5E-4

定点観測による不正アクセス分析システムの提案[1]

～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～

平井規郎[2]　鹿島理華[3]　東辰輔[4]　榊原裕之[5]　藤井誠司[6]　北澤繁樹[7]

三菱電機株式会社　情報技術総合研究所[8]

1. はじめに

近年インターネットを経由した不正アクセスが増加しているが、特にワームによる被害は深刻である。こうした不正アクセスによる被害を最小限に抑えるためには、できるだけ早い時点で攻撃を把握し、対策をたてる必要がある。ワームによる攻撃を検出する方法としては、ルールベースの相関分析による検出方法がよく知られているが、未知のワームを検出することは困難である。

そこで本稿では、ネットワーク上のトラフィックを観測し、センサーデータマイニングの分野で多く用いられる特異値分解(以下 SVD：Singular Value Decomposition)を用いた主成分分析を適用しその傾向の変化を捉えることで、早期に不正アクセスを検出する手法を用いたシステムの提案を行う。

2. 不正アクセス分析手法

2.1. ワームの特徴

ワームとしては 2003 年に発生した Blaster や 2004 年に発生した Sasser などがあるが、これらのワームに共通する特徴的な現象としてトラフィック量の異常増加が挙げられる[1]。これはワームが新たな感染先を検索するために大量のスキャンパケットを送信するためである。そこで、正常時においてネットワーク上を流れるパケットのトラフィック量などを記憶しておき、それからはずれた状態を捉えて異常を判定することにより、早い時点での不正アクセスの検出が可能になると思われる。

2.2. SVD による分析

SVD を用いた分析とは、相関関係にあるいくつかの要因を合成し、特徴量に変換して分析することである。

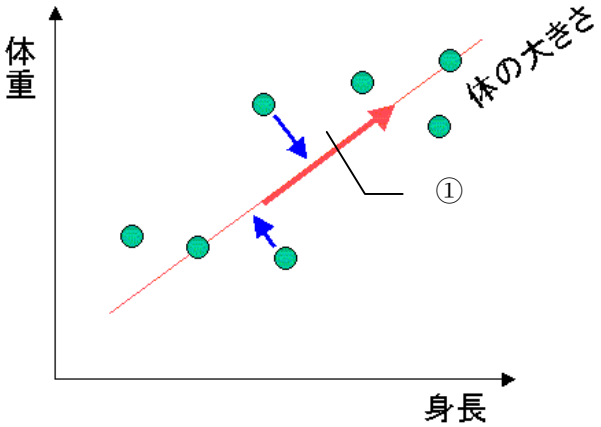

図 1：SVD の概念

図 1 において横軸は身長、縦軸は体重をあらわしている。身長と体重の関係を散布図に描くと図 1 のように右肩上がりの関係になり身長の増加によって体重も増加するという傾向にあることがわかる。この傾向を説明するためには新たに図 1 の線①で表した軸を考えればよい。このように変数を元のまま独立に扱うのではなく、「体の大きさ」のような総合的指標を導入することで、変数の特徴を容易に把握できるようにする手法が主成分分析であり SVD を用いて行うことができる[2]。

単位時間(たとえば 1 時間)あたりにネットワーク上を流れるパケットのトラフィック量から、ある一定の長さ(たとえば 12 時間)の変化を 1 単位時間ずつシフトしながら切り出し、それに対して SVD を適用し特徴量を抽出することで時系列データを容易に比較することが可能になる。

3. 提案システム

3.1. システム構成

不正アクセス分析システムの全体構成は[3]に紹介されており、この分析機能は異常検知機能に該当するものである。図 2 に異常検知機能の詳細機能概要を示し、各機能について説明する。

1 An Intrusion Detection System based on network stationary monitoring, – A network log analyzation method for detection worm attack. 2 Norio Hirai 3 Rika Kashima 4 Shinsuke Azuma 5 Hiroyuki Sakakibara 6 Seiji Fujii 7 Shigeki Kitazawa 8 MITSUBISHI ELECTRIC CORPORATION INFORMATION TECHNOLOGY R&D CENTER

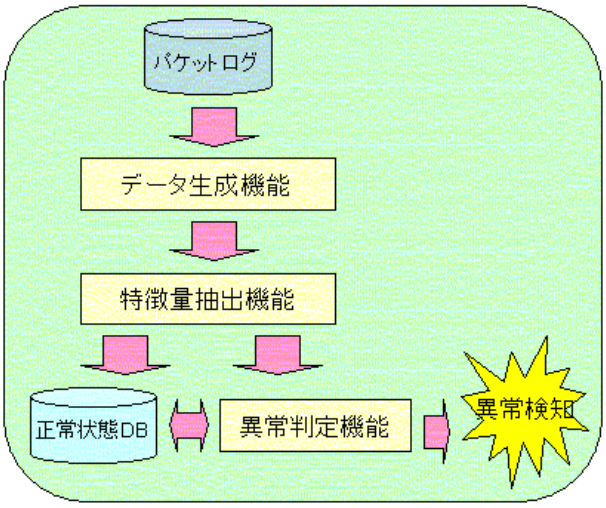

図 2:機能概要図

- **データ生成機能**
  定点観測装置により得られたパケットログを時系列データに変換し、一定の長さでシフトしながら切り出す機能。
- **特徴量抽出機能**
  データ生成機能により作成された行列に SVD を適用し、特徴量を抽出する機能。
- **異常判定機能**
  特徴量抽出機能により得られた新しいデータの特徴量を正常時データの特徴量 DB と比較し異常か否か判定する機能。

3.2. 処理の流れ

正常時に定点観測装置により収集されたパケットログデータは単位時間で集計され、時系列データに変換される。データ生成機能はこのデータから決められた期間で時間をシフトしながら時系列データを切り出すことにより長さの決まった時系列データの行列を作成する。次にこの正常状態のデータに SVD を適用することにより各時系列データの特徴量を正常状態の特徴量として学習する。新しく収集した監視対象の時系列データについても特徴量を抽出し、正常状態の特徴量と比較する。このときパケットログの時間的変化の傾向が正常状態と類似していれば、得られた特徴量も正常状態の特徴量の群と距離的に接近しており、異常状態であれば大きく離れる。図 3 では各時系列データから得られた特徴量が１つの点として表され、□が正常状態で得られた特徴量の群である。今監視対象のデータから得られた特徴量が●のように実線の楕円内に存在する場合には正常状態に近いため正常であると判定する。△のように正常状態から少し外れる場合には危険性が高いと判定し、×のようにさらに正常状態から外れた場合には異常であると判定する。この点線の範囲を実システム上でチューニングすることにより早期検出が可能になる。

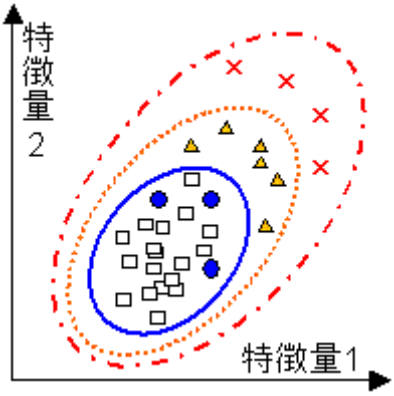

図 3：特徴量判定機能図

4. 評価

SVD によるワーム検出の妥当性を調査するため実際のトラフィック量のデータを用いて評価した。対象データとしては、インターネットに接続した Firewall で半年間にわたって採取した破棄パケットログを単位時間(1 時間)ごと、かつあて先ポートごとに集計した時系列データを用いた。この時系列データから長さ 12 時間で切り出したデータを作成し、これに SVD を適用し正常時と異常時の傾向を比較した。その結果 2004 年の 5 月に発生した Sasser ワームについて感染報告が発表された 5 月 1 日(日本時間)[1]より以前の 4 月 30 日に異常を検出することが可能であることがわかった。また Sasser 以外でトラフィック異常が報告された事例においても発表以前に異常を検出することが可能であることがわかった。以上の結果からワームの検出手法として SVD を用いることは妥当であると判断した。

5. まとめ

本稿では、SVD を使った主成分分析をネットワーク上の時系列データに適用することにより異常の早期検出が可能になることを示した。

今後は提案した手法をシステムに実装し、実際のデータを用いた評価を行うことにより、さらに提案した手法の有効性を検証していく。

参考文献

[1] @Police, “http:// www.cyberpolice.go.jp”
[2] 武藤眞介,“統計解析ハンドブック”,朝倉書店
[3] 榊原, 藤井, 北澤 他, “定点観測による不正アクセス分析システムの提案”, IPSJ 68 回全国大会予稿集