URLとDNSの異常特性に基づくC&Cトラフィック異常検出システム

YiHui Yan[1]　小出 洋[2]　櫻井 幸一[1]

概要：サーバー攻撃が増え続けるにつれて、ネットワークのセキュリティは重要な課題になっている。MIRAI および大多数の悪意のあるソフトウェアによる攻撃では C&C（Command and Control）機構を持っている。ボットネットを検出する有用な方法はいままでに多く存在するが，それらのほとんどは特定の攻撃動作に基づいている．一方，新しいタイプの悪意のある攻撃を検出するシステムと C&C トラフィックの検出方法に注目した研究はほとんど存在しない．本発表では、ボットネットの C&C 構造、プロトコルと通信隠蔽の方式を分析することにより，技術的な欠陥について議論し、C&C 間の通信に基づく異常検出システムを提案する。

キーワード：情報セキュリティ, C&C トラフィック, 異常検知, ボットネット, MIRAI

C&C Traffic Anomaly Detection system base on URL and DNS characteristics

YiHui Yan[1]　Hiroshi Koide[2]　Kouichi Sakurai[1]

Abstract: As server attacks continue to increase, network security has become an important issue. MIRAI and most malicious software attacks have a C&C (command and control) mechanism. There are many useful ways to detect botnets, but most are based on specific attack behavior. On the other hand, few studies have focused on new types of malicious attack detection systems and C&C traffic detection methods. In this presentation, we discuss technical flaws by analyzing the botnet C&C structure, protocols and communication concealment methods, and propose an anomaly detection system based on communication between C&C.

Keywords: Cybersecurity, C&C traffic, Anomaly detection, Botnet, MIRAI

1. はじめに

1.1 研究背景と目的

急速に発展した IoT(Internet of Things) は一連のセキュリティ問題を引き起こしている。特に、2016 年に、安全性の低い IoT デバイスを狙い、「史上最大級の DDoS 攻撃」[1] を惹き起こした Mirai[2] マルウェアは、我々に早急な IoT セキュリティ対策を呼びかけている。Mirai マルウェアの伝播方式をシミュレーションすることにより、伝播の原因を解析し、攻撃をどのように効果的に検出するかが課題となっている。業界評価と予測によると、図 1 に示すように、2020 年までに約 310 億個の IoT デバイスが存在し、同時に世界の IoT のグローバルビジネス価値を 8.9 兆ドルに押し上げます。Mirai は、ホームルーターや安全でない Telnet サービスを開く IP カメラなどの IOT デバイスを対象とする悪意のソフトウェアです。インターネット全体でマルウェアを効果的に自己増殖させ、DDOS 攻撃を開始して、侵害された接続デバイスを介してターゲットをオフラインにします。私たちが Mirai を選んだのは、他の IOT マ

[1] 九州大学　大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University
[2] 九州大学情報基盤研究開発センター
Information Infrastructure Research and Development Center, Kyushu University

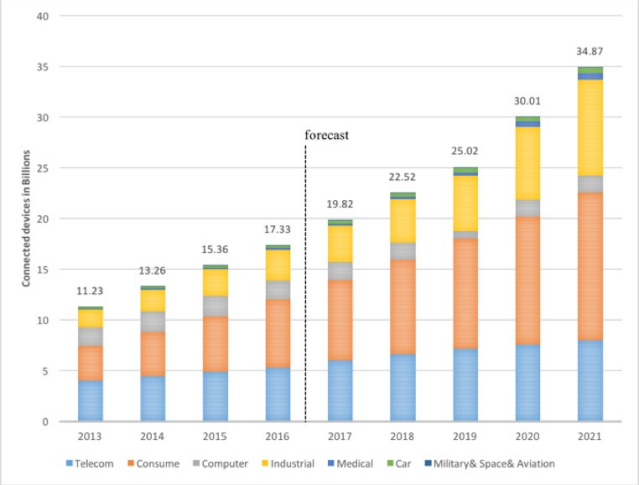

図 1 世界の IoT デバイスの数 2013-2021

ルウェアよりも大きな能力を持っていて大規模な破壊を引き起こす能力を持っていたことと、Mirai ベースの亜種は引き続き出現し、最大の IoT マルウェアファミリの 1 つになります。

DDOS 攻撃は、複数のソースからの偽のソースから正当なサーバーをフラッディングする悪意の試みです。悪意があるタスクを実行するために相互に連携する参加者を bot またはボンネットと呼ばれ、攻撃者が bot 管理を行うサーバを C&C(command and control server) と呼ばれます。

1.2 既存研究

1.2.1 Mirai の伝播と攻撃

このような攻撃を防ぐためには、システムにマルウェアを実際に実装して、詳細な情報と実用的なソリューションを取得する必要があります。具体的には、シミュレーションには、マルウェアがネットワークシステムに侵入する方法と、悪意目的のためにどのように実装されるかが含まれます。Mirai マルウェアの動作は、前輩の研究から知られていました。具体的は次のとおりです。

Mirai マルウェアは、最初に TCP ポート 23（○ 1）でランダムな IPv4 スキャンを実行することを開始します。これは Telnet プロトコルのデフォルトです。ポート 23 以外、一部の Phillips 製品がデフォルトの Telnet として使用する 2323 への少数のスキャンもあります。

利用可能なデバイスが見つかると、マルウェアは一般的に使用される 62 の工場出荷時のデフォルト認証情報でログインを試行します。暴力ログインが成功すると、感染したデバイスの情報 (例えばユーザー名、パスワード、IP アドレス) が Report Server(○ 2 と○ 3) を介して Loader に送信される。ロードプログラムは、Loader は悪意のプログラムを新しく開発されたデバイスに注入し、それらをボットに変換します。

すべてのボットは、C&C サーバーを介して攻撃者の制御下にあります（○ 5)。攻撃を開始するコマンド（○ 6）を受信すると、利用可能なすべてのゾンビマシンは、それに応じてターゲットをあふれさせます（○ 7)。Mirai ボットネットが DDoS-for-hire サービス（○ 9）を提供している。誰でもターゲットへの DDoS 攻撃ためにお金を払うことができます。以上の過程から、MIRAI および大多数の

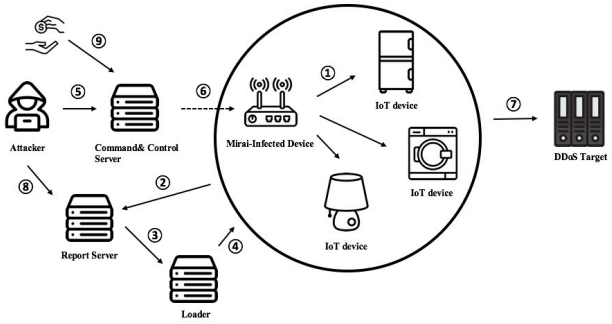

図 2 Mirai の伝播と攻撃の操作

悪意のあるソフトウェアによる攻撃では C&C 機構を持っているということが分かる。

1.2.2 既存の検出方法

現在、ボットネットの C&C 検出は、主にハニーポット、ネットワークトラフィックの監視、およびホストログ分析に基づいている。ハニーポットテクノロジーは、攻撃を欺いて攻撃動作の詳細を発見することにより、悪意のあるトラフィックを分析する。ハニーポット自体はデータを生成できない、C&C トラフィックの特性を効率的に収集できるが、ゾンビホストを効果的に検出できない。ネットワークトラフィックの監視と分析は、署名と異常検知の 2 つの主な方法でボットネットを効果的に検出できる。署名の動作原理は，トラフィックの特性を抽出することで既知の特性とマッチングすることである。そのため、既知のトラフィック特性を持つボットネットに対して良好な検出結果が得られる，ただし、トラフィックの抽出と分析は、悪意のある動作が発生した後ので、特性を継続的に修復する必要がある、常に新型マルウェアが発生する場合、メンテナンスの時効性が悪いだけでなく、新型マルウェアを検出することもできないため、署名はより安定したネットワーク環境で特定のタイプのマルウェア攻撃を監視するのに適している。異常検知技術は未知の悪意ある攻撃行為を検知でき、主に大トラフィック、高遅延や非一般なポートでトラフィックが発生するなどの異常特徴から悪意のある行動を発見できる。

ネットワークのトラフィック監視の位置によって，異常検出はホストベースの検出技術とネットワークベースの検出技術に分類される。ホストベースの異常検出技術は、単一ホスト上のネットワークトラフィックを分析することでゾンビネットワークを識別し、ネットワーク端末に配備す

る。Balram[3] たちは，ホストの疑わしいトラヒックを解析するゾンビ C&C トラヒック検出機構を提案し，ユーザホストが発生するトラヒック中の正常なトラヒックをフィルタリングし，残りのトラヒックを動的に行動パターン解析する。Taft[4] たちはゾンビのホストは定期的にサーバーに接続してコマンドを受け入れると仮定する。ホストの連続接続先のセーフリストを確立することにより、正当な連続接続がセーフリストに追加され、そうでない場合はブロックされます。この方法はセキュリティリストが正確に獲得でき、ネットワークが常に更新する必要がない場合に適用される。ネットワークベースの異常検出技術は、同一のネットワークに関連する複数のホストの類似行為によりゾンビネットワークを検出するものである。例えば、Zhang[5] たちはゾンビネットワーク監視システム BotMiner を提案した、複数のゾンビの一貫性行動をクラスタリングする、まず既知のセキュリティ特性に基づいて一部のトラヒックをフィルタリングし、C&C 通信トラフィックにおける類似した活動をクラスタリングする。類似したトラヒックがあると、Snort を用いて異常検出を行い，活動タイプによるクラスタリングが行われる、2 回のクラスタリングを検査し、ホストの類似性を相関させてボットネットを識別する。

検知技術の向上に伴い、ゾンビネットワーク技術もアップデートされている。ターゲットネットワークの既存のプロトコルを使用して C&C 通信プロトコルを構築するか、暗号化通信を使用してトラフィック特性を隠すことにより、元の効果的な検出方法のパフォーマンスが低下するか、無効にさえなります。しかし、ゾンビネットワークの攻撃技術がどのように更新されても、C&C トラフィックによる悪意の検知は、ゾンビネットワークの変化に対応する有効な方法となっている。

2. C&C 分析

2.1 C&C 構造

C&C サーバとゾンビの間の C&C トポロジー型により、中心構造、分散構造、ハイブリッド構造、ランダム構造の 4 つのタイプに分類する。

中心構造は最も簡単なゾンビネットワーク C&C 構造です、中央 C&C サーバはゾンビごとに直接接続されており、通信プロトコルに特別な要求がなく設定が容易で、コマンドを直接ゾンビに転送することができ、通信遅延性が低い。また，C&C サーバの一点失効によりロバスト性能が低いという欠点があり、C&C サーバを発見すればゾンビネットワーク全体を破壊することができるため、ゾンビネットワークは予備 C&C サーバを使用することにより、中心構造 C&C ゾンビネットワークのロバスト性をある程度向上させている。 Fast Flux ネットワークのように、ドメインシステム (DNS) は、通常ドメイン名と IP アドレスを関連付けるために使用され、記憶を容易にするために、DNS は、複数のドメイン名と 1 つの IP アドレスを関連付けることができ [6]、1 つのホームページに複数のホストを開くことができる。この通信機構は Fast Flux ネットワークに悪用され、複数のゾンビが 1 つの DNS ホストレコードを登録し、ログを解析して 1 つのゾンビ IP アドレスを返すと、これらのゾンビは C&C サーバに直接接続するのではなく、中間ホストとして C&C サーバにデータを転送し、プロキシとして機能する。この方法は，複数のサーバ間を高速に循環させることができ、ゾンビネットワークを検出することの難しさを増している。現在、HTTP によるドメイン生成技術を用いた中心 C&C 構造のゾンビネットワークが一般的である。分散構造は、図 4 に示すように、ネットワークにはゾンビのみ、C&C サーバはなく、どのゾンビも潜在的な C&C サーバであり、分散式は p2p プロトコルを採用しており、各ゾンビは少なくとも他のゾンビと接続しており、完全メッシュゾンビネットワークはロバストであり、いかなる数のゾンビを削除しても通信に影響しないが、数が増加するにつれて、ゾンビ直接の接続も倍に増加し、ほとんどの分散構造は完全なメッシュではなく、一般的に大型ゾンビネットワークにも使われていない。ハイブリッ

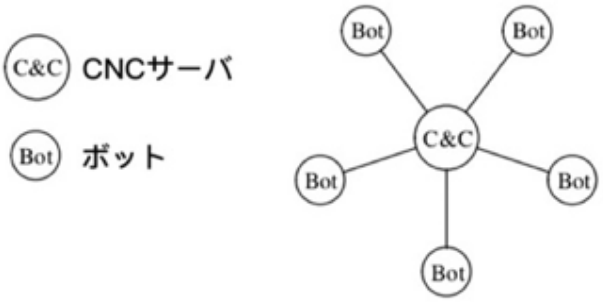

図 3 中心構造 C&C

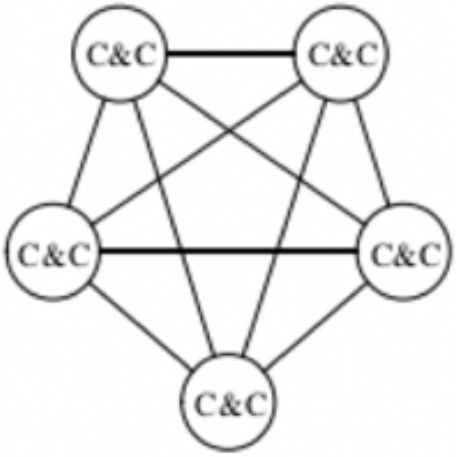

図 4 分散構造 C&C

ド構造は、中心構造と分散構造の両方の利点に合わせて、ゾンビが C&C サーバに直接接続するのではなく、p2p トポロジからなるエージェント層を経由して C&C サーバに接続し、視認性を低下させるためにワークプレースを追加することでタスクを実行し、ゾンビがエージェント層であるかワーク層であるかはゾンビネットワークの接続属性に依存し、層数を増加させることで C&C サーバの耐検知能力を向上させることができるが、メッセージが遅延することがある。もう 1 つのハイブリッド方式は一部は中心構造を用い、もう一部はによる p2p 分散構造を用いている。図 5 のように。ランダム構造、ゾンビは自発的に C&C サー

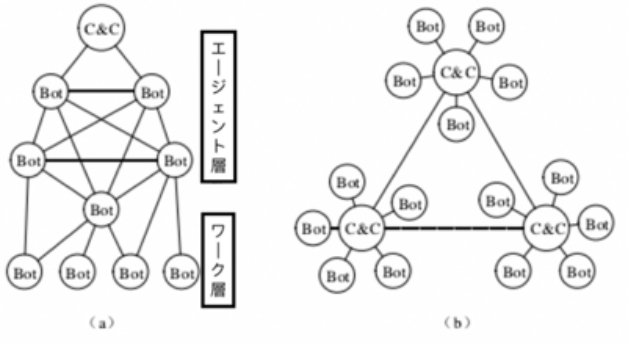

図 5 ハイブリッド構造 C&C

バーあるいは他のゾンビに連絡しないで、主人が自発的に連絡するのを待って、主人がメッセージを発信しようとする時、まずメッセージを暗号化して、すぐにネットワークをスキャンして、別のゾンビがメッセージを伝えることを発見して、ランダム構造の設計は簡単で、単一のゾンビは検出されてもネットワーク全体に関与しないが、遅延性が高く、効果的な伝達を保証することができない。ゾンビのランダムスキャン行為が発見された可能性が高い。

2.2 C&C プロトコル

現在は主に 2 種類のゾンビネットワーク C&C プロトコルがあり、1 つは既存のプロトコルであり、もう 1 つは特別に策定された新しいプロトコルである。

（1）IRC プロトコル。RC プロトコルはテキスト転送をサポートし、ゾンビネットワークのバイナリファイルの配布、構成および更新ファイルの配布に利用可能であり、IRC には中心構造の利点があるが、ファイアウォールなどのセキュリティ施設を介して有効に阻止でき、このプロトコルは一般的には使用わない、センター構造 C&C ゾンビネットワークはプロトコルとして HTTP を広く利用している。

（2）HTTP。要求応答プロトコルであり、クライアントは 1 つの要求を送信し、対応するサーバは 1 つの応答を返信する。各ゾンビは特定の命令に従う必要があり、グループ通信ができないため、遅延が高い。HTTP はトポロジーのために設計されている、ゾンビがメッセージを他のゾンビに転送し、これらのゾンビがまたこれらのメッセージを転送することによるループであり、どのようにゾンビネットワーク上で確実に通信を行うか、他のゾンビを探すことは攻撃者が検討すべき課題である。通常アプリケーション層と実際には命令の間に p2p 層を加えてメッセージの氾濫を防ぎ、ルーティング層を増やして信頼性の問題を解決することも可能であり、攻撃者が HTTP プロトコルを好んで利用するもう 1 つの理由は HTTP によるゾンビネットワーク C&C 通信により正常なトラフィックパターンに隠蔽できることである。

（3）p2p。分散構造で使用する、p2p ネットワークを個別にセットアップしたり、ネットワークの一部として悪意のある活動を隠すことができる。

（4）新しいプロトコル。いくつかのゾンビネットワークはデータグラムプロトコル UDP あるいは転送制御プロトコル TCP を利用して新しいプロトコルを設計して C&C 通信を行う、UDP はパケットサービスを提供し、単一のパケットを確実に転送することができない、TCP は信頼できるデータ転送プロトコルである。また、制御メッセージプロトコル ICMP に基づく C&C チャネルが出現している。

2.3 C&C 通信隠蔽と混淆の方式

隠蔽の方式

（1）チャンネルを非表示にします。通信用途でない手段を用いて通信を行うことにより、埋込みを実現する。例えば、TCP や IP で使用されていないヘッダビットは、情報を送信するために使用されてもよい [7]。

（2）通信を暗号化する。暗号化は、ネットワークパケット負荷分析に基づく検出方法を無効化することができ、暗号化アルゴリズムは、ストリームに基づいてブロックに基づいてもよいし、ストリームに基づいて単一のバイトを暗号化し、ブロックに基づいて固定サイズを暗号化することもできる。暗号化アルゴリズムは対称と非対称に分けられる. 対称暗号化は同じ鍵を用いて暗号化と復号を行い、非対称には 2 つの互いに異なる鍵を用いて暗号化と復号を行う。

（3）多重プロトコル。ゾンビネットワークは単一の C&C プロトコルの代わりに異なるプロトコルを用いており、この方法は圧縮、暗号化などの他の方式と結合することができ、多重プロトコル選択があり、ゾンビネットワークはターゲットネットワークの主要プロトコルを利用して多重プロトコル C&C 通信を行うことができる。

混淆の方式

（1）圧縮。アルゴリズムによって入力が減少し、暗号化と異なり、圧縮に鍵は不要であり、アルゴリズムが分かればデータを伸張することができるため、混淆の方式である。

（2）ステガノグラフィ。情報を通信に使用しないキャリアーに隠蔽し、隠蔽チャネル技術で採用されているキャリ

アーとは異なり、ステガノグラフィは画像やドキュメントなどのファイルをベクターとしている。以上の分析から、HTTP プロトコルは IRC や p2p プロトコルよりも広く利用されている。最新のゾンビネットワークは、動作が柔軟で、動的変化やリソース最小化利用などの特徴がある、攻撃者は大量の C&C 暗号化通信、反蜜缶技術などを利用して、ゾンビネットワークの検出の難しさを増している。難点は，ゾンビネットワークが既存のプロトコルを用いて C&C チャネルを構築することであり、通信によって発生するトラフィックは通常のトラヒックに似ており、次にゾンビによるトラフィックが小さいこと、さらにゾンビネットワークが暗号化方式を用いて C&C 通信を行う可能性があることである。

3. 異常検出システム

3.1 異常検出方法

HTTP ベースのゾンビネットワークは設計や管理上の便宜のためにほとんど非暗号化された C&C 通信形式をとっている。彼らはゾンビネットワークの C&C トラフィックの特性を隠し、識別の難しさを増加させた。それにもかかわらず、ボットネット C&C 通信には、模倣するのが難しい一般的なユーザー通信機能がまだある。これらの特性を使ってゾンビネットワークを検出することができる、ゾンビネットワークは C&C 通信において

HTTP リクエスト URL パケットと DNS 応答パケットが正常ユーザと異なるため，これらの異常を分析することでゾンビネットワークの C&C トラヒックを効率的に検出できることが分かった。

HTTP リクエスト URL パケットの異常特性

(1) リクエストの URL 数量が違う、正常な場合、同一サイト上の異なるページにアクセスすると、複数の異なる URL に対応して一定の数値範囲で URL 総数が高くなるのに対し、ソフトウェアエージェントが自動的に生成する HTTP ストリームが持つ URL 数は一般に低い。

(2) リクエスト URL の頻度が違う。通常、ソフトウェアエージェントは 1 つの URL を繰り返したり、異なる URL を生成するたびに。人手によるブラウジング活動のシミュレーションに成功することは困難でる、頻度の平均値や標準偏差をデータとして収集し、異常検出を行うことができる。

(3) リクエスト URL の長さは違う、正常な場合は URL の長さが異なり、ソフトウェアエージェントの場合、URL は暗号化時に長さが異なるにもかかわらず、一般に長さは同じであり、攻撃者は URL の長さをランダム化して回避しなければならないため、観察された URL の長さの標準偏差を判別データとして用いることで異常検出を行うことができる。

DNS 応答パケットの異常特性

(1) DNS ごとに IP アドレス数が異なる、攻撃者は様々な技術を用いて C&C ドメインの検出を逃れるため、ドメイン名関連の IP アドレスの変化が速くなる。また、DNS 応答パケットあたりの IP アドレス数は少なく、通常のサイトドメイン名は、DNS 応答パケット毎の IP アドレス数が比較的固定されている。ドメインの IP アドレスの変更頻度は、平均 TTL(Time To Live) 値でチェックできる。

(2) 関連付けプレフィックスの数が異なる、IP アドレスのプレフィックスも、1 つのドメインが正常ドメインであるか否かを判断することができる。正常なネットワーク中のホスト IP アドレスは通常ホスト名がないのバウンダリゲートウェイプロトコル (BGP) プレフィックスに属しているが、ゾンビネットワーク C&C 通信では IP アドレスプレフィックスは固定されていない。

3.2 異常検出アルゴリズム

異常検出は主に 2 段階に分けられ、第 1 段階では、ソフトウェアエージェントが自動生成した HTTP パケットブラウザから生成された HTTP パケットを正常なアクティビティで生成された HTTP パケットから分離し、ブラウザが自動的に発生するトラフィックを除くと、ソフトウェアエージェントが発生する HTTP トラフィックは、ゾンビネットワークの C&C 通信トラフィックである可能性が高い。第 2 段階では，半監督機械学習異常検出アルゴリズムを用いて、すべての IP アドレスから上位ドメイン名を抽出し、次にドメイン名ごとに必要な DNS 応答の特性を抽出し、タグ付けされたサンプルデータと未マーキングのデータトレーニングモデルを用いて、最後にこのモデルを利用してゾンビネットワークの C&C トラヒックを処理する。

このようなニーズに基づいて、チェビシェフ不等式に基づく異常検出アルゴリズムを検討し、教師なし方式にも半監視方式にも適用可能であり、データが未知であり、互いに独立している場合には、一般に良い検出結果が得られている。ODV(Outlier Detection Value) は異常検出値を表し，上限 (ODVU) または下限 (ODVL) はいずれのデータもこの範囲を超えると異常値と考えられ、通常 2 段階で ODV を計算する。第 1 段階では、次の段階で平均値と標準偏差を計算する際に、計算における偏差を除去するために、可能な異常値を主に除去する。第 2 段階では、データセットが ODV よりも極端なデータをすべて除去し、この段階のデータセットを用いて ODV を再計算し、ODV を完全なデータセットに適用して異常検出に用いる。

このほか、一方向サポートベクターマシンに基づく半監督的アルゴリズムと、教師なし機械学習によく用いられる最近傍局所異常因子に基づく異常検出アルゴリズム [8] も運用されていると考えられるが、今後の研究では、さらにアルゴリズムを比較する。

3.3 異常検出の過程

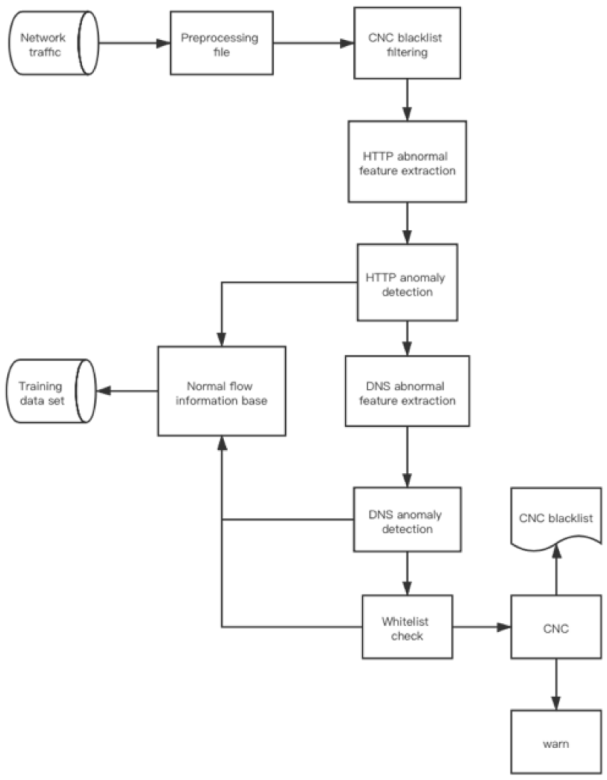

図 6 異常検出の動作過程

異常検知システムの動作過程を図に示すように、まず、データから HTTP リクエストパケットを抽出し，パケットの異常検出を行う。異常があれば、さらに DNS 応答パケットを抽出し、応答パケットの異常検出を行う。すべての過程で、通常のストリームのパケットであれば、パケットの情報をデータベースに保存するだけであり、後続の異常検出手順を経験する必要はない。悪意のあるネットワーク通信のパケットであれば、すべての段階の処理で最後に検出される。異常検出は特性の蓄積と発見の過程であるため、1 回目の特徴情報が少ないため検出に成功しない可能性がある。特性が蓄積されていくにつれて、ゾンビネットワーク C&C トラフィックは検出に成功する。また、異常検出はパケット全体を格納するのではなく少量の特性情報を記憶するだけであり、検出効率を大幅に向上させ、メモリのオーバヘッドを削減することができる。

4. まとめ

本研究では、MIRAI の伝播過程を分析することにより、ほとんどのマルウェア攻撃が C&C 機構を持つことが分かった。異なる検出技術を比較することにより，異常検出が新しいウイルスを検出できる有効な方式である。次に C&C 構造，プロトコルおよび通信隠蔽方式を分析して、ゾンビネットワーク C&C に存在する技術的欠陥をまとめた。異常検出では、C&C 通信におけるゾンビネットワークの異常特性を検研究した。HTTP リクエストの URL の数、周波数と長さは正常ネットワークトラフィックと異なり、DNS 応答では IP 数、TTL 平均値および関連プレフィックス数が正常ネットワークトラフィックと異なる。最後にこれらの特性を収集し、異常検出アルゴリズムを研究することにより、異常検出システムを提案した。今後の課題としては、異常検出のアルゴリズムをさらに比較すると、このシステムを実現し、誤報率および誤報原因を実験的に分析し，検出の正確性を向上させる。

参考文献

[1] Jelena Mirkovic and Peter Reiher, “A taxonomy of DDoS attack and DDoS defense mechanisms”, ACM SIG- COMM Comput. Commun. Rev.34, 2, 39-53, 2014.
[2] Manos Antonakakis, Tim April, Michael Bailey, Matthew Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Inv- ernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason , Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas and Yi Zhou,“Understanding the Mirai Botnet” In USENIX Security Symposium. 1093-1110, 2017.
[3] S.Balram, M.Wilscy. User Traffic Profile for Traffic Reduction and Effective Bot C&C Detection[J]. IJ Network Security, 2014, 16(1):46-52
[4] N.Taft, F.Giroire, J.Chandrashekar, et al. Exploiting temporal persistence to detect covert botnet channels[C]. International Workshop on Recent Advances in Intrusion Detection, Saint-Malo, France, 2009, 326-345.
[5] J Zhang, G GU, R Perdisci, et al. Botminer: Clustering analysis of network traffic for protocol and structure independent botnet detection[J]. 17th USENIX Security Symposium, 2008, 139-154.
[6] Domain names:Implementation specification[R]. http://tools.ietf.org/html/rfc883:Network Working Group.
[7] S.Zander, G.Armitage, P.Branch. A survey of covert channels and counter and counter measures in computer network protocols[J]. IEEE Communications Surveys and Tutorial, 2007,9(3):44-57.
[8] M M Breuning, H P Kriegel, R T Ng, et al. Exploiting temporal persistence of detect covert botnet channels[C]. International Workshop on Recent Advances in Intrusion Detection, Saint-Malo, France, 2009, 326-345.