L-032

統計処理トラフィックを用いた DoS 攻撃検知
Dos attack detections by applying statistically processed traffic

張 振 † Zhen ZHANG　渡辺 一平‡ Ippei WATANABE　宮内 充† Mitsuru MIYAUCHI

1. はじめに

インターネットの爆発的な普及に伴い不正アクセスも増加の一途をたどっており、ここ数年 DoS（Denial of Services)攻撃が顕著になっている。DoS 攻撃は、攻撃用ツールが Web 上で簡単に入手可能であり、Bot の出現や多地点からの分散攻撃などが被害増加の原因となっている。

不正なパケットを検出する侵入検知システム(IDS：Intrusion Detection System)では未知の攻撃が検知できない点とトラフィック量の増加に従ってパケットの取りこぼしによる検知率低下が問題点となる。そこで、大容量のネットワークではパケットの内容解析を行わずネットワークトラフィックの状態を監視し、統計値やその分布の異常を検知する IDS を用いるのが有利である[1]。

本稿では、すでに提案している[2]過去のトラフィックパターンを元に、将来のトラフィックを予測し、その結果と実際のトラフィックを比較することによりトラフィックの異常を検知する手法に統計処理を用いて、評価結果を示す。

2. 提案手法

2.1 トラフィック予測を用いた検知手法

本提案とは、もし将来のトラフィック量が分かっており、その値に十分な信頼性があるならば、観測したトラフィックパターンの異常の原因が DoS 攻撃によるものかどうかを判断することが可能となるという考えに基づいている。

従来、予測方法として、ニューラルネットワークモデルを用いた手法をまず取り組んだが、処理時間が長いため、実用できないという問題点があった。そこで、テンプレート方式としてパターンマッチングによる方法で処理時間の短縮に見通しを得たが、推定精度の向上が課題となった[2]。そのため、今回の提案手法は高速処理が期待できる回帰直線アルゴリズムを使い、トラフィックパターン処理部分では揺らぎの除去と長期周期考慮の統計処理を加え、高速予測かつ精度向上を目指す。図 1 では予測システムの流れを示す。

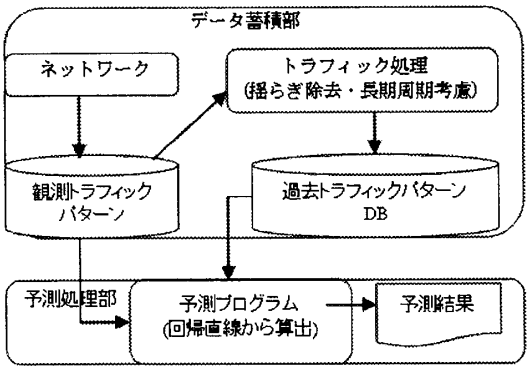

図 1 .予測システム

2.2 回帰直線を用いたトラフィック予測

本研究で提案するトラフィック予測は回帰直線を用いて行う。まず、予測に用いるトラフィックパターン（現在の観測トラフィックパターンと最も相関係数が高い物）を過去のトラフィックパターンデータベースから抽出する。次に、テンプレートとなるパターンと現在の観測トラフィックパターン間の回帰係数を求め、回帰直線から数学的に将来のトラフィックパターンを求める。式 1 に回帰係数計算式、式 2 に予測トラフィックパターン計算式をそれぞれ示す。

$$\hat{a}=\frac{n\sum_{i=1}^{n}x_i y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i}{n\sum_{i=1}^{n}x_i^{\,2}-(\sum_{i=1}^{n}x_i)^2} \qquad \hat{b}=\frac{\sum_{i=1}^{n}y_i-\hat{a}\sum_{i=1}^{n}x_i}{n} \qquad \text{（式 1）}$$

$$y=\hat{a}x+\hat{b} \qquad \text{（式 2）}$$

式 1 から回帰直線の回帰係数 a と b が求められる。式 2 では、テンプレートとなった過去のトラフィックパターンの x と式 1 求められた回帰係数 a、b から、将来のトラフィックパターン y が求められる。また、n はトラフィックパターンの長さである。計算は四則演算のみであり高速処理が期待できる。

2.3 トラフィック揺らぎの除去

予測を高精度で行うためにはトラフィックの特徴を抽出する必要がある。トラフィックパターンは通常人為的要因から発生されているので各周波数によって何らかの特性を持っている。高い周波数成分の発生は、突発的な事象により発生しているものと考えられる。突発的なトラフィックパターンの変化は、予測をする上で不必要である。DoS 攻撃を検知する目的ならば、トラフィック波形の揺らぎ細部までを一致させる必要はないため、ここでは統計処理により高周波数成分をカットすることを提案する。この揺らぎを除去するために移動平均法を用いてトラフィックパターンの高周波数を遮断する。式 $a_0'=\sum_{i=0}^{n-1}a_i/n$ の処理するデータサンプル数（フィルタ強度）n を可変することで周波数特性を変更できる。

2.4 周期の抽出と周期の除去

長期周期はトラフィック変動の中から不定期に発生している変動を取り除くことで得ることができる。そこで、周期の抽出を行うためには、数日間のトラフィックパターンを収集し、同時刻におけるサンプルごとの平均をとり、その値を時系列に沿って並べたものを長期的な周期成分とする。この長期的な周期は毎日繰り返されるトラフィックの変化には共通して現れる。すなわち、この周期成分はトラフィックパターンから除外してトラフィック予測を行うことが可能である。単純に観測したトラフィックパターンと周期パターンとの差分を求め、その結果を予測用のトラフィックパターンとする。

† 千葉工業大学 情報科学研究科
‡(株)インターネットイニシアティブ, IIJ

3. 実験

3.1 実験内容

実験はトラフィック容量が大きい JPIX[3] 5分間隔トラフィックパターンと、100Mbps 回線を想定して synTraff[4] で生成した 1 秒間隔トラフィックパターンを用いそれぞれ 5 分後、1 秒後のトラフィック予測を行う。メインの予測システム開発には高速性を重視し C 言語を用い、予測システムを制御するスクリプトは柔軟性が高い Perl 言語を用いた。表 1 に設定ファイルによって与えることができるパラメータを示す。

表1． 実験パラメータと可変内容

トラフィックパターンデータ長 (D)	3,5,7,10,15
比較するパターンの数 (P)	10,50,100,500,1000
長期周期考慮 (Long)	有/無
揺らぎ除去強度フィルタ (F)	0,3,5,7,10

3.2 実験結果

(1). 揺らぎ除去の有効性

synTraff 生成トラフィックをトラフィックパターンデータ長 10、フィルタ強度 3、比較するパターン数 1000 の条件で予測した結果を図 2 に示す。また各フィルタのパラメータにおける観測トラフィックパターン(観測 TP)と予測トラフィックパターン(予測 TP)の誤差から求めた標準偏差を図 3 に示す。なお、予測に要した時間は 1 日分 86400 回で約 5 秒であり、リアルタイム処理が可能な時間である。

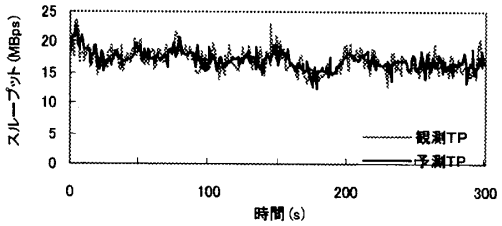

図 2．synTraff 生成トラフィックの予測結果

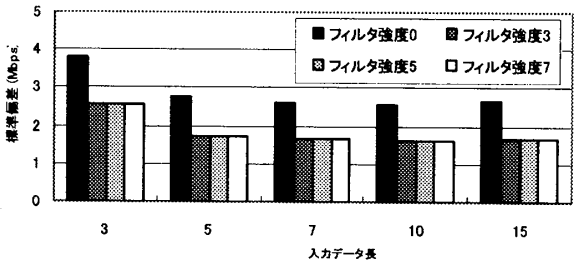

図 3. synTraff 生成トラフィックの予測誤差比較

図 2 から synTraff 生成トラフィックの予測では、トラフィックパターンの特徴を正確に認識して予測が行えることがわかった。図 3 から揺らぎ除去を行った場合の方が行わなかった場合(フィルタ強度 0)よりは予測誤差が小さくなった。また、フィルタの強度を高くしても、ほとんど予測精度に変化が見られなかった。これはトラフィック予測を阻害するような揺らぎはフィルタ強度が 3 程度で十分取り除くことが可能であるためといえる。今回の実験では、フィルタ強度で精度変化が見られなかったが、あまりフィルタ強度を上げすぎてしまうと予測を行う上で必要なパターンまで除去されてしまうおそれがあると考えられるので、フィルタ強度は 3 程度が最適値であると考えられる。実験結果から 100Mbps 回線において標準偏差 2Mbps 程度の誤差範囲で予測が可能となることがわかる。

(2). 長期周期考慮の有効性

ここでは長期周期を考慮したことにおりどの程度トラフィックの予測精度が向上したかについて考察する。図 4 に SynTraff 生成トラフィックおよび JPIX トラフィックの長期周期を考慮した場合としない場合の相関係数比較を示す。なお、フィルタ強度は 3、トラフィックパターンデータ長は 3 である。

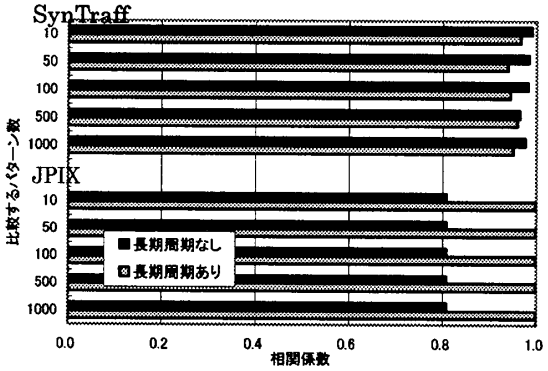

図 4. 長期周期有/無の比較

JPIX トラフィックの予測においては長期周期考慮した場合に相関が 0.2 程度増加しており非常に精度向上が行えるが、逆に SynTraff の場合には精度が若干低下してしまった。このことから予測するトラフィックによって長期周期の有効性が変わってくると言える。JPIX のようなたくさんのユーザが利用するネットワークでは大数の法則から突発的な揺らぎが減り、トラフィックの長期周期がトラフィックパターンに与える影響が大きくなることが考えられ、それを排除するアルゴリズムは有効である。それに対して、揺らぎの成分が多く含む SynTraff 生成トラフィックでは長期周期がトラフィックパターンに与える影響が少ないため、余計な処理を行うことになり、逆に本来トラフィックが持っていた特徴を掻き消してしまい、精度が低下すると考えられる。つまり、予測するトラフィックの特性によって長期周期を予測に用いるかどうかを判断する必要があると言える。

4. まとめ

本研究では DoS 攻撃検知に向けてトラフィック予測の精度向上手法として、回帰直線による予測、揺らぎの除去と長期周期を考慮した手法の提案とその実験を行い、提案手法の有効性を確認した。その結果、提案手法は予測精度向上に有効であり 100Mbps 回線においては 2Mbps 程度の誤差範囲で予測ができた。またトラフィックの特性によって、長期周期を考慮する有効性も示した。今後の展望として予測に回帰直線以外関数の導入や、実際に DoS 攻撃検知システムに予測システムを組込んだ精度の高い攻撃検知システムの作成などが挙げられる。

文献

[1].武井洋介,太田耕平,他 ,"ﾄﾗﾌｨｯｸﾊﾟﾀｰﾝを用いた不正ｱｸｾｽ検出及び追跡方式",信学論文誌,VOL.J84-B,No.8,P1464 (2001)
[2].平石陽太,渡辺一平,宮内充,"ﾄﾗﾋｯｸﾊﾟﾀｰﾝ自動生成におけるDoS 攻撃検知",信学技報 IN2004-62 pp.13-18, (2004-9)
[3].JPIX http://www.jpix.ad.jp/jp/techncal/traffic.html
[4].SynTraff http://www.cs.usask.ca/faculty/carey/software.html