ネットワークセキュリティと“不”自然科学

電子情報通信学会
情報ネットワーク科学研究会
2014年5月16日

森　達哉 (早稲田大学)
http://nsl.cs.waseda.ac.jp

NSL
Networked Systems Laboratory

本講演の目的ではないもの

- ネットワークセキュリティの解説

- 「不自然科学」という新しい研究領域（？）の提案

本講演の目的

- データ解析の立場でネットワークセキュリティの研究に取り組む中で感じた面白さを2つの事例を通じて紹介

- タイトルに込めた意図
  - セキュリティの問題では「自然」な事象よりはむしろ「不自然」な事象に興味がある
  - 不自然な事象の中にも一定の規則性が観察できることが多い（規則性の発見⇒対策技術につながる）
    - 発表者はそこに面白さを感じている

ネットワークセキュリティとデータ解析

- セキュリティ課題に対して有効な対策を立てる上ではまず敵を知ることが重要
  - *敵を知り己を知れば百戦危うからず　孫氏兵法*
- 敵を知る→手口の詳細な理解（プロファイリング）。技術的にはデータの解析がメイン。
  - 規則性やパターンを導き出すことで未知の脅威に備える
- 得られたプロファイリングにもとづき、対策を立てることが有効
  - [cf] ピッキング対策

2つのケーススタディ

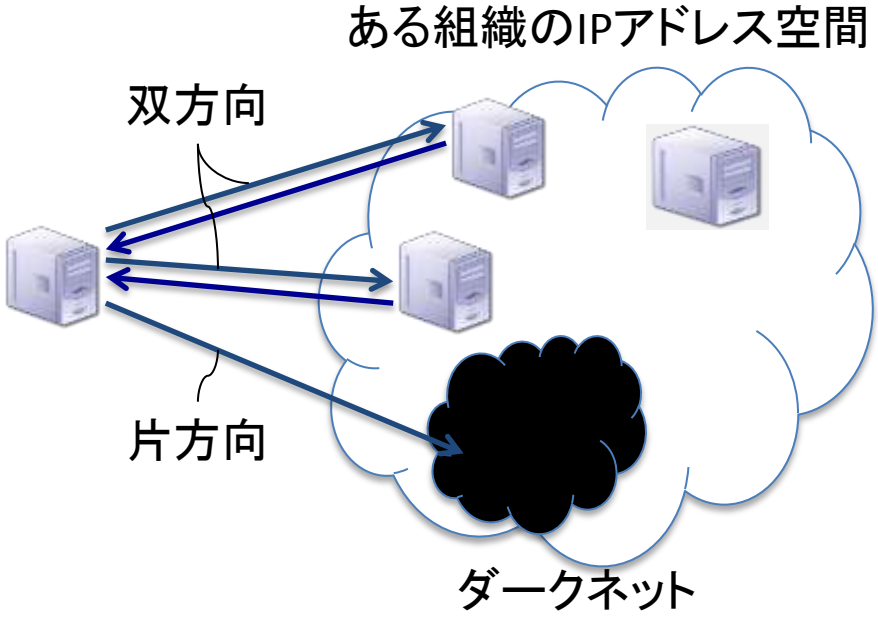
1. ダークネット解析（ネットワーク・感染端末）

1. モバイルアプリ解析（ソフトウェア・人間）

1　ダークネットの解析

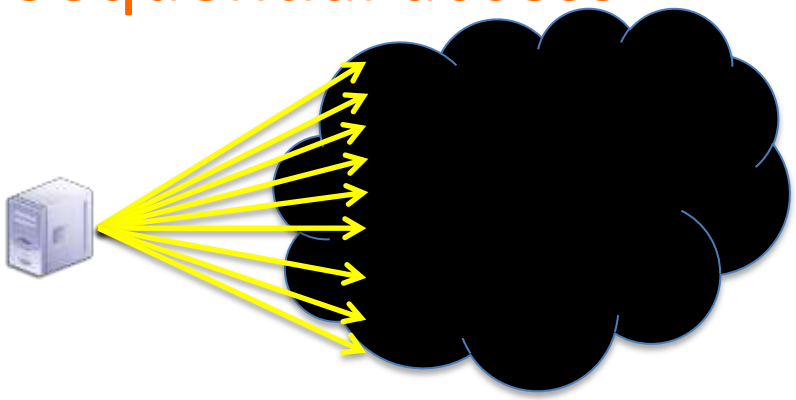
ダークネット

- インターネット上において到達性（経路）があるが，組織内で割り当てられていない未使用アドレス空間
- 組織外からはダークネットの存在はわからない
  ⇒攻撃者による脆弱性を持つホスト・サービスの探査活動をステルスに捕捉できる（トラップ，ネズミ捕り）



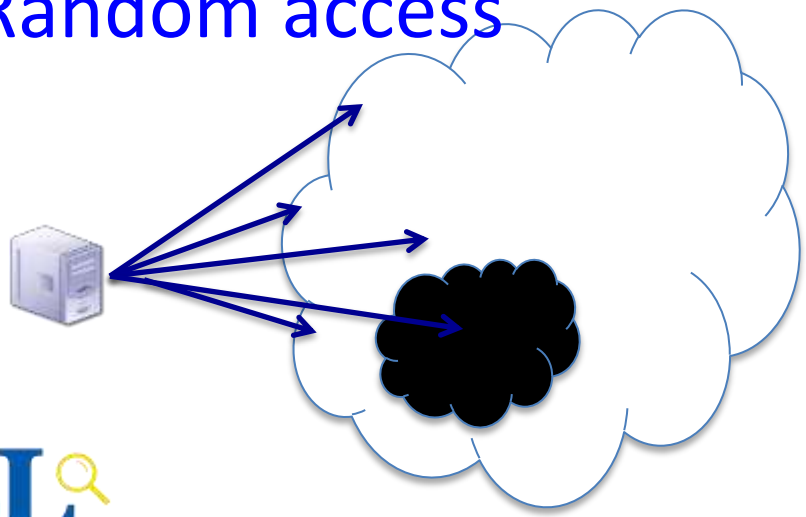
ダークネットへのアクセスパターン

Sequential access



ネットワークスキャン
ポートスキャン
⇒検知されやすい

Random access

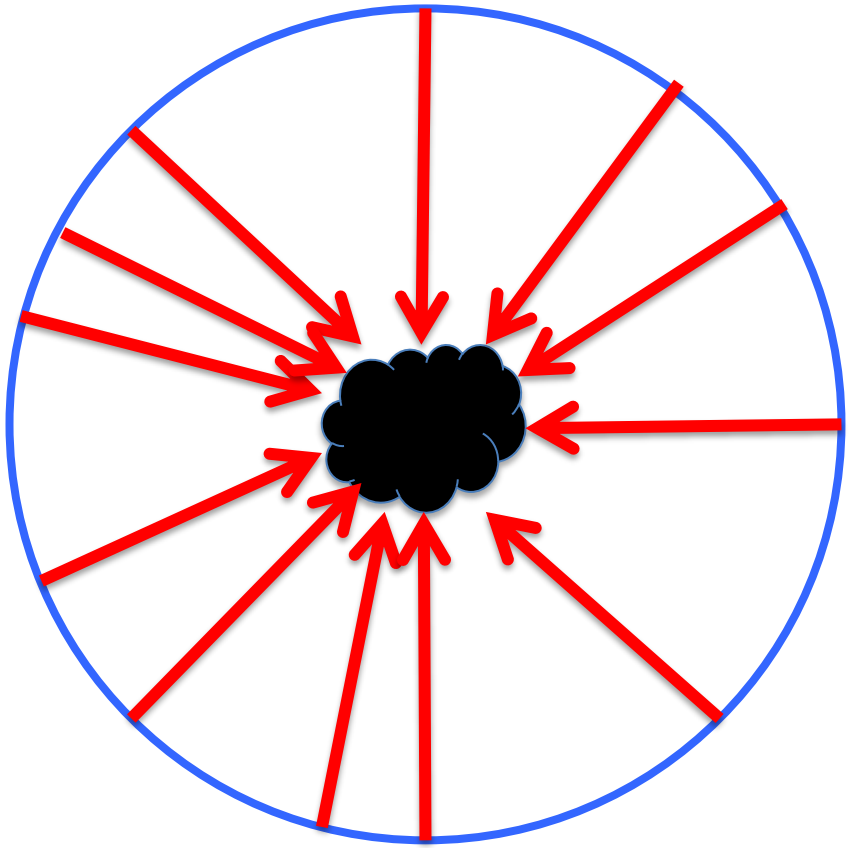


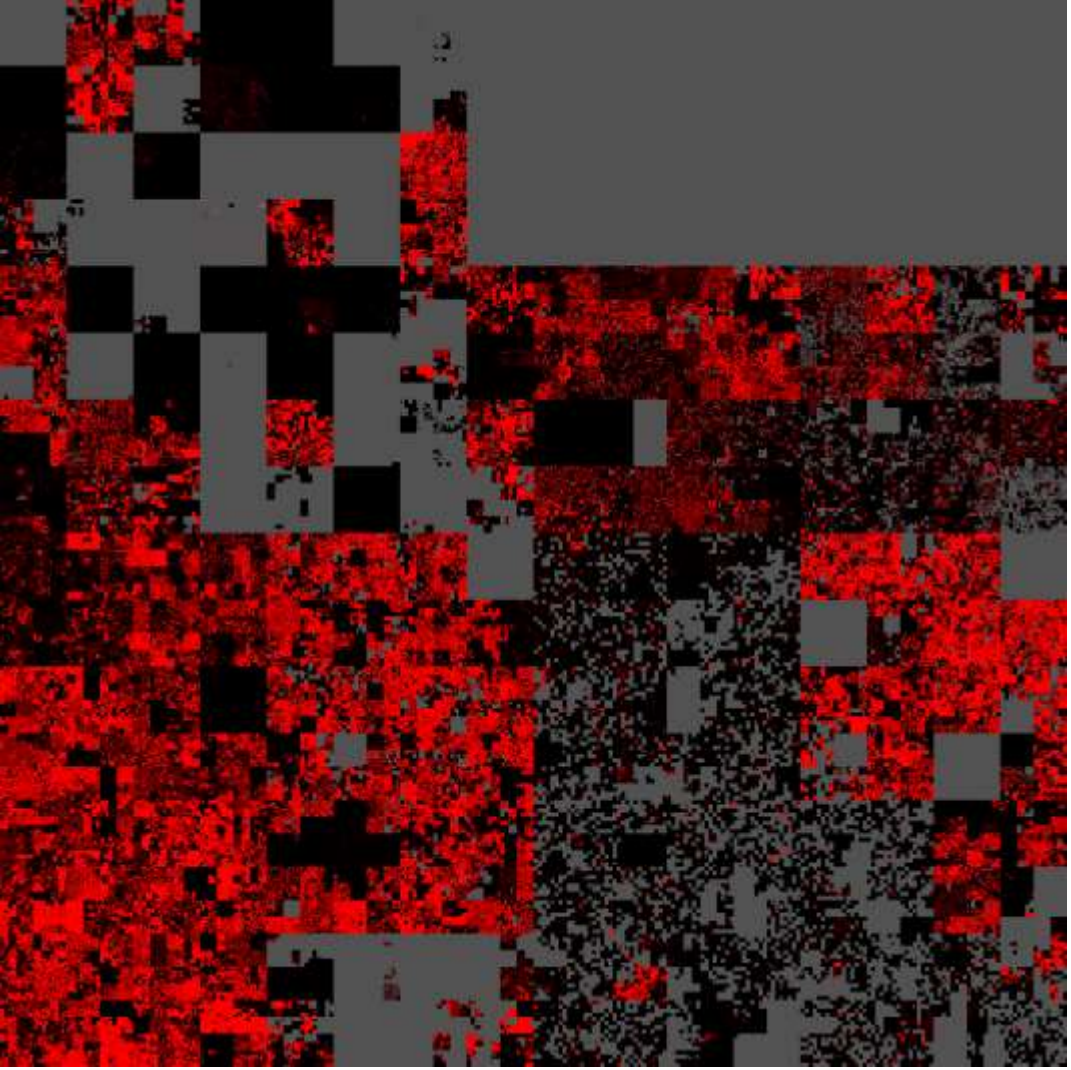
検知を逃れるために
ランダムにスキャン
⇒計測に確率的な要素が入る

NSL
Networked Systems Laboratory

インターネット背景放射

ダークネットに届くパケットはIPアドレス全空間から等方的に観測される？（実際は等方的ではなく偏りがある）





CAIDA Telescope

Red: observed

Grey: no route

Black: Unobserved→ darknet?

Dainotti, K. Benson, A. King, k. claffy, M. Kallitsis, E. Glatz, and X. Dimitropoulos, “Estimating Internet address space usage through passive measurements'', ACM SIGCOMM Computer Communication Review (CCR), vol. 44, no. 1, pp. 42--49, Jan 2014.
http://www.caida.org/publications/papers/2014/passive_ip_space_usage_estimation/supplemental/

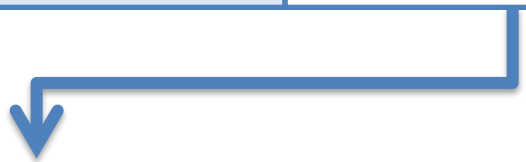
ダークネットの観測能力

IPv4 アドレス空間

32 bit = 2^{32}個

ある組織のアドレス空間 (/24)

Prefix = 133.9.XXX.0/24	8 bit = 2^{8}個



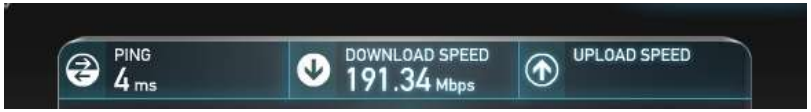
例えばこのアドレス空間をダークネットとして
使うとき全空間に対するカバー率 p = 1 / 2^{24}

NSL
Networked Systems Laboratory

ランダムスキャンホストがダークネットで観測されるための条件

- スキャンレート r [P/sec]
- スキャン時間 t [sec]
- ヒット確率 p (カバー率)
- n = rt 回試行したときに少なくとも一回サンプルされる確率 $q = 1 - (1-p)^{rt}$
- $p = 1 / 2^{24}$, t=86400 sec (1日) のとき,
  - r = 10 P/sec → q = 0.05
  - r = 100 P/sec → q = 0.40
  - r = 1 KP/sec → q = 0.99

Commodity hardware のスキャン能力

- zmap を利用した実験
- CPU: Intel Xeon E5-2620 v2 @ 2.10GHz
- Ethernet Controller: Intel i350 GbE
- 回線: フレッツ光ネクスト ハイスピードタイプ
  - 実効スループット計測値: 

- 達成したスキャンレート: 705 Kp/sec >> 1 Kp/sec
  - 1時間30分で全 IPv4 空間をスキャン完了

```
0:00 0%; send: 230 617 Kp/s (5.06 Kp/s avg);
0:01 0%; send: 705506 705 Kp/s (675 Kp/s avg)
0:02 0%; send: 1408796 703 Kp/s (688 Kp/s avg
0:03 0%; send: 2112459 704 Kp/s (693 Kp/s avg
0:04 0%; send: 2815722 703 Kp/s (696 Kp/s avg
0:05 0% (1h28m left); send: 3518860 703 Kp/s
0:06 0% (1h28m left); send: 4222585 704 Kp/s
0:07 0% (1h28m left); send: 4927669 705 Kp/s
0:08 0% (1h28m left); send: 5631238 704 Kp/s
```

ダークネットデータの解析事例

- nicter datknet 2013

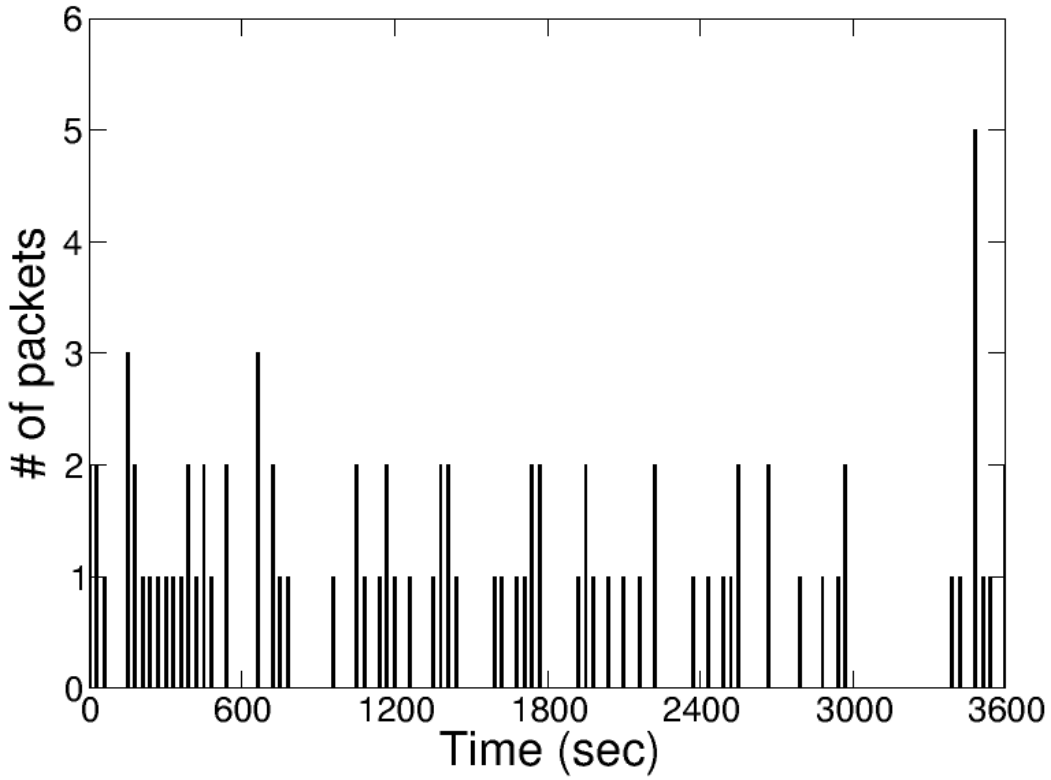
  - 情報通信研究機構が研究コミュニティに提供しているデータセットを利用
- ダークネットのサイズ
  - ある /16 を有する組織内のダークネット (/20)
  - 観測能力 → p = 1 / 4096
- 今回の解析対象データの計測期間
  - 2011年1月1日～2013年12月31日

以降の解析対象ホスト

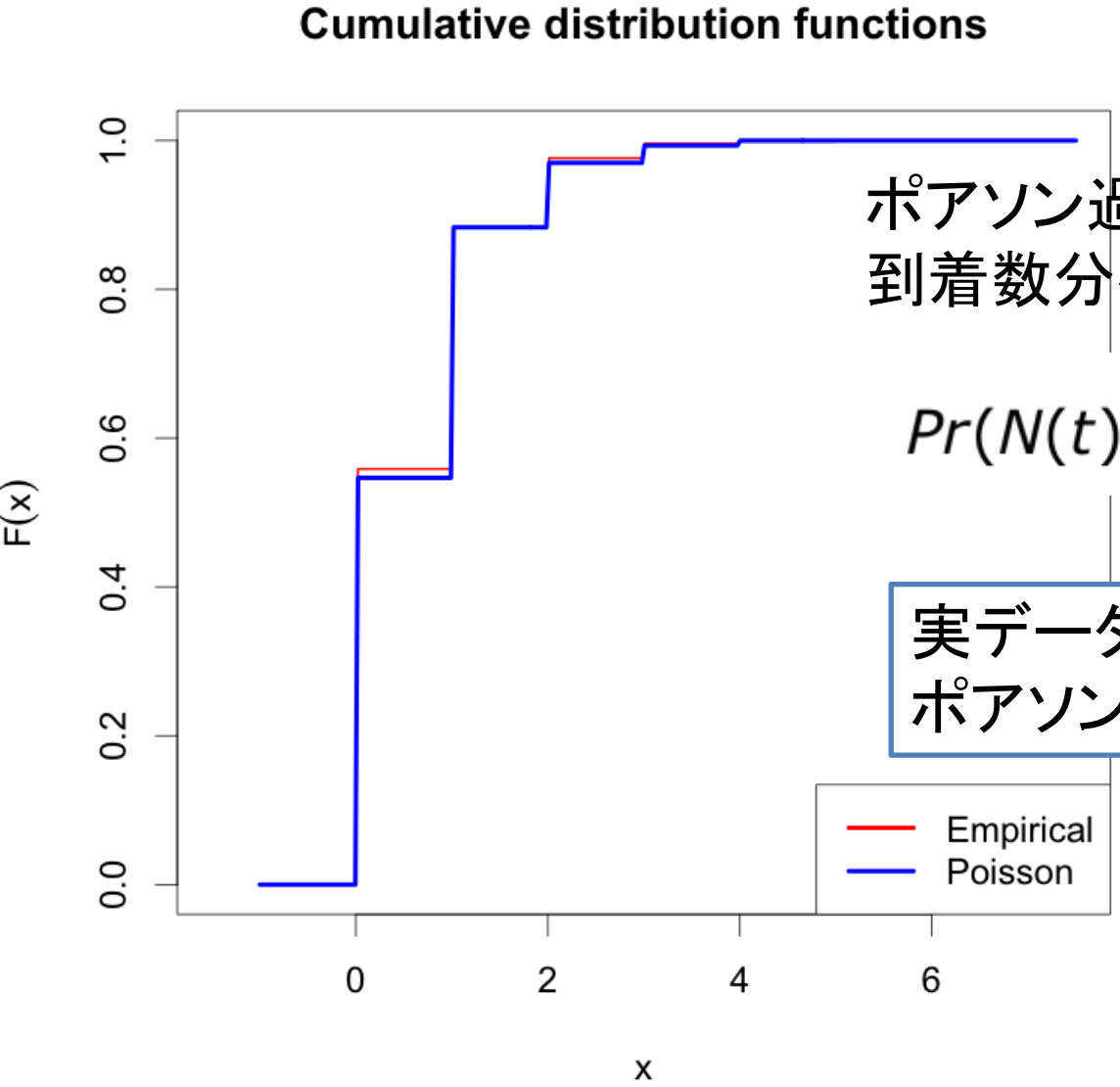
- 特にランダムスキャンをしかけてくるホストを対象とする
  - 1日1パケットのみ観測するケースが大半
- シーケンシャルスキャンを行ったホストはオフライン処理で除去
  - 実際の観測パケットの大半はシーケンシャルスキャンホストによるもの
    ⇒全体傾向を把握する上ではノイズとなる

素朴な疑問

ランダムスキャンホスト群が発生したパケットの到着に規則性はみられるか？



30秒間にダークネットに到着したパケット数の分布

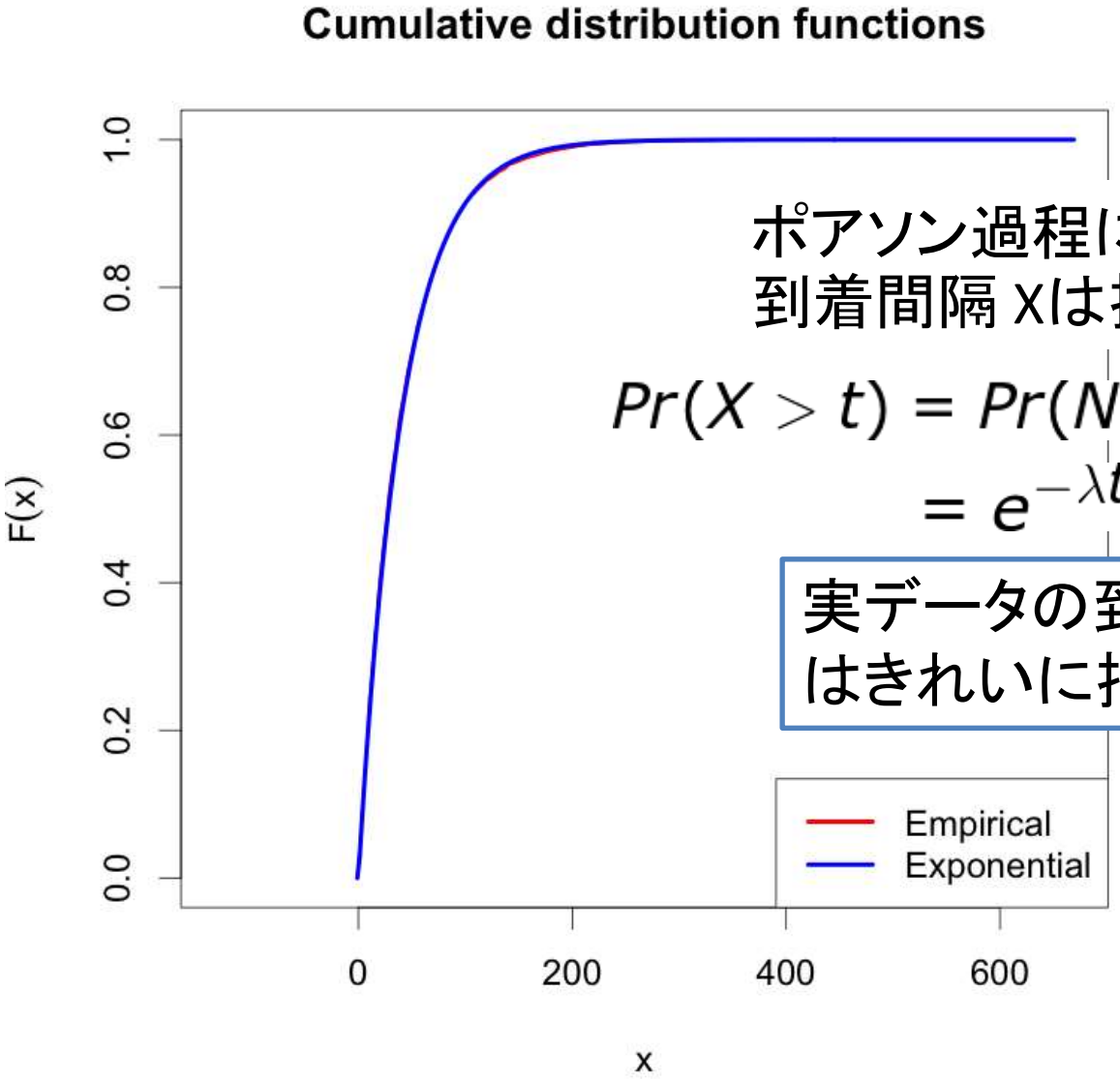


ポアソン過程＝ランダム到着では
到着数分布はポアソン分布する

$$Pr(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

実データの到着数分布は
ポアソンで良く近似される

ダークネットに到着したパケットの到着間隔分布

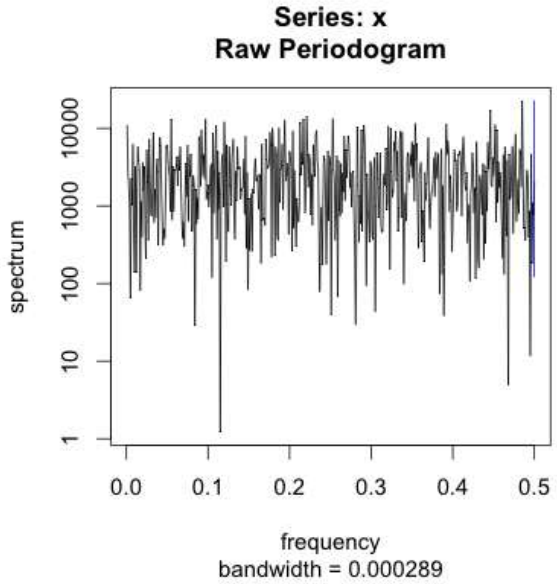


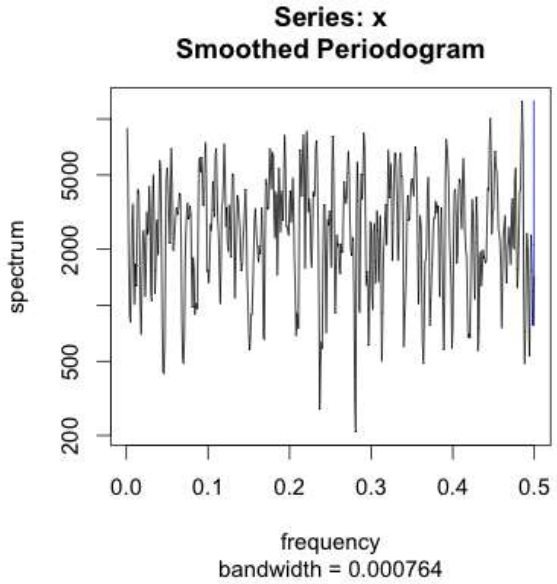
ポアソン過程にしたがう出来事の
到着間隔 Xは指数分布する

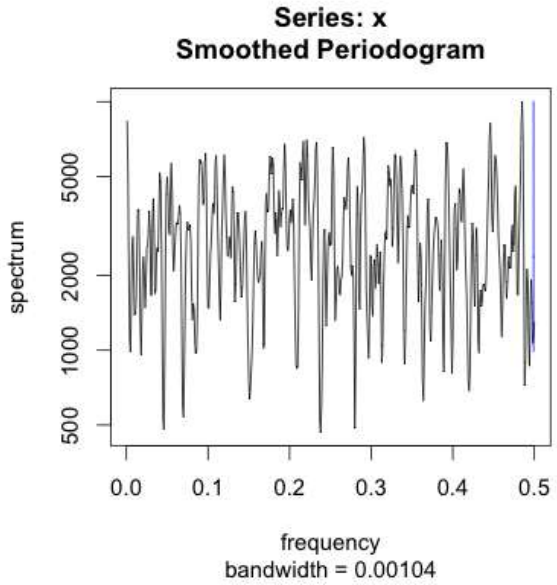
$$Pr(X > t) = Pr(N(t0+t) - N(t0)) = 0$$
$$= e^{-\lambda t}$$

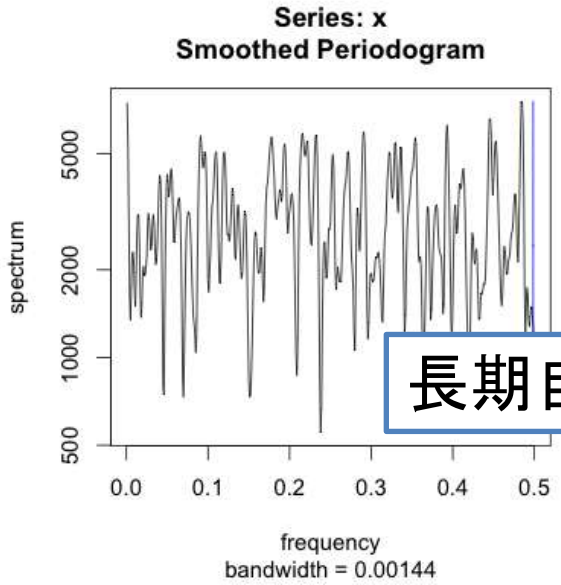
実データの到着間隔
はきれいに指数分布する

到着間隔列の時間相関構造





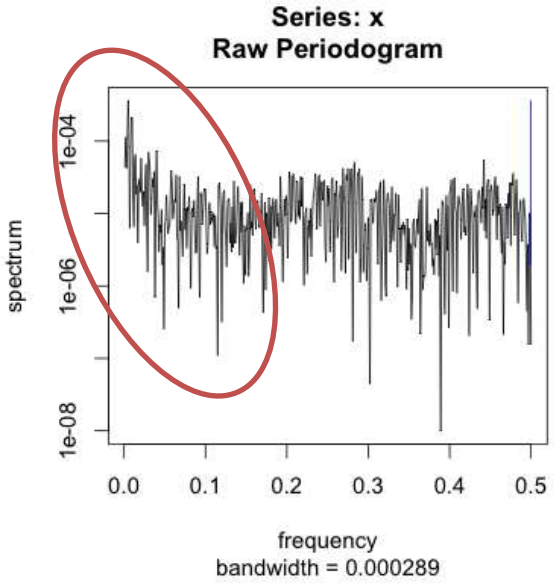


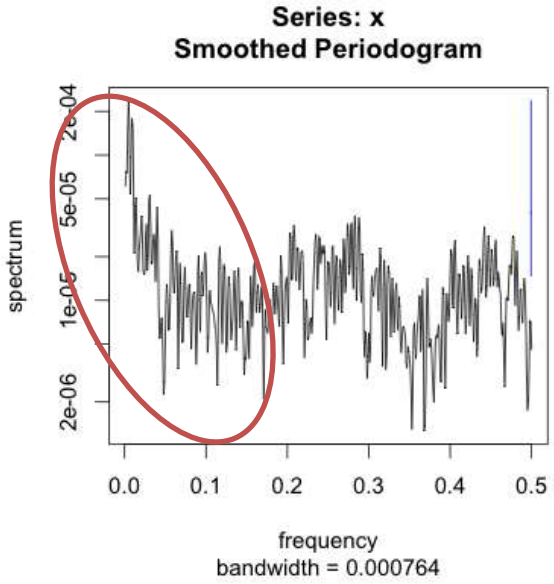


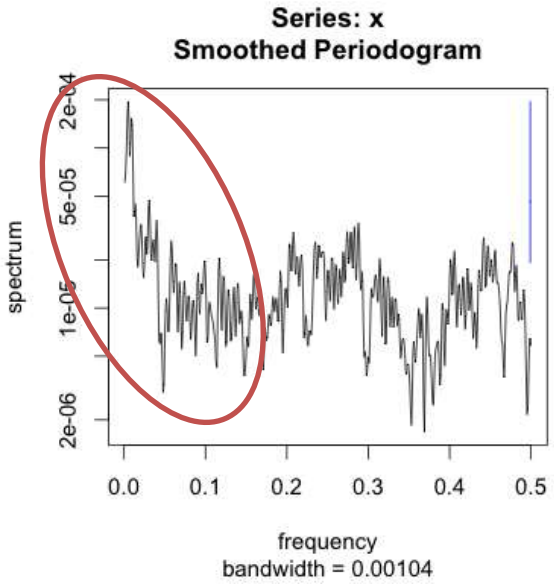
長期自己相関なし

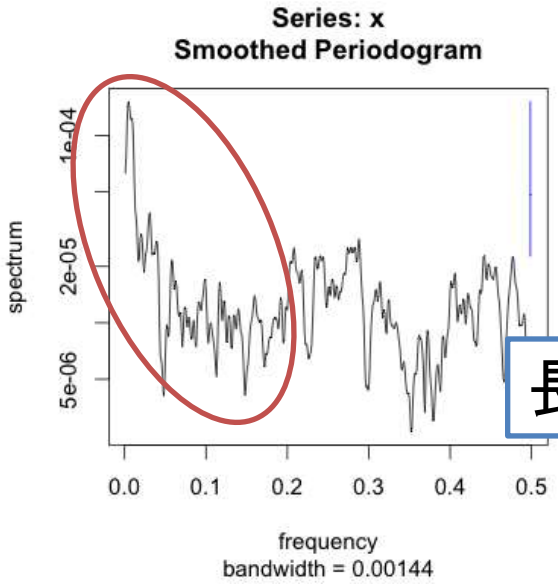
(cf) LANトラフィックの場合

Bellcore traffic dataset http://ita.ee.lbl.gov/html/contrib/BC.html









長期自己相関

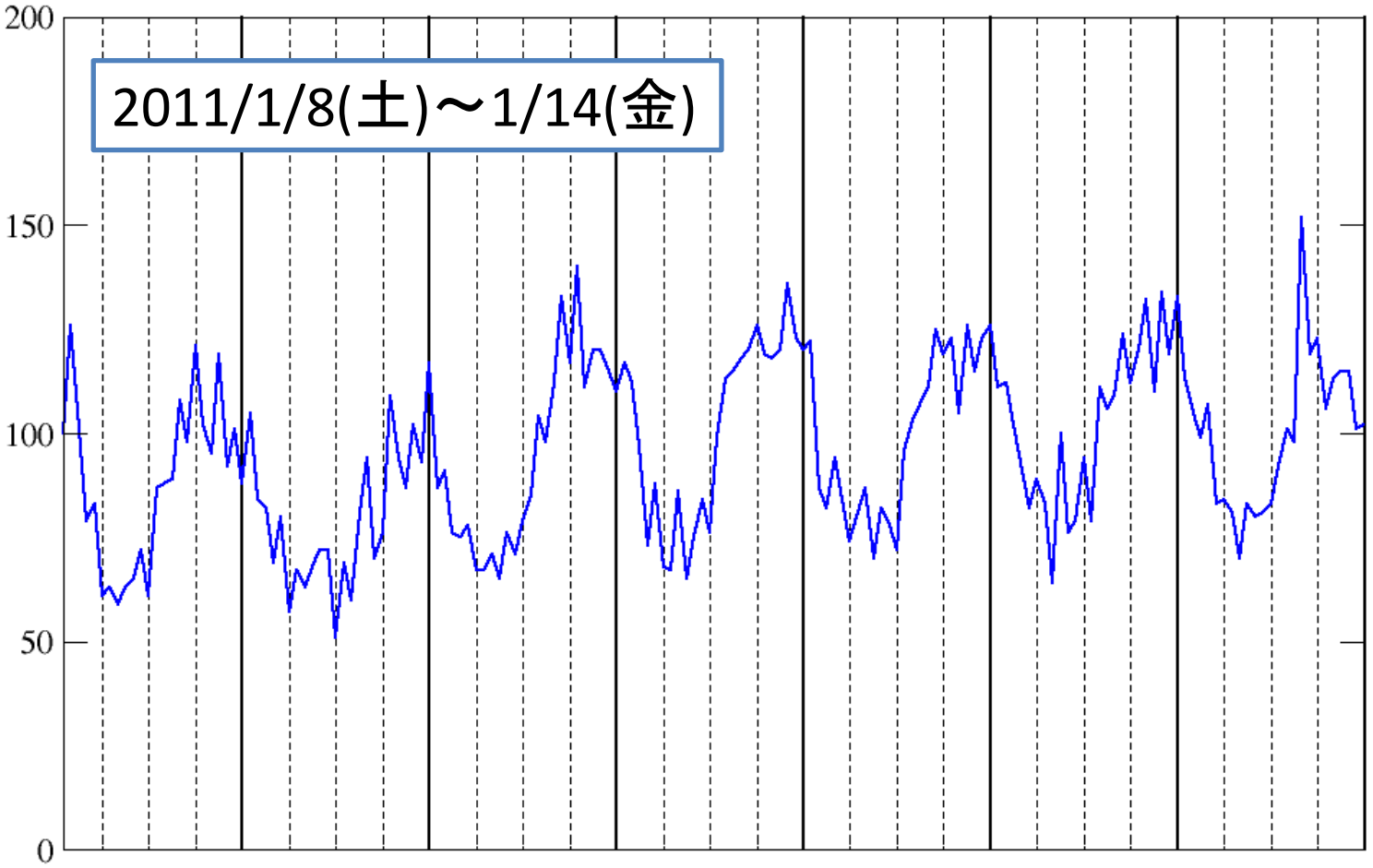
ランダムホスト群のパケット到着

- ランダムスキャンホスト群からダークネットに到来するパケットの到着過程はポアソン過程と同様の到着数分布，到着間隔分布を示した

※厳密には到着数分布がポアソン分布し，到着間隔分布が指数分布するのは計測した対象がポアソン過程であることの必要条件である．十分条件については「独立性」，「定常性」，「希少性」を示す必要あり（本発表では割愛）

非定常性と周期性

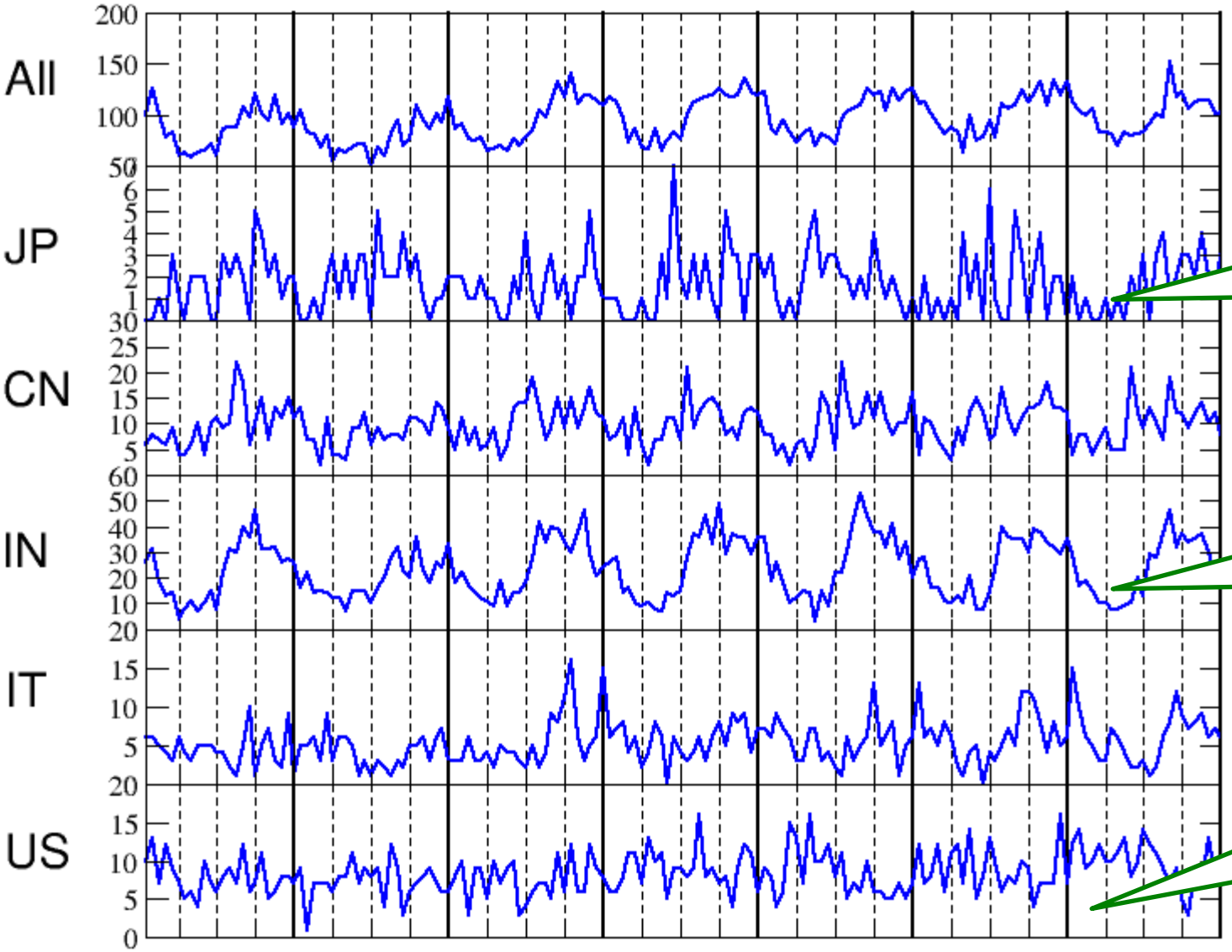
- 中～長期間で観測するパケット到着過程は非定常であり，特徴的な周期性を示す．下記は日変動の例．



2011/1/8(土)～1/14(金)

※時刻は日本時間

国ごとのパケット到着数⇒時差によるピークのずれ



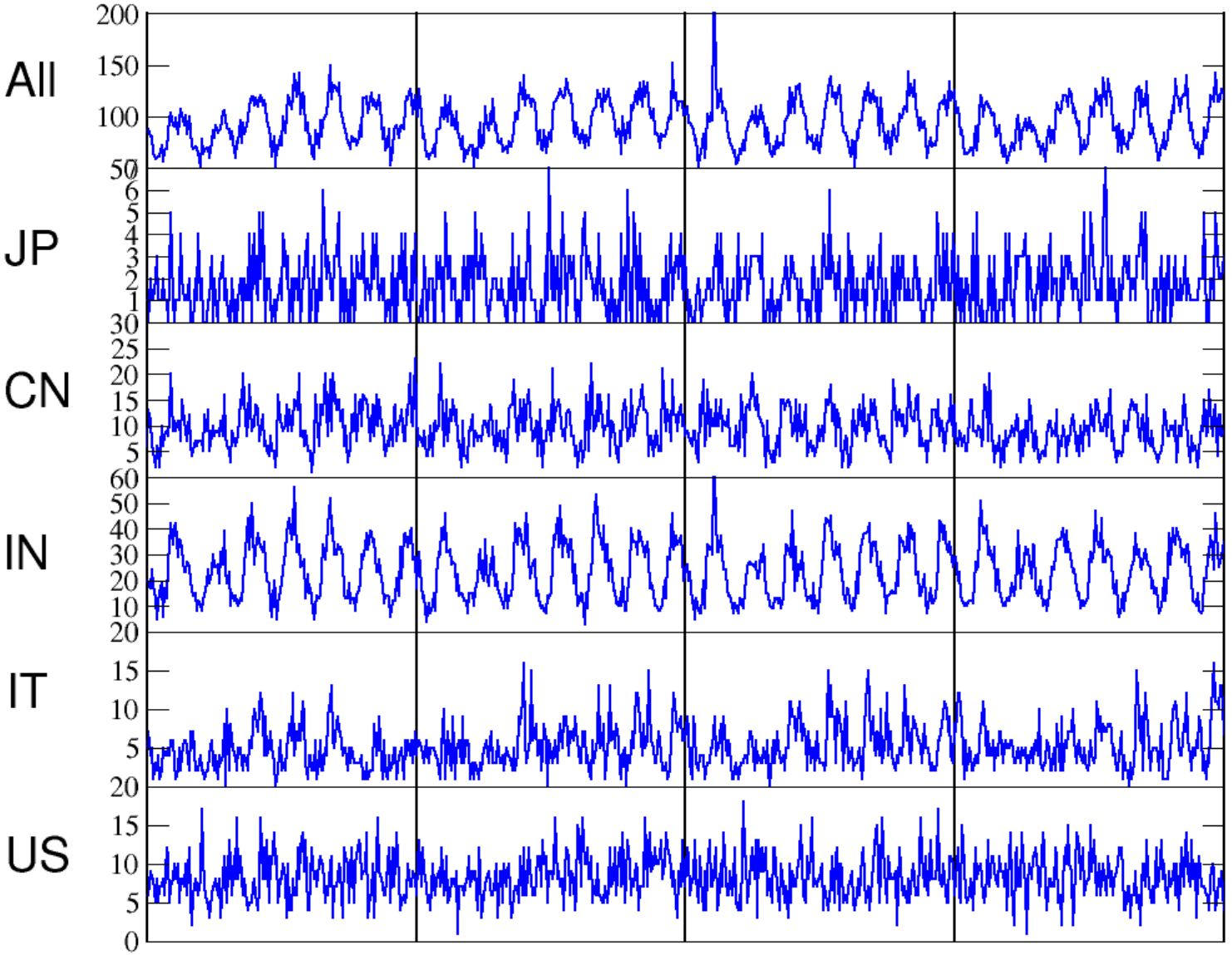
日本発は少ない

この時期インド発が多かった

周期性が低い

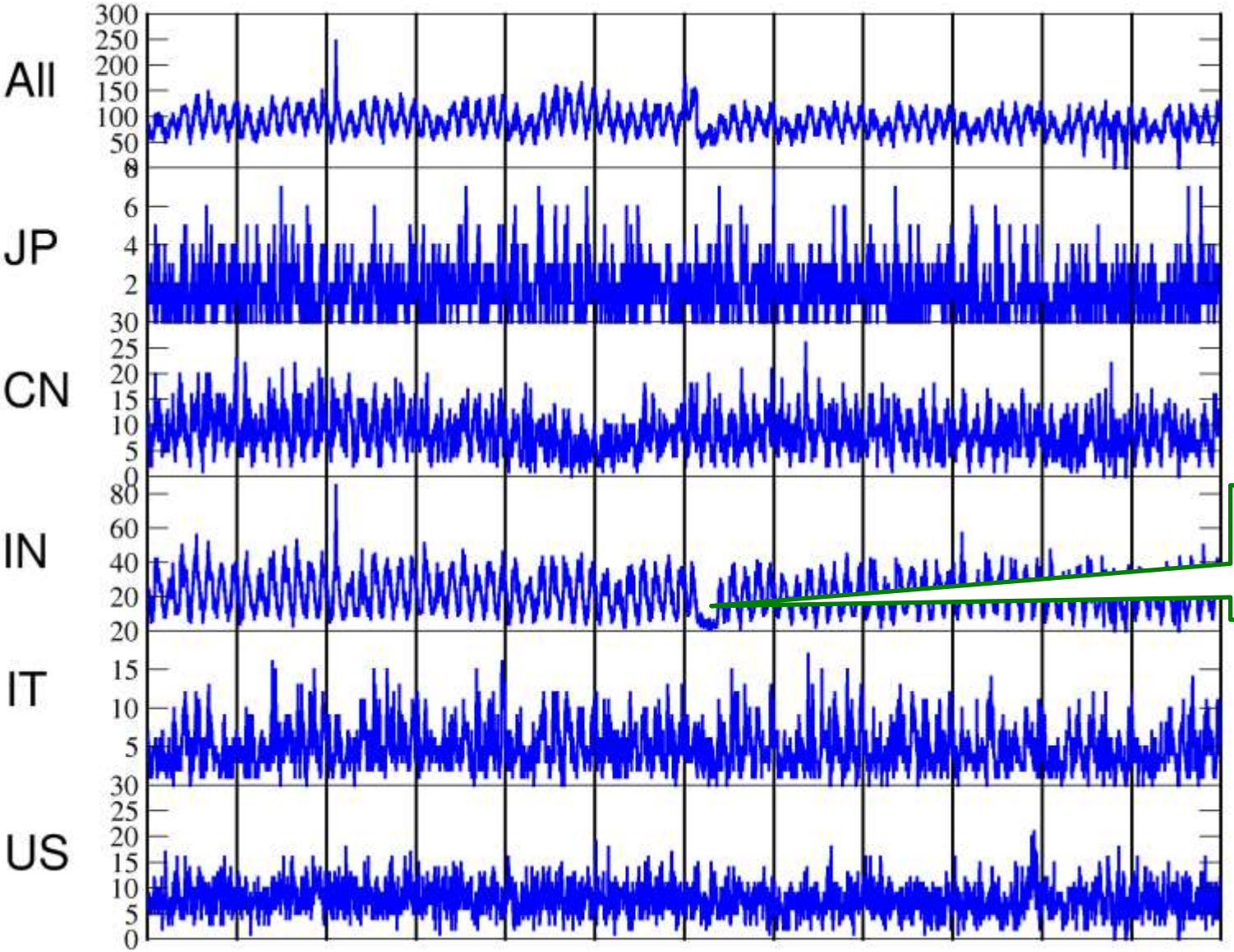
2011/1/1(土)～1/28(金)　(4週間)

※時刻は日本時間



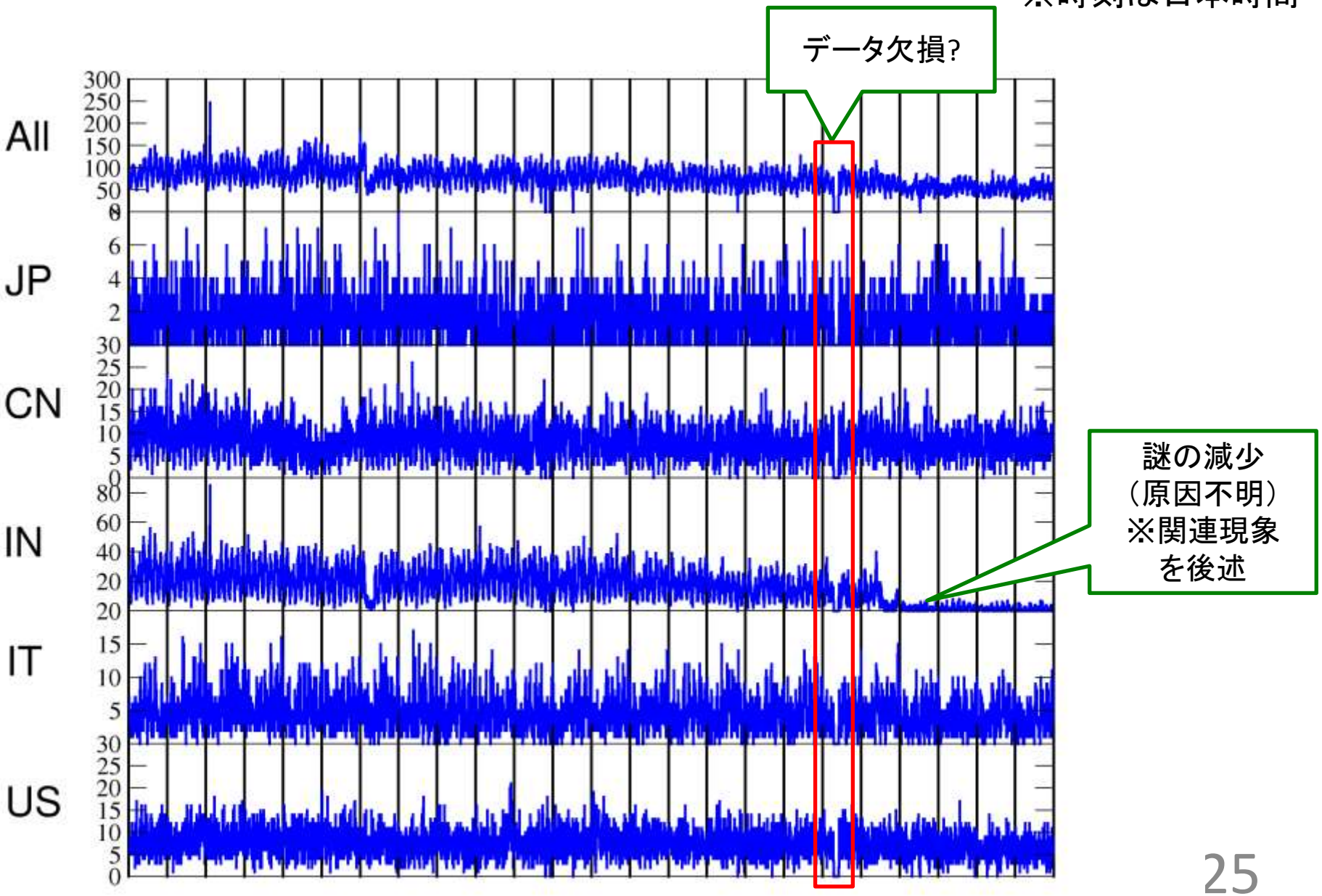
2011/1/1(土)～3/25(金)　(12週間)

※時刻は日本時間



2011/1/1(土)～6/10(金)　(24週間)

※時刻は日本時間

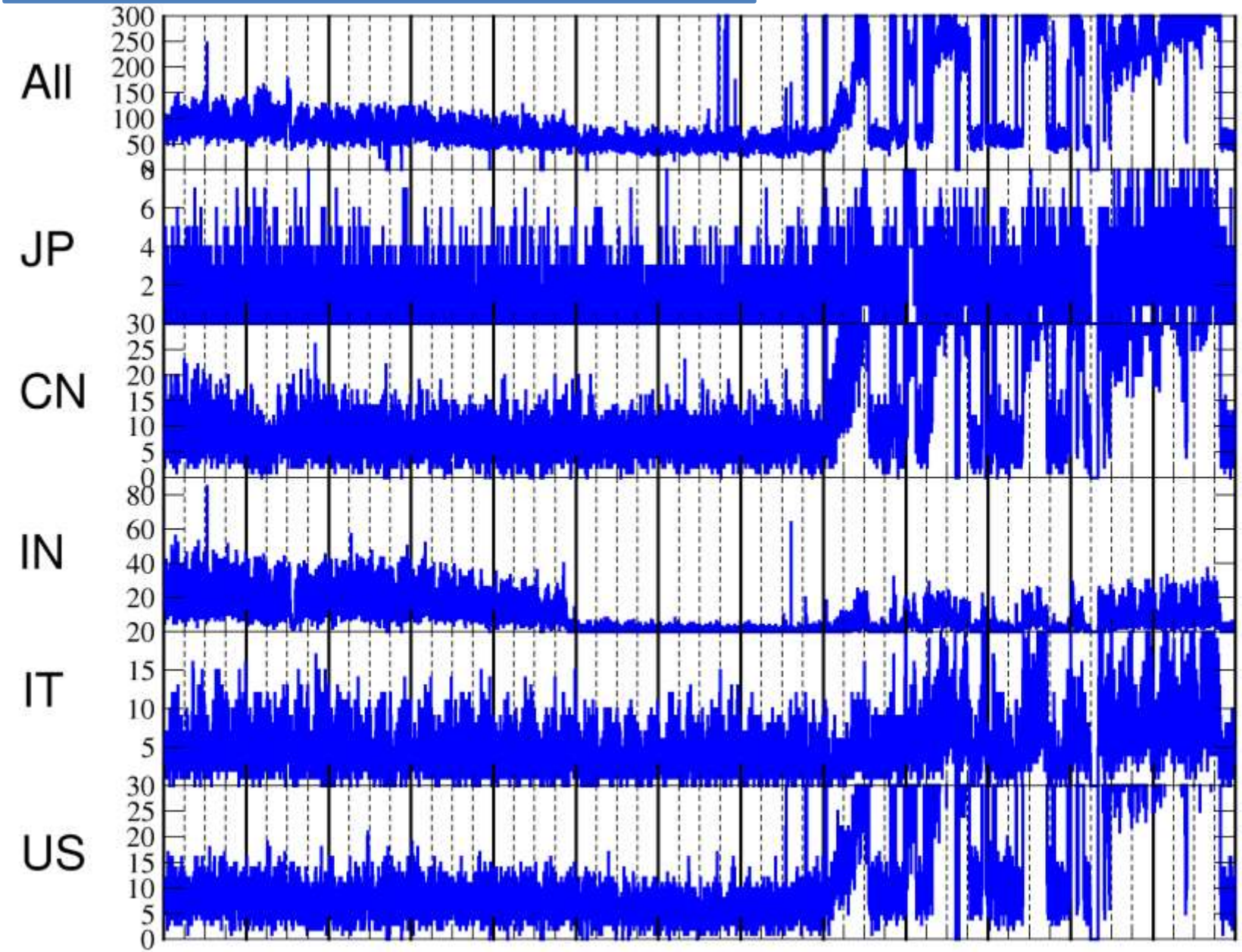


なぜ周期性を生じるのか

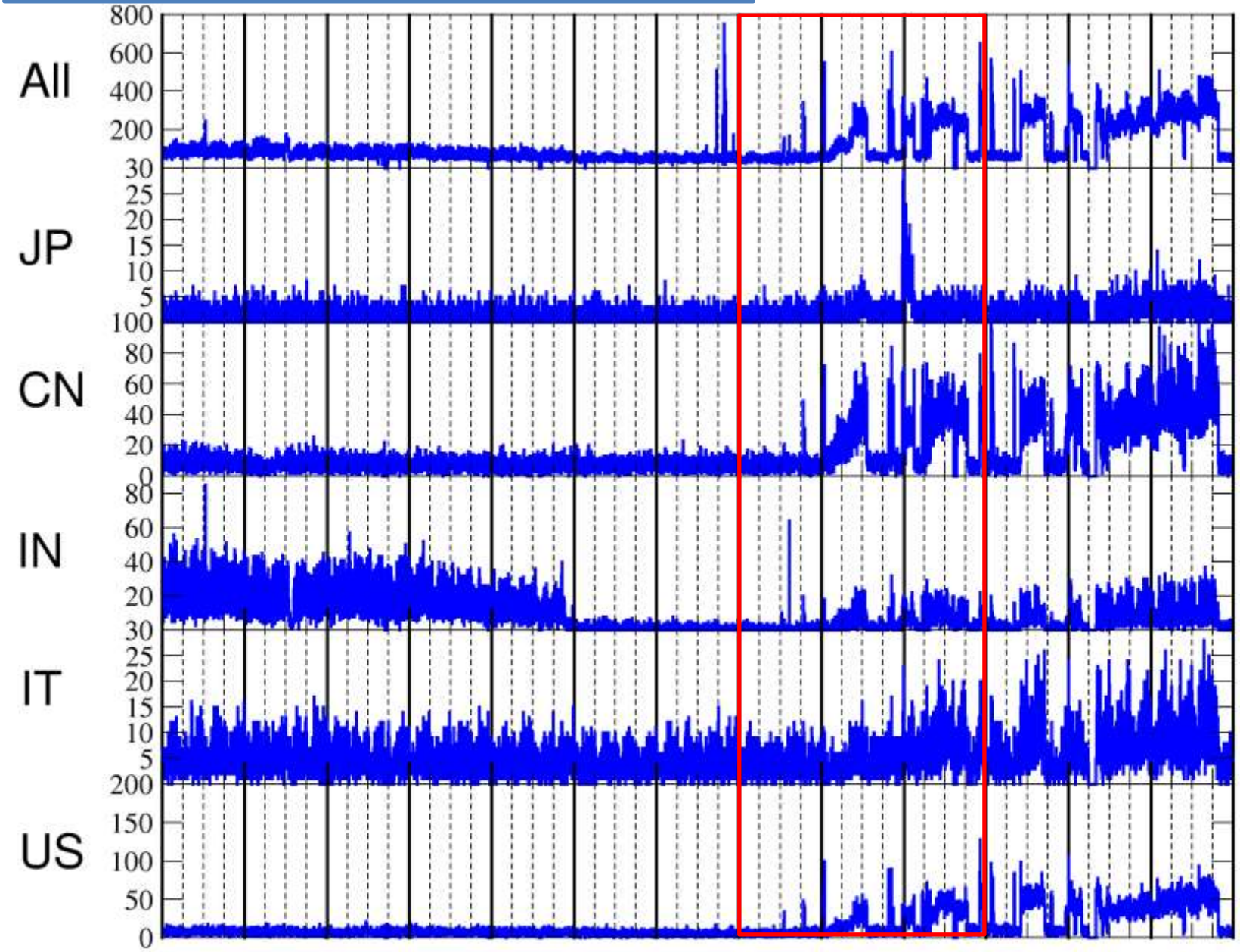
- 通信トラヒックの流量が日変動を示すのは良く知られた経験則
  - 人間活動の周期性とリンク
- 感染端末の活動＝PCがアクティブな時間帯に限られる⇒人間の活動時間帯と相関を持つ
- 同一周期(24/7)で振幅（端末数）と位相（時差）が異なる波の重ねあわせ
  - 特に振幅が大きい＝ボットが多数生息する地域のタイムゾーンで全体の傾向が決まる

平和の終焉

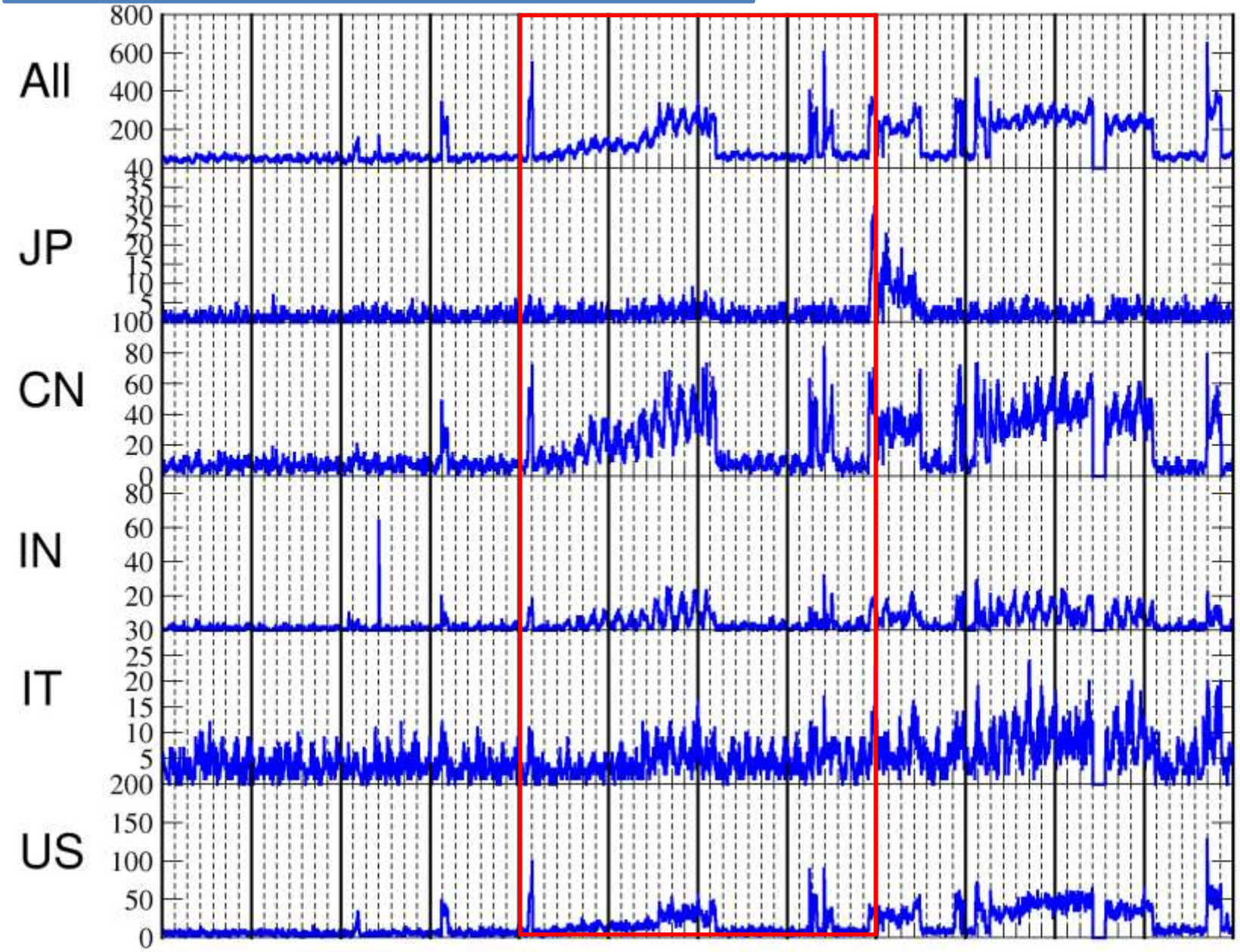
2011/1/1(土)～12/30(金)　(52週間)



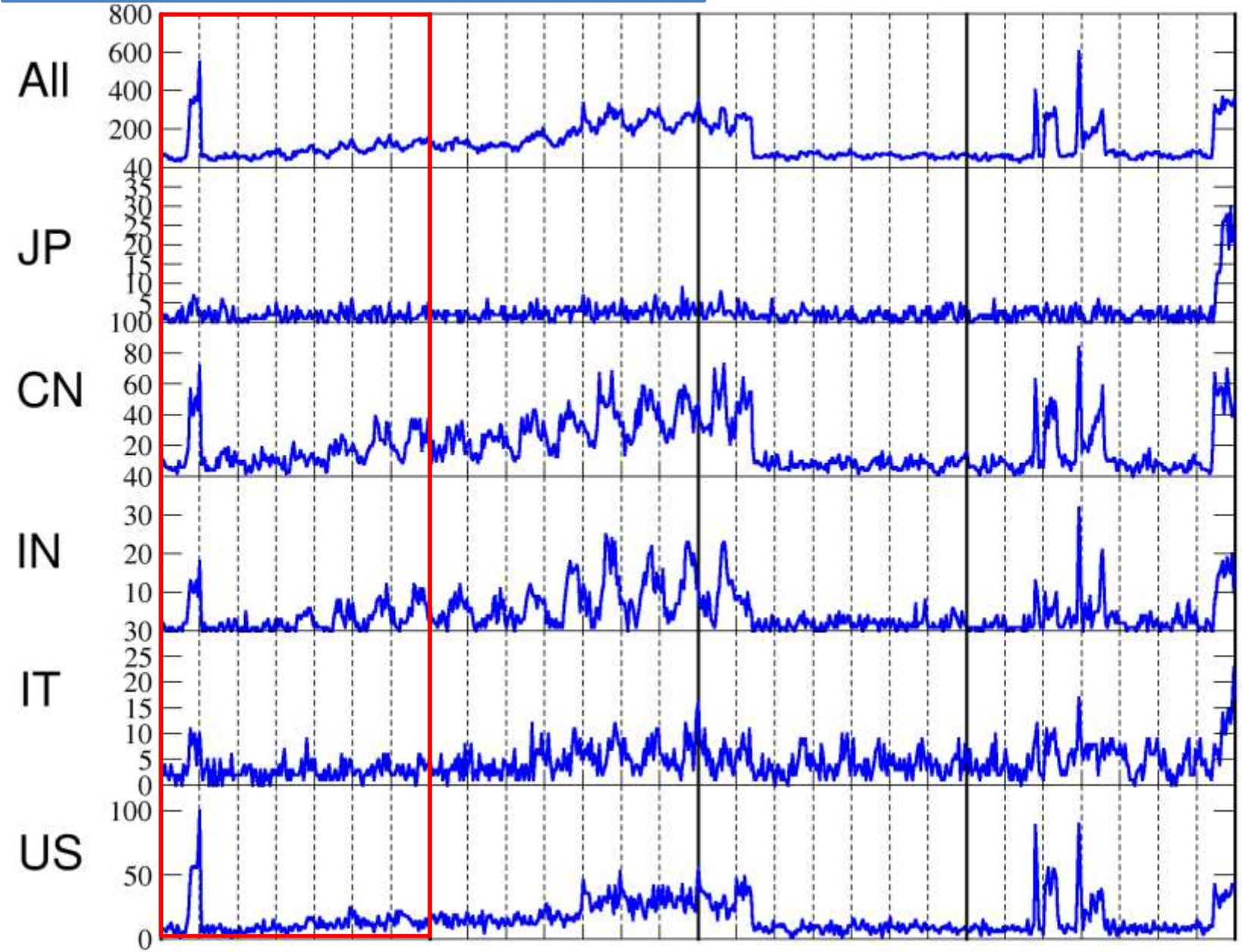
2011/1/1(土)～12/30(金)　(52週間)



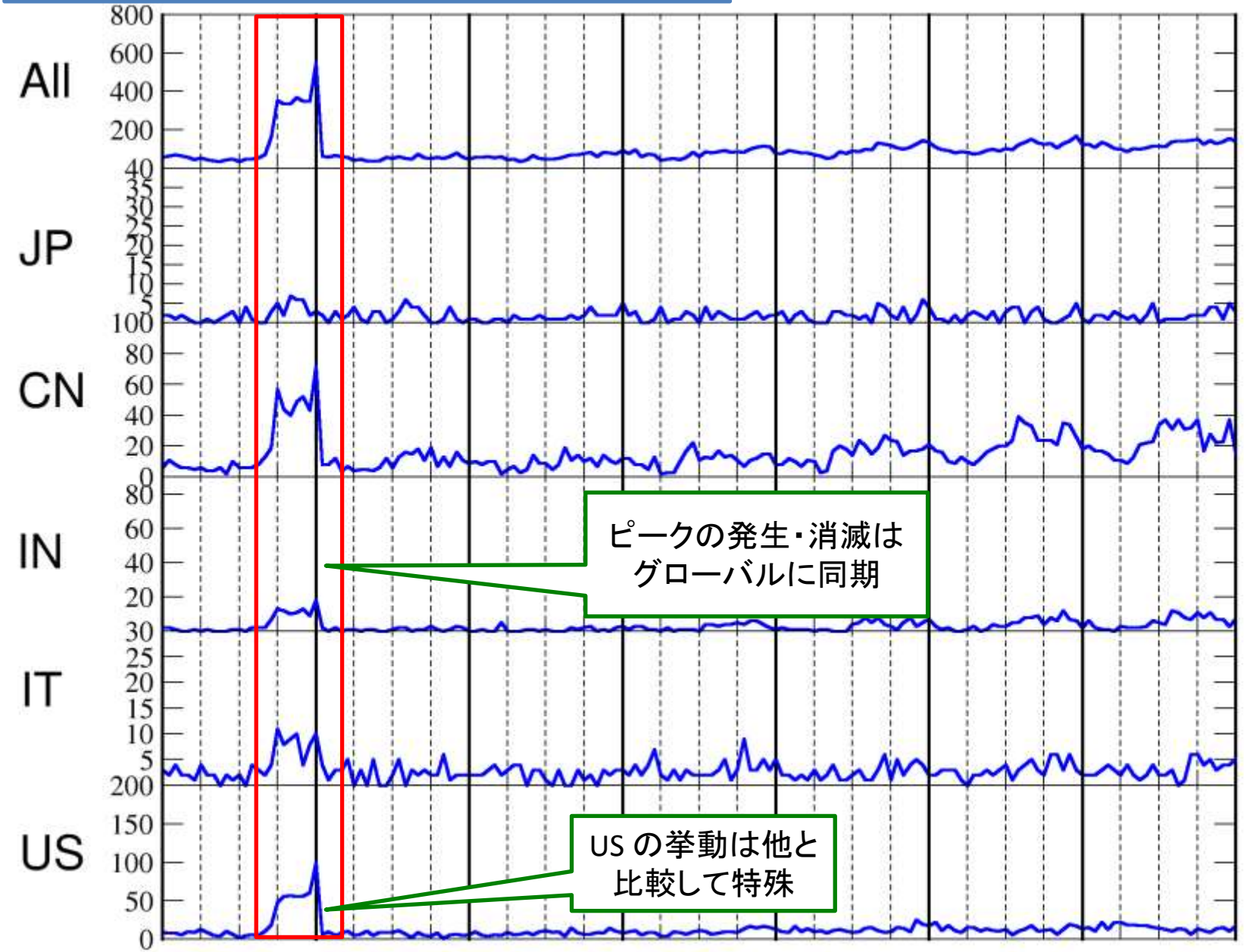
2011/7/16(土)～10/7(金)　(12週間)



2011/8/13(土)～9/9(金)　(4週間)



2011/8/13(土)～8/20(金)　(1週間)



流量急増の原因分析

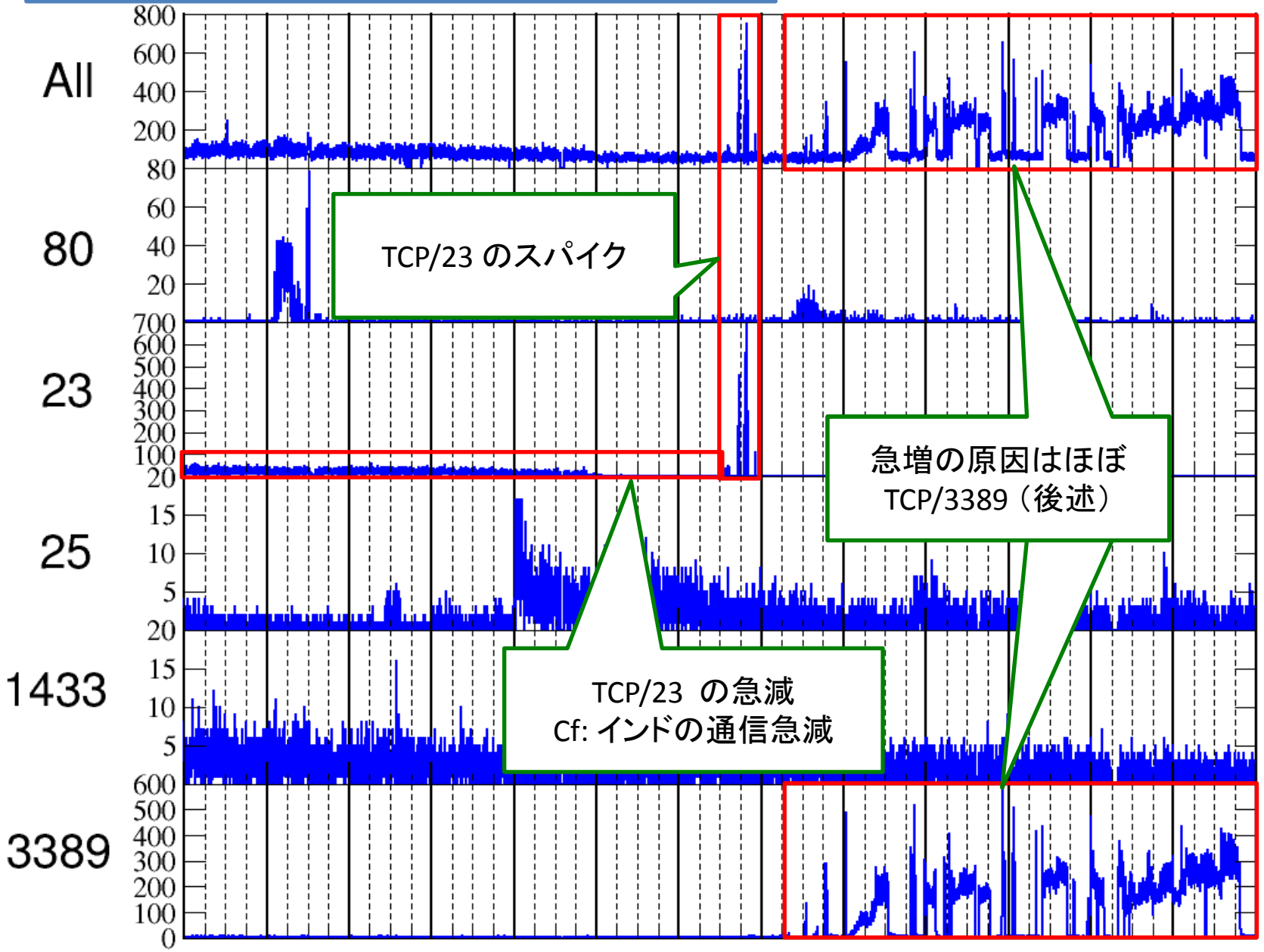
- 1つのTCPパケットからわかる情報として以下を使う（分析の軸を増やす）

- ポート番号の調査
  - どのサービスをターゲットとしたものか

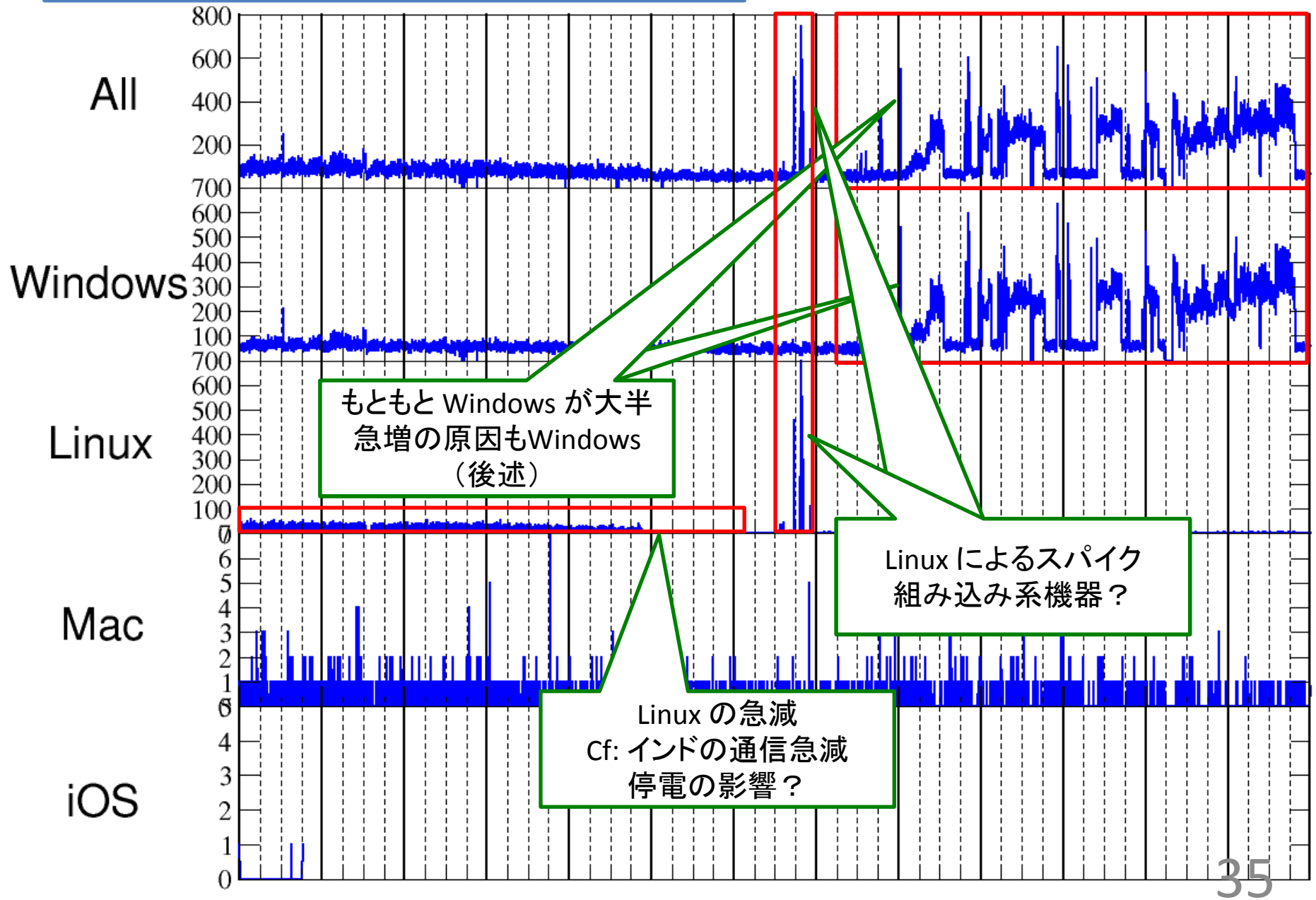
- OSの調査
  - どの OS に関連があるか
  - TCP fingerprinting により，パケット送信元の TCP/IP stack (OS依存) がわかる

2011/1/1(土)～12/30(金)　(52週間)

ポート番号別パケット数



2011/1/1(土)～12/30(金) (52週間) OS別パケット数

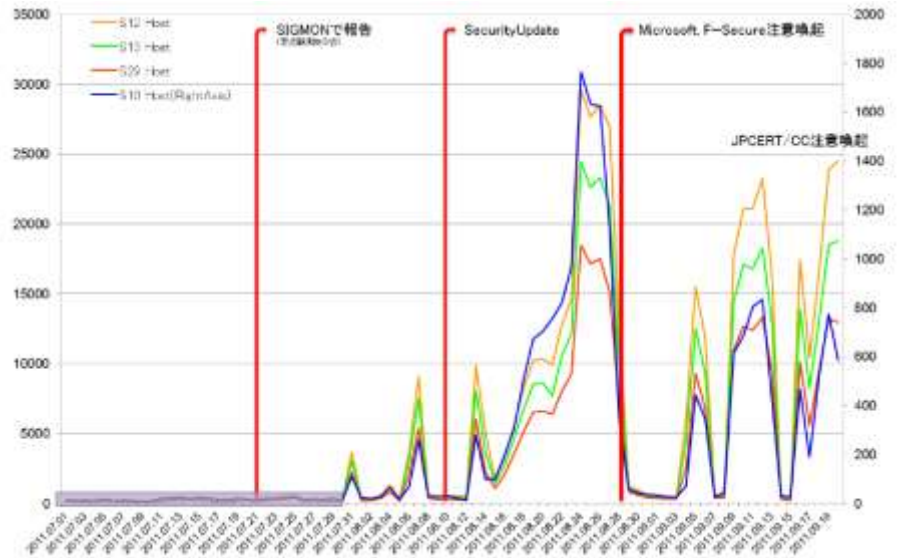


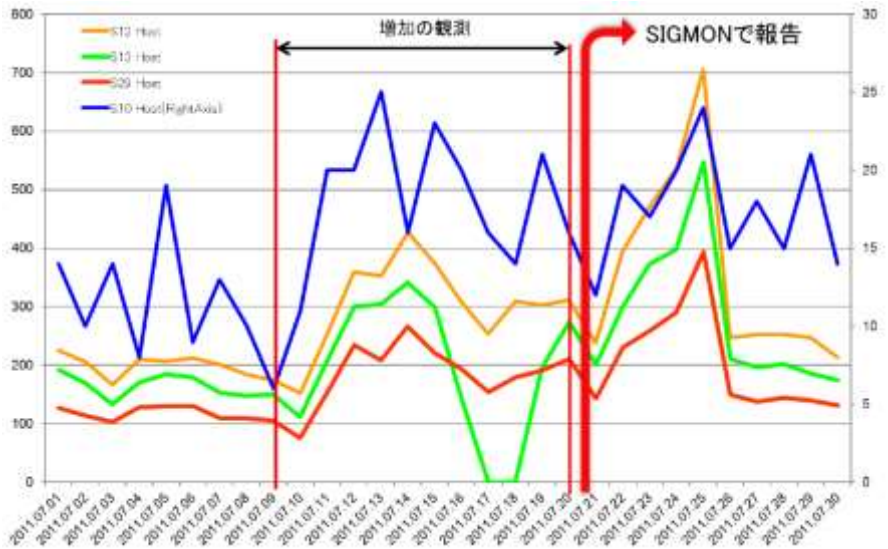
TCP/3389 + Windows → Worm:Win32/Morto

- Windows の Remote Desktop Services 上で感染するワーム
- Remote Desktop Protocol (RDP) の通信が TCP/3389 で行われることを利用
- 通信成立後は管理者権限でのログインを試みる
- 攻撃成立後はマルウェアを実行し，レジストリファイル改変，アンチウィルス無効化，特定のサーバ(C&Cと考えられる)への通信を行う

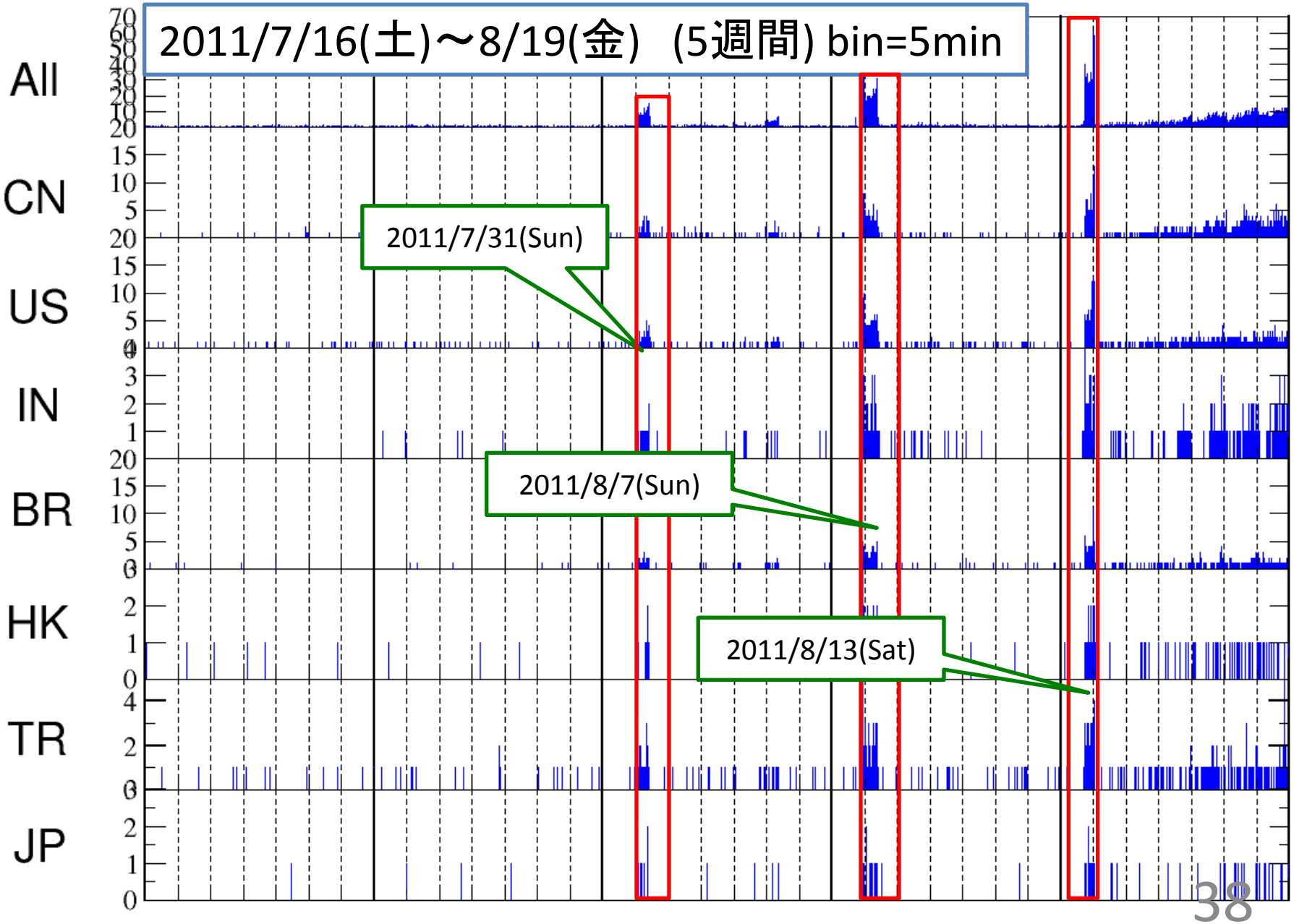
Win32/Morto の活動開始時期

- NICT笠間氏による nicter を用いた分析結果（下図）
  - http://www.iwsec.org/mws/2013/files/nicterdarknet_Dataset_2013.pdf
  - Outbreak が起こるよりも早い時点で予兆検知ができていた
  - ※今回分析した/20 のデータではそのトレンドはきれいに見えなかった
    ⇒アドレス空間がより広いダークネットが有効

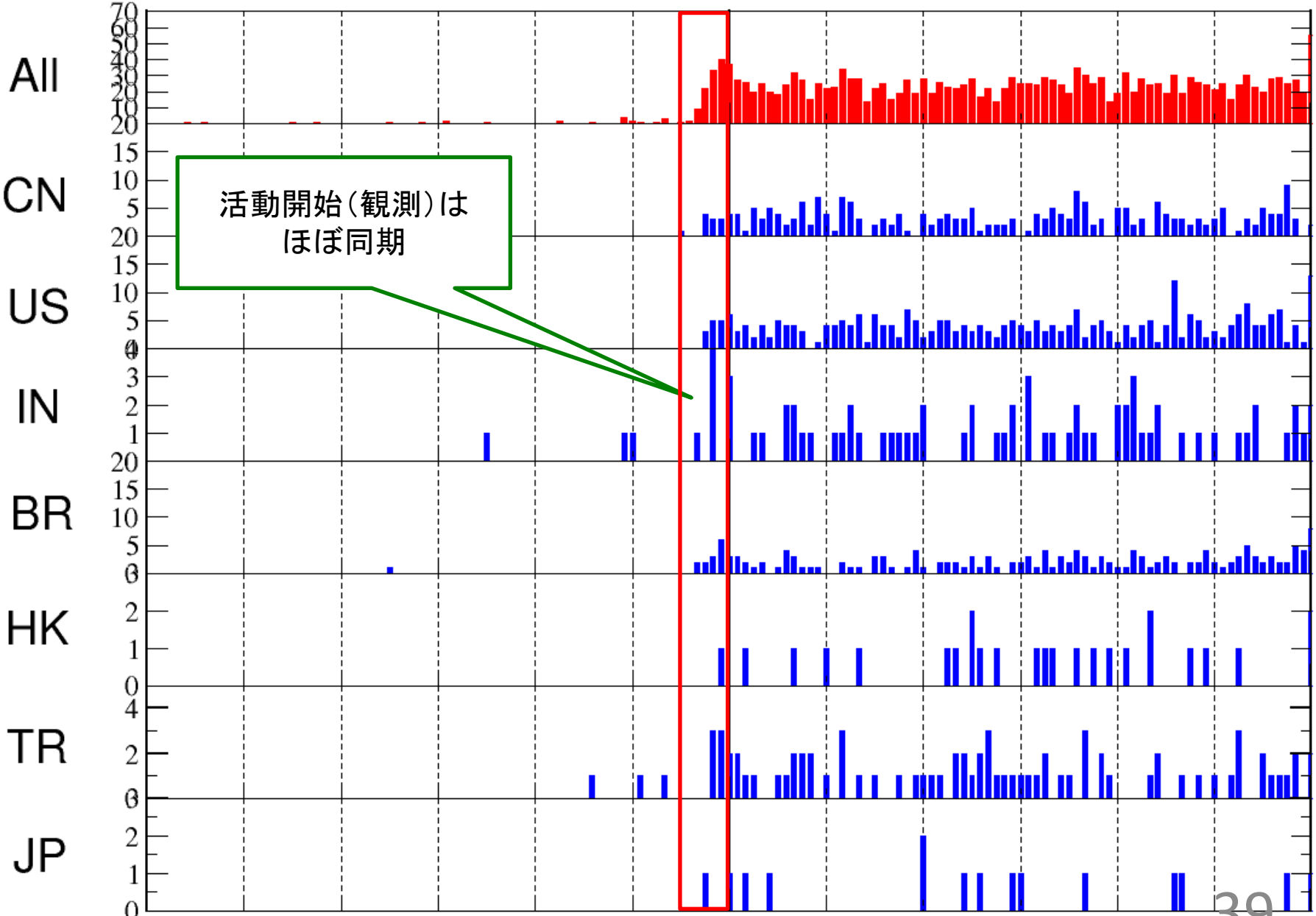




Outbreak 時の TCP/3389 + Windows



2011/8/13(土) 18:00～24:00 bin=5min



Questions

- 攻撃者のダイナミクス（間欠性）は何によって決まっているのか
  - 自然現象ではない
  - ボットのレンタル期間？オペレーションの都合？
- 時々出現する「不自然」なスパイクの要因
  - 双方向通信ではないため，原因推測できないことが多々ある（得られる情報が少ない）
  - 間接的に他の外部要因に結びつけるしかない

  →多変量解析による自動化ができないか？
- 他のダークネットとの相関
  - パケット送信先は意図的に変更できる⇒原理的に見えないものが存在
  - 複数の観測点の情報を総合的に見る（開口合成？）

その後の世界

講演当日投影資料

ダークネット通信に見られた規則性とセキュリティ対策への応用

- 短期間ではランダム到着の性質を有する
- 中期間では人間の社会活動に紐づく強い周期性
  - 周期性の崩れ⇒異常の可能性
- 特定の軸に沿った分類の有効性
  - 原因特定（原因は必ずある）
  - ヒューリスティックや経験則⇒自動化の需要
- 異常の発生パターン
  - 早期発見⇒早期対策
- 長期トレンド（今後の課題）
  - 定点観測により大きなトレンドを発見できるか？
  - 予測に活かせるか？

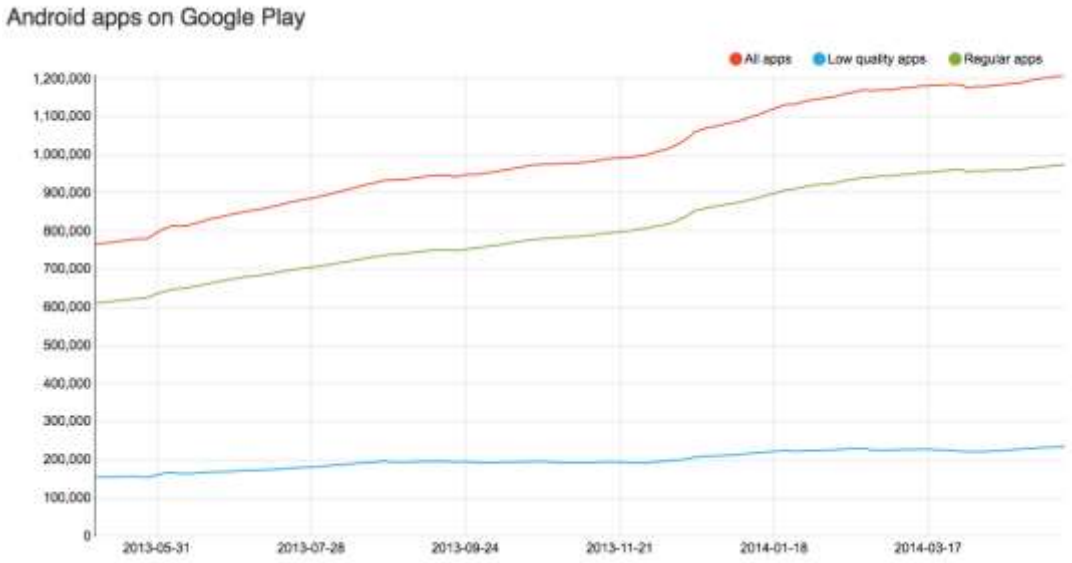
NSL
Networked Systems Laboratory

2　モバイルアプリの解析

モバイルアプリ

- 多くの場合，モバイルアプリは Google Play, Apple App Store 等のモバイルソフトウェア配信プラットフォームを経由して配布
- セキュリティ・プライバシーに関するポリシーの違い
  - Google (Android): 最終判断は開発者・ユーザに委ねられる
  - Apple (iOS): アプリ登録前に Apple 社で厳密な調査を行う（手動）
  - サードパーティマーケット: 玉石混交．問題のあるサイトも少なくない

モバイルアプリマーケットの規模感



出典: AppBrain Stats: http://www.appbrain.com/stats/number-of-android-apps

Table 1. Mobile App Store Downloads, Worldwide, 2010-2016 (Millions of Downloads)

	2012	2013	2014	2015	2016	2017
Free Downloads	57,331	92,876	127,704	167,054	211,313	253,914
Paid-for Downloads	6,654	9,186	11,105	12,574	13,488	14,778
Total Downloads	**63,985**	**102,062**	**138,809**	**179,628**	**224,801**	**268,692**
Free Downloads %	*89.6*	*91.0*	*92.0*	*93.0*	*94.0*	*94.5*

Source: Gartner (September 2013)

出典: **Gartner Says Mobile App Stores Will See Annual Downloads Reach 102 Billion in 2013**
http://www.gartner.com/newsroom/id/2592315

マルウェアの拡大

- Kaspersky によるサードパーティも含めた Android マーケットを対象にした調査
- 累積して 1000万以上のユニークなマルウェアを検出
  - 公式マーケットよりもサードパーティの方がアプリ数が多い

Number of the week: list of malicious Android apps hits 10 million

http://www.kaspersky.com/about/news/virus/2014/Number-of-the-week-list-of-malicious-Android-apps-hits-10-million

プライバシー窃盗・漏洩の問題

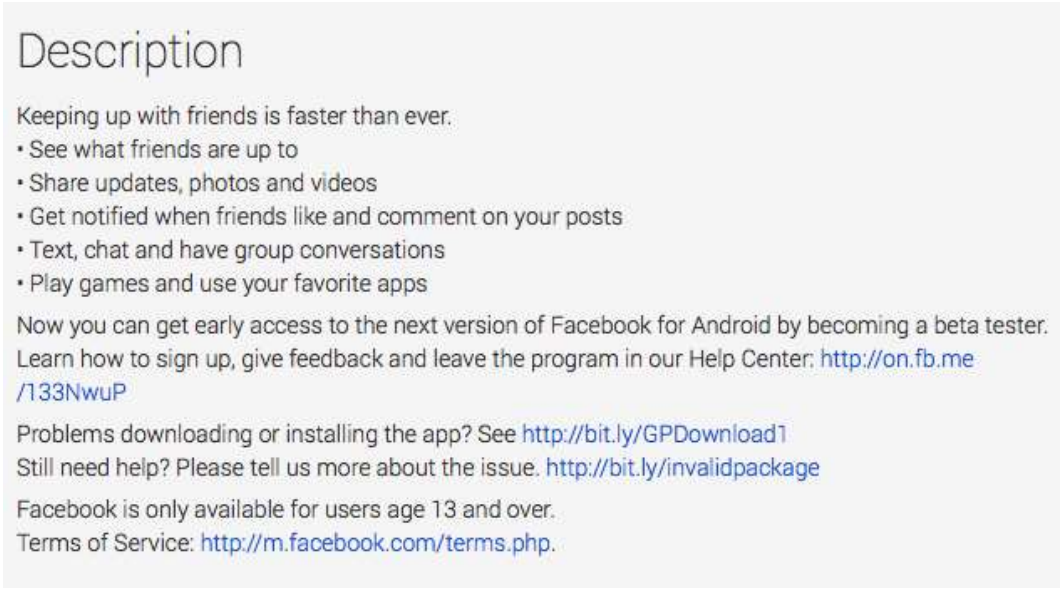
- マルウェアであるか否かに関わらず，ユーザの意図していないところでプライバシー情報を収集・濫用される問題
  - ロケーション
  - コンタクトリスト
  - SMS
  - カメラ（視覚）

プライバシー対策と課題

- 対策例
  - インストール時にアプリが使用する “Permission” を明示的に開示
  - Privacy Policy の記載（開発者側の対策）
- 課題
  - Permission に注意を払うユーザは少ない
    - インストール時にpermission に注意を払ったユーザは17%のみ．Felt et al. SOUPS 2012
  - Privacy Policy を正しく設置する開発者は少ない
    - 半数のアプリは Privacy Policy をつけていない
      FPF Report 2012

ユーザが注意を払う情報源

- アプリに対する Description (詳細説明)
  - 検索の対象
  - 全文とは言わずとも新規アプリであれば目は通す
  - 開発者はユーザの関心を引くように書く動機がある.



https://play.google.com/store/apps/details?id=com.facebook.katana

仮説

重要なプライバシー情報にアクセスする旨を Description に記述していない不自然なアプリには問題がある

⇒そのようなアプリは詳細検査する価値がある

※全アプリの詳細検査は原理的に不可能な規模

背後に潜むであろう普遍則

後ろ暗い事は隠しておく！

アプローチ

静的コード解析　＋　自然言語処理

アクセスする＝API/URIが存在

逆アセンブルコード

```
.super Landroid/view/SurfaceView;
.source "CameraSurfaceView.java"

# interfaces
.implements Landroid/view/SurfaceHolder$Callback;


# static fields
.field private static final SDF:Ljava/text/SimpleDateFormat;


# instance fields
.field private final ReadPreviewImage:Landroid/hardware/Camera$Previe

.field camera:Landroid/hardware/Camera;

.field cameraParameters:Landroid/hardware/Camera$Parameters;

.field prevdata:[B
```

言及する＝単語が存在

Description

详细描述：

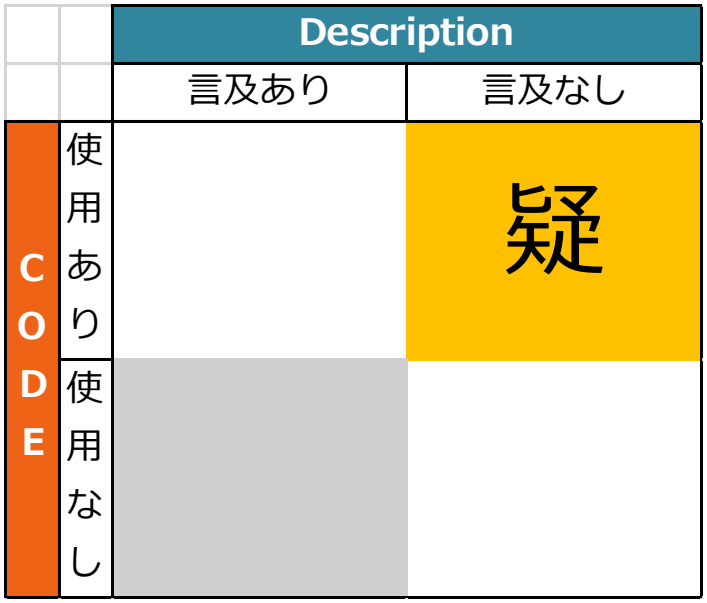
SuperSU（超级授权管理）是一款功能超强超级用户访问权限管理工具.Super
有应用的超级用户访问权限的高级管理,在另外一个超级用户管理工具的基础上,

主要功能：
- 超级用户访问权限提示
- 超级用户访问权限日志记录
- 超级用户访问权限通知
- 配置各应用的通知规则
- 永久取消ROOT权限
- 深度处理检测
- 可在Recovery中工作
- 可在Android系统未能正常启动时工作
- 可在不标准的Shell位置中工作
- 始终运行在幽灵模式中
- 提示时唤醒设备

NSL Networked Systems Laboratory

狙い

- コードと description に齟齬があるアプリケーションを自動で検出
  - コード解析: API call graph analysis
  - Text 解析:　NLP + Supervised ML



大量アプリの解析

- 収集検体数:
  - 英語 10万検体，中国語10万検体
  - Google Play およびサードパーティマーケット
- 教師ラベル
  - 英語: 1000検体，中国語1000検体
    - 各検体を3人がラベル付与
- Sensitive resource
  - Location, Contact list, Camera

講演当日投影資料

まとめ

- セキュリティ課題における「敵を知る」営み
  →不自然さに潜む規則性 (invariant) の発見と応用
- データ解析がメイン
  - 様々な解析手技を駆使
  - 時系列解析、機械学習、パターン認識、自然言語処理、コード解析、リバースエンジニアリング
- 研究の特徴
  - Moving target のため iteration が早い
  - 応用研究であり，出てきた結果が実益に供しやすい
    - その反面では一般性を確立する営みを大切にしたい

謝辞

- 貴重な講演の場をご提供頂きました情報ネットワーク科学時限研究専門委員会の皆様に感謝致します
- 貴重なダークネットデータをご提供頂いた情報通信機構に感謝致します
- ダークネット解析およびAndroid アプリ解析を実施した早稲田大学基幹理工学部の笹生憲君，渡邉卓弥君に感謝します．
- Android アプリ解析にご協力頂いたNTT セキュアプラットフォーム研究所の秋山満昭氏に感謝します．