機械学習型侵入検知システムの改善

木下研究室　　小野翔太郎 (201002684)

1　まえがき

近年、アクセス制御や暗号技術などのコンピューターセキュリティの発達により、安全な通信を行うことができる。しかし、セキュリティ技術の発達はその分攻撃者の手口が進化してることも表している。特に DoS 攻撃に対しては各地で現段階も研究が続けられている。

DoS 攻撃とは大量のパケットなどのリクエストを送信したり、受信側のバグを突いたりして企業や政府機関などのサービスそのものを妨害する攻撃手法である。対策として IDS の開発がされている。IDS は DoS 攻撃やウイルス攻撃などの様々の攻撃を検知する。しかし欠点があり、検知の仕方によっては未知の攻撃に対応できなかったり、通常の通信接続に対しても誤って検知してしまう場合がある。そのための研究として IDS 自体に機械的学習アルゴリズムを用いて学習機能を持たせることによって未知の攻撃を防ぐ研究がされている。しかし問題点として一部のプロトコルのみでしか実装されておらず、完全には性能評価されてはいない。またこれらの研究の学習には時間が掛かる。

そこで本研究の目的は従来の研究で用いられている遺伝的アルゴリズム (以下 GA) による特徴配列の生成のアルゴリズムの改良を加えることで IDS が検知にかかる処理時間を短くし、DoS 攻撃に対する検知率を高くする機械学習型 IDS を提案していくことである。改良点は特徴配列生成において遺伝子操作の部分を変えていく。従来の研究では一点交叉を用いてたが、最適な解にたどり着かないことも多く学習するまでの時間がかかり結果として検知するまでの時間がかかったり、誤検知もしやすかった。そのために今回の研究では二点交叉を用いて、交叉するために仕分ける箇所を二点に増やすことで最適な解に近づけやすくしていき学習するまでの時間を縮めていく。また今回は実装されていないプロトコルの中で TCP に並び DoS 攻撃が多発している UDP で実装をしていく。

2　提案方式

本研究では機械型学習に着目して GA を用いる。GA は一般的に学習アルゴリズムとして用いられる。GA が一般的な学習アルゴリズムとして用いられる理由は広い汎用性に加えて、確率的要素を含み得られた解から新たな手法や分析でき、さらには基本的にはある問題に対する評価関数のみを考えるだけでいいので最適化問題として応用性が高いためである。これらの特徴から今回は IDS にこの GA を適用させることで学習型 IDS として、様々な攻撃パターンがあり検知することが難しい未知の DoS 攻撃にも対応できる IDS を提案していく。

今回は各プロトコルの中で DoS 攻撃が多発している UDP において DoS 攻撃の特徴配列（DoS 攻撃のデータによく似た配列）生成に GA を用いる。また特徴配列生成の遺伝子操作において一点交叉の代わりに二点交叉を用いることで効率良く学習していく。

手順として、ベンチマークテストやデータマイニングなどに使われる KDD99 と呼ばれる米空軍基地のローカルエリアネットワークのダンプデータを使用する。今回はこのデータの中のトレーニングデータから GA を用いて評価関数によって適応度をつけ、DoS 攻撃の特徴配列を作り、同じく KDD99 のテスト用データと類似度を比較、検知して性能を評価していく。今回は従来の研究の評価関数を用いていく。

今回用いる評価関数の式は以下のようになる。

$$fitness = W1 \cdot \frac{|AandB|}{N} + W2 \cdot \frac{|AandB|}{A_N} \quad (1)$$

ここで $W1$ と $W2$ は重みである。$|AandB|$ は DoS 攻撃だと判定する条件 A と条件 B 両方に当てはまるネットワーク接続の数であり、N はネットワーク接続の総数であり、A_N は条件 A に当てはまるネットワーク接続の数である。この適応度が高ければ高いほど DoS 攻撃のデータに近いことがわかる。

特徴配列生成のアルゴリズム

1. KDD99 のランダムに選択されたトレーニングデータを入力、一つ一つのパケットを染色体とし、その集まりを初期世代集団とする。
2. 評価関数に用いられる重みパラメータの値を設定する。
3. 集団内の各染色体に評価関数によって適応度を与える。
4. 集団内の染色体を 2 つ選択して、二点交叉によって決められた位置の遺伝子を交換する。
5. 集団内の染色体をランダムに 1 つ選択して、その染色体のビット列を反転させる。
6. 世代数を＋1 世代とする。
7. 規定の世代数になるまで 36 を行う。
8. 最後の世代なったときの一番適応度の高い染色体を DoS 攻撃の特徴配列とする。

3　今後の課題

遺伝的アルゴリズムは基本的な計算手順はあるが現在もどの問題に対して適用できる一般的な計算手順は確立していない。そのため、それぞれの問題に対して各自で評価関数や遺伝子操作などを工夫する必要がある。しかしその分さらなる可能性を期待できる。

今回はまだ実装は行っていないが今後は IDS に他の一般的に用いられる学習アルゴリズムの中で最も認識性能が高いと言われている SVM を用いて学習機能を比較することで、遺伝的アルゴリズムの有用性を示していく。