攻撃元特定のための
トラヒックマトリクス推定手法

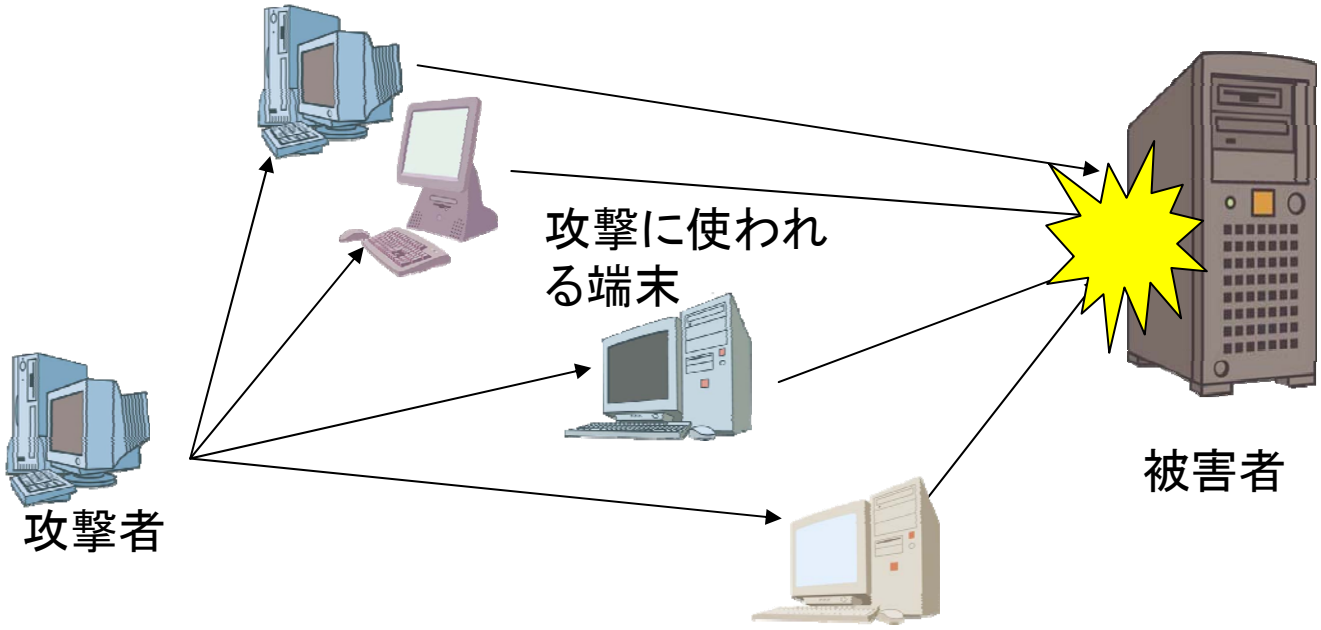
大下　裕一
大阪大学　情報科学研究科

目次

- DDoS攻撃対策と攻撃元特定の重要性
- 従来の攻撃元特定手法の問題点
- トラヒック計測による攻撃元特定
- 攻撃元特定のためのトラヒック推定手法
- 評価

DDoS 攻撃とは

- 攻撃者は複数の端末に攻撃プログラムを仕掛け大量のパケットを攻撃対象に送信する
- 近年攻撃の大規模化が問題となっている



DDoS 攻撃の対策の問題点

- DoS 攻撃では一般に攻撃パケットは偽装されている
  - 攻撃パケットを正確に識別できない
- 攻撃元が分散している
  - 一箇所での対策では、不十分

- 攻撃元を特定し、攻撃元で攻撃を遮断することが有効

既存の攻撃元特定手法

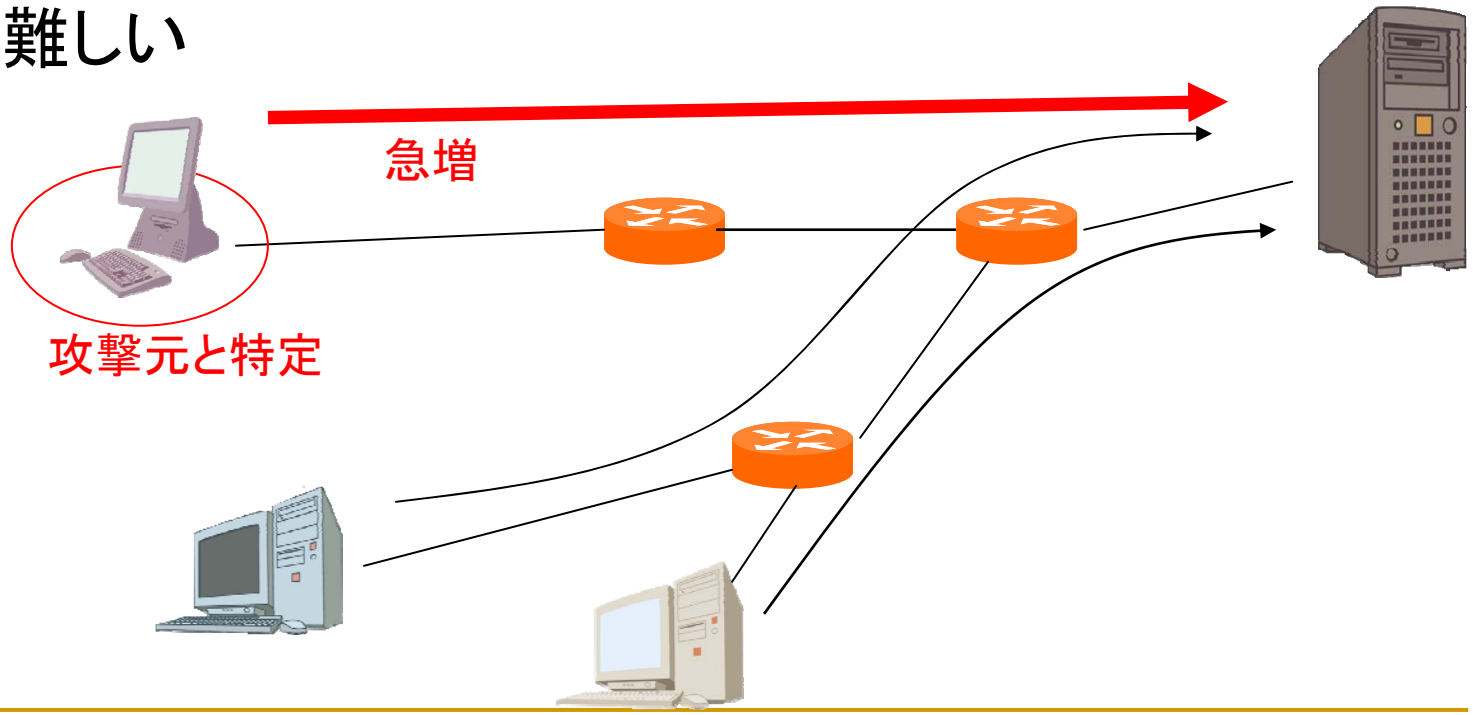
- 攻撃元を特定する既存の技術
  - ルータがパケット転送時に自身の識別情報を宛先に送る
    - ICMP traceback、パケットマーキング法
  - ルータにおいて転送したパケットのハッシュ値を保存
    - Hash-based traceback
- 問題点
  - ルータの置き換えが必要
  - 正常な通信相手と攻撃元の区別が難しい
    - パケットの送信元が特定できるものの、どのパケットが攻撃パケットかの判別は困難

トラヒック観測による攻撃元特定

- 被害者宛のトラヒックを急増させたものを攻撃元とみなすことが可能

  問題点

  送信元・宛先ごとのトラヒック量を直接観測するのは難しい

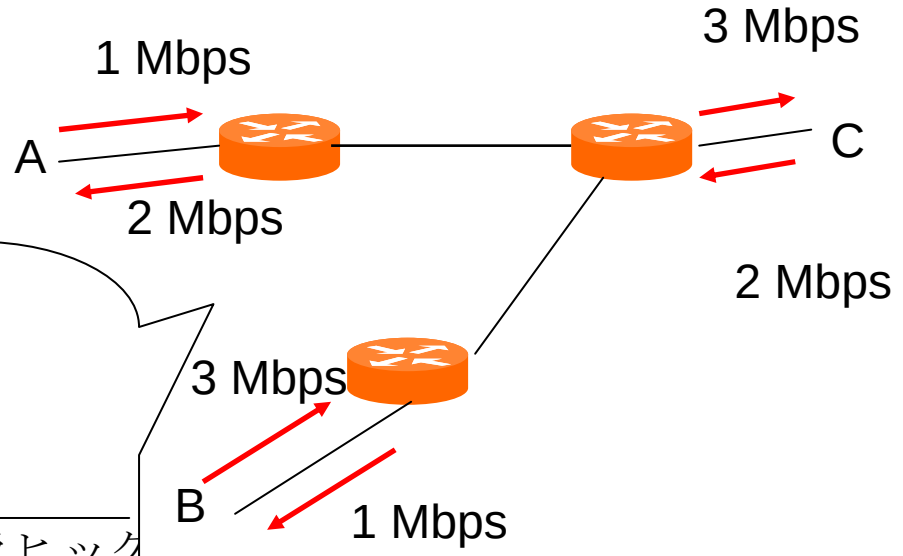


トラヒックマトリクス推定

- 送信元･宛先間のトラヒックを推定する手法
  - 入力
    - 各リンクのトラヒック量
      - SNMPをもちいて収集可能
  - 出力
    - 送信元宛先間のトラヒック量

既存のトラヒックマトリクス推定手法

- Gravity modelを用いた推定
  - 送信元・宛先間のトラヒック量は
    - 送信元、宛先、両方のトラヒック量に比例するとする



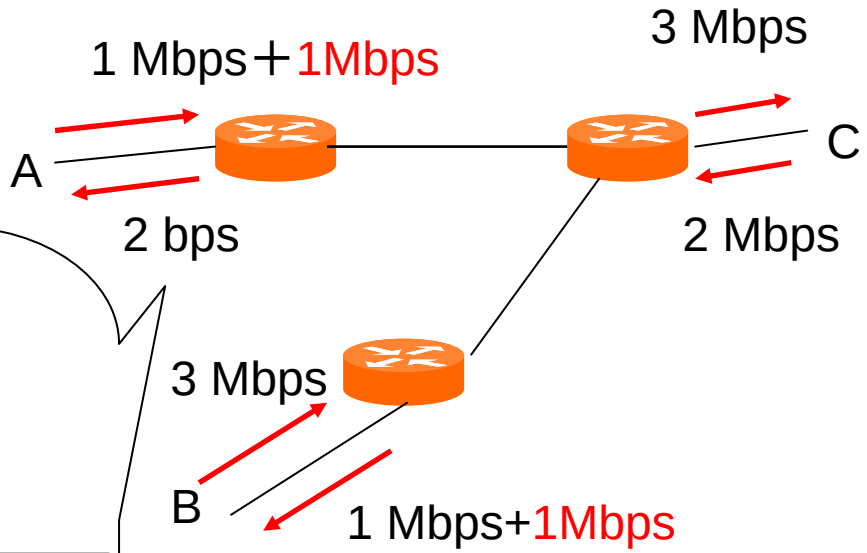
AからBへのトラヒック量は

$$A\text{からのトラヒックの総量} \times \frac{B\text{の外向きトラヒック}}{\text{Bの外向きトラヒック}+\text{Cの外向きトラヒック}}$$

$$= 1\text{Mbps} \times \frac{1\text{Mbps}}{1\text{Mbps}+3\text{Mbps}} = 0.25\text{Mbps}$$

既存のトラヒックマトリクス推定手法

- Gravity modelを用いた推定の問題点
  - トラヒックが増加した場合も、正常なトラヒックを含む全体のトラヒック量で比例配分される



AからB間のトラヒックが1bps
増えた場合でも

$$A\text{からのトラヒックの総量} \times \frac{B\text{の外向きトラヒック}}{\text{Bの外向きトラヒック}+\text{Cの外向きトラヒック}}$$

$$= 2\text{Mbps} \times \frac{2\text{Mbps}}{2\text{Mbps}+3\text{Mbps}} = 0.8\text{Mbps}$$

0.55 bps しか増加してない
ように推定される

研究の目的

- 攻撃元特定のためのトラヒックマトリクス推定手法
  - 攻撃元特定に有効となる、送信元・宛先ごとのトラヒック増加量を推定可能な手法
- 指針
  - リンクごとのトラヒックの増加量を用いた推定手法
    - 正常なトラヒックの影響を除外可能
    - トラヒックが急増した送信元からのトラヒックほど増加量が大きいと判断可能

トラヒック増加量推定の手順

- 各リンクごとに現在のトラヒック量の、正常なトラヒック量の平均との差を求める

$$G_n = X_n - \overline{X_n}$$

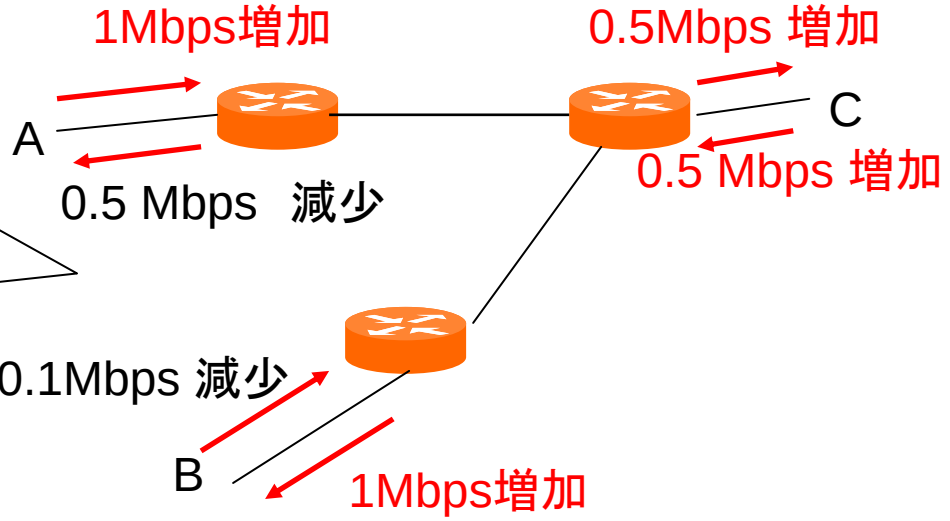
  - X_nはすべてのリンクの時刻 n でのトラヒック量をあらわす行列
  - $\overline{X_n}$は各リンクの時刻n以前の正常なトラヒック量の平均を表す行列
  - G_nは現在のトラヒック量と正常なトラヒック量の平均の差を表す行列
- 各リンクの正常なトラヒック量との差を入力とし、Gravity model を用いて推定を行う
- 途中の経路のトラヒック量の情報を反映させる
- 推定結果をもとに $\overline{X_{n+1}}$を定める

Gravity Model を用いた増加量推定

- 増加量をもとに推定
  - 入り口・出口ともにトラヒックが増加している場合
    - $$\text{入り口の増加量} \times \frac{\text{出口の増加量}}{\text{トラヒックが増加した出口の増加量の合計}}$$
  - 入り口・出口ともに減少している場合
    - $$-\left|\text{入り口の増加量} \times \frac{\text{出口の増加量}}{\text{トラヒックが減少した出口の増加量の合計}}\right|$$
  - それ以外
    - 0

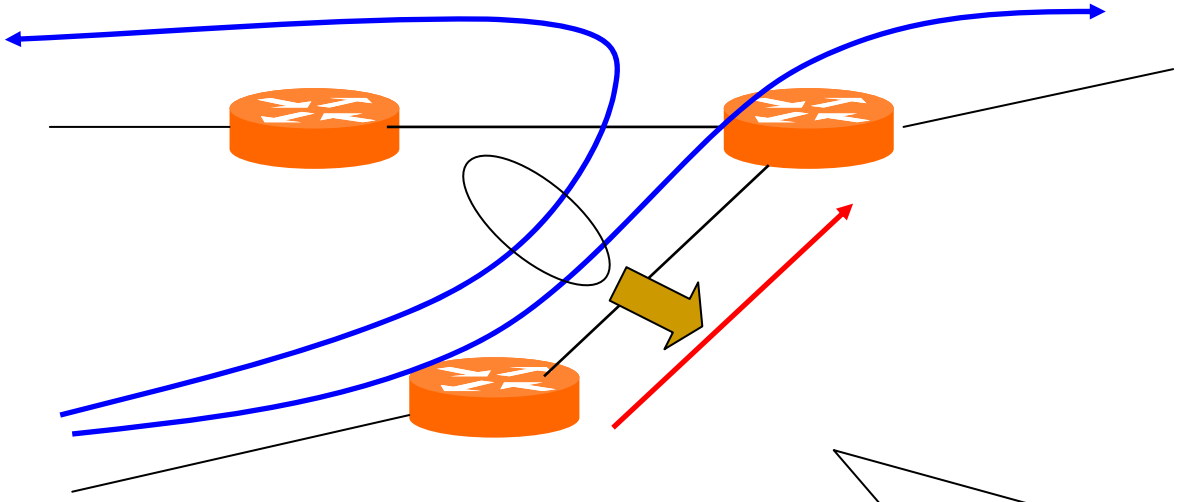
AからBへのトラヒック増加量は

$$1\text{Mbps} \times \frac{1\text{Mbps}}{1\text{Mbps}+0.5\text{Mbps}} = 0.66\text{Mbps}$$



トラヒックが急増している送信元ほど増加量が多いと推定される

各リンクのトラヒック量と送信元・宛先間のトラヒック量の関係



各リンクのトラヒック量はそこを経由するトラヒックの和となる

各リンクのトラヒック量と送信元・宛先間のトラヒック量の関係

- 各リンクのトラヒック量はそこを経由するトラヒックの和となる

$$G = AF$$

  - F は送信元・宛先間のトラヒック増加量を要素とするベクトル
  - G はリンクで観測されたトラヒック増加量を要素とするベクトル
  - A は要素 $a_{(i,j),k}$ が i から j へのトラヒックがリンク k を経由する場合は 1 それ以外は 0

途中のリンクのトラヒック情報の反映

- $G = AF$ の条件がなりたつように、Gravity modelで求めた結果に補正を加える
- 方法
  - 以下の計算を行い F を最終的な推定結果とする

$$F = F' + A^{-1}(G - AF')$$

  - F' はGravity Modelでの推定結果
  - A^{-1}はルーティング行列 A の擬似逆行列
  - G は各リンクで観測されたトラヒック増加量

正常なトラヒック量

- 定め方の方針
  - 各リンクごとの基準値を過去に観測されたトラヒック量をもとにした値に定める
    - トラヒックの急増の影響を受けない定め方
      - トラヒックの急増の影響が反映されてしまうと、その後の検出に悪影響を与える
    - リンク間での整合性のある値の定め方
      - $G = AF$ の関係がなりたつ必要がある

通常のトラヒック量の定め方

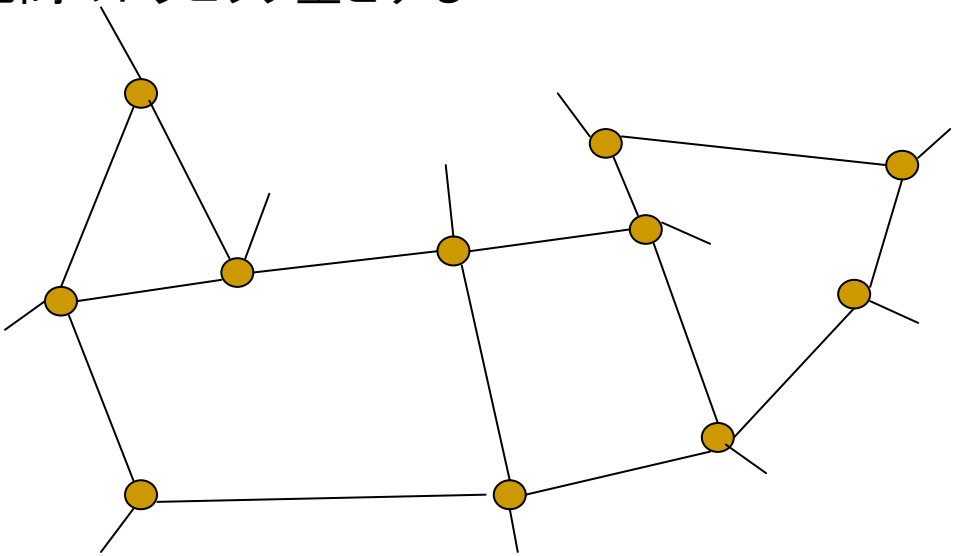
- 推定されたトラヒック量から、急増していない箇所のみを抜き出す
  - 急増していない箇所を抜き出した行列 $\hat{F}_n$ の要素を次のように定義
    - トラヒックが急増したフローに対する要素は0
    - それ以外の要素は推定されたトラヒック増加量と等しい
- すべてのリンクに反映し重み付き平均を求める

$$\overline{X}_{n+1} = \alpha(\overline{X}_n + A\hat{F}_n) + (1-\alpha)\overline{X}_n \quad (0 < \alpha < 1)$$

  - A は各送信元・宛先間のトラヒック量のリンクへのマッピングを表す行列
    - 要素 $a_{(i,j),k}$ が i から j へのトラヒックがリンク k を経由する場合は1 それ以外は0

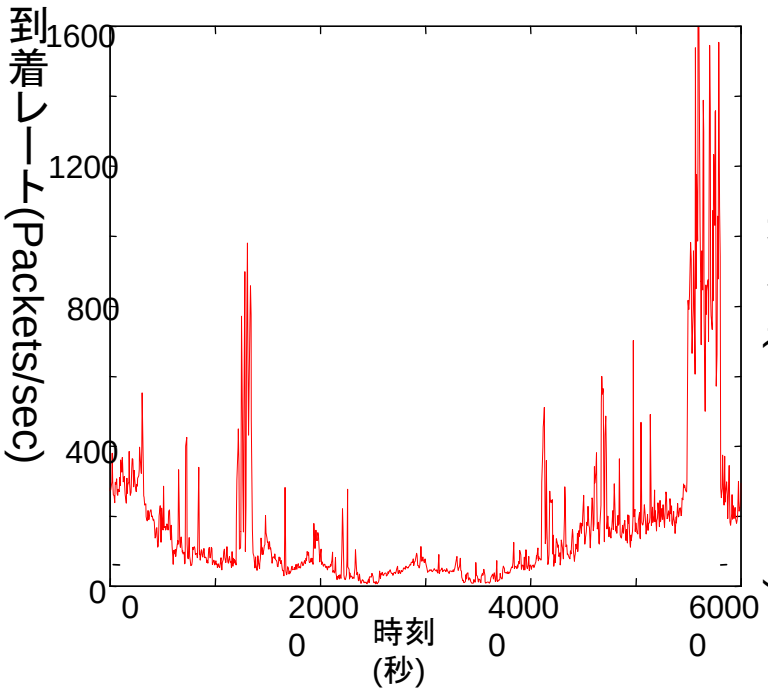
評価方法

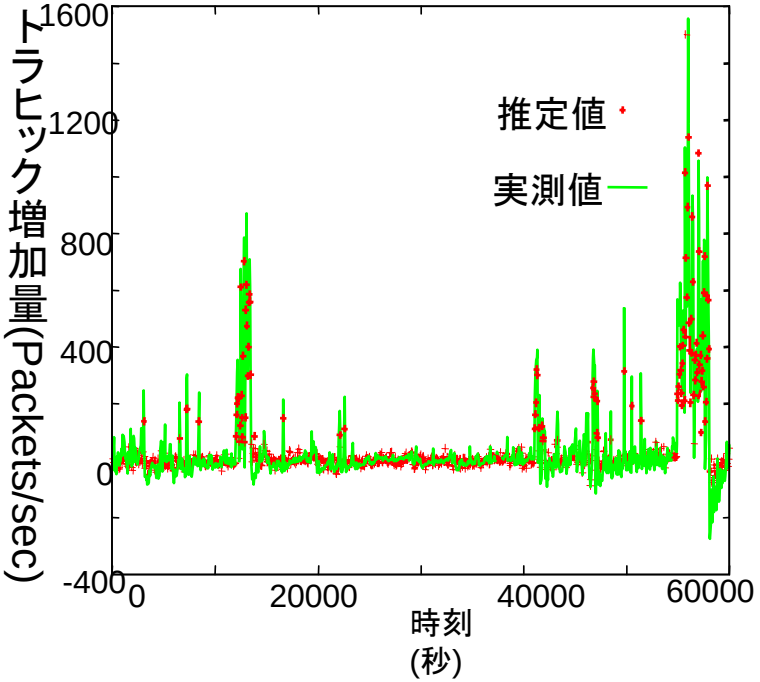
- シミュレーションによる評価
  - トポロジ
    - Abileneのバックボーントポロジ
  - トラヒックデータ
    - 大阪大学のゲートウェイでの観測されたデータ
      - 送信元アドレス上位16ビットごとに到着レートを観測
      - 観測されたアドレスを110個の組にわけ、その組の到着レートを送信元宛先間のトラヒック量とする



トラヒック増加量の時間変化と推定結果

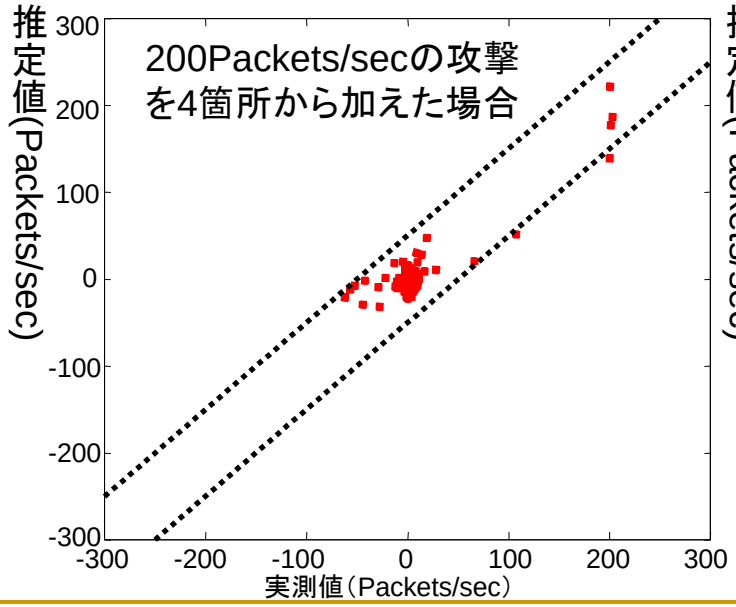
- あるフローについてトラヒックの増加量とその推定値を調べた
- トラヒックの増加量を調べることにより、日常的なトラヒックの変化の影響を除外可能
- トラヒックが急増した箇所は急増したと推定可能

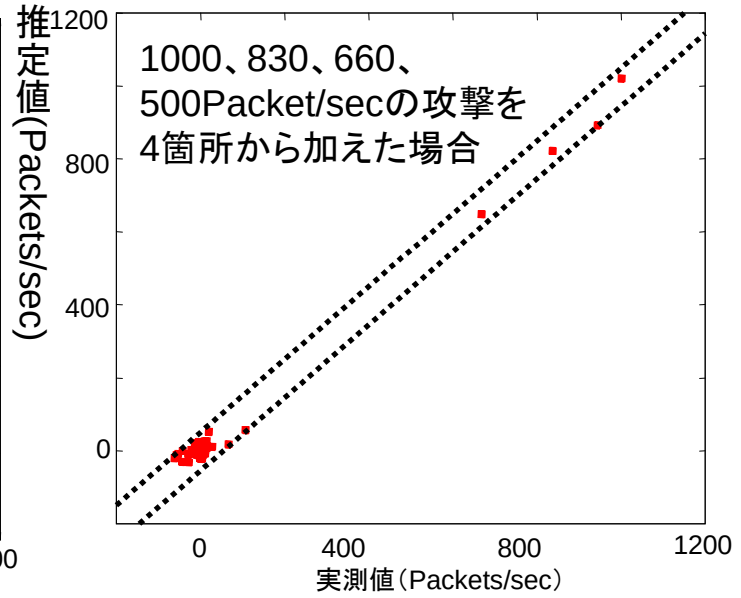




推定の精度

- 攻撃を加えた時間帯の推定精度を調べた
- 攻撃のレートが大きい場合でも
  ±50Packets/sec以内の誤差で推定可能





まとめ

- 攻撃元特定のためのトラヒックマトリクス推定手法を提案
- 提案手法では、トラヒックの増加量に特化することにより、攻撃元特定に有効な情報を提供可能
- シミュレーションにより、提案手法が正確にトラヒック増加量を推定可能なことを確認