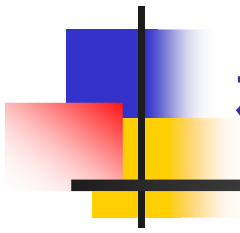
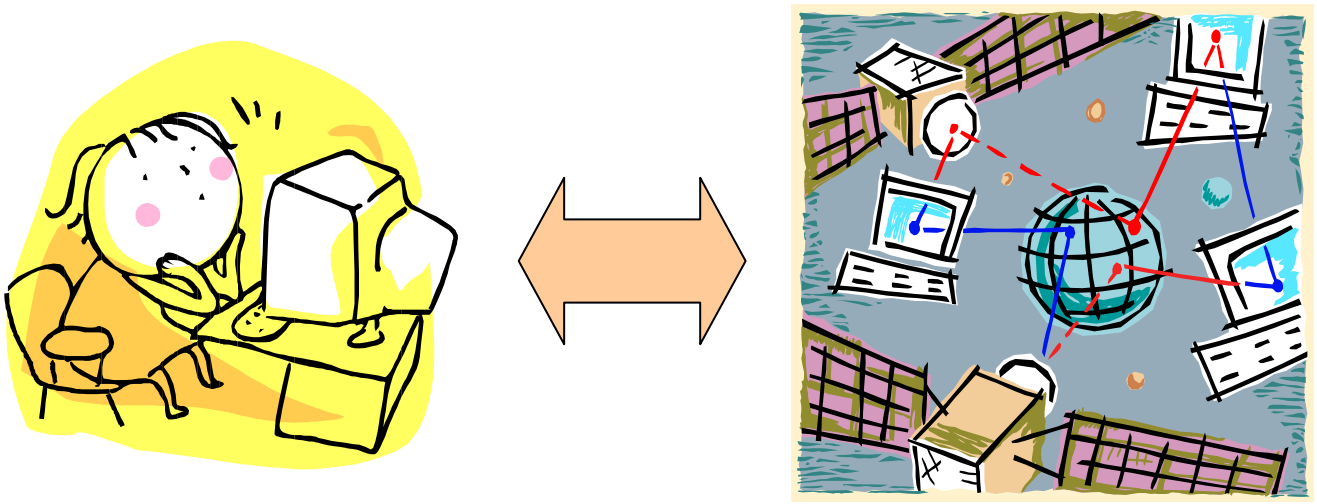


トラフィック特徴量の時系列データにおける相関特性を用いた変化点からの異常検出

平成19年度　後期雑誌会
B4　松本　亮介

研究背景



- 計算機・ネットワークは社会に必要不可欠
- ウィルスなどの発生　⇒　セキュリティが重要

研究背景

- 現実には様々な被害を受けている
- ネットワークセキュリティの確保
  - IDS(インシデント分析システム)への注目

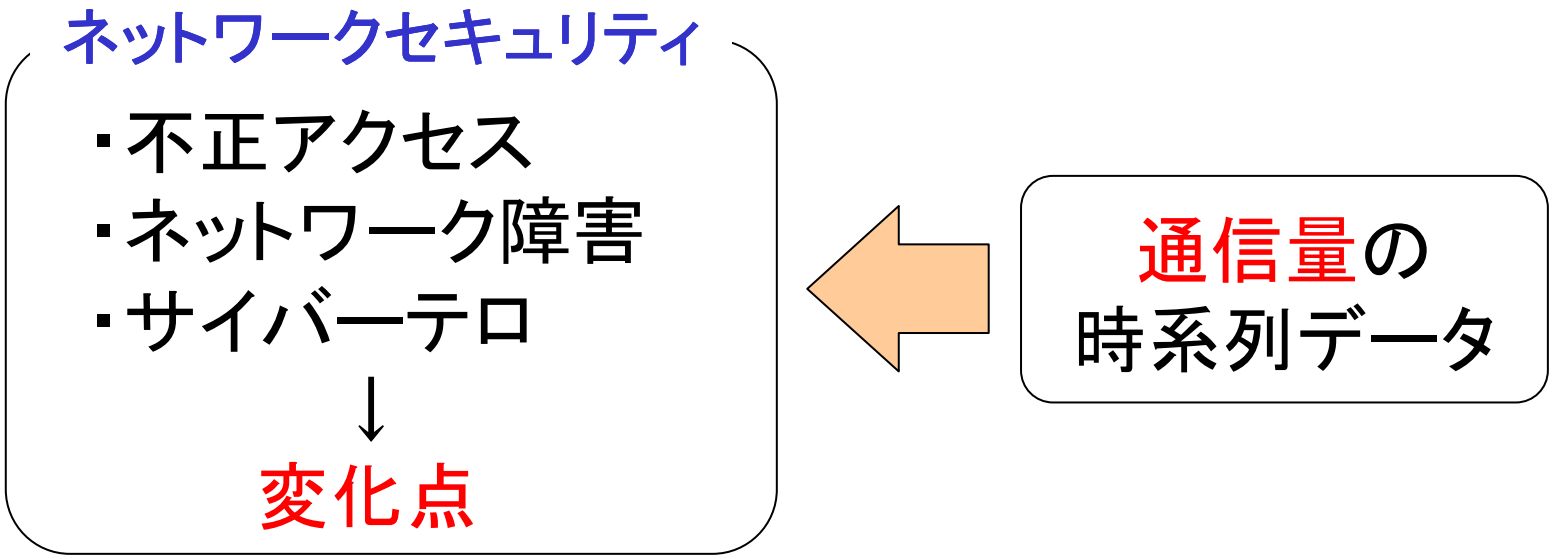
⇒ 各種ログデータからのインシデント候補を検出

リアルタイム検知
（インシデントを検出しアラートを出す）

⇒ ネットワーク管理者の監視を効率化

リアルタイム検知の研究

研究：時系列データからの外れ値特定による変化点検出



■ 変化点検出をセキュリティに利用

⇒ 変化点検出によってウィルス被害等を防ぐ

⇒ データマイニングの観点からセキュリティに応用

これまでの手法の問題点

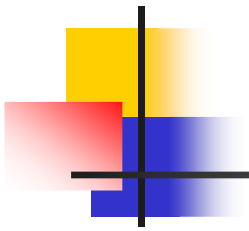
- 通信量やトラフィック特徴量を単一で使用
  - 異常状態が全体との関係において不明瞭
  - 単一のデータでは正常と異常の区別が困難

・悪質な攻撃であるDoS攻撃
・不正アクセスの準備段階で用いられるポートスキャン

⇩

一つ一つの通信は正常であることが多い

⇒ 検出が困難である攻撃法



本発表

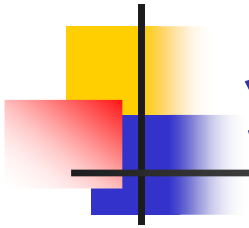
文献[4]によるChangeFinder（変化点検出エンジン）を用いて、トラフィック特徴量の相関特性を利用した手法を提案する。

⇒　単一の特徴量ではとらえられない異常を検出

⇒　DoS攻撃とポートスキャン検出の精度向上

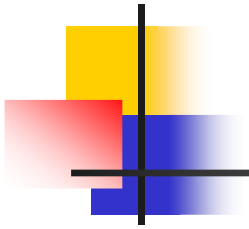
---

[4] J. Takeuchi and K. Yamanishi, “A Unifying Framework for Detecting Outliers and Change Points from Time Series,” IEEE transactions on Knowledge and Data Engineering, pp.482-492, 2006.



発表の流れ

- 提案手法の説明
- 実験と考察
- まとめ



目次

- 提案手法の説明
- 実験と考察
- まとめ

変化点検出における異常検出

■ これまでの手法

単一のトラフィック特徴から異常を検出

⇒ 正常と異常が混合された結果となる

[例1] HTTPにおいて大量の通信が起きた時
⇒ 不正アクセスによって生じている
⇒ ホームページを見ている人が大量にいる

[例2] 単一の特徴量で変化点を検出
⇒ 各時間において全体との関係が分からない
⇒ 単一の特徴量では変化点を異常とする設定が困難

提案手法による異常検出

■ 提案手法

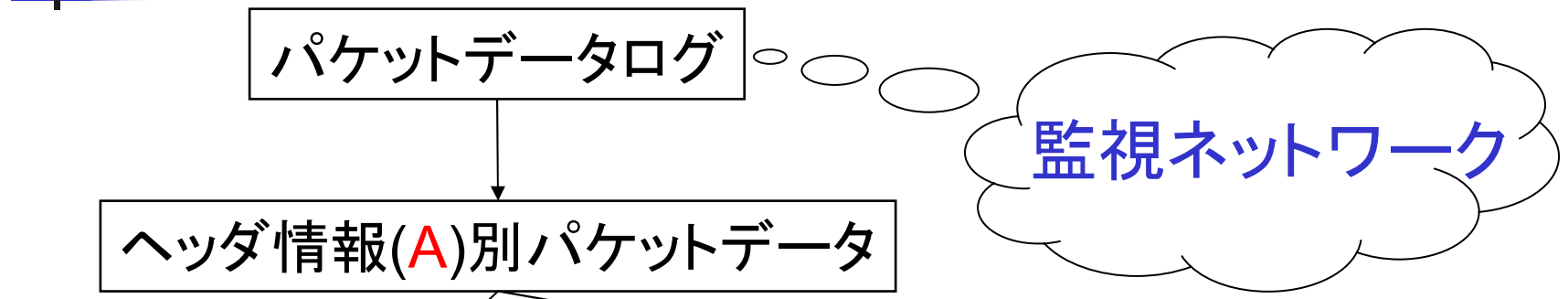
複数のトラフィック特徴の相関関係を利用

・複数のトラフィック特徴量の関係を考慮
⇒ 単一の特徴量では捉えられない変化を抽出

⇒ 正常と異常を区別できる可能性が高くなる

1、トラフィックから特徴量を抽出し時系列データ作成
2、時系列データによる相関係数時系列データ作成
3、相関係数時系列データから変化点検出 ⇒ 異常

トラフィック特徴量抽出の流れ



■ ヘッダ情報の種類

(A) サービス別、フラグ別、パケットサイズ範囲別

(B) 送信元IP、送信元ポート、送信先IP、送信先ポート、パケットサイズ

⇒ 本発表ではDoS攻撃とポートスキャンの検出精度向上を目標

DoS攻撃

■ DoS攻撃（HTTPへの攻撃と定義）

・大量の接続要求を送ることでサーバに過負荷を発生させる攻撃 ⇒ 一つ一つは正常な通信（検出が困難）

⇒機械的に行うことが多いため、ヘッダ情報に偏りが生じる

送信先IP＋送信先ポート＋データサイズ

通信量と偏りは通常ではゆるやかな相関関係

⇩

相関に急激な変化

DoS攻撃である可能性が高い

⇒ 通信量と偏りを時系列データに抽出

時系列データ処理(例：通常状態)

単位時間を1分として処理を行う

時間	送信元IP.送信元ポート > 送信先IP.送信先ポート: (パケットサイズ)
00:37:07	IP 172.16.114.50.http > 206.48.44.50.2222: . ack 5841 win 32120
00:37:17	IP 172.16.114.50.http > 206.48.44.90.2313: . ack 2921 win 32120
00:37:25	IP 206.48.44.40.2222 > 172.16.114.30.http: . (256) ack 8192 win 31744
00:37:25	IP 206.48.44.50.2222 > 172.16.114.40.http: . (1320) ack 8192 win 31744
00:37:38	IP 206.48.44.60.2222 > 172.16.114.70.http: . (156) ack 8192 win 31744
00:37:49	IP 206.48.44.90.2313 > 172.16.114.50.http: . (1460) ack 8192 win 31744
00:37:58	IP 206.48.44.90.2313 > 172.16.114.50.http: . (1460) ack 8192 win 31744

7

[総パケット数] = 7　　⇒ 通信量(HTTP)

[組み合わせ数] = 4
[組み合わせにおけるヘッダ情報要素を含むパケット数] = 5

$$\frac{[\text{パケット数}]}{[\text{組み合わせ数}]} = \frac{5}{4}$$

⇒ 偏り(DoS攻撃)

時系列データ処理(例：DoS攻撃)

時間	送信元IP.送信元ポート > 送信先IP.送信先ポート: (パケットサイズ)
00:38:07	IP 172.16.114.50.http > 206.48.44.50.2222: . ack 5841 win 32120
00:38:17	IP 172.16.114.50.http > 206.48.44.90.2313: . ack 2921 win 32120
00:38:25	IP 206.48.44.50.2222 > 172.16.114.50.http: . (320) ack 8192 win 31744
00:38:25	IP 206.48.44.50.2222 > 172.16.114.50.http: . (320) ack 8192 win 31744
00:38:38	IP 206.48.44.50.2222 > 172.16.114.50.http: . (320) ack 8192 win 31744
00:38:49	IP 206.48.44.90.2313 > 172.16.114.50.http: . (320) ack 8192 win 31744
00:38:58	IP 206.48.44.90.2313 > 172.16.114.50.http: . (320) ack 8192 win 31744

7

[総パケット数] = 7 ⇒ 通信量（HTTP）

[組み合わせ数] = 1

[組み合わせにおけるヘッダ情報要素を含むパケット数] = 5

$$\frac{[\text{パケット数}]}{[\text{組み合わせ数}]} = 5$$

⇒ 偏り（DoS攻撃）

⇒ 時系列データ

抽出した時系列データの相関関係

相関係数時系列データを作成 ⇒ 変化点を検出

ポートスキャン

■ ポートスキャン

・サーバのサービスポートに順番にアクセスし、侵入口となりうる脆弱なサービスポートがないかを調べる行為

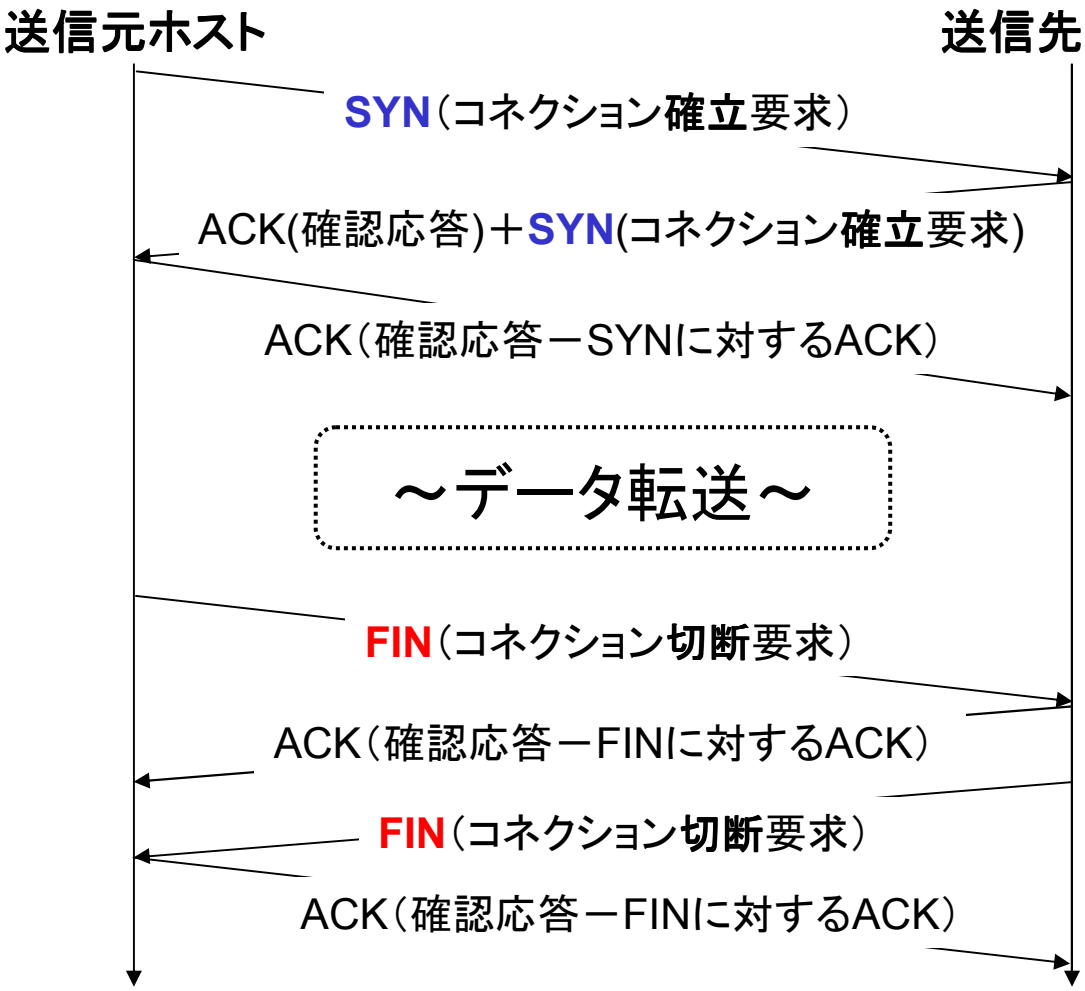
[特徴1] 送信元IPは偽装されない

[特徴2]サーバ管理者にスキャンを知られたくない

⇒ 接続を確立しない通信方法

⇒ 特徴量において相関関係が崩れる部分が発生

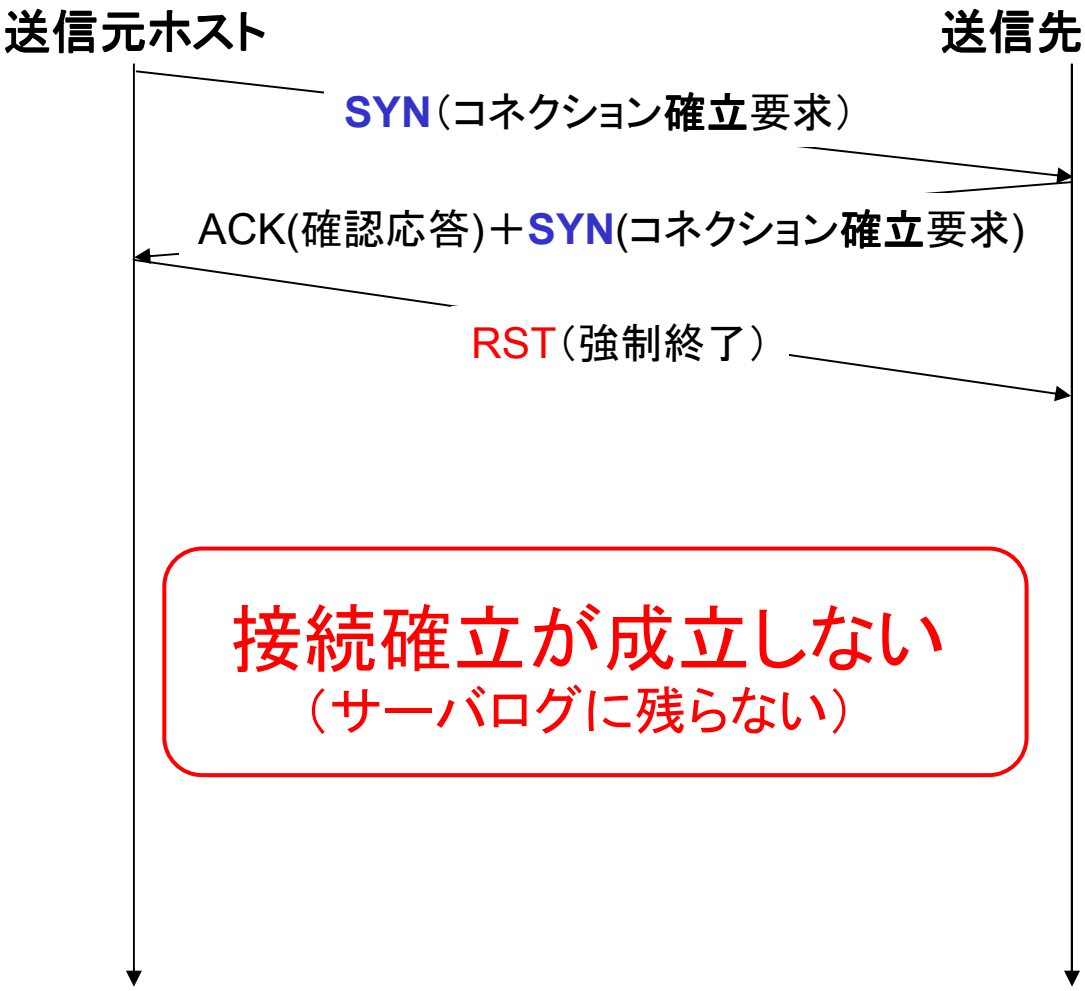
通常の通信方式



3way-handshake
接続確立と接続切断

SYNとFINは
強い相関関係

ポートスキャンの通信方式



接続確立が成立しない
（サーバログに残らない）

SYNフラグとFINフラグ
の対応がとれていない

SYNパケット数と
FINパケット数の
相関が崩れる

時系列データ処理(ポートスキャン)

・通信量(ヘッダ情報:SYN、FIN)

・SYNフラグの立ったパケット
・FINフラグの立ったパケット

単位時間(1分)当たりのパケット数をカウント

⇒ フラグ別のパケット数を特徴とした時系列データを抽出

抽出した時系列データの相関関係

相関係数時系列データを作成 ⇒ 変化点を検出

相関係数時系列処理

ピアソンの積率相関係数

・窓サイズNにおける時系列間の相関係数 ⇒ 1ずつ窓をずらす

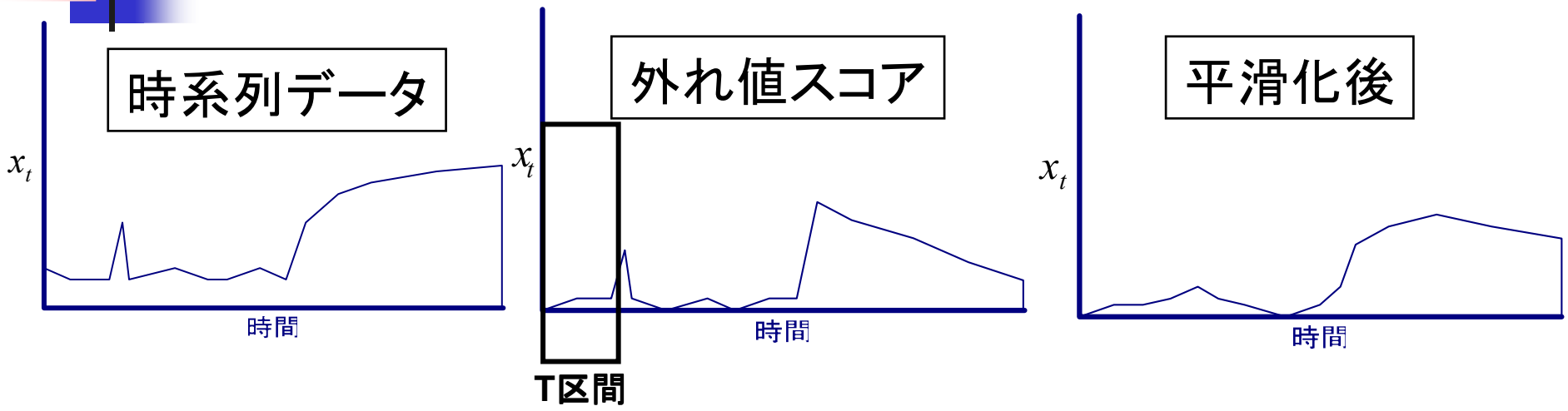
$$\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}\sqrt{\sum_{i=1}^{N}\left(y_i-\bar{y}\right)^2}}$$

$$(x,y)=\{(x_i,y_i)\}\ (i=1,2,\cdots.N)$$

$\bar{x},\bar{y}$：N内の相加平均

※ ピアソンの積率相関係数は、線形相関以外に対応していない計算法であるが、予備実験の結果十分特徴を得ることができていたため採用した。

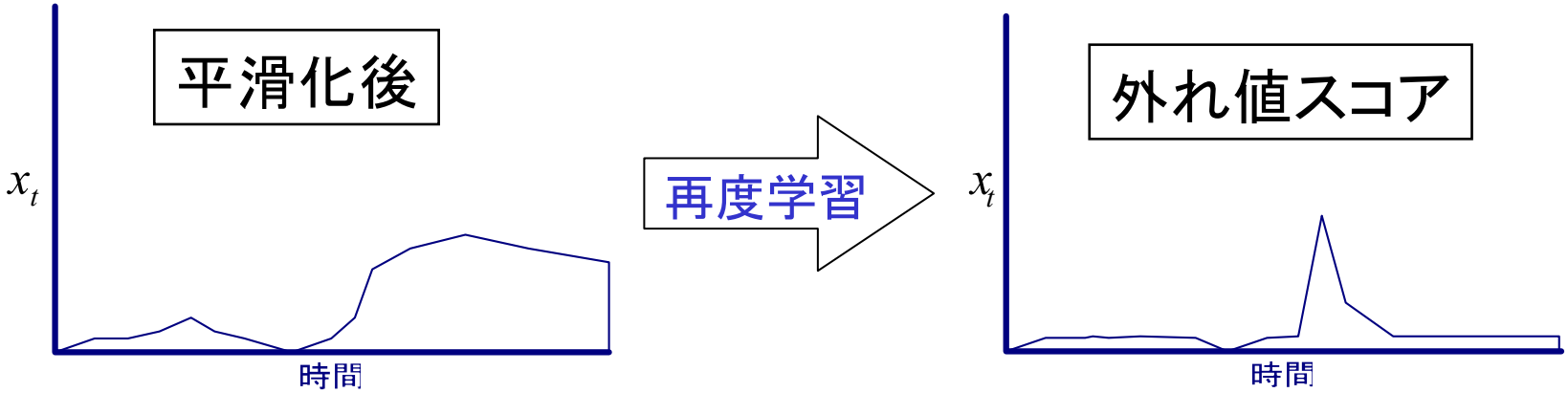
ChangeFinder[4]の流れ（第1段階）



新しいデータ毎に、その時点からT時点前までの平均をとる
⇒ T区間をずらすイメージ

平滑化後の外れ値スコア ＝ 時系列データ
⇒ 再度学習

ChangeFinder[4]の流れ（第2段階）

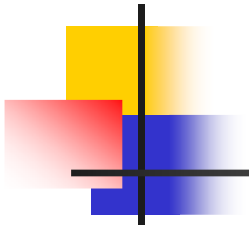


第1段階学習での外れ値スコア

⇩

第2段階学習では変化点スコア

⇒ 変化点スコアが高いほど変化点である可能性が高い



目次

- 提案手法の説明
- 実験と考察
- まとめ

実験条件

■MITのLINCOLN研究所が作成したIDS評価用データ
・IDSの性能を比較するリファレンスデータ

・Week5のTuesdayデータ(tcpdump形式)
-ポートスキャンとHTTPに対するDoS攻撃が含まれる

・攻撃時間（単位時間は1分）
- HTTPに対するDoS攻撃 ― 211分、795分
- ポートスキャン ― 616分

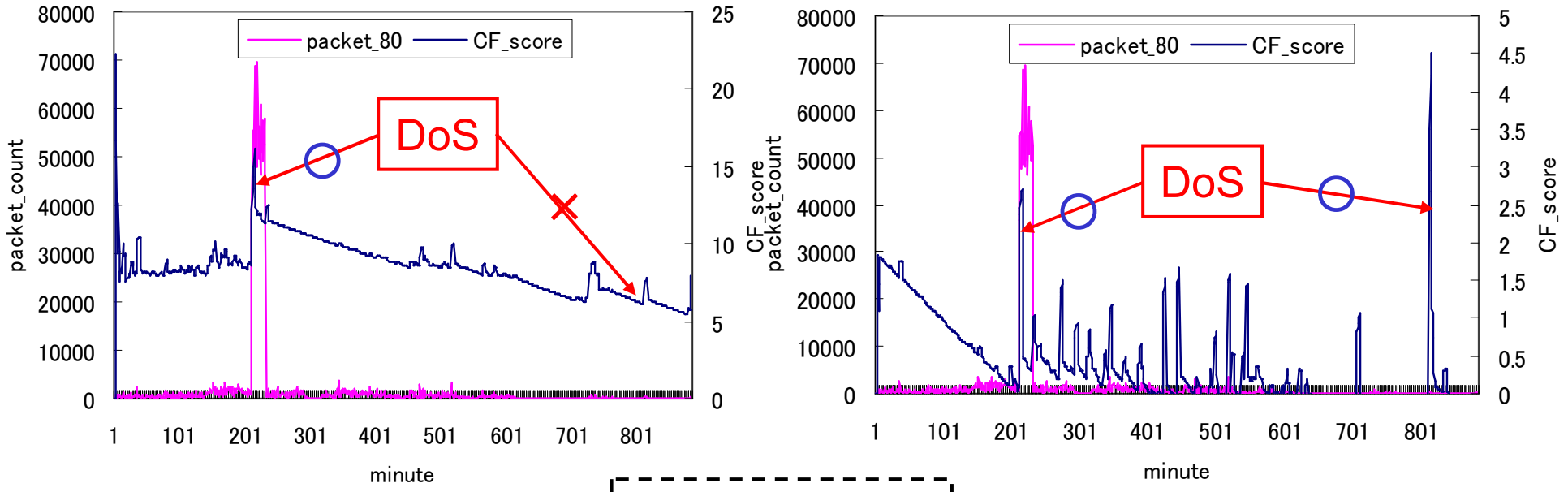
・相関係数の窓サイズ　N = 20

・実験
- [実験1]通信量（HTTP）と偏り（DoS攻撃）の相関
⇒ DoS攻撃の検出
⇒ 単一の特徴量を用いた手法[従来手法]との比較
- [実験2]通信量（SYN・FINフラグ）同士の相関
⇒ ポートスキャンの検出

実験結果(1)

■ HTTPサービスに対するパケットデータ

⇒ 通信量(HTTP)
⇒ 偏り(DoS攻撃)
) 相関関係の変化

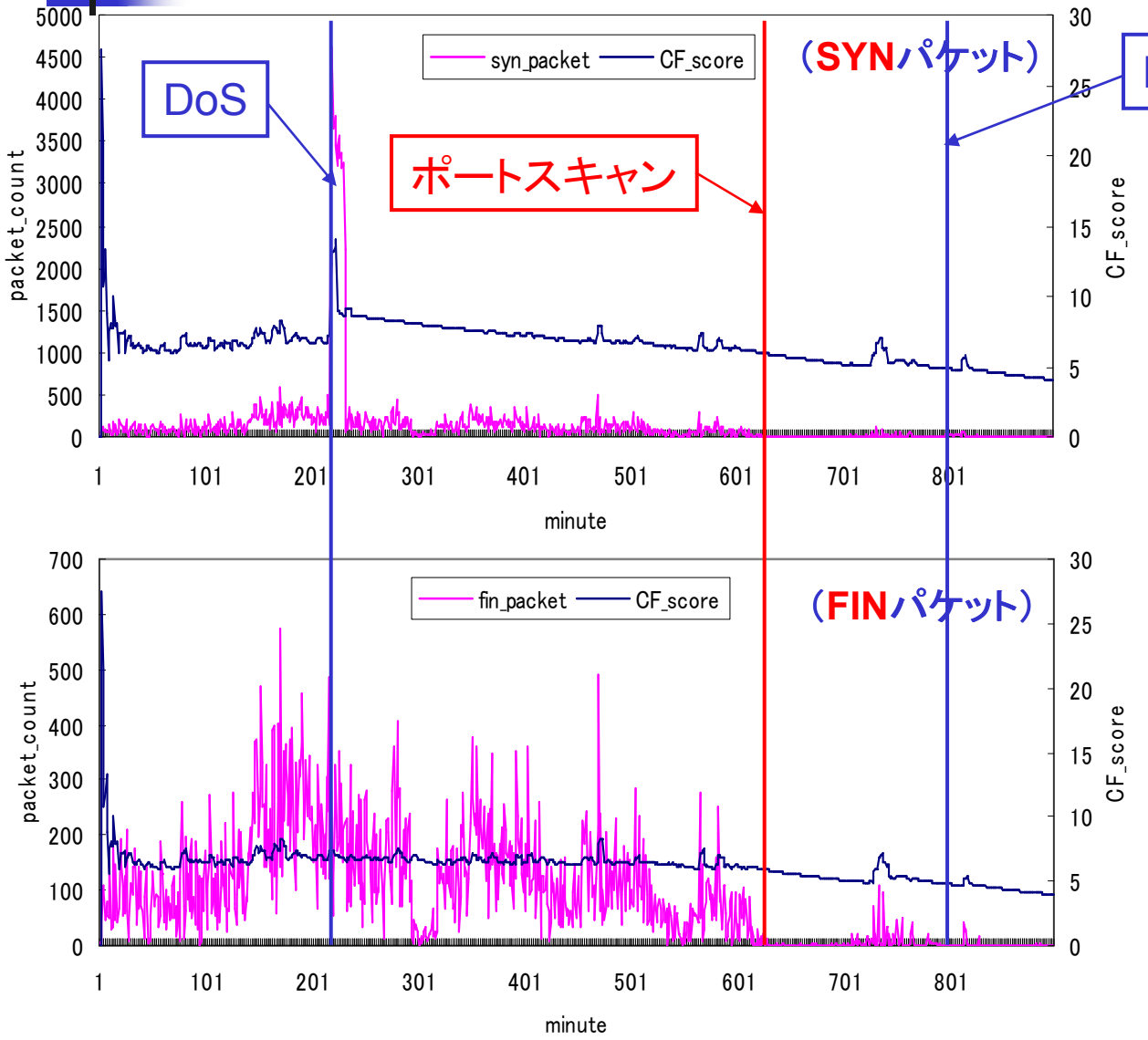


(従来手法)

(提案手法)

—:変化点スコア
—:通信量

実験結果(2) – 単一の特徴量

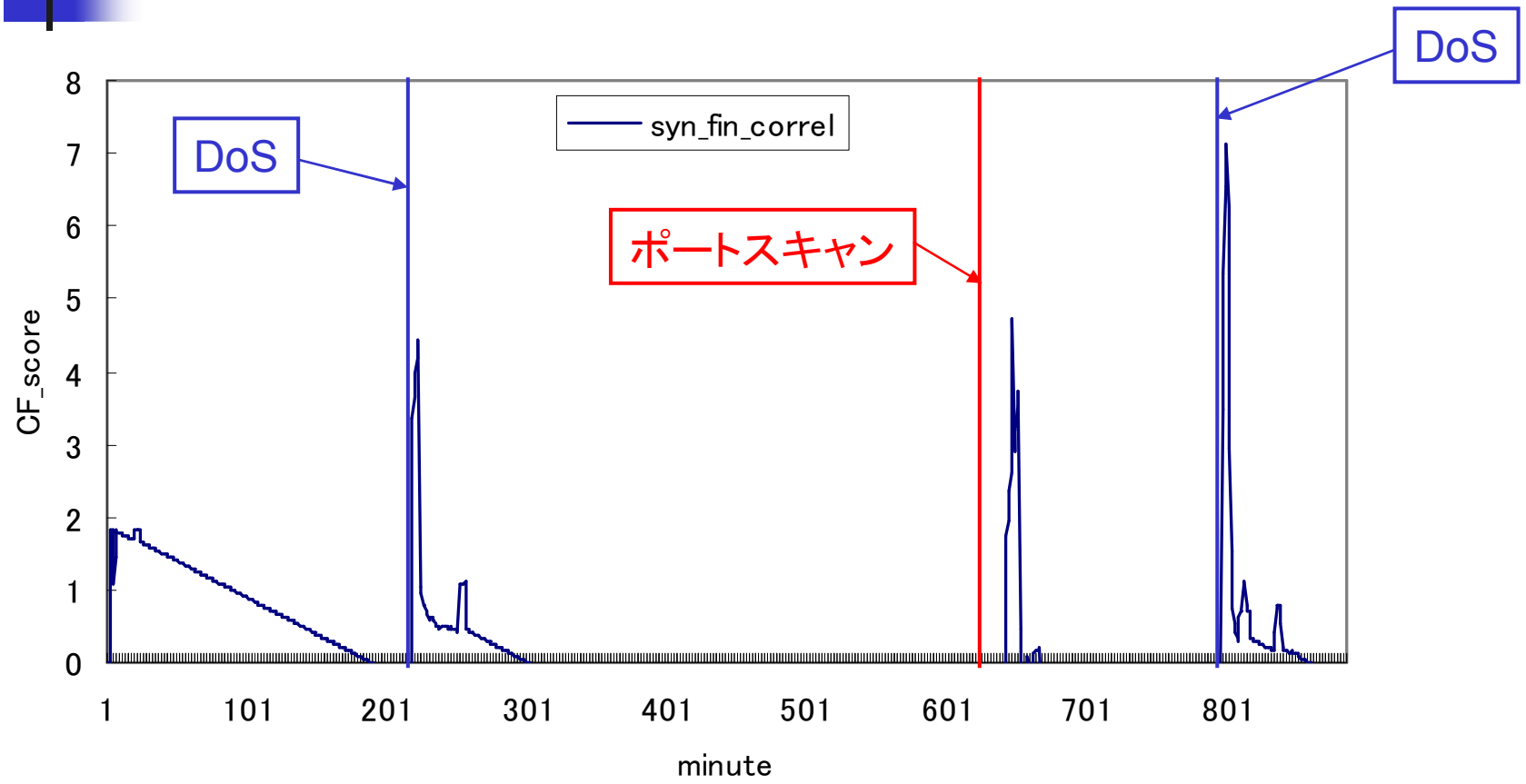


—:変化点スコア
—:通信量

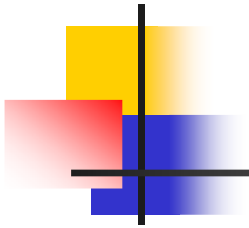
特徴量の相関

相関係数時系列データ

実験結果(2) – 提案手法

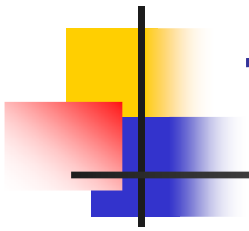


DoS攻撃もログに残らないような攻撃法をとっていた



目次

- 提案手法の説明
- 実験と考察
- まとめ



まとめ

- トラフィック特徴の相関特性を利用した手法を提案
  - トラフィック特徴を抽出
  - 特徴量時系列データにおける相関
  - 相関係数時系列データの変化点を検出

■ 特徴

⇒ 単一の特徴量ではとらえられない異常を検出

⇒ 通信方法から逸脱したアクセスを検出

■ 問題点

⇒ データ量の少ない部分で特徴を抽出し過ぎる点

今後の課題

■ 相関係数の窓サイズの適応的決定
- ⇒ 標準偏差が窓サイズにおいて0になる場合の処理
- ⇒ 標準偏差による窓の適応的決定

■ 組み合わせによる特徴量のルール改善
- ⇒ 相関関係の現れるヘッダの組み合わせを検証
- ⇒ サーバ側へのパケットに絞ることによる評価
- ⇒ サーバ負荷情報に対する特徴量の考慮

■ 変化点検出エンジンの多次元化と拡張
- ⇒ 複数の相関係数時系列データを扱えるよう拡張
- ⇒ 忘却係数と平滑化係数の適応的決定