

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

Development of Internet Attack Detection System based on Bayesian Estimation

石黒 正揮* 鈴木 裕信† 村瀬 一郎* 大野 浩之 ‡
Masaki Ishiguro Hironobu Suzuki Ichiro Murase Huroyuki Ohno

あらまし インターネット上の特定の IP アドレスにおいて観測されるポートスキャンログから、ベイズ推定に基づき広域的なネットワーク攻撃の活発化によるインターネットの危険状態を検知する手法を提案する。本手法では、ポートスキャン頻度の時系列データをトレンドからの差として観測し、危険状態に関する確信度の更新を繰り返すことにより推定精度を高める。これにより自動的に検知された攻撃は、その危険度に応じた警報として、ポートスキャン頻度系列グラフとともに、サーバー管理者などに通知する。攻撃状態は時間とともに動的に変化するため、未知母数自体を確率的に変動する量としてとらえるベイズ推定に適している。

キーワード ネットワーク攻撃、ポートスキャン、攻撃検知、ベイズ推定、ROC 分析

1 はじめに

近年、インターネット上のサーバーに対する攻撃による大規模な被害が相次いでいる。サーバーソフトウェアの脆弱性に対して、インターネット上でのエクスプロイト(攻撃法)や攻撃ツールの流通速度が速まったことにより、十分な対策がなされないまま攻撃を受けることで被害の拡大につながる事例が増加している。侵入検知システム(Intrusion Detection System(IDS))による対策は試みられているが[1]、IDSは、自サイトに侵入された形跡をもとに不正アクセスを検出するものであるため、先行対応への有効な手段には至っていない。一方、CERT Advisory, CVE等の脆弱性情報データベースに基づくソフトウェアの不具合対応の意識は高まっているが、脆弱性情報が公開された後、時間が経過してからの対策では被害の拡大を食し止めるには不十分であると見なされている。

本研究では、インターネット上の特定の IP アドレスにおいて観測されるポートスキャンデータを分析するこ

とにより、自サイトがネットワーク攻撃を受ける前に、インターネット上の攻撃の活発化による危険性を自動的に検知し、脆弱性情報等の公開や攻撃ツールの流通に先駆けて、サーバ管理者に警告通知を行うシステムを提案する。これによりサーバの先行対策による被害発生および拡大の防止に役立てることができる。

2 攻撃検知手法

特定の IP アドレスにおいて観測されるポートスキャン頻度の時系列データから、インターネット上の広域的な攻撃活動による危険状態の検出を行う手法を示す。ネットワーク状態の危険度は、時間と共に変化するため、ポートスキャンの観測に対して、ベイズ推定に基づき、危険度推定値を繰り返し更新することにより、危険度の動的な変動に対応した推定を行う。

ベイズ推定は、観測をもとに、推定する状態に対する確信度の更新を繰り返す。本手法では、ポートスキャン頻度をそのトレンド(移動平均)からの差として観測することにより、ネットワークの危険状態を推定する。我々が推定したいネットワークの危険状態を、“インターネット上の広域的な攻撃活動の活発化により、自サイトへの攻撃による被害が発生する可能性の高い状態”と定義する。

危険度の推定を行う対象時刻を t (図 1 中、推定時刻と表示) とする。時刻 t からそれ以前の一定期間 T (確信度更新区間と呼ぶ) のポートスキャン頻度および、その区間の各時刻における移動平均(トレンドと呼ぶ)、および、

* 株式会社三菱総合研究所 情報技術研究部, 〒 100-8141 東京都千代田区大手町二丁目 3-6, Information Technologies Research Dept. Mitsubishi Research Institute Inc., 3-6, Otemachi 2-Chome, Chiyoda-ward, Tokyo 100-8141

† 鈴木裕信事務所, 〒 151-0051 東京都渋谷区千駄ヶ谷 1-28-8-006, Hironobu SUZUKI Office, 1-28-8-006 Sendagaya, Shibuya, Tokyo 151-0051

‡ 独立行政法人通信総合研究所 情報通信部門非常時通信グループ, 〒 184-8795 東京都小金井市貫井北町 4-2-1, Emergency Communications Group, Information and Network Systems Division, Communications Research Laboratory, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795

各時刻の一定区間のポートスキャン頻度のトレンドとの差の標準偏差(局所標準偏差と呼ぶ)を観測し、各時刻において、1つ前の時刻の危険度に対する確信度(事前確率)を用いて、式(1)に基づきベイズ更新を行い、次の時刻の危険度(事後確率)を推定する。確信度更新区間 T において式(1)に基づくベイズ更新を繰り返すことで、時刻 t における危険度を推定する。

$$P(s_i|r) = \frac{P(s_i)P(r|s_i)}{\sum_j P(s_j)P(r|s_j)} \quad (1)$$

ここで、 $s_i (i = 0, 1)$ はネットワーク攻撃に危険状態の有無を表す。

$$\begin{cases} s_0 & : \text{危険状態} \\ s_1 & : \text{安全状態} \end{cases} \quad (2)$$

r はポートスキャン頻度のトレンドからの差を表す観測値、 $P(s_i|r)$ は、 r を観測した時に状態が s_i であると推定される確率を表す事後確率、 $P(s_i)$ は、観測前に状態が s_i である確率を表す事前確率、 $P(r|s_j)$ は、状態が s_j であるときに r が観測される尤度を表す。尤度関数 $P(r|s_j)$ は、危険状態であるときにネットワーク攻撃につながるポートスキャンが活発化することから設定することができる。本モデルにおいては以下のように定義する。

$$P(r|s_0) = \frac{r}{k\sigma_r + r} \quad (3)$$

$$P(r|s_1) = \frac{k\sigma_r}{k\sigma_r + r} \quad (4)$$

r は、ポートスキャン頻度のトレンドからの差、 σ_r 観測地点での、一定区間のポートスキャン頻度のトレンドからの差から求めた局所的な標準偏差、 k は、ベイズ更新における更新のスピードを定めるパラメータ(ベイズ更新偏差係数と呼ぶ)である。式(3)および式(4)は、0から1の区間の実数値をとり、ポートスキャン頻度のトレンドから差が正の方向に大きいほど、危険状態であることを意味する。ベイズ推定に基づく危険度推定の計算手順は、図3のようになる。

図2は、あるポートに対するスキャン頻度時系列データである図1に対して、各時刻におけるベイズ推定による攻撃状態に対する確信度の推移を示している。

3 広域攻撃予測システムの概要

広域攻撃予測システムは、インターネット上のいくつかのIPアドレスに対するポートスキャン動向を観測することにより、インターネット上の攻撃の活発化による広域的な危険性をいち早く検知し、自サイトでの被害が発生する前に、警告を通知するシステムである。本システ

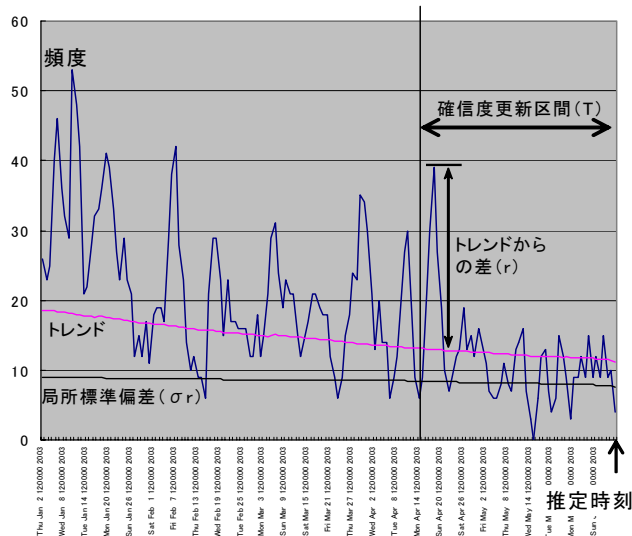


図1: ベイズ推定における観測データとパラメータ

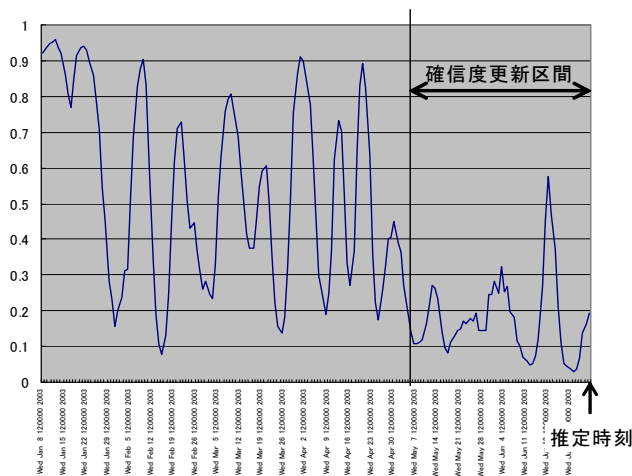


図2: ベイズ更新区間と各時刻のベイズ推定履歴

ムは、ポートスキャン頻度の時系列データをもとに、ベイズ推定に基づきインターネットの危険状態を検出する。

図4は、広域攻撃検知の概略を示した図である。全体システムは、ポートスキャンを検知するセンサボックス、センサボックスで検知されたデータを収集管理するログサーバ、危険度の推定を行う解析予測サーバ、解析結果をWEB上で表示あるいは携帯電話、メールなどに通知する通知システムから構成される。センサボックスは、Linux syslog を用いて取得したTCP/UDPポートアクセスを clscan データ標準形式 [4] で記録し、ログサーバに送信する。ログサーバは、複数のセンサボックスからの clscan データを収集し、解析予測サーバの要求に応じてデータを提供する。解析予測サーバは、第2章に示

1. ポートスキャン頻度時系列から、トレンドを求める。
2. 確信度更新区間の初期時刻 T_0 における危険度事前確率の初期値を設定する。
3. 確信度更新区間の時刻 t の事前確率および、トレンドに基づく観測値 (r) をもとに、ベイズ更新式 (1) に従い、時刻 $t + 1$ の危険度 (事後確率) を求める。
4. 確信度更新区間の最終時刻 T_f まで、ステップ 3 のベイズ更新を繰り返す。

図 3: 危険度推定の計算手順

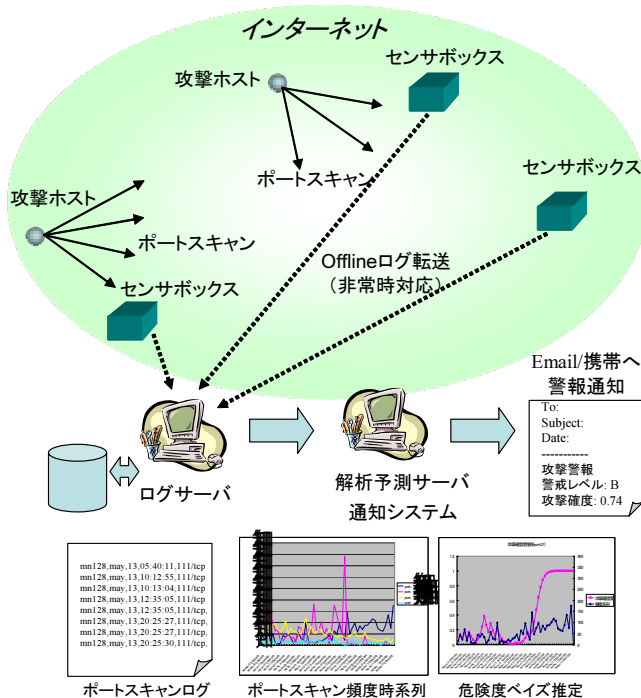


図 4: 広域攻撃検知の概要

した手法に基づき、危険状態を検出し、通知システムに送る。通知システムは、現在、WEB による予測結果の表示、ポートスキャン頻度時系列データのグラフ表示、i-mode, ez web による携帯電話での表示を実現している。

4 評価

本攻撃検知手法を、信号検出理論の分野で有効なことが知られている [2]ROC 分析 (Receiver Operating Characteristics Analysis) により評価する。一般に、正事例、負事例の判別精度の評価には、正答率や相関を用いるこ

とができるが、これらの方法は、正事例と負事例の比率強く依存し、一方の比率が極めて高い場合には、無条件に比率の高い方を予測すれば、評価値が高くなる問題点がある。ROC 分析の場合、負事例を誤って正事例と判断する偽陽性率 (False-Positive Fraction) と正事例を正しく正事例と判断する真陽性率 (True-Positive Fraction) の両面を考慮した総合的な尺度による評価が可能になる。真陽性率は、検出の感度に相当するものであり、偽陽性率は、特異な事例の比率を表す特異度に対応するものである。ROC 分析は、検出の閾値や事例の分布に依存しないものである。

本手法は、インターネット上で広域攻撃がある程度活発化した危険状態を検出するものである。このような定義による真の危険状態は、一般に、認識不可能なものであり、性能評価を行う上での困難の原因となっている。この評価においては、ポートスキャンデータから、人が見て危険であると見なされる時刻を設定し、それを真の危険状態の近似と見なすことにより、自動検知結果との比較分析を行う。

図 5 は、2003 年に JPCERT における緊急報告で注意喚起をされた脆弱性に該当するポートを含む比較的ポートスキャン数の多いポートを選び、ポートスキャン頻度の時系列推移を表示したものである。

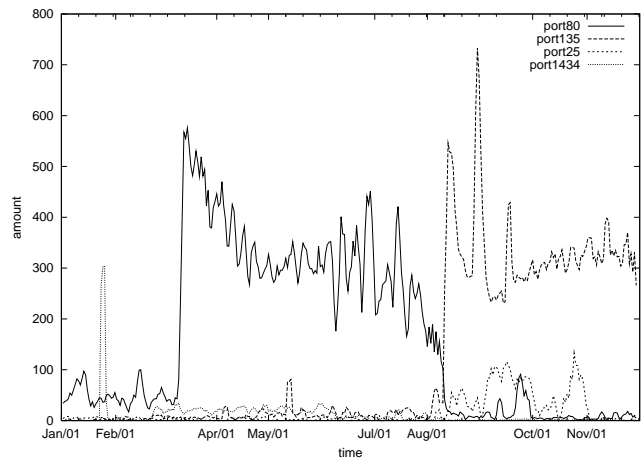


図 5: ポートスキャン頻度の時系列推移

ポート 80 は http サーバで使用され、2003 年 3 月 18 日に “Microsoft IIS 5.0 の脆弱性に関する注意喚起” (JPCERT-AT-2003-0003) のあったものである。ポート 135 は、Windows RPC サービスで使用されるもので、2003 年 8 月 15 日に、W32/Blaster ワームによって大規模な被害を生じた “TCP 135 番ポートへのスキャンの増加に関する注意喚起” (JPCERT-AT-2003-0006) のあったものである。ポート 25 は、メールサーバで使用されるもので、2003 年 3 月 31 日に “新たな sendmail の脆弱性に関する注意喚起” (JPCERT-AT-2003-0004) のあったものである。

ポート 1434 は、Microsoft SQL Server 2000 で使用されるもので、2003 年 1 月 27 日に “UDP 1434 番ポートへのスキャンの増加に関する注意喚起” (JPCERT-AT-2003-01-27) のあったものである。

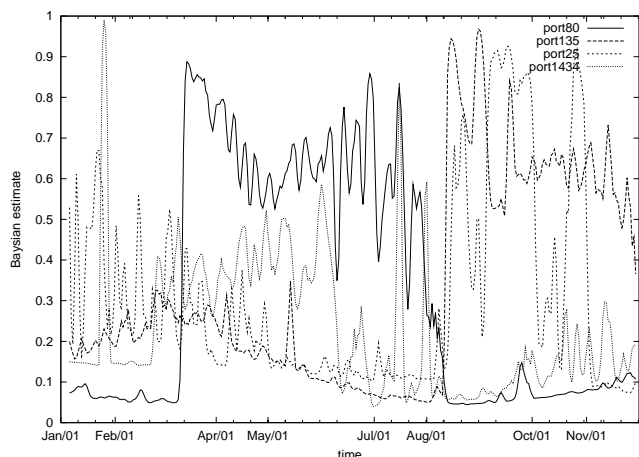


図 6: 危険状態推定値の時系列推移

図 6 は、図 5 のポートスキャン頻度に対して、各時刻における危険状態を推定したものである。

これらの結果のうち、データの対象期間である 2003 年において、CERT Advisory 等において、注意勧告のなされた脆弱性のうちの 1 つで、対象期間のポートスキャン頻度の時系列変動の複雑なポート 25(sntp) を対象として ROC 分析を行う。

分析対象データは、1 地点の IP アドレスにおいて 2003 年 1 月 1 日から 2003 年 12 月 1 日までの 11 ヶ月に観測された TCP/UDP ポートアクセスである。図 7 は、この期間における上記の危険状態 (図中 Positive cases) とそれ以外の安全状態 (Negative cases) に関して、攻撃検知によるベイズ推定値の分布を示したものである。

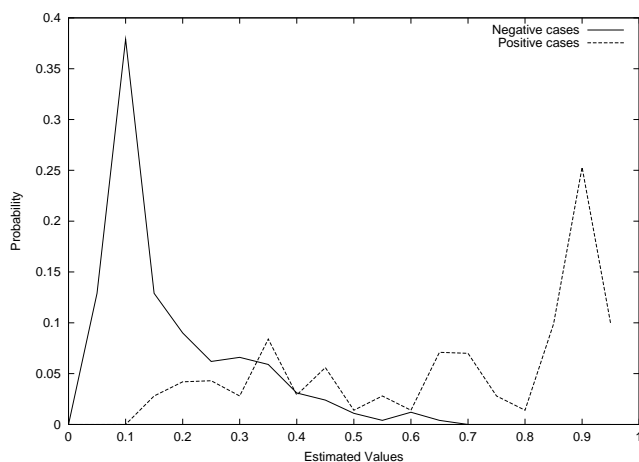


図 7: 危険状態の区分によるベイズ推定値の分布

このグラフより、危険状態におけるベイズ推定値の分布のピークは 0.9 ぐらいの高い値に位置し、安全状態におけるそれは 0.1 ぐらいの低い値に位置しており、両事例の判別指標として使えることを示している。

図 8 の横軸 (ベイズ推定値) の閾値に対して、危険状態と安全状態のそれぞれのグラフの閾値より右側の面積を求めると、True-Positive Fraction および False-Positive Fraction を求めることができる。ROC 曲線は、この閾値を変化させたときに得られる True-Positive Fraction および False-Positive Fraction を 2 次元上に分布させることにより描くことができる。

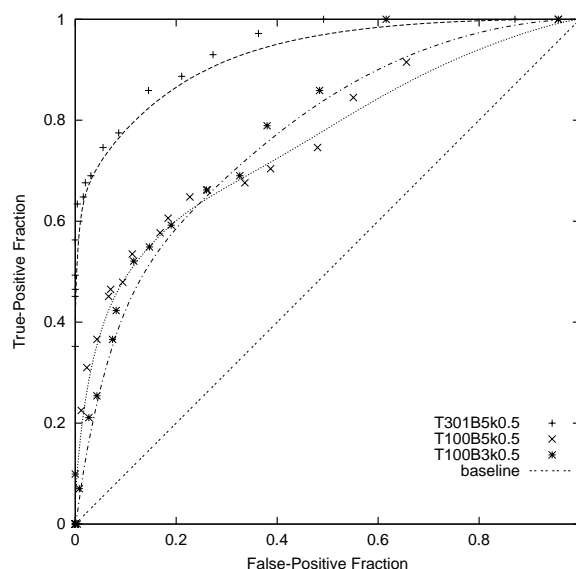


図 8: ポート 25 のベイズ推定に対する ROC 曲線

図 8 は、ポート 25 に関して、攻撃検知手法のいくつかのパラメータを変化させた時 (表 1)、上記と同様に True-Positive Fraction および False-Positive Fraction を求めることで得られる ROC 曲線を示している。

表 1: 攻撃検知手法の分析パラメータ値

図 8 凡例 ID	ベイズ更新 偏差係数	ベイズ更新 区間	トレンド 区間	Az 値
T301B5k0.5	0.5	5	301	0.95
T100B5k0.5	0.5	5	100	0.79
T100B3k0.5	0.5	3	100	0.80

ROC 曲線は、 $y = x$ 線上に位置し、上方に位置する程判別性能が高いことを示し、その性能評価は、ROC 曲線の下面積 (Az 値) によって行う。表 1 中の Az 値から、トレンド区間 301 日、ベイズ更新偏差係数 0.5、ベイズ更新区間 5 の場合の検知性能が良好であることが確認された。

5 まとめ

本研究では、インターネット上の特定の IP アドレスにおいて観測されるポートスキャンデータから、ベイズ推定に基づき広域的なネットワーク攻撃を検知するシステムを開発した。攻撃検知は、ポートスキャン頻度の時系列データをトレンドからのずれとして観測し、攻撃状態に関する確信度の更新を繰り返すことにより、危険状態を推定する。自動的に検知された攻撃は、その危険度に応じた警報として、ポートスキャン頻度系列グラフとともに、システム管理者などに通知される。ROC 分析により、危険状態の検出に有効なトレンド区間、ベイズ更新偏差係数、ベイズ更新区間などの検出パラメータを比較し、性能の高いパラメータを判定した。

参考文献

- [1] P. Porras and P. Neumann, “EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances”, In Proceedings of the Nineteenth National Computer Security Conference, October 1997.
- [2] Richard O. Duda et al., Pattern Classification, John Wiley & Sons, 2001
- [3] 石黒 正揮 他, “肝 X 線 CT 画像における診断特徴量に関する学習ルールを用いた腫瘍の良悪性判別手法”, 日本医用画像工学会論文誌 2001, 第 19 巻 1 号 pp.43-49
- [4] 鈴木裕信, clscan ホームページ, <http://www.pp-iiij4u.or.jp/h2np/h2np/lscan/>
- [5] 水友 仁史 他, “DDoS 攻撃予測システムの開発,” コンピュータセキュリティシンポジウム 2003, pp.97-102