NTT Secure Platform Laboratories
NTT セキュアプラットフォーム研究所

MWS2013意見交換会

D3M (D*rive-by* D*ownload* D*ata by* M*arionette*) 2013

秋山満昭

ネットワークセキュリティプロジェクト

研究コミュニティとデータセット

攻撃コード、マルウェア等

- iseclab （ http://www.iseclab.org/ ）
  - **コミュニティ内でのデータセット共有**（EURECOM, UCSB, Ruhr-universitat bochem **等が参加**）
- Honeynet Project （ http://www.honeynet.org/ ）
  - **コミュニティ内でのデータセット共有**

Wepawet (http://wepawet.iseclab.org/)
Anubis (http://anubis.iseclab.org/)

悪性URLリスト

- Malware domain list (http://www.malwaredomainlist.com/)
  - URL**を公開**
- Google safe browsing
  - URL**のハッシュ値のみ公開,**（NDA**により実際の**URL**共有？**）
- StopBadware （https://www.stopbadware.org/ ）
  - **一部の**URL**を公開，　パートナーシップ契約（有料）により詳細情報を共有**

・各研究コミュニティにおいて、データセット共有により研究開発が促進
・ただし、データセットを共有してもらうまでの道のりは長い、、、
・MWS**は日本人であればコミュニティへの参加が比較的容易**

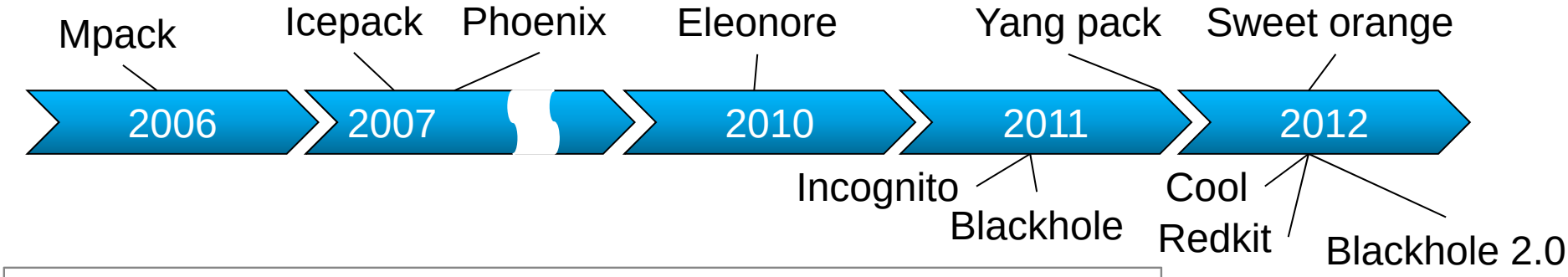
進化する攻撃手法

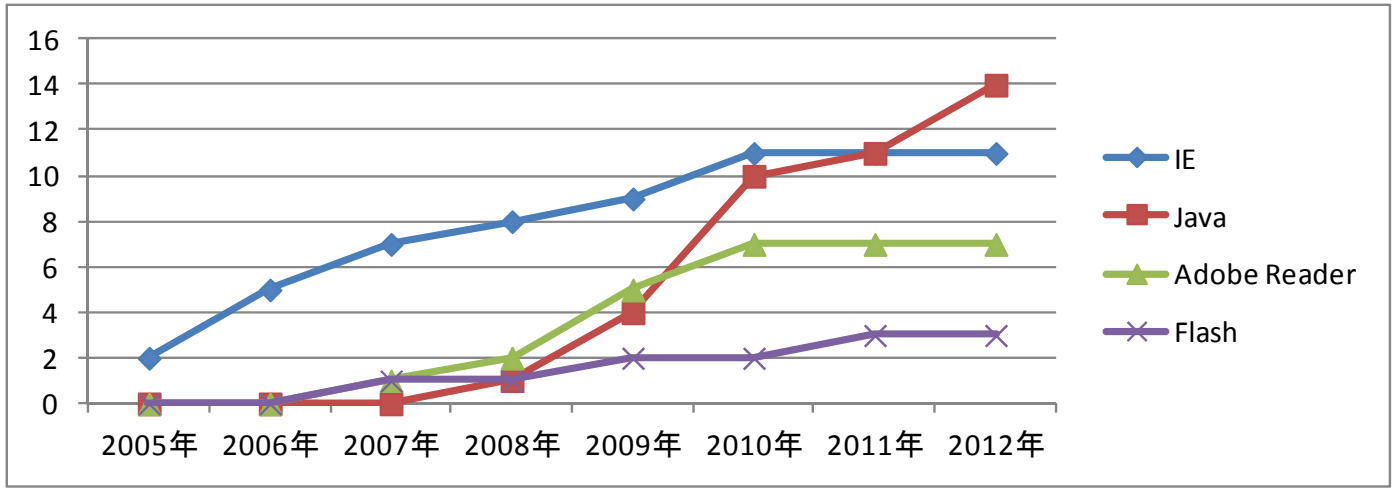
- マルウェア感染経路の変化
  - ドライブバイダウンロード攻撃（Webブラウザの脆弱性に対する攻撃）が主流
- 脆弱性の多様化
  - Webブラウザ（IE 6/7/8/9，FireFox，Opera）、
    プラグインアプリケーション（Acrobat 8/9，Flash 9/10/11，Java 6/7，．．．）
- 難読化手法の高度化による検知・解析妨害
  - HTML難読化，JavaScript難読化，PDF難読化，Java難読化
- 悪性サイトのクローキング
  - 自動転送（HTTPリダイレクト，iframeリダイレクト，JavaScriptリダイレクト，
    外部スクリプト読込，Traffic Direction System (TDS))
  - クライアントブラックリスト化によるアクセス拒否

・ 高度化・多様化するドライブバイダウンロード攻撃は、
データセットの収集自体も困難な状況になりつつある。
・ D3Mでは一連の攻撃通信およびマルウェアの通信が記録されており、
また、多様な攻撃手法やマルウェアが含まれている。

Exploit kit

- ドライブバイダウンロード攻撃を行う悪性サイトを構築するためのツールキット
  - 攻撃コード、難読化、クライアントブラックリスト化、リダイレクトコード生成などの一連の攻撃をコントロールパネルで分かりやすく支援
  - 数十種類のExploit kitがアンダーグラウンド市場で売買
  - 攻撃コードや難読化手法は定期的にアップデート

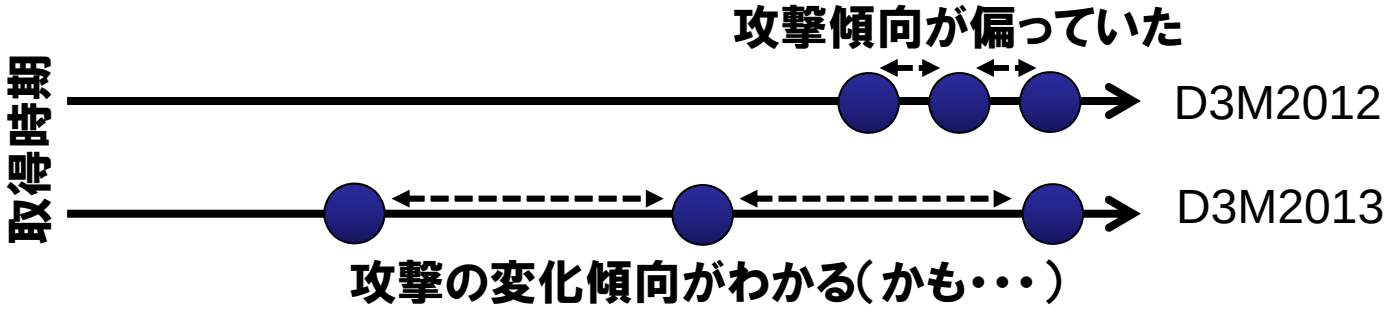




Exploit kitが標的とする主な脆弱性の累積

D3M 概要

- データセット内容
  - 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
    - 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
  - マルウェア通信データ
    - 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）
- 取得時期
  - 期間を空けて合計3回分提供する予定

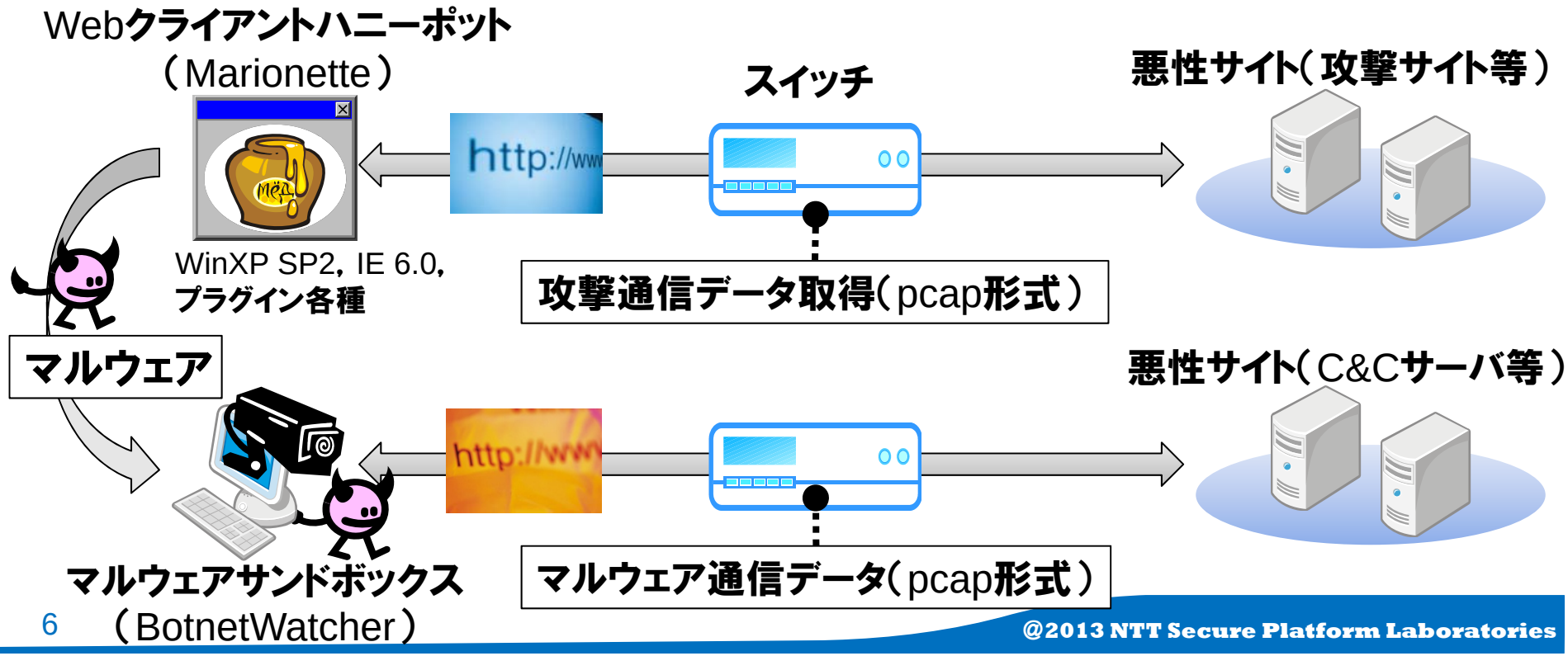


- *D3M2013* には *D3M2012* , *D3M2011*, *D3M2010* が同梱されています

D3M 取得環境

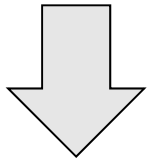
- ドライブバイダウンロード攻撃に関わるURLをブラウザに入力し、自動的に発生する一連のWeb通信、および感染するマルウェアの通信を記録
- 取得手順
  - 1. 公開ブラックリスト(※)をWebクライアントハニーポットで巡回
  - 2. 検知したURLを直ちに再巡回し、その際の通信データを記録
  - 3.　2で取得したマルウェア検体をマルウェアサンドボックスで解析し、その際の通信データを記録

(※) malwaredomainlist.com



D3M に含まれる情報

提供されるデータの形式：
・pcap(ドライブバイダウンロード通信，マルウェアの通信
・バイナリ(マルウェア検体）



- **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，・・・）
- **マルウェア検体**
- **マルウェアの通信**

どのような観点から研究を行えばよいか？

ここ2～3年のトップカンファレンスを調査すれば 世の中の技術動向が把握できる

- **学術系のトップレベルカンファレンス**
  - –ACM CCS
  - –IEEE S&P
  - –USENIX Security
  - –NDSS
  - –RAID
  - –ACSAC
  - –etc.

- **産業系のカンファレンス**
  - –Blackhat, CanSecWest, Recon, etc.

ドライブバイダウンロード攻撃対策研究のフェーズ，関係性，動向

- 攻撃検知/解析技術
  - 多様化する脆弱性、難読化技術、Webサイトクローキング等、攻撃コードの多フォーマット化(html, js, pdf, jar, swf, ・・・)の対策が研究されている
- クローリング技術
  - 悪性サイトの特徴に基づいた効率的なWeb空間探索技術が研究されている
- 悪性サイト情報／マルウェア収集技術
  - Webクライアントハニーポットの研究や、それを用いた大規模な実態調査が行われている
- マルウェア解析技術
  - ドライブバイダウンロード検体に限らず、広く研究されている
  - BHO化やMITBなどブラウザに寄生するマルウェアの解析が行われている

攻撃検知/解析技術 (1/2)

- JavaScript**解析**
  - JSUnpack *（http://jsunpack.jeek.org）*
  - JSAND *[WWW 2010] (UCSB, Wepawet)*

  - **抽象構文解析木による不正な**JavaScript**の特徴点抽出手法の提案**
    *[MWS2011]（**セキュアブレイン　神薗ら**）（MWS2011**優秀論文賞**）*
  - **難読化されたスクリプトにおける特徴的な構文構造のサブツリーマッチングによる同定**
    *[MWS2011]（**奈良先端大　Gregoryら**）*
  - **抽象構文木を用いた** Javascript **ファイルの分類に関する一検討**
    *[MWS2011]（**東大　宮本ら**）*
  - **抽象構文解析木の符号化による不正な**Javascript**の分類手法の提案**
    *[MWS2012]（**神大　上西ら**）*

  - ZOZZLE *[USENIX Security 2011] (Microsoft research)*
  - ROZZLE *[IEEE S&P 2012] (Microsoft research)*

攻撃検知/解析技術 (2/2)

- HeapSpray**検知**
  - NOZZLE *[USENIX Security 2007] (Microsoft research)*
  - Heap Inspector *[Blackhat USA 2011]*
- Flash**解析**
  - Analyzing and detecting malicious flash advertisements *[ACSAC2009](UCSB)*
  - FlashDetect *[RAID2012](UCSB)*
- PDF**解析**
  - **動的解析を利用した難読化**JavaScript**コード解析システムの実装と評価** *[MWS2010]（**セキュアブレイン　神薗ら**）（MWS2010**優秀論文賞**）*
  - Detection of Malicious PDF Files Based on Hierarchical Document Structure *[NDSS2013]*
- Jar**解析**
  - Jarhead *[ACSAC2012](UCSB)*

リダイレクト解析

- **検知を目指した不正リダイレクトの分析**
  *[MWS2010]（**富士通研究所　寺田ら**）*

- **パスシーケンスに基づく**Drive-by-Download **攻撃の分類**
  *[MWS2010]（**東海大　桑原ら**）*

- Analysis of Redirection Caused by Web-based Malware
  *[APAN Network Research Workshop2011]（**早大　高田ら**）*

- **通信可視化と動的解析の連携による攻撃解析支援**
  *[MWS2012]（**名工大　義則ら**）*

クローリング方法

- WebCop *[USENIX LEET 2010] (Microsoft research)*
  - **マルウェア配布**URL**のリンク・被リンクを辿ることで、マルウェア配布に関わる悪性な入口**URL**を発見**
- Structural Neighborhood URL Lookup *[SAINT2011]*
  - **既知の悪性**URL**の構造的な近隣を中心に検査することで効率的に未知の悪性**URL**を発見する方法**
- PoisonAmplifier *[RAID2012]*
  - SEO**を行う既知の悪性**URL**に対して、**SEO**特有の文字列を抽出してキーワード検索を行うことで、**SEO**を行う未知の悪性**URL**を発見する方法**
- EVILSEED *[IEEE S&P2012]*
  - **ハイパーリンク、**URL**構造、**SEO**、ドメイン登録情報、**DNS**クエリ情報などを使って未知の悪性**URL**を発見する方法**
  - **上記3論文の手法の合わせ技**

悪性サイト情報／マルウェア収集技術

- Webクライアントハニーポット
  - 高対話型
    - HoneyMonkey *[NDSS2006] (Microsoft research)*
    - Argos / Shelia *(VU Amsterdam)*
    - Capture-HPC *(Honeynet project)*
    - BLADE *[ACM CCS2010] (Georgia Tech)*
    - Marionette *[SAINT2012] (NTT SC研)*
  - 低対話型
    - HoneyC *(Honeynet project)*
    - PhoneyC *[USENIX LEET2009] (Honeynet project)*
    - Thug *(Honeynet project)*
  - ハイブリッド型
    - HoneySpider *(CERT Polska)*

その他の研究

- **ドメイン解析**
  - **ドメイン情報に着目した悪性Webサイトの活動傾向調査と関連性分析** *[MWS2010]（**九大　福島ら**）*

- **総合的な悪性サイト検出**
  - Prophiler *[IEEE WWW 2011] (UCSB)*
  - ARROW *[IEEE WWW 2011] (Microsoft research)*

- **ブラウザ防御手法**
  - IceSheild *[RAID2011] (Ruhr-University Bochum)*

実態調査

- A Crawler-based Study of Spyware in the Web
  [NDSS2006]

- Know Your Enemy: Malicious Web Servers
  (http://www.honeynet.org/papers/mws/, 2007) (Honeynet project)

- All Your iFRAMEs Point to Us
  [USENIX Security 2007] (Google)

- Manufacturing Compromise: The Emergence of Exploit-as-a-Service
  [ACM CCS2012]

セキュリティにおける理想的な研究開発サイクル
検知手法 ＞ 実態調査 ＞ 検知手法改良 ＞ 実態調査 ＞ ・・・

まとめ

- 一般的にセキュリティ研究のためのデータセットを収集すること自体難しいが、MWSではさまざまなデータセットがすでに提供されている

- ドライブバイダウンロード攻撃対策研究
  - “攻撃検知・解析技術”、“クローリング技術”、“ハニーポット技術”、“マルウェア解析”などの分野がある

- 観測手法と観測結果に基づく手法改良のサイクルを継続的に回すことが重要