Innovative R&D by NTT

# MWS 2017 意見交換会

## D3M (**D**rive-by **D**ownload **D**ata by **M**arionette) データセット説明

NTTセキュアプラットフォーム研究所

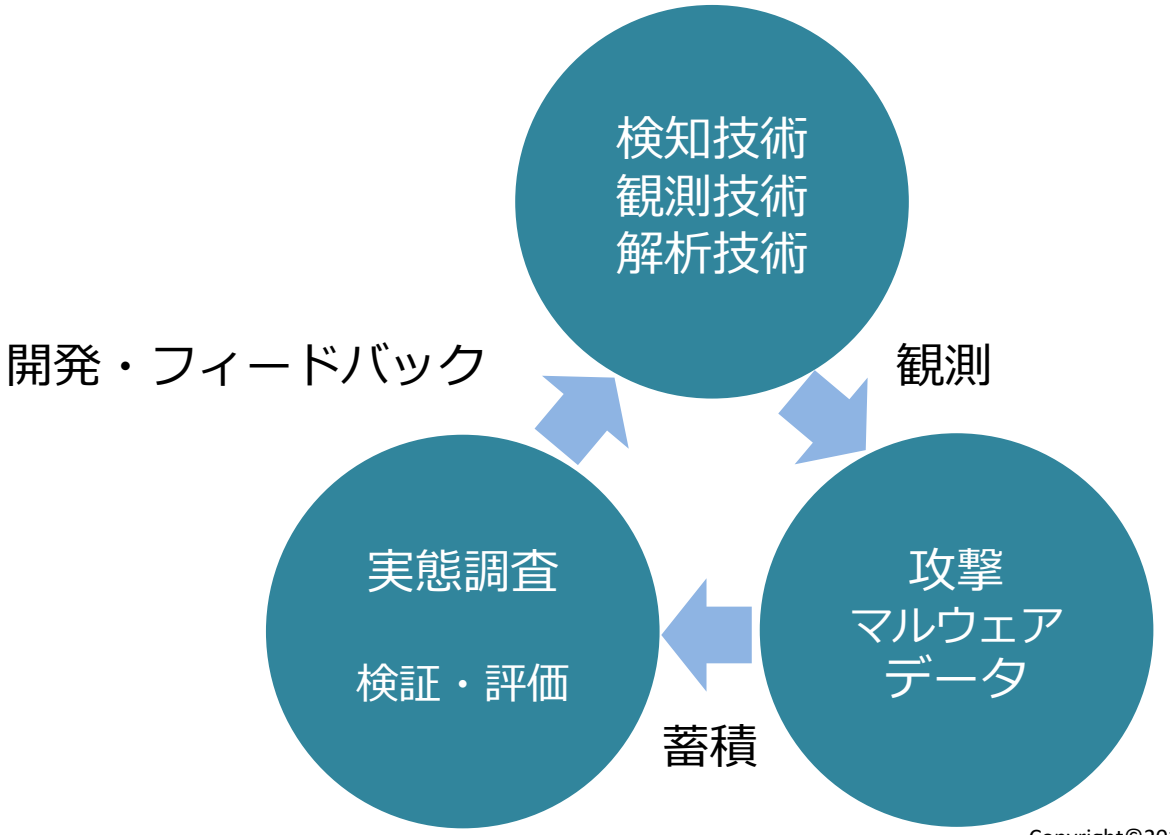
高田 雄太、秋山満昭

2017年06月06日

# はじめに：データセット提供の意義



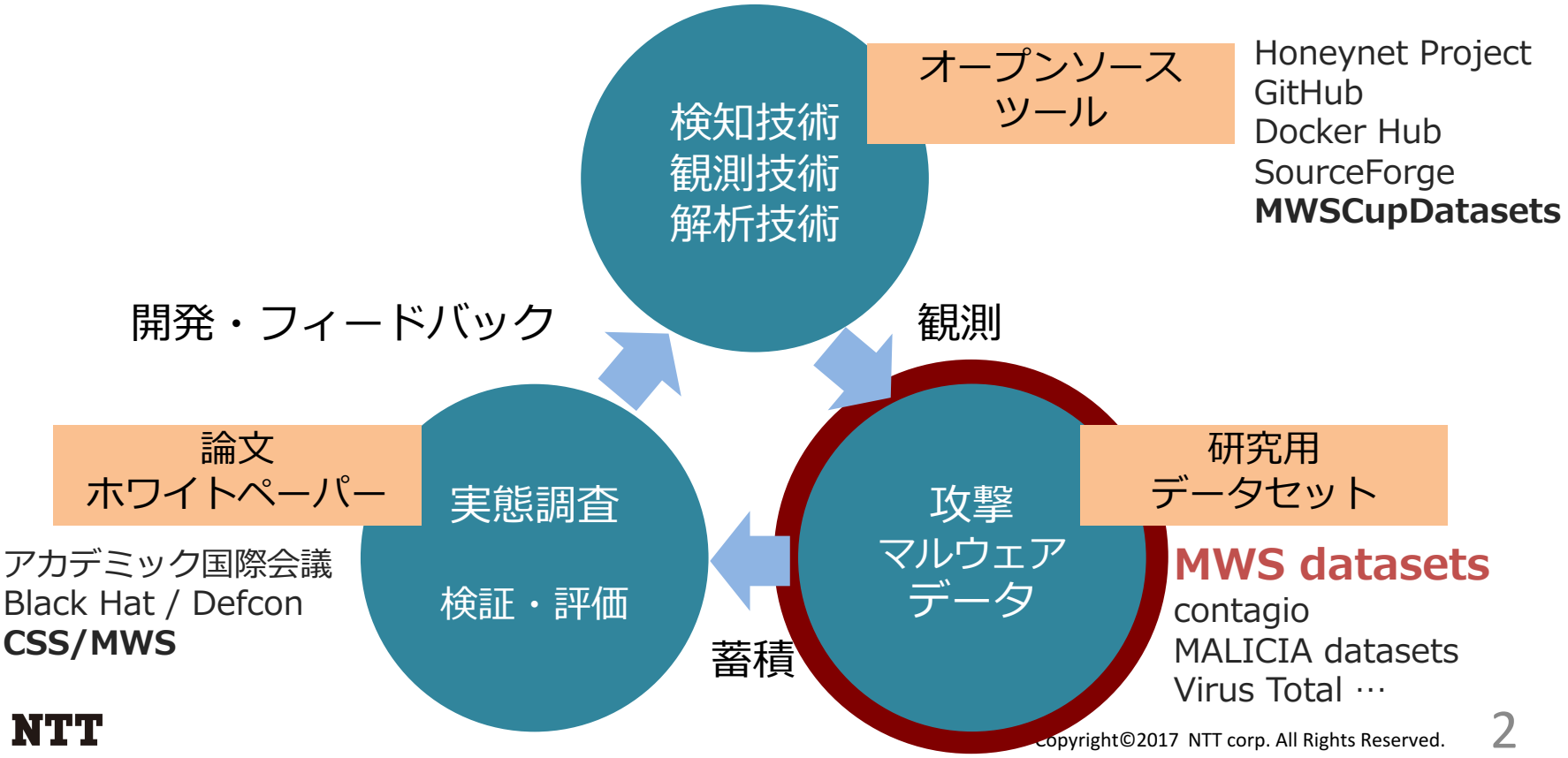
- 研究開発のサイクルを加速させ、日々進化するサイバー攻撃に対抗
  - **サイクルの循環を加速させるには？**



# 研究開発サイクルを加速させるために



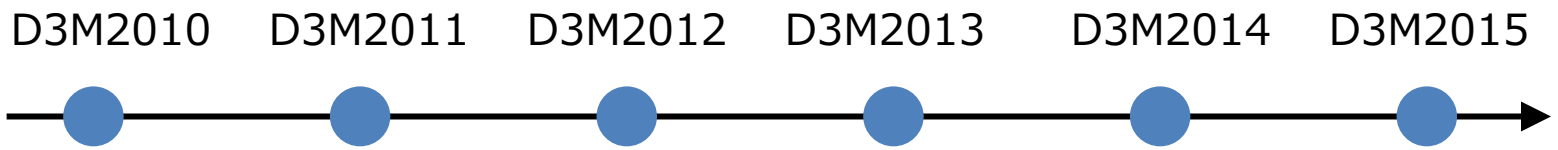
- 近年、各フェーズをサポートする情報やツールが充実化
  - もちろん、さらなる充実化は必要
- これらを有効に活用し、**研究開発を加速**



# D3M の概要



- データセットの内容
  - 攻撃通信データ
    - 悪性 URL を巡回した際に得られたドライブバイダウンロード攻撃の通信データ
  - マルウェア情報
    - ドライブバイダウンロード攻撃によってホスト上にダウンロードされた実行形式のファイルなど
  - マルウェア通信データ
    - 実行形式のファイルを取得して24時間以内にマルウェアサンドボックス上で実行した際の通信データ
    - マルウェアサンドボックスはインターネットに接続可能（攻撃通信は遮断）

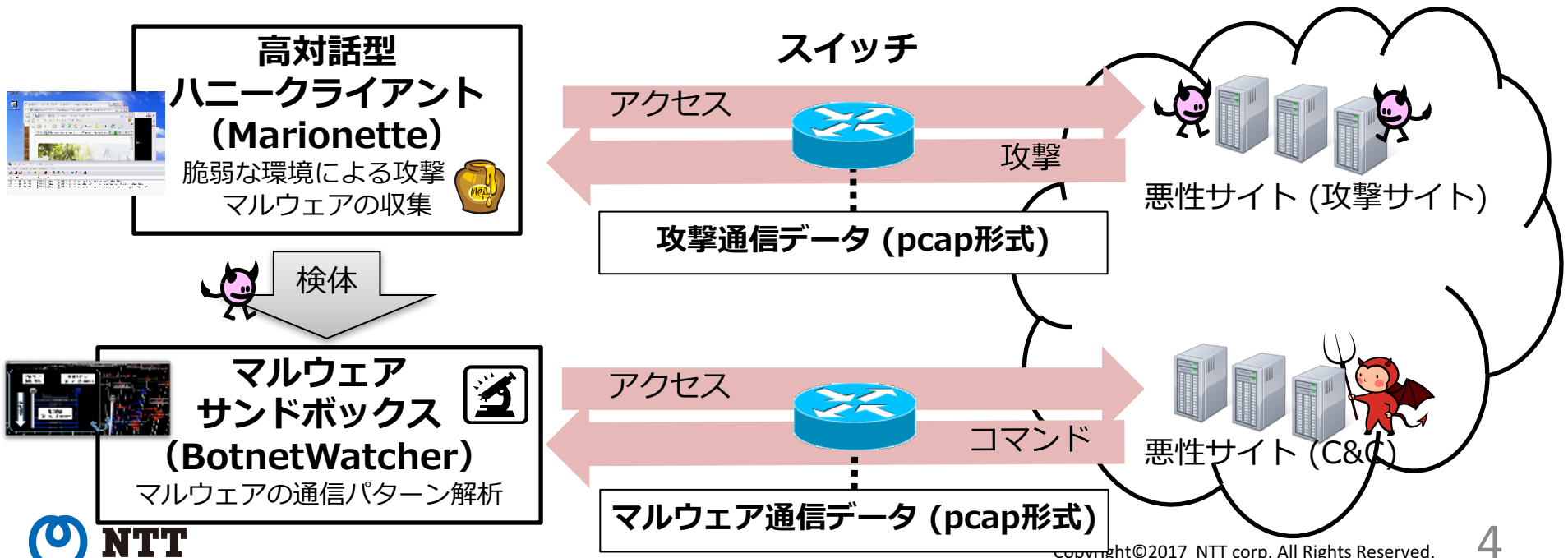



D3M には D3M2010～2015 が含まれており、
数年にまたがった解析により攻撃の変化・傾向を観測できる（かも）

# D3M の取得環境



- ドライブバイダウンロード攻撃に関連する URL をブラウザへ投入し、自動的に発生する一連の Web 通信、ダウンロードした実行ファイルの通信を収録
  1. 独自に収集した悪性 URL リストを高対話型ハニークライアント Marionette で巡回
  2. 検知した URL を直ちに再巡回し、その際の通信データを記録
  3. 2. で取得した実行形式のファイルを、マルウェアサンドボックス Botnet Watcher で解析し、その時の通信データを記録



# D3M に含まれる情報



- 提供されるデータ形式
  - pcap（ドライブバイダウンロード攻撃通信、マルウェア通信）

- 攻撃を行う URL, ドメイン名, IP アドレス
- 難読化された JavaScript
- 攻撃コード（HTML, JavaScript, PDF, Jar, …）
- C&C サーバとの通信
- など

# 最新のD3Mデータセットについて



- これまでのデータセット (D3M2010-2015) を参考に、是非**独自の**攻撃検知/観測基盤を構築していただき、最新データの収集に挑戦してもらいたい！
  - **MWS Cup 2016 に参加した各チームの解析ツールを提供: MWSCup2016Datasets**
    - 既に GitHub アップロード済みスクリプトも！
      - https://github.com/h-uekawa/drive-by-finder
  - 多種多様な組織が所属する MWS におけるデータセット共有、データセットを活用した研究推進を活発にしていきたい

**今後のデータセットは(Cup等を通じて)**
**皆さんで協力して作っていきましょう！**
異なる観測点・環境でのデータ収集は重要

# “意見交換” 会ということなので...



- 実は自前でデータセットを収集/活用している という組織・個人はいらっしゃいますか？
- データ収集/提供する際の障壁とは？
  - 解析の仕方が分からない？
    - MWS で作った解析ツールで得たデータ提供は可能か？
- どのようなデータセットがあると良い？
  - 似たようなデータセットがあっても別に良い
    - e.g., MWS Cup 2016 Datasets のスクリプトを実行/改良してみる、 MWSセッション内で結果比較してみる等
  - インジケータも貴重な情報でしょう
    - e.g., IOC, マルウェアハッシュ, コードシグネチャ

**続きはBoFで**

# その他の公開データ



- Malware Traffic Analysis: 悪性 pcap ファイル
  - http://malware-traffic-analysis.net/
- Malware domain list: 悪性 URL
  - http://www.malwaredomainlist.com/
- hphosts: 悪性ドメイン名
  - http://www.hosts-file.net/

# 研究動向



- Webサイト巡回
  - [NDSS'16] Cache, Trigger, Impersonate: Enabling Context Sensitive Honeyclient Analysis On-the-Wire
- リダイレクト分析
  - [Security'15] WebWitness: Investigating, Categorizing, and Mitigating Malware Download Paths
  - [SecureComm'16] Website Forensic Investigation to Identify Evidence and Impact of Compromise
- Exploit kit / JavaScript 解析
  - [NDSS'15] EKHunter: A Counter-Offensive Toolkit for Exploit Kit Infiltration
  - [CODASPY'16] Detecting Malicious Exploit Kits using Tree-based Similarity Searches
  - [DSN'16] Kizzle: A Signature Compiler for Detecting Exploit Kits
  - [WWW'17] J-Force：Forced Execution on JavaScript

# おわりに



- 「検知観測技術 > データ収集 > 実態調査」の研究開発サイクルを加速し、サイバー攻撃に対抗

MWS では様々なデータセットが提供されているが、さらなる充実化のために **MWS 全参加組織の協力**の下、**データセット作成**と**論文投稿**のご協力をいただきたい