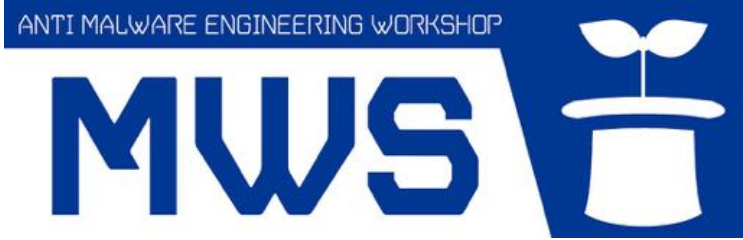


MWS 2021 ポストミーティング

MWSをやってて博士とったらNISTで研究までやってきた話

畑田 充弘

NTT Com-SIRT

自己紹介

- NIST Guest Researcher (2019.3 – 2021.9)
  - Computer Security Division, Information Technology Laboratory
  - ボットネットによる攻撃の予兆検知(NIST TN 2111)、NVD Analysis
- NTTコム (2003.4 – 2018.5、2021.12 -)、日本電信電話 (2018.6 – 2021.11)
  - サイバーセキュリティの研究開発、人材育成、NTT Com-SIRT立ち上げ、Team V、国際イベントでのThreat Hunting、team enu
- 博士(工学) (2018.2)、早稲田大学招聘研究員 (2018.12 – 現在)
- ICT-ISAC Japan国内外連携WG/情報共有WGメンバ、CSDE Steering Committee
- IEICE ICSS研究会専門委員、IPSJ CSEC研究会専門委員、MWS組織委員/実行委員
- NTT Group Certified Security Principal、CISSP、RISS

アジェンダ

1. MWSのはじまり
2. 博士課程へ
3. NISTへ
4. ふりかえり

1. MWSのはじまり

きっかけ

仕組み作るか？

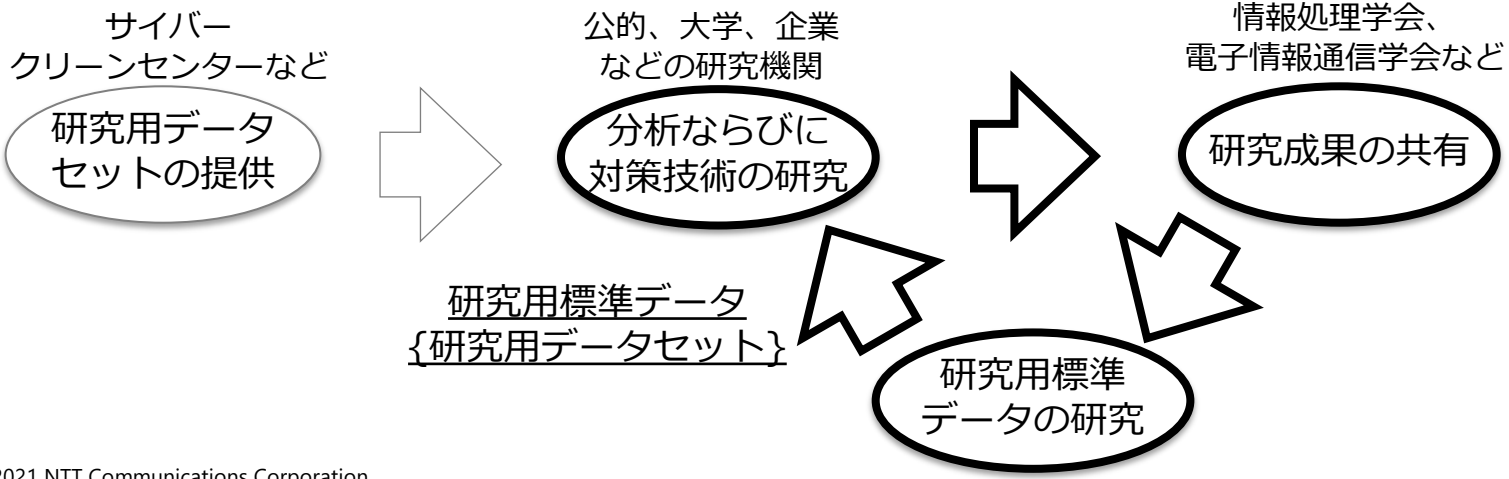
ANTI MALWARE ENGINEERING WORKSHOP
MWS

http://www.iwsec.org/mws/

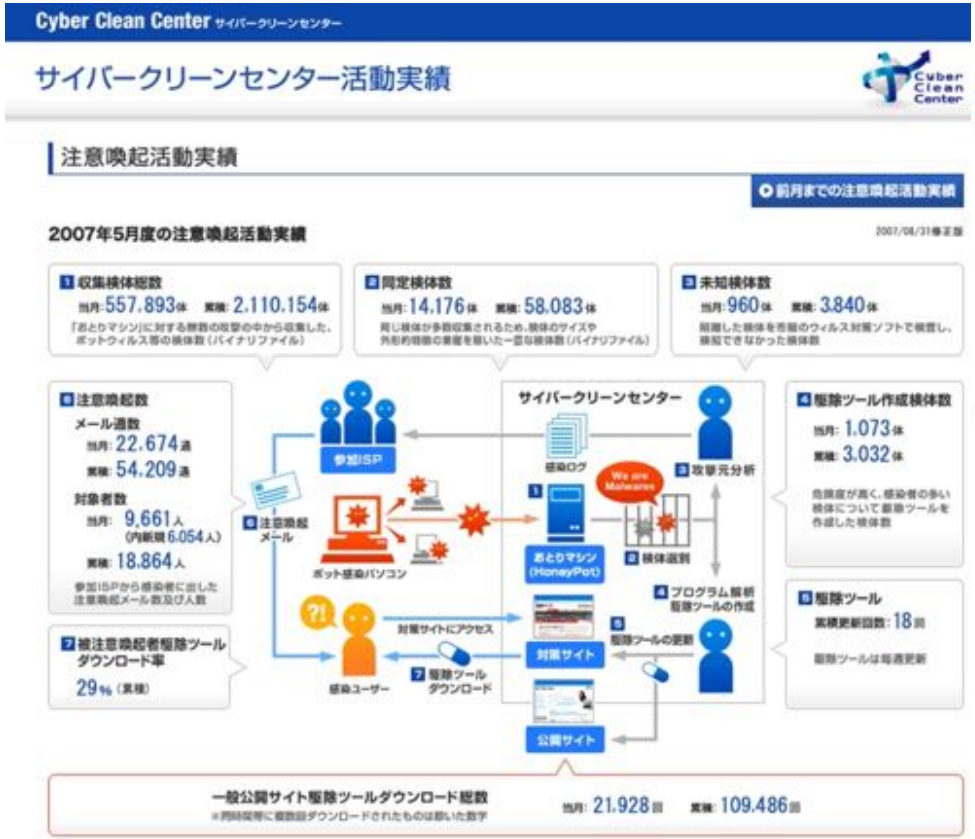
NTT Communications

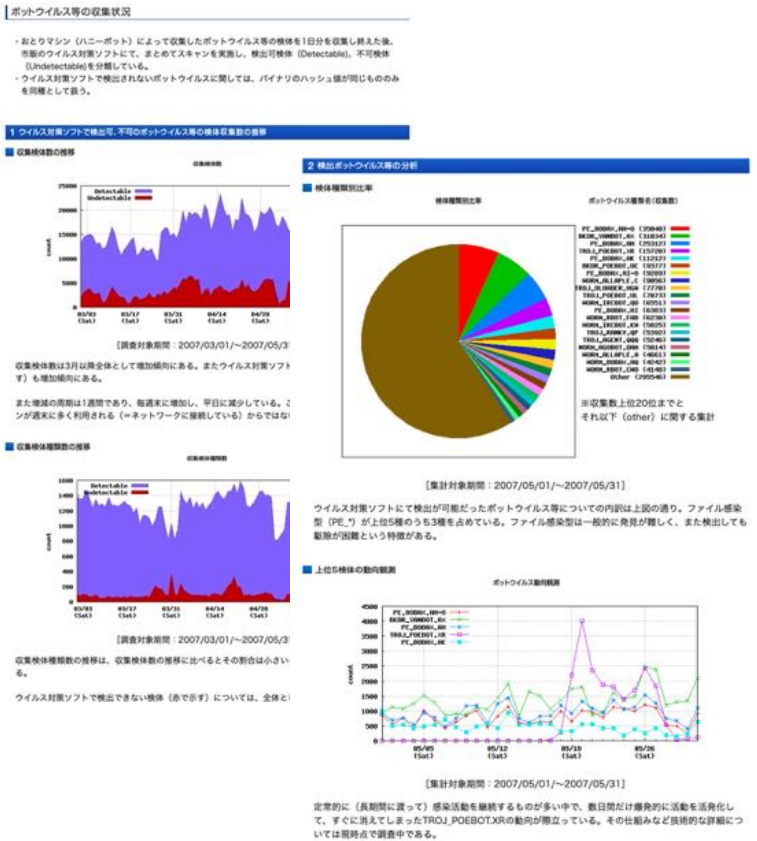
マルウェア対策研究人材育成ワークショップ（2008～）

- マルウェアの機能や運用の高度化、インシデントの把握が困難
- 先端研究者に限らず、CSIRT等の実務者もマルウェアに関する知識が必要
- 教材となる共通のデータセットがない、研究用データの収集が困難



Cyber Clean Center

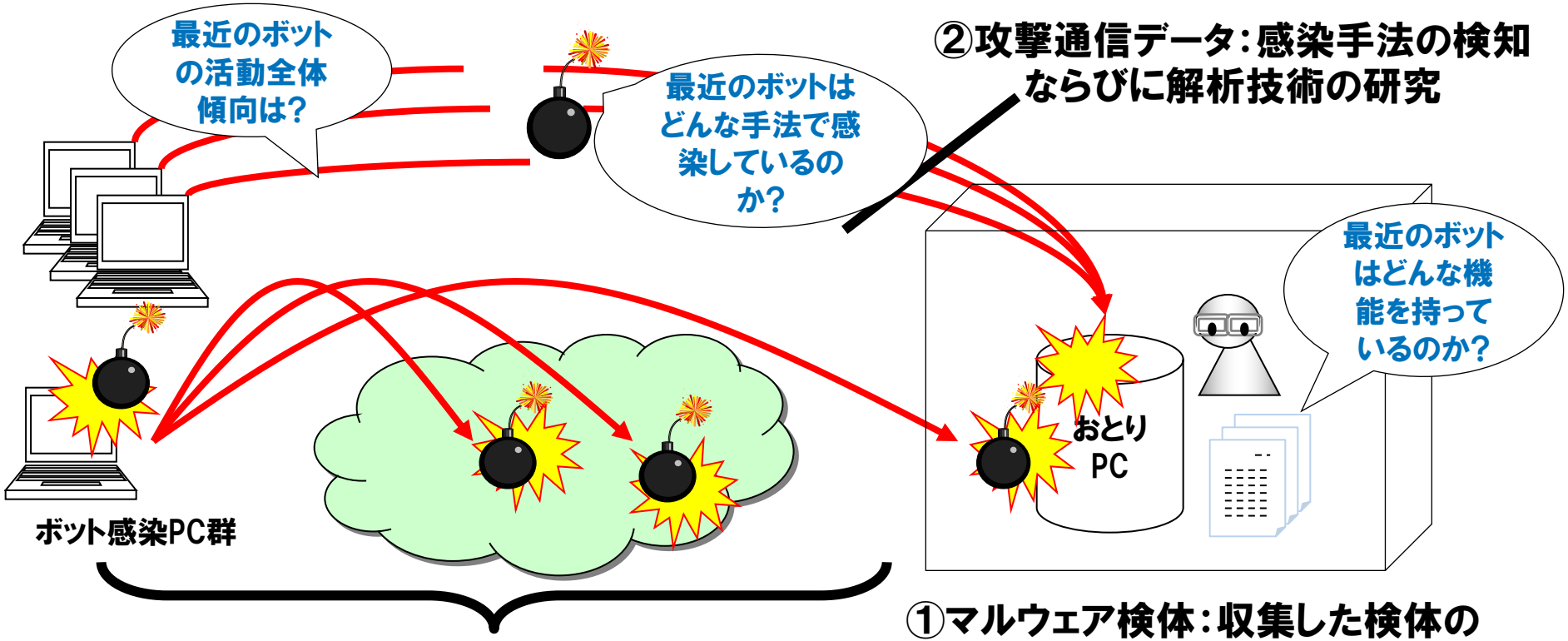




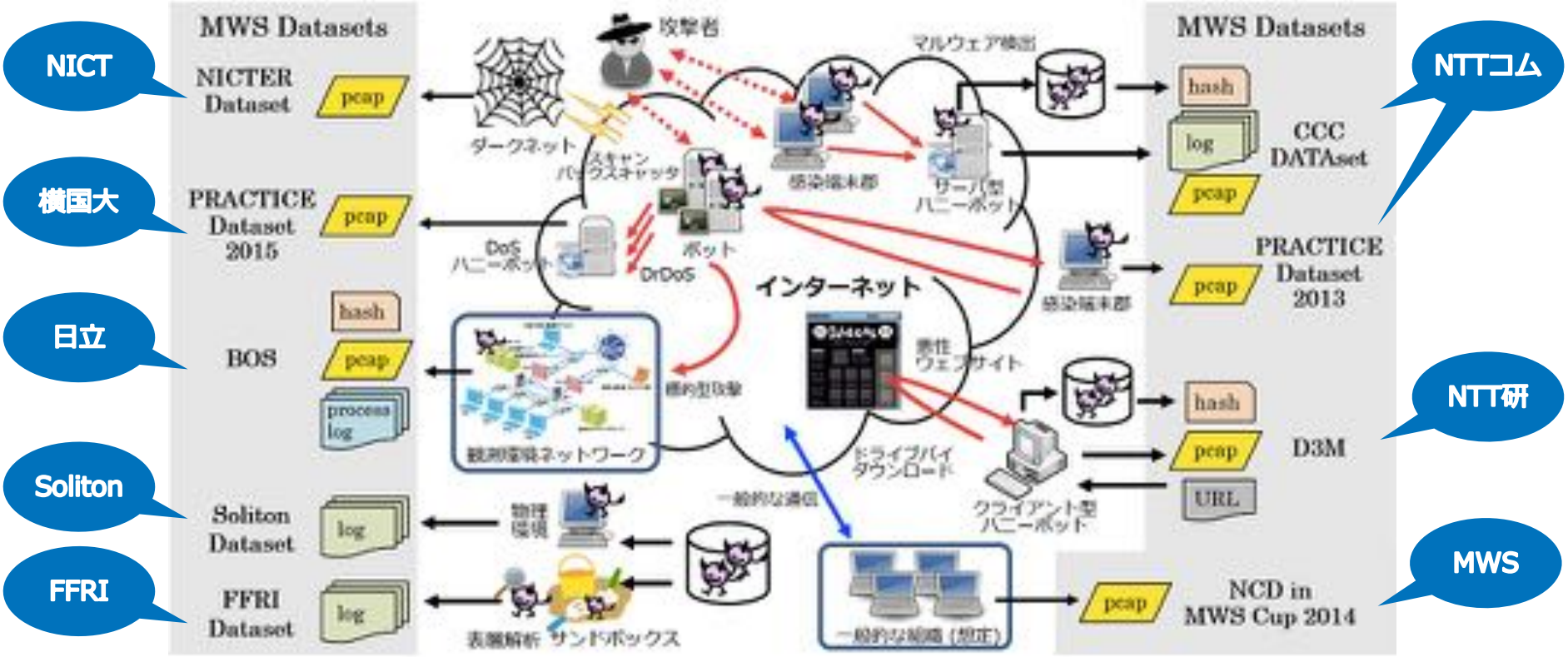
Copyright 2021 NTT Communications Corporation

https://www.telecom-isac.jp/ccc/report/200705/0705monthly.html

CCC DATAset



MWS Datasets 2021



MWS Cup 2009 (第1回)

■課題1： 攻撃通信データを探し出せ。

競技用CD-ROMに収められている (W) が5個，(B) と (B') で5個の合計
apファイル）のうち、(B) と (B') を探し出し、その番号を解答欄に記入せよ。
拡張子を除いた数字を用いよ。
（注：）
(W) ＝ マルウェアに感染していないPCの通信データ。攻撃を受けても感

■課題2： マルウェア名を言え。

課題1で探し出した各通信データファイルにおいて、「ホストが感染して
から選び、それぞれの解答欄に記入せよ。尚、各解答欄には一つの
(m0) BKDR_MYBOT.AH (m1) BKDR_RBOT.ASA (m2) PE_BO

■課題3： 今後の通信パターンを予測せよ

課題1で探し出した通信データファイルの、「今後の通信パターン
ものを一つ選び、解答欄に記入せよ。（注：）一つの選択肢が、複
(a0) 何もしない (a1) ICMPスキャン (a2) 位置情報データベース
(a3) C&C接続 (a4) C&C接続、特定ホストへのポートスキャン
(a5) Microsoft Update を実行 (a6) 連鎖感染、外部へシステム
(a7) 連鎖感染、外部 Webサイトへ DoS攻撃
(a8) 連鎖感染、C&C接続、外部へ DNSクエリ（MXレコード）
(a9) 連鎖感染、TCP(135)スキャン、C&C接続

MWS Cup 2009 総合優勝：東京電機大学三原チーム

MWS Cup 2009 技術賞：東海大学人海戦術チーム

MWS Cup 2009 芸術賞：Internet Security Centerチーム

MWS Cup 2021

MWS Cup

- マルウェアによ
- MWS Datasets
- 課題構成

種別	内容
事前課題	ツール
課題 1	Exploit
課題 2	マルウ
課題 3	マルウ
課題 4	フォレ

cup2021_hackathon 事前課題（ハッカソン）の成果展示＆投票所です。

123

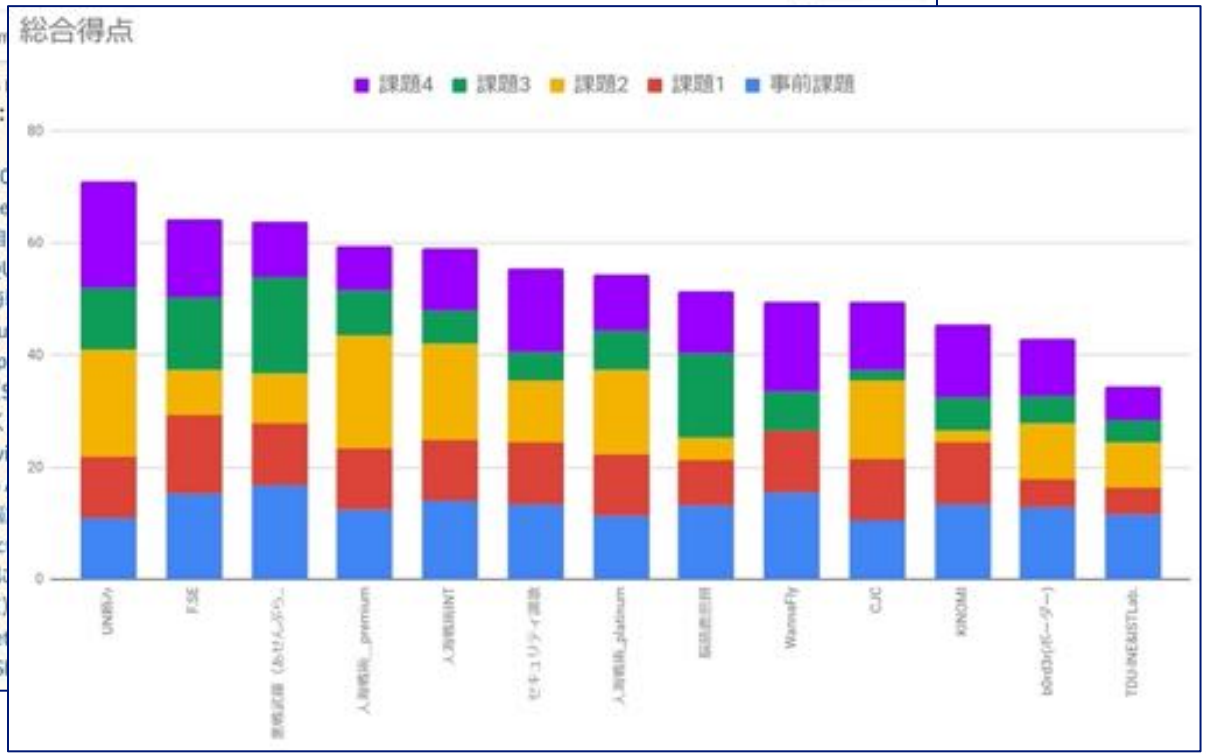
2 Pinned + Add a bookm

松木_NFLabs 10:35

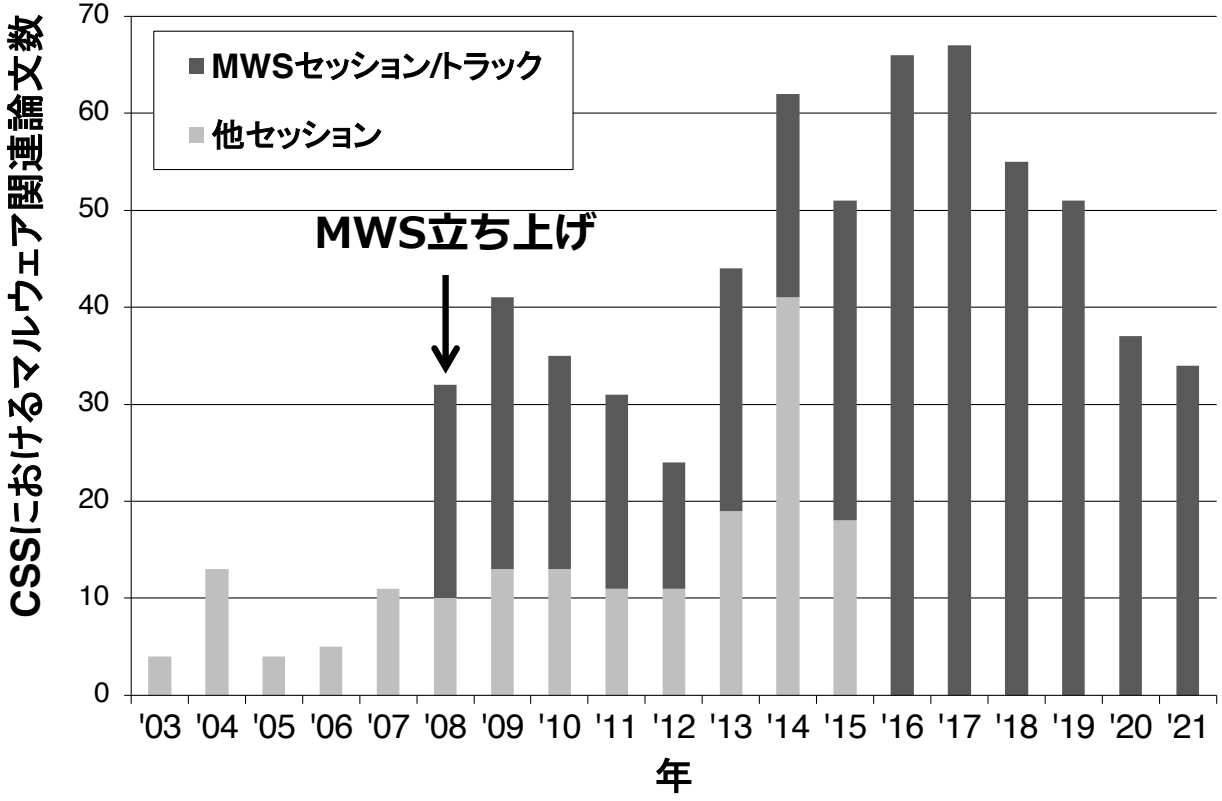
事前課題成果一覧：

い。

1. FFRI Dataset 20
2. Cyber FrontLine
3. 不審ファイル自
4. JOKAKUSHI QU
5. sedu-vm（人海
6. MISParser（Blu
7. pwndbg-percep
8. GPCChecker（5
9. ころなけんさく
10. MAAMBOT-Tw
11. Brilliant Search
12. CtFinder（卍脳
13. Phishing Protec
14. n-gram+GloVe
15. REN-A.I.（F.SE
16. MAID -Mail De
17. Avocado Safe S



CSSにおけるマルウェア対策系論文数



2. 博士課程へ

きっかけ

「お前、(試験的に)博士行くか？」

学位論文: 博士（工学）

- Towards Efficient Incident Handling Triage: Automated Threats Classification and Data-centric Talent Development
  - インシデント対応の効率的なトリアージに向けた脅威分類の自動化とデータ主導の人材育成
  - 脅威分類の自動化では、マルウェアが行う通信のモデル化による新種マルウェアの分類、ならびにアクセスするドメイン群の類似度に着目したPUAの分類
  - データ主導の人材育成では，実用的なデータセットの整備により，マルウェア対策に関わる人材育成

博士の価値

- やりぬく力（問題設定、調査、計画、実装、評価、プレゼン）
- グローバルな信用（専門家による査読を複数本クリア）

モチベーション

- これまでの業務経験をもとに、がっつり研究したかった
  - 現実に感じてきた問題
    - 真に未知のマルウェアをどう判定するか
    - グレーなものにどう優先順位をつけるか
  - 新たなロールモデルの開拓

生活

- 休止したこと
  - 都リーグ社会人サッカー
  - 読書
  - CTF（SECON/DEFCONは例外）
- 始めたこと
  - 水泳（息子のスクール中）
- 変化
  - 体重+2kg

これから博士を目指す人へ

- なぜ学位をとりたいか、その後どうするかを明確に
- 5～10年目くらいがよさげ
  - 実務を経験した上での研究への欲求
- 実行可能な計画と自信が必要
  - 博士取得要件の把握、国際会議・ジャーナルの査読プロセス
  - 入学前の研究業績、先生との研究計画相談
- 履修科目・必要単位・研究室行事を加味した行き先選択
- 1-2年目は業務2:研究3、3年目は業務3:研究2

3. NISTへ

きっかけ

「そんなん、想像できるやろ？」

NISTのミッション

National Institute of Standards and Technology

- 米国国立標準技術研究所、商務省配下の研究機関
- 1901年にNational Bureau of Standardsとして設立

経済安全保障と生活品質向上に資する方法で、計測、標準、技術を進歩させることによって、米国の技術革新と産業競争力を強化

https://youtu.be/2j9BGVKbzS4

概観

本部: メリーランド州
ゲイザースバーグ市

職員: 3,400+

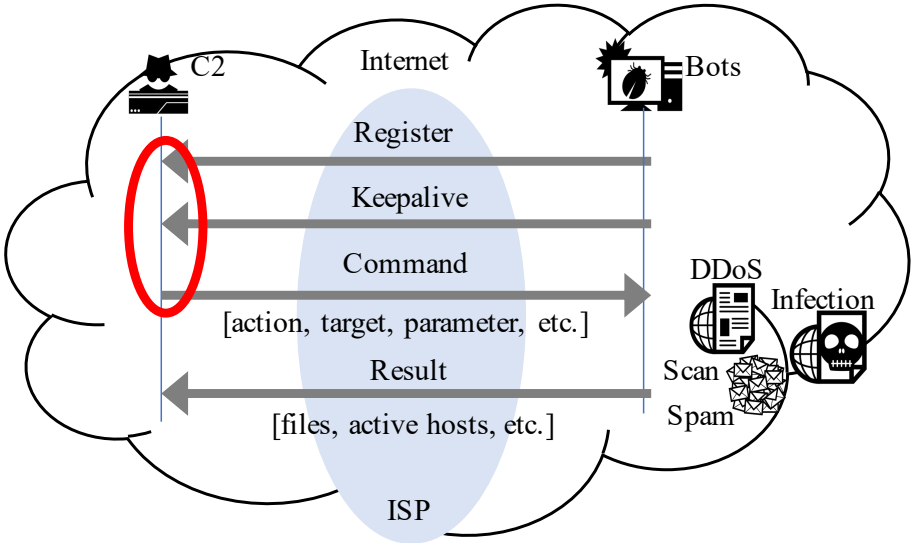
協力: 3,500+

フロー分析でボットネット攻撃予兆検知

NIST Technical Note 2111

ISPのフローデータを利用して、既知のC2を出入りするトラフィックの変化から、スパムメール増加の予兆を77%の精度で検知

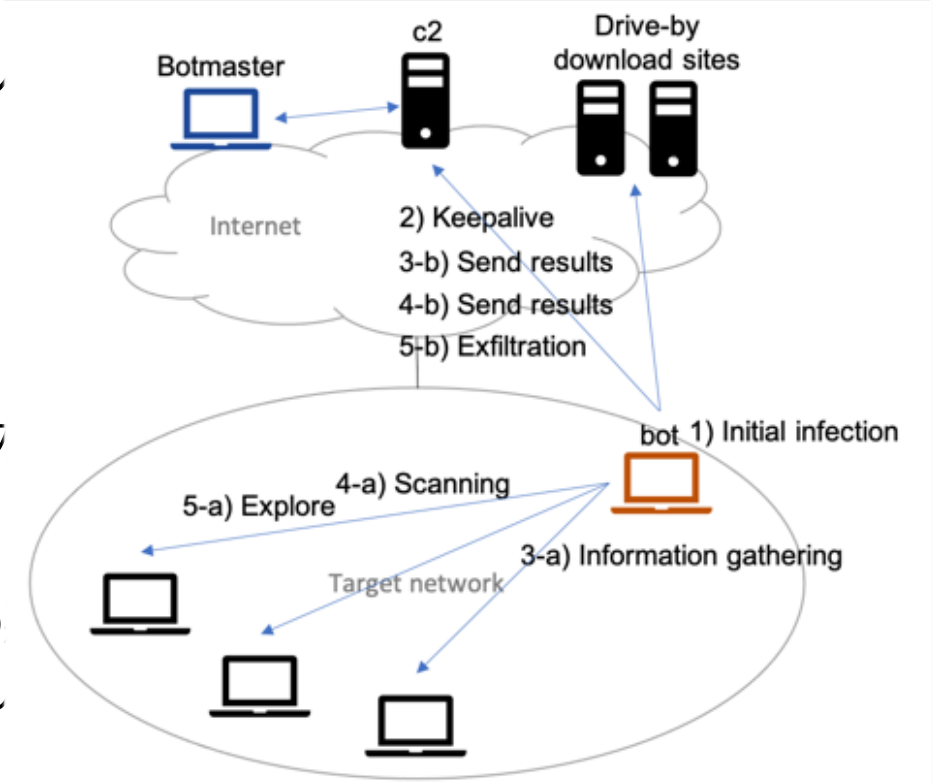
- アノテーション自動化
- LSTM、GPUクラスタ (Enki)
- Explainable (SHAP)



アノマリ検知の評価フレームワーク

NIST Technical Note 2142

- 悪性トラフィックの生成と各種
  ラメータ
- 良性トラフィックと悪性トラ
  フィックの混合
- アノマリ検知システムの評価指



組織内NWでのボット感染から情報漏洩の
性トラフィック生成シミュレータも開発し
上記TNの拡張で論文投稿中

セキュリティ自動化のための標準

資産管理

- Software identification (SWID) tag
- Common Platform Enumeration (**CPE**)

脆弱性管理

- National Vulnerability Database (**NVD**)
- Common Vulnerability Scoring System (CVSS)
- Vulntology

設定管理

- Security Content Automation Protocol (SCAP)
- National Checklist Program (NCP)
- Common Configuration Enumeration (**CCE**)

計画・評価

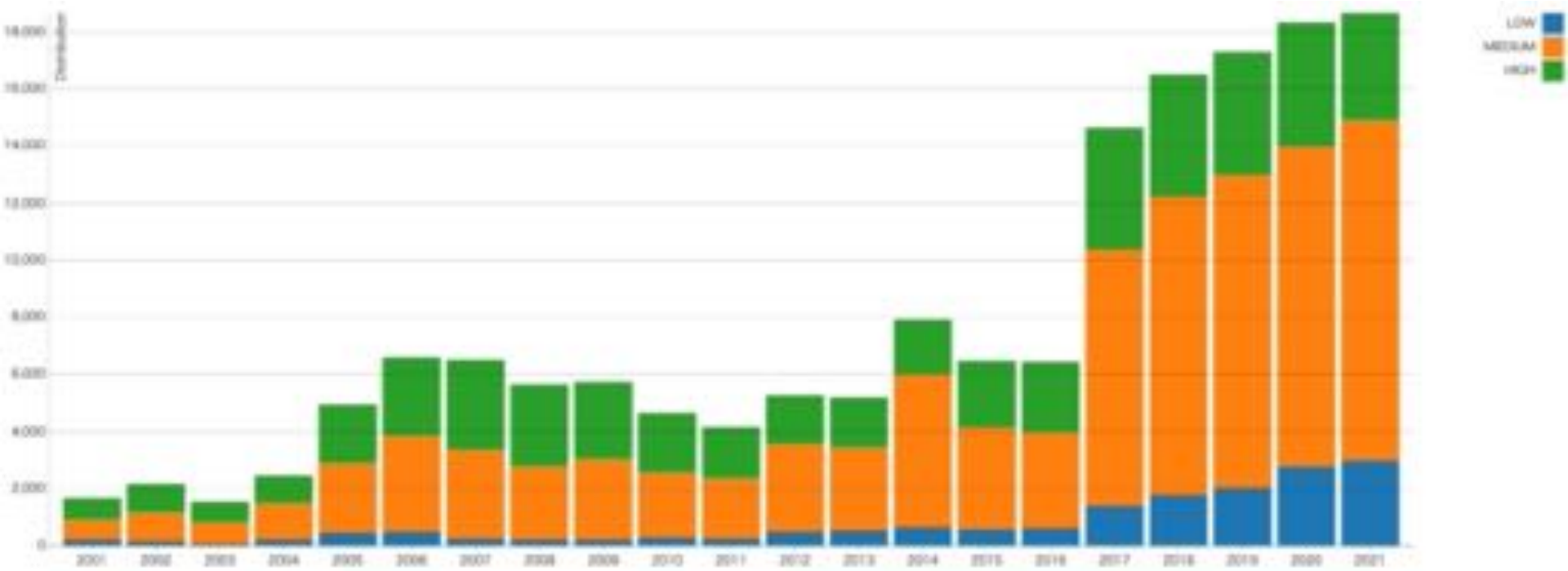
- Open Security Control Assessment Language (OSCAL)

National Vulnerability Database

標準に基づく脆弱性管理データの米国政府のリポジトリ

- CVE-ID
- 脆弱性の説明
- CVSS v3/v2 ベーススコア
- 参考情報、タグ（URL、種別：ベンダアドバイザリー等）
- CWE: Common Weakness Enumeration（共通脆弱性タイプ）
- CPE: Common Platform Enumeration（共通ﾌﾟﾗｯﾄﾌｫｰﾑ一覧）

NVD: CVE件数（CVSS v2ベーススコア）



https://nvd.nist.gov/general/visualizations/vulnerability-visualizations/cvss-severity-distribution-over-time as of Dec. 15, 2021

NVD: チームの役割

https://github.com/CVEProject/cvelist

MITRE/CNAから CVEデータを受領

- CVE-ID
- 脆弱性の説明
- 参考情報

NVD分析

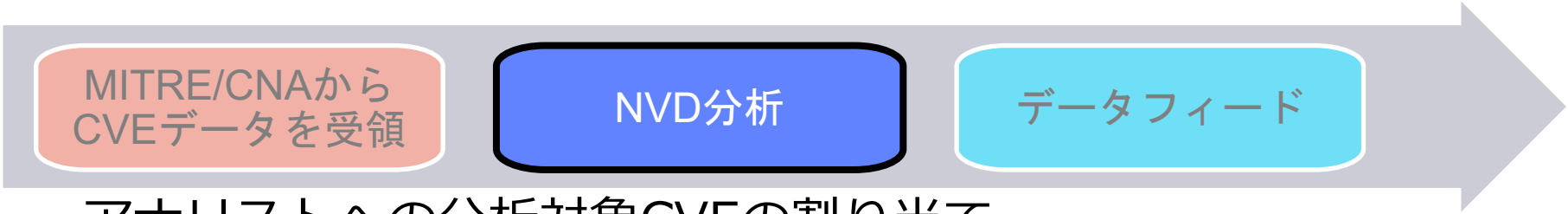
- CVSS
- CWE
- CPE
- 参考情報、タグ

データフィード

- JSON
- RSS
- API
- :

https://nvd.nist.gov/vuln/data-feeds

NVD分析プロセス



- アナリストへの分析対象CVEの割り当て

1. CVEデータのレビュー、参考情報のサイト確認・追加、タグ
2. CWE選択
3. CVSS 2.0/3.1スコア算出
4. CPE割り当て

- シニアアナリストによるレビュー

1. 参考情報、タグ

脆弱性の内容をできるだけ正確に把握し、CWE/CVSS/CPEの分析に必要な情報を揃える（透明性担保のため公開情報のみを参照）

例: CVE-2021-44228

Hyperlink	Resource
http://packetstormsecurity.com/files/165225/Apache-Log4j2-2.14.1-Remote-Code-Execution.html	Third Party Advisory, VDB Entry
http://packetstormsecurity.com/files/165260/VMware-Security-Advisory-2021-0028.html	
http://packetstormsecurity.com/files/165261/Apache-Log4j2-2.14.1-Information-Disclosure.html	
http://packetstormsecurity.com/files/165270/Apache-Log4j2-2.14.1-Remote-Code-Execution.html	
http://www.openwall.com/lists/oss-security/2021/12/10/1	Mailing List, Mitigation, Third Party Advisory
http://www.openwall.com/lists/oss-security/2021/12/10/2	Mailing List, Mitigation, Third Party Advisory

2. CWE

単純化した脆弱性のカテゴリのビュー（CWE-1003）から選択

分析当時に情報不足やカテゴリにマッチしないものもある

例: CVE-2021-44228

Weakness Enumeration

CWE-ID	CWE Name	Source
CWE-502	Deserialization of Untrusted Data	NIST Apache Software Foundation
CWE-20	Improper Input Validation	Apache Software Foundation
CWE-400	Uncontrolled Resource Consumption	Apache Software Foundation

3. CVSS 2.0/3.1スコア

対応の優先度判断の指標となることも多い脆弱性の深刻度

CVEデータ、参考情報をもとに各要素を選択しベーススコア算出

例: CVE-2021-44228

4. CPE

CVEデータ、参考情報をもとにベンダ名、プロダクト名からCPE辞書を検索し、CPEを組み合わせて割り当て

- 命名規則: NIST IR 7695 (CPE v2.3)
  - *cpe:2.3:part:vendor:product:version:update:edition:sw_edition:target_sw:target_hw:language:other*
- Configurations
  - 単体、特定のHW上で動作するOS、インストール済のソフトウェア
- CPE Match String Range:　バージョンの範囲指定
  - [“From”、”Up to”] x [“including”、”excluding”]

開発環境で1.5ヶ月間、日々分析

アナリストリーダーからのレクチャ

CVSS、CWE、CPEの仕様をあらためて知る

1日十数件の分析、都度チャットで質問

後日、公開されたNVDと比較して答え合わせ


COVID -19
Vaccination Clinic


COVID-19
I GOT VACCINATED!
PROTECTED TOGETHER.
NIST
ASPR
ASPR
NIST

問題提起、分析

3点の問題提起を行いチームで議論

1. CPEで表現できる粒度の適正化
   - Windowsで特定のCU以前に脆弱性がある場合、CPEの”update”でCUを識別しているため、単体の修正パッチ（KB）適用有無が表現できない
2. 同一端末に関連する複数の脆弱性がある場合のCVSS考慮
   - ローカルでの権限昇格の脆弱性がある端末に、RCEの脆弱性があった場合に、CVSS環境スコアのModified Attack VectorをNetworkにするといった関連付けにCPEを活用
3. CPE割り当ての効率化

理想：CVEデータ->CPE


```
{
  "data_type": "CVE",
  "data_format": "MITRE",
  "data_version": "4.0",
  "CVE_data_meta": {
    "ID": "CVE-2020-10002",
    "ASSIGNER": "product-security@apple.com",
    "STATE": "PUBLIC"
  },
  "affects": {
    "vendor": {
      "vendor_data": [
        {
          "vendor_name": "Apple"
          "product": {
            "product_data": [
              {
                "product_name": "watchOS",
                "version": {
                  "version_data": [
                    {
                      "version_affected": "<"
                      "version_value": "7.1"
```



```
  "description": {
    "description_data": [
      {
        "lang": "eng",
        "value": "A logic issue was addressed with improved state management. This issue is fixed in macOS Big Sur 11.0.1, watchOS 7.1, iOS 14.2 and iPadOS 14.2, iCloud for Windows 11.5, tvOS 14.2, iTunes 12.11 for Windows. A local user may be able to read arbitrary files."
      }
```


cpe:2.3:o:apple:watchos:*:*:*:*:*:*:*:*	Up to (excluding) 7.1
Show Matching CPE(s)	

現実：“version_value”の例

"Android-8.0”

"See provided reference”

"10 Version 1803 for x64-based Systems”

"All Versions < V1.04”

"12.1.0-12.4.0"

“18.3R2-S4“

“V500R001C30”

“All versions before 9.0.2”

“Ver.1.6”

“See advisory https://www.(略)”

CPE検索自動化の試み

versionの表記方法はベンダ固有も多く、正確な抽出は困難

“vendor_name”と”product_name”から部分CPE（*cpe:2.3:part:vendor:product:*）の候補を自動的にリストアップするだけでも実作業は大幅に楽になる

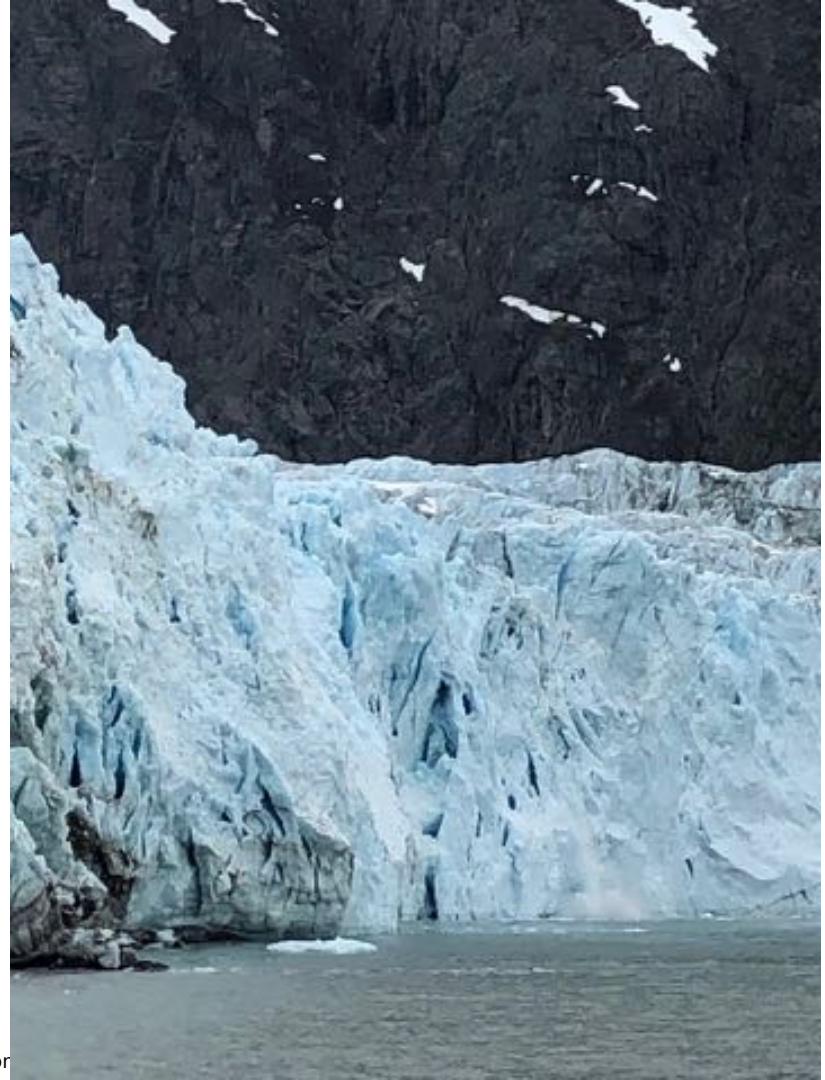
1. 編集距離による類似度を用いた検索
2. 自然言語処理と機械学習を用いた他クラス分類

編集距離による類似度を用いた検索

同日取得のCPE辞書が検索対象

NVDのCPEとの一致率で評価	一致率
• ほぼ完全一致検索 • 大文字を小文字に、” ”を”_”に置換	13.4%
• ファジー検索 • ＋ “corporation”等ストップワード、レーベンシュタイン距離	36.3%

類似度が高いが不一致 ⇒ CPE辞書のメンテナンス対象



Product Detail	Face Value (¥ JPY)	Face Value (USD) 1 USD= (JPY) ¥103.06	Handling Fee ($USD)	Unit Price ($USD)	Price Cat.	Qty	Total ($USD)
TOFBL29 Football Men's Quarterfinal Saturday, 7/31/2021, 19:00 - 22:00 Venue: Saitama Stadium	¥5,800	$56.28	$11.26	$67.53	D	4	$270.12

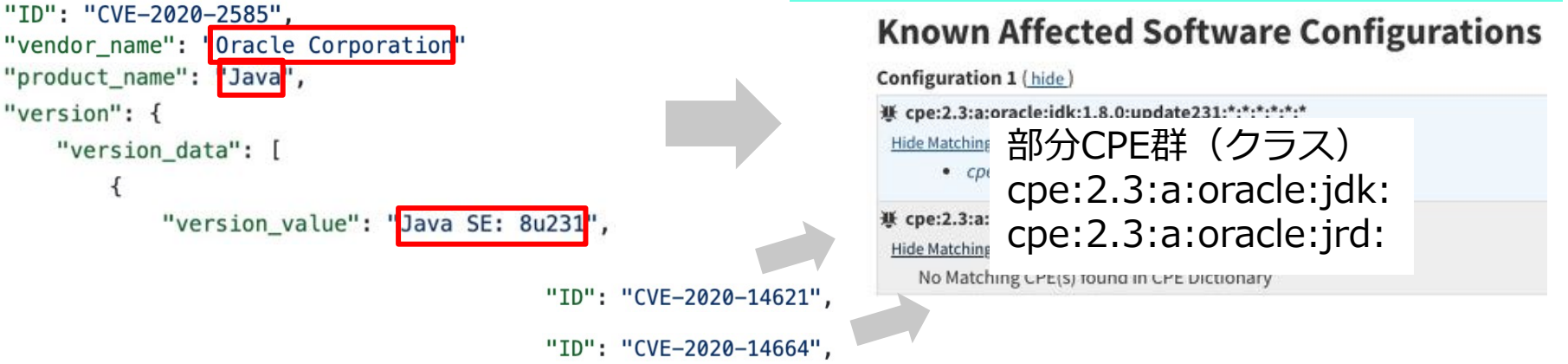
TICKET PURCHASER AGREES THAT TICKETS MAY NOT BE RE-SOLD OR USED TO PROMOTE ANY PERSON, ENTITY PRODUCT OR SERVICE.

NLP/MLを用いた他クラス分類

“vendor_name”、”product_name”、”version_value”の文字列を連結、置換とストップワードを適用、TF-IDFベクトル化

NVDのデータから、部分CPE群を正解クラスとして、ランダム・フォレストで学習・テスト

仮説）ソフトウェアコンポーネントを複数のプロダクトに流用していることが多いと考えられるため、部分CPEは同じパターン



評価結果

1クラスに含まれるデータ数	CVE数	クラス数	正解率
1	10,601	3,712	62.8%
2	8,127	1,238	82.4%
3	7,113	731	87.9%
4	6,477	519	88.3%
5	6,013	403	89.3%

継続的な評価 → 一貫性あるCPE割り当て

NVDに携わってみて

自動化できる範囲を広げていく

- 100%を求めない

継続的なMeasurement（測定）は重要

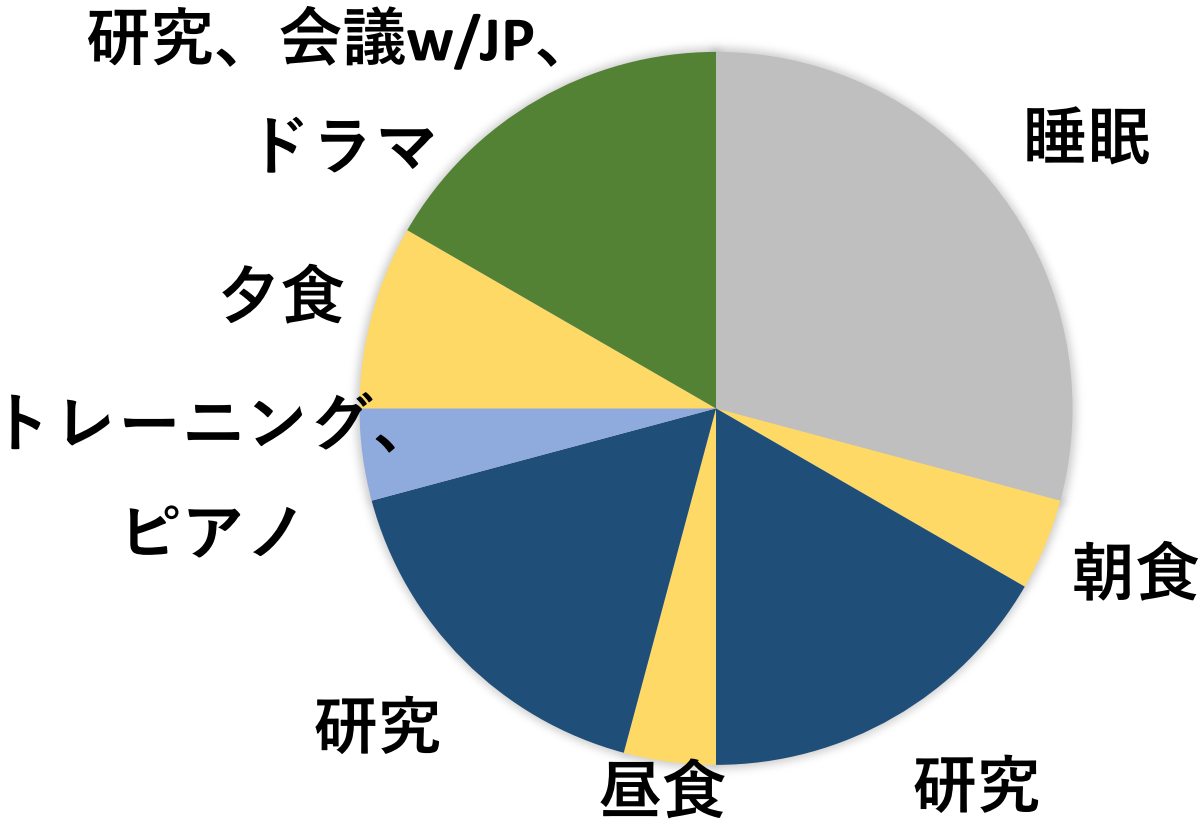
- 自動化に向けたデータ品質 *1
- 自動化の効果

理想と現実のはざまにはチャンスがたくさん

- 脆弱性情報の共有は世界への貢献
- より良い使い方のためのフィードバック

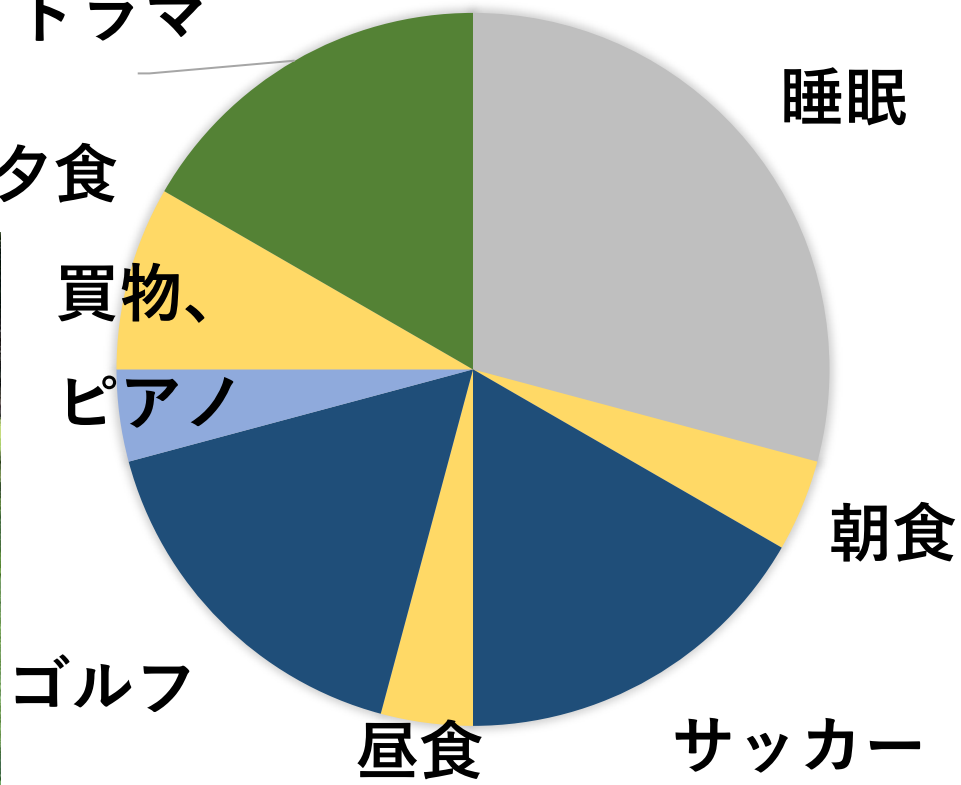
*1 佐々木ら: “脆弱性管理のためのCPEマッチング手法の提案” CSS2021

平日のある日の過ごし方



Cooking in the US
アイテム564件

週末のある日の過ごし方



4. ふりかえり

思いがあったからきっかけを掴める

- MWS: 仕組み作るか？
  - 膨大なデータを有効活用したい
- 博士: 行くか？
  - ガッツリ研究したい
- NIST: 想像できるやろ？
  - 世界に挑戦したい、入社動機を実現したい

実績の積み上げが自信になる

- MWS
  - データセット整備・分析、コミュニティ運営
- 博士
  - 機械学習、論文、GRIT
- NIST
  - フロー分析、NVD、米国生活

その他

- 生活全般
  - 1年目：試行錯誤
  - 2年目：活動の幅の拡大
  - 3年目：日常

- 英語
  - 当初、ミーティング前に喋る内容を考えたり下書き
  - そのうち、適当に喋って、適当に聞き流すように

mitsuhiro.hatada@ntt.com