Anomaly 型不正アクセス分析システムの正常域データの変動への対応について

榊原裕之† 河内清人† 北澤繁樹† 藤井誠司†
†三菱電機株式会社 情報技術総合研究所

1. はじめに

筆者らは未知の不正アクセスを早期に検知するため，正常な状態のアクセス数（本稿では正常域データと呼ぶ）からの増加変動を主成分分析により発見する Anomaly 型の不正アクセス分析システムの開発に取り組んでいる[1][2].

本システムを運用し監視サービスを実現する場合，監視対象のネットワークにおいてシステム規模の拡大などにより正常なアクセス数が増加することがある．しかし，本システムの分析においては，増加する前の正常域データを使用して分析を継続すると正常なアクセスを検知する課題がある.

本稿では，正常域データの変動と課題について説明し対策について提案を行う.

2. 検知方式と正常域データの変動の課題

本システムにおける不正アクセスの検知方式と正常域データの関係について説明し，正常域データの変動と検知への影響の課題について述べる.

2.1. 検知方式と正常域データの関係

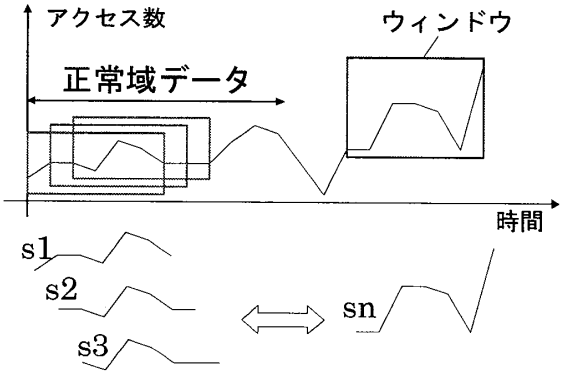

図 1 検知方式

本システムではアクセス数の正常な状態に対する増加変動を不正アクセスの兆候として検知する．図 1の正常域データに対しウィンドウをずらしながら切り出した s1,s2,s3・・・（ここではパターンと呼ぶ）に対して，最新のアクセス数を含む sn を比較し，類似を判定する．類似していなければ，sn は正常域データに含まれない異常なパターンとして判定する．類似の判定においては，主成分分析によりパターンの主成分得点を算出し，主成分得点間のマハラノビス距離を比較する．正常域データに含まれる各パターンの主成分得点と sn のそれの距離が閾値を超えた場合，類似していないと判定する[1].

最新のアクセス数が増加傾向でも減少傾向でもパターンとして正常域データに存在しなければ類似していないと判定するため，上記の非類似の判定に加え最新のアクセス数が正常域データの平均＋標準偏差を超えたことをもって増加を判定する.

2.2. 正常域データの変動

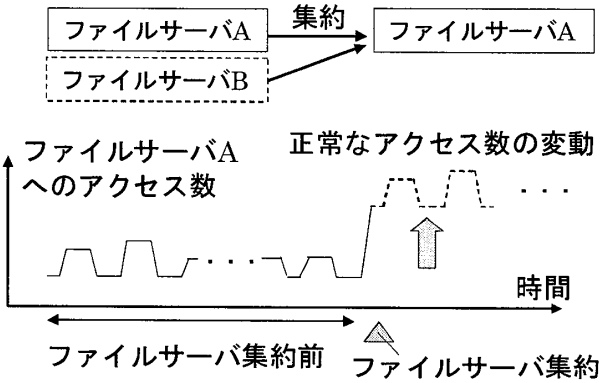

図 2 正常域データの変動の例

図 2にアクセス数の正常な変動の例について説明する．例えば，ファイルサーバ A へのアクセス数を監視する場合において，システム規模の拡大によりファイルサーバ A に別のファイルサーバ B が集約されると，ファイルサーバ A への正常なアクセス数が増える．これは正常なアクセス数の変動であり，本稿では正常域データの変動と呼ぶ.

2.3. 正常域データの変動の課題

正常域データの変動後に変動前の正常域データを使い分析を続けた場合，観測されるアクセス数は変動前の正常域データよりも大きいため増加変動として判定され検知が発生する．しかし，この検知は正常なデータを検知しているため誤りであり，変動後は変動した正常域データを基準にアクセス数の増加の分析を行うべきである.

これを実現するため，図 3において T の期間アクセス数を観測し新しい正常域データとして蓄積してから分析を再開する方法がある．しかし，例えば 1 ヶ月の正常域データを必要とする場合 T は 1 ヶ月となりその期間は分析を行えない．つまり，正常域データの変動後は，変動した正常域データを基準に分析を行う必要があるが，この正常域データの速やかな準備と分析の再開が課題である.

On Measures for Variation of Normal Condition Data in Anomaly Intrusion Detection System
Hiroyuki Sakakibara†, Kiyoto Kawauchi†, Shigeki Kitazawa†, Seiji Fujii†
† Information Technology R&D Center, Mitsubishi Electric Corporation

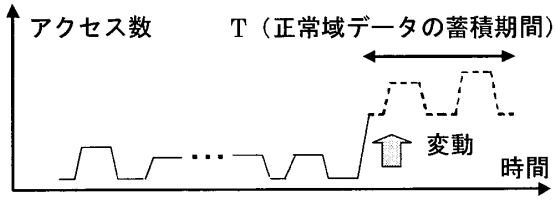

図 3　正常域データ変動後の正常域データの蓄積

3.　正常域データの変動の課題への対策

本稿では，前述の課題への対策として予測の正常域データを生成・使用することで速やかな分析の再開を実現する方式を提案する．

3.1.　提案する方式の概要

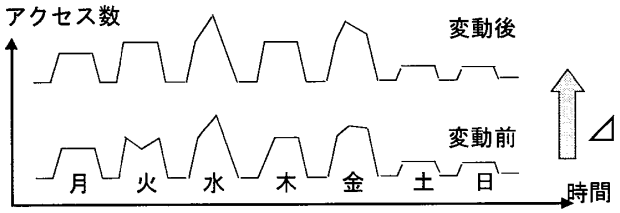

図 4　アクセス数の底上げ

変動後に長期間に渡り正常域データを蓄積せずに分析を再開するためには，代わりとなる擬似的なデータを利用する必要がある．筆者らの観測実験では，システム規模の拡大によりアクセス数が変動した場合，昼間アクセスが多く夜間は少ないといった日内変化，平日はアクセスが多いが土日は無いといった週内変化の傾向を保ちながら，アクセス数が底上げする傾向が見られた（図 4）．従って，変動前のアクセス数全体に対して一定の値を足すことで，日内・週内変化の傾向を保ちながらアクセス数の底上げを行ったデータを実データの代わりに利用することが有効であると判断した．これを予測正常域データと呼ぶ．また，分析の再開時は予測正常域データを使用するが，正常域データとして必要な期間アクセス数の観測・蓄積が行われた後は，蓄積されたアクセス数を正常域データとして利用する．

3.2.　予測正常域データの生成

予測正常域データを生成する際に底上げする大きさ（以降⊿と呼ぶ）を求める必要がある．⊿は予め分からないため，変動後に短期間アクセス数を観測し⊿を求める方法を提案する．変動が発生した 1 日のアクセス数を観測しその平均 μ_a と，変動前のアクセス数の平均 μ_b との差を⊿とすることにした．筆者らの観測実験では，アクセス数は異なる曜日間では差が大きかったため，変動前のアクセス数全体の平均を μ_b としてしまうと実際の⊿と差異が発生する可能性がある．逆に，同じ曜日間では差は小さかったため，変動が発生した曜日と同じ曜日の変動前のアクセス数の平均を μ_b とすることにした．$\Delta=\mu_a-\mu_b$ を変動前のアクセス数全体に足すことで予測正常域データを生成する．

3.3.　予測正常域データの選択

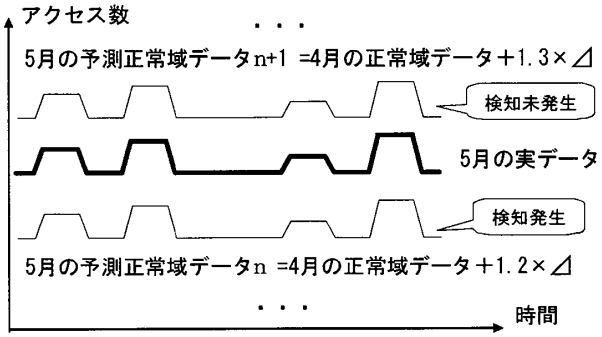

図 5　複数の予測正常域データと選択

予測正常域データを利用する場合，実データとの間に誤差が生じるが，可能な限り誤差の小さいものを使用する必要がある．本節では，予め複数の予測正常域データを準備し，この中から誤差の小さいものを選択する方式を示す．

まず，3.2に従い⊿を生成する．次に，⊿に複数の倍率をかけたものを用意し，各々に変動前のアクセス数を足す．例えば，図 5は，5 月に変動が発生した場合の複数の予測正常域データの生成の例である．5 月の予測正常域データ n は 4 月の正常域データに⊿×1.2 を，n+1 は⊿×1.3 を足したものである．この様に複数の倍率を⊿に掛け予測正常域データを複数生成する．次に，予測正常域データ生成後の 1 日間において，複数の予測正常域データを用いて分析を行う．検知が発生した予測正常域データは実データより小さいと考えられるため破棄し，残った中で一番小さい予測正常域データを採用する．図 5の例では，5 月の予測正常域データ n 以下は検知が発生したため破棄し，残った中で一番小さい n+1 を採用することで 5 月の実データと誤差の小さい予測正常域データを選択する．

以上の様に，提案方式では短期間の観測から実データとの誤差が小さい予測正常域データを準備し速やかな分析の再開が可能になる．

4.　おわりに

開発システムにおける，正常な時系列データの変動の課題を示し，予測正常域データを用いる解決方法を提案した．今後はプロトタイピングを行い，実際の監視環境における正常域データの変動を実験データとして，提案方式の有効性を評価する予定である．

参考文献

[1] 定点観測による不正アクセス分析システム, 榊原, 北澤, 藤井ほか, CSEC-35, pp63-68(2006)
[2] 主成分分析を用いたネットワーク異常検知システムの運用評価, 大野, 藤井ほか, DICOMO2007, pp1151-1154