標的型攻撃検知システムの評価

北澤 繁樹　　祢宜 知孝　　河内 清人　　榊原 裕之　　藤井 誠司

三菱電機株式会社 情報技術総合研究所
〒 247-8501 神奈川県鎌倉市大船 5-1-1

あらまし　近年，“標的型攻撃” の脅威が深刻化している．標的型攻撃は，攻撃対象を特定の組織や人に限定した攻撃である．標的型攻撃は，巧妙な手口によってユーザにマルウェアを実行させようとすると伴に，アンチウィルスソフトウェアの回避措置が取られていることも多く，アンチウィルスソフトウェアでは十分な対策が難しい．そこで，本論文では，企業内部ネットワークに標的型攻撃によって侵入したマルウェアを早期検知する方式について述べる．提案方式では，これまで我々が開発してきた SW-PCA（Sliding Window - Principal Component Analysis）を用いて，時系列データの傾向の変化を検知する．評価では，提案方式が，従来方式よりも企業内部ネットワークで発生した標的型攻撃を早期に検知できることを，実測データ（CCCDATAset2009）を用いて示す．

Evaluations of A Targeted Attack Detection System

Shigeki Kitazawa　　Tomonori Negi　　Kiyoto Kawauchi　　Hiroyuki Sakakibara
Seiji Fujii

Mitsubishi Electric Corporation, Information Technology R&D Center
5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501, Japan

Abstract　Recently, the threat of “The Targeted attack” has become aggravated. The Targeted Attack is an attack that limits the attack target to a specific organization and the person. In the targeted attack, a user is made to execute the malware by the sleight of hand. Moreover, the evasion measures of Anti-virus software are often taken. Accordingly it is very difficult to take measures to the malware by Anti-Virus Software. In this paper, we propose the new Targeted Attack Detection Sysntem. Our system can detect network behaviors of the malware. We show that our system is very efficient to detect the behaviors of the malware by the real observed data(CCCDATAset2009).

1　はじめに

近年，標的型攻撃による脅威が深刻化している [1, 2]．標的型攻撃は，攻撃対象を特定の組織や人に限定した攻撃である [2]．文献 [1] によれば，標的型攻撃を受けた経験がある企業は 8.2% に上っている．ただし，標的型攻撃では，マルウェアをユーザに実行させるために，巧妙な仕掛け（ソーシアルエンジニアリング，文書ファイルへの埋め込みなど）がなされていることもあり，被害者は攻撃を受けたこと自体に気づきにくく，被害が表面化しにくいといった側面がある．したがって，実際に被害を受けている組織はさらに多いと推測される．

また，標的型攻撃では，アンチウィルスソフトウェアによる対策を回避するようゼロデイ攻撃が使用されることも多く，アンチウィルスソフトウェアでは十分な対策は難しい．これは，標的型攻撃による被害の多くが社員のマルウェアへの感染である [1] ことからも裏付けられる．

上記の課題を解決するため，我々は，これまで，インターネットで発生した未知ワームの早期検知を目的として開発してきた “セキュリティー攻撃予兆分析システム” [3, 4] をベースとして，“標的型攻撃検知システム” を開発した．標的型検知システムでは，企業内部ネットワークを監視対象として，監視対象上で観測されるマルウェアによる活動をいち早く検知して，対策をとることを目的としている．

本論文では，まず 2 節において，関連研究について述べる．次に，3 節では，開発した標的型攻撃検知システムについて説明する．4 節，5 節で，標的型攻撃検知システムの評価方法とその結果について述べ，6 節で，得られた評価結果に関して考察する．最後に，7 節で本論文をまとめる．

2　関連研究

ネットワークトラフィックを観測した時系列データを分析して，マルウェアの発生を検知する手法の1つに，文献 [5] がある．

文献 [5] においてトラフィックの時系列データの分析に用いられている ChangeFinder [6, 7]（以下，CF）は，MS.Blast や Sasser などのネットワーク感染型ワームや DoS 攻撃の早期検知に有効であることが示されている．

以下に，CF の分析ステップを簡単に説明する（詳細は，文献 [6, 7] を参照）．

Step1 第 1 段階学習
AR（AutoRegressive）モデル（次数 k）を用いたオンライン忘却型アルゴリズム（SDAR：Sequentially Discounting AR model learning）を用いて学習し，時系列データの確率密度関数の列を求める．求めた確率密度関数を用いて，外れ値スコア（Shanon 情報量）を計算する．

Step2 平滑化
得られた外れ値スコアの時系列をウインドウサイズ T で移動平均を計算する．

Step3 第 2 段階学習
Step2 で得られた時系列データに対して，再度 SDAR を用いて学習し，各時刻における変化点スコアを計算した結果を，ウインドウサイズ T' で移動平均を求める．

CF の特徴は，第 1 段階の学習では時系列中の外れ値しか検出できないところを，外れ値スコアの平滑化を通じて，ノイズに反応した外れ値を除去し，2 回目の学習によって本質的な変動のみを検出できるようにしたところにある [7]．

3　標的型攻撃検知システム

3.1　概要

我々は，これまで，インターネットで発生した未知ワームの早期検知を目的とした，“セキュリティー攻撃予兆分析システム” の研究開発を行ってきた [3, 4]．

セキュリティー攻撃予兆分析システムでは，マルウェアの活動によって，ネットワークの振る舞いに変化が観測されることに着目し，トラフィックの時系列データを主成分分析（Principal Component Analysis，以下 PCA）によって分析して，トラフィックの異常な変化を検知する [8, 9]．PCA は，多数の変量から少数の変量へ次元を縮退し，その次元を縮退した変量からデータを評価する分析手法である．PCA を適用することによって，変数の特徴を容易に評価・把握可能となる．

本論文では，これまでに開発した “セキュリティー攻撃予兆分析システム” をベースとして，企業内部のネットワークを監視して，近年，脅威が増している標的型攻撃を検知することを目的として開発した，標的型攻撃検知システムを提案する（以下，提案方式と呼ぶ）．標的型攻撃を早期に検知することにより，迅速に対策をとることが可能となるため，標的型攻撃による被害を最小限に抑えることができる．

次節以降で，提案方式について詳細を説明する．

3.2　システム構成

提案方式の構成を，図 1 に示す．

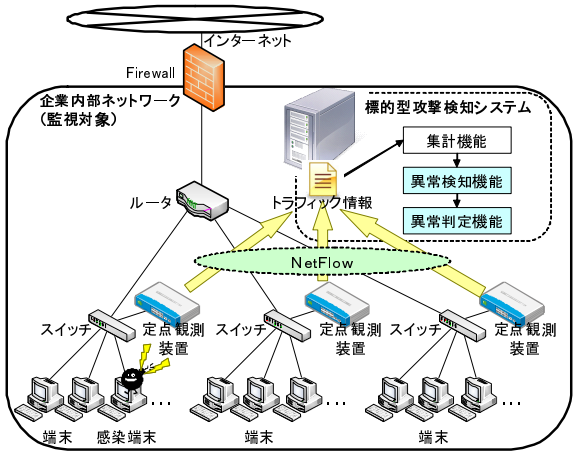

図 1: システム構成

提案方式では，監視対象となる企業内部ネットワーク上に配置した定点観測装置から NetFlow [10] を用いてトラフィック情報を収集する．

収集されたトラフィック情報は，提案方式の各機能（集計機能，異常検知機能，異常判定機能）によって処理され，最終的に，監視対象上で，マルウェアの活動が発生しているかどうかを判定する．

3.3　分析の流れ

本節では，提案方式における分析の流れを説明する（図 1 の集計機能，異常検知機能，異常判定機能）．

集計機能では，収集したトラフィック情報を，設定された集計条件（発信元 IP アドレス，宛先 IP アドレス，宛先ポート番号）にしたがって単位時間あたりの発生数やデータ転送量を集計する．

異常検知機能では，集計機能で作成した時系列データをウインドウサイズ（w）で，1 つずつずらしながら分割して，$(l-w+1)$ 個のパターン（w 次元行ベクトル）を抽出する（スライディングウインドウ方式，図 2 参照）．ここで，l は，学習期間中に観測した時系列データの個数を表している．抽出したパターンの変化を PCA によって分析することにより，時系列データの傾向の変化を検知する [9]（以下，SW-PCA と呼ぶ）．

時刻 t における SW-PCA の分析では，$(l-w+1)\times w$ 行列 $\mathbf{X}_t$ を作成する．切り出された最新のパターンとなる w 次元行ベクトルを $\mathbf{s}_t$ と置くと，作成される行列 $\mathbf{X}_t$ は，以下のようになる．

$$\mathbf{s}_t \quad = \quad (y_{t-w+1}, y_{t-w+2}, \cdots, y_{t-1}, y_t)$$

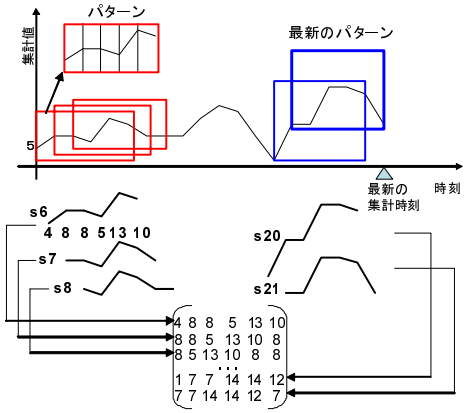

図 2: スライディングウインドウ方式（$w = 6$）によるパターンの抽出例

$$\mathbf{X}_t = \begin{pmatrix} \mathbf{s}_w \\ \mathbf{s}_{w+1} \\ \vdots \\ \mathbf{s}_{t-1} \\ \mathbf{s}_t \end{pmatrix}$$

作成した行列 $\mathbf{X}_t$ に対して PCA を適用することによって，各行の特徴量（主成分得点）を抽出する．

異常判定機能では，最新のパターンが学習パターンに含まれるか否かを，最新のパターンの特徴量の学習パターンの特徴量の集合からのマハラノビス距離を計算することによって判定する（図 3）．マハラノビス距離を用いるのは，学習パターンの特徴量の分布が，一様であるとは限らないため，ユークリッド距離で比較するよりも，分散を考慮したマハラノビス距離の方が，集合からの距離を正しく計ることができるためである．

異常としての判断は，学習パターンの特徴量のうち，マハラノビス距離が最大であるものを MD_{max} と置き，MD_{max} を β 倍（β:1 以上の実数）した値と最新のパターンの特徴量のマハラノビス距離（MD_{new}）を比較し，$MD_{new} > \beta MD_{max}$ の場合に，“学習パターンに属さない”（＝異常が発生している）と判断する．

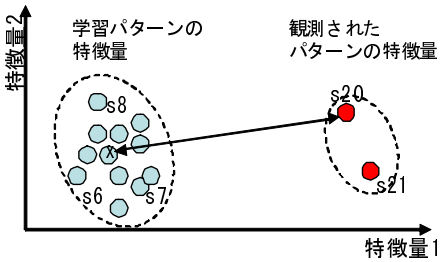

図 3: パターンの特徴量の比較

表 1 に，SW-PCA における分析パラーメータの一覧をまとめる．

表 1: SW-PCA における分析パラメータ一覧

パラメータ名	説明
w	ウインドウサイズ
l	学習した時系列データ個数
β	マハラノビス距離の閾値の係数

4 評価

4.1 評価手法

本論文では，実験により，企業内部ネットワークにおけるマルウェアの活動開始時刻と分析による検知時刻の差分をとることによってどれだけ早期に検知できるかを評価する．差分が小さい（最小値は 0）ほど，早期に検知したことを表す．

実験では，マルウェアに感染した端末のトラフィックとして，CCCDATAset2009 [11] に含まれる攻撃通信データ（パケットダンプデータ）から抽出した攻撃トラフィックを用いる．攻撃通信データには，任意に選択された 2 台のハニーポットが送受信した攻撃通信データが含まれている．

また，企業内部ネットワークの通常トラフィックとして，通常業務で使われている社内で観測したトラフィックを用いる．観測したネットワークには，Windows 系の端末が接続されており，ネットワークファイル共有も使用されている．

観測したトラフィックと，CCCDATAset2009 の攻撃通信データから抽出した攻撃トラフィックとを合成することで評価用データを作成して分析する．これにより，企業内部ネットワークの端末 1 台が標的型攻撃によってマルウェアに感染した状況を模擬し，マルウェアの活動を早期に検知できるか否かについて評価する．

4.2 攻撃トラフィック

攻撃トラフィックは，CCCDATAset2009 に含まれている，攻撃通信データから抽出する．そこで，まず，攻撃通信データの分析を行った．攻撃通信データには，マルウェアによる複数の活動が記録されていたが，その中で，2009 年 3 月 13 日の honey003 の通信に含まれていた，マルウェアの感染およびその後の活動と思われる通信に着目した．以下，時系列に沿って詳細を説明する．

00:29:53 xxx.xxx.113.235 からの ICMP（Echo Request）受信し，応答（Echo Reply）

00:29:54 xxx.xxx.113.235 からの TCP/139 の SYN パケット受信し，コネクション確立

00:30:00 xxx.xxx.113.235 との SMB セッションを匿名（anonymous）で確立

00:30:15 xxx.xxx.113.235 からの TCP/9988 宛 SYN パケット受信し，コネクション確立

00:30:18 xxx.xxx.113.235 から TCP/9988 経由で，データ（マルウェア本体）を受信完了

00:30:21 外部 IP アドレスに対して ICMP（Echo Request）によるネットワークスイープ開始．ICMP

（Echo Reply）を返した外部 IP アドレスに対して TCP/139 宛の接続および TCP/445 宛の接続を試行．

00:42:02 感染活動停止

上記のような活動をするマルウェアとしては，2007 年に発見された W32.Rahack.W が報告されている [12]．

本ケースでは，感染元ホスト（xxx.xxx.113.235）からのスキャンを受けてからわずか 28 秒で，感染活動が始まっている点，および，スキャンレートが高い（平均 33pps）点から，アウトブレイク型のマルウェアであるといえる．

honey003 を発信元とするマルウェアによる一連の振る舞いについて，攻撃通信データを元に NetFlow によりトラフィック情報を作成し，10 分ごとに集計したものを，図 4 に示す．

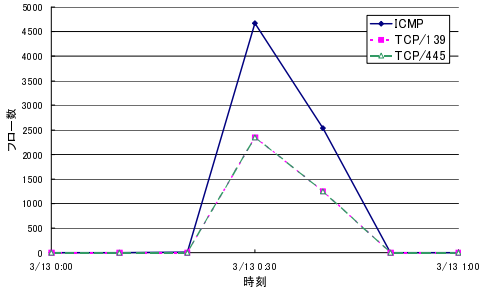

図 4: 生成された時系列データ

4.3 通常トラフィック

通常トラフィックとして，通常業務で使われている社内のクラス C ネットワークの入出力を 60 日分（10 分集計で 8640 点）観測した．なお，その間，セキュリティインシデントは発生していない．

観測で得られた時系列データ（ICMP，TCP/139，TCP/445）を，それぞれ，図 5，図 6，図 7 に示す．図に示すように，日常的にトラフィックが発生しており，トラフィックに周期性（平日 5 日，休日 2 の 1 週間）が見られる．

なお，これらのトラフィックは，マルウェアによって悪用されることが多く，企業内部ネットワークの監視においては一般に監視対象となるトラフィックである．

4.4 評価用データの作成

4.2 節で抽出したマルウェアの活動によるトラフィックと，4.3 節で観測した内部ネットワークのトラフィックを合成して評価用データを作成する．これは，監視しているクラス C ネットワークに接続された 1 台の端末が標的型攻撃により，マルウェアに感染した状況を模擬している．

攻撃トラフィックの埋め込みは，通常トラフィックの時刻による変動を考慮して，以下の組み合わせで行った．

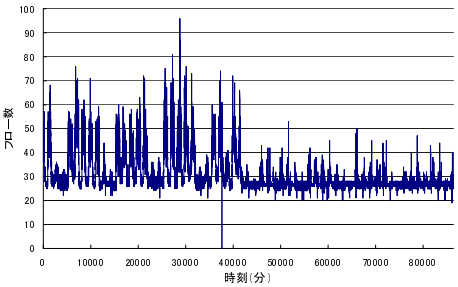

図 5: 通常トラフィック（ICMP）

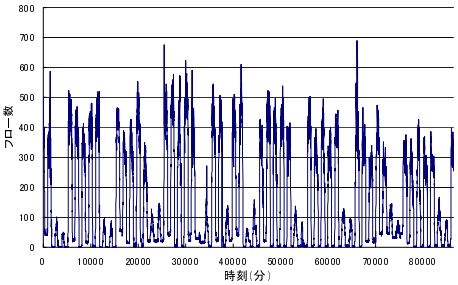

図 6: 通常トラフィック（TCP/139）

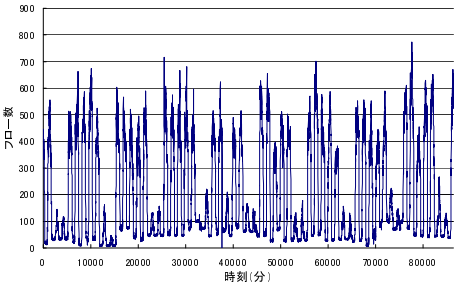

図 7: 通常トラフィック（TCP/445）

a) トラフィックが増加し始めた時刻
b) トラフィックがピークに達した時刻
c) トラフィックが減少し終えた時刻

上記，組み合わせを，任意に選んだ平日と休日に対して埋め込んで計 6 種類の評価用データを作成する．図 8 に平日の TCP/445 のトラフィックに対して，a（トラフィックが増加し始めた時刻）に TCP/445 の攻撃トラフィックを合成して作成した評価用データを示す．

4.5 分析パラメータ設定

提案方式では，分析を行うにあたり，表 1 に示した分析パラメータを設定する．まず，学習期間は，これまでの評価実績に基づいて，30 日間（$l = 4320$）とした．したがって，作成した評価用データの前半 4320 点は，学習データとなる．

β の値は異常を検知する感度に影響を与える．評価では，$\beta = 1.0$ に設定し，最大感度とした．

w の値は，一般に企業の勤務時間が 8 時間であることを考慮して，8 時間分の 24 点を採用した．

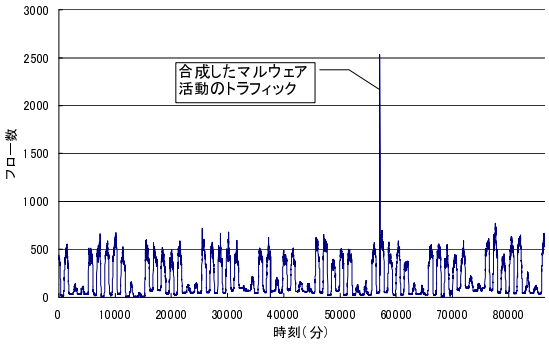

図 8: 評価用データの例（TCP/445，a: トラフィックが増加し始めた時刻）

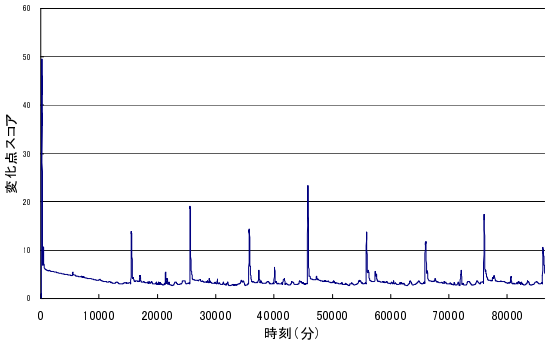

図 9: CF の変化点スコア（TCP/445）

また，本論文では，評価結果の比較対象として，2 節で述べた CF を選択した．CF は，文献 [6, 7, 13] を参考に MATLAB [14] で実装した．

CF における分析パラメータは，文献 [6, 7, 13] によれば，忘却パラメータ（$0 < r < 1$），AR の次数（k），移動平均のウインドウサイズ（T, T'）がある．文献 [7] によれば，忘却パラメータは値が小さいほど，過去データの重みが増す．また，移動平均のウインドウサイズは 5 から 10 に設定することが多く，大きいほど検知は遅れるが，外れ値がフィルタリングされて変化点だけが検知できるようになる．

評価では，文献 [13] で，TOPIX の実データの分析で使われていた値を採用し，$r = 0.005$，$k = 4$，$T = T' = 5$ とした．

上記の設定を行い，社内における観測で得られた攻撃データを含まない時系列データ（図 5，図 6，図 7）を用いて，誤検知が発生するか否かを調査するため，分析を行った．

結果として，提案方式では誤検知が発生しなかった．CF では，休日明けの平日，トラフィックが増加するタイミングで，変化点スコアの上昇が見られた（図 9）．そこで，本論文では，CF にてこれらを誤検知しないよう，休日明けの平日午前を基準として ICMP，TCP/139 ，TCP/445 の変化点スコアの閾値を，それぞれ，26，25，26 と定めた．

表 2: （実験 1） 合成時刻と検知時刻の差分（分）

		平日			休日		
		a	b	c	a	b	c
ICMP	提案方式	0	0	0	0	0	0
	CF	60	60	60	60	60	60
TCP/139	提案方式	0	0	0	0	0	0
	CF	80	80	70	70	70	70
TCP/445	提案方式	0	0	0	0	0	0
	CF	80	80	-	70	70	70

表 3: （実験 2）合成時刻と検知時刻の差分（分）

		平日			休日		
		a	b	c	a	b	c
ICMP	提案方式	0	0	0	0	0	0
	CF	60	60	60	60	60	60
TCP/139	提案方式	10	10	10	10	10	10
	CF	-	-	-	-	-	-
TCP/445	提案方式	20	20	10	20	30	20
	CF	-	-	-	-	-	-

5 評価結果

実験では，まず，4.4 節で作成した評価用データを，提案方式，ならびに，CF を用いて分析した（実験 1）．加えて，4.2 節で抽出したマルウェアの振る舞いが，アウトブレイク型であったため，より，見つかりにくいマルウェアの振る舞いとして，スキャンレートを 1/5 に抑えたマルウェアの振る舞いのトラフィックについても 4.4 節と同様の手順で合成し，作成した評価用データについて分析を行った（実験 2）．1/5 としたのは，TCP/139，TCP/445 の通常トラフィックのピーク以下に抑えるためである．

提案方式，ならびに，CF で分析を行った結果，攻撃トラフィックを合成した時刻と検知時刻の差分を，表 2，表 3 に示す．なお，表において，“ - ” は，検知漏れを表している．

まず，表 2 について，提案方式では，全てのトラフィック（ICMP，TCP/139，TCP/445）の監視において，マルウェアの振る舞いを合成した時刻で異常検知しているのに対して，CF では，60 ～ 80 分遅れて検知していることが分かる．

次に，表 3 について，提案方式では，TCP/139，TCP/445 のトラフィックの監視において，10 ～ 30 分の遅れが検知に表れている．これに対し，CF では，ICMP については，60 分程遅れて検知しているが，TCP/139，TCP/445 では変化点スコアの閾値を越えなかった（検知漏れ）．

以上の結果から，提案方式は，CF よりも早期にマルウェアの振る舞いを検知できる．加えて，スキャンレートを抑えたマルウェアについても CF よりも検知性能が高いといえる．

6 考察

評価の結果，提案方式は CF に比べて，以下の点で優れていることが分かった．

1. マルウェアの振る舞いを発生直後に検知可能であること
2. スキャンレートを抑えたマルウェアに関しても検知が可能であること

まず，1 について，提案方式では，SW-PCA によって，学習期間に観測された時系列データから，グラフの形状をパターンとして学習している．したがって，企業内部ネットワークのように，比較的安定したトラフィックが発生する環境においては，学習したパターンからの逸脱が顕著にあらわれるため，異常の発生と同時に検知が可能である．

一方，CF では，アルゴリズムの中でノイズ成分を除去するために，移動平均によって平滑化を行っている．このため，マルウェアの振る舞いの変化が観測された時点よりも遅れて変化点スコアに現れるためであると考えられる．

次に，2 について，CF では，スキャンレートを抑えたマルウェアの検知（実験 2）において，TCP/139，TCP/445 のトラフィックの分析では，検知ができなかった．これは，4.5 節でも触れたように，CF では，休日明けの平日午前のトラフィックの変動に変化点スコアが敏感に反応していたため，閾値を下げることができなかったことが挙げられる．仮に閾値を下げたとしても，誤検知が多発してしまうため，検知結果の信憑性が損なわれてしまう．したがって，CF を用いる場合，内部ネットワークの TCP/139 や TCP/445 のトラフィックのように，特徴の異なるデータ（平日と休日）が混在して現れる時系列データを分析した場合には，誤検知や検知漏れに繋がる．

これを裏付ける結果として，平日，休日を問わず，ほぼ安定している ICMP のトラフィックの分析では，遅れは生じているもののスキャンレートを抑えたマルウェアの活動を検知できている．

7 まとめ

本論文では，我々が開発した標的型攻撃検知システム（提案方式）について，検知性能を評価した．

評価では，CCCDATAset2009 の攻撃通信データに記録されていたアウトブレイク型マルウェアの活動を抽出し，社内で観測した実トラフィックと合成することによって，標的型攻撃により，内部ネットワークの端末がマルウェアに感染した場合のトラフィックを模擬した評価用データを作成し，評価した．加えて，アウトブレイク型マルウェアのスキャン活動を 1/5 にスキャンレートを抑えたマルウェアの活動の検出についても評価を行った．

評価の結果，提案方式は，既に提案されている CF と比較して，内部ネットワークで発生したマルウェアの活動を早期に検知できることを示した．

参考文献

[1] 有限責任中間法人 JPCERT コーディネーションセンター（JPCERT/CC）:「標的型攻撃について」(2008). http://www.jpcert.or.jp/research/2008/inoculation_200808.pdf.

[2] 独立行政法人情報処理推進機構（IPA）：情報セキュリティ白書 2009 第 II 部 10 大脅威攻撃手法の『多様化』が進む (2009). http://www.ipa.go.jp/security/vuln/documents/10threats2009.pdf.

[3] 榊原裕之，藤井誠司，北澤繁樹，平井規郎，鹿島理華，東 辰輔：定点観測による不正アクセス分析システムの提案，情報処理学会 第 68 回全国大会 (2006). 5E-3.

[4] 榊原裕之，北澤繁樹，大野一広，藤井誠司：定点観測による不正アクセス分析システム，情報処理学会研究報告 (2006). CSEC-35-13.

[5] 竹内純一，佐藤靖士，力武健次，中尾康二：変化点検出エンジンを利用したインシデント検知システムの構築，暗号と情報セキュリティシンポジウム 2006（SCIS2006）(2006).

[6] Takeuchi, J. and Yamanishi, K.: A Unifying Framework for Detecting Outliers and Change Points from Non-Stationary Time Series Data, *IEEE transactions on Knowledge and Data Engineering*, Vol. 18, No. 4, pp. 489–489 (2006).

[7] 山西健司：データマイニングによる異常検知，共立出版 (2009).

[8] 平井規郎，鹿島理華，東 辰輔，榊原裕之，藤井誠司，北澤繁樹：定点観測による不正アクセス分析システムの提案～ワーム攻撃による異常検出のためのネットワークログ分析手法～，情報処理学会 第 68 回全国大会 (2006). 5E-4.

[9] 鹿島理華，藤森敬悟，平井規郎，榊原裕之，大野一広，藤井誠司：定点観測による不正アクセス分析システム～不正アクセス検知のためのネットワークログ分析手法～，情報処理学会 第 69 回全国大会 (2007). 2F-4.

[10] Claise, B., Sadasivan, G., Valluri, V. and Djernaes, M.: [RFC 3954] Cisco Systems NetFlow Services Export Version 9 (2004). http://www.ietf.org/rfc/rfc3954.txt.

[11] 畑田充弘，中津留勇，寺田真敏，篠田陽一：マルウェア対策のための研究用データセットとワークショップを通じた研究成果の共有，マルウェア対策研究人材育成ワークショップ 2009 (2009). MWS2009.

[12] Symantec: W32.Rahack.W テクニカルノート. http://www.symantec.com/ja/jp/security_response/writeup.jsp?docid=2007-011509-2103-99&tabid=2.

[13] 竹内純一，山西健司：忘却型学習アルゴリズムを用いた外れ値検出と変化点検出の統一的扱い，2002 年情報論的学習理論ワークショップ（IBIS2002）(2002).

[14] The MathWorks, Inc.: MATLAB 2009a. http://www.mathworks.co.jp/.