Innovative R&D by NTT

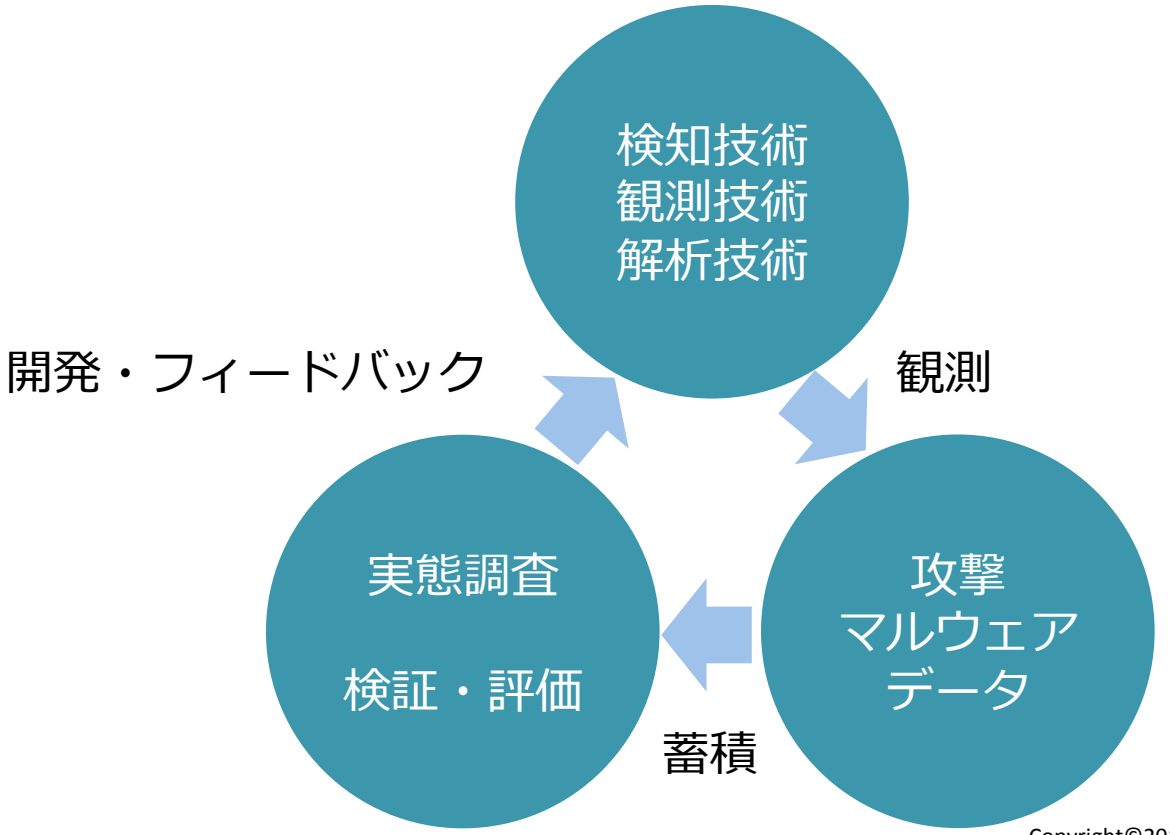
MWS 2016 意見交換会
D3M (**D**rive-by **D**ownload **D**ata by **M**arionette) データセット説明

NTTセキュアプラットフォーム研究所
秋山 満昭、高田 雄太

2016年05月30日

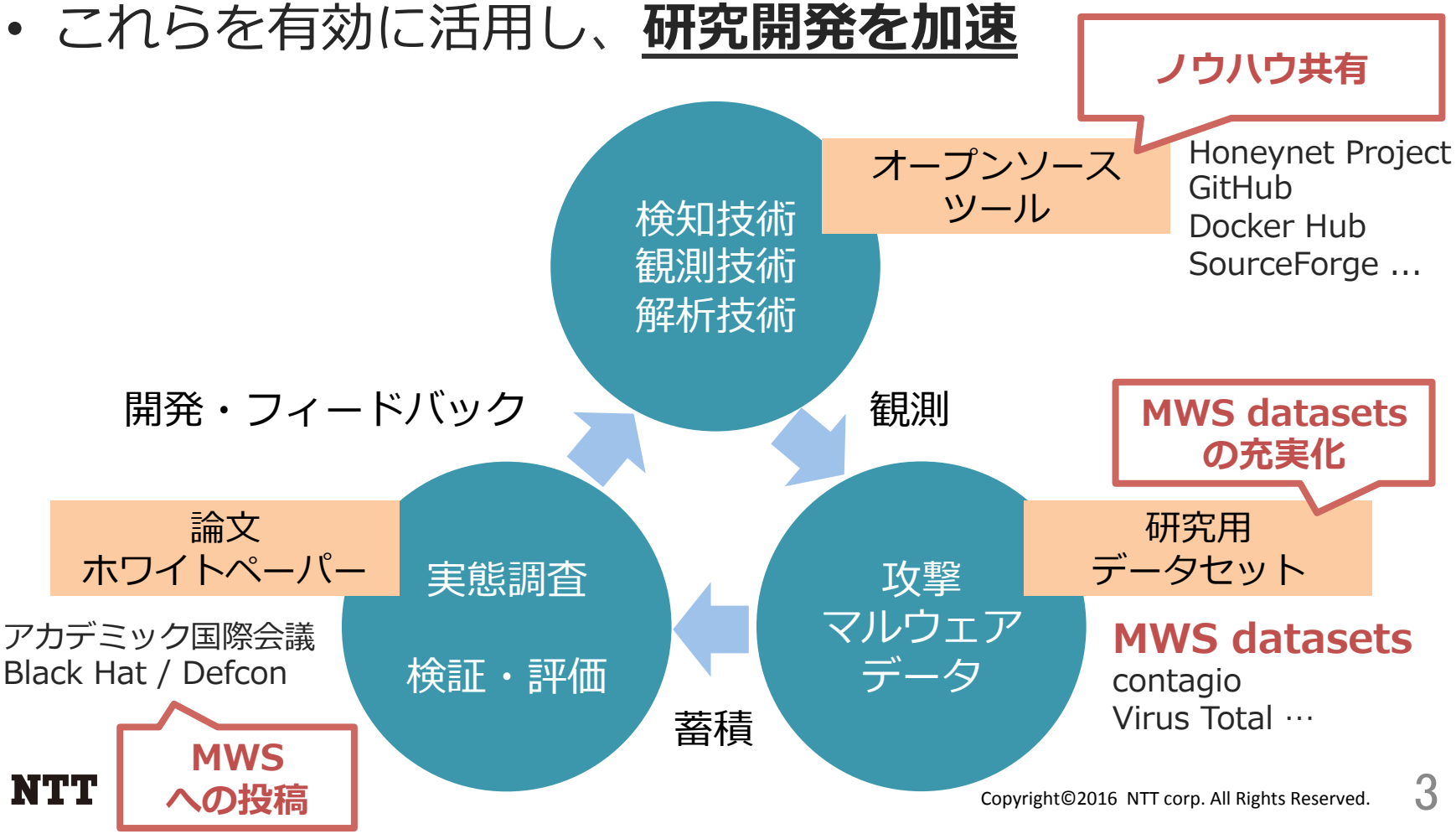
はじめに：データセット提供の意義

- 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？**



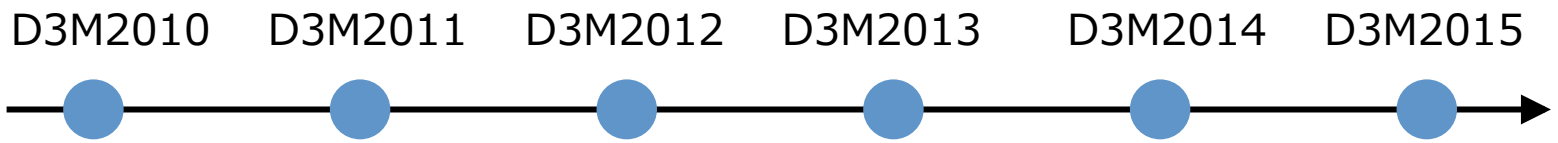
研究開発サイクルを加速させるために

- 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

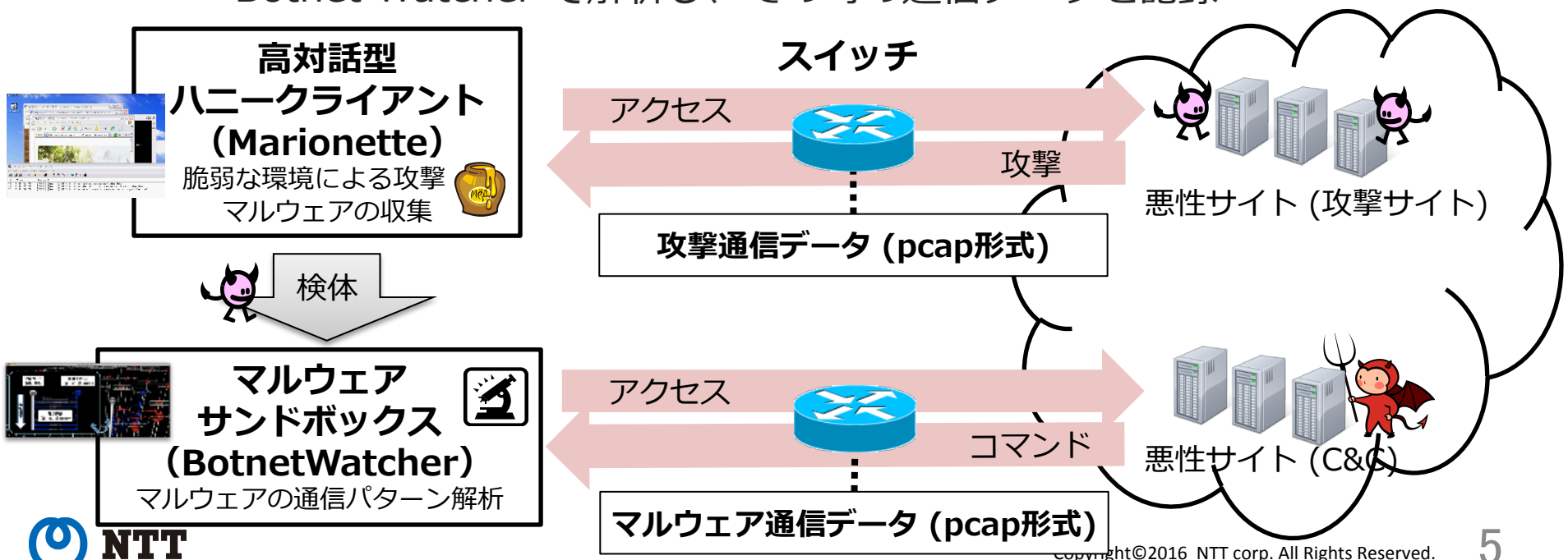
- データセットの内容
  - 攻撃通信データ
    - 悪性 URL を巡回した際に得られたドライブバイダウンロード攻撃の通信データ
  - マルウェア情報
    - ドライブバイダウンロード攻撃によってホスト上にダウンロードされた実行形式のファイル情報など
  - マルウェア通信データ
    - 実行形式のファイルを取得して24時間以内にマルウェアサンドボックス上で実行した際の通信データ
    - マルウェアサンドボックスはインターネットに接続可能（攻撃通信は遮断）



D3Mには これまで提供されたデータセット全て 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

D3Mのデータ量

D3M	URL数	pcap サイズ
2010	554	130 MB
2011	283	33 MB
2012	158	29 MB
2013	46	14 MB
2014	56	11 MB
2015	299	300 MB

D3M提供形態の変更

- 今までのデータセット（D3M2010-2015）
  - 目的：攻撃手法の理解，基本的な研究の着手
  - 利用形態：今まで通りMWSと契約して入手

- 最新のデータセット
  - 目的：難関国際会議などを目的にした発展的な研究
  - 利用形態：MWSと契約した上でさらに、
    ①提供元組織(NTT)へのインターン
    ②提供元組織(NTT)との共同研究

まずは今までのデータセットを用いた研究
をお勧めします

他のデータセット

- Malware Traffic Analysis
  - http://www.malware-traffic-analysis.net/
  - マルウェアやExploit-Kitの通信データ（pcap）、IDSログなどを配布
  - データセットの解説（図解付き）
  - データセットの解析チュートリアル

最近のExploit-Kitの手法や手動解析のコツを勉強できる。
実際の攻撃通信なので、検知手法の評価にも利用できるはず。

研究動向 ('14～)

- 巡回URLリスト生成/Web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
- Exploit kit / JavaScript 解析
  - [CODASPY'14] Webwinnow: Leveraging Exploit Kit Workflows to Detect Malicious URLs
  - [DIMVA'14] PExy: The Other Side of Exploit Kits
  - [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

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