初期段階における Remote Access Trojan の検知手法

蒋 丹　　面 和成

北陸先端科学技術大学院　情報科学研究科
923-1292　石川県能美市旭台 1-1
{jiangdan, omote}@jaist.ac.jp

あらまし　近年の標的型攻撃において，RAT (Remote Access Trojan) を代表とするマルウェアが組織に侵入することを完全に防ぐことは難しい．組織に侵入した RAT はスパイ系のマルウェアであり，遠隔操作によりひそかに組織の機密情報を盗む可能性があるため，RAT の不正アクセスをできるだけ早く検知することが重要である．本稿では,「接続初期段階」のネットワークパケット情報のみで RAT を検知するという新たな手法を提案する．具体的には，RAT と正常アプリケーション通信の目的の違いに着目し，通信特徴を抽出して機械学習（決定木，ランダムフォレスト）を行い，RAT の不正アクセスを高精度で検知するものである．本実験評価では，93.7%以上の精度を達成し，本手法が RAT の検知に有効であることが明らかになった．

Detection Method of Remote Access Trojan in an Early Stage

Dan Jiang　　Kazumasa Omote

School of Information Science, Japan Advanced Institute of Science and Technology
1-1, Aasahidai, Nomi-shi, Ishikawa, 923-1292, JAPAN
{jiangdan, omote}@jaist.ac.jp

Abstract　In recent years, targeted e-mail attack lets malware get into the Intranet of an organization more easily. RAT (Remote Access Trojan) is a kind of spy malware. After it invade the organization's network, it will monitor and control the victim's PC remotely, waiting for an opportunity to steal the confidential information. Therefore, it is important to detect RAT's network access in the Intranet as fast as possible. In this paper, we proposed an idea of "early stage of a session". We extract the features from the different purpose of the communication between normal applications and RAT. Afterward, we used Decision Tree and Random Forest to test the detection model and used 5-Fold cross-validation to evaluate it. The result of accuracy is over 93.7%, which shows that it is valid to detect RAT in the early stage.

1　はじめに

ネットワーク技術の進歩に伴い，個人情報の安全性問題は次第に表面化された．標的型攻撃は電子メール，USB，Web（ドライブバイダウンロード）の 3 つの侵入方法がある．例えば，IPA の公開資料 [1] によると，標的型メール攻撃による組織へのスパイ，諜報活動の脅威は 2014 年 10 大脅威の第一位になった．RAT (Remote Access Trojan) は諜報活動を行うマルウェアの一種であり，標的型攻撃による組織へ侵入した後，自主的にインターネットにいる攻撃者の C&C (Command and Control) サーバに接続して，トンネリング通信を張る．攻撃者は RAT を通じ

て，被害者 PC のファイル閲覧，ファイルダウンロード，画面キャプチャ，キーロガーなど様々な遠隔操作ができるため，被害者の組織の機密情報が漏れるリスクが高い．しかしながら，標的型攻撃において，RAT が組織に侵入することを完全に防ぐことが極めて難しい．したがって，RAT の不正アクセスをできるだけ早く検知することが重要である．

ネットワークベースで RAT を検知するには，膨大な組織ネットワークの通信データをリアルタイムで処理する必要がある．そのため，通信パケットの中身を見るのではなく，TCP のヘッダから簡単に抽出できるネットワーク情報を用いることが望ましい．故に，組織のネットワーク全体を管理する利便性やコストを考慮した上で，ネットワークベースの軽量な検知手法を用いることが重要である．なお近年では，RAT と P2P サービスやクラウドサービスである正常アプリケーションとの区別は難しいことに注意が必要である．

本研究では「接続初期段階」のネットワークパケット情報で RAT を検知する新たな手法を提案する．具体的には，RAT と正常アプリケーション通信の目的の違いに着目して通信特徴を抽出し，機械学習（決定木，ランダムフォレスト）を行い，RAT の不正アクセスを高精度で検知するものである．RAT は諜報活動を行うために目立たない通信をする傾向があると考えられるが，正常アプリケーションはそのような心配はない．故に，RAT は初期段階での通信量が正常アプリケーションより少ない傾向にあり，「接続初期段階」での特徴の差分で RAT を検知できると考える．例えば，初期段階の存続時間の実験データを見ると，正常アプリケーションが 0.03~97.11(s) になるが，RAT が 0.14~1.16(s) になる．

本稿では，2 章で既存研究について述べ，3 章で RAT および機械学習の手法を説明する．4 章で提案手法の詳細を述べ，5 章でその実験を行い，6 章で本研究に関する考察を行う．最後に，7 章はまとめと今後の課題について述べる．

2 既存研究

既存のネットワークベースのマルウェア検知に関する研究では，基本的に送受信パケット数，データ量，セッションの存続時間などの共通特徴を使っている．本章では新しい特徴または手法に関する研究を主に述べる．

Shicong Li らの研究 [2] では，RAT の攻撃者が毎回指令の結果を分析するのに時間がかかるため，「パケットの平均間隔時間が正常通信より長い」という新しい特徴を抽出した．検知精度は約 90%に達したが，情報漏えいを防ぐことが難しい．なぜなら，その検知は TCP ハンドシェイクから FIN/RST パケットまでの情報が必要であるが，RAT は侵入後，攻撃者が意図的に接続を切らない限りセッションがずっと張られる傾向があるため，FIN/RST パケットが発生する前に一部の機密情報がすでに洩れる可能性があるためである．

山田らの研究 [3] は RAT が組織内部において拡散等の諜報活動を検知するものである．しかし，イントラネット内の他の PC への諜報活動を行わないような RAT は検知できない．山内らの研究 [4] は正常の通信がボットネットのロボット的な通信よりランダム性があるため，「アクセス回数」と「アクセス時間の標準偏差」二つの新しい特徴を抽出して，HTTP 型ボットネットの C&C トラヒックを検知する．しかし，RAT の攻撃者はロボットプログラムでなく人間の場合が多く，RAT の攻撃アクセスにもランダム性があると考えられ，この方式では正常アプリケーションと RAT の区別が難しい．

Liang Yu ら [5] は「正常アプリケーションがマルチスレッドの場合が多い一方で RAT はそうではない」という違いから，「並行コネクションの割合」などの新しい特徴を抽出した．しかし，これらの特徴の抽出はホストのプロセス情報が必要ため，完全なネットワークベースの検知手法ではない．ネットワークベースとホストベースのハイブリッドによって確かに検知精度が上がるが，情報収集の時間及び資源のコストは組織においてのリアルタイム検知には適切ではないと考える．

Buyun Qu らの研究 [6] ではプロトコルの通

信交換のパターンから，最初の 6 個程度のパケット情報から，約 80%の精度で正常通信を分類できた．しかし，その研究は正常通信の分類であり，RAT の検知に有効かどうかは分かっていない．Vahid ら [7] は同じネットワーク行為パタンを持つ暗号化通信アプリケーションの分類に「パケットの PSH フラグの回数」という特徴を加えたが，RAT と同じく，P2P サービスやクラウドサービスの正常アプリケーションの通信パケットも PSH の回数が多く，区別できそうな情報量が含まれないと考える．

3 準備

3.1 RAT (Remote Access Trojan)

RAT は遠隔操作のマルウェアで，一般的には被害者であるサーバサイドと攻撃者であるクライアントサイドから構成される [2]．RAT は被害者に気づかれないようにトンネリング通信を張って，密かにリモート制御し悪意のある活動を行う．

代表的な RAT である Poison Ivy の攻撃者側の制御画面（図 1）を確認すると，ファイルの操作やキーロガー，スクリーンキャプチャだけでなく，デバイス/プロセス/レジストリの管理など様々な不正を実行できる．

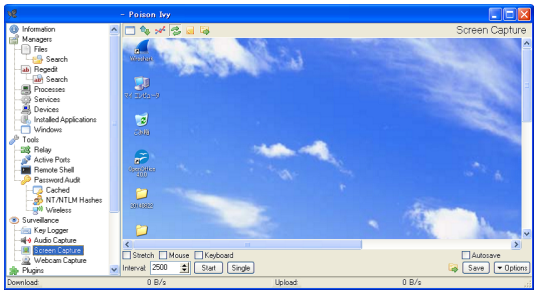

図 1: Poison Ivy の制御画面

RAT の二つのサイドの間の通信は二つのパターンがある．一つは攻撃者から接続の請求を発生するパターンであるが，ファイアウォールを乗り越えられないため，このパターンの RAT は少ない．もう一つは被害者からリクエストを発生するものであり，ファイアウォールを突破し，さらに Proxy サーバを経由することもできるため，RAT のメインの通信パターンである．

3.2 機械学習

機械学習とは，膨大なデータから情報を収集し，常時に更新するデータに合わせて潜在的な規則を把握した上で，データを処理するという手法である．研究には教師ありの機械学習アルゴリズム決定木とランダムフォレストを用いる．

3.2.1 決定木

決定木（Decision Tree, 以下 DT とする）は訓練サンプルの属性により分岐のノードを決め，再帰的に分岐属性ノードの選択を行って木を生成する手法で，全体のストラクチャは基本的に if-then の構造であり，非線形の識別境界を得る．非線形の識別境界はサンプルが混ぜる時識別精度も高く保つ利点がある．決定木アルゴリズムの代表は ID3, C4.5, CART(classification and regression tree) がある．CART は不純度の減り方が一番大きいという規準でノードを選ぶ．不純度はあるノードの分割規則を構成する時の候補点を表す指標であり，Gini 係数で計算する．

$$Gini\ Index = 1 - \sum_{c} p^2$$

その中，c はすべてのクラス，p はある特定のクラスに分類する確率である．

3.2.2 ランダムフォレスト

ランダムフォレスト (Random Forest, 以下 RF とする) は木の弱識別器を組み合わせるバギング方法の改良版で，学習データが少し変化してもより安定な識別器を構成することができる．RF は訓練サンプルからブートストラップサブサンプルを M 個生成し，ランダムの属性でサブ決定木を作成する．新しいテストデータに対して，すべてのサブ決定木に入力し，結果を多数決により出力する．RF の構成は図 2 のようになる．

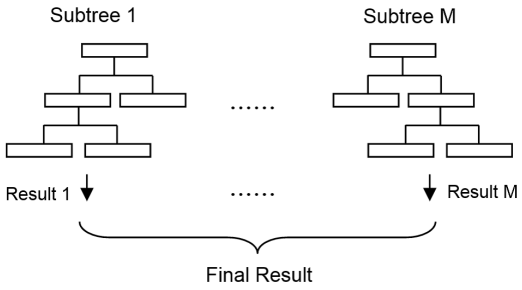

図 2: ランダムフォレストの一般構造

4 提案手法

RAT が組織に侵入した場合，外部の攻撃者と通信する不正アクセスを早い段階で検出し遮断することが望ましい．早期に検知し，窃取した情報を外部に転送する前に遮断することができれば，RAT の APT(Advanced Persistent Threat) 攻撃による組織の情報漏えいリスクも減少すると考える．我々は，接続の初期段階において，ネットワーク情報だけで RAT の不正アクセスを検知する新たな手法を提案する．ここで，接続初期段階を以下のように定義する．

定義 1 (接続初期段階) 接続初期段階は，*TCP 3-way* ハンドシェイクから始まり，間隔時間が t 秒未満の連続した *TCP* パケット列である．

接続初期段階が RAT の検知に有効だと考える理由は次の通りである．RAT 及び正常アプリケーションの接続初期段階は，目的の通信のための準備通信である．つまり，RAT の「目立たないように通信する」という正常アプリケーションの通信行動と根本的に異なる点がこの準備通信に存在すると考えられる．たとえ RAT の通信の仕組みが調整されたとしても，初期段階で正常アプリケーションと同じく多くのパケットを交換したり初期段階を長くしたりすることは「目立たないように通信する」という根本的な特徴に反すると考えられる．

本手法は，組織のゲートウェイでトラヒックを監視し，三つのフェーズ（特徴抽出フェーズ，学習フェーズ，検知フェーズ）で RAT の不正アクセスを検知する．

4.1 特徴抽出フェーズ

特徴抽出フェーズでは，初期段階と処理した Pcap ファイルから表 1 のネットワーク特徴を取得し算出する．具体的な処理は図 3 の通りである．まず特徴変数を初期化し，TCP パケットを一個ずつ読み込み，間隔時間を計算する．これが閾値 t 以下場合は 1~5 の特徴数値算出のプロセスに進み，計算完了後にまた新しいパケットを読み込む．間隔時間が閾値 t より長い場合は 6~8 の特徴数値を算出し，1~5 の特徴の最終的な値と共にネットワーク特徴ベクトルを出力する．

表 1: ネットワーク特徴

特徴	説明
PacNum	合計パケット数
Duration	初期段階セッションの存続時間
OutByte	Outbound のデータ量
OutPac	Outbound のパケット数
InByte	Inbound のデータ量
InPac	Inbound のパケット数
OB/OP	Outbound での一つのパケットの平均データ量 (OutByte/OutPac)
IB/IP	Inbound での一つのパケットの平均データ量 (InByte/InPac)

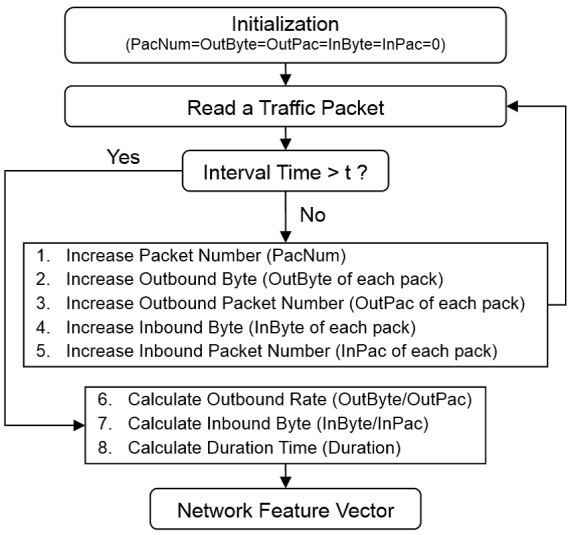

図 3: 特徴抽出フェーズ

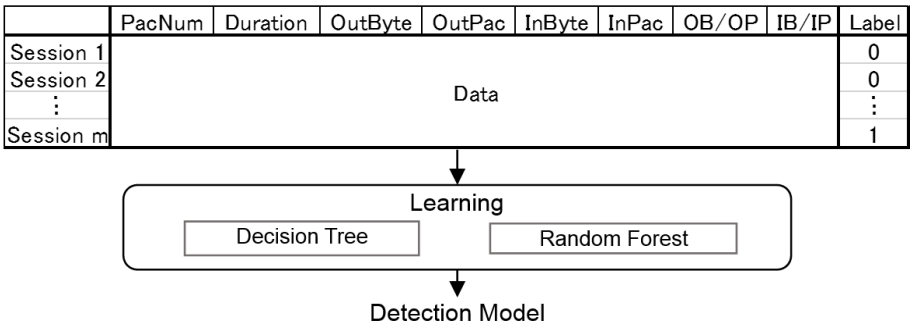

図 4: 学習フェーズ

4.2 学習フェーズ

学習フェーズでは，正常アプリケーションと RAT のデータから抽出した特徴ベクトル（ラベル付き）を DT と RF に入力し，検知モデルを作成する（図 4）．ラベルには 0 と 1 二つの値があり，0 が正常アプリケーション，1 が RAT を表す．ラベルを付けるのは教師あり機械学習アルゴリズムに使われるためである．

教師あり学習は主に分類（classification）に使い，答えがある訓練サンプルのにより導出された検知モデルによって，新しいテストデータのラベルを予測できる．

送信元 IP アドレスと宛先 IP アドレスをペアにして，一つのセッションとする．全てのセッションの特徴を統合し，ラベルを付けて 9 次元のベクトルを作成し，機械学習アルゴリズムを用いることによって検知モデルが作成される．

4.3 検知フェーズ

検知フェーズでは，組織の情報からリアルタイムに 8 次元のネットワーク特徴ベクトル（ラベルなし）を生成し，検知モデルに入力して検知の結果を出力する（図 5）．結果が 1 であれば RAT の通信であることを意味し，結果が 0 であれば正常通信であることを意味する．

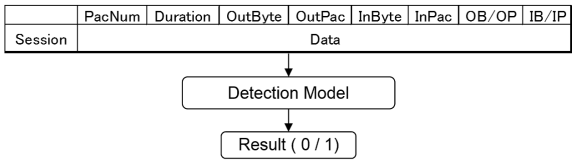

図 5: 検知フェーズ

5 実験

5.1 目的

本実験では，DT と RF を使い，5-Fold 交差検定で算出した Accuracy（精度），FPR（誤検知率），FNR（見逃し率）の指標で RAT の検知効果を評価し，接続初期段階が RAT の検知に有効であるかを検証する．

5.2 実験データ

本実験は 10 種類の RAT と 10 種類の正常アプリケーション（表 2）の通信データを利用する．正常アプリケーションは，RAT の通信特徴と近いと考えられるクラウドサービスおよび P2P サービスを対象とする．

5.3 実験手順

5.3.1 前処理

RAT の通信データは仮想環境上で動作させることによって収集し，正常アプリケーションの通信データは大学の実ネットワークから収集

表 2: 10 種類の正常アプリケーション

名称	説明
Dropbox	クラウドサービス
Skype	IM ツール
Chrome	Web ブラウザ
YahooMessenger	IM ツール
Firefox	Web ブラウザ
BitComet	P2P ダウンロードツール
Teamviewer	P2P 遠隔管理ツール
TorBrowser	匿名 Web ブラウザ
BitTorrent	P2P ダウンロードツール
PPTV	P2P ビデオ共有ツール

した．仮想環境は 2 台の実験 PC を接続させ，Windows の環境で RAT のサーバサイドとクライアントサイドを用意し，閉じられたネットワークの中でその間の通信を収集した．両方の通信データは Wireshark を用いて収集し，セッション単位に分けた．RAT の通信が 10 セッション，正常アプリケーションの通信が 165 セッション，合わせて 175 セッションの通信データを収集した．

次に，セッション毎に分割した Pcap ファイルから，研究対象外である次の通信データを削除した．本研究の目的が，機密情報をインターネット側の攻撃者に漏れることを防ぐことであるため，LAN 内の端末同士間の通信は対象外とする．また，RAT が双方向の通信を行うため，単一方向の通信や単一パケットの通信も全部対象外とする．

5.3.2 特徴抽出

前処理した Pcap ファイルを入力とし，初期段階の特徴抽出を行う．具体的には，パケットが到着する度に，ヘッダ情報から方向を判断し，PacNum, OutByte, OutPac, InByte, InPac 特徴変数の値を更新する．初期段階終了後は，最終的な値で OB/OP と IB/IP を算出し，特徴ベクトルを作成する．今回の実験では，初期段階の間隔時間を $t = 1$ とする．なお，特徴抽出全体のプログラムは Python で実装した．

5.3.3 機械学習

機械学習の実装も Python を用いて行った．Python には scikit-learn という機械学習のライブラリがある．scikit-learn には，代表的な様々な機械学習アルゴリズムが実装されている．本実験では，scikit-learn における Decision Tree Classifier と Random Forest Classifier の関数を使った．また，正常アプリケーションまたは RAT を表すラベルを付けた 9 次元の特徴ベクトルを学習させる．

5.3.4 交差検定

scikit-learn ライブラリは学習や予測だけでなく，交差検定などの評価の機能も備える．本実験では，K-Fold 交差検定を使う．具体的には次の 3 つのステップを行う．

1) データを重ならないように K 部に分ける．
2) $K-1$ 部を訓練データとして学習し，残りの 1 部をテストデータとして検知する．
3) 2) を K 回行って，評価の指標を求める．

175 セッション分のサンプルデータを平均的に分けるため $K = 5$ とし，KFold 関数で 5-Fold 交差検定を行う．一つの部分には (165/5) 個正常通信サンプルと (10/5) 個 RAT サンプルで構成され，合計 35 個のサンプルがある．

予測の結果を評価するために，Accuracy, FPR, FNR を以下のように算出する．

$$Accuracy = \frac{\text{正しく検知したサンプルの数}}{\text{サンプルの総数}}$$

$$FPR = \frac{\text{”1”と誤って検知した正常サンプルの数}}{\text{正常サンプルの総数}}$$

$$FNR = \frac{\text{”0”と誤って検知した } RAT \text{ サンプルの数}}{RAT \text{ サンプルの総数}}$$

表 3: 特徴パターンの検知精度の結果

Item	Accuracy	FPR	FNR	Feature			
DT	0.942	0.042	0.300	OutByte	OutPac	OB/OP	-
RF	0.954	0.036	0.200				
DT	0.942	0.042	0.300	OutByte	OutPac	OB/OP	PacNum
RF	0.942	0.042	0.300				
DT	0.965	0.024	0.200	OutByte	OutPac	OB/OP	InPac
RF	0.948	0.042	0.200				
DT	0.937	0.054	0.200	OutByte	OutPac	OB/OP	IB/IP
RF	0.954	0.036	0.200				

5.4 実験結果と評価

8 個の特徴の総計 255 通りの組み合わせで 5-Fold 交差検定を行った．8 個の特徴の組み合わせの結果を分析すると，最も検知精度が良い組み合わせパターンは表 3 の通りである．特に共通する「OutByte + OutPac + OB/OP」は，DT では Accuracy = 0.942，FPR = 0.042, FNR = 0.3, RF では Accuracy = 0.954, FPR = 0.036, FNR = 0.2 であった．これらの特徴は全て Outbound に関する特徴なため，Outbound に関する特徴が早期段階で RAT の不正アクセスの検知に有効であることが明らかになった．

全 255 通りの特徴の組み合わせの中，一番精度が高いパターン「PacNum + OutByte + OB/OP」は，DT では Accuracy = 0.931，FPR = 0.06, FNR = 0.2, RF では Accuracy = 0.971, FPR = 0.018, FNR = 0.2 であった．FNR として誤判定された RAT は Cerberus と Poison Ivy であり，これらの共通点は他の RAT より接続初期段階において多くの通信量を発生したことであった．

6 考察

6.1 接続初期段階について

接続初期段階を過ぎると通信データから RAT を区別することが難しくなる傾向が見られた．なぜなら，通信特徴の差分が小さくなるためである．具体的に表 4 は，Outbound データ量 (byte) を例とし，セッションの接続初期段階と最初の 100 パケットまでの違いを表している．正常アプリケーションは，通信データが複数のセッションあるため，ここの数値は平均値である．この表から，初期段階で RAT と正常アプリケーションのデータ量を比較すると，両者の差が明らかである．なぜなら，RAT では接続初期段階のデータ量が先頭 100 パケットまでのデータ量より少ないが，正常アプリケーションでは接続初期段階のデータ量が先頭 100 パケットまでのデータ量より多いものが半数以上あり，RAT のデータ量が正常通信と比べて明らかに少ないことが読み取れる．

6.2 特徴について

本研究で扱うネットワーク特徴はパケットのヘッダから直接抽出できるものであり，ペイロードも見ないため，暗号化通信でも膨大な通信量でも軽量に対応できる．また，実験で算出したデータ（表 5）を見ると，Outbound に関する OutByte と OutPac が確かに RAT と正常アプリケーションを区別する有力な特徴であることが分かる．

7 まとめ

本研究では，RAT が「目立たないように通信する」という根本的な特徴から,「接続初期段階」で検知する新たな手法を提案した．実験の結果，93.7%以上の検知精度 (Accuracy) に至り，早期

表 4: Outbound データ量 (byte) が接続初期段階と先頭 100 パケットまでの違い

RAT	初期段階/先頭 100 パケット	正常アプリケーション	初期段階/先頭 100 パケット（平均）
Bandook	303/2319	Dropbox	9148/10161
Bozok	308/4472	Skype	1626/1776
Cerberus	516/3055	Chrome	6458/2498
Nuclear	343/2420	YahooMessenger	18954/9888
Poison Ivy	1368/3612	Firefox	3387/2504
Turkojan	294/2977	BitComet	65827/2677
Gh0st	355/3175	Teamviewer	5232/5623
Netbus	129/74696	TorBrowser	90435/31593
OptixPro	221/1390	BitTorrent	1094/4271
ProRat	130/4583	PPTV	63471/3700

表 5: 特徴の考察

特徴	種類	傾向
PacNum	N	78%が **10pack** 以上
	R	20%が **10pack** 以上
Duration	N	64%が **0.5s** 以上
	R	20%が **0.5s** 以上
OutByte	N	93%が **500byte** 以上
	R	20%が **500byte** 以上
OutPac	N	89%が **6pack** 以上
	R	10%が **6pack** 以上
InByte	N	78%が **400byte** 以上
	R	10%が **400byte** 以上
InPac	N	54%が **6pack** 以上
	R	10%が **6pack** 以上
OB/OP	N	83%が **90byte/pack** 以上
	R	40%が **90byte/pack** 以上
IB/IP	N	76%が **100byte/pack** 以上
	R	10%が **100byte/pack** 以上

* N: Normal, 正常アプリケーション , R: RAT

段階で RAT の不正アクセスの検知が可能であることが示された．

接続初期段階の存続時間を見ると，正常アプリケーションが 0.03~97.11(s) になるが，RAT が 0.14~1.16(s) になる．RAT の存続時間が短い範囲に収まることは，RAT が小さい通信量で自分の存在を隠そうとする意図が見られる．故に，RAT と正常アプリケーションの通信特徴が「接続初期段階」での差分が明らかであり，初期段階での検知が有効であることが明らかになった．今後の課題として，FNR 指標がまだ高いことが挙げられる．

参考文献

[1] IPA 情報処理推進機構，”2014 年版情報セキュリティ10 大脅威”，2014 https://www.ipa.go.jp/files/000037151.pdf

[2] Shicong Li, Xiaochun Yun, Yongzheng Zhang, Jun Xiao, Yipeng Wang,“ A General Framework of Trojan Communication Detection Based on Network Traces”, IEEE NAS, 2012

[3] 山田正弘, 森永正信，海野由紀，鳥居悟，武仲正彦，“組織内ネットワークにおける標的型攻撃の諜報活動検知方式”, SCIS, 2014

[4] 山内一将, 川本淳平，堀良彰，櫻井幸一，“機械学習を用いたセッション分類による C&C トラヒック抽出”, SCIS, 2014

[5] Liang Yu, Peng Guojun, Zhang Huanguo, Wang Ying,“ An Unknown Trojan Detection Method Based on Software Network Behavior”, Wuhan University Journal of Natural Sciences, 2013

[6] Buyun Qu, Zhibin Zhang, Li Guo, Dan Meng,“ On accuracy of early traffic classification”, IEEE NAS, 2012

[7] Vahid Aghaei Foroushani, A. Nur Zincir-Heywood,“ Investigating Application Behavior in Network Traffic Traces”, IEEE CISDA, 2013