2Y-2

マルウェアが生成する不正パケットの分析によるシグネチャファイル自動生成機構の設計と実装

重松邦彦† 水谷正慶† 武田圭史‡ 村井純‡

†慶應義塾大学政策・メディア研究科 ‡慶應義塾大学環境情報学部

1. はじめに

本研究では，マルウェアの通信に共通して見られるネットワークトラフィックの特徴をメタデータの形式で集積し，既存のマルウェアの通信の検知に用いるとともに類似する亜種の検知に用いる方式について検討した．また同時にマルウェアに感染していない平常時の通信データとの比較検証によりフォールポジティブの低減をはかった．

2. マルウェアシグネチャの課題

従来のウイルス対策ソフトウェアでは主に3つ存在する.マルウェア実行ファイル自体に出現する特定のデータパターンを特定するパターンマッチングによる検知方式や，検査対象となるプログラムの動作をエミュレートして，その振る舞いの特徴が「ルール」と合致するかによって検知をするヒューリスティック検知方式，実行中のプログラムの挙動を調べて，検出するビヘイビア検知方式である.マルウェア本体に出現するデータ列をシグネチャとして用いるパターンマッチング検知方式は，実行効率と誤検知を低く検知するアプローチとしてもっとも広く利用されている．マルウェアシグネチャは発見されたマルウェアの分析に基づき作成されるため，マルウェアの出現からシグネチャが作成され検知に用いられるには一定の時間がかかる.また，マルウェア自体のファイルが少しでも変更されると検出されない可能性が高くなる.このためマルウェアを検知するシグネチャの効率的な生成は，指数関数的に増加するマルウェアを検知する上で有用である.この目的を達成するためのシグネチャ自動生成のアプローチが提案されている．

3. 関連研究

マルウェアのシグネチャの自動生成に関するアプローチは大きく以下の3種類に分類できるものと考える．

1 種類目は，パケットのヘッダ情報やペイロードの頻度，IP アドレスの散らばりを統計的に分析する手法である． EalryBird[1]では，マルウェアの不変な文字列からワームのシグネチャを自動生成するために，パケットのコンテントの頻度や散らばりを用いる．

2 種類目は，ペイロードの文字列に着目し，データマイニングを使って類似度からマルウェアの不正パケットを検出する手法である．Honeycomb[2]は，本質的に疑わしいトラフィックを集めるためにハニーポットを使い，LCS（Longest Common Subsequence）アルゴリズムを適用することによってシグネチャを生成する．また，Anomalous[3]は文字傾向の出現頻度による IDS システムで，False Positive は 1%以下という結果を出している．

3 種類目は，マルウェアの特性を抽象化し，他への感染癖や周辺アドレスへの攻撃癖などの特徴を抽象化する手法である．ヒューリスティックビヘイビア型の検知手法を用いた AntiBot[4]などがある．

4. 解決方法

4.1. 要件定義

増え続けるマルウェアに対してシグネチャの自動生成による解決策は通信データをマルウェア特有の通信と判断するためにメタデータ化しシグネチャ生成を行う必要がある．まず，それらの作業を迅速にルール化するためには，以下の５つの条件を満たす必要がある．

(1)処理の自動化

マルウェアに感染してないパケットとマルウェア特有のパケットを切り分ける必要がある．パケットの情報を手動で判断するためには相応の時間がかかり現実的ではない．自動化することで，ある程度マルウェアの通信と判断できるアルゴリズムが必要となる．

(2)処理結果の保存

処理結果をファイルに保存し，後で参照できることが望ましい．できるだけ汎用性のある形で保存し，後で異なる条件で通信データを特定できる必要がある．

(3)通信データの特徴を捉えた結果

処理前には得られなかった通信データの特徴が，処理の結果から得られていることが望ましい．

(4)検索可能な結果

処理の結果得られた通信データの情報群は研究者が一度に把握できることが望ましい．任意の特徴を持つ通信データとそうでない通信データを選別できることで，意味付けを修正することも可能となる．

(5)シグネチャ化しやすいフォーマット形式

パケットのヘッダ情報の特徴とペイロードの文字列を比較する．

4.2. 解決策

感染活動やポートスキャンなどの際に発生する大量の通信を適切に扱うため，自動的に通信を集約化する機能を実装し，集約した情報を元にシグネチャファイルを生成する.通信データをまずメタデータ化して，そのデータをインプットデータとして，シグネチャファイルの生

「Automated Generation of Signatures for Malware Detection by Analyzing Network Traffic」

「Kunihiko Shigematsu† , Masayoshi Mizutani† , Keiji Takeda‡ , Jun Murai‡

‡ Faculty of Environment and Information Studies, Keio University 5322 Endo Fujisawa Kanagawa , 252-0882 , Japan

†Graduate School of Media and Governance, Keio University, 5322 Endo Fujisawa Kanagawa, 252-0882, Japan

成を行う.

4.3. アプローチ

以上の５つの条件を満たす解決策として，通信データから抽出したメタデータをデータベースに格納する方法を提案する．データベースの設計によって(3)を満たす．データベースは情報を保存し，それらの情報から任意の情報を検索できることから(2)と(4)を満たすことができる．さらに，パケットの解析プログラミングをすることで，(1)を満たすことができる．抽出した情報からシグネチャ形式のフォーマットに変形するために(5)を実現する.

5. メタデータ化

マルウェアの通信と思われるデータを抽出し，大量の処理の効率化や自動化を実施するためにデータベースを用いる.研究室でハニーポットを構築し，マルウェアを実行し，通信データを一定期間取得する.この一連の流れを自動化し，メタデータ化すればマルウェアのシグネチャに活用できることが期待される.

本稿ではマルウェアの通信データ毎の特徴，また同じフロー内で発生した同じ TCP/IP 通信ポート番号の組み合わせによる通信データ毎の特徴が含まれることが望ましい．データベースは，ファイルテーブル，フローテーブル，セッションテーブルの 3 つのテーブルから構成する.

ファイルテーブルには通信データの特徴を保存する．フローテーブルでは，同じ IP アドレスの組み合わせの IP 通信フロー毎にデータが格納されている．フローテーブルに通信先ホストの IP アドレスや送受信パケット数等の特徴を格納することで通信相手を調べることができる.

セッションテーブルには，同じフロー内で発生した複数の通信について，監視対象のホストが使用した TCP/UDP 通信ポートと通信先のホストが使用した通信ポートの組み合わせを１つのセッションとみなし，セッション毎の特徴を格納している.セッションテーブルを調べることで，使用した通信ポート番号に関わらず HTTP や IRC などの特定の通信が存在したかどうかを判別する.

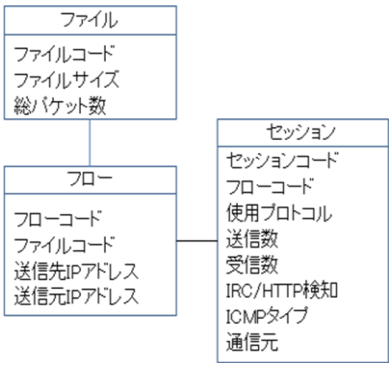

図１　データベースの構成

6. シグネチャ化

シグネチャは snort のルールファイルを適用する.オープンソースの侵入検知システムで，シグネチャをシンプルなフォーマットに従って作成することができる.セッション情報からシグネチャに必要な情報を抽出し，マルウェアの通信と断定できれば，シグネチャを生成することができる.

7. 実験結果

本研究で利用した通信データは，研究室のハニーポットで収集した 974848 バイト分のデータを利用し，セッションテーブルの情報からシグネチャを作成した．さらにセッションテーブルで通信先ホストのポート番号 139 番や 445 のパケットを検索したところ，該当したセッションは 80 あった．このセッションデータを元に snort のシグネチャを生成し該当するパケットの抽出の検出に成功した．外部への TCP ポート 139 と 445 がフィルタリングしていたため，全ての通信先ホストから応答が観測されなかった.この一連の特徴的な通信から，このデータセットが他のマルウェアと区別できる特徴的なマルウェアの感染活動の通信を含んでいると推測できる.

8. 今後の課題

良いシグネチャの条件として，脆弱性が悪用される条件を基に検知・防御するタイプのシグネチャが望ましい.多数の亜種が存在するマルウェアも１つのシグネチャで防御可能となる.今回は，マルウェアによる通信データからシグネチャを生成する試みを行った．マルウェアに感染していない一般環境でもマルウェアを正しく判定する必要がある．先行研究にあるようにペイロードの文字列とパケットのヘッダ情報を組み合わせた高精度のシグネチャを生成する必要がある.

9. 参考文献

[1] Singh, S., Eitan, C., Varghese, G., Savage, S.: Automated worm fingerprinting. In: 6th Symposium on Operating Systems Design and Implementation (OSDI), December (2004)
[2] Kreibich, C., Crowcroft, J.: Honeycomb: creating intrusion detection signatures using honeypots. SIGCOMM Comput. Commun. Rev. 34(1), 51- 56 (2004)
[3] Christodorescu, M., Jha, S., Seshia, S., Song, D., Bryant, R.: Semantics-aware malware detection. In: Proceedings of the IEEE Symposium on Security and Privacy(2005)
[4]Symantec.Norton.antibot. http://www.symantec.com/norton/antibot, 2008.