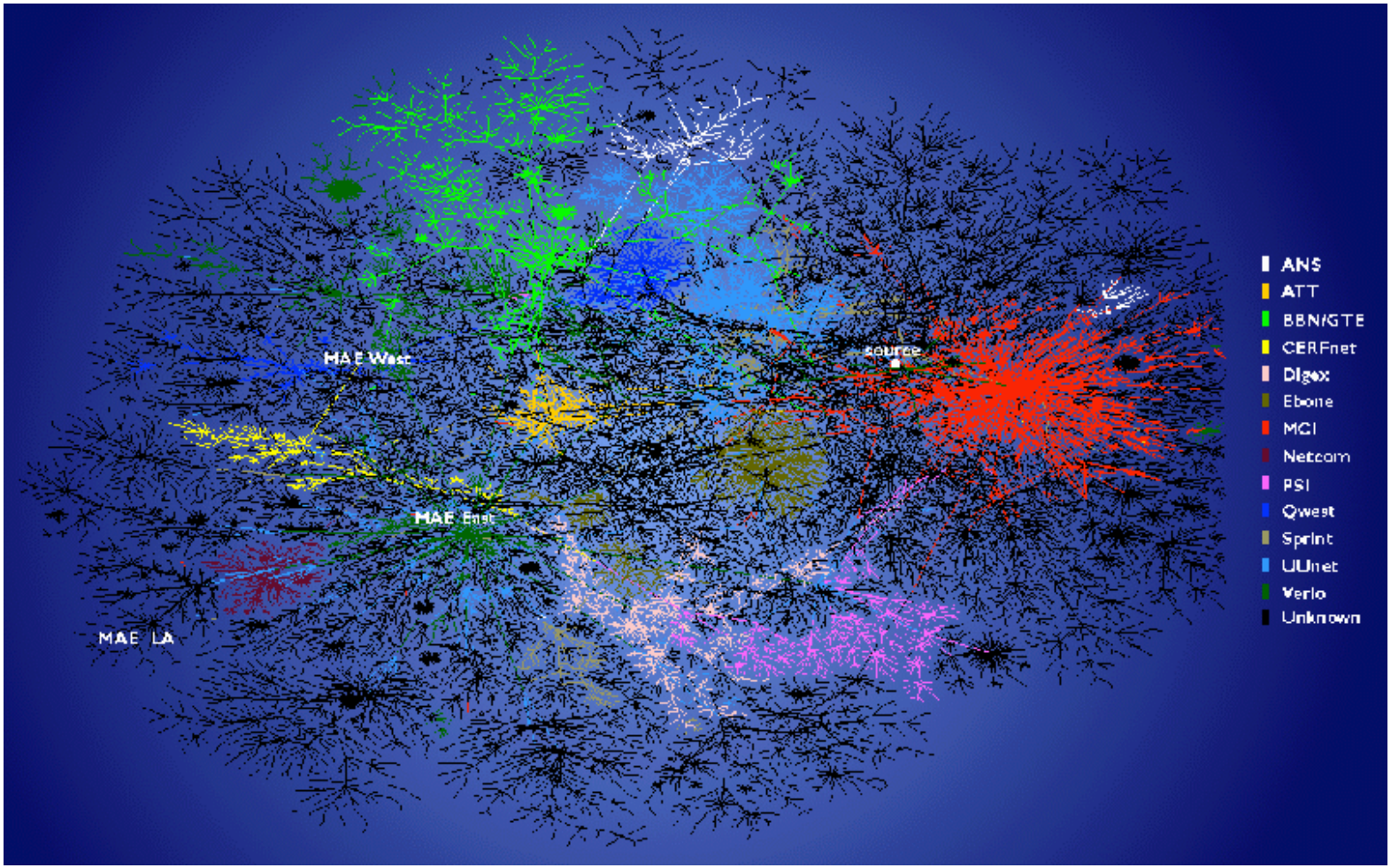
研究100連発
インターネットの諸々を測定する

国立情報学研究所
福 田　健 介

National Institute of Informatics



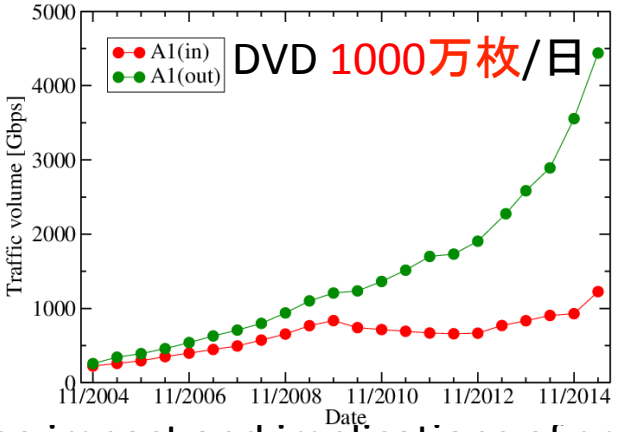
B.Huffakar, E.Nemeth, k.claffy,
Otter: A general-purpose network visualization tool Proc. INET99, 1999

20のトピック

- 日本の固定インターネットのデータ量は?
- スマホで無線LANどのくらい使ってる?
- インターネットには何が流れている?
- このコンピュータは何をしている?
- スマホで使っているアプリをあてる
- インターネット上に囮を作る
- 普通と違うのが異常
- 画像認識で異常を探す
- 三人よれば文殊の知恵
- 人手で異常検出の精度を上げるには
- 分割して統治せよ
- 有名サイト似た偽サイトを探す
- 迷惑メールはどうやって送られてくる?
- オンラインRPGでの悪い人探し
- ビルはどれだ省エネできる?
- 東日本大震災時の学術ネットワーク
- みんなの力でスマホの診断
- 大規模サーバはどれだけ効率的?
- トポロジを意識した負荷分散
- 集合知による新しいインターネットセンサー

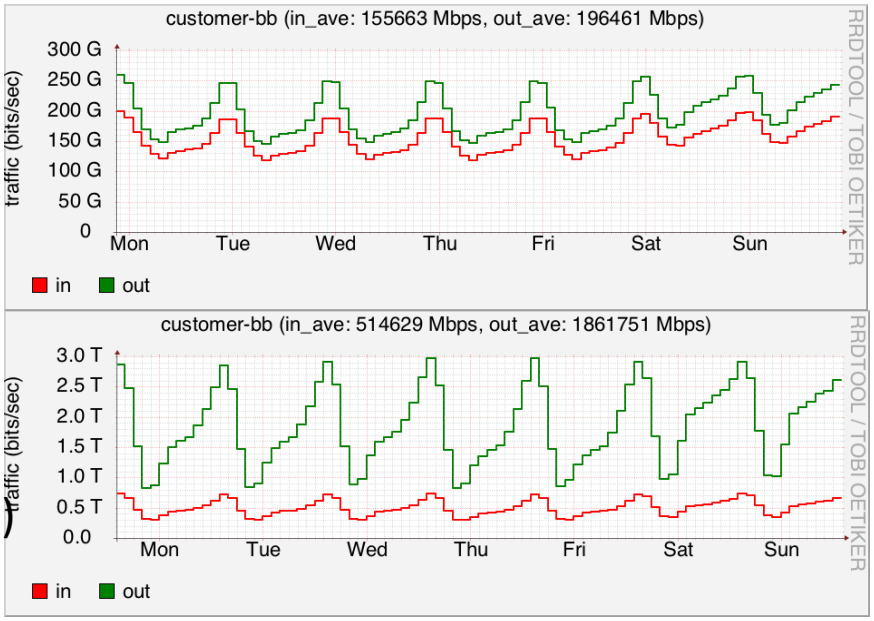
1. 日本のインターネットのデータ量は?

- 国内6ISP，総務省と2004年から国内ブロードバンドインターネットトラフィックの収集
  - 総トラフィック量の推定
  - 使用パターンの変化



2005 (p2p)

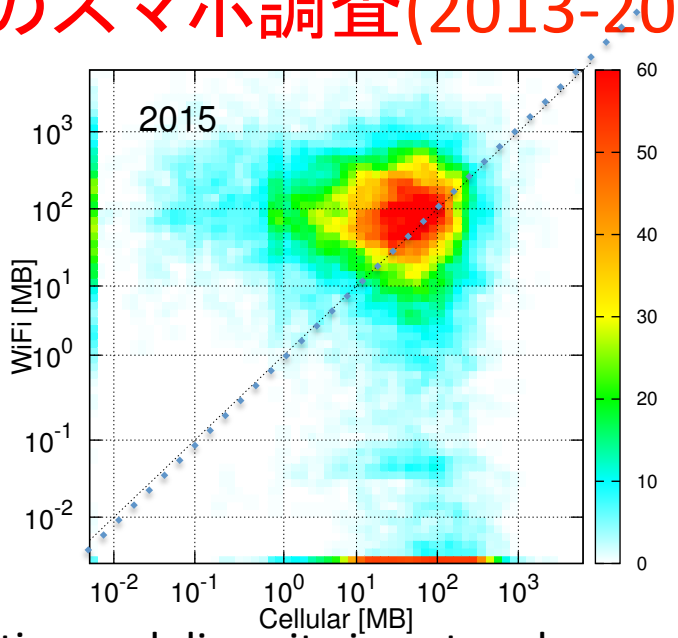
2015 (video)

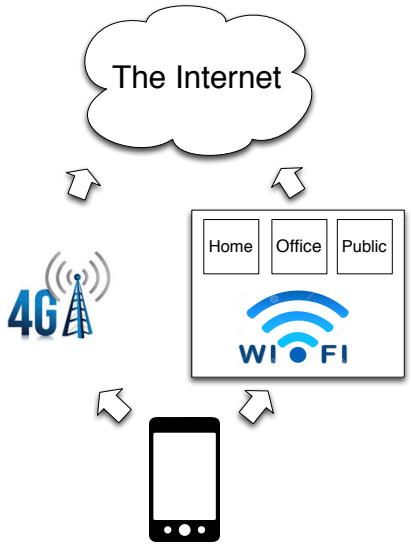


The impact and implications of growth in residential user-to-user traffic (SIGCOMM'06)

2.スマホで無線LANをどのくらい使ってる?

- 携帯キャリア・固定キャリアともにわからない
- 東京圏1500ユーザのスマホ調査(2013-2015)
  - アプリケーション
  - トラフィック量
  - オフロード率

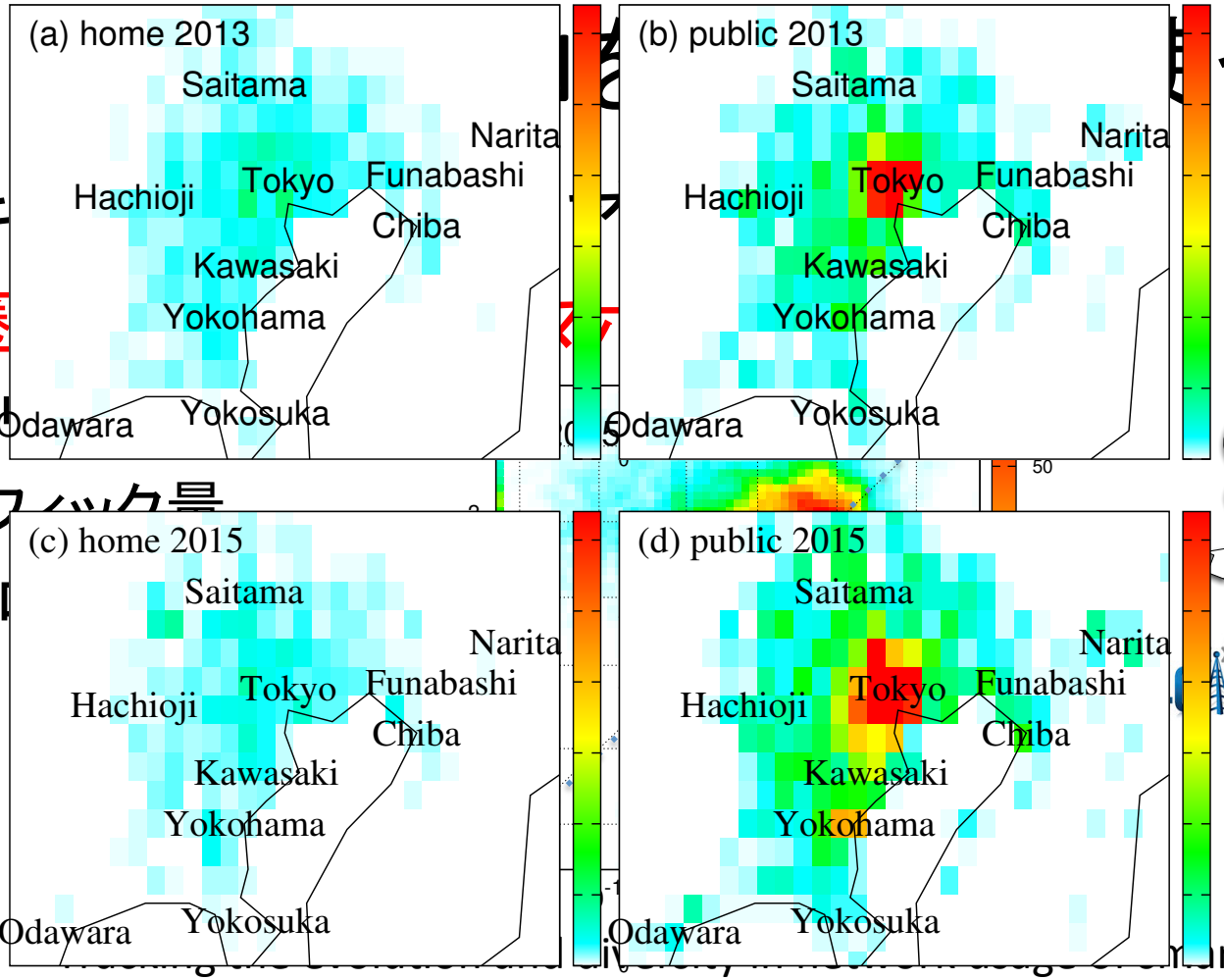


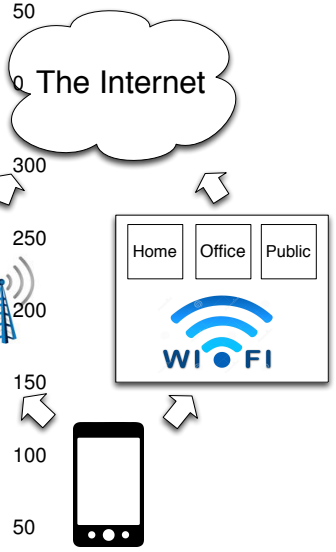


Tracking the evolution and diversity in network usage of smartphones (IMC'15)

2.スマ　　ってる?

- 携帯キ
- 東京圏
  - アプ
  - トラフィック量
  - オフ

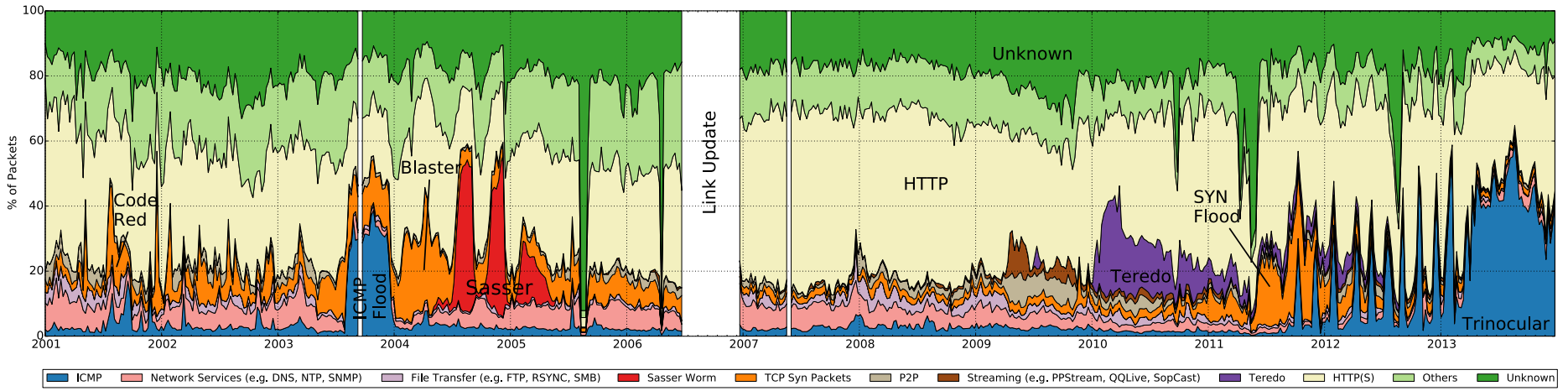




martphones (IMC'15)

3.インターネットには何が流れている?

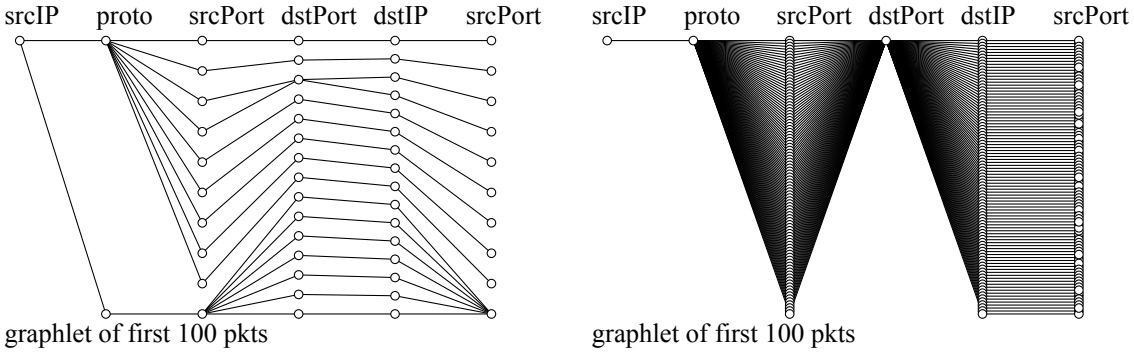
- アプリ推定はプライバシや暗号化の問題のため難しい
- バックボーントラフィックでの機械学習を用いたアプリ推定



A streaming flow-based technique for traffic classification applied to 12+1 years of internet (Tele. Sys. in 2015)

4.このコンピュータは何をしている?

- ホストの提供サービスをトラフィックから推定
  - 通信パターンによる分類木
  - グラフレット

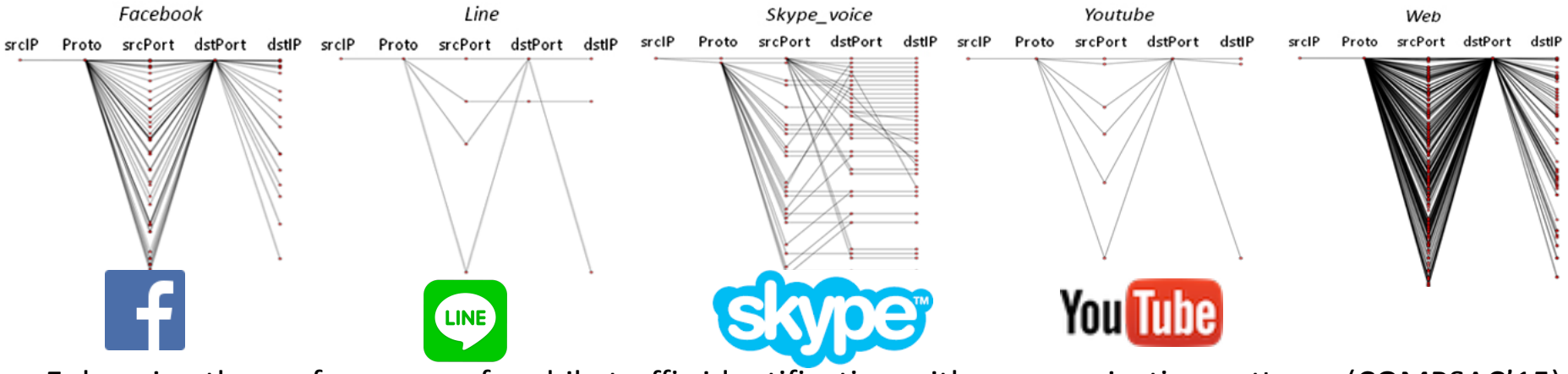


P2Pソフト　　スキャン(攻撃)

Synoptic Graphlet: Bridging the gap between supervised and unsupervised profiling of host-level network traffic (ToN in 2013)

5.スマホで使っているアプリを当てる

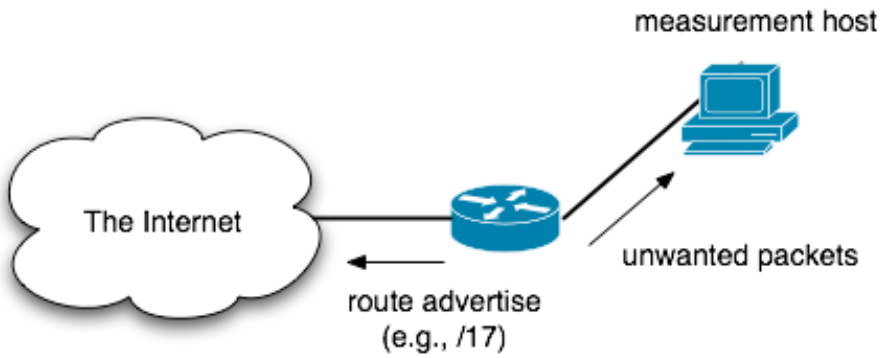
- 教師有り機械学習によるスマホアプリ推定
  - パケットサイズ
  - 通信パターン



Enhancing the performance of mobile traffic identification with communication patterns (COMPSAC'15)

6.インターネット上に囮を作る

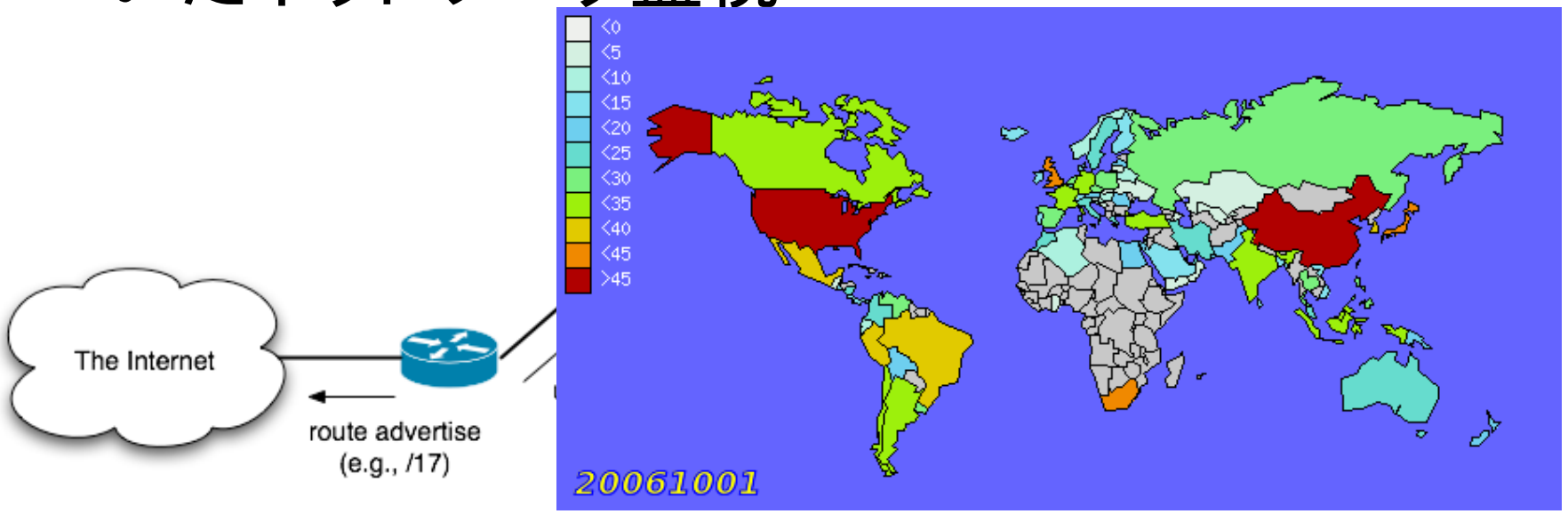
- ダークネット(ホストのいないネットワーク)を用いたネットワーク監視



Correlation among piecewise unwanted traffic time series (Globecom'08)

6.インターネット上に囮を作る

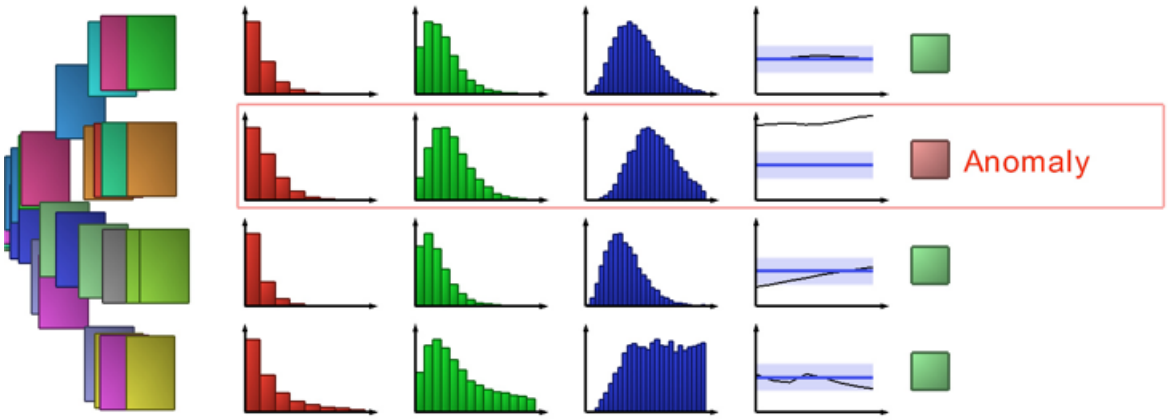
- ダークネット(ホストのいないネットワーク)を用いたネットワーク監視



Correlation among piecewise unwanted traffic time series (Globecom'08)

7.普通と違うのが異常

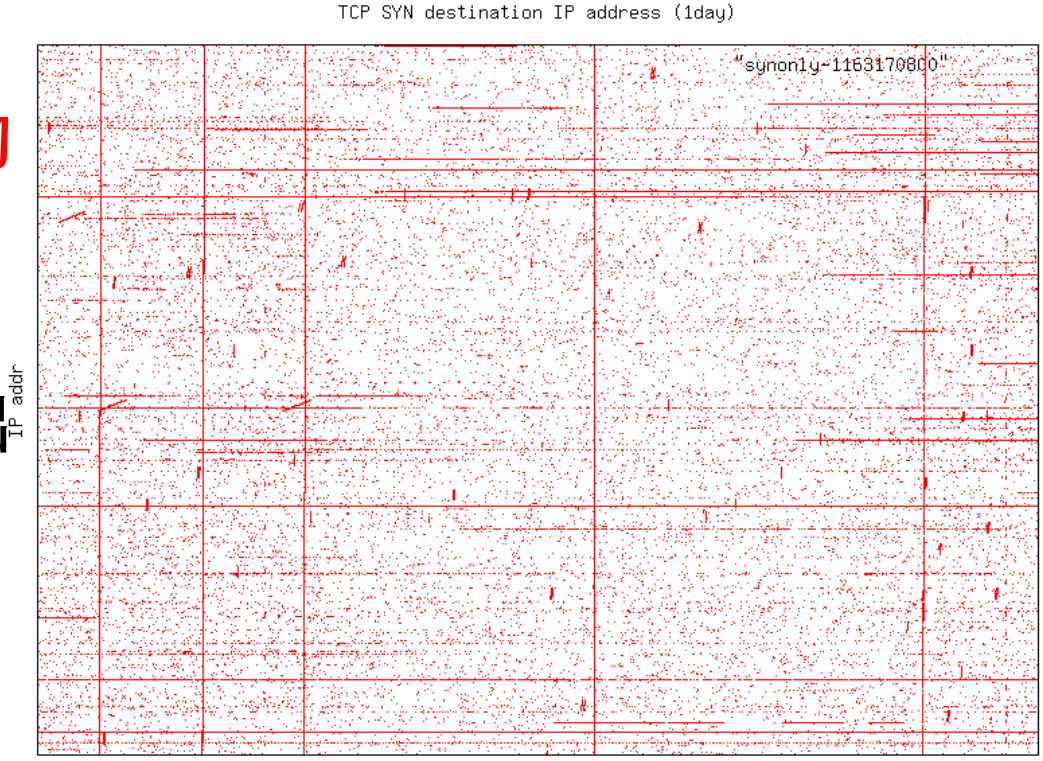
- ハッシュとモデルによるインターネットトラフィック異常検出(攻撃, ウィルス, 設定間違い,



Extracting hidden anomalies using sketch and non-Gaussian multiresolution statistical detection procedures (LSAD'07)

8.画像認識で異常を探す

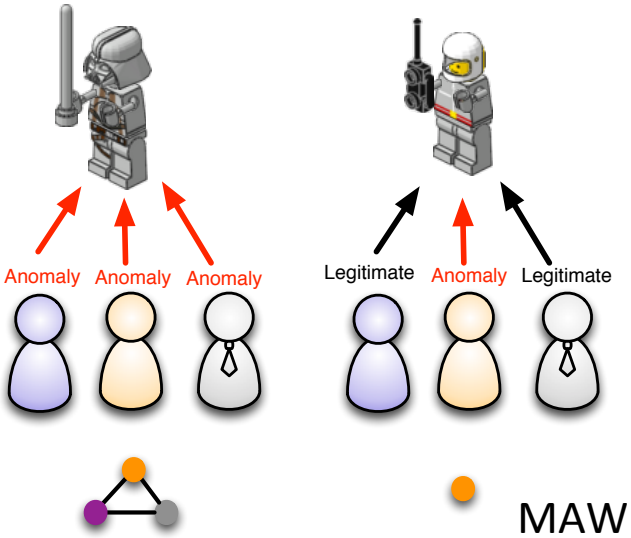
- 攻撃パターンを時空間的な構造へ変換し画像として検出
- ハフ変換による直線抽出

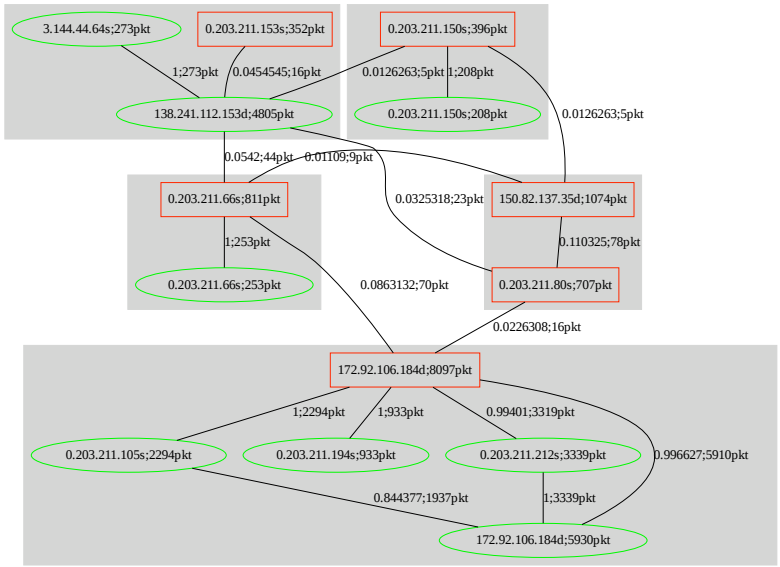


A Hough-transform-based anomaly detector with an adaptive time interval (ACR in 2013)

9.三人居れば文殊の知恵

- 複数の異常検出器の組み合わせによる精度向上
  - グラフマイニング+SVD

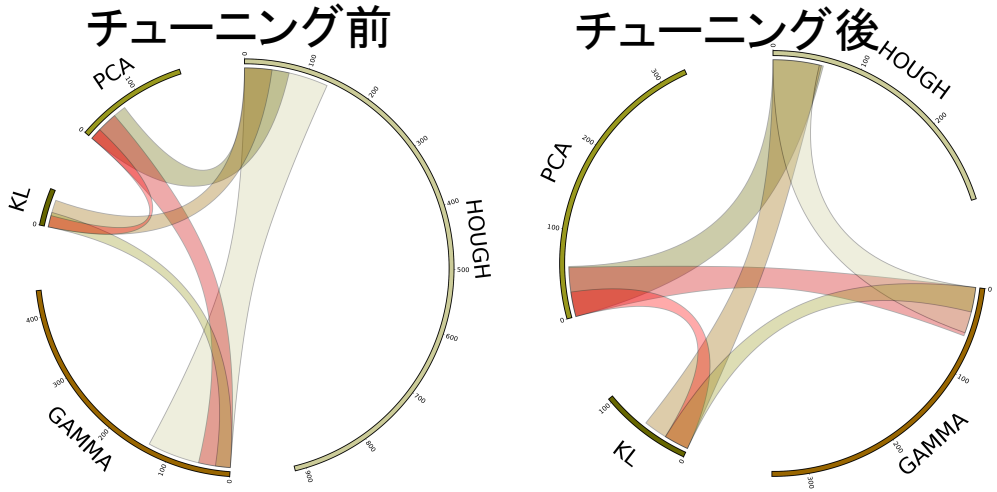




MAWILab: Combining diverse anomaly detectors for automated anomaly labeling and performance benchmarking (CoNEXT'10)

10.人手で異常検出の精度を上げるには?

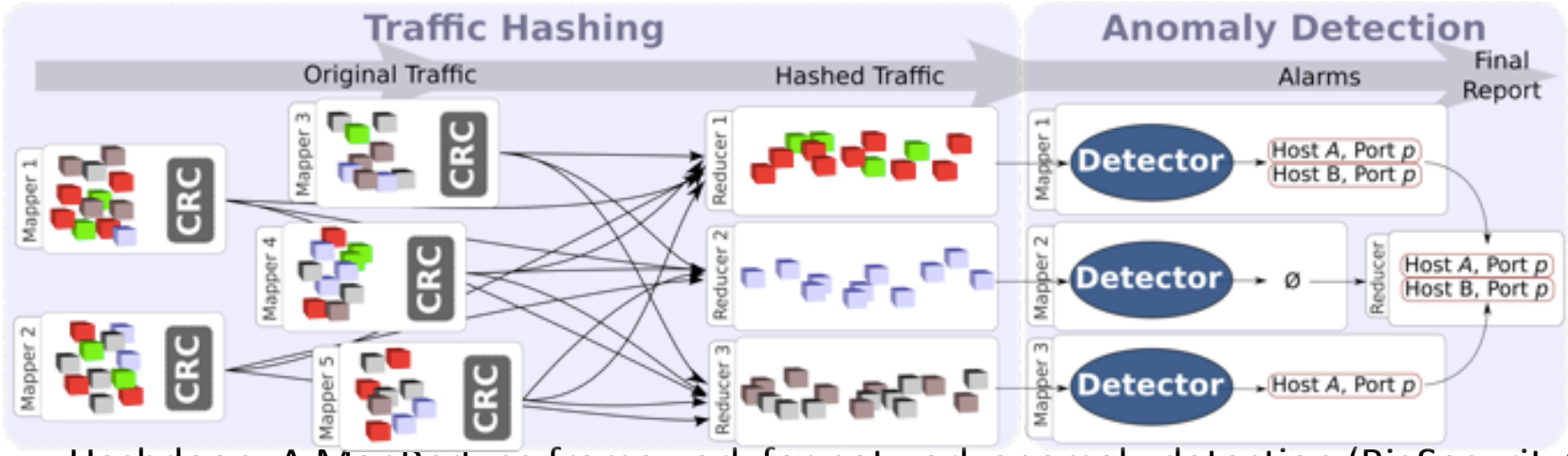
- 複数の異常検出器のパラメータチューニング
- 視覚化(コード図)によるサポート



Visual comparison of network anomaly detectors with Chord diagrams (SAC'14)

11.分割して統治せよ

- 大規模分散計算基盤を用いた解析
- ビッグデータをハッシュによりスモールデータへ
- 10倍速く3倍賢く



Hashdoop: A MapReduce framework for network anomaly detection (BigSecurity'14)

12.有名サイトに似た悪いサイトを探す

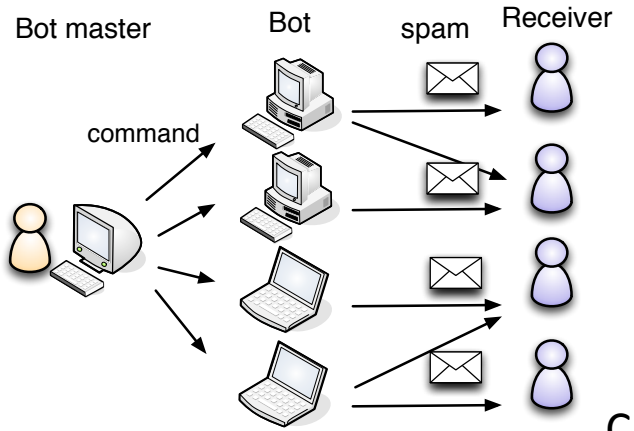
- インターネットを悪いことに使う人もいる
  - 有名サイトに似た偽サイトを登録
  - 迷惑メールを送ったり，フィッシング詐欺に使われる
- 機械学習を用いて悪いサイトを自動検出
  - 約2000サイトを検出(登録者2名)

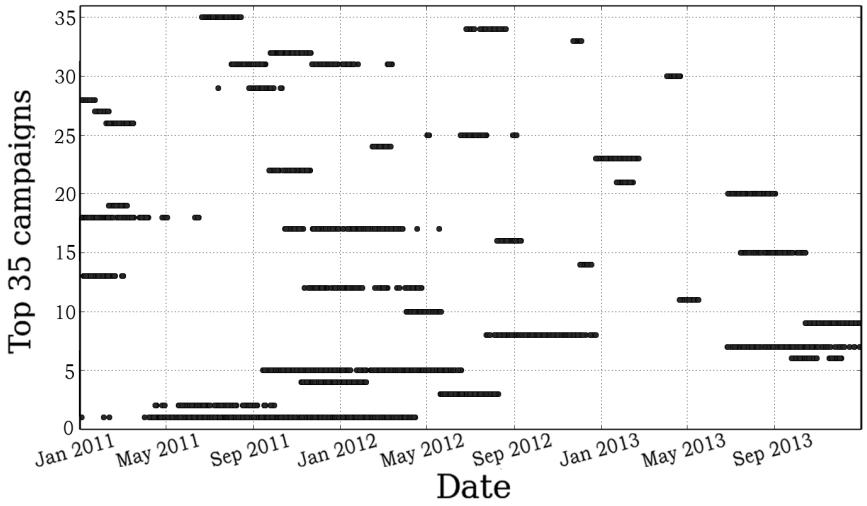
Domain names	
www.akivcsgree.jp	www.yrjtohjmbga.jp
mail.gtasomgree.jp	www.bsyhdjaskwheatmixi.jp
yrtwetwamixi.jp	www.lkjaysaddlebrowngree.jp
mayonnaisembga.jp	ns1.djbngree.jp

Towards classification of DNS erroneous queries (AINTEC'13)

13.迷惑メールはどうやって送られてくる?

- 似た文面を持つ迷惑メールをグループ化
- 多くはウィルスに感染したホスト(bot)から
- 大きなキャンペーンが多数

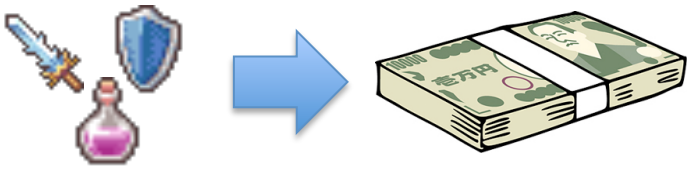




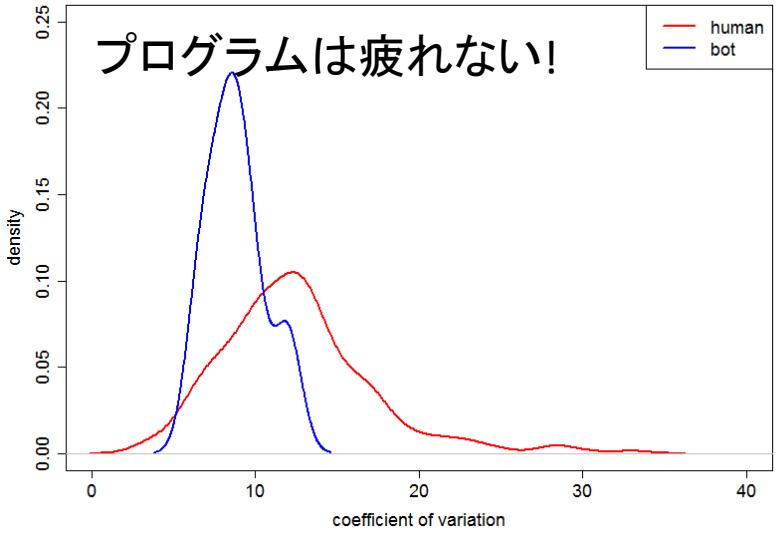
Clustering spam campaigns with fuzzy clustering (AINTEC'14)

14.オンラインRPGでの悪い人探し

- 人間ではなくプログラムがゲームマネー&アイテム稼ぎ
- オークション等で換金
- 人とは違う動きを検出する



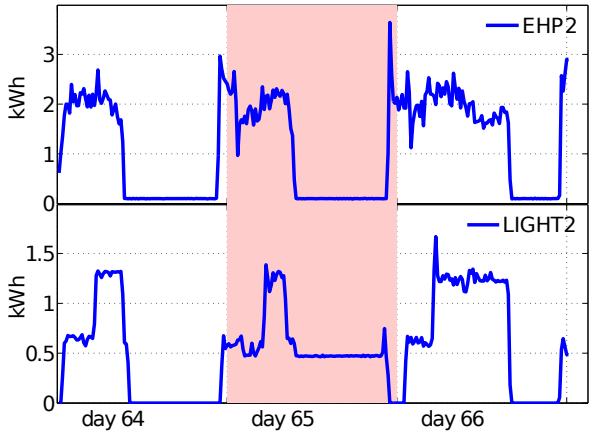
Real Money Trading



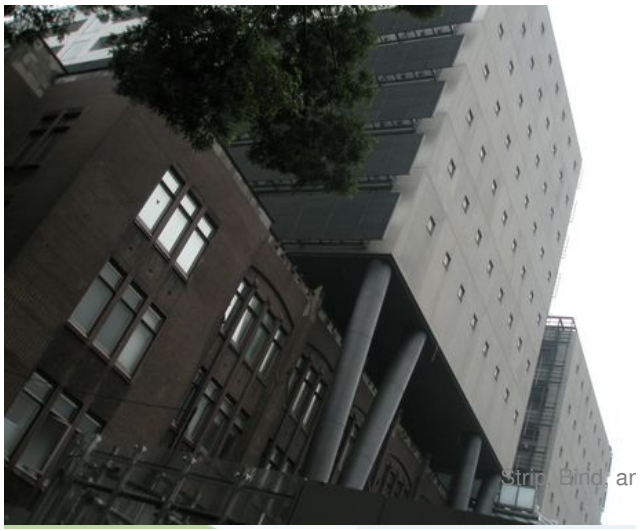
An analysis of players and bots behaviors in MMORPG (AINA'13)

15.ビルはどれだけ省エネできる?

- 多数(>100)のセンサーの時系列から異常を検出
  - 例: 冷房と暖房が同時に入ってる!



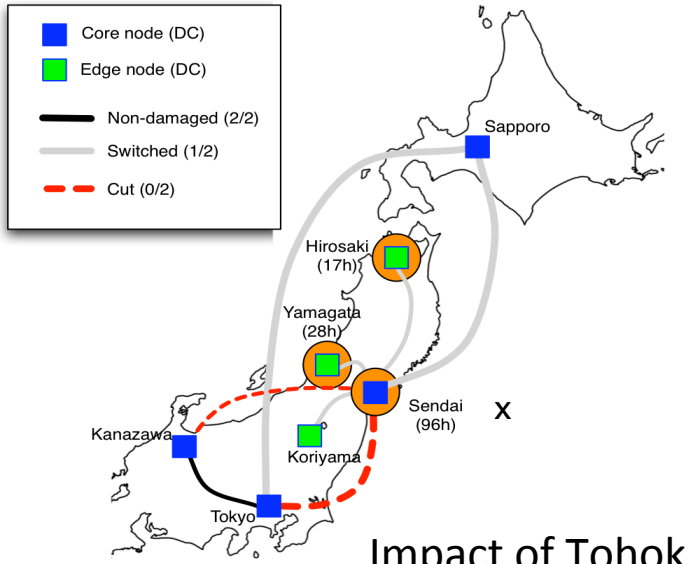
電灯の消し忘れ



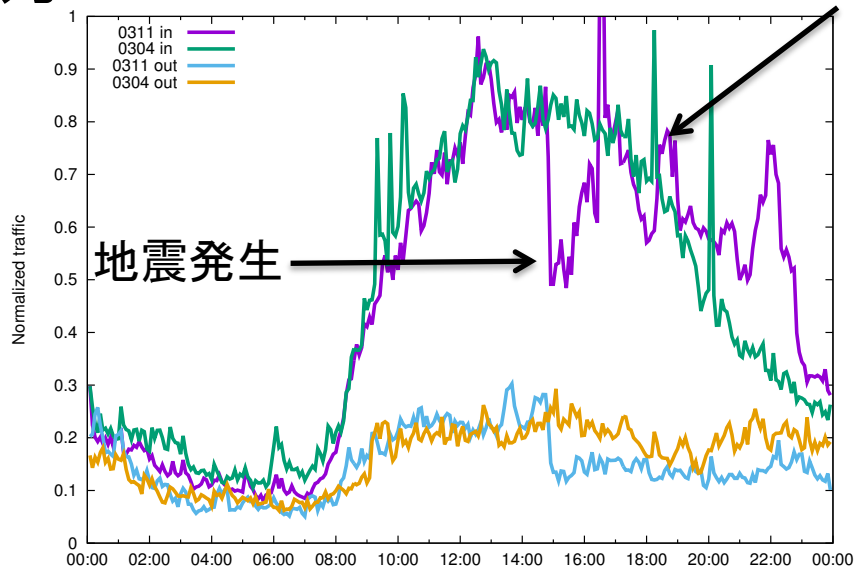
Strip, Bind, and Search: A method for identifying abnormal energy consumption in buildings (IPSN'13)

16.東日本大震災時の学術インターネット

- 学術ネットワーク(SINET)は冗長構成で耐えていた
- 想定外の新しい使われ方



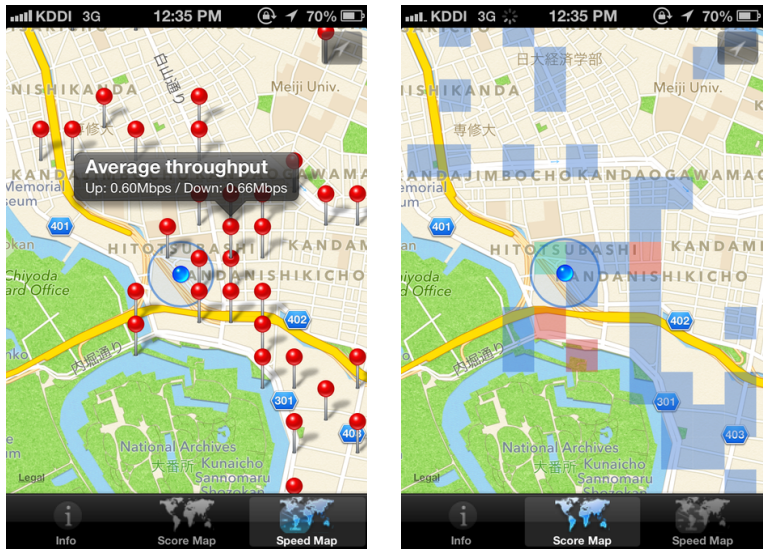
テレビ(勝手に)再配信



Impact of Tohoku earthquake on R&E network in Japan (CoNEXT WoID'11)

17.みんなの力でスマホの診断

- ネットワークが遅い原因の究明
- みんなで測定して結果をアップロード
  - 次の角まで行けば快適になるかも?

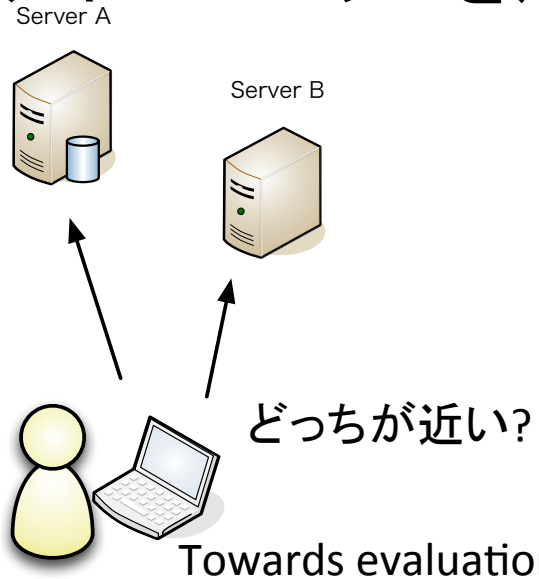


(a) QoS (b) QoE

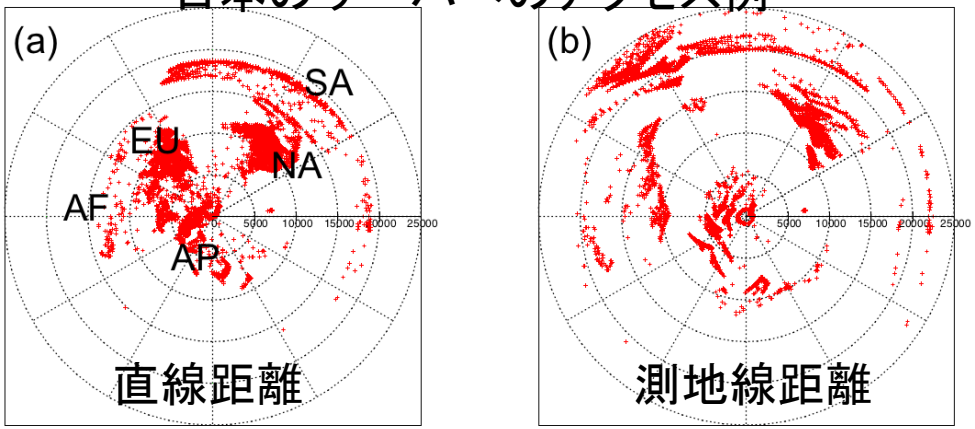
Social benchmarking of QoS & QoE in cellular data networks (CQR workshop'13)

18.大規模サーバはどれくらい効率的?

- 世界中の数百万のクライアントからアクセス
- ファイバーマップを用いた負荷分散の効率解析



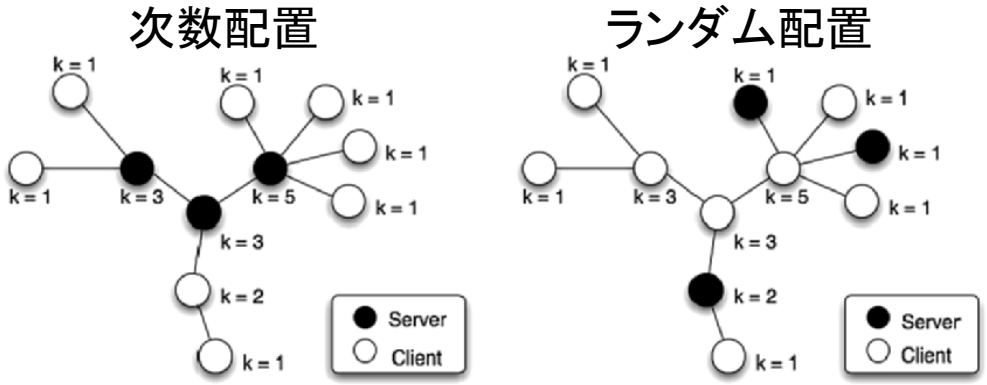
日本のサーバへのアクセス例



Towards evaluation of DNS server selection with geodesic distance (NOMS'14)

19.トポロジを意識した負荷分散は可能?

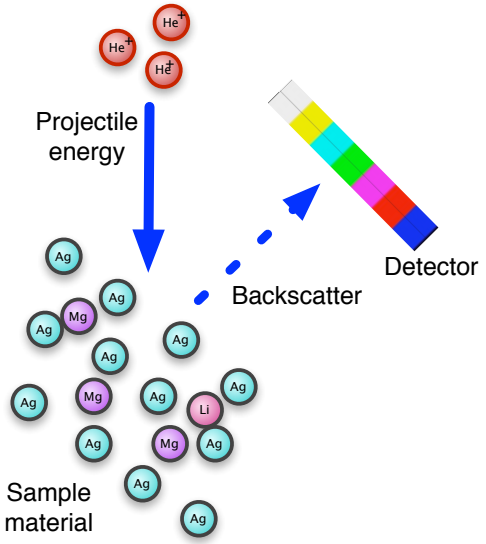
- インターネットは分散システムのため全容はわからない
- 周りを見ただけで全体の効率を良くできるか?
- 次数(つながり具合)を見ることである程度は可能



Dependency of network structures in agent selection and dependency (IAT'06)

20.集合知による新しいインターネットセンサ

- DNS反射波を用いた異常検出センサ
  - 大きなイベント(迷惑メール，スキャン)発生時に自動生成されるDNSクエリを使用
  - DNS(名前解決)サーバは世界中に遍在
  - 個々のクエリの情報量は少量 -> 集合知
  - プライバシに配慮



Detecting malicious activity with DNS backscatter (IMC'15)