4X-04

簡易ハニーポットによるダークネット観測とペイロード分類手法の提案

鈴木　悠太†　　後藤　洋一†　　中村　康弘†
†防衛大学校　情報工学科

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．

ダークネットを観測する方法にはパッシブ型のブラックホールモニタリングとアクティブ型のハニーポットモニタリングがある．前者は攻撃状態に影響を与えることなく観測が可能である反面，セッションを開始しないためペイロードを取得することができず，後者はペイロード取得が可能である．

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捕獲し，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，本稿ではまず，ペイロードパターンの取得アルゴリズムを検討した．有意な文字列パターンを抽出する提案手法について述べ，実データを用いて実験した結果を示す．

2　関連研究

角田ら[1]はワームによって生じるフローのペイロード内容に類似性があることに着目し，複数のフローペイロードから共通して得られるトークンを種別分類用のシグネチャとして用いることでワームを自動分類する手法を提案した．和泉ら[2]はパケットのペイロードを8ビット単位のコードに分割し，それらの出現頻度をクラスタリングして，ヒストグラムとヒストグラム間の距離から通信内容の類似性について調査した．これらの手法はどちらもワームなどの不特定多数のホストに感染を試みる拡散型不正アクセスを対象としており，それ以外の攻撃に関しては考慮されていない．

本稿では，ペイロードから特定の長さ以上の任意の文字列を抽出し，深さ優先探索アルゴリズム（バックトラック法）を基に宇野ら[3]が開発したLCMにより選択した頻出する文字列を特徴量として自動抽出する手法を提案する．

なお，深さ優先探索アルゴリズムとは，ある解を求めるとき可能性のある手順を順に試し，その手順では解が求められないと判明した時点で1つ前の状態に戻って別の手順を試すアルゴリズムである．

3　提案手法

提案手法の流れを図1に示す．

まず，簡易なハニーポットを用いてダークネットに到達する送受信パケットを捕獲する．IPアドレス毎のペイロードのバイナリ部分について特定の長さ以上の文字列を抽出する．このとき，抽出した文字列の集合をアイテム集合とする．

ペイロードから文字列を抽出する際，文字列の長さを短くしすぎると1つの文字列から意図しない複数の単語を抽出してしまう．例えば,「NT LM 0.12」とい

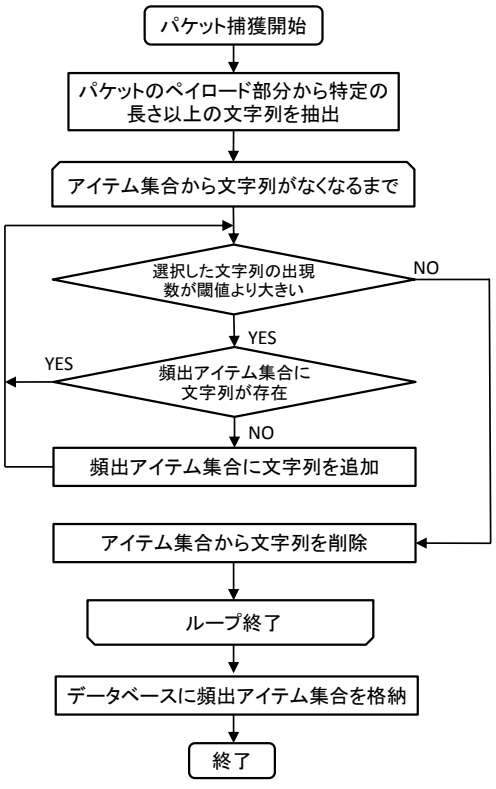

図1: フローチャート

A playload classification method using honeypot observation on Darknet
†YUTA SUZUKI　†YOICHI GOTO　†YASUHIRO NAKAMURA
†Department of Computer Science, National Defense Academy

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.

う文字列を抽出しようとしても，抽出する文字列の長さを指定しなければ「NT」「LM」「0.12」という3つの文字列に分割されて抽出する．これを回避するためにはある一定以上の長さを指定する必要があり，本稿においては文字列の長さがN以上のものを抽出することで意図しない文字列の抽出を防ぐこととする．

次にアイテム集合からLCMにより頻出する文字列を選択する．その文字列の出現数が閾値よりも大きければ再帰呼出しを行い，頻出アイテム集合に追加，その文字列の出現数が閾値よりも小さければ再帰呼出しを行わず，その文字列をアイテムから削除する．これらの処理を抽出した全ての文字列に対して実施する．

その後，頻出した文字列，パケットの取得年月日，出現数を記録するためにデータベースに格納する．新たにパケットを取得したときは上記の処理を実施した後，結果を同様にデータベースに格納する．

4 実験結果

第3項に述べた提案手法の適用結果の一部を図2に示す．図2は2014年11月25日に防衛大学校のダークネットより取得した46433個のパケットから，N=10としたときに抽出された842713個の文字列を基にCSVファイルで出力したものである．頻出した文字列，観測した日時，出現数を取得することで，一定期間内により多く出現した文字列やある文字列の出現状況を確認することができる．

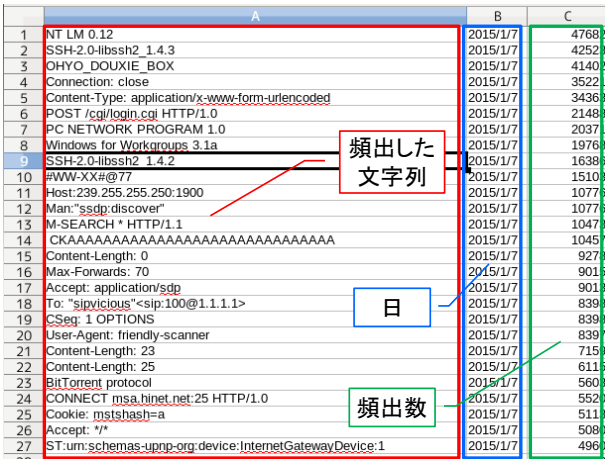

図2: 出力結果

5 まとめと今後の課題

本稿では，簡易なハニーポットを用いてダークネットを観測，初期ペイロードを捕獲し，この際の通信特徴およびペイロードパターンを自動取得することにより攻撃タイプを判別することを目的に，有意な文字列を抽出する手法を提案した．今後，この文字列パターンのマッチング処理を実装して攻撃タイプの自動判別を行う．また，新たな攻撃を検知する上で過去に出現したことのある文字列よりも初めて出現した文字列の方がより重要であると考えられ，初めて出現した文字列に対して強調する方法について検討することが今後の課題である．

参考文献

[1] 角田裕，和泉勇治，関部然，根元義章．フローペイロードの共通部に着目したワームフローの自動分類に関する検討．電子情報通信学会技術研究報告 Vol.106 No.153 Page.17-22, 2006.7

[2] 和泉勇治，辻雅史，根元義章．パケットペイロードのクラスタリングによる拡散型不正アクセス検知方式に関する一考察．電子情報通信学会技術研究報告 Vol.105 No.280 Page.19-24, 2005.9

[3] Takeaki Uno, Tatsuya Asai, Yuzo Uchida, and Hiroki Arimura. An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases. Lect Notes Comput Sci Vol.4835 Page.402-414, 2007

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.