2F-4

定点観測による不正アクセス分析システム[1]

～不正アクセス検知のためのネットワークログ分析手法～

鹿島理華[2]　藤森敬悟[3]　平井規郎[4]　榊原裕之[5]　大野一広[6]　藤井誠司[7]

三菱電機株式会社　情報技術総合研究所[8]

1. はじめに

近年増加しているワーム、Dos 等のネットワーク経由の不正アクセスに対応するために、筆者らは、定点観測による不正アクセス分析システムにおける、ネットワークログ分析の開発に取り組んでいる。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検知を行う。

不正アクセス分析システムでは、早期に不正アクセスを検知し被害の拡大を未然に防ぐために、収集したデータをリアルタイムに分析する必要がある。そこで、本稿では、これに対応するために行った分析手法の改良について述べる。

2. 不正アクセス分析手法

2.1. 不正アクセスの SVD による分析

不正アクセスの一つにワームがある。代表的なワームとしては 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、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して起きるトラフィック量の異常増加が挙げられる[1]。そこで、実際のトラフィック量を用いて評価し、ワームの検出手法として SVD を用いることが妥当であること確認した。[2]

2.2. 不正アクセス分析システムでの不正検知

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。要因の数を次元とすると、要因の合成とは、SVD により次元を削減することである。

不正アクセス分析システムでは、ネットワーク上を流れるパケットのトラフィック量を単位時間（たとえば 1 時間）で集計し時系列データに変換する。次にこのデータから一定の期間（たとえば 12 時間）を、1 単位時間ずつシフトしながら切り出すことにより時系列データの行列を作成する。この行列に対して SVD を適用し特徴量を抽出する。

時系列データには定常状態であると既に判定されている期間があり、その特徴量が抽出されているものとする。そこで、新しいデータが追加されるとそれに対応する特徴量を抽出し、定常状態の特徴量と比較する。新しいデータの時間的変化の傾向が定常状態と類似していれば、得られた特徴量も定常状態の特徴量の群と接近しており、異常状態であれば大きく離れる。これを利用し、時系列データの傾向の変化の兆候を捉える。

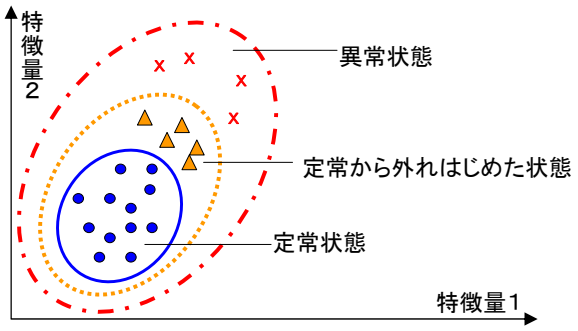


図 1：特徴量判定機能図

3. 分析手法の改良

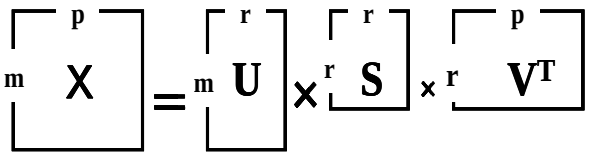
3.1. SVD による特徴量の抽出

SVD は図 2 に示す行列演算である。時系列データから切り出して作成したデータ行列が、行列 X に対応する。

新しいデータが収集されると、行列 X にもデータ（行）が追加される。一方、ネットワークの状態は常に変化しているので、最新のデータに対する過去のデータの影響を排除する必要がある。つまり、行列 X から古いデータに対応する行を削除する必要があり、SVD の行列演算を再

1 An Intrusion Detection System based on network stationary monitoring, – A network log analyzation method for detection unlawful accesses. 2 Rika Kashima 3 Keigo Fujimori 4 Norio Hirai 5 Hiroyuki Sakakibara 6 Kazuhiro Ono 7 Seiji Fujii 8 MITSUBISHI ELECTRIC CORPORATION INFORMATION TECHNOLOGY R&D CENTER

計算する必要がある。



r=ランク

図 2:行列演算 SVD

3.2. SVD のアルゴリズムの改良

不正アクセス分析システムでは、複数のポート毎にトラフィック量を監視する必要がある。これは、ポート毎にそのトラフィック量の変化の傾向が異なるためである。このため、SVD の行列演算は監視対象の個数分、同時に行う必要がある。

一方、SVD の計算量は、行列Ｘのデータ量に依存するため、行列Ｘの全体のデータ量に対し追加と削除が僅かであっても、行列Ｘ全体に対し再計算を行う必要があり、システムへの負荷が大きくなるという課題があった。

そこで、これを解決するために、更新された分だけ再計算するようにアルゴリズムの改良を行い、データの追加については追加分のみ、削除については削除分のみの計算で高速に更新することが出来るアルゴリズムを[3][4]により開発し、不正アクセス分析システムに適用した。

4. 評価

改良したアルゴリズムによる効果を評価するために、性能測定を実施した。行列Ｘに１行を追加、あるいは削除する場合に、行列Ｘ全体に対し SVD の再計算を行う従来の手法と、改良 SVD により追加分だけ再計算する手法、削除分だけ再計算する手法とを比較した。また、今回開発したアルゴリズムの処理性能は図２の r の大きさに依存するので、比較は、図２の p に比べて r がかなり小さい場合（p=120, r=10）（ケース①）と、r が p と同じ場合（p=119, r=119）（ケース②）で行った。比較結果を図３と図４に示す。グラフの x 軸は行列Ｘの行数、y 軸は処理時間を示す。

①の場合、改良 SVD による処理性能は追加・削除の場合共に従来の手法に比べ、22 倍から 35 倍の性能向上がみられた。一方、②の場合は、追加・削除共に、従来の手法と比べ約 1.5 倍の性能向上がみられた。

このことから、改良 SVD は従来の手法に比べ高速であり、r が p より小さいほどその効果が大きいことがいえる。

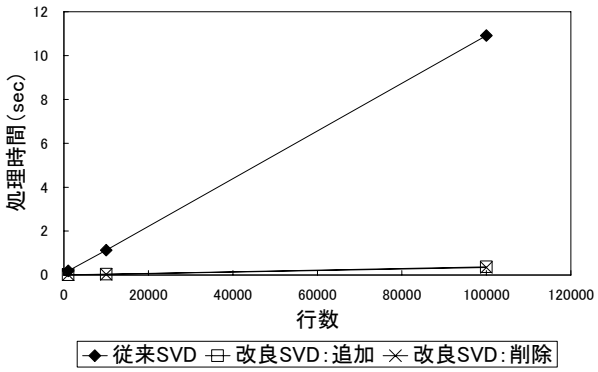


図 3:ケース①

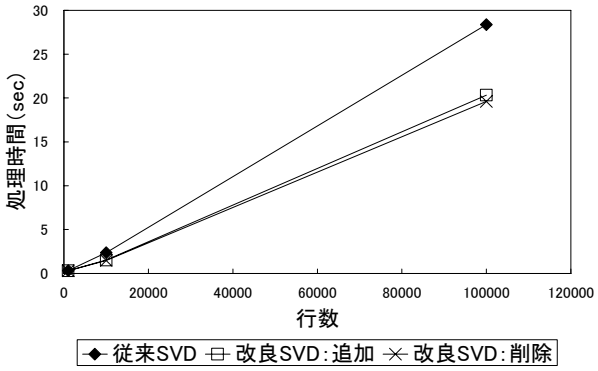


図 4: ケース②

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にあたり、リアルタイムに分析するために行ったアルゴリズムの改良が、性能面で効果があることを示した。

今後は今回開発したアルゴリズムを実装したシステムで、実際のデータを用いた評価を行うことにより、さらに有効性を検証していく。

参考文献

[1] @Police, “http:// www.cyberpolice.go.jp”
[2] 平井, 鹿島, 東 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
[3] M. Brand, ″Incremental Singular Value Decomposition of Uncertain Data with Missing Values″, European Conference on Computer Vision, May 2002.
[4] M. Brand, ″Fast Online SVD Revisions for Lightweight Recommender Systems″, [appendix A Low-rank modifications], SIAM International Conference on Data Mining, May 2003.