ダークネットへ到達するパケットの初期ペイロードの自己組織化分類手法

鈴木 悠太 †　　中村 康弘 †

† 防衛大学校
239-0811 神奈川県横須賀市走水 1-10-20
em53034@nda.ac.jp, yas@nda.ac.jp

あらまし ダークネットに到達するコネクション要求に応答を返すことで初期ペイロードを取得することができる．この初期ペイロードを種類毎に分類する手法を提案する．取得した初期ペイロードの Fuzzy Hash 値を特徴量とし，自己組織化マップを用いて分類を行うことにより，シグネチャを用いず観測結果のみから大まかな分類を行うことができる．

A self-organized classification method of Darknet arrival packet payloads

Yuta Suzuki†　　Yasuhiro Nakamura†

†National Defence Academy
1-10-20 Hashirimizu, Yokosuka, Kanagawa 239-0811, JAPAN
em53034@nda.ac.jp, yas@nda.ac.jp

Abstract It is possible to obtain the initial payload by returning a response to the connection request to Darknet. In this study, classification of this payload, and the method will be proposed analyze of the intention of these communications. The Fuzzy Hash value of the acquired initial payload is a characteristic quantity and performing classification by using a Self-Organizing Map, it is possible to perform rough classification from the only observation without using a signature.

1　はじめに

ダークネットには常に多数のパケットが着信している．これは，マルウェアが感染拡大のためにランダムなアドレスにパケットを送信することなどが原因である．このため，ダークネットに到達するパケットを観測することで攻撃の兆候を発見することができる．ダークネットに到達するパケットは多種多様であり，全てを手動で解析するには多大な時間を要する．

この研究では，攻撃の多くが TCP を用いることに着目し，簡易なハニーポットを用いてダークネットを観測，TCP 接続後の初期ペイロードを取得する．そして，通信特徴およびペイロードパターンを取得することにより攻撃タイプを自動判別することを目的とする．このために，本稿ではダークネットに到達するコネクション要求に応答を返すことで取得した初期ペイロードを取得する．取得したペイロードの Fuzzy Hash 値を特徴量とし，自己組織化マップ (Self-Organizing Map: SOM) を用いて種類毎に分類する手法を提案する．

本稿の構成は以下の通りである．まず，第 2 章で関連研究，第 3 章で提案手法を説明する．その後，第 4 章で実ネットワーク上のダークネットに対して提案手法を適用した結果を示し，第

5 章でまとめと今後の課題について述べる．

2 関連研究

小出ら [1] は，OS の機能を使わずに独自のネットワークスタックを用いた通信を行うマルウェアやツールは TCP/IP ヘッダやアプリケーションプロトコルに固有の特徴を持つ場合があることに着目し，特定のヘッダフィールドに固有値が設定されている通信パケットを抽出することでダークネット上で観測される通信を分類する手法を提案している．この手法はシグネチャ候補からヘッダパターンを選択してシグネチャを作成しているが，ヘッダパターンがいくつかの値に集中しているとき複数の値をシグネチャパラメータ値として選択している．このときの判断を手動で行っているため，自動で分類できないことがある．

椛島ら [2] は，DoS/DDoS 攻撃を想定した不正アクセスを検知するために，パケット中の複数の情報の出現頻度を入力ベクトルとして，自己組織化マップ上でそれらの情報を統合してパケットを解析する手法を提案している．この手法は一定の個数のパケットをまとめて 1 つのデータとしており，リアルタイムでの解析が難しい．そのため，1 つのパケット毎に入力ベクトルを作成すれば，よりリアルタイムに近い分類ができると考える．

3 提案手法

本手法では，取得した初期段階のパケットの中からペイロードを含むパケットを選択し，ペイロードとヘッダ情報を抽出する．このとき，ペイロードについては Fuzzy Hash 値を算出した後に数値化する．数値化したペイロードとヘッダ情報を入力ベクトルとして自己組織化マップを作成し，初期ペイロードを種類毎に分類する．提案手法の流れは以下のとおりである（図 1）．

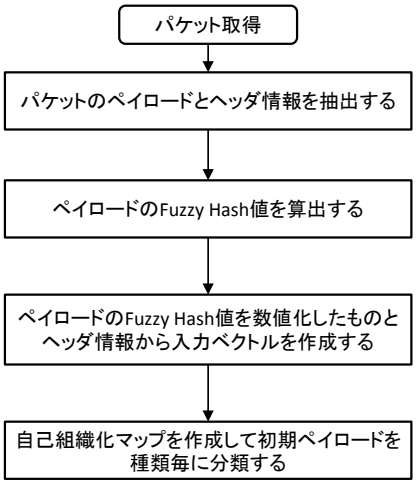

図 1: 提案手法の流れ

3.1 特徴量の抽出

パケットのペイロードおよびヘッダ情報を特徴量とする．パケットは同種の通信においてペイロードに共通のパターンを含んでいることから，Fuzzy Hash を適用して類似する特徴量を算出する．Fuzzy Hash には CTPH（Context Triggered Piecewise Hashes）[3] を算出する ssdeep[4] を使用する．ssdeep は変換する文字列をブロックサイズ毎に区切り，それぞれについてハッシュ値を生成し，以下の書式で結果を出力する．

blocksize:hash1:hash2

blocksize はハッシュを計算するバイト数，hash1 は 64 文字のハッシュ値，hash2 は 32 文字のハッシュ値である．同一内容のブロックであれば同一のハッシュ値を得ることができる．類似する部分データを含む場合，類似するハッシュ値を得ることができる．本手法においては，hash2 を使用する．

また，ヘッダ情報は送信元ポート番号，宛先ポート番号，パケットサイズを特徴量として使用する．

3.2　入力ベクトルの作成

自己組織化マップの入力ベクトルは数値である必要がある．送信元ポート番号，宛先ポート番号，パケットサイズについては，その値を入力ベクトルとする．ペイロードの Fuzzy Hash 値については，そのままでは入力ベクトルとして使用することができないため，以下の計算により数値化する．まず，hash2 の先頭から 1 文字ずつ 10 進数に変換する．1 文字目の値は，10 進数に変換した値にする．2 文字目以降の値は，10 進数に変換した値と 1 つ前に求めた値に 64 を掛けた値を足した値にする．

3.3　自己組織化マップの作成

4.3 で作成した入力ベクトルを基に自己組織化マップを作成して，初期ペイロードを種類毎に分類する．

自己組織化マップは，T. Kohonen によって開発された中間層のない 2 階層型の教師なし学習のアルゴリズムを用いるニューラルネットモデルである．自己組織化マップでは，多次元データを圧縮して，低次元のマップを描く．自己組織化マップは，入力層と出力層に分かれ，入力層で提示された情報は出力層の全てのユニットに伝えられる．そこで，入力情報と接続重みとが互いにどれだけ似ているかを出力層のユニット間で競争する．競争の結果，一番似ていたユニットを勝者と呼び，そのユニットの重みを調節して入力情報にさらに近づける．また，勝者のユニットに加え，その近隣のユニットの重みも，距離に応じて入力情報に近づけるように変化させる．この結果，近隣のユニットは類似した重みを持つ傾向が生じ，多次元のデータを予備知識なしでクラスタリングすることができる．

4　検証実験

実ネットワーク上のダークネットで取得したパケットを用いて提案手法を検証した．実験は 2014 年 11 月 25 日 00 時 00 分から同日 00 時 01 分の 1 分間に取得したパケットを用いて行った．自己組織化マップの作成には R 言語で実装されているパッケージ「kohonen」を使用する．

結果を図 2 に示す．図 2 内の値は，個別のパケットに付与した番号，破線は各ペイロードが共通のパターンを含んでいたユニット，太線は各ペイロードが共通のパターンを含んでいなかったユニットである．

図 2 内の座標（1,3）に属しているパケットは宛先ポート番号とパケットサイズは同一で，hash2 は類似している．これらは全て

”OHYO_DOUXIE_BOX”

という文字列を含んでおり，正しく分類することができた（表 1）．

一方，図 2 内の座標（5,5）に属しているパケットは宛先ポート番号とパケットサイズは類似しているが，hash2 は一部類似していない（表 2）．一部類似していない hash2 を持つパケットは他のパケットと共通のパターンを含んでおらず，正しく分類することができなかった．これは，正しく分類するための入力ベクトルを選択できておらず，送信元ポート番号，宛先ポート番号，パケットサイズ，数値化した hash2 以外の入力ベクトルを追加する必要があると考える．

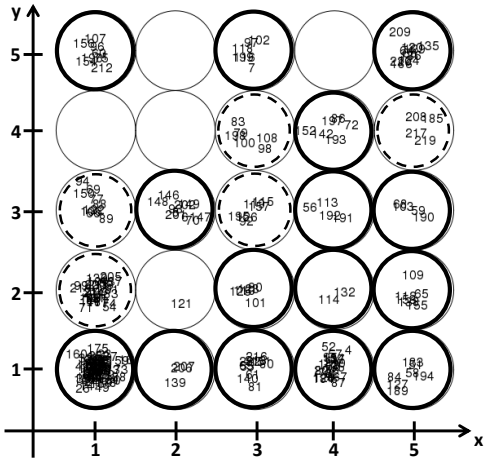

図 2: 実験結果

5　まとめと今後の課題

本稿では，初期ペイロードを種類毎に分類するために，ペイロードの Fuzzy Hash 値とヘッダ

表 1: 正しく分類できたユニットの例

No.	IP address	src port	dst port	bytes	hash	Base64
66	112.84.76.169	3622	80	122	5.06E+14	rDHXMWUJ
69	113.98.117.119	2942	80	122	3.24E+16	rDHXMWENi
77	116.113.183.48	1320	80	122	2.07E+18	rDHXMWHbfT
88	121.229.103.204	10977	80	122	3.24E+16	rDIbMC/3w
89	121.23.77.131	2093	80	122	5.43E+23	rDHXMWUx1qQhL
94	122.231.244.68	1306	80	122	8.49E+21	rDIbMC/OkGYG
106	14.112.209.112	1074	80	122	2.07E+18	rDHXMWDrug
122	180.139.99.77	2670	80	122	8.49E+21	rDHXMWUnKiqc
150	218.80.207.241	2962	80	122	5.06E+14	rDHXMWNy

表 2: 正しく分類できなかったユニットの例

No.	IP address	src port	dst port	bytes	hash	Base64
53	108.240.198.68	55420	51183	122	3.80E+23	OXa9RL85EIqGn
64	112.117.165.26	4863	51443	122	2.43E+25	OXa9RL9VjOXOQb
120	177.98.63.160	3745	51183	134	1.56E+27	OXa9RLVrW5EewKn
124	182.141.163.139	40027	51443	122	9.95E+28	OXa9RL9VjOXOL4Zs
129	185.21.216.135	41959	51183	122	1.56E+27	OXa9RL9VjOXOL4Zs
135	192.95.6.8	47096	51183	172	6.03E+39	EvEUKNISwKzmnXmV5UG1Nn
136	192.95.6.8	48105	51183	134	2.43E+25	OXa9RLVrW5EIBH
186	46.107.226.170	57423	51183	134	3.80E+23	OXa9RLVrW5EIN
209	67.185.195.49	50386	51183	91	5.68E+12	Qkh64ce
220	90.191.126.83	62992	51183	211	3.52E+43	d8XWvnGrVVzsns/Q9okDhy10

情報を入力ベクトルとして自己組織化マップを作成しパケットを分類する手法を提案した．本手法を実ネットワーク上のダークネットに適用した結果，宛先ポート番号，パケットサイズ，Fuzzy Hash 値が同一または類似していれば初期ペイロードを種類毎に分類できることがわかった．

一方，本稿で選択した入力ベクトルが類似または同一であっても正しく分類できないパケットが存在することもわかった．今後は入力ベクトルのとり方と Fuzzy Hash の数値化の方法について検討する必要があると考える．

参考文献

[1] 小出駿, 鈴木将吾, 牧田大佑, 村上洸介, 笠間貴弘, 島村隼平, 衛藤将史, 井上大介, 吉岡克成, 松本勉, 通信プロトコルのヘッダの特徴に基づく不正通信の検知・分類手法, コンピュータセキュリティシンポジウム 2014 論文集, Vol.2014, No.2, pp.48–55（2014）

[2] 椛島健, 堂園浩, 自己組織化マップによる統計的情報を用いたネットワーク IP パケット解析手法, 信学技報, Vol.109, No.252, pp.7-12（2009）

[3] J. Kornblum, Identifying almost identical files using context triggered piecewise hashing, in Digital Investigation, Vol.3, Supplement, pp.91-97（2006）

[4] J. Kornblum, Fuzzy Hashing and ssdeep, Sourceforge, 2012, http://ssdeep.sourceforge.net/