多地点分散型ハニーポットによる攻撃発生および被害予報システムの検討

須藤　年章　　畑田　充弘

NTT コミュニケーションズ株式会社
sudo@mfeed.ad.jp　　m.hatada@ntt.com

あらまし　数年前までは，脆弱性探索や攻撃活動はインターネット全体に対して網羅的に行われることが多く，小規模な検出システムでもインターネット全体の状況をとらえることが可能であった．しかし，近年，検出の回避や攻撃の効率性の向上などのため攻撃対象となる IP アドレスは限定的なものとなり小規模な検出システムではインターネットで発生している問題の多くを検出することができない．本稿では，複数のハニーポットシステムで収集したデータを利用し，各検体の通信特性に着目した攻撃の識別，影響度の考察を行うとともに，AS，IP アドレス等の違いによる検体検出状況を組み合わせることにより，より効率的な感染予測システムの検討を行う．

Examination of attack generation and damage forecast system by multi point decentralized honeypot

Toshiaki Sudoh　　Mitsuhito Hatada

NTT Communications Corporation
sudo@mfeed.ad.jp　　m.hatada@ntt.com

Abstract The vulnerability search and the attack activity were able to be often done to the entire Internet covering it, and to catch the situation of the entire Internet even with a small-scale sensing system several years ago. However, Internet Protocol address to be attacked for the evasion of detection and the improvement of the efficiency of the attack, etc. becomes limited in recent years and many of problems that occur on the Internet cannot be detected with a small-scale sensing system. In this text, a more efficient infection forecast system is examined by identifying the attack that pays attention to the communication property of each specimen material by using the data collected with two or more honey pot systems, considering the influence level, and combining the specimen material detection situations by differences between AS and Internet Protocol address, etc.

1.　はじめに

古くは Blaster 等のワームの活動，近年ではボットによる脆弱性スキャンなどは IP アドレスを網羅的に対象とすることが多く，そのような攻撃特性のため小規模な検知システムでもインターネット全体で発生している事象を検出することが可能だった．ところが最近では，ボットの隣接ネットワークのみに限定した攻撃から，CnC の命令による無関係な特定ネットワークのみを攻撃対象とした攻撃などにより攻撃対象が限定されるようになり，小規模な検出システムでは，限定されたアドレス空間に対する攻撃しか検出できなくなっている．このため公開される攻撃情報が自ネットワークとはまったく関係ない場合や，逆に一部のネットワークのみでしか観測されない事象に関しては，必要な対策情報などがまったくない場合があり被害を拡大させてしまう．

本研究では CCC DATAset 2009[3]の攻撃元データ（以降，攻撃元データと呼ぶ）を利用し，マルウェアの種類の違いによる捕捉特性およびマルウェアに感染することによって発生する影響トラフィックを解析し，さらに独自のハニーポットのデータを用い，AS ごとの捕捉特性を解析することで，各 AS への広がりやどのような影響を与えるのか予測する仕組みについて検討する．

2. 脆弱性スキャントラフィックの変化

攻撃に利用するボットを作成するために端末をマルウェアに感染させる手法は，近年では感染効率の向上や検知を妨害することなどを目的とするようになり，OS やアプリケーションの脆弱性を利用するものから，利用者の多いその時々の流行のインターネットサービスを利用するもの，ソーシャルエンジニアリングを利用するものなど古くから利用されている手法から新しいものまで様々な手法が次々と生み出されている．その手法を分類すると次のようになる．

(1) ネットワーク経由での脆弱性スキャン
(2) 悪意のあるソフトウェアを実行させる
(3) ブラウザの脆弱性を利用した Web 経由での感染

近年は，その効率の高さからブラウザおよび関連するアプリケーションの脆弱性を利用した感染手法が広く用いられているが，攻撃の基本機能としてネットワーク経由での脆弱性スキャン活動は手法を変えつつ継続的に行われている．脆弱性スキャン活動には以下のような機能パターンがある．

(1) IP アドレス網羅的攻撃（シーケンシャル）
(2) IP アドレス網羅的攻撃（ランダム）
(3) 感染端末の IP アドレスを元にした隣接攻撃
(4) 対象を特定した攻撃（CnC からの命令）

(1)は始点 IP アドレスからアドレスをインクリメントしながら攻撃をしていくもの，(2)はランダムにアドレスを決めながら攻撃するもので，そのアルゴリズムに差異はあるもののインターネットに対して網羅的に攻撃トラフィックが送出されるため検出しやすい．(3)はボットの基本機能として多く観測されるものであり，そのボットの近接ネットワークには，そのボットが感染に利用された脆弱性と同じ脆弱性をもつ端末が多く存在するであろうという想定のもと感染効率を高めるためにおこなわれるものである．

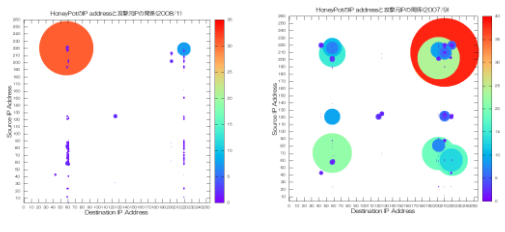

図 1 スキャントラフィックの変化

図 1 の右側のグラフが隣接ネットワークへの攻撃で左側のグラフが特定ネットワークへの攻撃の例である．このような攻撃手法の変化により小規模な検出システムや，darknet を利用したシステムなどでの捕捉効率は非常に悪くなっている．またそれ以外に送信元に着目した場合は，ボットだけではなく無料のウェブサービス，ホスティングサービスの利用の拡大により，送信元特定の意味を失わせるような攻撃が増えている．

3. マルウェアごとの捕捉特性

表 1 に，同一のマルウェアが観測期間中に捕捉された頻度について示す．1 年間でも 10 回以下のものが 97.51%存在するのは，大量の亜種が，限定期間にのみ活動していたものと推定される．

表 1 マルウェア捕捉頻度

マルウェア捕捉頻度	割合[%]
10 以下	97.51
10～100	2.00
100～1000	0.44
1000～2000	0.02
2000 以上	0.03

表 1 のとおり各マルウェアの捕捉頻度の差は非常に大きいため本研究では，捕捉頻度 10 以下，100 から 1000，1000 以上の 3 つの捕捉頻度カテゴリに分け，その中から任意に選択したマルウェアについて，ハニーポット毎の捕捉特性およびそのマルウェアに感染することによって発生する攻撃の影響を考察する．表 2 に解析に利用したマルウェアとそのハッシュ数を示す．

表 2 解析に利用したマルウェア

マルウェア名	ハッシュ数
PE_VIRUT.AV	9863
BKDR_VANBOT	231
TROJ_AGENT	192
WORM_SDBOT	177
TSPY_KOLABC.CH	11

3. 1 CCC ハニーポットの捕捉特性

CCC DATAset2009[3]攻撃元データにはハニーポットの識別子は含まれているが，IP アドレスや AS に関する情報はないため，単純に別 NW に接続されたハニーポットのマルウェア捕捉特性の違いとして解析する．ここでは表 2 の中から二種類のマルウェアの解析結果を例としてあげる．

グラフはX軸が時間を，Y軸がハニーポットの番号を表し，プロットの色の違いがハッシュ値の違いを表す．

3. 1. 1 PE_VIRUT.AV

TROJ_AGENT の捕捉特性を図 2 に示す．すべてのマルウェアに共通的にX軸に平行に表れる長期間捕捉特性とY軸に平行して表れる短期蔓延特性が見られる．2008 年 5 月以降継続観測されており特に 30～40 番のハニーポットが集中的に捕捉していることがわかる．また一ヶ月から二ヶ月周期でほぼすべてのハニーポットで数日間のみ同時捕捉される新規ハッシュが存在することがわかる．定期的なバージョンアップが行われている典型的な例と想定される．またこのマルウェアも同一のハッシュが時期をあけて再度観測されている．また，一度捕捉されたハッシュが数カ月以上のちに同一ハニーポットもしくは別ハニーポットで捕捉される特性を持つマルウェアもある．図2では2008年6が上旬に一週間程度の期間すべてのハニーポットで網羅的に検出されたハッシュが三カ月に一部のハニーポットで約二ヵ月間再度捕捉されている．この原因としては攻撃者がそのマルウェアの再利用を行った，攻撃ツール，ボットの再利用時に操作を失敗したなどさまざまな要因が考えられる．この特性を詳細に解析するためには，期間を空けた観測されたハッシュの実際の効果，影響，被害の調査が必要と考えられるが，ここでは，実際の観測状況として短期間で攻撃が終了したと思われたマルウェアが数ヶ月から数年後の再利用される危険性があることを示す．つまり最初に検出した時に影響度が小さなものでも数カ月以上たって影響度を増すということも懸念される．

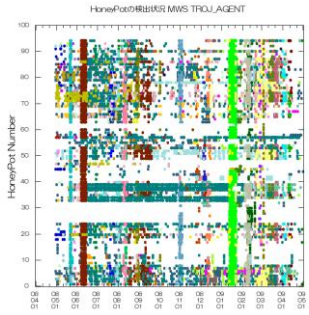

図2 TROJ_AGENT

3. 1. 2 TSPY_KOLABC.CH

TSPY_KOLABC.CH の捕捉特性を図3に示す．このマルウェアは網羅的に二ヶ月間捕捉され続けたハッシュと同時発生したが一ヶ月間のみ捕捉されたハッシュが目立つ．ただし後者のハッシュはそれ以後も離散的に観測されつづけているが，その意味，影響については不明である．

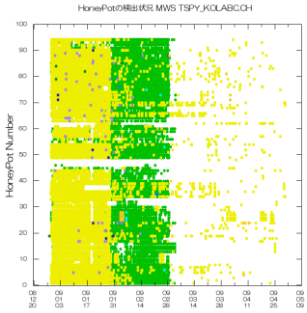

図3 TSPY_KOLABC.CH

3. 2 マルウェア毎の発生トラフィック特性

次にそれぞれのマルウェア毎に感染の発生によるネットワークリソースへの影響度を予報するためのトラフィック特性について述べる．トラフィック特性として，各ハッシュに対応する検体を 3 分間の動的解析にかけた動作記録データの一部である通信記録を用いる．疑似環境における通信記録であるため実際のトラフィック量はわからないため，ここではセッション数によりトラフィック特性を表し，DNS, IRC, HTTP, SMTP を対象とする．

図4はPE_VIRUT.AV の解析データである．検体収集数のグラフは, X 軸は時間を, Y 軸は検体収集数を表す．その他のグラフはX軸が時間を，Y 軸はハッシュ毎のセッション数を表す．なお，全てのハッシュに対応する動作記録データは得られていない．

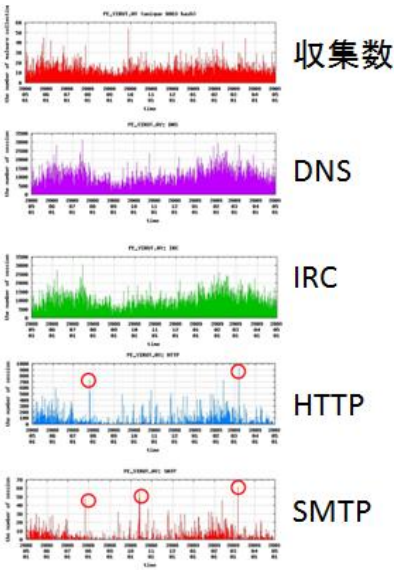

図4 PE_VIRUT.AV の発生トラフィック特性

PE_VIRUT.AV はハッシュ数も多くハッシュによっては感染後の動作が異なるための時期によってトラフィック特性に差が表れている．DNS，IRC はこのマルウェアに感染した際の動作として必ず発生するもので，あるが，攻撃その他に利用される HTTP や SMTP には特徴的なパルスが表れている．

これらの結果から PE_VIRUT.AV は感染時に必ず発生する DNS，IRC，HTTP 通信によって発生するセッション数の影響が非常に大きく，その後の攻撃に利用された SMTP のセッション数影響がわずかに発生すると想定できる．したがって感染が広がった場合は，DNS サーバーへの影響が最も大きく，次に IRC，HTTP などのセッション増によるネットワーク設備への影響が予測することができる．

その他のマルウェアについても同様な解析を行い，発生セッション数と発生頻度を利用した単純に影響度を数値化したものを表 3 にまとめる．それぞれの数字は大きいほど影響度が大きいことを表す．

表 3　各マルウェアの影響度評価

マルウェア	DNS	IRC	HTTP	SMTP
PE_VIRUT.AV	100	56	26	2
BKDR_VANBOT	37	37	0.7	0
TROJ_AGENT	11	11	15	0
WORM_SDBOT	31	31	7.5	0
TSPY_KOLABC.CH	50	50	8.8	0.6

実際のネットワークへの影響に関しては定量的な評価が必要であるが，ここでは単純にこの解析で算出された数字をもとに発生予測されるトラフィック規模として 5 段階評価を行いまとめたものが表 4 である．

表 4　各マルウェアの影響度指標

マルウェア	DNS	IRC	HTTP	SMTP
PE_VIRUT.AV	A	B	C	D
BKDR_VANBOT	C	C	E	-
TROJ_AGENT	D	D	D	-
WORM_SDBOT	C	C	D	-
TSPY_KOLABC.CH	B	B	D	E

4.　AS 別捕捉特性

次に，ISP15 社に設置した別のハニーポットを利用し，AS の違いによる捕捉特性についてのべる．本稿では AS1～AS15 という識別子で個々の AS を表現する．

4. 1　マルウェア毎の AS 別捕捉特性

この解析ではハッシュごとに各 AS のハニーポットの捕捉特性を一つのグラフに表している．これによりそのマルウェアが各 AS でどのようなタイミングでどのような規模で観測されるかを可視化する．ここでは二つのマルウェアを例として解析する．

4. 1. 1　BKDR_VANBOT

BKDR_VANBOT の中から一つのハッシュについてその特性を図 5 に示す．

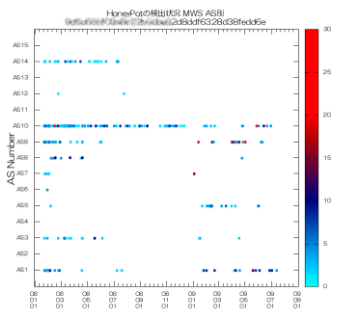

図 5　BKDR_VANBOT

AS9 で最初に検出した後，11 時間後に AS10，さらに 10 時間後に AS1，さらに 3 時間後に AS3，AS7，AS9 で捕捉している．AS10 で長期に捕捉され続けているがそれ以外の AS については初期捕捉から一年後に再度活発に捕捉されている．

4. 1. 2　TSPY_KOLABC.CH

TSPY_KOLABC.CH の中から一つのハッシュについてその特性を図 6 に示す．

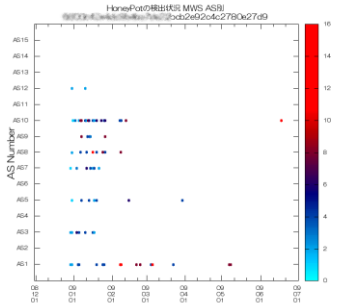

図 6　TSPY_KOLABC.CH

AS1，AS3，AS7 で最初に検出し一日後に AS5，AS8，AS12 で検出しさらに一日後に AS10 に広がっている．こ

のマルウェアもまったく検出できない AS がある．また最初の検出から 2 ヵ月後に AS1 で活発に捕捉し，さらに 3 ヵ月以上あけて AS1 で再度，さらにその後 AS10 で再捕捉されている．

4．2　AS 毎のマルウェア捕捉タイミングに関する解析

このように，AS の違いによりマルウェアの捕捉性，捕捉タイミングなどに違いがある．これらの特性をさらに解析することでハニーポットの捕捉性の向上や AS の関係から次にどの AS に感染が広がるかなどについての予測を立てることができないか検討する．

4．2．1　単独 AS での捕捉性

前述のとおり同じマルウェアにより全体の捕捉頻度は大きく異なり表 1 に示す通り全体の 97%が 10 回以下しか補足されていない．図 7 にマルウェアの全体捕捉頻度と捕捉できた AS の数の関係を示す．

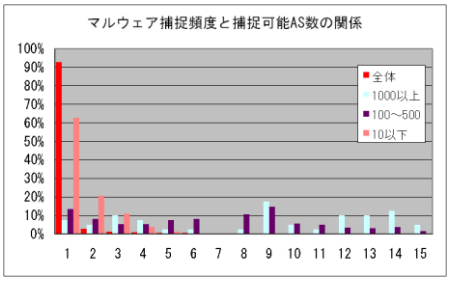

図 7　マルウェア捕捉頻度と捕捉可能 AS 数の関係

捕捉頻度カテゴリ別にみると，1000 回以上捕捉されたマルウェアのうち 1AS のみでしか検出できなかったものは 6.25%，9AS 以上で捕捉できたのは 62.51%となる．逆に 10 回以下のものについては 62.66%が 1AS でしか捕捉できない．全マルウェアでみると 92.95%は 1AS でしか検出されていない．このように大規模感染型のマルウェアの 93.75%は複数の AS で検出され，小規模なものほど一部の AS でのみ検出されるということを表している．

この結果からさらに単独で捕捉することの多かった AS を図 8 に示す．

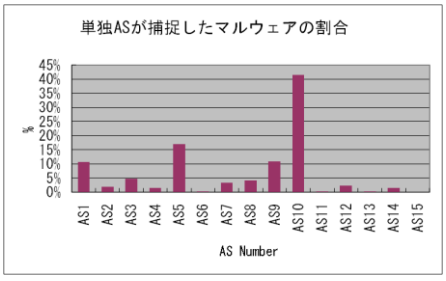

図 8　単独の AS での検出率

ハニーポットの捕捉効率の観点からは，上位 3AS でも 70%程度をしめるためこれらを有効に利用することで効率の向上が考えられるが，残りの 30%が網羅されない状況はマルウェア対策には大きな問題であるため，その他の AS もカバーする必要がある．

4．2．2　AS 毎のマルウェア早期捕捉度

次に図 9 に各 AS がマルウェアを最も早く捕捉したかについて示す．ここでも同様にマルウェアの全体での捕捉頻度を 3 カテゴリに分類し，それぞれについて解析した．いずれのカテゴリでも AS10 が早期捕捉できている．AS4，AS9 は大規模なものほど早期捕捉でき AS5，AS10 は小規模なものの早期捕捉できている．また，数は少ないがその他 AS についても他の AS よりも早期捕捉できるマルウェアが存在することを示している．

これらの特性を利用して各 AS に設置する捕捉システムの規模を検討することで捕捉効率を上げることができると想定される．たとえば全体的に捕捉効率の高い AS10，AS1，AS5，AS9 により多くの捕捉システムを設置することによって捕捉効率を上げることができるのではないかと想定される．

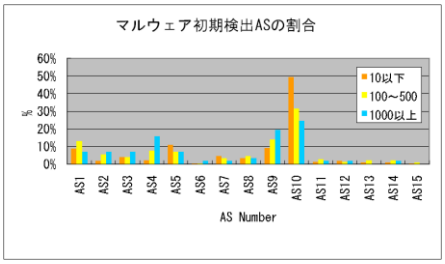

図 9 マルウェア初期検出 AS の割合

4．2．3　AS 毎のマルウェア捕捉時差

次にマルウェアが，ある AS で初期捕捉された後別の AS（第二捕捉 AS）で捕捉されるまでの時差について解析する．まず，検出頻度の高い上位 50 件のマルウェアに関して初期捕捉 AS と第二捕捉 AS の関係をまとめると図 10 のようになる．

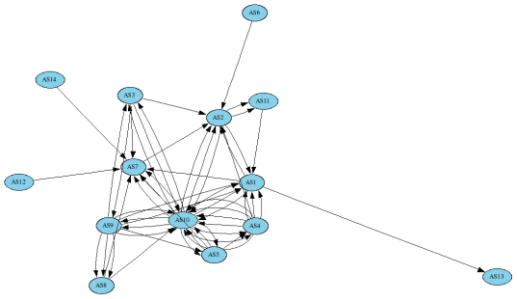

図 10　AS 間の捕捉タイミングの関係

矢印の向きが伝染方向を表し，AS 間の距離は時差を表している．AS6，AS12，AS14 は初期捕捉 AS としてのみ，AS13 は第二捕捉 AS としてのみ機能しており，その他の AS に関しては相互に依存しあっている．AS13 は AS1 からのみ伝染するがかかる時間は非常に長いということを表す。

これらの関係をわかりやすくするために第二捕捉 AS に着目して解析する．表 4 は第二捕捉 AS としての AS10，AS1 とその初期捕捉 AS との関係を示す．

AS10 について見ると全 15AS 中 7AS で捕捉されたのち AS10 で捕捉されている．したがってこの 7AS で捕捉されたことが次に AS10 で捕捉される可能性があることの指標として利用することができるのではないかと考えられる．

さらに AS10 に対する初期捕捉 AS として認識された度数の高い AS については，単純に，その AS で，あるマルウェアが捕捉された場合は AS10 に伝染する可能性を高い考えることができる．表 4 の例では AS4 が最も高い．

また初期捕捉 AS と第二捕捉 AS のマルウェア捕捉の平均時差も要素の一つと考えられる．平均時差が小さいほど関係性が強いと考えることができるが，AS1 と AS9 の関係のように 634.5 時間のような極端な場合もあり，前述のとおり，数か月以上の時差がある例も多々あるため，その要素の有効性の判断は難しい．

これらの検討から，この例では，7AS 中，度数の一番多い AS4 が捕捉したあと平均 12.6 時間で AS10 でも捕捉する可能性が最も高いと考えることができる．その他の 7AS での捕捉についても確率は下がるが AS10 での捕捉可能性があると考える．

表 4 初期補足 AS と捕捉時差の例

AS10			AS1		
初期捕捉 AS	時差 (h)	度数	初期捕捉AS	時差 (h)	度数
AS1	1	1	AS2	4	1
AS2	20	1	AS4	5.2	3
AS3	1	1	AS9	634.5	2
AS4	12.6	5	AS10	11	2
AS5	13.5	2	AS11	20	1
AS8	26	1			
AS9	11.5	2			

同様に AS1 とその初期捕捉 AS との関係については度数の一番高い AS4 が捕捉したあと 5.2 時間で AS1 でも捕捉する可能性が高いと考えることができる．

5. 予報システムに関する考察

ここまで解析した結果から，あるマルウェアが特定の AS で捕捉された後，どの AS に広がり，どのような影響が発生するかの予報についての一例として，表 5 に BKDR_VANBOT を用いた被害予報の例を示す．AS9 で捕捉された場合，12 時間以内に AS10，24 時間以内に AS1，AS3，AS7 に広がり，DNS への影響度が大きく，IRC，SMTP 等により発生するセッション増に注意の必要があると予測される．

表 5 BKDR_VANBOT の予報

初期捕捉 AS	第二捕捉 AS	第三捕捉 AS	初期捕捉からの時間 (h)	被害予測
AS9	AS10		11	DNS B
		AS1	21	IRC B
		AS3	24	HTTP D
		AS7	24	SMTP C

6. まとめ

本研究では AS に分散配置したハニーポットを利用することで，AS 毎に捕捉できるマルウェアの違いや，捕捉タイミング，AS 間の関係などの AS に依存したマルウェア捕捉特性を示し，その差異を利用したマルウェア感染の拡大の予測についての検討を行い一例を示した．今後は，マルウェアのバリエーションを増やすなどサンプルを増やし予測指標のより詳細な解析や，自動システム化についての検討を行っていく．

謝辞

本研究の一部は総務省の「スパムメールやフィッシング等サイバー攻撃の停止に向けた試行」の一環で行った調査・研究結果をまとめたものである．また，本研究を進めるにあたり，有益な助言と協力を頂いた Telecom-ISAC Japan の関係者各位に深く感謝致します．

参考文献

[1] http://project.honeynet.org

[2] http://www.ccc.go.jp

[3] 畑田充弘，中津留勇，寺田真敏，篠田陽一：”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