トラフィック特徴量の相関特性を用いた異常検出

知能情報講座　松本　亮介

1. まえがき

近年，ウィルスや DoS 攻撃などのネットワークを介した攻撃が増え，ネットワークの運用・管理においてセキュリティへの対応が重要になっている．そこで，インターネット上で起こるセキュリティインシデントを分析する IDS(インシデント分析システム)が注目されている．IDS においては，各種ログデータなどからインシデント候補をリアルタイムで検知することが重要である．現在，データマイニングの観点から時系列データの変化点検出を行うことにより，インシデントを検出する手法[1]など，さかんに研究が行われている．本研究では，トラフィック情報に含まれるヘッダ情報から，対象ネットワークの特徴を抽出し，その特徴量の相関特性から時系列データを作成し，変化点検出を行うことで，異常を検出する手法を提案する．

2. 提案手法

2.1 トラフィック特徴量

異常検出を行う前処理として，トラフィックの特徴を反映した時系列データを抽出する．トラフィック情報から，ヘッダ情報(フラグ，パケットサイズ，サービス)に着目して，ヘッダ情報の値に対応するパケットを単位時間（1 分）でカウントし時系列データを得る．同様に，ヘッダ情報(送信元 IP，送信先 IP，送信元 Port，送信先 Port，パケットサイズ)の組み合わせに着目して，任意の組み合わせに対応するパケットを単位時間でカウントし，その中に含まれる組み合わせの種類数で除算することで，組み合わせの種類が占める比率を計算し時系列データを得る．

2.2 相関係数時系列データ

2.1 で求めた特徴量は，相関関係を持つものが含まれていると考えられる．そこで，求めた特徴量を用いて，特徴量間に生じる相関を計算し，時系列データを得る．ここでは，ピアソンの積率相関係数を用いる．まず，2 つの特徴量時系列データに対して，窓サイズ N を決定し，2 つの時系列データ間の N 区間内における相関係数を求める．そして，N 区間を 1 づつずらすことにより，2 つの時系列データから同じ時系列要素数で構成された 1 つの相関係数時系列データを得る．これによって，通常の通信では起こりにくい通信が現れた場合，相関関係が崩れることや，急激に強い相関関係を持つことがあり，その時点でデータの特性が大きく変化する時系列データを得ることができる．そこで，この時系列データの変化点を検出することで，異常であるトラフィックが検出できる．

3. 変化点検出エンジン

時系列データから変化点を検出する際には，変化点検出エンジンである ChangeFinder[1]を用いた．ChangeFinder は二段階の学習過程を行う．第一段階学習では，まず各時刻において，AR モデルに忘却機能と逐次学習機能を実装した SDAR アルゴリズムを用いて学習する．学習データから一区間先のデータを予測し，予測データと一区間先の実データを比較することで，予測データとのずれを計算し，その値を外れ値スコアとする．第二段階学習では，外れ値スコアを再度 SDAR アルゴリズムで学習することで，各区間での変化の程度を表す変化点スコアを得る．変化点スコアは値が大きいほど，変化点である可能性が高い．この手法は，二段階学習を行うことで，外れ値の影響を緩和し，時系列データの本質的な振る舞いを知ることができる．

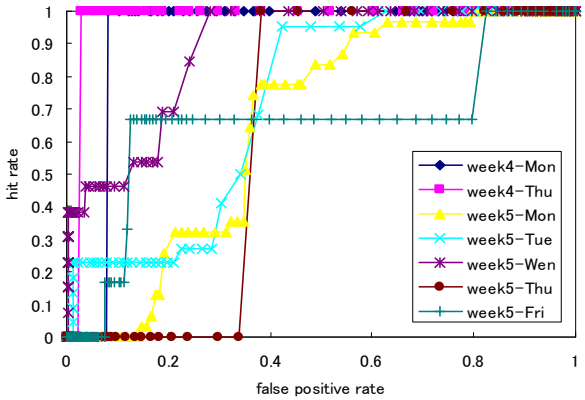

図 1：DoS 攻撃の検出率と誤検出率(HTTP)

4. 実験

本研究では，検証実験を行うためのデータとして，MIT の LICOLN 研究所が作成した IDS 評価用データの一部を使用した．検出するインシデントは，Prove 攻撃と DoS 攻撃とし，DoS 攻撃は主要ポートである HTTP，SMTP，Telnet に対する攻撃を対象とした．図 1 は HTTP に対する DoS 攻撃の ROC カーブであり，閾値の変動に対する検出率と誤検出率の関係を示している．縦軸は検出率，横軸は誤検出率を示している．図 1 から，DoS 攻撃の特徴をうまく抽出することができたと考えられる．

5. むすび

本研究では，トラフィック特徴量の相関特性の変化に着目し，その変化点を検出することにより，異常とする手法を提案した．実験により，これまで検出困難であった DoS 攻撃の特徴をうまく抽出することができた．検出の遅延を誤検出としないように閾値設定を行うことで，さらに精度を向上できると考えられる．

参考文献

[1] J. Takeuchi and K. Yamanishi, “A Unifying Framework for Detecting Outliers and Change Points from Time Series,” IEEE transactions on Knowledge and Data Engineering, vol.18, no.4, pp. 482-492, 2006.