多種多様な攻撃に用いられるIPアドレス間の相関解析

千葉 大紀† 八木 毅‡ 秋山 満昭‡ 森 達哉†† 後藤 滋樹†

†早稲田大学基幹理工学研究科 169-8555 東京都新宿区大久保 3-4-1
{chiba,goto}@goto.info.waseda.ac.jp

‡NTT 情報流通プラットフォーム研究所 180-8585 東京都武蔵野市緑町 3-9-11
{yagi.takeshi,akiyama.mitsuaki}@lab.ntt.co.jp

††NTT サービスインテグレーション基盤研究所 180-8585 東京都武蔵野市緑町 3-9-11
mori.tatsuya@lab.ntt.co.jp

あらまし マルウェアの活動による攻撃は，日々複雑化や高度化が進んでいる．それに伴い，マルウェアの攻撃対象はOSの脆弱性だけではなく，Webブラウザや Webアプリケーションの脆弱性にまで及んでいる．このような多種多様な攻撃を特定・分析するために，攻撃対象に応じたハニーポットが研究開発され，攻撃情報が収集されている．しかし，配置可能な各ハニーポット数にはコストに起因した限界があり，より効率的に攻撃情報を収集する必要がある．そこで本稿では，特定のIPアドレスが複数種類の攻撃に使用される可能性に着目し，多種多様なハニーポットで収集したIPアドレスを横断的に解析する技術を提案するとともに，IPアドレス間の特性を調査した結果を報告する．

Correlation Analysis Between IP Addresses Used in Variety of Attacks

Daiki Chiba† Takeshi Yagi‡ Mitsuaki Akiyama‡ Tatsuya Mori ††
Shigeki Goto†

†Graduate School of Fundamental Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, JAPAN
{chiba,goto}@goto.info.waseda.ac.jp

‡NTT Information Sharing Platform Laboratories
3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585, JAPAN
{yagi.takeshi,akiyama.mitsuaki}@lab.ntt.co.jp

††NTT Service Integration Laboratories 3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585, JAPAN
mori.tatsuya@lab.ntt.co.jp

Abstract Malware now attacks not only vulnerabilities in the Windows operating system, but also those of web applications on servers and web browsers on personal computers. To collect attack information for protective use, many types of honeypots have been developed. However, attack information that can be collected by each type of honeypot is limited, so it is necessary to collect it more efficiently. Our proposal performs correlation analysis of IP address structures in attack data collected by various types of honeypots, and detects IP addresses of attackers which are used for various attacks. Using IP addresses allows us to detect certain types of malware infection attacks efficiently.

1　はじめに

近年，マルウェア感染において攻撃対象となる脆弱性が多様化している．従来，OS の脆弱性を対象とする攻撃が多かったが，Web ブラウザの脆弱性を対象とした端末への攻撃や，Web アプリケーションの脆弱性を対象としたサーバへの攻撃が増加している [1]．これらの攻撃への対策に向けて，各脆弱性に対応した多種多様なハニーポットで攻撃を収集する必要がある．さらに，攻撃の多様化はマルウェア感染が成功する確率を高めており，攻撃者に悪用されるボットの数も増加している．一方，設置可能なハニーポット数はコスト的に制限される．このため，複数種類のハニーポットから効率的にボットの情報やマルウェア検体を収集する必要がある．そこで本稿では，特定の IP アドレスが複数種類の攻撃に使用される可能性に着目し，多種多様なハニーポットで収集した IP アドレスを横断的に解析する技術を提案するとともに，IP アドレス間の特性を調査した結果を報告する．

2　ハニーポットによる攻撃情報の収集

2.1　多種多様なハニーポットの構成

マルウェア感染攻撃では，攻撃対象の脆弱性によって攻撃方法が異なる．このため，図 1 に示すように，多種多様なハニーポットを用いることで各種攻撃情報を収集する．

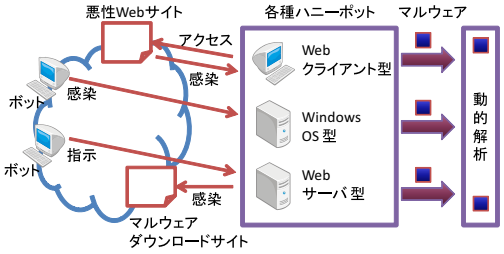

図 1: 攻撃情報収集

Windows OS への攻撃を収集するハニーポット [2][3] では，マルウェアに加え，攻撃元の IP アドレスを特定できる．また，Web アプリケーションへの攻撃を収集する Web サーバ型ハニーポット [4] では，マルウェアに加え，攻撃元の IP アドレスと，攻撃対象にマルウェアをダウンロードさせるためのマルウェアダウンロードサイトの IP アドレスや URL を収集できる [5]．一方，Web ブラウザへの攻撃を収集する Web クライアント型ハニーポット [6] では，Web サイトを巡回することで，アクセスした際にリダイレクトさせられる Web サイトや攻撃を送信する Web サイトの URL や IP アドレス [7] を収集できる．

各ハニーポットで収集した情報は，ユーザ環境とインターネット間に配置されるセキュリティアプライアンスにおいて，ブラックリスト情報として活用できる．例えば，Web クライアントを収容するセキュリティアプライアンスでは，Web クライアント型ハニーポットで収集した URL や IP アドレスとの通信をフィルタする．また，Web サーバを収容するセキュリティアプライアンスでは，Web サーバ型ハニーポットで収集した URL や IP アドレスとの通信をフィルタする．さらに，各ハニーポットで収集した検体をマルウェア動的解析器 [8] を用いて解析することで，Command and Control (C&C) サーバの URL や IP アドレスをブラックリスト情報として特定できる．

2.2　ハニーポットで収集した攻撃者 IP アドレス分布の調査

あるハニーポットで収集した URL や IP アドレスが，複数種類の攻撃に使用されていれば，一種類のハニーポットで収集した情報を複数種類の攻撃に共通したブラックリスト情報として効率的に活用できる．そこで，複数種類の攻撃で使用される IP アドレスの有無を確認するために，Windows OS 対応ハニーポット DenDenHoney[3] と Web クライアント型ハニーポット Marionette[6] と Web サーバ型ハニーポット Web Phantom[5] で収集した攻撃元 IP アドレス，悪性 Web サイトの IP アドレスおよびマルウェアダウンロードサイトの IP アドレスの一

表 1: 分析対象の攻撃者 IP アドレス概要

データセット	収集した攻撃情報	収集期間	IP アドレス数
DenDenHoney (DDH)	WinOS への攻撃元 IP アドレス	2011-05-10〜2011-07-21	21,111
Marionette (Mari)	悪性 Web サイトの IP アドレス	2010-11〜2011-02	12,245
WebPhantom (WPa)	Web サーバへの攻撃元 IP アドレス	2009-12-24〜2011-07-24	5,912
WebPhantom (WPm)	Web サーバからの誘導先 IP アドレス	2009-12-24〜2011-07-24	992

致性を調査した．これらの IP アドレスの総称を攻撃者 IP アドレスと定義する．各ハニーポットで収集した攻撃者 IP アドレスの概要を表 1 に示す．さらに，各ハニーポット間で重複して観測された攻撃者 IP アドレス数を表 2 に示す．表 2 に示すように，393 の IP アドレスが複数種類のハニーポット間で重複して観測されていた．この結果，複数の攻撃に共通的に使用される攻撃者 IP アドレスの存在を確認できた．しかし，複数種類の攻撃に対するブラックリストを一種類のハニーポットで効率的に生成するためには，あるハニーポットで収集した攻撃者 IP アドレスが異なる種類のハニーポットで収集される可能性を推定する手法が必要となる．

表 2: 重複 IP アドレス数

	DDH	Mari	WPa	WPm
DDH	-	56	18	2
Mari	-	-	76	157
WPa	-	-	-	96
WPm	-	-	-	-

3 異なるハニーポットで収集した IP アドレス間の相関解析

3.1 概要

本稿では，IP アドレス間の構造的な距離に着目し，異なるハニーポットで収集した IP アドレスが密集する領域の IP アドレスは，複数のハニーポットで観測される可能性が高いと推測した．そこで，異なるハニーポットで収集した攻撃者 IP アドレスの近接性と特徴を調査した．具体的には，各ハニーポットで収集した IP アドレスを，ヒルベルト曲線に基づく 2 次元グラフ上に配置した．さらに，2 次元グラフ上のマンハッタン距離に応じて攻撃者 IP アドレスをクラスタリングし，異なるハニーポットで収集した攻撃者 IP アドレスが混在するクラスタの特徴を調査した．詳細を以下に示す．

3.2 攻撃者 IP アドレスの 2 次元グラフ化

ヒルベルト曲線は，再帰的に定義される空間充填曲線である．ヒルベルト曲線は，U 字型の形状を基本図形として，図 2 に示すように，以下の式の再帰的な組み合わせで描画される．

$$\mathrm{DRU}(n) = \mathrm{RDL}(n\text{-}1) \downarrow \mathrm{DRU}(n\text{-}1) \rightarrow \mathrm{DRU}(n\text{-}1) \uparrow \mathrm{LDR}(n\text{-}1)$$

$$\mathrm{LUR}(n) = \mathrm{ULD}(n\text{-}1) \leftarrow \mathrm{LUR}(n\text{-}1) \uparrow \mathrm{LUR}(n\text{-}1) \rightarrow \mathrm{DRU}(n\text{-}1)$$

$$\mathrm{ULD}(n) = \mathrm{LUR}(n\text{-}1) \uparrow \mathrm{ULD}(n\text{-}1) \leftarrow \mathrm{ULD}(n\text{-}1) \downarrow \mathrm{RDL}(n\text{-}1)$$

$$\mathrm{RDL}(n) = \mathrm{DRU}(n\text{-}1) \rightarrow \mathrm{RDL}(n\text{-}1) \downarrow \mathrm{RDL}(n\text{-}1) \downarrow \mathrm{ULD}(n\text{-}1)$$

D:Down, L:Left, R: Right, U: Up

ここで，n はヒルベルト曲線の次数を示しており，式中の矢印は各方向への線分の描画を示している．

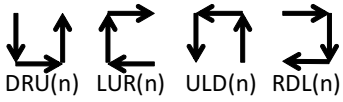

図 2: 描画ルール

ヒルベルト曲線上への IP アドレスの配置に関しては，IP アドレスの隣接構造を保持しつつ近傍の IP アドレスを空間的に近い順に配置する手法 [9] が検討されている．本稿では，IPv4 アドレスの第 1〜3 オクテットの情報を 12 次の

ヒルベルト曲線上に配置することで 2 次元グラフを作成する．配置結果の一部を図 3 に示す．

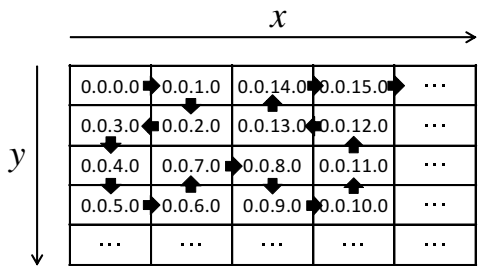

図 3: IP アドレスのヒルベルト曲線上への配置

3.3 攻撃者 IP アドレスのクラスタリング

2 次元グラフ上に IP アドレスを配置することで，IP アドレス間の距離を座標によって定義できる．本稿では，ヒルベルト曲線を用いた格子上に IP アドレスを配置していることから，距離関数としてマンハッタン距離を適用する．k 次元ベクトル点 $A(a_1, a_2, \cdots, a_k)$，点 $B(b_1, b_2, \cdots, b_k)$ があるとき，AB 間のマンハッタン距離 d は次式で定義できる．

$$d(A, B) = \sum_{i=1}^{k} |a_i - b_i|$$

ただし，IP アドレスが距離的に隣接していたとしても，複数の AS に割り当てられた各 IP アドレスの特徴はネットワーク構造的に異なる．そこで，点 A と点 B に相当する IP アドレスが異なる AS 番号を持つ場合は $d(A, B) = \infty$ とすることで，ネットワーク構造を距離に反映させる．距離に基づいて IP アドレスをクラスタリングし，異なるハニーポットで収集した IP アドレスが混在するクラスタを調査する．本稿で適用した階層的クラスタリングの手順を以下に示す．ここでクラスタ間の距離は，最遠隣法を用いて定義する．最遠隣法とは，各クラスタから抽出した IP アドレスの最長距離をクラスタ間の距離とする方法である．

1. IP アドレス間の距離 d を計算する．
2. 距離 d が最小の IP アドレス間でクラスタを生成する．
3. 生成したクラスタと他クラスタおよび他 IP アドレスに対して，距離が最小の 2 つを結合してクラスタを生成する．
4. すべてのクラスタ，IP アドレスが結合されるまで 1〜3 を繰り返す．

上記のクラスタ形成過程の一部を図 4 に示す．図 4 では，縦軸が IP アドレス間の距離 d を示しており，横軸に IP アドレスを配置し，距離 3 において縦軸で各 IP アドレスを結合することで，デンドログラムを生成している．ここで，距離に応じて木構造を分割し，一定の距離以内の IP アドレス群をクラスタとして抽出する．各クラスタに異なるハニーポットで収集した攻撃者 IP アドレスが混在する場合，このクラスタを混在クラスタと定義し，異なる種類の攻撃が近傍の IP アドレスを用いて実施されていると判断する．

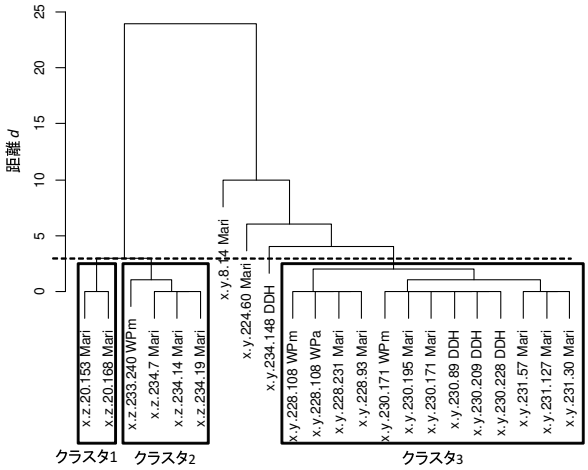

図 4: クラスタ形成過程の例

4 攻撃者 IP アドレスの分析

4.1 概要

提案の相関解析では，クラスタ化する IP アドレス間の距離に応じて，クラスタ内の IP アドレスの特性が変化する．そこで，距離に応じたクラスタ数とクラスタ内 IP アドレス数の変化を調査した．なお，今回は，表 1 に示すデータを用いた評価実験 1 と，CCC DATAset 2010・

2011[10] と D3M 2010･2011[10] を用いた評価実験 2 を実施した．評価実験 2 で用いた IP アドレスの概要を表 3 に示す．さらに，Windows OS への攻撃を収集した結果である CCC DATAset と Web クライアントへの攻撃を収集した結果である D3M において重複した IP アドレス数を表 4 に示す．

表 3: 各データセットの IP アドレス数

データセット	収集期間	IP アドレス数
CCC2010	2010-03	13,337
CCC2011	2011-01	8,775
D3M2010	2010-03	257
D3M2011	2011-02	151

表 4: 重複する IP アドレス数

	CCC 2010	CCC 2011	D3M 2010	D3M 2011
CCC2010	-	24	2	0
CCC2011	-	-	0	0
D3M2010	-	-	-	20
D3M2011	-	-	-	-

4.2 実験結果

各評価実験において，クラスタ化する際の距離に対する，総クラスタ数や混在クラスタ数および混在クラスタ内の総 IP アドレス数を調査した結果を表 5 と表 6 に示す．一般的に，距離の増加に伴ってクラスタ内の IP アドレスは増加する．このため，距離に対して，総クラスタ数は単調減少し，混在クラスタの発生確率は単調増加する．表 5 では，距離 6 において混在クラスタ数が最大値を示している．これは，1 つの混在クラスタ中に含まれる IP アドレス群の範囲が広がり，全体の混在クラスタ数が減少したためだと考えられる．さらに，表 5 では，混在クラスタ内の総 IP アドレス数は距離に応じて増加している．総 IP アドレス数の増加は，異種ハニーポット間で重複して観測される可能性が低い IP アドレスが混在する原因となる可能性もある．このため，距離 6 によってクラスタリングすることで，異種ハニーポット間で重複して観測される可能性が高い IP アドレスを効率的に抽出できると考えられる．なお，表 6 では表 5 のような現象を確認できない．これは，表 1 と表 3 に示すように，データ内に含まれる IP アドレス数が異なっているためだと考えられる．以上から，相関解析の対象となる IP アドレス数に応じて，適切な距離で IP アドレスをクラスタリングすることで，異種ハニーポット間で重複して観測される可能性が高い IP アドレスを効率的に抽出できると考えられる．

表 5: 評価実験 1 結果

距離	総クラスタ数	混在クラスタ数	混在クラスタ内 総 IP アドレス数
0	24,295	1,183	4,058
1	21,480	1,722	6,075
2	18,963	2,122	7,951
3	17,172	2,294	9,411
4	15,732	2,388	10,567
5	14,625	2,445	11,509
6	13,761	2,466	12,245
7	13,088	2,456	12,783
8	12,451	2,438	13,187
9	11,880	2,440	13,747
10	11,429	2,430	14,177

表 6: 評価実験 2 結果

距離	総クラスタ数	混在クラスタ数	混在クラスタ内 総 IP アドレス数
0	15,619	9	35
1	13,073	10	39
2	11,119	11	44
3	9,667	16	61
4	8,608	17	68
5	7,823	20	78
6	7,179	23	85
7	6,671	25	94
8	6,217	27	100
9	5,846	27	100
10	5,527	31	112

4.3 考察

本調査により，攻撃者が使用する IP アドレス空間において，複数のハニーポットで収集した攻撃者 IP アドレスが密集する空間の存在が明らかになった．この現象は，特定の IP アドレス空間における，複数種類のマルウェア検体への多重感染や，攻撃者によるボットの有効活用に起因

して発生しているものと考えられる．データセットのIPアドレス数が示すように，各ハニーポットで収集できる攻撃者IPアドレス数には大きな偏りがある．具体的には，近年脅威が増大しているWebブラウザやWebアプリケーションの脆弱性を対象とした攻撃と比較して，Windows OSの脆弱性を対象とした従来の機械的な攻撃件数が非常に多い．このため，Windows OSの脆弱性への攻撃を送信するIPアドレスに対して提案の相関解析を適用することで，悪性Webサイトやマルウェアダウンロードサイトおよびマルウェア検体などの攻撃者情報を効率的かつ効果的に収集できる可能性が高いと考えられる．

5 おわりに

本稿では，様々な脆弱性を対象とした多種多様な攻撃への対策を効率的かつ効果的に実現するために，異なるハニーポットで収集したIPアドレスの相関性を解析した．

ヒルベルト曲線に基づき攻撃者が使用するIPアドレスを2次元グラフ上に配置し，マンハッタン距離に応じてクラスタリングすることで，異なるハニーポットで収集したIPアドレスが混在する可能性が高いIPアドレス空間を抽出できた．さらに，効率的な抽出に向けて，クラスタリングの際に適用すべき最適なマンハッタン距離が存在する可能性を特定できた．

本提案の相関解析などを利用し，Windows OSの脆弱性への攻撃の情報を用いて，Webブラウザや Webアプリケーションに対する攻撃への対策を加速させることで，多種多様な攻撃からユーザを保護可能なネットワークを構築できると考えられる．

謝辞

本研究は，NTT情報流通プラットフォーム研究所での実習において得られた成果である．

参考文献

[1] SANS, ”Sans institute infosec reading room,” http://www.sans.org/reading_room/whitepapers/malicious/bots-botnet-overview_1299

[2] Dionaea, http://dionaea.carnivore.it/

[3] 青木一史, 川古谷裕平, 秋山満昭, 岩村誠, 針生剛男, 伊藤光恭. ”能動的攻撃と受動的攻撃に関する調査および考察,” 情報処理学会論文誌, Vol.50, No.9, pp.2147-2162, Sep 2009.

[4] The Honeynet Project, ”Web Application Honeypot,” http://www.honeynet.org/gsoc/project8

[5] T. Yagi, N. Tanimoto, T. Hariu and M. Itoh, ”Design of Provider-Provisioned Website Protection Scheme against Malware Distribution,” IEICE TRANS.COMMUN., VOL.E93-B, NO5, pp1122-1130, May, 2010.

[6] M.Akiyama, K.Aoki, Y.Kawakoya, M.Iwamura and M.Itoh, ”Design and Implementation of High Interaction Client Honeypot for Drive-by-download Attacks,” IEICE TRANS.COMMUN., VOL.E93-B, NO5, pp1131-1139, May, 2010.

[7] 独立行政法人情報処理推進機構, ”Gumblar 攻撃に対する技術の現状と課題,” http://www.ipa.go.jp/security/fy22/reports/tech1-tg/a_07.html

[8] 青木一史, 川古谷裕平, 岩村誠, 伊藤光恭, ”半透過性仮想インターネットによるマルウェアの動的解析,” マルウェア対策研究人材育成ワークショップ2009, 2009年10月.

[9] B. Irwin and N. Pilkington, ”High Level Internet Scale Traffic Visualization Using Hilbert Curve Mapping,” VizSEC, pp147–158, 2007.

[10] 畑田充弘, 中津留勇, 秋山満昭, ”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.