資料2

サイバーセキュリティ向上のための AI活用に関する研究開発

国立研究開発法人 情報通信研究機構
サイバーセキュリティ研究所
サイバーセキュリティ研究室

高橋健志

本日の内容

1. サイバーセキュリティ研究室の簡単な紹介
2. サイバーセキュリティ分野におけるAI技術活用の現状
3. NICTの研究開発活動

サイバーセキュリティ研究室 研究マップ

インシデント分析センタ
NICTER

対サイバー攻撃アラートシステム
DAEDALUS

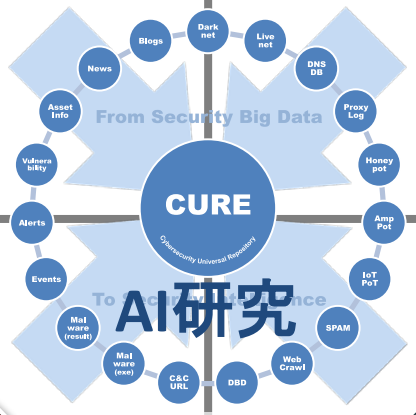
Passive

ネットワーク可視化システム
NIRVANA

サイバー攻撃統合分析プラットフォーム
NIRVANA改

Global (無差別型攻撃対策)

(標的型攻撃対策) Local



サイバーセキュリティ
ユニバーサル・リポジトリ
CURE

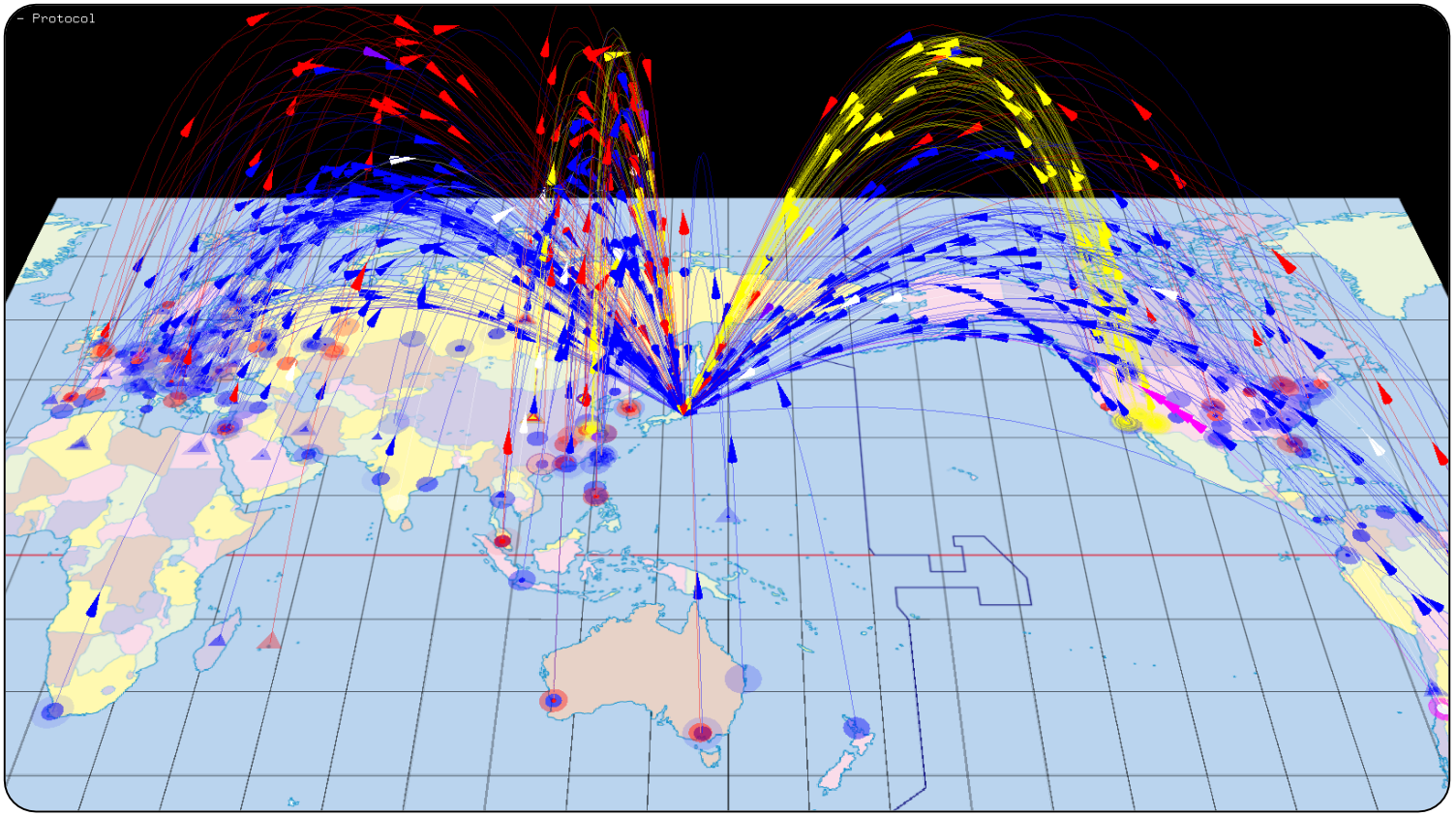
Active

委託研究
Web媒介型攻撃対策フレームワーク
WARPDRIVE
(ワープドライブ)

サイバー攻撃誘引基盤
STARDUST
(スターダスト)

NICTER

- 大規模サイバー攻撃観測・分析システム
- 国内外で30万の未使用IPアドレス“ダークネット”を観測
- 無差別型攻撃の大局的な傾向把握に有効



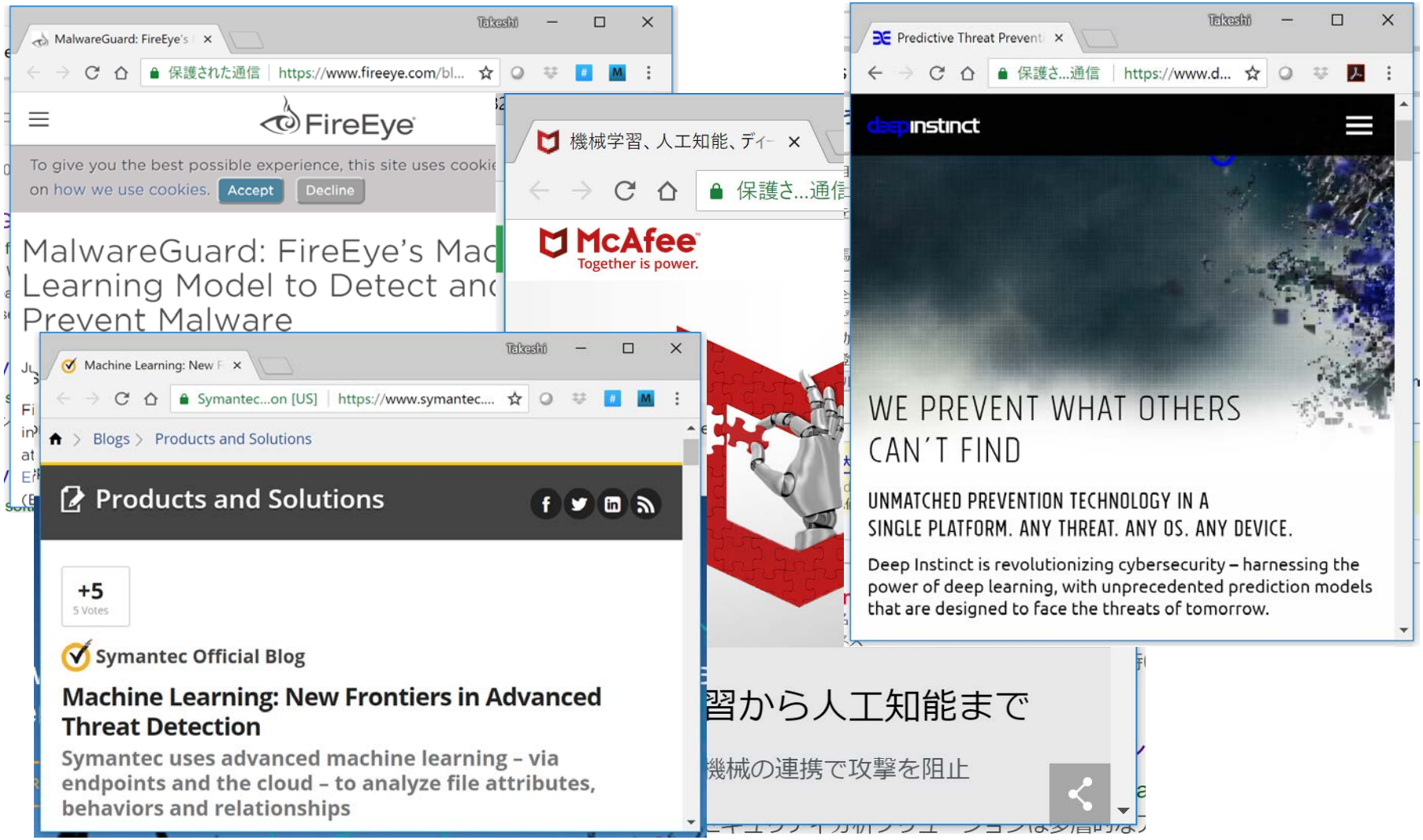
本日の内容

1. サイバーセキュリティ研究室の簡単な紹介
2. サイバーセキュリティ分野におけるAI技術活用の現状
3. NICTの研究開発活動

セキュリティ対策の自動化は世界的な潮流

- Cyber Grand Challengeでは、全ての攻防戦が、コンピュータにより自動で実施され、ヒトは見ているのみ
- 自動で脆弱性を発見し、パッチを作成し、対処
- カーネギーメロン大学の ForAllSecure チームの「Mayhem」というシステムが優勝。優勝賞金 200 万ドル（約 2 億円）を取得

深層学習等のAI技術活用を謳う商品は多数存在



但し、商品やアルゴリズムの詳細は明かされないため、詳細は不明

世界中がサイバーセキュリティへのAI活用を検討

近年になって、世界中の著名な研究組織がAIの適用可能性を模索
USENIX Security 2018にてAI関連の報告を実施した組織は下記の通り

欧州
- EPFL
- Frauhofer FKIE
- Max Planck Institute for Informatics
- RWTH Aachen University
- Siemens CERT
- Universidade de Lisboa

イスラエル
- Bar-Ilan Uniersity

アジア
- Chinese Academy of Science
- Beijing Jiaotong University

米国
- Boston University
- Columbia University
- Florida Institute of Technology
- Google Inc
- Indiana University
- Iowa State University
- MIT
- UC Santa Barbara
- University of Chicago
- University of Delaware
- University of Illinois
- University of Maryland
- Virginia Tech

世界中がサイバーセキュリティへのAI活用を検討

近年になって、世界中の著名な研究組織がAIの適用可能性を模索
CCS 2018にてAI関連の報告を実施した組織は下記の通り

欧州
- Lancaster University
- University College London

アジア
- Inha University
- Peking University
- Zhejiang University
- The Hong Kong Polytechnic University
- Chinese Academy of Sciences
- Hanyang University
- National University of Singapore

米国
- University of Central Florida
- Florida International University
- Northwest University
- Lehigh University
- The Pennsylvania State University
- Virginia Tech
- University of Pennsylvania
- Symantec
- UC Riverside
- UC Berkeley
- University of Illinois at Urbana-Champaign
- University of Massachusetts

最近報告されている研究内容

トラフィックの異常検知&マルウェア検知
(long standing area)
- Explainable system
- Performance improvements /real-time operations

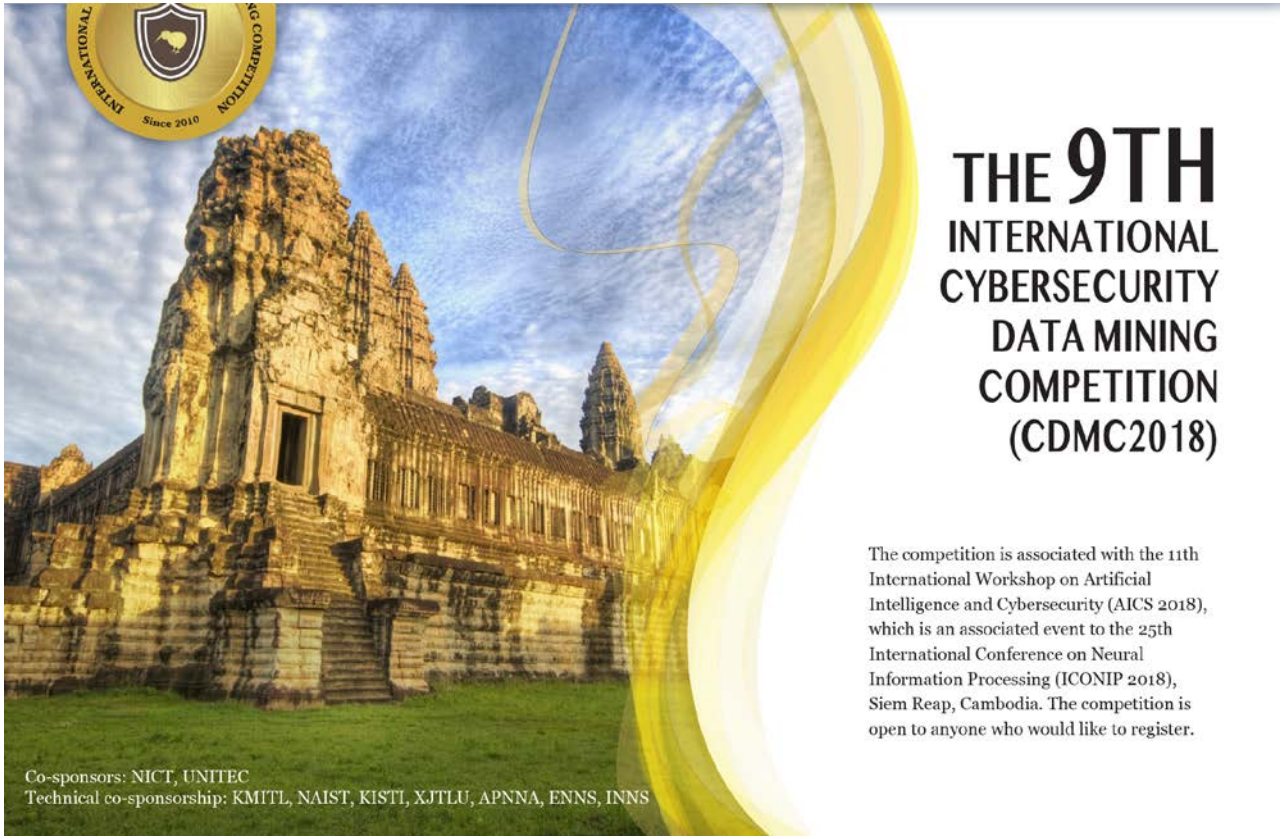
各種コンピューティングシステムへの攻撃
- Solving captcha
- Malfunctioning voice recognition systems

非匿名化 (プライバシーに対する攻撃)
- Code Authorship Identification
- Document author attribute classification
- Identification of account pertaining review comments

対策・防衛技術
- Program debloating (minimize vulnerabilities)
- Watermarking DNN
- Event prediction

機械学習の脆弱性
- Poisoning attacks
- Vulnerabilities of transfer learning
- Attribute inference attacks
- Model reuse attack

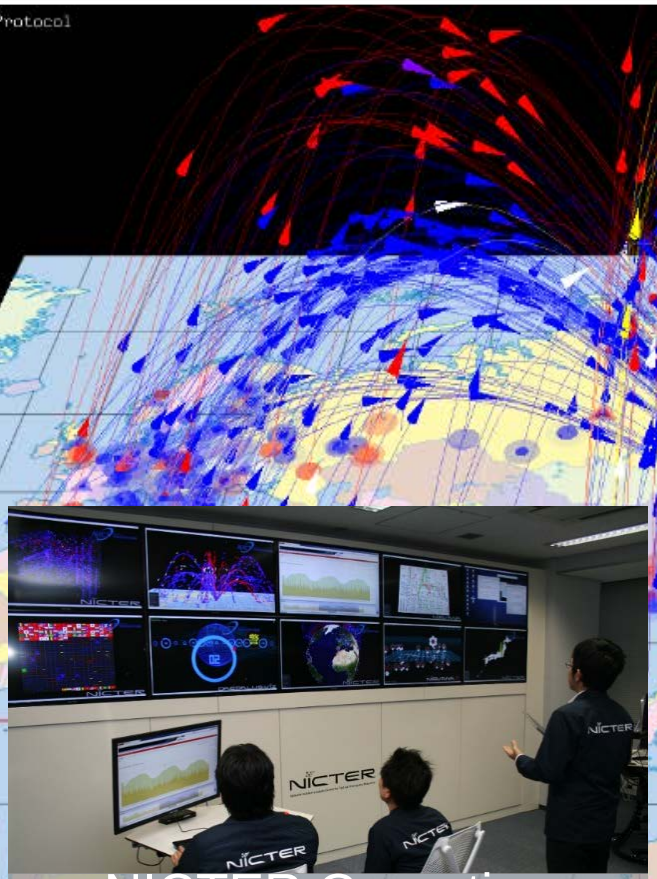
NICTはAI x cybersecurityの重要性をいち早く認識



- 11th International Data Mining and Cybersecurity Workshop (DMC), 2018: ICONIP併催
- 9th International Cybersecurity Data Mining Competition (CDMC), 2018: DMC併催

独自のネットワーク観測技術を用いてデータを蓄積

- ✓ 大規模なダークネット空間を観測
- ✓ NICTERやDAEDALUSなどのシステムを構築



NICTER Operation Room

我々のデータセット

カテゴリ	蓄積データの具体例
ダークネット関連データ	未使用IPアドレス空間に送られたトラフィックデータ。Pcapファイル、統計情報、悪性ホスト情報などを含む
ライブネット関連データ	NICT内部のトラフィックデータ。Pcapファイル、フローデータ、セキュリティ機器により生成されたセキュリティアラートなどを含む
マルウェア関連データ	マルウェア検体、静的解析結果、動的解析結果、など
スパム関連データ	スパム(ダブルバウンス)メールデータ、統計情報、など
Android関連データ	Androidアプリケーションパッケージファイル、カテゴリや説明文などのアプリのメタデータ、など
ブログ・記事	ツイート、セキュリティベンダーブログ、など
Webクローラ	URLリスト、Webコンテンツ、それらの評価結果、など
ハニーポットデータ	高対話型/低対話型ハニーポットから得られたデータ
商用インテリジェンスデータ	VirusTotal、SecureWorks、Anubis、DomainTools、Malnet、Team5などから購入したマルウェアをホストしているサイトの情報、ボットやC&Cのリスト、ドメイン履歴データ、検体、脅威レポートなど

本日の内容

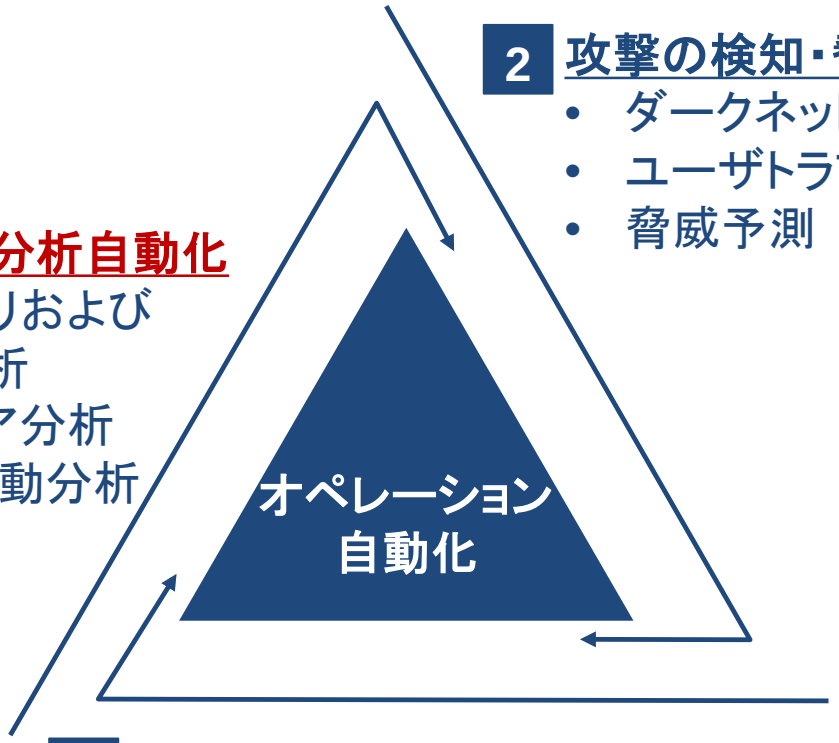
1. サイバーセキュリティ研究室の簡単な紹介
2. サイバーセキュリティ分野におけるAI技術活用の現状
3. NICTの研究開発活動

我々が現在注力しているドメイン

1

マルウェア機能分析自動化

- Androidアプリおよびマーケット分析
- IoTマルウェア分析
- マルウェア自動分析ツール開発



2 **攻撃の検知・脅威予測**

- ダークネット分析
- ユーザトラフィックの異常検出
- 脅威予測

3 **インシデント対応の優先順位の自動判定**

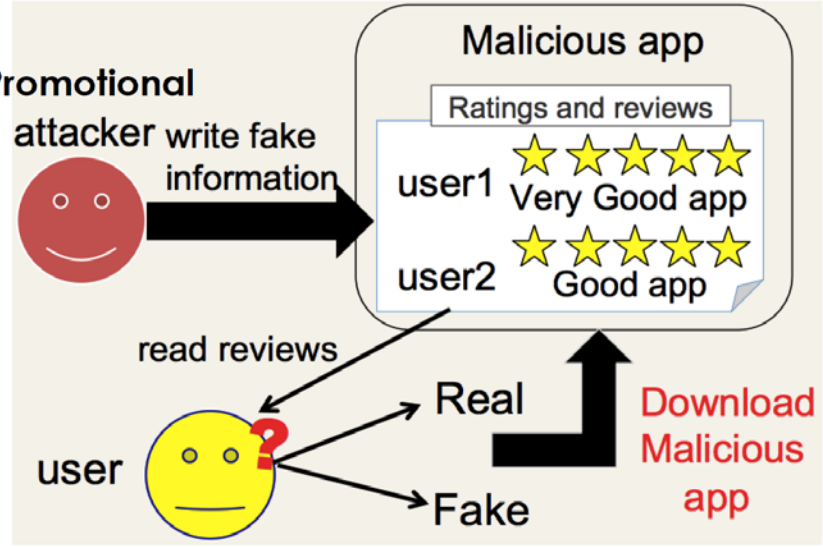
- アラートスクリーニング
- 脆弱性の分析

Androidプラットフォーム上でのマルウェア検知

1. *Androidマルウェアの検知及び分類*

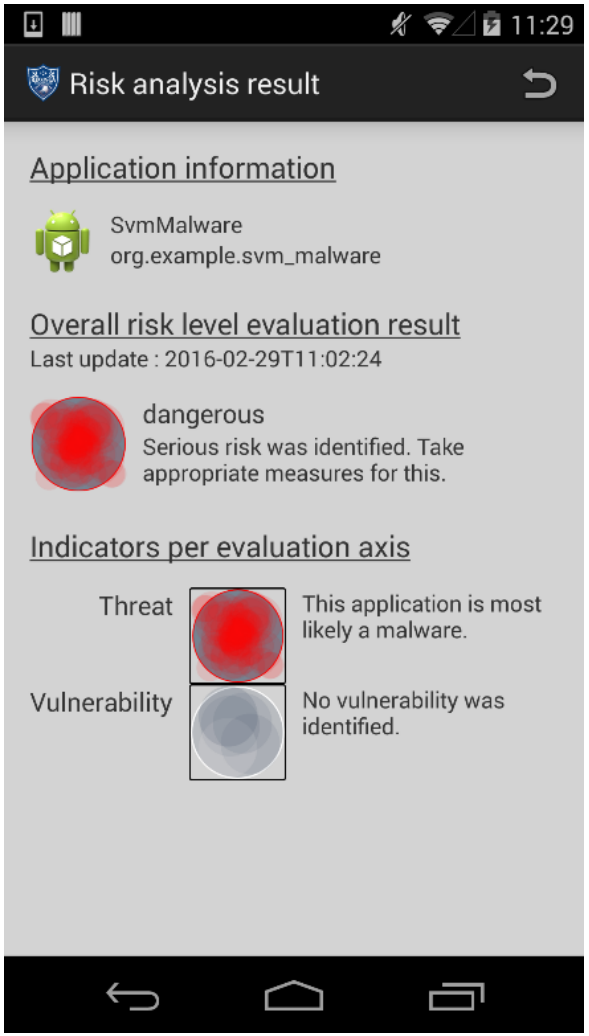
   a. 入力とする特徴情報を生成 (静的解析、動的解析、Web情報収集)
   b. 機械学習(SVM)および特徴選択技術の活用(explainable)
   c. ニューラルネットワーク/深層学習の活用(non-explainable)
   d. コード分析を回避した検知率の向上

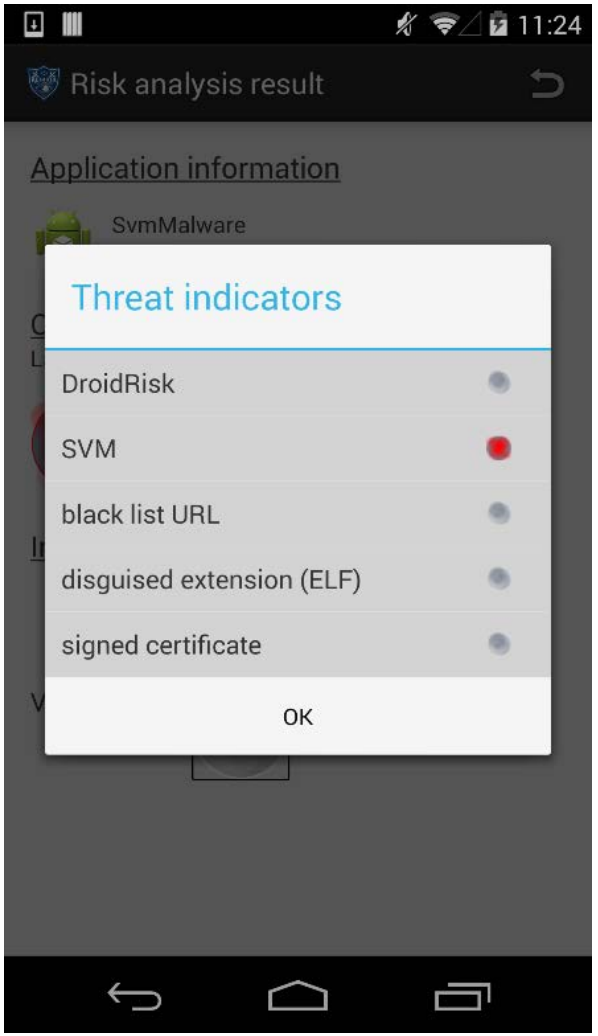
2. *promotional attackとdemotional attackをマーケット上で検知*

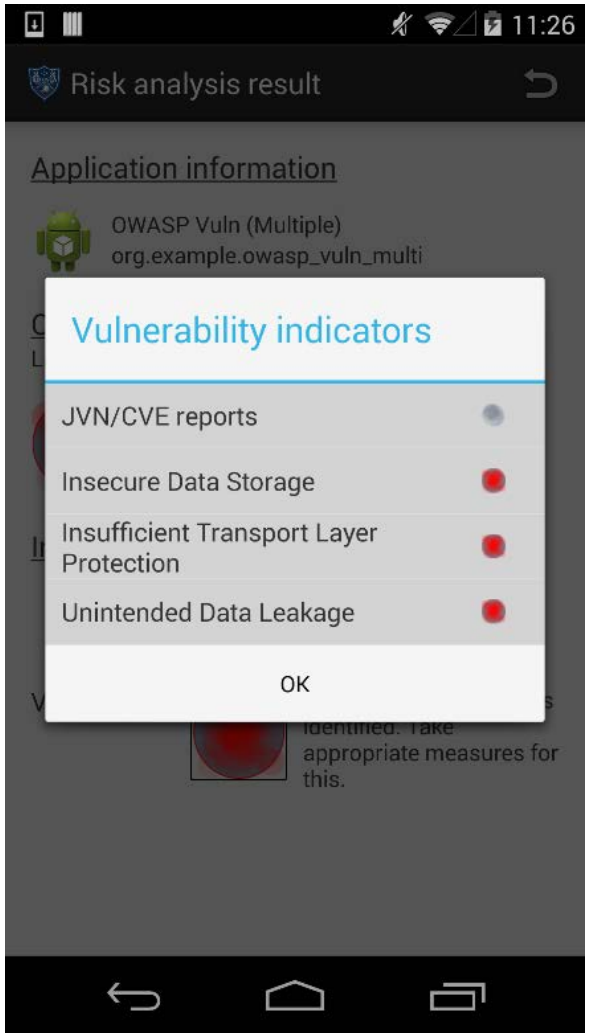


Source: T.Takahashi, T.Ban, "Android Application Analysis using Machine Learning Techniques," Intelligent Systems Reference Library, 2019.

マルウェアの特定をわかりやすくユーザに掲示



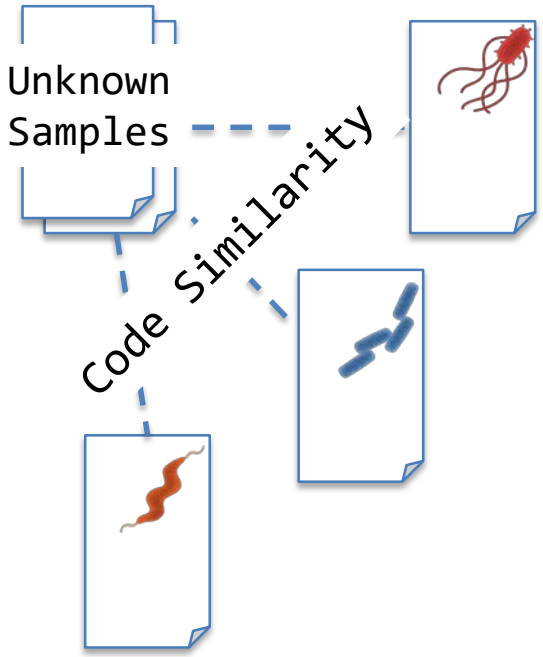


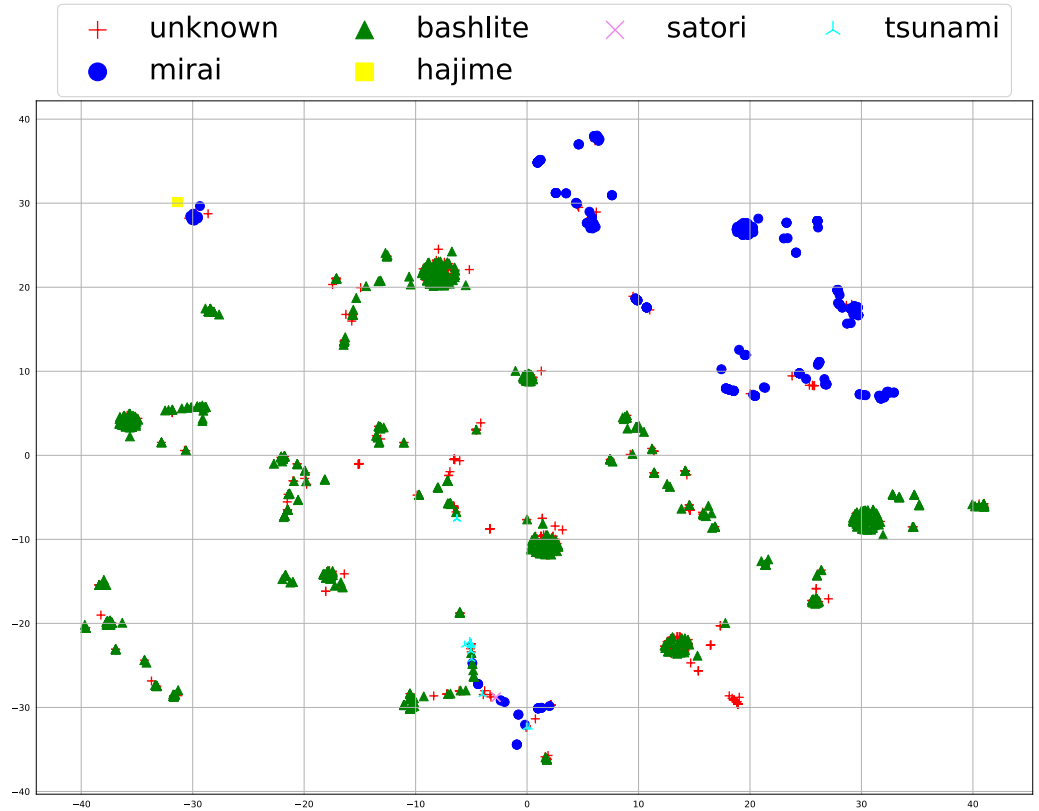


本研究のデータセットはWebから入手可能 (要申請) http://mobilesec.nict.go.jp.

収集したマルウェアの帰属するファミリーを分析

未知のマルウェアサンプルを複数のファミリーに分類。分類することにより、
これらのサンプルの効率的な分析に貢献。





Samples mapped on a two-dimensional plane with T-SNE

我々が現在注力しているドメイン

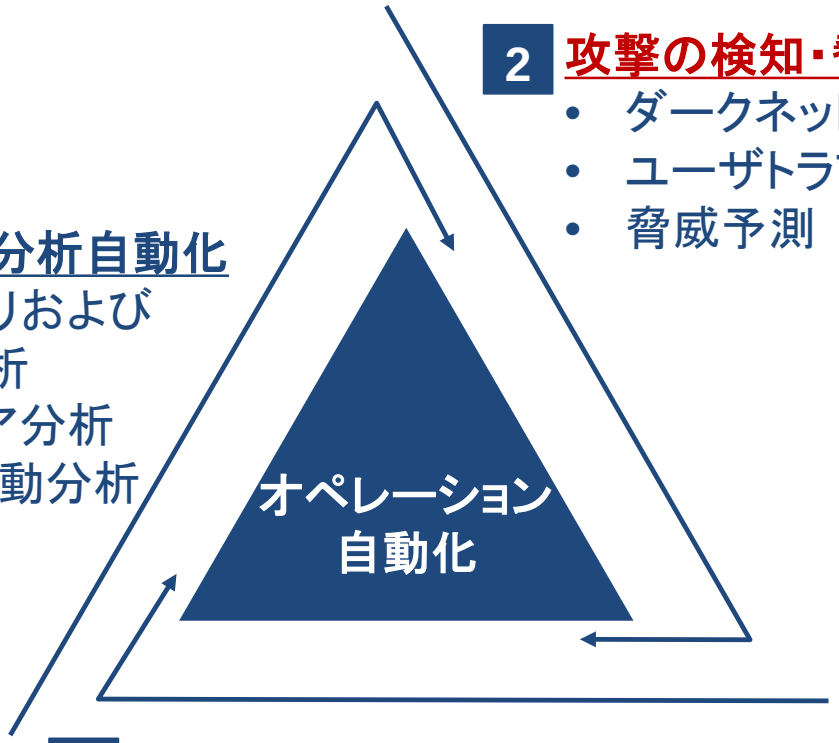
1

マルウェア機能分析自動化

- Androidアプリおよびマーケット分析
- IoTマルウェア分析
- マルウェア自動分析ツール開発

2 **攻撃の検知・脅威予測**

- ダークネット分析
- ユーザトラフィックの異常検出
- 脅威予測

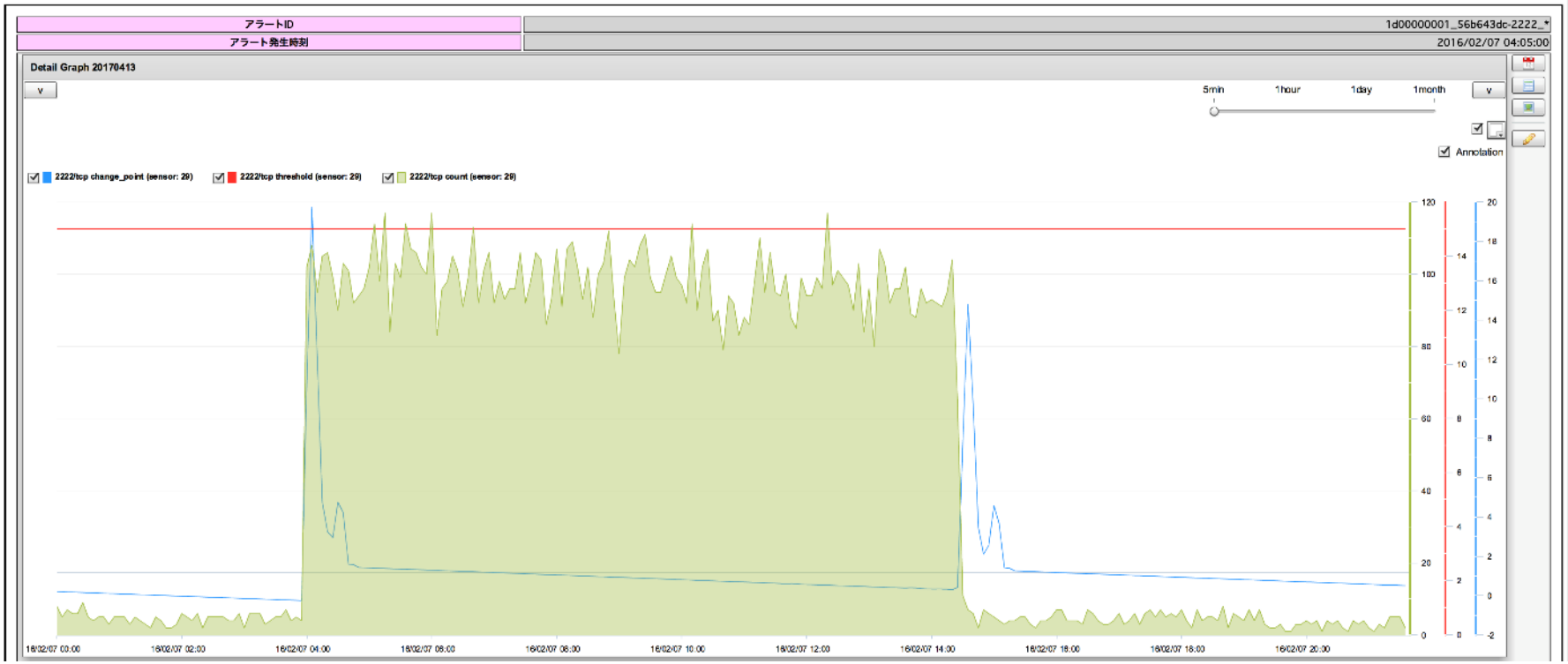


3 **インシデント対応の優先順位の自動判定**

- アラートスクリーニング
- 脆弱性の分析

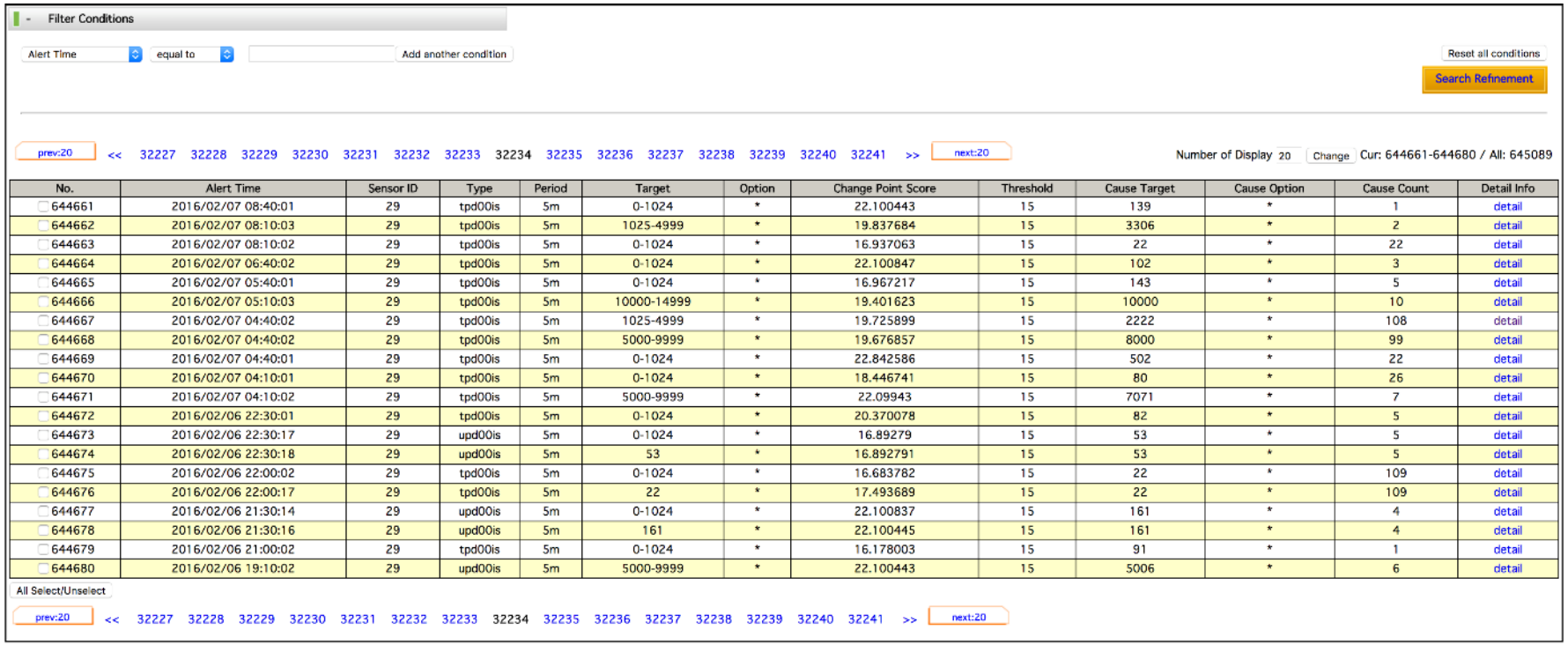
ホストの協調動作を検知

同一のボットネット内にあるボットの活動は協調性を示すことが多い (C&Cサーバの指示で一斉に動作するため). 多数のホストからほぼ同時にトラフィックが観測された場合には、それらのホストがボットである可能性が考えられる



上図では、横軸は時間を、縦軸はダークネット空間で観測されたパケット送信元数を示している。このケースでは、特定の時刻に、とあるマルウェアの活動が活性化され、また停止されるケースが観測できる。

検知時にはアラートを自動生成



Filter Conditions

Alert Time | equal to | Add another condition | Reset all conditions | Search Refinement

prev:20 << 32227 32228 32229 32230 32231 32232 32233 32234 32235 32236 32237 32238 32239 32240 32241 >> next:20

Number of Display 20 Change Cur: 644661-644680 / All: 645089

No.	Alert Time	Sensor ID	Type	Period	Target	Option	Change Point Score	Threshold	Cause Target	Cause Option	Cause Count	Detail Info
644661	2016/02/07 08:40:01	29	tpd00is	5m	0-1024	*	22.100443	15	139	*	1	detail
644662	2016/02/07 08:10:03	29	tpd00is	5m	1025-4999	*	19.837684	15	3306	*	2	detail
644663	2016/02/07 08:10:02	29	tpd00is	5m	0-1024	*	16.937063	15	22	*	22	detail
644664	2016/02/07 06:40:02	29	tpd00is	5m	0-1024	*	22.100847	15	102	*	3	detail
644665	2016/02/07 05:40:01	29	tpd00is	5m	0-1024	*	16.967217	15	143	*	5	detail
644666	2016/02/07 05:10:03	29	tpd00is	5m	10000-14999	*	19.401623	15	10000	*	10	detail
644667	2016/02/07 04:40:02	29	tpd00is	5m	1025-4999	*	19.725899	15	2222	*	108	detail
644668	2016/02/07 04:40:02	29	tpd00is	5m	5000-9999	*	19.676857	15	8000	*	99	detail
644669	2016/02/07 04:40:01	29	tpd00is	5m	0-1024	*	22.842586	15	502	*	22	detail
644670	2016/02/07 04:10:01	29	tpd00is	5m	0-1024	*	18.446741	15	80	*	26	detail
644671	2016/02/07 04:10:02	29	tpd00is	5m	5000-9999	*	22.09943	15	7071	*	7	detail
644672	2016/02/06 22:30:01	29	tpd00is	5m	0-1024	*	20.370078	15	82	*	5	detail
644673	2016/02/06 22:30:17	29	upd00is	5m	0-1024	*	16.89279	15	53	*	5	detail
644674	2016/02/06 22:30:18	29	upd00is	5m	53	*	16.892791	15	53	*	5	detail
644675	2016/02/06 22:00:02	29	tpd00is	5m	0-1024	*	16.683782	15	22	*	109	detail
644676	2016/02/06 22:00:17	29	tpd00is	5m	22	*	17.493689	15	22	*	109	detail
644677	2016/02/06 21:30:14	29	upd00is	5m	0-1024	*	22.100837	15	161	*	4	detail
644678	2016/02/06 21:30:16	29	upd00is	5m	161	*	22.100445	15	161	*	4	detail
644679	2016/02/06 21:00:02	29	tpd00is	5m	0-1024	*	16.178003	15	91	*	1	detail
644680	2016/02/06 19:10:02	29	upd00is	5m	5000-9999	*	22.100443	15	5006	*	6	detail

All Select/Unselect

prev:20 << 32227 32228 32229 32230 32231 32232 32233 32234 32235 32236 32237 32238 32239 32240 32241 >> next:20

1. 我々のプロトタイプは上記のようにアラートを自動生成するが、リアルタイム動作およびfalse positive/negativeの最小化が課題
2. 現在、我々はglasso、NMF、テンソル分解を用いて本課題にアプローチ

テンソル分解を用いたボットネットの活動検知事例

1. 著名なセキュリティ関連ブログにて本件が報告される前に、我々は協調動作を検知
2. NICTERシステムが顕著なトラフィック量の増加を検知する前に協調動作を検知

我々が現在注力しているドメイン

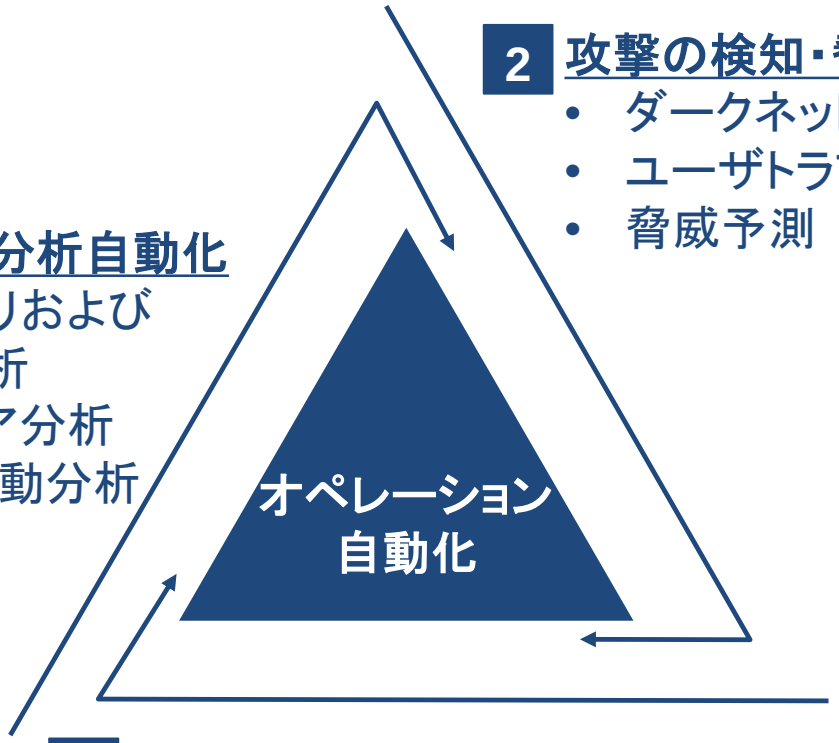
1

マルウェア機能分析自動化

- Androidアプリおよびマーケット分析
- IoTマルウェア分析
- マルウェア自動分析ツール開発

2 **攻撃の検知・脅威予測**

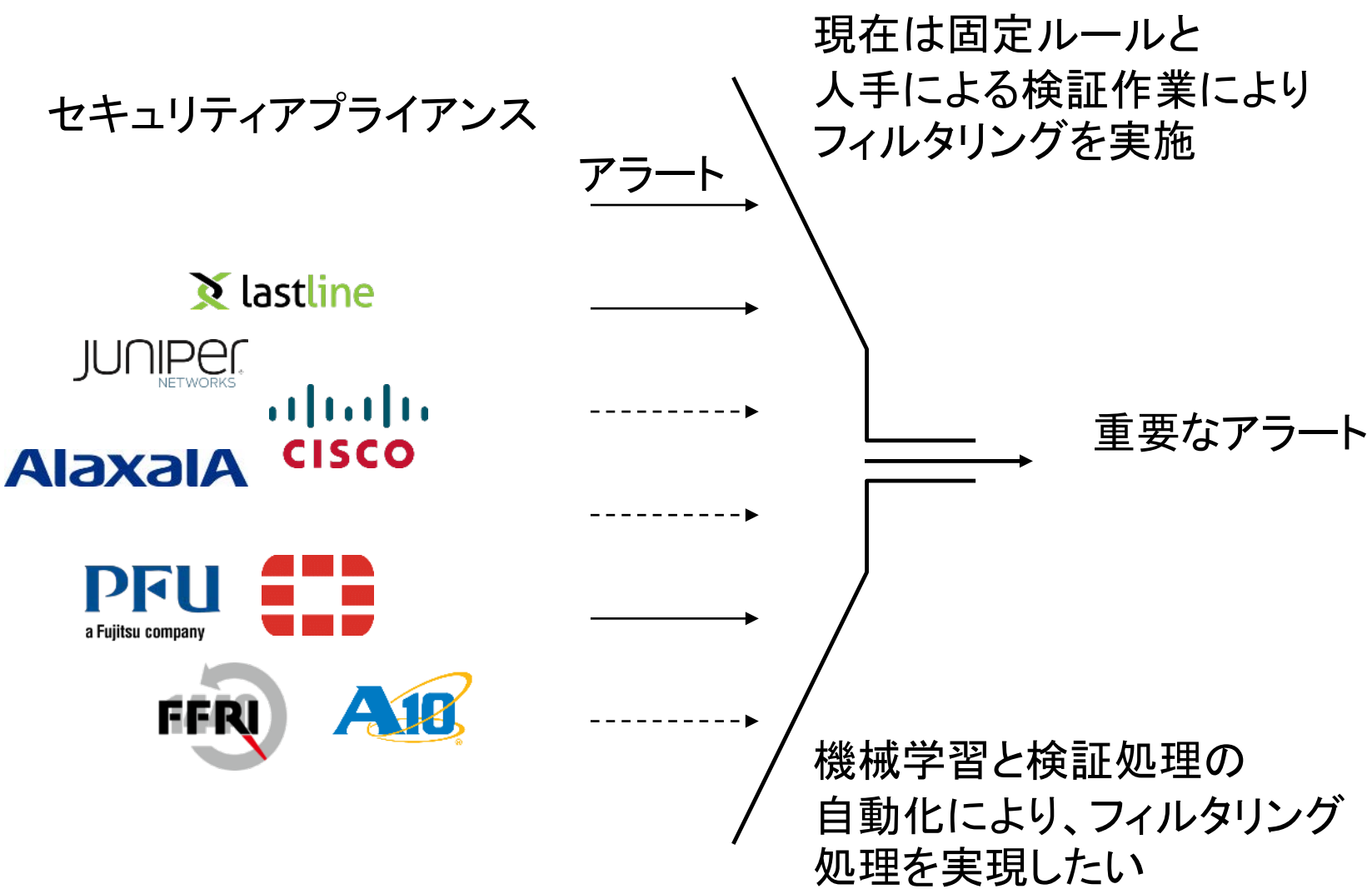
- ダークネット分析
- ユーザトラフィックの異常検出
- 脅威予測



3 **インシデント対応の優先順位の自動判定**

- アラートスクリーニング
- 脆弱性の分析

アラートのスクリーニング及び優先順位付け

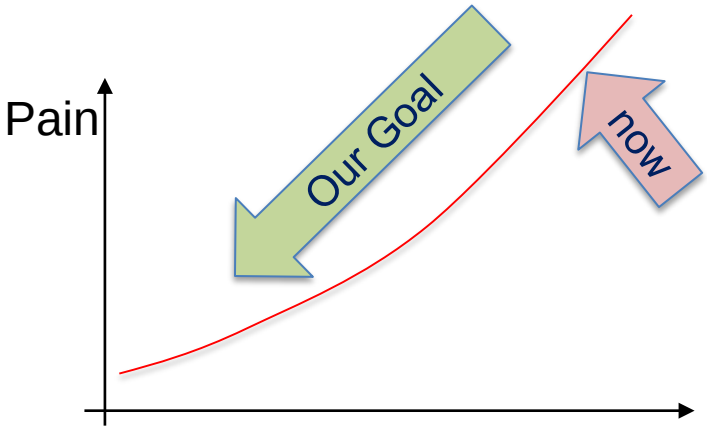


セキュリティオペレータの負荷を軽減したい

アラート
~ 100,000/day

精査が必要な
アラート
~ 100/day

対処が必要な
アラート
~ 5 /day



専門家が1日4時間もの時間を費やしているのが現状

主な発表論文

1. H.Kanehara, Y.Murakami, J.Shimamura, T.Takahashi, D.Inoue, N.Murata, "Real-Time Botnet Detection Using Nonnegative Tucker Decomposition," ACM SAC, 2019.
2. B.Sun, T.Ban, S.Chang, Y.Sun, T.Takahashi, D.Inoue, "A Scalable and Accurate Feature Representation Method for Identifying Malicious Mobile Applications," ACM SAC, 2019.
3. T.Takahashi, T.Ban, "Android Application Analysis using Machine Learning Techniques," Intelligent Systems Reference Library, 181 - 205, 2019.
4. S.Chang, Y.Sun, W.Chuang, M.Chen, B.Sun, T.Takahashi, "ANTSdroid:Using RasMMA Algorithm to Generate Malware Behavior Characteristics of Android Malware Family," IEEE PRDC, 2018.
5. L.Zhu, T.Ban, T.Takahashi, D.Inoue, "Employ Decision Value for Binary Soft Classifier Evaluation with Crispy Reference," ICONIP, 2018.
6. R.Iijima, S.Minami, Z.Yunao, T.Takehisa, T.Takahashi, Y.Oikawa, T.Mori, "Poster: Audio Hotspot Attack: An Attack on Voice Assistance Systems Using Directional Sound Beams," ACM CCS, 2018.
7. T.Takahashi, B.Panta, Y.Kadobayashi, K.Nakao, "Web of cybersecurity: Linking, locating, and discovering structured cybersecurity information," Int J Commun Syst. 2017.

現在我々が感じていること

1. 我々はAIのcybersecurity応用が相当難しいのは分かっているものの、その重要性にはいち早く気づき、10年前から検討をしてきている
2. 機械学習をNICTERシステム内で自動で動かせるようになるまで10年
3. しかしながら未だに課題が多い
    - リアルタイム性の問題 (そもそもデータの次元数が高い)
    - 判定根拠が良く見えない形式の深層学習結果などは、そのまま実運用へ適用するのはセキュリティ分野ではリスクが高い
4. 今は、AIのサイバーセキュリティ活用の可能性をみんなが試している時代
5. このタイミングで実際に使える機械学習の研究開発を強化をし、同時にデータを継続的に蓄積することで、10年後に打っていけるコアコンピタンスを育成していきたい