# NICTER DARKNET DATASET データ解析のノウハウ

早稲田大学 森達哉 研究室

芳賀夢久，笹生憲

2015/05/13

MWS意見交換会

# 目次



- NICTER Darknet Dataset について
- ダークネットの研究事例
- データ解析のノウハウ

# 目次

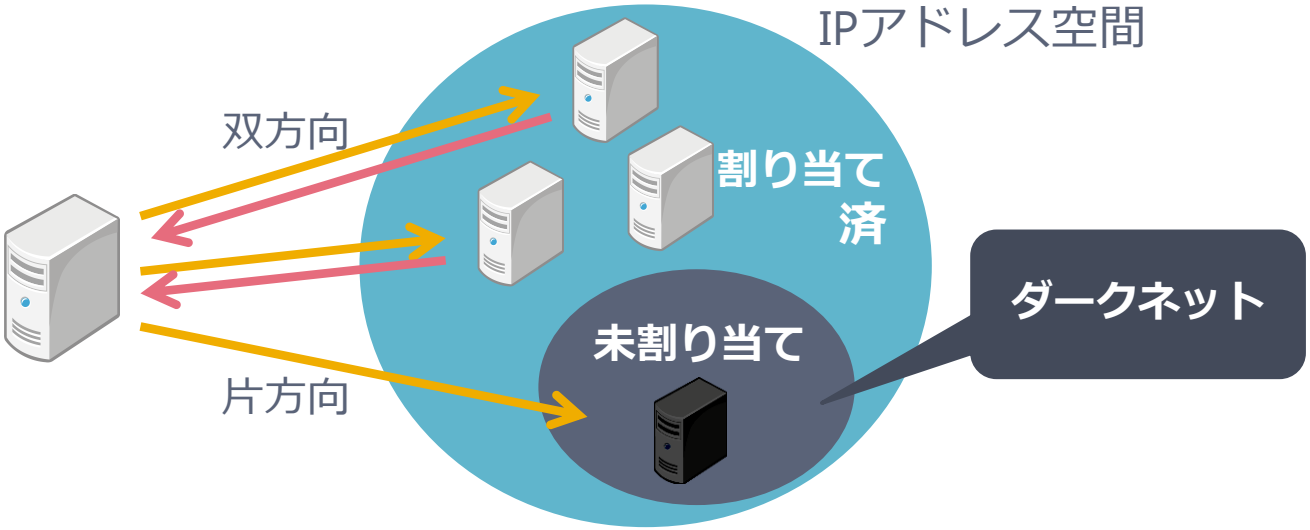


- NICTER Darknet Dataset について
- ダークネットの研究事例
- データ解析のノウハウ

# ダークネットとは



- **ダークネット**
  - インターネット上で**到達可能**かつ**未使用**のIPアドレス空間
  - 外部からのアクセスに対して一切のレスポンスを返さない



# Dataset

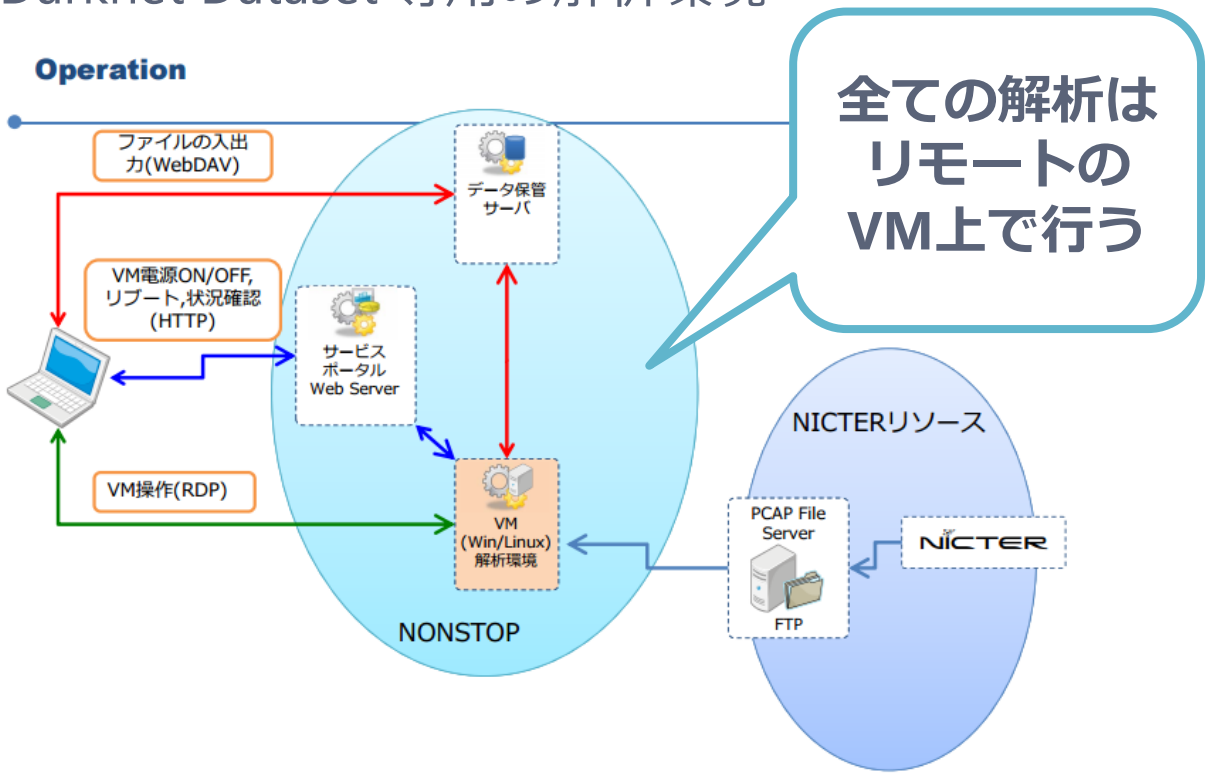


- **NICTER Darknet Dataset**
  - 情報通信研究機構により提供された通信データ
  - 観測対象はある1組織(/16)のダークネット(**/20**)
  - PCAP形式のデータを**NONSTOP**上で解析

# NONSTOP



- **NONSTOP (Nicter Open Network Security Test-Out Platform )**
  - NICTER Darknet Dataset 専用の解析環境




引用: http://www.iwsec.org/mws/2014/files/NICTER_Darknet_Dataset_2014.pdf

# 目次



- NICTER Darknet Dataset について
- **ダークネットの研究事例**
- データ解析のノウハウ

# ダークネットの研究事例

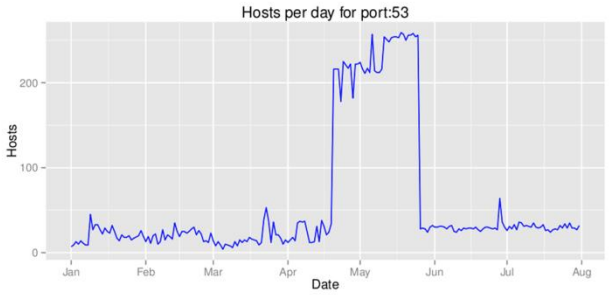


- 笹生，森，後藤，”**通信源ホストの分類を利用したダークネット通信解析**“，MWS 2013, 2013年10月
- 笹生，”**TCP ヘッダーと通信パターンの分類を利用したダークネット通信解析**”，The 10th IEEE Tokyo Young Researchers Workshop, 2013年12月
- 芳賀，笹生，森，後藤，”**リフレクター攻撃における増幅器探索通信の解析**”，MWS 2014，2014年10月
- 芳賀，笹生，森，後藤，"**インターネット計測はダークネット解析のノイズになるか？**"，IA/ICSS 6月研究会，2015年6月

# ダークネットの研究事例

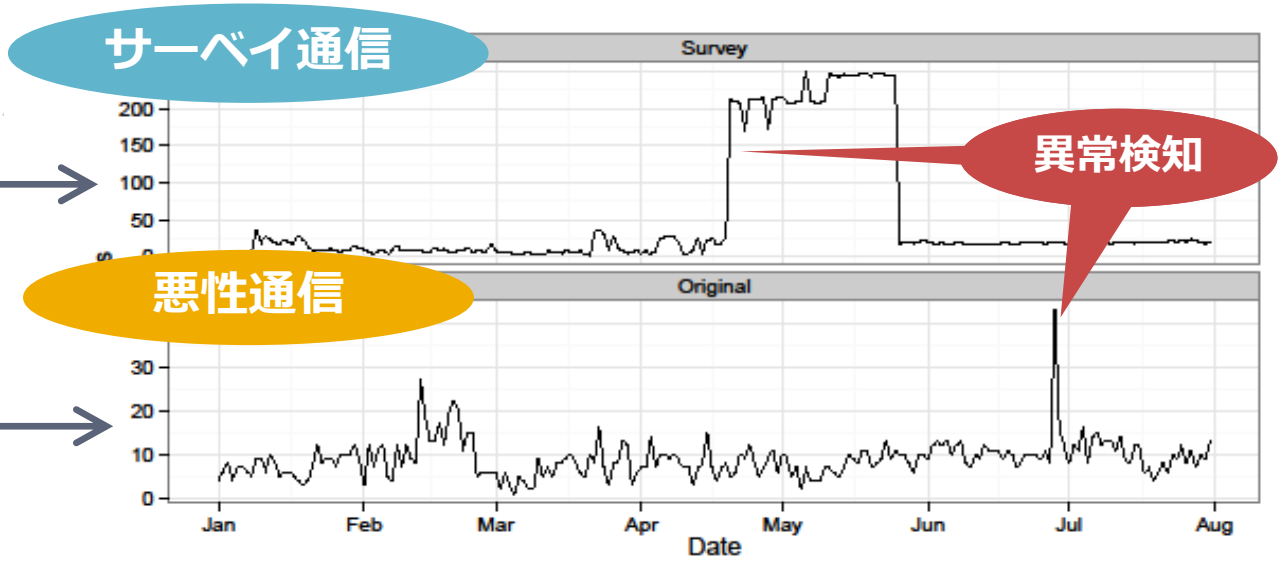


- **インターネット計測はダークネット解析のノイズになるか？**




ダークネットのDNSの通信を分類

分類
**(症例対照研究)**



# 目次



- NICTER Darknet Dataset について
- ダークネットの研究事例
- **データ解析のノウハウ**

# データ解析の流れ

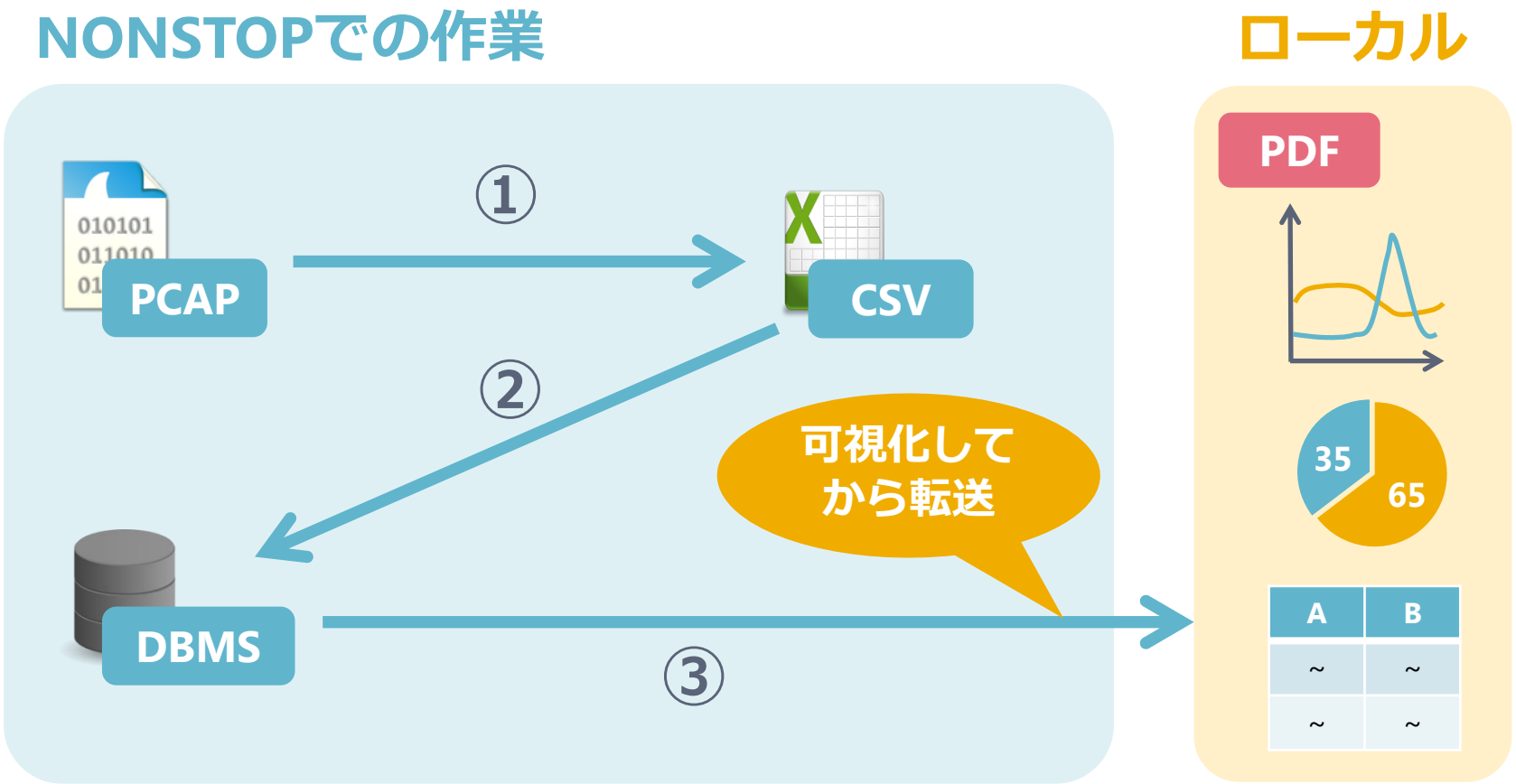


1. **生データ(pcap)の前処理**
   - tshark, p0fなどを用いてCSVファイルに出力
2. **中間データの管理**
   - DBMS: MySQL, SQLite
   - 適宜 外部データなどを追加（IPアドレスのBlacklistなど）
3. **分析**
   - 可視化ツール: ggplot2
   - 基本的な統計解析: R
   - より進んだ統計解析: 症例対照研究，異常検知(SDAR)

# データ解析の流れ



- PCAPデータから図表の出力まで



# データベースについて



- **スキーマ例**

| frame_date | frame_time | ip_src | ip_dst | srcport | dstport | … |
|---|---|---|---|---|---|---|
| 日付 | 時間 | 送信元IP | 送信先IP | 送信元ポート | 送信先ポート | |

- **SQL文例 (SQLite)**
  - 1日毎の80番ポート宛にパケットを送信したユニークなホスト数を知りたい場合

**SELECT frame_date, COUNT(DISTINCT ip_src) FROM table WHERE dstport == “80” GROUP BY frame_date;**

# ツールについて



- **tshark**
  - WiresharkのCUI版
  - パケットのヘッダ情報をCSV形式で出力可能
  - フィルタ機能に優れ，出力するパラメータを細かく指定できる
  - コマンド例（HTTPS をターゲットとした probe packet の場合）

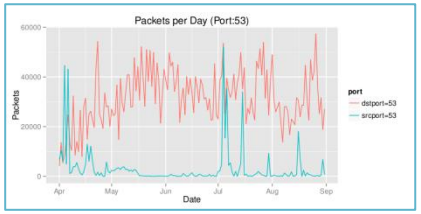
    **tshark -r file.pcap -R “tcp.dstport==443” -T fields -e frame.number -e frame.time -e ip.src -e ip.dst -e tcp.srcport -e tcp.dstport -E separator=¥|**

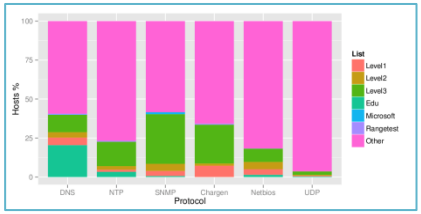
- **p0f (Passive OS Fingerprint)**
  - パケットのTCP/IPヘッダを元にOSを推定する
  - 基本的なヘッダ情報+OS情報をCSV形式で出力可能
  - http://lcamtuf.coredump.cx/p0f3/

# ツールについて



- **Python**
  - ライブラリが豊富 (GeoIP, changefinder, sqlite3, など)
  - 基本的な分析モデルを扱え，数値計算，文字列操作に強い
- **R**
  - 「データフレーム」により統計解析がしやすい
  - データや分析結果の可視化に優れている
  - 可視化にはggplot2 パッケージを使用

# 分析手法の例



- **送信元のOS推定**
  - Passive OS fingerprint (p0f)
- **送信元のロケーション推定**
  - Geo IP
- **スキャンツール(Zmapなど)の分別**
  - スキャンツールの動的解析によりシグネチャを作成
- **スキャンパターンの分析**
  - 送信元/先IPアドレスを元に協調型のスキャンなどを特定

# NOSTOPでのデータ解析のコツ



- **データベースがあれば解析がしやすくなる**
  - **NICTER提供のデータベース(MySQL)を利用してみる**

- **データ整形にかける時間をできるだけ短縮する**

  限られたマシンリソースを効率的に使うために
  - tsharkなどのフィルタ機能を使って余分なパケットを削ぎ落とし，PCAPデータのサイズを小さくしておく
  - 汎用的なデータベース設計を心がける
  - データ整形はスクリプトを用いてできるだけ自動化する
  - メモリ効率を意識したプログラミングを行う

# OSINT (Open Source Intelligence)



- 「**合法的に入手できる公開情報**」を
  「**合法的に調べ突き合わせ分析する**」手法
  - 報道，インターネット，書籍など
- **例**

**2014年7月**
**ダークネットにおける**
**23/TCPポートの通信が**
**増加**

**検索**

**ルーターを踏み台にした**
**スキャンの可能性(警察庁)**


HOME > レポート > 調査・ホワイトペーパー

ビル管理システムに対する探索行為が継続―インターネット治安情勢(警察庁)

2014年8月29日(金) 08時00分

ツイート 9 いいね！ 2 g+1 0 チェック ブックマーク 0 Pocket 4 メールマガジン(無料会員登録)

▸【特集】警察庁
├金融機関のフィッシングサイトが増加、暗号化されていないケース目…
├2月の特殊詐欺状況を発表、1月と当面同水準で推移(警察庁)
└複数のNoSQLデータベースに対する探索行為、高水準で推移(警察庁)

Ads by Google ▸ Port scan ▸ Bcp対策 ▸ Bcp事例

警察庁は8月27日、@policeにおいて2014年7月期のインターネット観測結果等を発表した。7月期では、「ビル管理システムに対する探索行為の継続」「宛先ポート23/TCP（に対するアクセスが増加」をトピックに挙げている。ビル管理システムに対する探索行為の継続については、ビル管理システムで使用される通信プロトコル用標準規格「BACnet」に定義された47808/UDPへのアクセスを分析したところ、6月下旬頃から検知した「ReadPropertyMultiple」のパケットを、今期も継続して観測している。

宛先ポート47808/UDPに対するアクセス件数の推移

宛先ポート23/TCPに対するアクセス件数の推移

また、7月下旬頃から宛先ポート23/TCP（に対するアクセスが増加している。23/TCP（はTelnetに使用されるポートであり、遠隔でネットワーク機器等に接続する際に使用されるもの。このポートに対する

**http://scan.netsecurity.ne.jp/article/2014/08/29/34745.html**

# ご清聴ありがとうございました.