ダークネットトラフィックデータの解析による サブネットの脆弱性判定に関する研究

コンピュータセキュリティシンポジウム 2013
2013年10月23日

西風 宗典†　班 涛‡　小澤 誠一†

†神戸大学大学院　工学研究科
‡独立行政法人情報通信研究機構

研究目的

ダークネットへのポート別トラフィックに基づくサブネットの脆弱性判定

マルウェア感染状況の把握によるサブネット利用者への警告

- ネットワークに接続するコンピュータの増加
- 通信の高速化による急速なマルウェア感染拡大
- 新種，亜種の出現や自己の存在を隠すマルウェアの登場により検出が難化

マルウェア感染状況の把握が困難化

- ダークネットの通信トラフィックを用いる.
- 感染単位としてサブネットを用いる.

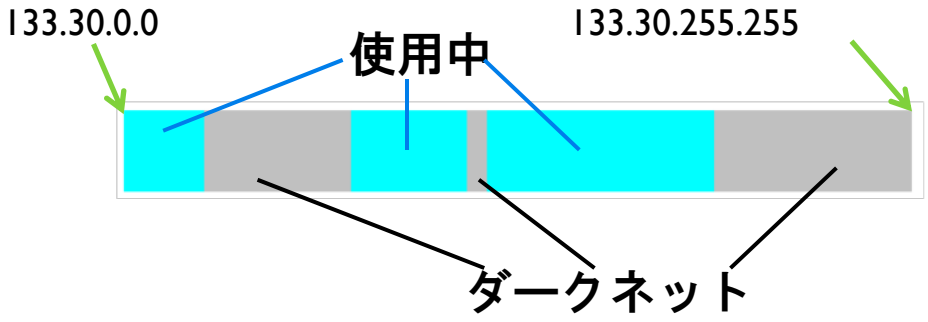
ダークネット

実際にはコンピュータが存在しないIPアドレス

- ダークネットを宛先としたパケットとは…
  - マルウェアによるネットワークスキャン，感染行為
  - DDos攻撃のバックスキャッタ
  - 設定ミス

よって

- ダークネットへの通信は主にマルウェアによるものと考えられる



サブネットの脆弱性判定

IPv4のアドレス空間をサブネットに分割

↓

各サブネットの通信トラフィックを特徴ベクトルに変換

↓

生成した特徴ベクトルを用いてクラスタリング

↓

クラスタリング結果の可視化，解析

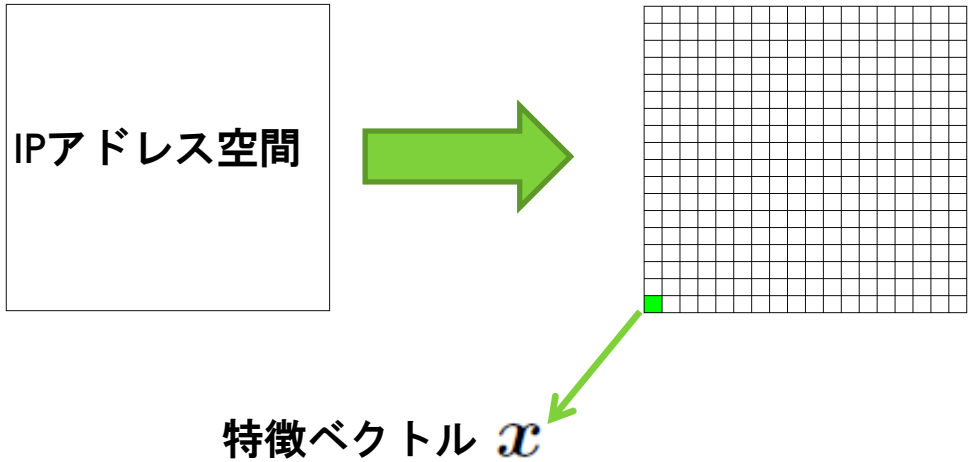
↓

サブネットの脆弱性判定

特徴ベクトル（I）

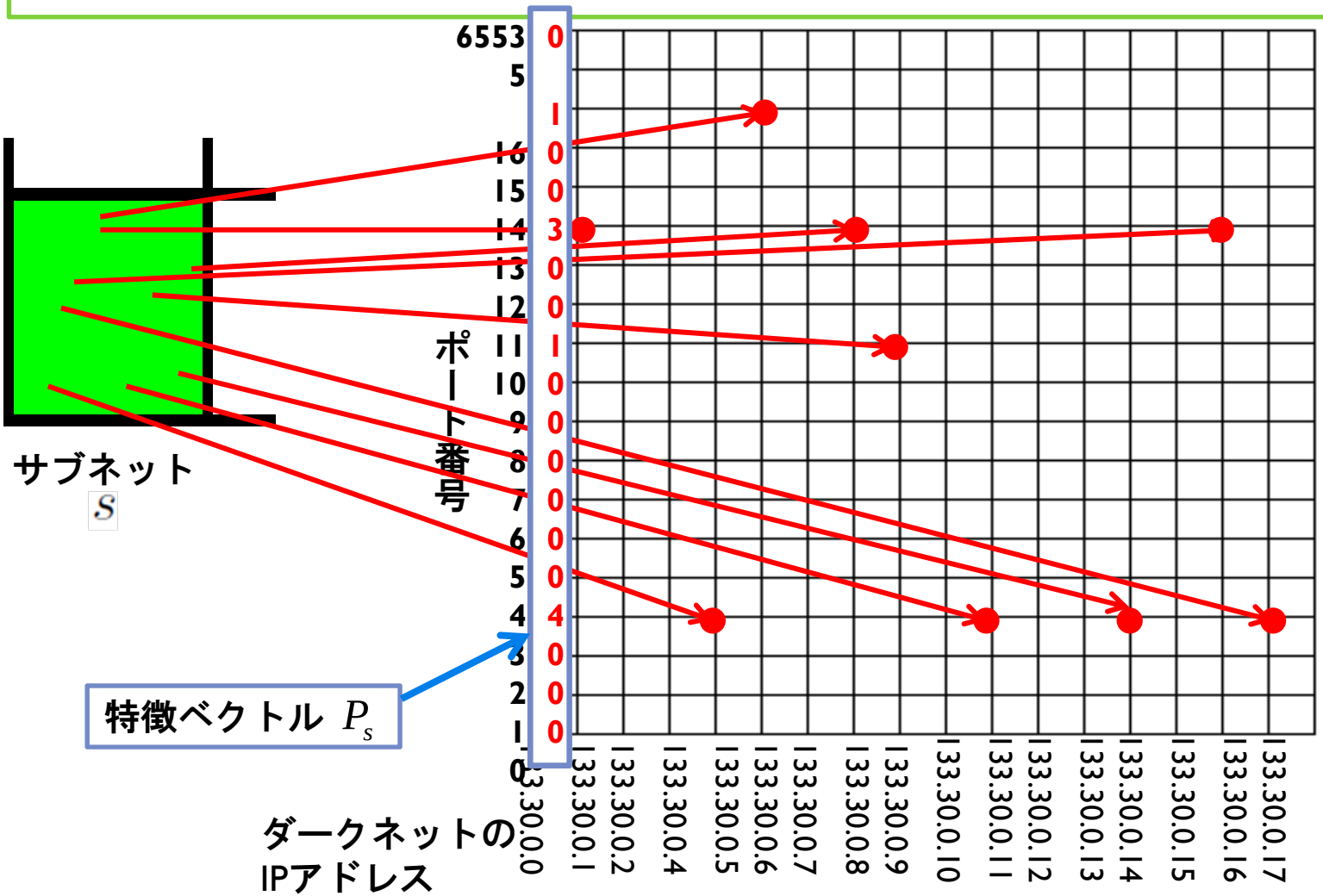
IPアドレス空間をサブネットに分割

約42億あるIPv4空間をサブネットに分割し，各サブネットのポート別通信トラフィックを特徴ベクトルに変換



特徴ベクトル（2）

各サブネットの通信トラフィック(宛先IPアドレスとポート)特徴ベクトルに変換



特徴ベクトルの正規化

各サブネットの通信トラフィック(宛先IPアドレスとポート)特徴ベクトルに変換

ベクトル P_s 0 1 0 0 3 0 0 1 0 0 0 0 0 0 4 0 0 0

ポート別攻撃先IP数

特徴ベクトル $x_s = \frac{P_s}{N_s}\log(N_s+1)$

理 由

正規化処理

ポート別攻撃先IP総数 N_sで正規化

攻撃強度 $\log(N_s+1)$で重みづけ

攻撃先IP数の絶対値そのものに意味はなく，ポート別の攻撃先IP数の分布が重要

攻撃先IP総数の多いサブネットを重視

攻撃パターンのクラスタリング

最長距離法によるクラスタリング

最長距離法

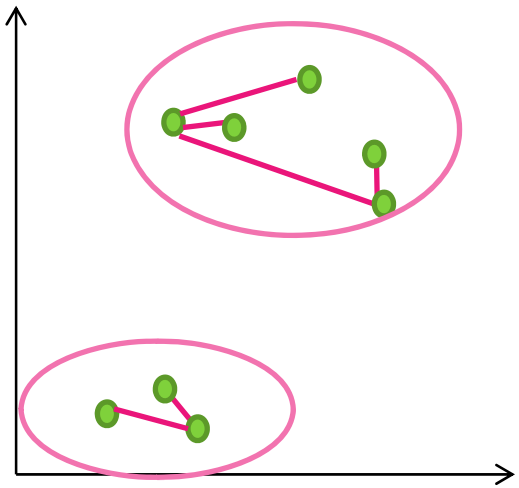
クラスタ間距離の近いものから統合していく手法の一種

クラスタ間距離

クラスタに属する
ベクトル間距離の最大値

$$\mathrm{d}_1(C_i, C_j) = \max_{a,b}(\mathrm{d}_2(\boldsymbol{x}_{ia}, \boldsymbol{x}_{jb}))$$

$$\boldsymbol{x}_{ia} \in C_i, \boldsymbol{x}_{jb} \in C_j$$



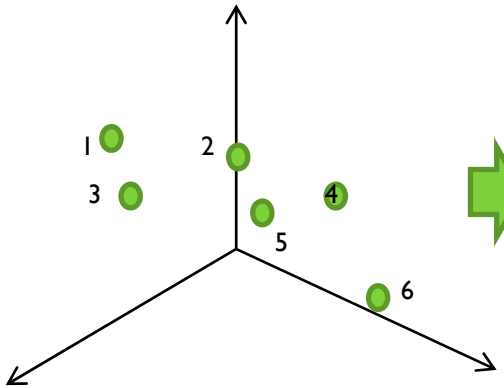
クラスタリング結果の可視化

Isomapによる次元削減

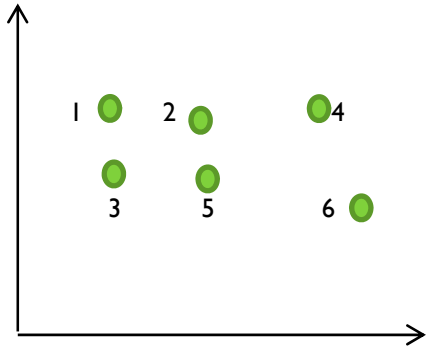
ポート数(65536)次元の特徴ベクトルから次元削減をし，可視化する.

Isomap

近傍のベクトル同士の距離の求め，その距離情報の誤差が最小となるよう低次元にマッピングを行う手法で非線形に次元圧縮.
何次元に圧縮するかにより誤差が変化する.



	1	2	3	4	5	6
1	0	1.024	1.852	2.233	3.784	4.014
2	1.024	0	1.998	1.982	3.547	3.722
3	1.852	1.998	0	0.927	1.587	2.587
4	2.233	1.982	0.927	0	1.257	1.654
5	3.784	3.547	1.587	1.257	0	1.374
6	4.014	3.722	2.587	1.654	1.374	0



実験方法

- IPアドレス空間は2^{16}（65536）に分割
  クラスB（国の機関や大企業の大きさ）

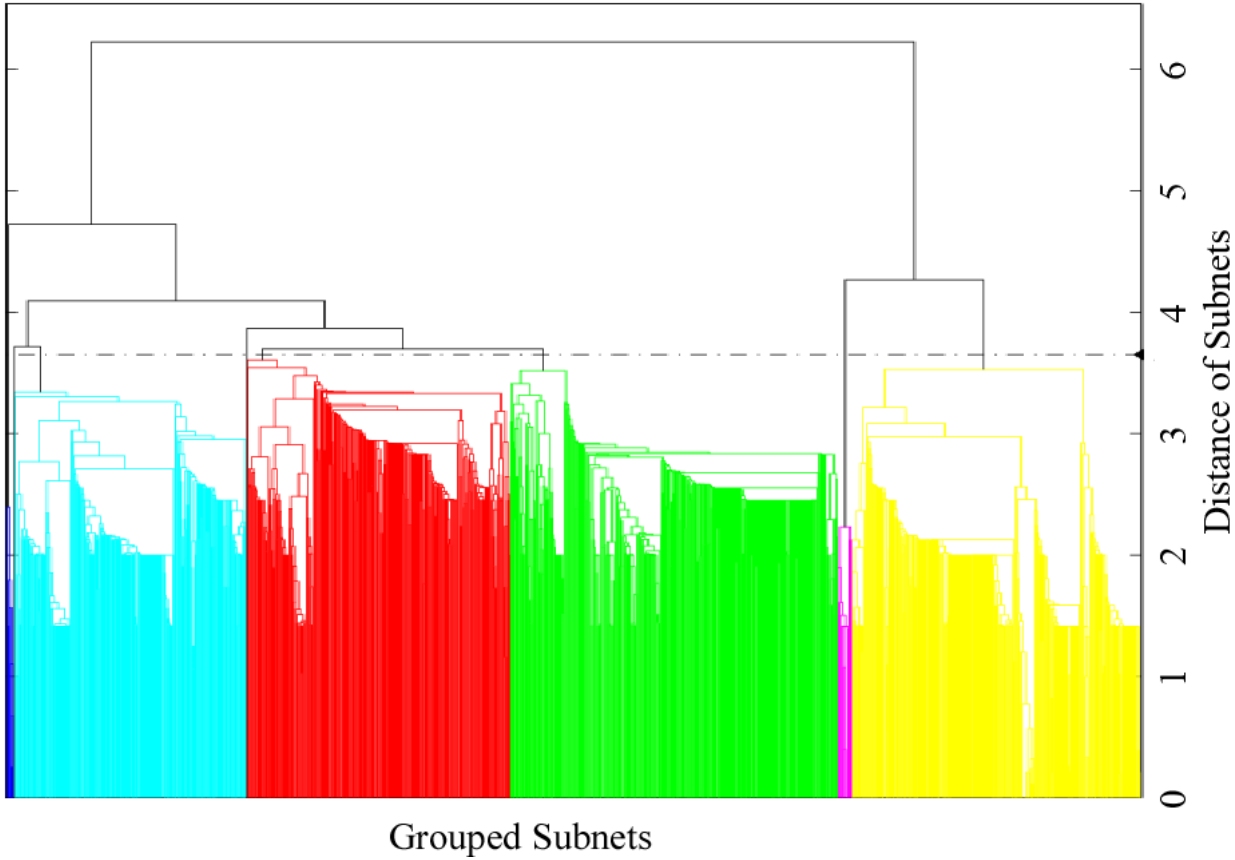
- 使用したデータ
  ダークネットIPアドレス4096個に届いた，2011年1月のデータ（約686万パケット）

- 4個以上のIPからの送信を確認した1296のサブネットを使用

- Isomapのベクトル同士の距離は10近傍まで使用

攻撃パターンによるクラスタリング結果

- 最長距離法によるクラスタリング結果
- 距離差による閾値設け，色分けして表示（クラスタ数：8）

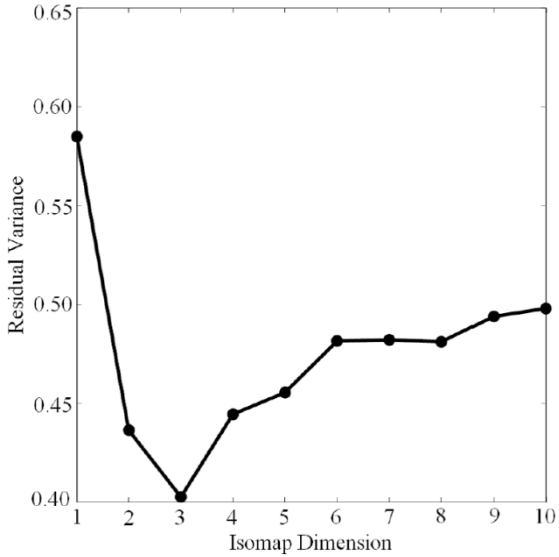


クラスタリング結果の可視化

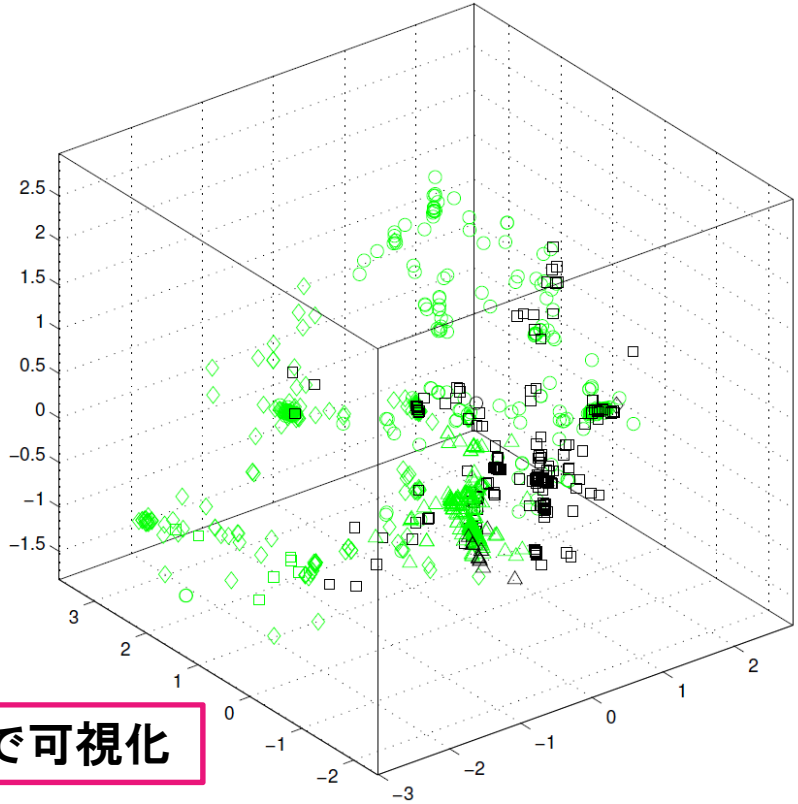
Isomapによる次元削減での可視化

ポート数長65536次元から可視化できる次元に変換

変換後の次元数と元データとの誤差



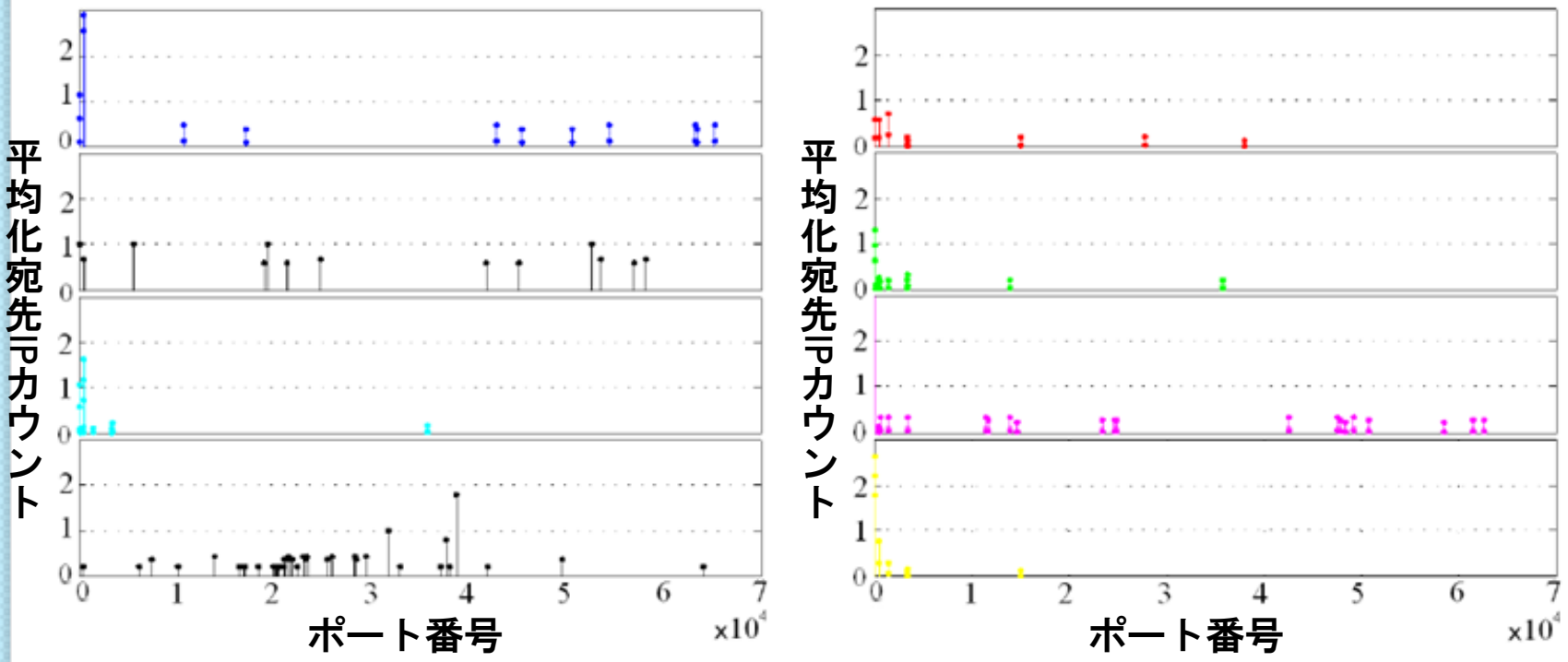
3次元での可視化



3次元のとき誤差が小さい → 3次元で可視化

クラスタ中心の攻撃パターン

各クラスタの特徴ベクトルの[平均]と[平均+標準偏差]



攻撃先ポートのランキング

各クラスタと攻撃したポート

ポート＼クラスタ	1	2	3	4	5	6	7	8
1位	445	23	445	38732	1433	23	23	23
2位	23	5580	23	31750	445	3389	1433	445
3位	10736	19338	3389	37659	23	445	586	1433
4位	42866	52639	35771	13882	14979	1433	3389	3389
5位	54434	445	1433	21490	3306	3306	11462	14979

クラスタにより主に攻撃しているポートが異なっている．

- 代表的なポートとマルウェアの関係
- 23　telnet．リモートサーバを端末から操作するプログラム．セキュリティに問題
- 445　ファイル共有用のポート．脆弱性があり，Conficker が攻撃
- 1433　Microsoft SQL Server の通信用．SQL Slammer が攻撃
- 3389　リモートデスクトップ接続．Morto が攻撃
- 3306　MySQL（リレーショナルデータベース管理システム）．MySQL Bot などが攻撃

まとめと課題

- まとめ
- 本研究では，マルウェア感染状況の把握によるサブネット脆弱性判定を目的とした．マルウェア感染状況の把握のためにサブネットごとに通信トラフィックを特徴ベクトルに変換し，クラスタリングを行う手法を提案した．
- 実験結果から各クラスタに含まれるサブネットにおいてどのマルウェアが支配的かを考察した．

- 課題
- サブネット脆弱性判定
  - IP空間，または地図と関連付け，それによる感染状況を把握
  - 各クラスタの危険度の判定
- マルウェアの挙動によるクラスタリング
- さらに詳細なサブネットの分割
- 他の期間のデータでの有効性の検証，時系列を考慮した解析

ご清聴ありがとうございました。