通信をログとして残す際，ログが冗長になる傾向があることに加え，IDS を設置する環境に合わせたルールのチューニングが必要になる．

そこで本論文では，CCC DATAset 2011[2]の攻撃通信データを用いて BOT の攻撃からマルウェアに感染するまでの通信を分析する．そして，分析結果より攻撃からマルウェア感染までの IDS のルールを抽出する. そして，抽出した IDS のルールを組合せる事で BOT 通信の検知に最適なモデルを提案し，マルウェアの感染を防ぐ．また，我々が先に提案した改ざんサイト自動検知システム DICE[6]の新たな機能として提案した検知モデルを加えることで，改ざんされた WEB ページからのマルウェア感染と BOT からの感染に対応した新たな検知システムを提案する．

2. CCC DATAset 2011 を用いた分析

2.1 ボットの行動モデル

水谷らはボットネットには一定の行動モデルがあることを実証している．[3]ここで述べられている行動モデルを図 1 に示す．図 1 より S_0 から S_2 に遷移するまでの行動が一定であることが分かる．

そこで, 本章では図 1 の S_2 までの行動を IDS の Snort[4]により検知することで，BOT による攻撃を明確にする．

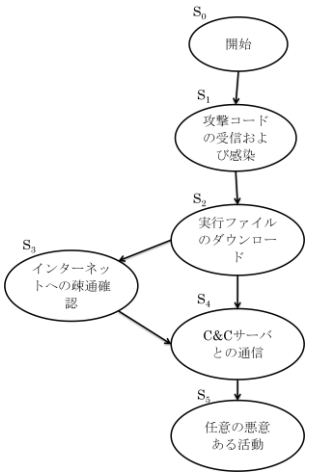

図 1　ボットの行動モデル

2.2 CCC DATAset 2011 の概要

本論文では主に CCC DATAset 2011 の攻撃通信データを用い，マルウェア感染の時間を得るため CCC DATAset 2011 の攻撃元データを用いる．攻撃通信データはハニーポット 2 台(H_1,H_2)への通信を 2010 年 8 月 18 日から 8 月 31 日と 2011 年 1 月 18 日から 1 月 31 日の計 28 日間に取得したパケットキャプチャデータである．この 2 台のハニーポットの OS は Windows XP SP1 であり，定期的にクリーンな状態にリセットされる．攻撃元データは 2010 年 5 月 1 日から 2011 年 1 月 31 日までの 9 ヶ月間に国内の複数 ISP に設置した 72 台のハニーポットにより観測されたマルウェア検体に関するログである．

2.3 攻撃元データの分析

攻撃通信データの期間に取得されたマルウェア数を知るため，攻撃元データの分析を行なう．分析の結果を表 1 に示す．2010 年では 342 検体, 2011 年では 399 検体, 計 741 検体のマルウェアをダウンロードしている．

表 1　攻撃元データの分析

	マルウェア検体数	
	2010年8月18日～31日	2011年1月18日～31日
H_1	170	185
H_2	172	214
合計	342	399

2.4 攻撃通信データの分析

本研究ではマルウェア感染までの攻撃に着目するため，OS がクリーンな状態からの通信が必要になる．そこでハニーポットの感染の有無に関わらずクリーンな状態にリセットされる特性を利用する．Windows XP は起動時に NTP サーバ(time.windows.com)にアクセスし時刻を同期ため，攻撃通信データからハニーポットが NTP サーバにアクセスし, NTP サーバからの応答が返ってきた時間で攻撃通信データを分割する．また攻撃通信データはハニーポット 2 台の通信が単一ファイルとなっているため，ハニーポット毎に分割した．これにより，3042 個のファイルが生成された．この生成されたファイルをスロットと呼ぶ.

生成されたスロットに対して Snort を適応する．Snort のルールは 2.8.5.2 Ruleset を使用し，マルウェアの検知には”This program cannot by run in DOS mode”と”Windows Program”がパケット内部に存在した場合検知するようローカルルールとしてそれぞれ定義し使用した．そして，Snort を適応した結果と，攻撃元データからマルウェアに感染した通信と感染に起因した攻撃のみを抽出した．結果を次章で解説する.

3.　分析結果

3.1 分析結果の概要

Snort による検知結果から H_1 では 350 検体，H_2 では 383 検体の計 733 検体の検知に成功し，残りの 8 検体は検知することが出来なかった．検知できなかった検体については後述する．次節からより詳細な分析結果を述べる.

3.2 検知ルール

ここでは，Snort により検知したルールとその合計をマルウェアダウンロードに起因した攻撃(表 2)とマルウェア感染(表 3)をハニーポット毎に示す．SID とは Snort で用いられるシグネチャ ID である.

表 2 より最も多かった攻撃は SID: 2466 である．このルールは，Windows のファイル共有の脆弱性を狙った攻撃である．このルール以外のルールでも「NETBIOS SMB」に関するルールがほとんどであり，これら全て Windows のファイル共有の脆弱性を狙った攻撃である．この攻撃は 2005 年から存在しているため[5]，現在でも古くから狙われている脆弱性である事がわかる．また，2010 年にはあまり見られなかった，SID:2465 が 2011 年から増加している．逆に SID:2351 は 2011 年に減少している.

表 3 より最も多いルールは SID:14415 である．これは”This program cannot by run in DOS mode”がパケット内に存在した際のルールであり, TCP プロトコルを用いたマルウェアのダウンロードが最も多かった．SID:100000241 は BOT による通信で使われていたため，マルウェアに感染した後の通信と考えられる.

表 2 攻撃ルール

SID	攻撃ルール	2010年 H_1	2010年 H_2	2011 H_1	2011 H_2	合計
2466	NETBIOS SMB-DS IPC$ unicode share access	144	146	129	141	560
2465	NETBIOS SMB-DS IPC$ share access	1	1	45	66	113
2351	NETBIOS DCERPC ISystemActivator path overflow attempt little endian unicode	18	18	4	2	42
537	NETBIOS SMB IPC$ share access	3	3	1	1	8
2492	NETBIOS SMB DCERPC ISystemActivator bind attempt	0	2	2	0	4
538	NETBIOS SMB IPC$ unicode share access	1	1	1	0	3
100000241	COMMUNITY BOT Internal IRC server detected	0	1	1	1	3

表 3 感染ルール

SID	感染ルール	2010年 H1	2010年 H2	2011 H1	2011 H2	合計
14415	DOS TCP	149	154	179	209	691
14421	win Shell	15	11	1	1	28
14419	DOS UDP	3	7	3	1	14

3.3 検知ルールに対応したポート番号

ここでは，攻撃に利用されたポート番号とマルウェアダウンロードに使われたポート番号の結果を検知ルールと対応させて表 4，表 5 に示す．

表 4 より最も多かいポート番号 445 は SID:2466,2465 のルールで使われていた．

表 5 より最も多かったポート番号 1028 は SID:14415 による感染がほとんどであった．また，感染に用いられるポート番号の大部分は 1025～1051 であることが表 5 より分かる．

表 4 攻撃に利用されたポート番号

		2010年 H1	2010年 H2	2011年 H1	2011年 H2	合計
攻撃ポート番号	135	18	18	4	2	42
	139	4	6	4	1	15
	445	145	147	174	207	673
	1029	0	0	0	1	1
	1031	0	1	1	0	2

表 5 感染に利用されたポート番号

		2010年 H1	2010年 H2	2011年 H1	2011年 H2	合計
ポート番号	1025	1	0	0	0	1
	1026	0	1	0	0	1
	1027	15	9	4	1	29
	1028	117	104	111	85	417
	1029	19	44	47	87	197
	1030	7	3	6	19	35
	1031	2	0	4	7	13
	1032	1	1	3	4	9
	1033	1	1	2	4	8
	1037	0	0	1	0	1
	1051	0	0	1	0	1
	1176	0	1	0	0	1
	1192	0	0	0	1	1
	2158	0	0	0	1	1
	3601	0	1	0	0	1
	9988	4	5	3	2	14

3.4 攻撃から感染までの IP アドレス

2010 年の通信から攻撃から感染までの IP アドレスはほとんどのものが同一であり，その全てがユニークな IP アドレスであった．

しかし，2011 年では攻撃から感染までの IP アドレスで異なるものが出てきた．また攻撃に複数回使用される IP アドレスや感染に複数回使用される IP アドレスが出てきた．この結果の一部を表 6 に示す．

表 6 より最も攻撃の多かった IP アドレスは同一のものであることが判明した．また，攻撃に使われた IP アドレスは第一セグメントが同じものが多いことが判明した．

感染に使用された IP アドレスでは「G.10.179.100」が最も多く，H_1 では 66 回，H_2 では 45 回であった．この結果，最近では攻撃ホストとマルウェアダウンロードホストを分け，一箇所からマルウェアをダウンロードするように変化していることが判明した．

3.5 検知ルールの組合せ

攻撃から感染までに Snort で検知したルールの組合せを表 7 に示す．表 7 より，攻撃から感染までの組合せとして最も多かったものは SID:2466 から SID:14415 である．

表 6 攻撃に複数回利用された IP アドレス

2011 H_1	
攻撃IPアドレス	総計
A.30.156.196	10
B.172.36.98	3
B.200.39.165	3
C.145.7.101	2

2011 H_2	
攻撃IPアドレス	総計
A.30.156.196	20
C.164.213.58	4
D.224.5.47	3
E.9.186.108	3
C.163.128.67	2
F.146.95.139	2
F.145.106.246	2
F.148.114.132	2
A.29.109.196	2
B.200.39.165	2
D.217.26.113	2

表 7　攻撃から感染までの流れ

感染までのルール組合わせ							
攻撃ルール		感染ルール	2010年 H_1	2010年 H_2	2011年 H_1	2011年 H_2	合計
537	→	14415	3	3	1	1	8
538	→	14415	1	1	1	0	3
2351	→	14421	15	11	1	1	28
2351	→	14419	3	7	3	1	14
2465	→	14415	1	1	45	66	113
2466	→	14415	144	146	129	141	560
2492	→	14415	0	2	2	0	4
100000241	→	14415	0	1	1	1	3

3.6　攻撃から感染までの間隔

3.5 で示した検知ルールの組合せの時間間隔を分析した．その結果を表 8 に示す．この結果から，攻撃から感染までの平均時間は 20 秒以下となっているが，攻撃から感染まで最大で 73 秒掛かっている通信も存在する．また組合せ毎に特徴が存在するのが分かる．例えば，「537→14415」の組合せは 9 秒から 14 秒の間で攻撃から感染が行われており，「538→14415」の組合せでは 2 秒から 5 秒の間で攻撃から感染が起きている．これは BOT が決められたアルゴリズムによって動作しているため，このような偏りが出ていると考えられる．

表 8　攻撃から感染までの時間間隔

	攻撃→感染までの時間間隔			
攻撃ルール組合せ	平均時間	標準偏差	最小時間	最大時間
538→14415	3.397	0.824	2.750	4.560
537→14415	11.744	2.372	9.060	14.600
2492→14415	10.848	1.853	8.940	12.700
2466→14415	4.298	3.407	0.220	58.800
2465→14415	3.825	7.228	0.580	73.100
2351→14421	1.005	2.636	0.250	14.800
2351→14419	0.761	1.074	0.290	4.410
100000241→14415	18.600	6.450	10.700	26.500

3.7　検知漏れの検体に関する考察

検知漏れした通信に関して該当パケットを確認したところ，Snort で攻撃を検知している場合がほとんどであった．これまで分析してきた通信は BOT がハニーポットに対してマルウェアを送信していたが，検知漏れした通信ではハニーポットから BOT に対してマルウェアをダウンロードしていた．ここでハニーポットが接続を試みたポート番号は 21，80 番などであり，ハニーポット自らマルウェアのダウンロードを試みたため，Snort では正常な通信と判定されたため誤検知したものと考えられる．

4.　提案

4.1　Snort ルールを用いた検知モデル

3 章で分析した結果より，マルウェア感染までを Snort のルールを用いて発見する検知モデルを提案する．提案するモデルを図 2，図 3 に SID に対応するルールを表 9 に示す．

このモデルは表 7 の検知ルールの組合せから作成した．SID:100000241 は IRC を利用した BOT 通信であり，この通信はマルウェアに感染後に発生する通信のため提案モデルから除外した．また，各ノードに振られている時間は表 8 の時間間隔から標準偏差が 2.5 以下の組合せに対して設定している．これ以外のノードに対しては特に時間を設けない．ルールによる検知の理由として，3.4 より 2011 年から増加した攻撃と感染が異なる IP アドレスで動作していることから，ルールによる検知とした．

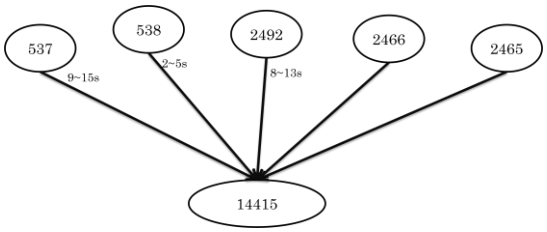

図 2　感染ルール 14415 による検知モデル

図 3　攻撃ルール 2351 による検知モデル

表 9　SID と攻撃ルール

SID	ルール
537	NETBIOS SMB IPC$ share access
538	NETBIOS SMB IPC$ unicode share access
2351	NETBIOS DCERPC ISystemActivator path overflow attempt little endian unicode
2465	NETBIOS SMB-DS IPC$ share access
2466	NETBIOS SMB-DS IPC$ unicode share access
2492	NETBIOS SMB DCERPC ISystemActivator bind attempt
14415	DOS TCP(BOT->honey)
14419	DOS UDP(BOT->honey)
14421	win Shell(BOT->honey)

4.2　DICE の概要

改ざんサイト自動検知システム DICE とは我々が先の研究で提案したシステムである．[6]このシステムは SQL Injection などによって改ざんされた WEB サイトをユーザが WEB サイトへアクセスする際に複数の特徴を用いて検知し，通信を遮断するシステムである．

4.3　DICE への機能追加

4.2 で改ざんサイト自動検知システム DICE の概要を説明した．ここでは，DICE への機能追加のためのシステム構成図を図 4 に示す．現在，DICE はプロキシサーバとして機能しており，その内部に Snort を実装することで IDS としての機能を持たせる．Snort が検知した通信をログに書き込む際，中間に検知モデルを置くことで Snort が検知した通信を検証し，検知モデルに合致するものであれば，DICE でその通信を遮断する仕組みである．これにより，IPS としての機能を持たせる事が出来る．

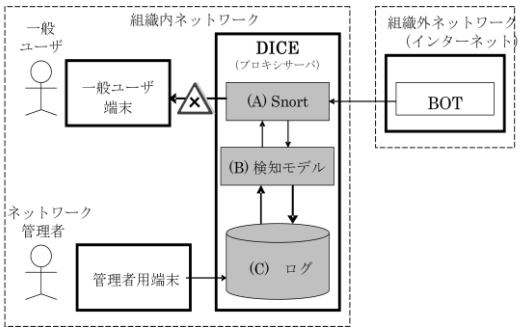

図 4　システム構成図

5.　おわりに

本論文では，CCC DATAset 2011 の攻撃通信データを用いて BOT の攻撃からマルウェアに感染するまでの通信を分析し，分析結果から攻撃からマルウェア感染までの IDS のルールを抽出した．そして，抽出した IDS のルールを組合せる事で BOT 通信の検知に最適なモデルを提案し，改ざんサイト自動検知システム DICE の新たな機能として，提案した．

今後の課題として，今回提案した Snort ルールによる検知モデルが，新種の BOT に対応出来るかを確認するため継続的な調査が必要になる．検知モデルを改ざんサイト自動検知システム DICE に実装し，実験を行う事も必要になってくる．

参考文献

[1] 「ボットネット」の正体を探る / SAFETY JAPAN [特集] / 日経 BP 社 http://www.nikkeibp.co.jp/sj/2/special/63/index2.html

[2] 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 2011 Datasets~, MWS2011(2011 年 10 月)

[3] 水谷正慶，他：通信の状態遷移に着目したボット活動の調査，マルウェア対策研究人材育成ワークショップ 2008(MWS2008),p.25-30

[4] Snort：http://www.snort.org/

[5]インターネットセキュリティインテリジェンスブリーフィング/日本ベリサイン株式会社 https://www.verisign.co.jp/basic/isib/pdf/isib_200506.pdf

[6]田中達哉，他：改ざんサイト自動検知システムの開発と評価，コンピュータセキュリティシンポジウム 2010(CSS2010)，p.531-536