機械学習によるサイバーセキュリティとプライバシー保護データマイニングへの取組み

小澤　誠一

神戸大学 数理・データサイエンスセンター
大学院工学研究科

自己紹介

小澤 誠一 (ozawasei@kobe-u.ac.jp)
所属：数理・データサイエンスセンター & 工学研究科
HP: http://www2.kobe-u.ac.jp/~ozawasei/index-j.html
第2次ブームのときからずっと人工知能を研究

◎**研究内容**

ニューラルネット，機械学習，パターン認識，ビッグデータ解析、サイバーセキュリティ，ソーシャルネット解析，スマート農業

◎**主な研究テーマ**

1) 機械学習のサイバーセキュリティへの応用
（ダークネットトラフィック解析を使ったサイバー攻撃検知，SNSやDark/Deep webからのサイバー情報収集など）
2) 機械学習を用いたSNS上のレピュテーションマネージメントに関する研究（炎上検知など）
3) スマート農業に向けた農作物の画像センシング手法の開発
4) 暗号データに対する機械学習アルゴリズムの開発
（プライバシー保護データマイニング）

サイバーセキュリティにおける機械学習の応用

- **マルウエア解析(静的/動的)**
  - ✓ マルウエア検知・分類
- **不正侵入検知**
  - ✓ 異常検知
  - ✓ 攻撃分類・検知
  - ✓ ハニーポット観測・分析
- **ダークネット分析**
  - ✓ 異常検知
  - ✓ 攻撃分類・検知
  - ✓ ボットネットの活動検知・分析
- **Webベース攻撃検知**
  - ✓ 悪性サイト・悪性スパムメール分析
  - ✓ 悪性JavaScript検知
- **サイバー攻撃情報の収集・分析**
  - ✓ 表層Web解析（SNS、セキュリティブログ・レポートなど）
  - ✓ 深層Web解析（ダークマーケット、ダークフォーラムなど）

機械学習のPros と Cons

Pros

- 大量かつ高次元の観測データから知識獲得できる.
- 観測データの追加学習による攻撃の変遷に合わせたアダプティブな異常値検出，分類，予測が可能
- 一定の耐ノイズ性や補間能力（汎化能力）が期待できる.
- 24時間，365日働き続ける.
- 高次元データの分布を可視化して，直感的な解釈を人間に与えられる.
- 機械学習で判定可能なものは自動化し，管理者の負担を軽減

機械学習のPros と Cons

Cons

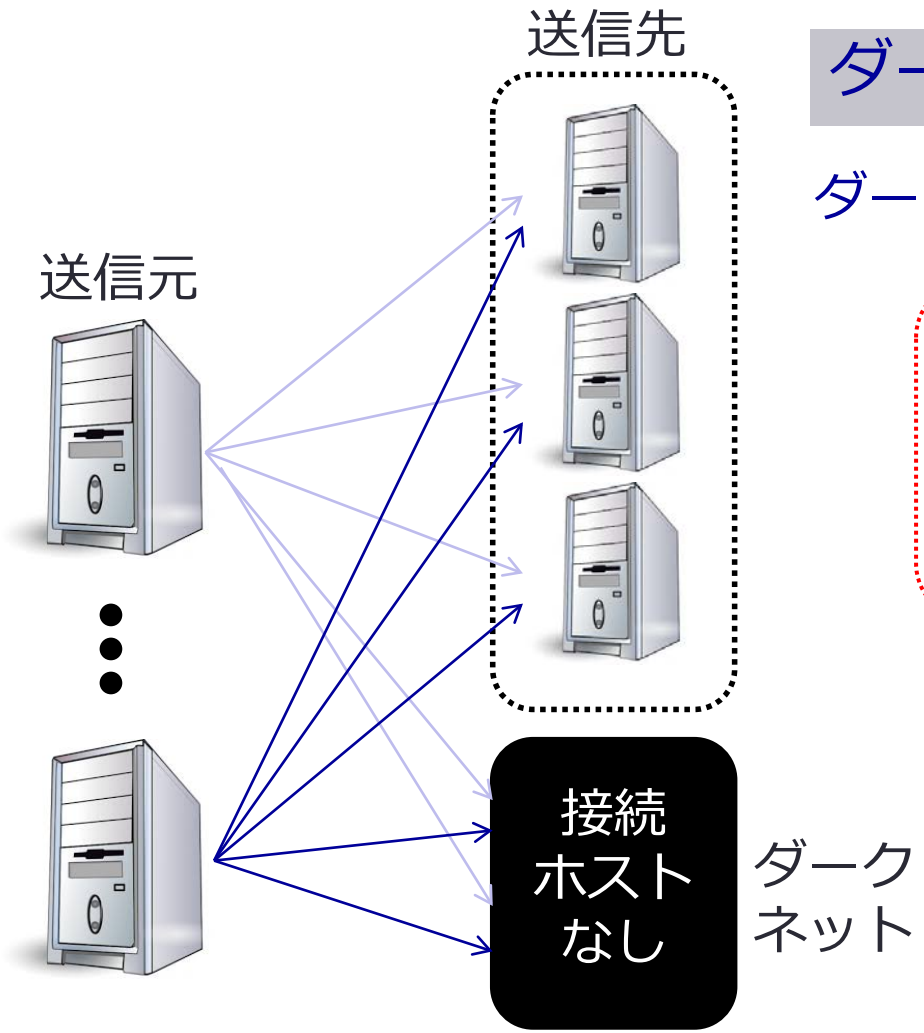
- 攻撃に関連したデータの収集は容易でない.
  - クラス分布の偏り（攻撃事例は極端に少ない）
- クラスラベルが与えられない.
  - 基本的に専門家が付与するが，工夫で，ある程度の自動化が可能．一方，インシデント情報などとの突合でしか得られない場合もあり.
- パケット（ヘッダ情報とペイロード）から目的に応じた特徴量の定義が必要
- 騙されやすい（adversarial setting)
  - Evasion attack：難読化（暗号化，画像ベース）
  - Poisoning attack：訓練データの操作，ラベルの反転など

ダークネット分析事例（1）

DDoSバックスキャッタ検知

JSPS 科研費基盤研究（B）
「サイバー攻撃のリアルタイム検知・分類・可視化のためのオンライン学習方式」

ダークネットとは？



ダークネット＝未使用IP群

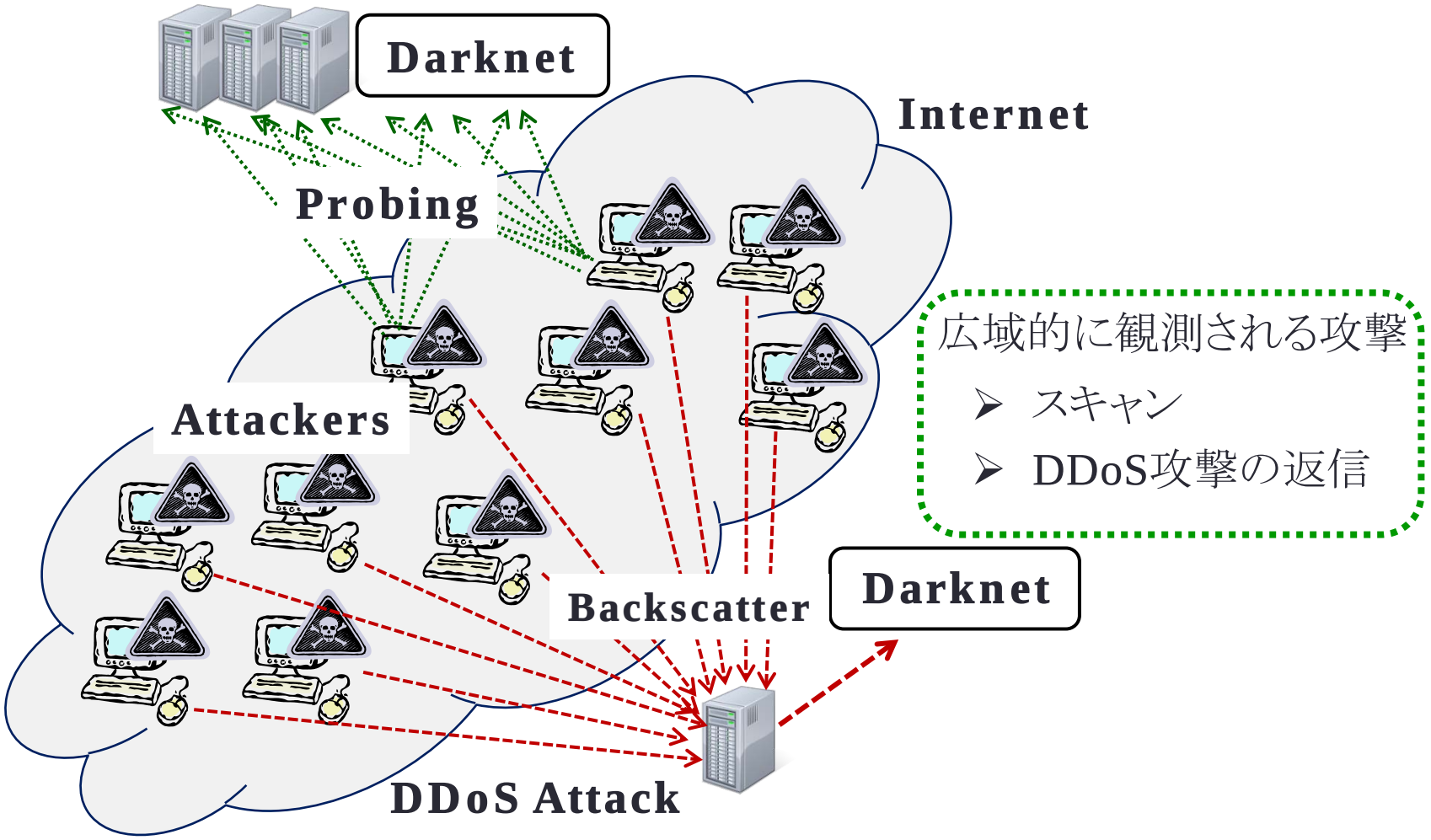
ダークネットにパケットが届く理由

- 設定ミス
- スキャン
- DDoS攻撃ターゲットホストからの返信パケット

サイバー攻撃に関連した通信

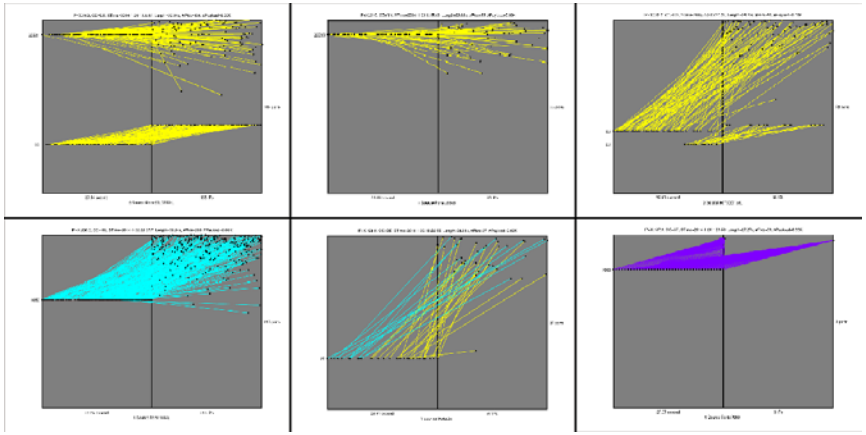
サイバー攻撃の広域観測

観測可能な攻撃は？

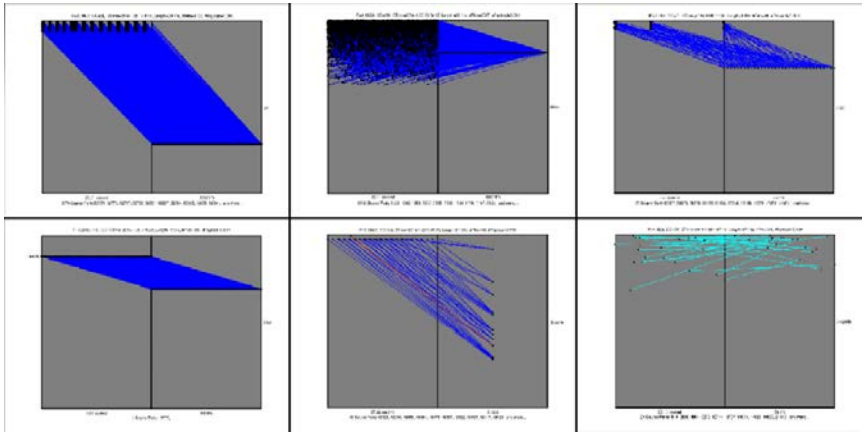

Darknet
Internet
Probing
Attackers
Backscatter
Darknet
DDoS Attack
広域的に観測される攻撃
スキャン
DDoS攻撃の返信

ダークネット・トラフィックの特徴

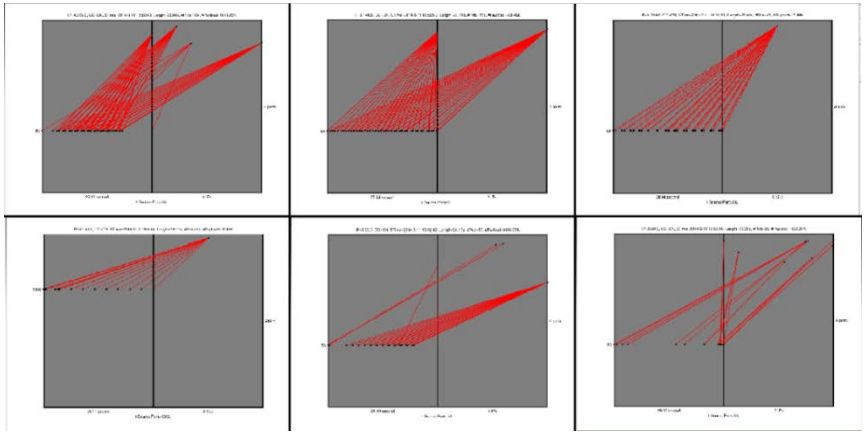
DDoSバックスキャッタ

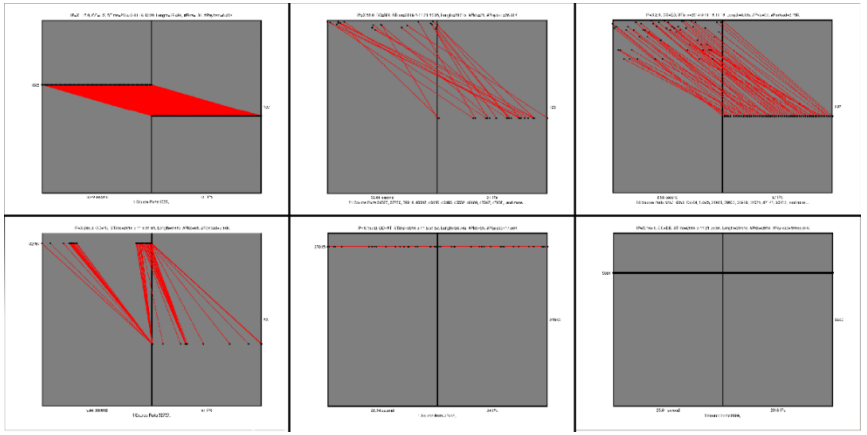


スキャン



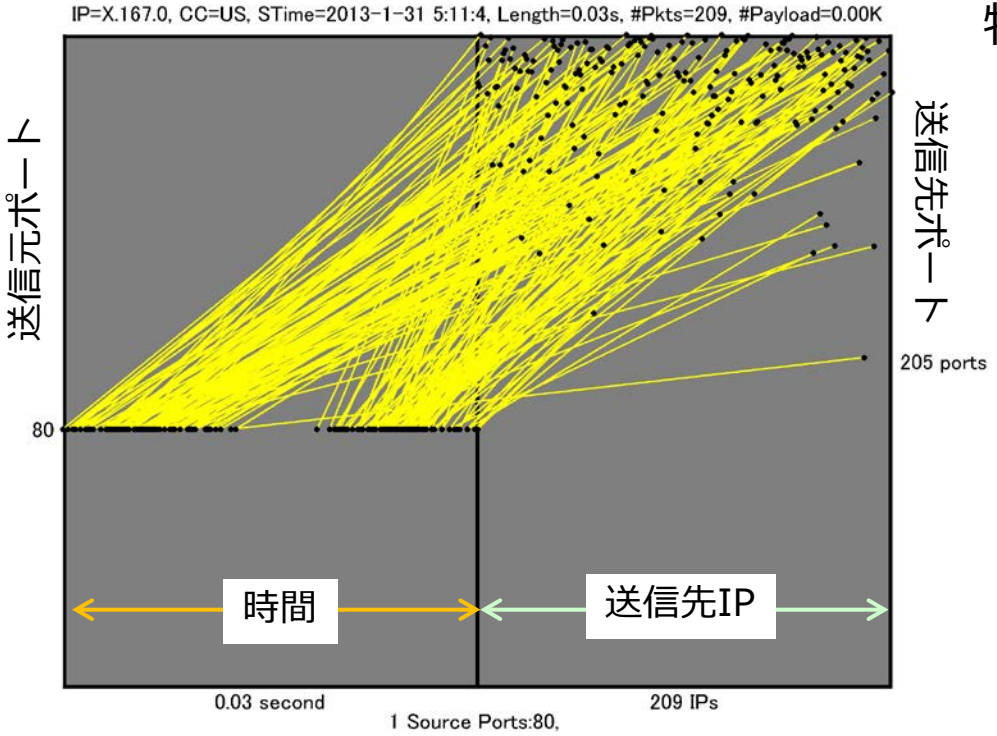
その他





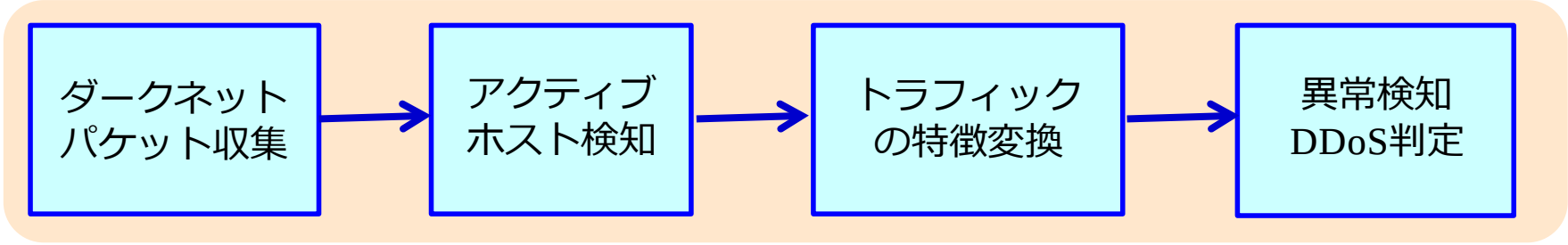
provided by NICT

DDoSバックスキャッター検知

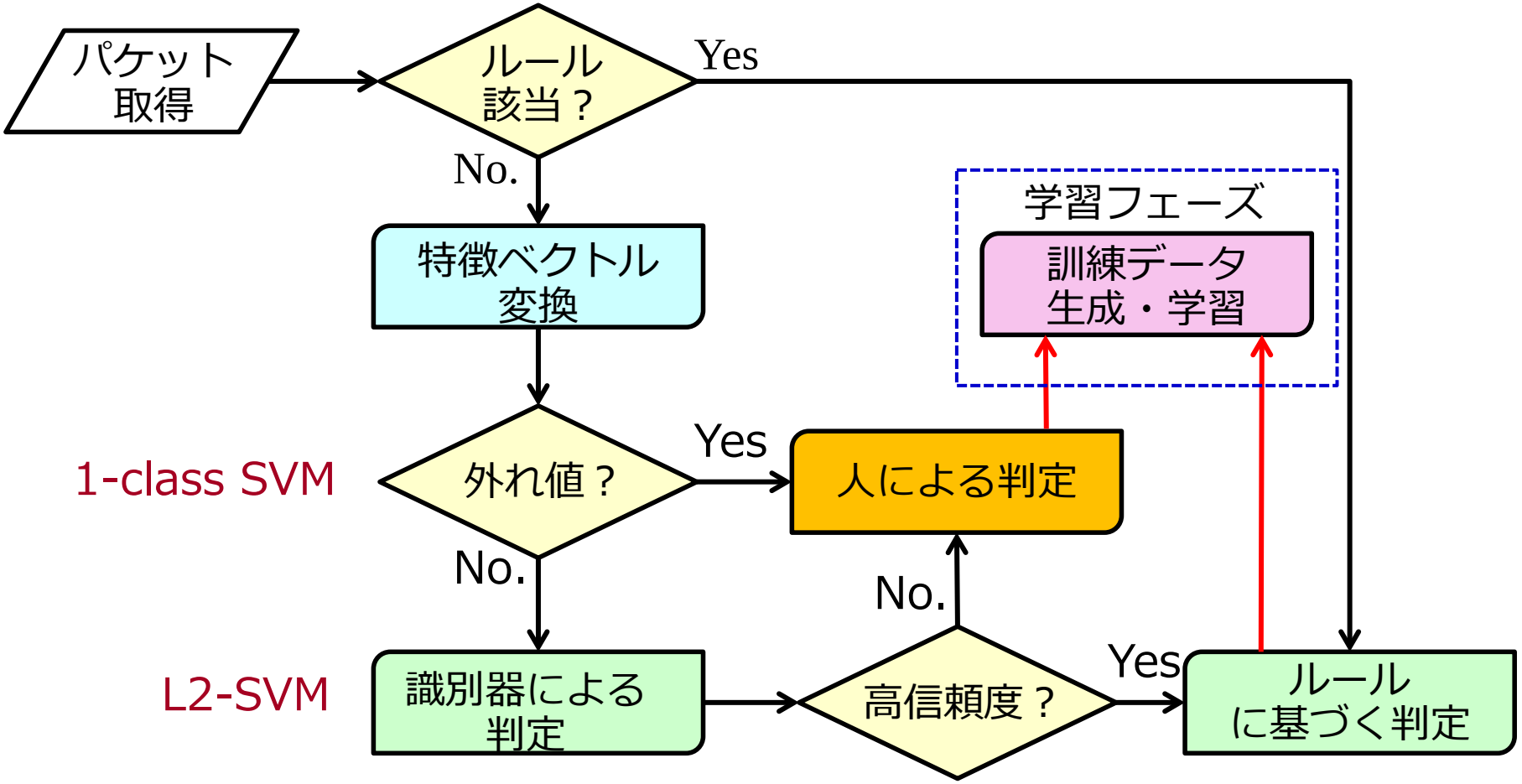


特徴ベクトル (17個)

- パケット数
- パケット間隔（平均・標準偏差）
- 送信元ポート数
- 送信元ポートのパケット数（平均・標準偏差）
- 送信先IP数
- 送信先IPへのパケット数（平均・標準偏差）
- 送信先ポート数
- 送信先ポートのパケット数（平均・標準偏差）
- 送信先ポートパケット間隔（平均・標準偏差）
- プロトコル数
- ペイロードサイズ（平均・標準偏差）



DDoSバックスキャッタ判定のフロー



訓練データのラベリング

（１）ルールに基づくラベリング（自動）

- 既知の攻撃特徴をルール化
- ルールで判定されたものは、そのままラベル情報に使用

（２）識別器によるラベリング（自動）

- 信頼度の高い識別器の判定結果をラベル情報に使用

→ 省略可

（３）専門家によるラベリング（手動）

- ルール化されておらず、識別器でも信頼度の低い予測ができないものは専門家が判定

(1) (2) も訓練データに使う理由

- オンライン学習における忘却防止
- トラフィックパターンを抽象化した特徴量を定義すれば、ポート番号やスキャンパターン、プロトコルなどの簡単な変更があっても攻撃検知が可能になる。

DDoSバックスキャッタの判定ルール

ルール	判定結果
80/TCP かつ 制御フラグ= SYN-ACK, RST-ACK, or RST	DDoS
53/UDP かつ ヘッダに接続要求か応答を示す情報あり	DDoS
TCP SYNフラグ=1	非DDoS
BitTorrent Protocol	非DDoS

ダークネット分析事例（2）

頻出パターンマイニングによるホスト分析

JSPS 科研費基盤研究（Ｂ）
「サイバー攻撃のリアルタイム検知・分類・可視化のためのオンライン学習方式」

頻出パターンマイニング

相関ルール

$$X \Rightarrow Y$$

前提部　　結論部

Xを満たすときYの条件も頻繁に満たす.

評価指標

最低支持数:X ⇒ Yの条件を満たすルールの数

最低確信度:X ⇒ Yの条件が成立する確率

相関ルール

$$X \Rightarrow Y$$

前提部　　結論部

Xを満たすときYの条件も頻繁に満たす.

T1{パン, 牛乳}
T2{おにぎり, お茶}
T3{パン, ジュース}
T4{おにぎり, お菓子, お茶}
T5{カップ麺, お茶}
T6{おにぎり, チキン, お茶}
T7{おにぎり, チキン}

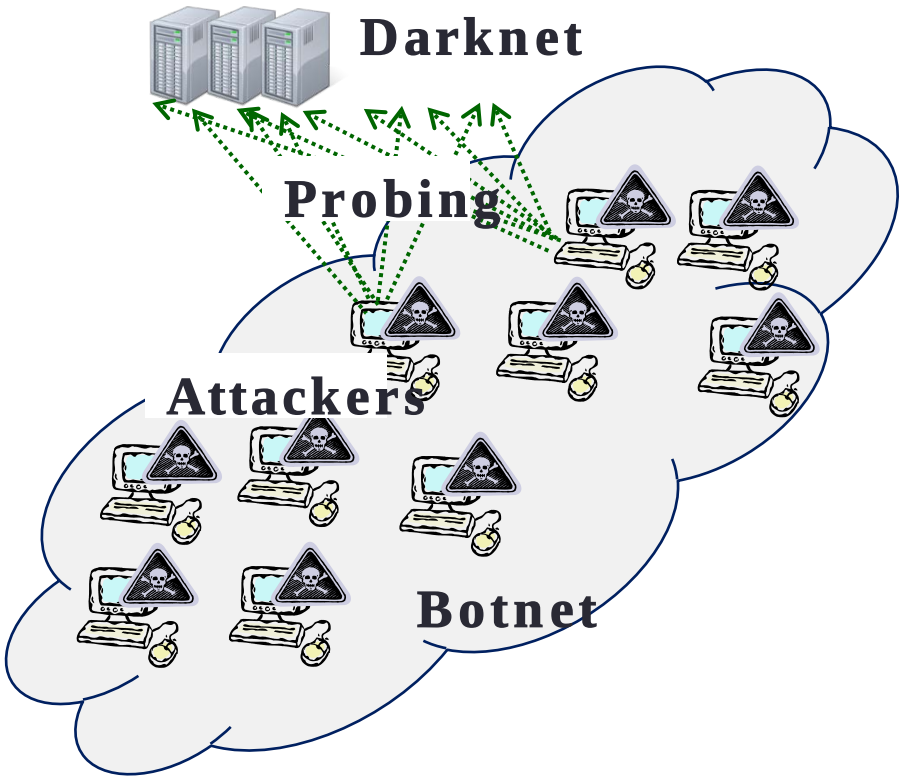
例：
{おにぎり} ⇒ {お茶}

おにぎりを買う人は、
高い頻度でお茶を買う

頻出パターンマイニングによるダークネット観測

- TCP SYN パケット
- 2016年7月1日から2016年9月15日
- NICT /16 ダークネットセンサー

観測パケット数:
1,840,973,403
ユニークホスト数:
17,928,006



相関ルール学習

Miraiソースコード
公開周辺

2016年の7月～9月の3ヶ月間

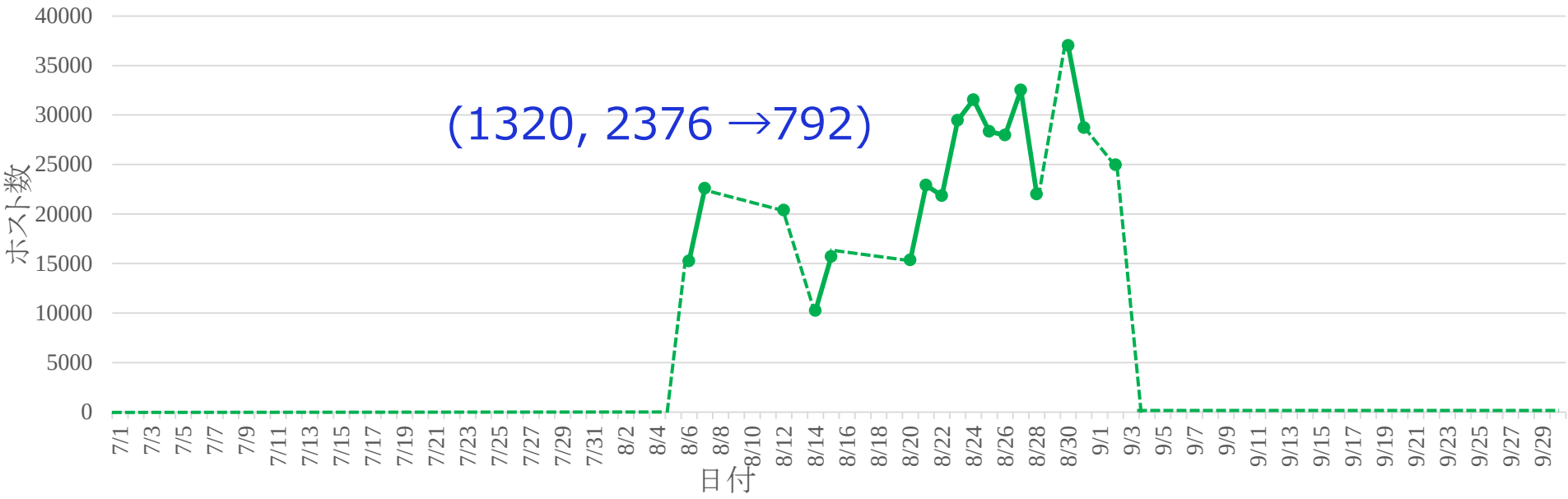
相関ルール学習の条件：
最低ホスト数: 1000host
最低確信度: 90%

▼

TCP ウィンドウサイズ
パケット優先度 (ToS)
送信先ポート番号

で特徴的な相関ルールが出現.

TCP ウィンドウサイズ

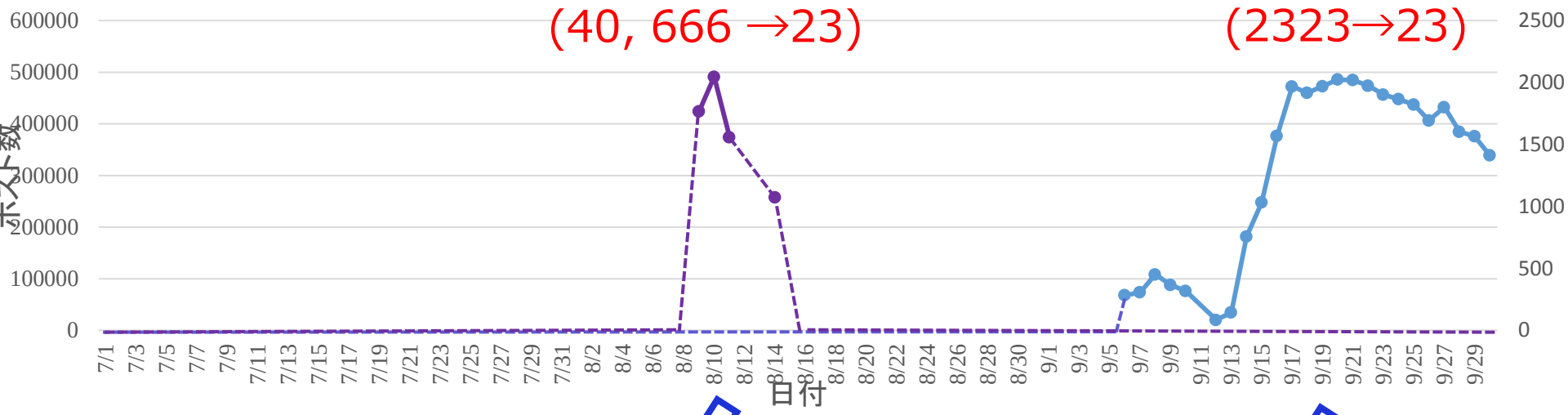


出現期間:8/6～9/3

最大ホスト数:37029(8/30)

平均 約2万host

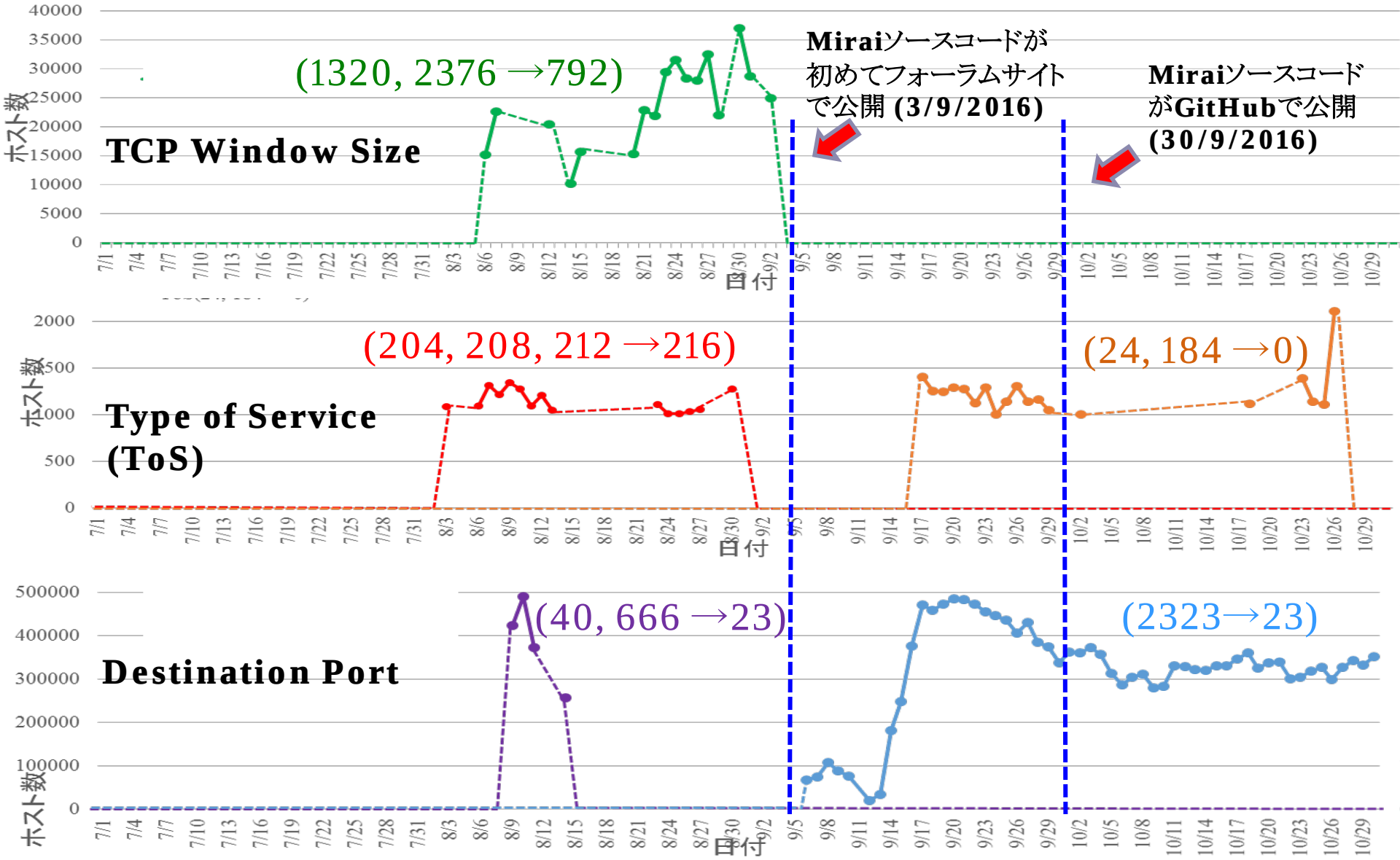
送信先ポート番号



出現期間:8/10～8/15
最大ホスト数:2046(8/11)

出現期間:9/6～
最大ホスト数:485572(9/20)

IoT マルウエア*Mirai* の特徴変遷



相関ルールと Mirai の関係

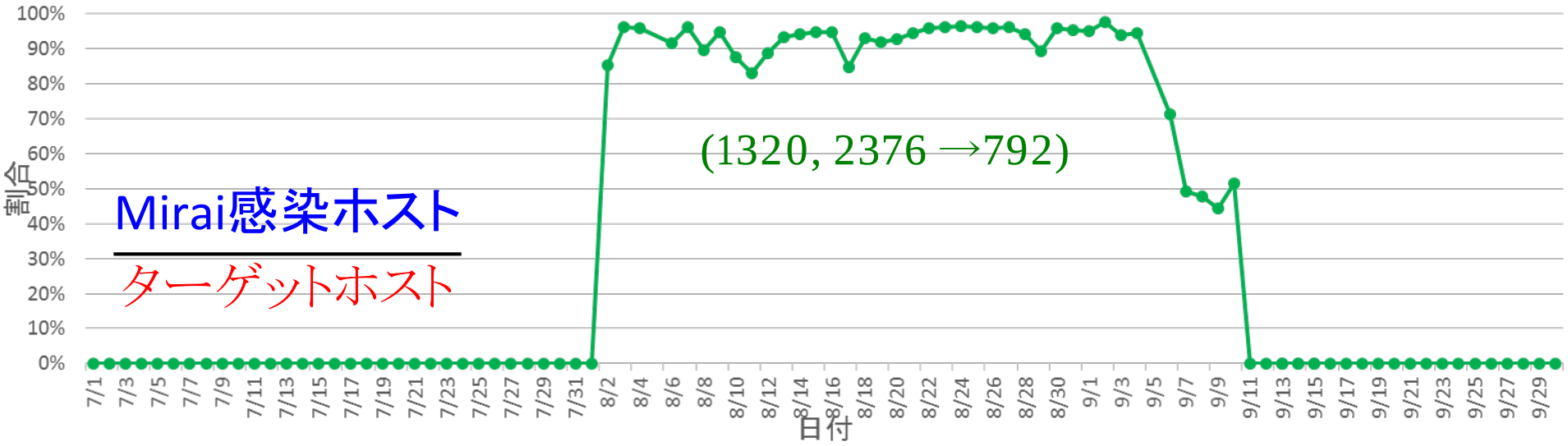
相関ルールに当てはまるプローブ活動を行った
ターゲットホストについて

1. シーケンス番号＝宛先IPアドレス
2. 宛先ポート番号＝23
3. 送信ポート番号＞1024

上記の条件を全て満たすパケットが
90%を超えたMirai感染ホストと定義

ターゲットホストのMirai感染を調べる.

TCP ウィンドウサイズ

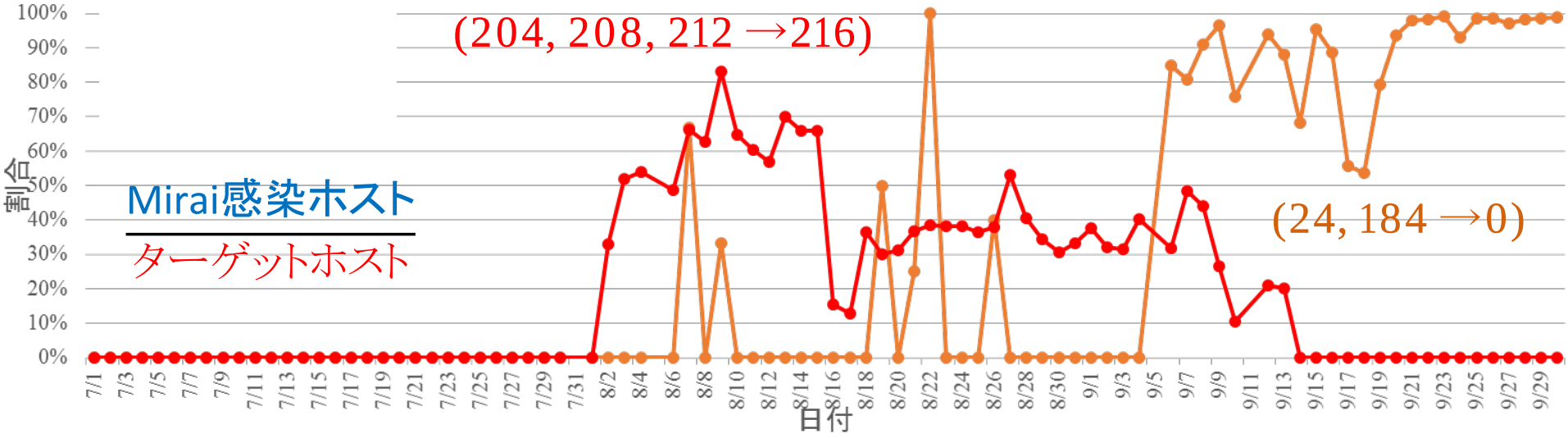


(792, 1320, 2376)

出現期間:8/2～9/10
90%近い一致を示した

ターゲットホスト: 792,1320,2376の3つ全てのTCP window sizeを用いるホスト

パケットの優先度(ToS)



(204, 208, 212, 216)

出現期間:8/2～9/14
一致は30%程度

(0, 24, 184)

出現期間:8/7～
相関ルールが出現した
9/17以降は100%近い一致

送信先ポート番号

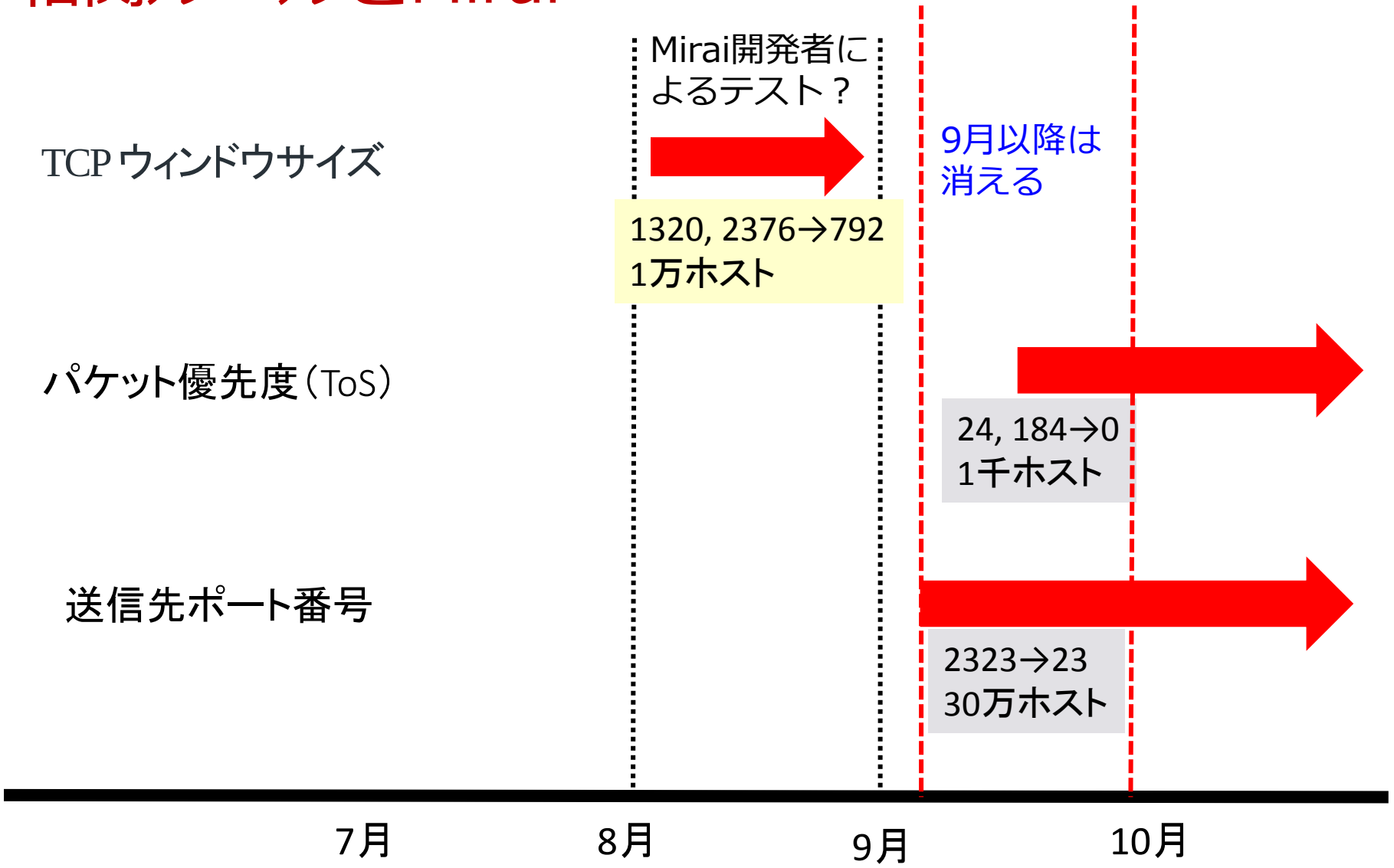
(40, 666, 23)

Miraiとの関連なし.

(2323, 23)

出現期間:9/5～
90%近い一致を示した

相関ルールとMirai



ダークウェッブ アングラマーケット監視

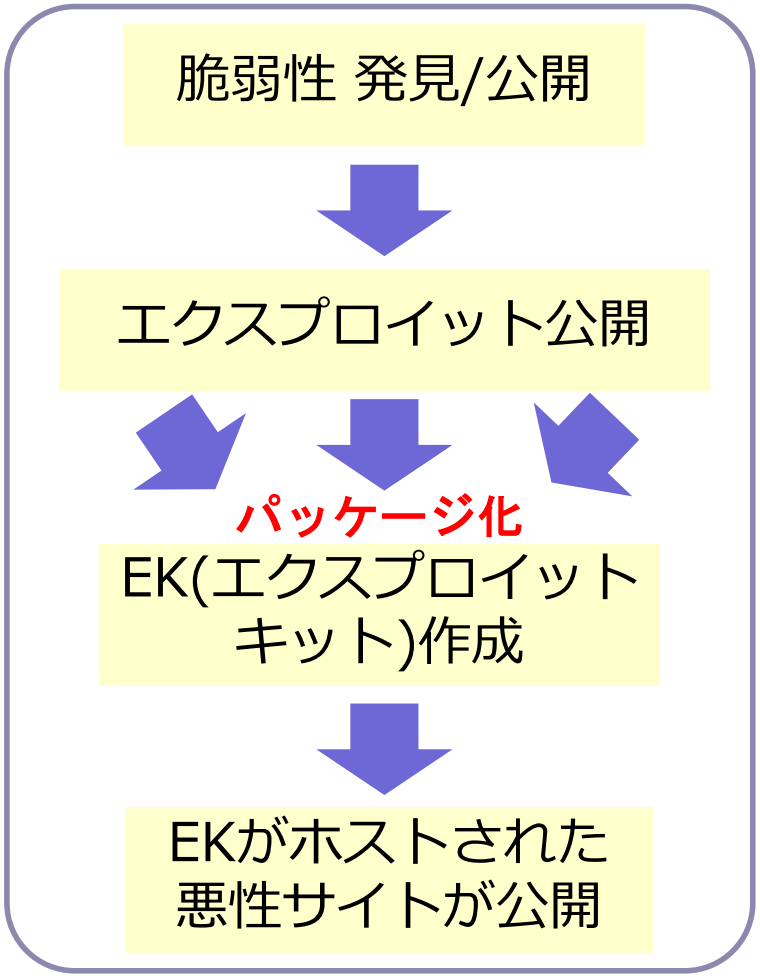
AI Web Crawlerの開発と商品分析

NICT委託研究
「Web媒介型攻撃対策技術の実用化に向けた研究開発」

Web媒介型攻撃の網羅的な観測・分析に基づくユーザ環境のセキュリティ高度化 (WarpDrive Project)

サイバー攻撃発生ステップ

- 脆弱性を用いてマルウェアが配布されるまでに複数ステップが存在する。
- エクスプロイットやEK，マルウエアの取引は一部ダークウェブで行われる。



必ずしも順序通りでなく同時発生も

脆弱性: 設計のミス等によって生じるセキュリティ上の欠陥

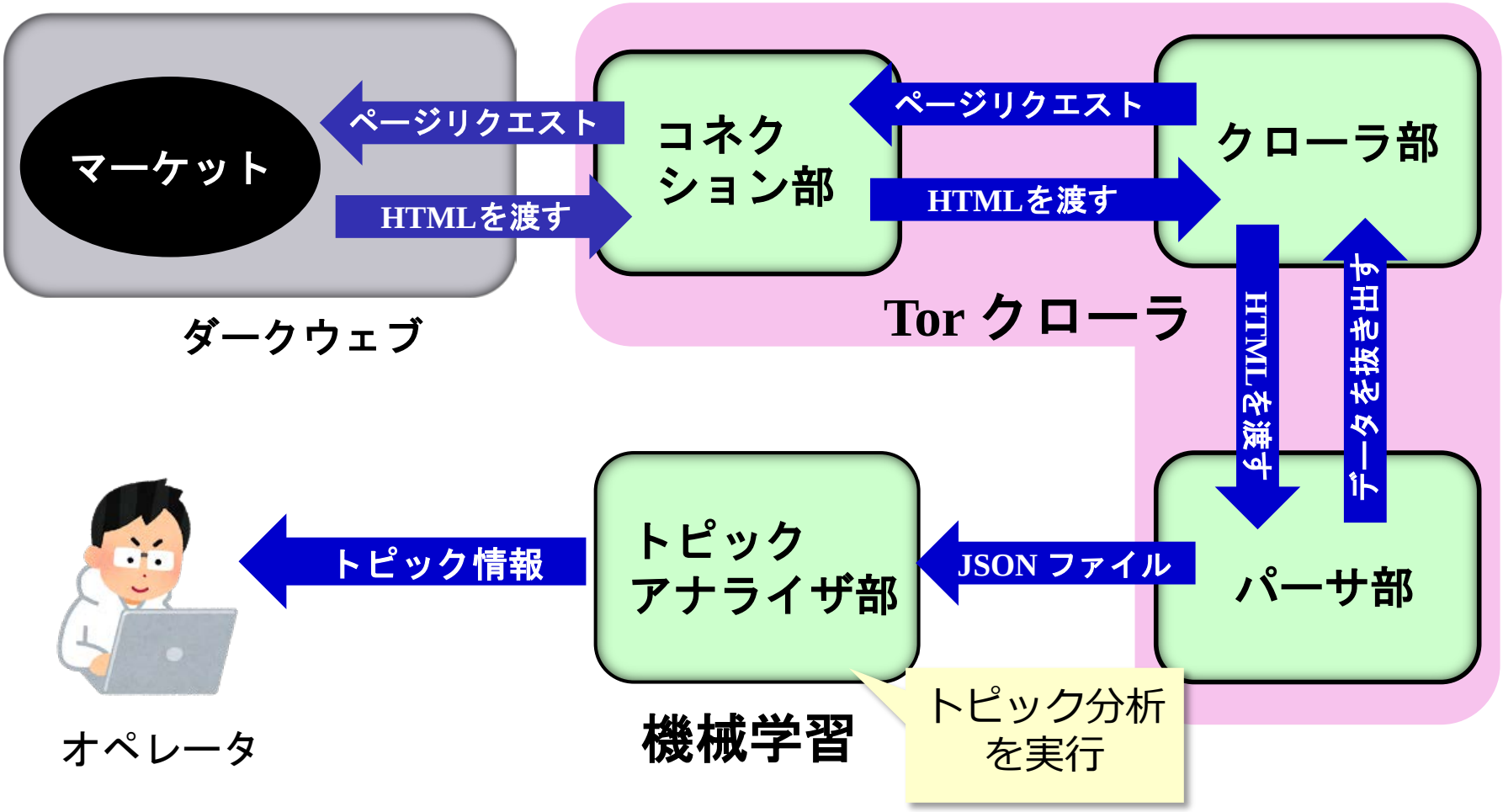
エクスプロイット: 脆弱性を悪用し、任意の不正操作が行えるソース

EK: 対象の脆弱性を特定し、その脆弱性を利用して不正操作を行うプログラム

エクスプロイット，EK，マルウェアなどサイバー攻撃関連商品の取引は，一部，ダークウェブで観測可能

Darkweb AI Crawler

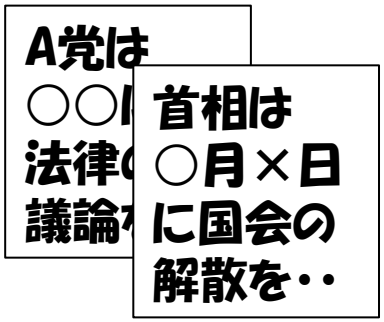
- 前頁の目的を実現するためのシステムについて説明する。
- ダークウェブ内のマーケットからクローラにより情報を収集
- 収集した情報を機械学習により分析。



トピックモデル

- アナライザ部で使用する**トピックモデル**について解説。
- 文書集合の中からトピックと各文書が属するトピックが推測できる。

トピックとは



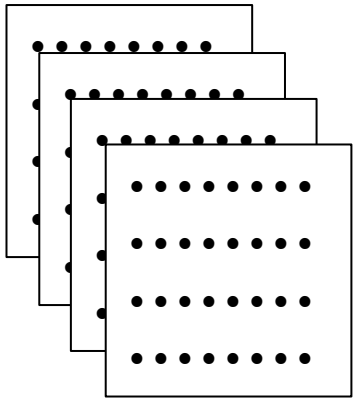
語彙集合

首相　総理大臣
国会　法律
野党　法案…

上のように政治系の文書集合で
あれば、首相、国会といった語彙が
同時に存在することが多い。

このような共起関係のある
語彙の集合をトピックと呼ぶ。

入出力



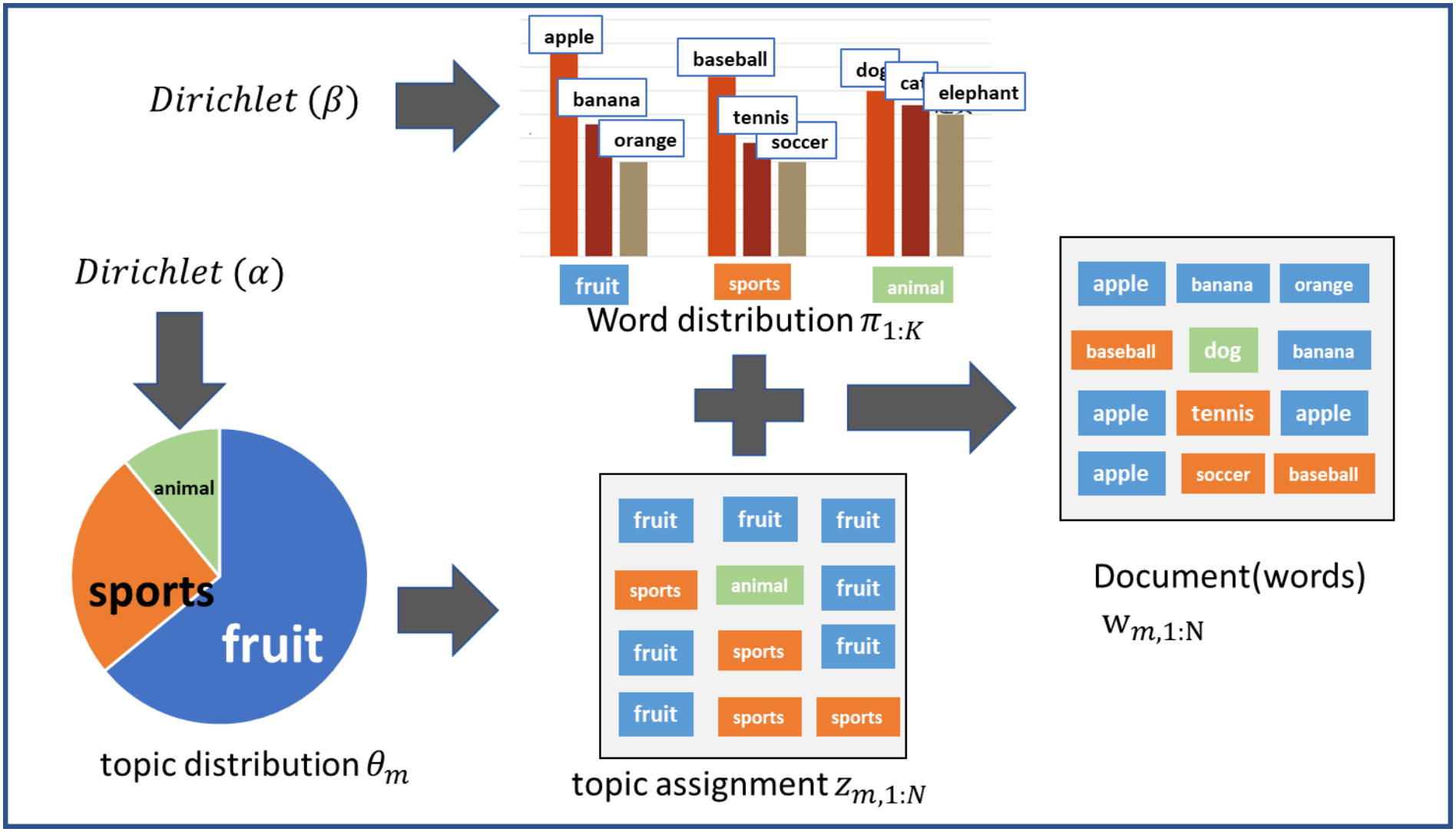
文書集合
（ラベルなし）

Topic1:国会　首相
Topic2:電車　駅
Topic3:スマホ　通信

トピック

各文書の
トピックの割合

Latent Dirichlet Allocation (LDA)



実マーケットへの適用(1/3)

AlphaBay

AlphaBayは2014年に設立されたマーケットであり、閉鎖される直前には40万人もの利用者が存在し、トップマーケットとして高い影響力を持っていた。

しかしながら2017年7月5日に運営者が逮捕され現在は閉鎖されている。

Hansaマーケット

AlphaBayの閉鎖後、多くのユーザが流入したマーケットがHansaマーケットである。

その後トップマーケットとなったが、AlphaBay閉鎖から僅か2週間後に同様に閉鎖することとなった。
(この一連の流れはoperation bayonetと呼ばれる作戦によるものであった。)

実マーケットへの適用(2/3)

トピック#	代表単語	代表商品
トピック1	carding software money free bitcoin need btc guide bank paypal	BTC Stealer 4.3 and Mass Address Generator 1.2+, Professional Carding Software
トピック2	file files advanced hacking computer use ransomware network victim windows	Advanced System Protector, KilerRat v10.0.0 Full
トピック3	account login paypal hacking bitcoin vpn cashout bank cc btc	VPN Account For Life ironsocket.com, BITCOIN STEALER
トピック4	password com hack recovery phone www forensic passwords accounts Software	GET ACCESS TO ANY PHONE BYPASS PASSCODES RETRIEVE ALL THE DATA ON THE PHONE, Professional Phone HACK software
トピック5	http php id exploit time make money tools file use	Voice changer(Android,Windows), SILENT EXPLOIT SETUP SERVICE

実マーケットへの適用(3/3)

トピック#	代表単語	代表商品
トピック1	carding money guide method make paypal cashout free tutorial cc	All in One-Carding/Money Making/Hacking, Overnight Money Making Machine!
トピック2	file files advanced hacking computer use ransomware network victim windows	Bugtroid - Android MEGA TOOLS PACK, DIAMOND RAT
トピック3	account login paypal hacking bitcoin vpn cashout bank cc btc	Get Access To Any Phone Bypass Passcode, ACCOUNT HACKING PROGRAM
トピック4	password com hack recovery phone www forensic passwords accounts Software	True Online Anonymity Kit, TRUE ONLINE ANONYMITY KIT
トピック5	http php id exploit time make money tools file use	TRAFFIC ENCRASER INCLUDE ATOMIC EMAIL SENDER AND OTHER TOOLS FOR Facebook,Backlink,Keyw, DK Brute - Bruteforce RDP, SSH, SMB, pop3, pop3s, VNC, FTP

プライバシー保護データマイニング

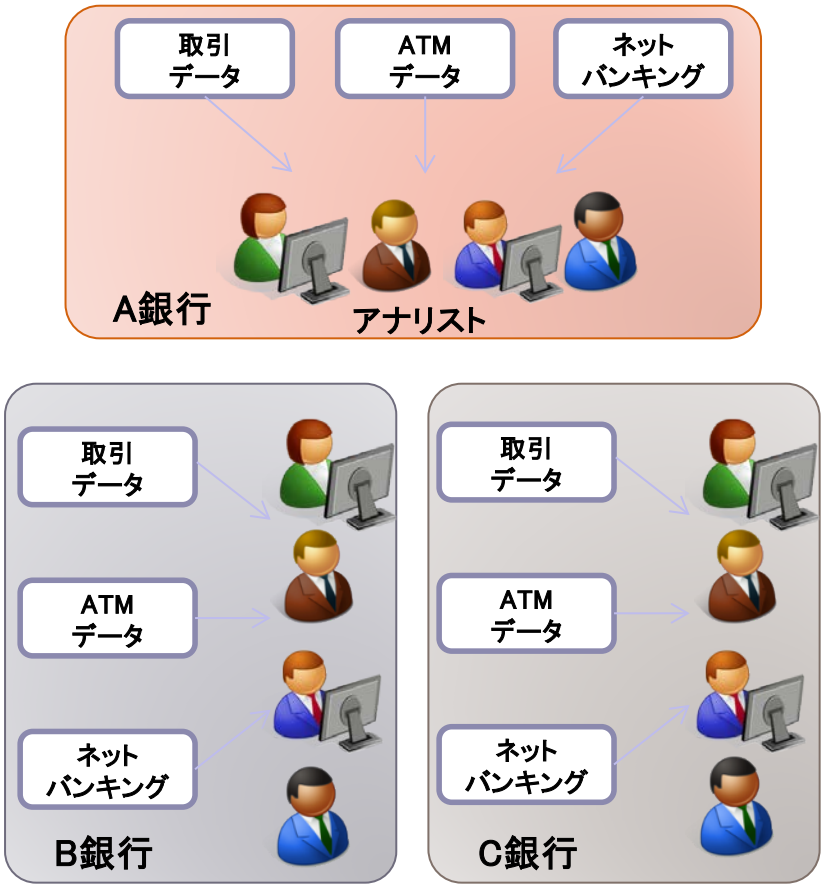
暗号化データのまま学習できるニューラルネット

CREST
「イノベーション創発に資する人工知能基盤技術の創出と統合化」（研究総括：栄藤　稔）

複数組織データ利活用を促進するプライバシー保護データマイニング （代表：盛合志帆）

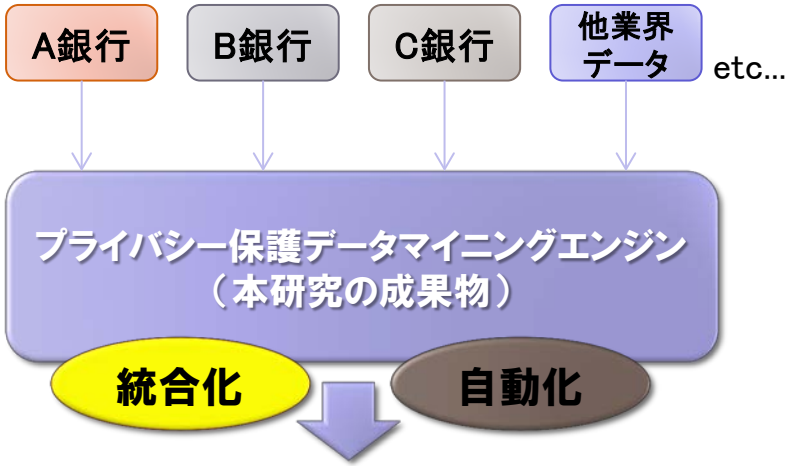
現状とめざす構想

現状



個々の銀行内で分析

めざす構想



不正送金の検知
適正利率の導出

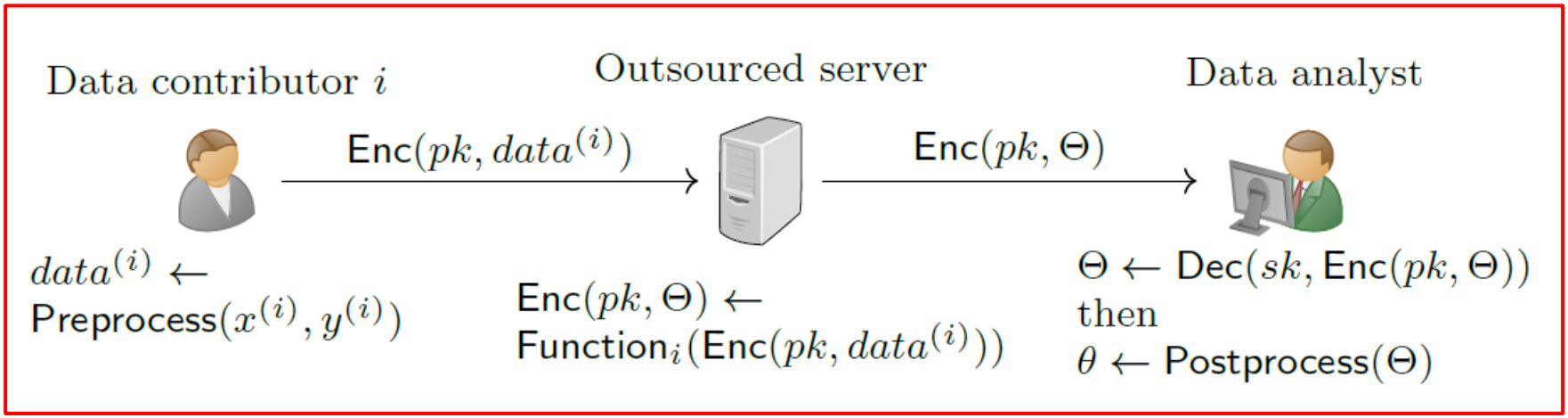
- 調査コスト削減
- 調査属人化の回避
- 調査精度の向上
  - ➢ 今まで見つからなかった検知が可能に！

研究領域「イノベーション創発に資する人工知能基盤技術の創出と統合化」
研究課題名「複数組織データ利活用を促進するプライバシー保護データマイニング」（盛合志帆）

PPDMスキーム

加法準同型暗号

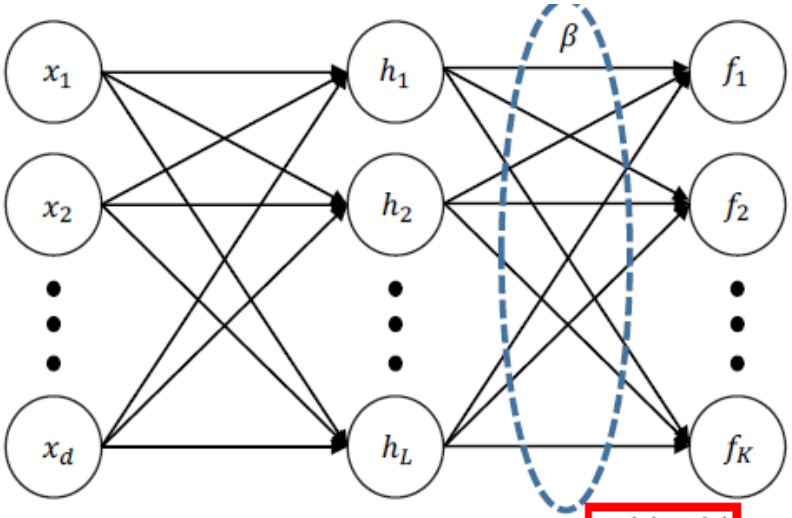
$$m_1 \cdot m_2 = \mathsf{Dec}(sk, \mathsf{Enc}(pk, m_1) \odot \mathsf{Enc}(pk, m_2)),$$



Privacy Preserving Extreme Learning Machine (PP-ELM)

Shohei Kuri, Takuya Hayashi, Toshiaki Omori, Seiichi Ozawa, Yoshinori Aono, Le Trieu Phong, Lihua Wang, Shiho Moriai, "Privacy Preserving Extreme Learning Machine Using Additively Homomorphic Encryption," Proc. of The 2017 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2017), pp. 1350-1357, 2017.

プライバシー保護ELMモデル



$$\beta = \left(\frac{\mathrm{I}}{\lambda} + H^T H\right)^{-1} H^T Y \qquad (N \gg L)$$

(N : *The number of data records*,
L : *The number of hidden nodes*)

$$H^T H = \begin{bmatrix} \sum_{i=1}^{N} h_1^{(i)} h_1^{(i)} & \cdots & \sum_{i=1}^{N} h_1^{(i)} h_L^{(i)} \\ \vdots & \cdots & \vdots \\ \sum_{i=1}^{N} h_L^{(i)} h_1^{(i)} & \cdots & \sum_{i=1}^{N} h_L^{(i)} h_L^{(i)} \end{bmatrix}$$

$$H^T Y = \begin{bmatrix} \sum_{i=1}^{N} h_1^{(i)} y_1^{(i)} & \cdots & \sum_{i=1}^{N} h_1^{(i)} y_K^{(i)} \\ \vdots & \cdots & \vdots \\ \sum_{i=1}^{N} h_L^{(i)} y_1^{(i)} & \cdots & \sum_{i=1}^{N} h_L^{(i)} y_K^{(i)} \end{bmatrix}$$

性能評価

評価データセット ： ４つのベンチマークデータ
Glass, Digits, Satellite, Shuttle
暗号方式： SPHERE （LWEベース 準同型暗号）

Datasets	PP-ELM L= 300	PP-Logistic ovr	Logistic ovr
Glass	0.684 +/- 0.089	0.596 +/- 0.099	0.604 +/- 0.070
Digits	0.965 +/- 0.021	0.889 +/- 0.037	0.925 +/- 0.027
Sattelite	0.875 +/- 0.007	0.758 +/- 0.019	0.827 +/- 0.018
Shuttle	0.997 +/- 0.001	0.873 +/- 0.002	0.933 +/- 0.002

（L：#Hidden Units）

+0.04～0.12

さいごに

セキュリティ×AIへの期待（課題）

1. 知的で高速、高精度な攻撃監視（異常検知・分類・予測・可視化）と自律学習機能（追加学習、オンライン特徴抽出、自動データ収集、自動ラベリングなど）
2. 専門家とコラボレーションしながら高度なサイバー攻撃を検知・防御できる知的プラットフォーム
3. 膨大な量のオープンデータからサイバー攻撃に関連した情報の自動モニタリング
4. 騙されないAI・機械学習
5. プライバシー漏洩のないデータからの知識獲得