ベイズ推測を用いた不正侵入イベント増減予測

石田千枝[†] 島田英一[††] 荒川豊[†]
竹森敬祐[†††] 笹瀬巌[†]

昨今，侵入検知システム（Intrusion Detection System: IDS）を用いたネットワーク監視が進められる中，攻撃による異常を把握するためのイベント分析に関する研究がさかんに行われている．攻撃への対策をあらかじめ整えておくためには，今後活発化する攻撃を事前に把握しておく必要がある．しかし，既存のイベント分析は起きている攻撃の現状を把握することはできるが，未来の攻撃頻度の状態を知ることはできない．そこで本論文では，IDS から出力されるイベントについて，攻撃の周期や増減度の推移に注目した，2 種類 4 パターンの不正侵入イベント増減予測アルゴリズムを提案する．提案アルゴリズムは予測計算にベイズ推測を用い，あるイベントの過去の発生頻度の推移を単位時間ごとに学習して，1 つ未来の単位時間の増減確率を算出している．実装した予測システムについて，実運用されている IDS イベントを用いて，未来のイベント頻度が増加する確率を計算し，その正解率について評価する．その結果，攻撃ごとにイベントの増減特性が異なることを明らかにし，攻撃推移の特徴に即した予測アルゴリズムを適宜選択することで，高い正解率を達成できることを示す．

Forecast of Increasing or Decreasing Intrusion Event Counts Using Bayesian Inference

Chie Ishida,[†] Eiichi Shimada,[†] Yutaka Arakawa,[†]
Keisuke Takemori[†††] and Iwao Sasase[†]

An intrusion detection system (IDS) is an important tool to detect and to analyze network attacks. The analysis techniques of IDS events are actively researched, since it is important to make use of results of analysis in understanding attack trends. To aim at a quick response in security operation, it is important to find the attacks that gets larger in the future. However, conventional approaches cannot indicate future fluctuation of attacks. In this paper, we propose forecast algorithms for increasing or decreasing the event counts. We consider two algorithms by focusing on an attack cycle and a fluctuation range of the event counts. Our algorithms use Bayesian Inference for calculating the conditional probability based on past-observed event counts to forecast increasing or decreasing the event counts. We implement the forecasting system and evaluate it with real IDS events. Experimental evaluations show that each attack has discrete characteristic of fluctuation and our proposed algorithms can forecast increasing or decreasing the event counts by selecting most appropriate algorithm for each attack.

1. はじめに

近年，ネットワークの拡大，高速化によって多くのコンピュータがインターネットに常時接続されるようになっている．これにともない，コンピュータウイルスや Denial of Services（DoS）攻撃といったネットワークを介した攻撃の被害も増大している．ネットワークやシステムに対して行われるこのような攻撃を検知するツールとして，IDS が注目されており，攻撃による異常を把握するための IDS イベント分析に関する研究がさかんに行われている[1)~4)]．しかし多くの IDS は，誤検知によって多量のイベントを出力してしまう傾向があり，危険をともなう注意すべき重要なイベントの変動を見落としてしまう問題がある．多量なイベントを効率良く処理する方法として，イベント頻度の平均と標準偏差を用いてイベントの異常度を数値化する手法[5)]が提案されている．これによって，頻度はそれほど大きくないが増加率の高い危険なイベントを容易に発見できるようになった．また，イベントの過去の発生頻度の推移を学習して現在の異常度を数値

† 慶應義塾大学
Keio University
†† 株式会社 NTT データ
NTT DATA Co.
††† 株式会社 KDDI 研究所
KDDI R&D Laboratories, Inc.

化する手法[6]や，広域ネットワーク上でのイベント頻度の活発化に注目した危険度評価手法[7]などが提案されている．これらによって，異常な攻撃の把握は可能になった．攻撃への対策をあらかじめ整えておくためには，今後活発化する攻撃を事前に把握しておく必要がある．しかし，既存のイベント分析は起きている攻撃の現状を把握することはできるが，未来の攻撃頻度の状態を知ることはできない．

そこで本論文では，IDS によって検知される不正侵入イベントの増減予測アルゴリズムを提案する．予測アルゴリズムとして，イベント頻度を直接計算に用いることができ，かつ算術過程が運用者に理解されやすいという利点から，ベイズ推測を用いることにする．これにより，イベントの過去の増減パターンを単位時間ごとに学習しておき，現時点の攻撃状態から，次の単位時間でイベントが増加する確率を予測する．具体的には，Attack Signature と呼ばれる既知の侵入手法を検知したイベント頻度の周期性や増減度の推移に注目して，2 種類計 4 パターンの予測アルゴリズムを考える．そして予測システムの実装を行い，実運用されている IDS イベントを用いて，未来のイベント頻度が増加する確率を計算し，その正解率について評価を行う．その結果，攻撃ごとにイベントの増減特性が異なることが判明し，攻撃推移の特徴に即した予測アルゴリズムを適宜選択することで，高い正解率を達成できることを示す．

以下，2 章において，既存の IDS イベント分析に関する研究とその問題点について述べ，本研究における要件をまとめる．3 章で攻撃イベント数の増減予測アルゴリズムを提案し，4 章で実装を行う．5 章で実データを用いた評価を行い，最後に 6 章で結論を述べる．

2. IDS イベント分析のシステムとその課題

本章では，広域ネットワークを監視する IDS イベント分析システムの構成について説明し，その分析手法に関する従来の諸研究について述べ，課題をまとめる．

2.1 IDS イベント分析システム

図 1 に IDS イベント分析システムの構成を示す．IDS agent が各地のネットワークに対して行われる攻撃を検知するエンジンであり，IDS manager が IDS イベントの分析を行う部分である．IDS manager は，Event collecter，Event data base（DB），Event analyzer，Interface によって構成されている．各地のネットワークに設置された IDS agent から出力されたイベントは，Event collector によって収集され，Event DB に保存される．Event analyzer は IDS イベントを分析し，その結果を Interface に出力して，運用者に通知する．

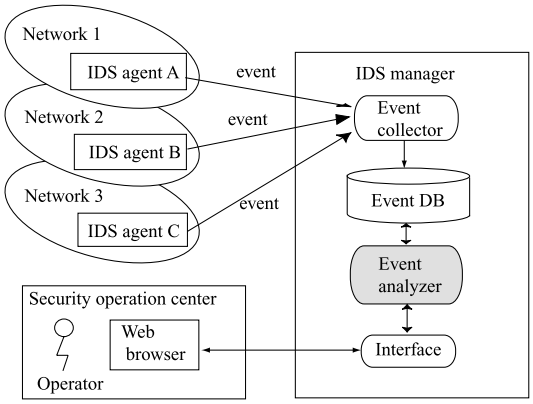

図 1 IDS イベント分析システムの構成
Fig. 1 Framework of analysis system for IDS events.

2.2 IDS イベント分析の従来研究と問題点

活発化する攻撃への対策を図るためには，IDS から出力されるイベントの危険度を把握する必要がある．主な研究として，個々の IDS イベントを，既知の侵入手法に分類し，攻撃者の行動手順を作成する手法がある[6]．これは，現在のイベントに見られる特性が，あらかじめ作成しておいた行動手順にあてはまる確率についてベイズ推測を用いて評価することで，その危険度を算出している．また，インターネット上で観測される攻撃先を探索するためのポートスキャンイベントの頻度の推移を，ベイズ推測によって学習して，ある時点の観測値と推測した値との乖離を異常値として危険状態を評価する手法がある[7]．これにより，学習値よりも活発化した危険な状態を的確に検知することができるようになる．

しかしながらこれらの研究は，起きている攻撃の現状を把握することはできるものの，今後活発化する攻撃への対策をあらかじめ整えておくための情報を提供するものではない．より積極的に攻撃からシステムを守るためには，IDS イベントの増減を予測する機能が必要となる．攻撃によってイベントの増減特性が異なることが考えられるため，様々な攻撃に対応してイベントの増減を予測できることが重要である．また，運用者に迅速な対策を促すためには，分析結果だけでなくその計算過程も理解できるものでなければならない．

2.3 予測手法

攻撃の未来の推移を高い精度で予測するには，攻撃の過去の変動のパターンを学習して，現在の状態から今後の攻撃が増加する確率を求めることになる．過去のパターンを学習して予測する手法として，状態遷移

モデルを利用して事象の発生過程を予測するマルコフモデル[8),10)]や，観測値の時系列に従って，予測値と観測値の誤差を補正しながら予測していくカルマンフィルタ[9)]がある．しかし，マルコフモデルにおいてはネットワーク IDS で取得されたイベントからでは，予測に用いる遷移モデルの作成が困難であるという問題がある．また，カルマンフィルタにおいては攻撃を数式化して適用するのが難しいという問題がある．ベイズ推測を用いて，ホストにログインしているユーザの行動を予測している研究[11)]もあるが，ネットワーク上の攻撃を予測するものではない．

2.4 要　　件

以上の背景と問題点より，予測のための要件を以下にまとめる．

- 要件 1：様々な攻撃に対応して予測できること．
- 要件 2：予測手法が運用者に理解されやすいこと．
- 要件 3：精度の高い予測を行えること．

3. 提 案 方 式

本章では，IDS によって検知されるイベントの増減を予測するためのアルゴリズムを提案する．要件 2 を考慮して，イベント頻度を直接計算に用いることができ，かつ算術過程が運用者に理解されやすい分析手法としてベイズ推測を用いる．ベイズ推測に対して，イベント頻度の周期性や増減度の推移に注目して 2 種類計 4 パターンのアルゴリズムを提案，適用する．

3.1 ベイズ推測

ベイズ推測は観測値に基づいて条件付確率を計算する方法であり，以下の式で表すことができる．

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \qquad (1)$$

式 (1) は，2 つの独立な事象 X，Y について，Y が与えられた場合に X が起こる条件付き確率を表す．本研究において事象 X をイベント頻度が未来で増加するという事象とし，事象 Y は現在のイベント状態とする．ここで Y は，予測アルゴリズムによって異なる学習の条件となる．式 (1) より求められる P(X | Y) は Y の条件の下に今後イベント頻度が増加する確率となり，0〜100%の値で算出される．

3.2 提案アルゴリズム

式 (1) の Y に該当する事象を変えて 2 種類のアルゴリズムを提案する．イベントの中から Attack Signature ごとに集計したイベント頻度の推移に注目すると，その増減に時間や曜日による周期が見られる Attack Signature があると推測する．また，一見ランダムに増減が変化しているように思われる Attack Signature の中には，その増減の程度が未来の増減に影響しているものがあると推測する．これらの特徴に注目して，過去のイベント頻度の推移を単位時間ごとに学習しておき，式 (1) を用いて次の単位時間のイベント頻度が増加する確率を求める．

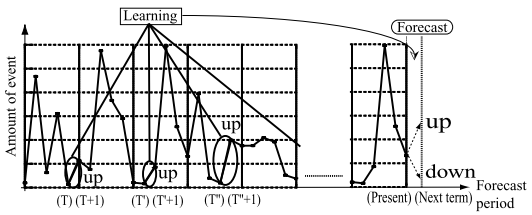

図 2 予測アルゴリズム 1 の例
Fig. 2 An example of forecast algorithm 1.

3.2.1 周期を考慮したアルゴリズム（予測アルゴリズム 1）

ここでは，イベント頻度の周期を考慮した予測を行う．1 日の時間帯の中では日中や夜はイベント頻度が高く，早朝はイベント頻度が低くなるといった時間による周期や，平日はイベント頻度が高く，土曜日と日曜日はイベント頻度が低くなるといった曜日による周期を持つ Attack Signature がある．**図 2** に予測アルゴリズム 1 の例を示す．図 2 は予測アルゴリズム 1 において曜日による周期に着目した場合の図である．図 2 の (T)，(T')，(T") は同じ曜日となり，(T+1)，(T'+1)，(T"+1) はその翌日の曜日となる．学習の段階で，T 日の曜日から翌日の曜日にかけてイベント頻度が増加した日数と減少した日数をそれぞれ過去のイベントからカウントする．予測の段階で，今日の曜日が事象 Y として与えられたときに，学習した過去のイベント頻度の推移に基づいて，次の日のイベント頻度が増加するという事象 X が起こる確率を式 (1) より求める．このとき，今日が日曜日であると仮定する．式 (1) 右辺の分母 P(Y) は T 日が日曜日である確率（この場合 1/7）となり，分子の P(X) は全日数に対する翌日にイベント頻度が増加した日数から計算される増加確率，P(X | Y) は翌日にイベント頻度が増加した日数に対してその中で日曜日である日数から計算される条件付き確率となる．これらの値を式 (1) に代入し，求められる P(X | Y) が，今日が日曜日である場合に翌日のイベント頻度が増加する確率となる．また，時間による周期に着目すると，T 時の時刻を事象 Y とした場合，(T+1) 時のイベント頻度が増加する確率を求める．

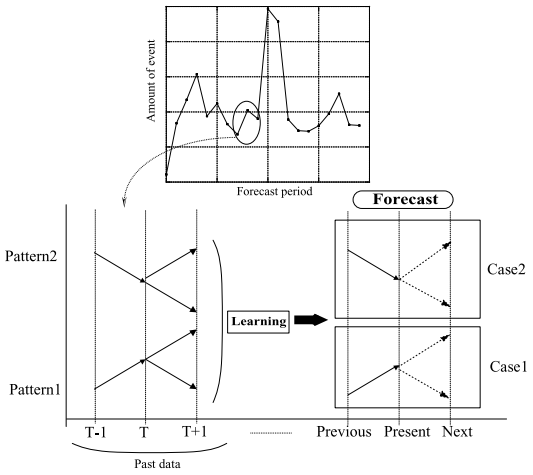

図 3 予測アルゴリズム 2-1 の例
Fig. 3 An example of forecast algorithm 2-1.

3.2.2 増減の度合いを考慮したアルゴリズム（予測アルゴリズム 2）

ここでは，イベント頻度の増減の度合いを考慮した予測を行う．ある時間単位 (T−1) から T にかけてのイベント頻度の増減の度合いを事象 Y として与える．事象 Y は (T−1) のイベント頻度に対する T のイベント頻度の割合である．さらに，増減の度合の範囲の場合分けを細かくして予測することとし，事象 Y の場合分けのパターンとして以下の 3 つのパターンを適用する．

- 事象 Y を 2 パターン設定（予測アルゴリズム 2-1）．
- 事象 Y を 4 パターン設定（予測アルゴリズム 2-2）．
- 事象 Y を 8 パターン設定（予測アルゴリズム 2-3）．

図 3 に予測アルゴリズム 2-1 の例を示す．ある時間単位 (T−1) から T にかけてのイベント頻度の増減の度合いによって，(T+1) のイベント頻度が増加した回数と減少した回数をそれぞれ過去のデータより集計する．ここで，1 つ前の予測期間から現在にかけてのイベント頻度の増減の度合いが事象 Y として与えられたときに，学習した過去のデータを用いて次の予測期間のイベント頻度が増加するという事象 X が起こる確率を式 (1) より求める．このとき，使用するアルゴリズムは予測アルゴリズム 2-1 とし，予測の時間単位を 1 日とする．式 (1) 右辺の分母 P(Y) は全日数に対して前日から今日にかけてイベント頻度が増加する確率となる．分子の P(X) は全日数に対する翌日にイベント頻度が増加した日数から計算される増加確率，P(X | Y) は翌日にイベント頻度が増加した日数に対してその中で前日から今日にかけてイベント頻度が増加した日数から計算される条件付き確率となる．これらの値を式 (1) に代入し，求められる P(X | Y) が，前日から今日にかけてイベント頻度が増加した条件において，翌日のイベント頻度が増加する確率となる．

4. 実　　装

ここでは，ネットワーク上で実際に運用されている IDS から出力されるイベントの増減予測の精度について検証するためのシステムを実装する．

4.1 実装内容

予測システムの実装に使用した計算機は，CPU が Pentium4 3 GHz，メモリ容量が 1 GByte であり，Windows XP OS 上で SQL Server を利用して実装した．使用言語は Visual Basic Scripting Edition（VBScript）である．システムの操作および結果の出力には Web ブラウザを用いることにして，Web サーバ上から SQL DB を操作する Active Server Pages を実装する．この計算機環境で，翌日の予測に必要な 1 カ月分の約 62 万イベントを用いて 30 秒〜1 分程度で予測計算を処理できる．

IDS のイベントは，図 1 で示す手順に従い DB に登録される．DB に保存したイベントから，予測計算に必要なパラメータを抽出する．IDS から出力されるイベントには，検知日時，Attack Signature, Source/Destination Port，Source/Destination IP，イベント検知数，通信プロトコルなどの情報が含まれている．本研究では提案アルゴリズムを適用する IDS イベントのパラメータとして Attack Signature に着目して予測計算を行うものとする．Attack Signature の一覧の中から選択した 1 種類の Attack Signature に対して，予測期間ごとのイベント検知数を取得し，ベイズ推測を適用する．実装した各予測アルゴリズムのパラメータを示す．

4.2 実装条件

提案予測アルゴリズムに対して以下のような値を設定して予測計算を行う．

- 予測アルゴリズム 1
  曜日による週間的な周期に着目し，日曜日を起点として，T 日の曜日を事象 Y として与える．事象 Y を日曜から土曜までの 7 パターン設定．
- 予測アルゴリズム 2-1
  事象 Y を 2 パターン設定．
  - Y1：T 日のイベント検知数が前日のイベント検知数に対して増加．
  - Y2：T 日のイベント検知数が前日のイベント検知数に対して減少．
- 予測アルゴリズム 2-2

事象 Y を 4 パターン設定.
- Y1：T 日のイベント検知数が前日のイベント検知数に対して 2 倍以上増加.
- Y2：T 日のイベント検知数が前日のイベント検知数に対して 2 倍以内増加.
- Y3：T 日のイベント検知数が前日のイベント検知数に対して 2 倍以内減少.
- Y4：T 日のイベント検知数が前日のイベント検知数に対して 2 倍以上減少.

- 予測アルゴリズム 2-3

事象 Y を 8 パターン設定.
- Y1：T 日のイベント検知数が前日のイベント検知数に対して 10 倍以上増加.
- Y2：T 日のイベント検知数が前日のイベント検知数に対して 10 倍以内 5 倍以上増加.
- Y3：T 日のイベント検知数が前日のイベント検知数に対して 5 倍以内 2 倍以上増加.
- Y4：T 日のイベント検知数が前日のイベント検知数に対して 2 倍以内増加.
- Y5：T 日のイベント検知数が前日のイベント検知数に対して 2 倍以内減少.
- Y6：T 日のイベント検知数が前日のイベント検知数に対して 5 倍以内 2 倍以上減少.
- Y7：T 日のイベント検知数が前日のイベント検知数に対して 10 倍以内 5 倍以上減少.
- Y8：T 日のイベント検知数が前日のイベント検知数に対して 10 倍以上減少.

4.3 予測計算

4.3.1 予測アルゴリズム 1

予測アルゴリズム 1 では，条件 Y が日曜日から土曜日までの 7 パターンとなるため，式 (1) における P(Y) は 1/7 となる．P(X) は全日数分の前日に比べてイベント検知数が増加した日数から計算でき，Y1〜Y7 それぞれにおいて，条件が成り立つ全日数分の，その中で前日に比べてイベント検知数が増加した日数から $P(X \mid Y)$ が求められる．

4.3.2 予測アルゴリズム 2

予測アルゴリズム 2-1 では，条件 Y は前日に比べてイベント検知数が増加した場合と減少した場合の 2 パターンとなるため，式 (1) における P(Y) はそれぞれ，全日数分の前日に比べてイベント検知数が増加した日数および減少した日数から計算できる．P(X) は全日数分の前日に比べてイベント検知数が増加した日数から計算でき，Y1，Y2 それぞれにおいて，条件が成り立つ全日数分の，その中で前日に比べてイベント検知数が増加した日数から $P(X \mid Y)$ が求められる．予測アルゴリズム 2-2，2-3 の場合も，条件 Y のパターン数が 4 パターン，8 パターンで，予測アルゴリズム 2-1 と同様の計算過程になる．

図 4 予測結果の実装画面
Fig. 4 Screen of forecast result.

4.4 実装画面

図 4 に実装予測結果画面を示す．図 4 は Attack Signature のうち TCP port scan について，9 月 1 日以降次の日のイベント頻度が増加する確率を計算した結果である．1 列目は Attack Signature 名，2 列目は予測期間，3 列目はイベント頻度，4 列目は次の予測期間にイベント頻度が増加する確率，最後の列は予測の結果を示す．たとえば増加すると予測して，実際に次の日のイベント頻度が前日よりも増加している場合予測正解となり，正解を示す “○” が記される．また，減少すると予測して，次の日のイベント頻度が前日よりも増加している場合予測外れとし，不正解を示す “×” が記される．ここで本論文での増減判断の閾値は 50%としている．増加確率が 50%以上の場合は増加すると判定し，増加確率が 50%未満の場合は減少すると判定する．

5. 評価

本章では，評価条件を述べ，予測期間を 1 日と 1 時間にした場合のイベント頻度の増減に関する判定の正解率を評価する．

5.1 評価条件

評価に用いる IDS イベントは 2003 年 7 月 1 日から 11 月 30 日の期間にネットワーク上で実運用されている IDS から収集したログである．その中で，今回の評価対象になる 4 種類のイベント数の合計は約 312 万であり，1 日あたり平均 5 千となる．IDS は一般に広く利用されている市販のネットワーク型 IDS [12] を用い，学術ネットワーク内のある 1 つのクラス C 規模のネットワークゲートウェイに設置する．7 月 1 日から 8 月

表 1 予測期間を 1 日とした場合の正解率
Table 1 Hitting rate of forecast by the day.

	Signature	正解率			
		予測 1	予測 2-1	予測 2-2	予測 2-3
A	HTTP port probe	**70.00%**	47.78%	54.44%	54.44%
B	TCP port scan	52.22%	63.33%	**70.00%**	65.56%
C	UDP port probe	50.00%	**67.78%**	66.67%	62.22%
D	FTP PORT restricted	48.89%	66.67%	**67.78%**	63.33%

表 2 予測期間を 1 時間とした場合の正解率
Table 2 Hitting rate of forecast by the hour.

	Signature	正解率			
		予測 1	予測 2-1	予測 2-2	予測 2-3
A	HTTP port probe	58.29%	**60.32%**	56.99%	58.33%
B	TCP port scan	53.75%	**66.71%**	64.58%	64.44%
C	UDP port probe	51.90%	**89.54%**	85.46%	85.46%
D	FTP PORT restricted	45.69%	**76.30%**	66.11%	65.93%

31 日までの単位時間ごとのイベント頻度の増減を学習し，9 月 1 日から 1 つ未来の単位時間のイベント頻度の増減の予測を行う．9 月 1 日以降は，単位時間経過するごとにイベント頻度の増減が過去のデータに加わる．増加確率の閾値を 50%とし，予測の正解率を求める．また，評価に用いる Attack Signature は，本ネットワーク上で提供している重要なサービスに関係するものや DoS 攻撃に利用される Attack Signature を選択する．

5.2 予測期間を 1 日とする場合の予測正解率の評価

表 1 に予測期間を 1 日とする場合の予測正解率を示す．表 1 において，Attack Signature “A”（HTTP port probe）は週間的な周期がみられた Attack Signature である．“B”（TCP port scan），“C”（UDP port probe），“D”（FTP PORT restricted）は週間的な周期性を持っておらず，イベント頻度の増減が任意に変動しているようにみえる Attack Signature である．それらの中で “B”，“D” はイベント頻度の増減の変動の幅が大きく，“C” は小さい．表 1 の太字で示された正解率が，各 Attack Signature における最高正解率を示す．表 1 の “A” に着目すると，予測アルゴリズム 1 の正解率が高くなっている．予測アルゴリズム 1 は周期に注目したアルゴリズムであり，実際に周期性のある Attack Signature の予測に有効であることが分かる．また “B”，“C”，“D” を比較すると，“B”，“D” は予測アルゴリズム 2-2 の正解率が一番高く，“C” は予測アルゴリズム 2-1 が一番高い．これは，増減の変動の幅が日によって大きく変化する場合は，増減の度合いの場合分けを多くした方が，学習データがより正確に分類され，正解率が向上しているためである．一方，予測アルゴリズム 2-3 の正解率が最も高い Attack Signature はない．これは，増減の度合いの場合分けを多くしすぎると，それぞれの場合分けの範囲の発生頻度が下がってしまうので，学習データに適さなくなるためである．

5.3 予測期間を 1 時間とする場合の予測正解率の評価

表 2 に予測期間を 1 時間とする場合の予測正解率を示す．表 2 は，予測の期間を 1 日から 1 時間に変えて，1 時間後のイベント頻度の増減の予測を行った結果である．表 2 の Attack Signature は，予測期間の違いによる正解率を比較するために表 1 と同じにする．また，1 時間ごとの予測アルゴリズム 1 の条件は，0 時を起点として，時刻 T を事象 Y として与える．事象 Y を 0 時から 23 時までの 24 パターンを設定する．表 2 の予測アルゴリズム 1 についての正解率をみると，すべての Attack Signature において 40〜50%台の正解率へと低くなっている．これは，Source IP に海外の IP アドレスも多く含まれていることから，日本国内に対する時間帯の異なる海外からの攻撃の影響により，“A” から “D” の Attack Signature に時間による周期がみられなくなるためである．また，すべての Attack Signature において，予測アルゴリズム 2-1 の正解率が最も高いという結果になっている．時間単位での予測についても，増減の度合いの場合分けを多くしすぎると，それぞれの場合分けの範囲の発生頻度が下がってしまうので，学習データに適さなくなるためと考えられる．

5.4 最適アルゴリズム選択方式

評価の結果，攻撃ごとにイベントの増減特性が異な

表 4 128 個のアドレス空間の予測正解率
Table 4 Hitting rate of the 128-bit address space.

	Signature	正解率			
		予測 1	予測 2-1	予測 2-2	予測 2-3
A	HTTP port probe	75.28%	50.56%	53.93%	58.49%
B	TCP port scan	51.69%	49.44%	57.30%	52.63%
C	UDP port probe	44.94%	66.29%	68.54%	65.38%
D	FTP PORT restricted	–	–	–	–

表 5 64 個のアドレス空間の予測正解率
Table 5 Hitting rate of the 64-bit address space.

	Signature	正解率			
		予測 1	予測 2-1	予測 2-2	予測 2-3
A	HTTP port probe	64.04%	58.43%	62.92%	56.52%
B	TCP port scan	58.43%	55.06%	62.92%	58.90%
C	UDP port probe	42.70%	65.17%	60.67%	59.70%
D	FTP PORT restricted	–	–	–	–

表 3 最適アルゴリズム選択予測の正解率
Table 3 Hitting rate of forecast by selecting most appropriate algorithm.

	Signature	正解率
A	HTTP port probe	68.24%
B	TCP port scan	77.65%
C	UDP port probe	67.86%
D	FTP PORT restricted	67.78%

ることが判明した．ここで，イベントの増減特性も日々変化していることが考えられる．そこで，あるイベントの増減確率を算出するときに，その時点から過去の統計で最も正解率が高い予測アルゴリズムを自動的に選択して 1 つ未来の単位時間を予測する手法について提案し，表 1 と同じデータを用いて評価する．**表 3** に単位時間を 1 日にしたときの最適アルゴリズム選択手法による予測の正解率を示す．表 1 と比較して，Attack Signature “A” に関しては正解率がわずかに下がってしまったが，“B” に関しては期間によって予測アルゴリズム 2-1，2-2，2-3 の間で最適アルゴリズムの移動があり，その結果 1 割程度正解率が改善された．また，“C”，“D” に関しては，最適アルゴリズムの移動はほとんどみられず，表 1 と表 3 で正解率はほとんど変化しなかった．したがって，攻撃によっては最適な予測アルゴリズムが変動しており，自動的に最適な予測アルゴリズムを選択すると良い結果が得られることが分かった．

5.5 ネットワーク規模を変化させた場合の予測正解率の評価

本論文の評価に用いた 256 個の IP アドレス空間に対する攻撃ログの中から，先頭から 1/2 の 128 個，先頭から 1/4 の 64 個の IP アドレス空間に対する攻撃ログを用いて予測正解率に関する評価を行う．このときログのサイズは，元のサイズの約 1/2，1/4 である．予測期間を 1 日としたときの，128 個のアドレス空間の予測正解率を**表 4** に，64 個のアドレス空間の予測正解率を**表 5** に示す．ただし，FTP PORT restricted については，ホストアドレス空間を小さくすることによってイベント検知数が極端に少なくなってしまい，評価データとして適さないため，省略する．表 4，表 5 の結果を，256 個の IP アドレス空間の場合の予測正解率である表 1 と比較すると，アルゴリズムや Attack Signature によって正解率が高くなったり低くなったりといった変化はみられるものの，IP アドレス空間が小さくなることによって，正解率が顕著に下がったり上がったりする変化は見られないことが分かる．これは，ネットワーク規模を変化させてもイベント検知数の増減の特徴は変化しないため，学習に大きな差異が現れないためと考えられる．したがって，本予測アルゴリズムは，ネットワーク規模にほとんど依存していないことが分かる．

5.6 MS Blaster に関する予測実験

実験に用いたデータセットには MS Blaster が大流行した 2003 年 8 月中旬から下旬が含まれており，MS Blaster に焦点を当てて予測実験を行うことは本研究の効果に対して重要であると思われる．MS Blaster は TCP135 ポート（Microsoft RPC）に対して攻撃を行うことから，MS Blaster の攻撃対象として該当すると思われる “MSRPC TCP port probe” という Attack Signature に注目して，2003 年 8 月 1 日から 31 日にかけて予測を行う．このとき，予測開始時点で

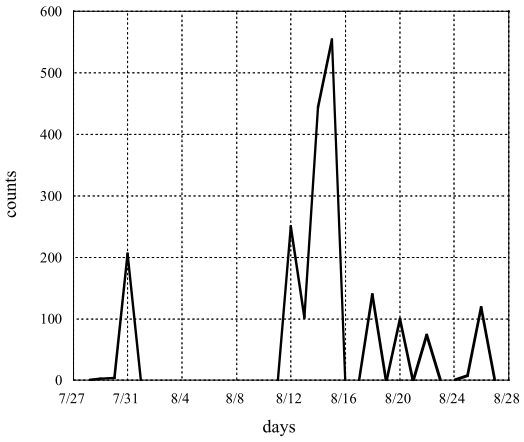

図 5 MSRPC TCP port probe のイベント検知数
Fig. 5 Event counts of MSRPC TCP port probe.

の学習期間は 7 月 1 日から 31 日までの 1 カ月間となる．評価の結果，予測アルゴリズム 2-1 で 83.33%の正解率となった．7 月から 8 月にかけての “MSRPC TCP port probe” のイベント検知数のグラフを**図 5** に示す．図 5 より，7 月末に予兆的な検知があった以外はしばらく検知されず，MS Blaster が発生したとされている 8 月 12 日以降から数日にわたって検知され始め，8 月末にはまたほとんど検知されない状態に戻っていることが分かる．Microsoft によって脆弱性が発表されたのが 7 月 17 日であり，7 月 25 日にはこの脆弱性を利用したエクスプロイトコードがインターネット上で公表されているので，図 5 にみられる 7 月末のイベント検知は，MS Blaster の原型の予兆である可能性が高いと考えらる．この場合 8 月中旬の攻撃発生を予測できることが重要となるが，実際の評価結果はイベント検知数が急に増加する箇所を予測できていない．これは，本研究における予測アルゴリズムは日々のイベント検知数の増減の特徴を学習して予測するものなので，1 カ所に設置した IDS からのみの学習データではあまりに少なく，予測のための十分な学習が困難であるためと考えられる．今後の課題として，多点に設置した IDS からのログを学習して，ネットワーク全体を広域監視することでより効率の良い予測を行うことを検討する．

5.7 考　　察

攻撃の周期と変動の大きさに注目して 2 種類 4 パターンの予測アルゴリズムを提案し，攻撃ごとにいくつかの予測アルゴリズムで適切にイベントの増減を予測できるようになった（要件 1 の達成）．また，ベイズ推測を用いた予測は，運用者にとって計算過程が理解しやすく，IDS から出力されたイベントを直接計算に用いることができるという利点がある（要件 2 の達成）．さらに，攻撃ごとに最適な予測アルゴリズムを自動的に選択する手法について検討し，高い正解率を達成できることが分かった（要件 3 の達成）．

本研究の予測結果である 67〜78%という数字の妥当性を検討するために，一般に広く知られている○か×かの 1 ビット予測の代表的なものとして気象庁による天気予報の結果と比較する．気象庁の予報による明日の降水の有無の適中率の例年値[13] は，地方別の年平均が 77%〜85%となっており，月平均では最も低い場合で 71%の的中率となっている．このことから，本研究の予測の正解率は，一般に広く受け入れられている気象庁の予測結果に近い値となっており，運用者の信頼をある程度得ることができるレベルと考えられる．

6. 結　　論

本研究では，IDS から出力されるイベントについて，攻撃の周期や増減度の推移に注目した 2 種類 4 パターンの不正侵入イベント増減予測アルゴリズムを提案した．提案アルゴリズムでは予測計算にベイズ推測を用い，あるイベントの過去の発生頻度の推移を単位時間ごとに学習して，1 つ未来の単位時間の増減確率を算出した．実装した予測システムについて，実運用されている IDS イベントを用いて，未来のイベント頻度が増加する確率を計算し，その正解率について評価を行った．その結果，攻撃ごとにイベントの増減特性が異なることが判明し，そこで，過去の統計で最も正解率が高い予測アルゴリズムを自動的に選択して 1 つ未来の単位時間を予測する手法について検討し再評価を行った結果，期間によって最適アルゴリズムが変動する攻撃があり，その場合 1 割程度正解率を改善できることが分かった．現段階ではイベントの増加と減少の 1 ビット情報を予測しているが，どの程度の数が増減するかという情報が重要になると考えられる．学習の時点で，単純に増減のみではなく，増減の程度をより精細に学習することで，今後のイベント数についてより具体的な予測が可能になる．効果的な運用のためにも増減の程度の予測については今後の課題とする．

本予測アルゴリズムにより，様々な攻撃の特徴に合わせたイベント頻度の推移を予測することができるようになり，安全なネットワーク運用に資することが期待される．事前に未来の攻撃の増減が分かれば，場合によってはトラヒック制御のためのフィルタリングを行うことで，正常なアクセスを円滑にすることができる．また事後として，予測と異なり大きく増加した場合に異常と判断することができ，未知攻撃の検知に寄

与できる．

謝辞 本研究の一部は慶應義塾大学 COE「アクセス網高度化光電子デバイス技術」プログラム，および KDDI との共同研究によって行われた．関係者に深謝する．

参 考 文 献

1) 沢田篤史，高倉弘喜，岡部寿男：開放型大規模ネットワークのための IDS ログ監視支援システム，情報処理学会論文誌，Vol.44, No.8, pp.1861–1871 (2003).
2) 高田哲司，小池英樹：見えログ：情報視覚化とテキストマイニングを用いたログ情報ブラウザ，情報処理学会論文誌，Vol.41, No.12, pp.3265–3275 (2000).
3) Kemmerer, R.A. and Vigna, G.: Intrusion detection: A brief history and overview, *Computer*, Vol.35, Issue4, pp.27–30 (2002).
4) Manikopoulos, C. and Papavassiliou, S.: Network intrusion and fault detection: A statistical anomaly approach, *Communications Magazine*, Vol.40, No.10, pp.76–82, IEEE (2002).
5) 竹森敬祐，三宅 優，中尾康二，菅谷史昭，笹瀬 巌：セキュリティデバイスログ分析支援システムの広域監視への適用，*CSS2003*, pp.397–402 (Oct. 2003).
6) Burroughs, D.J., Wilson, L.F. and Cybenko, G.V.: Analysis of Distributed Intrusion Detection Systems Using Bayesian Methods, *Proc. IEEE International Performance Computing and Communications Conference*, (Apr. 2002).
7) 石黒正揮，鈴木裕信，村瀬一郎，大野浩之：ベイズ推定に基づくインターネット攻撃検知システムの開発，*SCIS2004* (Jan. 2004).
8) Cho, S.-B.: Incorporating Soft Computing Techniques Into A Probabilistic Intrusion Detection System, IEEE Trans. Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol.32, No.2, pp.154–160 (2002).
9) 川本大輔，柱尾正敬，川瀬徹也，佐藤玲司，笹瀬 巌：2 方向の角速度による等速・等角速度予測を行うカルマンフィルタ，電子情報通信学会論文誌，Vol.J86-B, No.2, pp.154–160 (2003).
10) 蓮井亮二，白石善明，森井昌克：イベント依存モデルを用いた被害予測システムの実装，*CSEC2004*, pp.91–96 (July 2004).
11) Pikoulas, J., Buchanan, W.J., Mannion, M. and Triantafyllopoulos, K.: An Agent-based Bayesian Forecasting Model for Enhanced Network, *Engineering of Computer Based Systems* (*ECBS*) *2001, 8th Annual IEEE International Conference and Workshop*, pp.247–254 (Apr. 2001).
12) ネットワーク型 IDS．http://www.iss.net/
13) 気象庁．http://www.data.kishou.go.jp/yohou/kensho/reinen.html

(平成 16 年 9 月 3 日受付)
(平成 17 年 9 月 2 日採録)

石田 千枝（学生会員）
平成 16 年慶應義塾大学理工学部情報工学科卒業．現在，同大学大学院修士課程在学中．主として，インターネットセキュリティに関する研究に従事．電子情報通信学会会員．

島田 英一
平成 15 年慶應義塾大学理工学部情報工学科卒業．平成 17 年同大学大学院修士課程修了．同年（株）NTT データ入社．主として，通信ネットワークおよびインターネットセキュリティに関する研究に従事．

荒川 豊（学生会員）
平成 13 年慶應義塾大学理工学部情報工学科卒業．平成 15 年同大学大学院修士課程修了．現在，同大学院博士課程在学中．主として，通信ネットワークおよびインターネットセキュリティに関する研究に従事．IEEE，電子情報通信学会各会員．

竹森 敬祐（正会員）
平成 6 年慶應義塾大学理工学部電気工学科卒業．平成 8 年同大学大学院修士課程修了．同年 KDD（株）入社．平成 16 年同大学院博士課程修了．現在，（株）KDDI 研究所．主として，通信ネットワークおよびインターネットセキュリティに関する研究に従事．平成 14 年度本学術奨励賞受賞．電子情報通信学会会員．

笹瀬　巌（正会員）

昭和 54 年慶應義塾大学工学部電気工学科卒業．昭和 59 年同大学大学院博士課程修了，工学博士．同年カナダオタワ大学工学部電気・ポストドクトラルフェロー，昭和 60 年同大学講師，昭和 61 年慶應義塾大学理工学部電気工学科助手，昭和 63 年同大学専任講師，平成 4 年同大学助教授，平成 11 年同大学理工学部情報工学科教授，現在に至る．主として，ディジタル通信，通信ネットワーク，光通信理論，マイクロ波通信，非線形通信システム，通信理論，符号理論に関する研究に従事し，これまで原著論文 221 編，国際会議論文 322 編を発表．昭和 59 年度 IEEE ComSoc 学生論文賞，昭和 62 年度井上研究奨励賞，昭和 63 年度安藤博記学術奨励賞，昭和 63 年度篠原記念学術奨励賞，平成 8 年度電子情報通信学会交換システム研究会優秀論文賞受賞．現在，IEEE Communications Society Asia Pacific Board Director，電子情報通信学会通信ソサイエティ副会長，ネットワークシステム研究会委員長，情報理論とその応用学会評議員，総務省情報通信技術審議会専門委員，新エネルギー・産業技術総合開発機構電子・情報技術審議会委員等を務める．IEEE Senior Member，電子情報通信学会，情報理論とその応用学会各会員．