2007 年 5 月 8 日 (火)
1031060322 松本　亮介

ＴＣＰフロー情報を用いた ネットワークトラフィックからの異常検出（概略）

0 はじめに

※この資料は，参考文献を元に，参考文献の概略を作成したものです. この研究を私が実際に行ったわけではありませんので注意してください.

1 流れ

膨大なネットワークトラフィックから異常なトラフィックを検出するために，TCP フロー情報を用いた異常検出手法について提案する．TCP フロー情報は，MIT の LINCOLN 研究所が作成した IDS 評価用のデータを用いる．TCP フロー情報を用いると，IP パケット情報だけでは取得することができないような，コネクションあたりの通信時間や通信量などの特徴を抽出できる．また，不正アクセスを，攻撃先ホストの脆弱性を調査するポートスキャン，アクセスの集中によりサービスを妨害する DoS 攻撃，外部からの不正侵入やホストののっとりを行う R2L, U2R 攻撃の三種類に大別し，抽出した特長を用いてそれぞれに適した手法を適用することにより，効率的に異常を検出する手法を提案する.

2 手法

まず，攻撃先ホストの脆弱性を調査するポートスキャンに対しては，送信元の IP および送信先の IP, PORT に着目して, ポートスキャンを検出する手法を用いる.

基本的に，ポートスキャンは，以下の３つの攻撃特性を持っている場合が多く，正常な通信では同一のサーバから複数のサービスを短時間に利用することは珍しい.

1. 送信元・送信先 IP アドレスの組合せが同一
2. ポート番号がバラバラなアクセス
3. 大量の通信が発生

これらの特徴に着目し，単位時間内に任意の送信元，送信先 IP アドレスの組み合わせを含む通信が大量に発生し，さらに送信先ポート番号のばらつきが閾値を超える場合は，該当する TCP フローをポートスキャンとして検出する．ここで，サーバ側が公開するサービス数とアクセスを受け付けるサービスポート数は正の相関を持っている．そのため，公開サービス数を増加させない限り，アクセスを受けるサービスポート数は上限が決まっている．その値の上限を閾値とする.

次に，アクセス集中を意図的にさせることでネットワークサービスの妨害をする DoS 攻撃に対しては, 送信先 IP，PORT 及び TCP フロー内における総 IP パケット数に着目して，DoS 攻撃を検出する手法を用いる.

DoS 攻撃の基本的な以下の３つの特徴に注目して検出する.

1. 短時間に大量のアクセスが発生
2. アクセス対象は特定ホストの特定ポート
3. 大量に発生する通信内容は同じ

これらの特徴に着目し，単位時間内に任意の送信先 IP アドレス，ポート番号，TCP フロー内総 IP パケット数の組み合わせが同一である TCP フローが閾値を超える場合，同一の組み合わせを持つ TCP フローを DoS 攻撃として検出する．正常なトラフィックでも同一の組み合わせとなる TCP フローは複数出現する場合がある．そのため，本稿では実験的に最適であった値を閾値として使用した.

最後に，外部からの不正侵入やホストの乗っ取りを行う R2L，U2R 攻撃に対しては，時系列モデルを用いた手法を用いる.

用いる時系列モデルは AR モデルで, 特徴の時間的変化を予測する．次に，その予測された特徴に対して信頼できる範囲を学習データとの類似性によって動的に決定する．最後に，観測された特徴がその信頼できる範囲を超えた場合，異常なトラフィックと

して検出する．AR モデルとは，時系列の現時点の値を過去の値の線形結合によって推定するモデルである．AR モデルでは，IP アドレスのような値の大小が意味を持たないような指標に関しては予測ができない．そのため，扱う特徴は単位時間内の総フロー数，単位時間内の総 TCP フローサイズ，単位時間内の総 TCP フロー継続時間，単位時間内の総 IP パケットとする．検証データと学習データの類似性を測るための指標として，JSD を用いる．

3 結果

ポートスキャンに関しては，正常を異常と検出する割合が 0.0003%のとき，すべてのポートスキャンを検出することができている．

DoS 攻撃に関しては，半分程度しか異常を発見することができていない．実際の DoS 攻撃では，HTTP に関してみると，攻撃ごとに異なるページを閲覧している．そのため、TCP フロー内における通信量も IP パケット数も毎回異なっているなど，検出が困難な DoS 攻撃が多い．さらに，正常通信であっても自動リロードでは常に同じ TCP フロー数となるため，提案の DoS 攻撃検出手法では正常を異常とする確率が高くなっている．

時系列を用いた手法に関しては，HTTP 以外では大体管理者の監視負荷を助けるほどの結果は得られていると考えてよい．HTTP では，全体的なトラフィック量が多い．このため，異常としか考えられないような文字列があったとしても，それはテキストであるためにサイズが小さい．そのため，他のサイズが大きい HTML ファイルや画像ファイルによる正常な TCP フローとの統計的差異が出ない．このことから、異常を発見することが難しくなっていると考えられる．

参考文献

[1] 橋爪拓，“TCPフロー情報を用いたネットワークトラフィックからの異常検出”