Vol.2014-CSEC-66 No.19
Vol.2014-SPT-10 No.19
2014/7/3

マルウェア対策のための研究用データセット ～MWS Datasets 2014～

秋山 満昭[1,a)]　神薗 雅紀[2,3]　松木 隆宏[4]　畑田 充弘[5]

概要：マルウェアの脅威に対し様々な研究対策が盛んに行われているが，近年の脅威は攻撃の多様化や高度化により，研究を進める上で基礎となる”研究素材の収集と共有”が困難な状況が続いている．このような状況に対して，必要となる情報を収集して客観的な評価と研究成果の共有を容易にするためのデータセット（MWS Datasets 2014）を作成した．本稿では，MWS Datasets 2014 を構成する D3M 2014，FFRI Dataset 2014，NICTER Darknet Dataset 2014，および昨年度の MWS Datasets から引き続き提供される CCC DATAset および PRACTICE Dataset の概要を説明する．

キーワード：データセット, マルウェア, MWS Datasets 2014, D3M, FFRI Dataset, NICTER Darknet Dataset, CCC DATAset, PRACTICE Dataset

Datasets for Anti-Malware Research ～MWS Datasets 2014～

Abstract: Many security researches have continued to take countermeasures against malware threats. However, diversification and evolution of the recent attack make it increasingly difficult to collection and sharing of dataset for security research. For such a problem, anti-Malware engineering WorkShop (MWS) collected data related with malware threats and made the datasets (MWS Dataset 2014) for studies to evaluate the proposals and share the research achievements. In this paper, we introduce an overview of MWS Datasets 2014 comprised of D3M 2014, FFRI Dataset 2014, and NICTER Darknet Dataset 2014. CCC DATAset and PRACTICE Dataset are additionally contained in the datasets.

Keywords: Dataset, malware, MWS Datasets 2014, D3M, FFRI Dataset, NICTER Darknet Dataset, CCC DATAset, PRACTICE Dataset

1. はじめに

高度かつ複雑化したサイバー攻撃が世界的な問題となってり，各組織が個別に対策することはもちろんのこと，通信事業者，サービスプロバイダ，さらには国家レベルでの対策が急務となっている．特に，マルウェアに感染したホストは様々なサイバー攻撃を引き起こすことから，マルウェア対策やそこから派生する様々な研究が盛んに行われている．しかし，”共通の研究素材がないこと” および ”研究素材の収集の困難さ” が近年のマルウェア対策研究における推進の阻害要因としてある．

共通の研究素材とは，研究開発した技術の評価に用いるマルウェアや，マルウェア感染に関わる攻撃通信データ，マルウェア感染後の通信データなどのことを指し，可能な限り網羅的に，かつ攻撃の進化に合わせて適切に選択されたものが望ましい．従来では，研究素材となるこのようなデータは，主に研究者が収集環境を構築して自ら作成し，個々の技術の有効性や妥当性を評価してきた．このため，同じ研究テーマに取り組んだ場合でも，研究素材が異なる

[1] 日本電信電話株式会社，NTT セキュアプラットフォーム研究所
NTT Secure Platform Laboratories, Nippon Telegraph and Telephon Communication Corporation
[2] 株式会社セキュアブレイン，先端技術研究所
Advanced Research Laboratory, Securebrain Corporation
[3] 独立行政法人 情報通信研究機構
National Institute of Information and Communications Technology
[4] 株式会社 FFRI
FFRI, Inc.
[5] エヌ・ティ・ティ・コミュニケーションズ株式会社
NTT Communications Corporation
[a)] akiyama.mitsuaki@lab.ntt.co.jp

ために，その研究を相互に比較し適切に評価することが困難であった．

もう一つの阻害要因は，研究素材そのものが収集困難になってきていることである．攻撃者は検知回避手法や解析妨害手法を用いてサイバー攻撃やマルウェア感染を行い，またそれが年々高度化しているためである．例えば，ドライブバイダウンロード攻撃を行う Web サイトは様々な解析および検知を回避する機能を有しており，情報を収集する環境によって検査したとしても期待した情報を取得することができず，その結果として定性的にも定量的にも研究素材としての収集が難しくなっている．また，ボットの C&C サーバとの通信を収集する場合においても，近年の C&C サーバは短期間で活動を停止するため，期待した通信データを継続的に収集することが困難である．

高度かつ複雑化したサイバー攻撃に対峙していくため，われわれはマルウェア対策研究コミュニティを組織し，そのコミュニティ内で研究用データセットを共有することで研究を促進し，また研究成果を共有する場として「マルウェア対策研究人材育成ワークショップ（MWS）」を 2008 年から毎年開催してきた [1–6]．研究用データセットはコミュニティのメンバーによって任意に毎年更新され，多種多様なデータセットを過去から最新のデータまで併せて共有している．今年は以下のデータセットから構成される MWS Datasets 2014（図 1）を作成し，MWS2014 [7] を開催する．

(1) **D3M 2014**
Web 感染型マルウェアの観測データ
(2) **FFRI Dataset 2014**
マルウェアの動的解析データ
(3) **NICTER Darknet Dataset 2014**
ダークネットパケットデータ
(4) **CCC DATAset**[*1]
待受型ハニーポットで収集したデータ
(5) **PRACTICE Dataset**[*1]
マルウェアの長期観測データ

本稿では，2 節にて関連研究として他のデータセットや研究コミュニティを紹介し，3 節以降に MWS Datasets2014 の各データセットの概要を述べる．なお，CCC DATAset および PRACTICE Dataset に関しては文献 [8–11] ですでに説明されているため，本稿では説明を省略する．

2. 関連研究

2.1 研究用データセット

セキュリティに関連する研究用データセットのうち，非商用かつ比較的新しいものを以下に紹介する．

- CAIDA Data [12]
ネットワーク運用に関わる通信ログのデータセット
- MAWILab [13]
サンプリングで保存された通信リポジトリにラベル付けしたデータセット
- PREDICT Dataset [14]
ネットワーク運用およびセキュリティ装置に関わる通信ログのデータセット
- MALICIA Dataset [15]
ドライブバイダウンロード攻撃行う悪性 Web サイトから入手したマルウェア検体のデータセット
- Honeynet Project hpfeeds/hpfriends [16]
各種ハニーポット，サンドボックスの解析ログを集めたデータセット
- Contagio Malware Dump [17]
各種ファイルフォーマットの正規ファイルおよび悪性ファイル
- Android Malware Genome Project Dataset [18]
マルウェアファミリ毎に分類された Android マルウェア検体

これら以外にも研究用データセットは存在するが，データセット作成が 10 年以上前のものや，データセット提供を終了しているものが多い．2.2 – 2.4 節では，現状の研究用データセットにおける問題点を踏まえて議論する．

2.2 データセット入手の容易性

多くのデータセットにおいて，そのデータセット入手のためにはコミュニティへの加入が必要であり，加入の際に契約締結もしくは審査が行われる．政府がスポンサーとなっているコミュニティや地域性の高いコミュニティが多く，例えば PREDICT は米国の政府（国土安全保障省，DHS）や米国の大学が主体，iSecLab [19] は欧州の大学やセキュリティ研究所および企業が主体となっている．このようなコミュニティに対して，日本の学術機関や企業が単独で加入しデータセットを入手するためには，多大なコミュニケーションコストを必要とする．

一方，MWS は日本の学術機関や企業を中心とするため，MWS コミュニティへの参加は容易である．また，今後はコミュニティ間で連携を計ることにより，相互に研究用データセットの共有を行うことが MWS に求められる．

2.3 データセットの継続性

通信形態やプラットフォームの変化にともないサイバー攻撃やマルウェア感染手法は日々進化するため，研究用データセットには数年にわたる継続性が求められる．しかし，研究用データセットに継続性がない場合，つまりデータセットの更新がなく最新の傾向を反映できていない場合，研究用途としての活用は難しい（例えば，DARPA Intrusion Detection Data Sets [20] は 1998 年から 2000 年

[*1] 2014 年はデータセットの内容更新は無く，2013 年およびそれ以前のデータセットが提供される

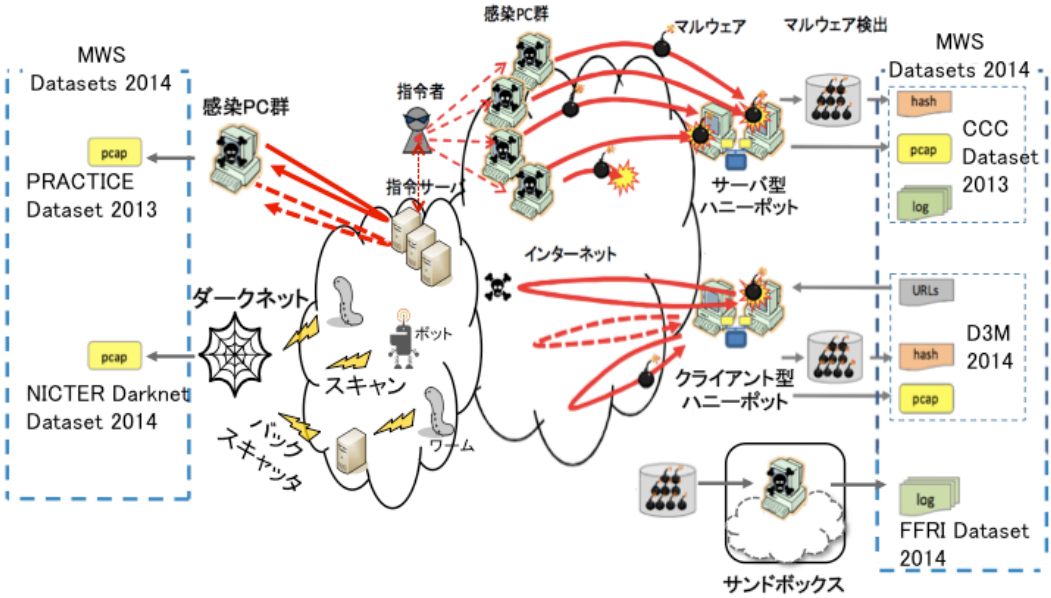

図 1 MWS Datasets 2014 の概要

に作成されたものである）．データセットの継続性を担保するためには，収集環境の整備とデータ作成者へのインセンティブが必要である．

MWS でも同様に，個々のデータセット提供者の収集環境に依存してデータセットの更新や共有の停止が発生することがあるため，コミュニティとしてデータセットの継続性を担保するための仕組みを検討および運用する必要がある．

2.4 データセットの網羅性

多種多様なサイバー攻撃に対して多角的かつ全域的な分析を実施するためには，データセットの種類および観測点の網羅性が求められる．CAIDA Data や PREDICT Dataset は様々な組織で収集した数十種類のデータセットを提供することでデータセットの種類と観測点の網羅性を向上させている．

MWS はマルウェア感染に着目して，感染前活動，感染時，感染後の各データセットを提供できている．観測点の網羅性については，さらにデータセット提供者やデータセット取得環境を増やすことで向上させたい．

3. D3M 2014

D3M（**D**rive-by-**D**ownload **D**ata by **M**arionette）2014 は NTT セキュアプラットフォーム研究所の高対話型の Web クライアント型ハニーポット（Marionette [21] [22]）で収集したドライブバイダウンロード攻撃に関連するデータである．ドライブバイダウンロード攻撃とは，近年マルウェアの主要な感染経路となっている Web ブラウザおよびそのプラグインの脆弱性を利用して制御を奪い，マルウェアを強制的にダウンロードおよびインストールさせる攻撃である．D3M にはドライブバイダウンロード攻撃に関する，攻撃通信データ，マルウェア検体，およびその通信データを収録した Web 感染型マルウェアの観測データ群である．Marionette は脆弱性に対する攻撃を受けるが，ダウンロードされたマルウェアの実行は許可しない．このため取得したマルウェアは動的解析システム（BotnetWatcher [23]）にて収集されてから 24 時間以内に解析される．

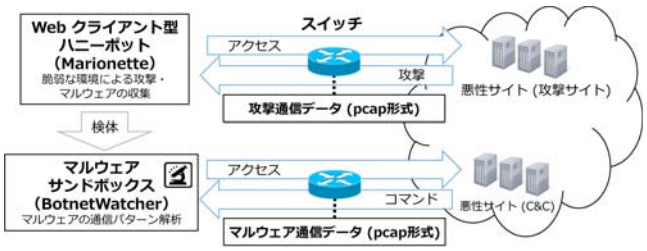

図 2 D3M の取得環境

このハニーポットで観測された一部のデータを D3M 2014 として提供する．D3M 2014 は感染手法の検知ならびに解析技術の研究のための”攻撃通信データ”，マルウェアの解析技術のための”マルウェア検体（ハッシュ値）”，および”マルウェア通信データ”から構成される．データセットの取得環境を図 2 に示す．

なお，過去のデータとの傾向を比較分析することができるよう，D3M 2010–2013 も提供している．以下，それぞれのデータについて概要を述べる．

3.1 攻撃通信データ

Web クライアント型ハニーポットの通信を tcpdump でパケットキャプチャした PCAP 形式のファイルである．ハニーポットの OS は Windows XP，Web ブラウザは Internet Explorer，プラグインは Adobe Reader，Flash Player，

WinZip，QuickTime，Java をあらかじめインストールしてあり，何れも脆弱性を含むバージョンでありセキュリティパッチは未適用である．ハニーポットはインターネット接続されており，パケットキャプチャは上流のネットワーク装置で行っている．データ収集日は 2013 年 4 月 12 日，2013 年 8 月 30 日，2014 年 4 月 10 日，および 2014 年 4 月 11 日であり，日毎に 1 ファイル，計 4 ファイルである．攻撃通信データは，あるブラックリストに登録されている URL を巡回し，攻撃を検知した URL を再度巡回した際の通信である．このため，攻撃コードが含まれる可能性が高く，かつ雑音（正常な Web サイトとの通信など）が少ない通信データになっている．巡回時に Web ブラウザに入力した URL を参考情報として付与している．なお入力 URL から派生してアクセスされる URL（リダイレクト，スクリプト読み込み，画像読み込み）は通信データに含まれるが，前述の URL リストには含まれない．

3.2 マルウェア検体

Web クライアント型ハニーポットで収集した Web 感染型マルウェアのハッシュ値をテキスト形式で記載したファイルである．3.1 節で収集した攻撃通信データに含まれる検体である．

3.3 マルウェア通信データ

3.2 節で収集した検体を 24 時間以内に動的解析システムで解析した際の通信のフルキャプチャデータである．動的解析システムはインターネットに接続した環境でマルウェアを 10 分間動作[*2]させており，ボットなどの遠隔制御されるマルウェアの動的解析が可能である．なお，外部ホストやネットワークに対する攻撃は動的解析システム内の仮想インターネット環境に転送することで，解析時の安全性を担保している．

4. FFRI Dataset 2014

FFRI Dataset 2014 は株式会社 FFRI が独自に収集した計 3,000 件のマルウェアを，動的解析することで得られたマルウェアの解析ログ群である．D3M は攻撃通信データやマルウェアの通信データ，およびマルウェアそのものをデータセットとしているが，FFRI Dataset はマルウェアの端末内での挙動に着目する．データセットの仕様について，以下に概要を述べる．

4.1 マルウェア

マルウェアはすべて PE (Portable Executable) 形式，かつ Windows プラットフォーム上で実行可能なファイルである．

[*2] 解析システムの都合上，2013 年 8 月 30 日の解析ログには解析時間が 10 分間のものと 30 分間のものが混在する．

2014 年 1 月から 2014 年 4 月の期間に Web クローリング等によって広く世界中から収集された比較的新しいマルウェアであり，2014 年 4 月時点で，少なくとも 10 社以上のアンチウィルス製品にてマルウェアと判定されていたものを選定した．収集された全数からランダムサンプリングを行っており，その内訳は収集時点におけるインターネットのマルウェアの感染トレンドを反映していると考えられる．当該マルウェア検体を利用した評価により研究成果の現実的な有効性を確認することを目的として選定されている．なお，データセットはこれらマルウェアの動的解析の結果であり，当該マルウェア自体は含まない．また，FFRI Dataset 2014 には FFRI Dataset 2013 の全データが含まれている．

4.2 動的解析

前述のマルウェアをオープンソースのマルウェア解析ツールである Cuckoo Sandbox [24] を用いて動的解析し，解析ログを生成している．Cuckoo Sandbox は，仮想化された Windows ゲスト内にマルウェアをコピー，実行，実行時挙動の記録，ゲスト環境の復元などの一連の解析動作を自動化するソフトウェアパッケージである．マルウェアの動的解析は，ネットワーク接続を有する専用のマルウェア解析環境上に Cuckoo Sandbox による解析システムを構築し，1 検体あたり 120 秒間実行した．ゲスト OS は Windows 7 である．また Cuckoo Sandbox は VirusTotal [25] と連携する機能を有しており，解析対象ファイルのハッシュ値に基づいて VirusTotal に問い合わせを行うことで，当該時点での各アンチウィルス製品での検知状況を取得することができる．本データセットの解析ログは，解析を実施した 2014 年 4 月時点での当該検出状況を含んでいる．表 1 に解析ログに含まれる具体的な項目の概要をまとめる．

また，同じ検体について FFR yarai analyzer Professional を用いて 60 秒間動的解析し，解析ログを生成した．FFR yarai analyzer Professional は FFRI のマルウェア自動解析ツールであり．プロセス毎の API 呼び出し履歴の取得機能，耐解析機能を持つマルウェアを解析する機能を有している．解析ログは，複数のログファイルから構成される．表 2 に FFR yarai analyzer Professional の解析ログの概要を示す．

5. NICTER Darknet Dataset 2014

NICTER Darknet Dataset 2014 は D3M 2014 と同様に攻撃通信データを提供する．他のデータセットと大きく異なる点は，ダークネットと呼ばれるインターネット上で到達可能かつ未使用の IP アドレス空間に届いた通信データという点である．

表 1　解析ログに含まれるデータ項目

項目	概要
info	解析の開始，終了時刻等
yara	yara [26] の有する標準ルールセットとの照合結果
signatures	ユーザ定義シグネチャとの照合結果（未使用）
virustotal	VirusTotal に登録されている各アンチウィルス検出結果
static	マルウェアファイルの静的情報（セクション構造，インポート API 等）
dropped	マルウェアが実行時に生成したファイルに関する情報
behavior	マルウェアが実行時に呼び出した API，引数，返り値等の情報
processtree	マルウェアが実行時に起動したプロセスの階層情報
summary	マルウェアが実行時にアクセスしたファイル，レジストリキー等の情報
target	解析対象となったマルウェアファイルの情報（ファイルサイズ，ハッシュ値等）
debug	動的解析時の Cuckoo Sandbox のデバッグログ
strings	マルウェアファイルに含まれる文字列情報
network	マルウェアが実行時に発生した通信情報

表 2　FFR yarai analyzer Professional 解析ログデータ項目

ログファイル	データ項目
analyzed_info.json	ファイルハッシュ（SHA1），ファイル名，ファイルサイズ
filetrace.log.json	ファイルアクセスイベント情報
regtrace.log.json	レジストリアクセスイベント情報
network.log.json	ネットワークアクセスイベント情報
api_logs/*.json	プロセス毎の API ログ（PID，TID，API 名，引数，返り値等）

5.1　ダークネット

ダークネットとは，インターネット上で到達可能かつ未使用の IP アドレス空間のことを指す．独立行政法人情報通信研究機構では，インターネット上におけるセキュリティインシデントの迅速な状況把握や原因究明および対策導出を目的としたインシデント分析センタ NICTER（Network Incident analysis Center for Tractical Emergencey Response） [27–29] の研究開発を推進しており，約 21 万 IP アドレスのダークネットに届くパケットを常時観測および分析している．

ダークネットに届くパケットの多くはネットワークを経由して感染を広げるタイプのマルウェアによるスキャンや，マルウェア自身がペイロードに含まれている UDP パケット，マルウェア同士が P2P ネットワークを確立するためのランデブー用のパケット，送信元 IP アドレスを詐称した DDoS 攻撃を受けているサーバからの応答（SYN-ACK）であるバックスキャッタなど，何らかの不正な活動に起因している．そのため，ダークネットに届くパケットを分析することで，インターネット上で発生している不正な活動の傾向把握が可能になる．2013 年は Spamhaus Project に対する攻撃をはじめ多くの DDoS 攻撃が発生し，特に，これらの DDoS 攻撃においては DNS や NTP を利用した DRDoS（Distributed Reflection Denial of Service）攻撃が多く利用されたことが特徴的であった．攻撃の詳細については割愛するが，これらの攻撃を行うためには攻撃に利用可能な DNS サーバや NTP サーバをインターネット上で探索する必要があるため，そのスキャン活動がダークネットでも観測されている．

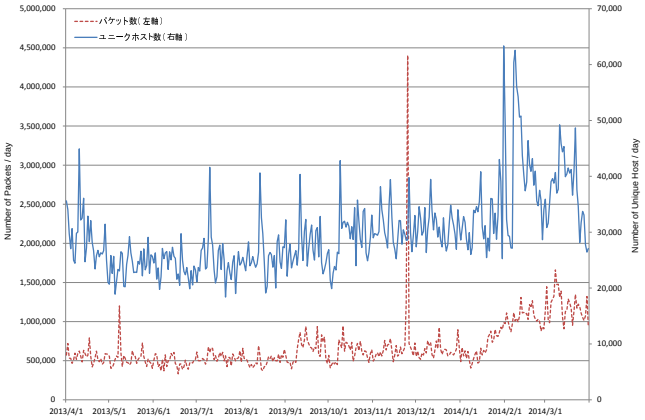

図 3　ダークネットで観測したパケット数およびユニークホスト数

5.2　ダークネット通信データ

前述のダークネット環境で観測されたトラフィックデータの一部を NICTER Darknet Dataset 2014 として提供する．観測環境のダークネットへ届いたパケットに対しては応答は行っていないため，本データセットには外部からダークネットに対するパケットしか含まれていない．また，観測地点を秘匿する目的で，データセットの宛先 IP アドレスの第 1 および第 2 オクテットは適当な値に置換している．

観測期間については 2011 年 4 月 1 日から 2014 年 3 月 31 日を基本とし，2014 年 4 月 1 日以降のトラフィックデータに関しても順次提供を行っている．参考までに図 3 に提供データセットにおける 2013 年 4 月 1 日から 2014 年 3 月 31 日までの日毎のパケット数とユニークホスト数（攻撃元ホスト数）の推移を示す．図 3 を見ると平均的に横ばい状態ではあるが，2014 年以降にパケット数，ユニークホスト数ともに増加傾向を示している．なお，2013 年 11 月 25 日にパケット数に急激なピークが現れているのは，あるブラジルの 1 ホストから 1 つの宛先 IP アドレスの 18991 ポートに対して大量の UDP パケットが送信されているからであるが，その原因は不明である．

5.3　NONSTOP

NICTER Datknet Dataset の提供には，NICTER で開発した NONSTOP（**N**ICTER **O**pen **N**etwork **S**ecurity **T**est-

Out **P**latform）[30] を活用する．NONSTOP は各種サイバーセキュリティ情報（ダークネットトラフィック，マルウェア検体，スパムメール，マルウェア解析結果など）を遠隔から安全に利活用するためのプラットフォームであり，いわゆる PaaS（Platform as a Service）の形態として開発が進められている．

利用を希望するユーザは，SSH クライアントとあらかじめ発行された認証用 IC カードを利用して NONSTOP へのアクセスを行い，研究内容に応じて提供される仮想マシン内で必要なサイバーセキュリティ情報にアクセスし分析を行うことになる．そのため，分析用に独自開発したツール等はローカルから仮想マシン内へファイル転送することで仮想マシン内での実行が可能である．また NONSTOP 内にリポジトリを用意することで，必要な各種ライブラリ等についてインストール可能としている．一方，提供したサイバーセキュリティ情報のうち外部への転送を禁止している情報の流出を防ぐ目的で，仮想マシンからローカルへのファイル転送に関しては，複数のフィルタ機構による検査，転送ファイルの一定期間の保存などが行われている．

6. MWS Datasets 利用状況

MWS Datasets を利用し，研究成果を共有する場として，”マルウェア対策研究人材育成ワークショップ (MWS)” を 2008 年から毎年開催しており，多くの研究成果が発表されている．過去の MWS Datasets と，MWS で発表された研究における利用内訳を表 3 に示す．CCC DATAset は従来のネットワーク感染型マルウェアのデータセットであり，それを利用した研究は年々減少している．一方で，Web 感染型マルウェアを含むデータセットである D3M や FFRI Dataset，昨年から急増している DRDoS 攻撃を含む NICTER Darknet Dataset などを用いた研究が増加している．実際のサイバー空間における攻撃やマルウェアのトレンドの変化に伴い，研究対象も徐々に変化していることが定量的にわかる．MWS としては，このような攻撃手法やマルウェアのトレンドの変化をカバーできるデータセットを継続的に提供し続ける必要がある．

なお，MWS Datasets を利用した研究発表は MWS だけに留まらず，多数の国際会議や論文誌等への掲載を確認している．

7. まとめ

最新のサイバー攻撃に対応可能な研究人材を育成するためのコミュニティである MWS は，マルウェア対策研究に必須となる研究用データセットを継続的に作成および共有している．本稿では最新のデータセットである MWS Datasets 2014 についてその概要を述べた．これら研究用データセット自体が研究者間で共通言語として役割を担うことや，研究用データセットを用いて研究開発した技術等の共有により，人材育成を含む本研究分野の発展に寄与することが期待できる．今後は，最新の脅威を見据えた研究用データセットの拡充ならびにデータセットの利用環境の構築および提供など，包括的なフレームワークを検討するとともに，評価用として利用可能なよりよい研究用標準データの作成に向け検討していきたい．

表 3 MWS Datasets を用いた MWS での論文発表数（MWS2008 – 2013 まで）

MWS Datasets	’08	’09	’10	’11	’12	’13
CCC (マルウェア検体)	5	7	6	5	7	3
CCC (攻撃通信データ)	9	14	5	6	2	0
CCC (攻撃元データ)	8	6	5	4	0	0
MARS[*3]	–	–	1	1	0	–
D3M	–	–	4	3	3	9
IIJ MITF[*3]	–	–	–	1	–	–
FFRI	–	–	–	–	–	5
PRACTICE	–	–	–	–	–	3
NICTER Darknet	–	–	–	–	–	6
データセット説明	0 [*4]	1	1	1	0 [*4]	1
合計	22	28	22	20	13	25
() 内は学生発表の件数	(8)	(15)	(10)	(9)	(9)	(10)

一部，複数のデータセットを利用した論文あり．”–” は提供なし．

[*3] 2014 年現在は配布していない．
[*4] MWS ホームページ等でデータセットの概要を説明している．

参考文献

[1] マルウェア対策研究人材育成ワークショップ 2008 (MWS2008) http://www.iwsec.org/mws/2008/
[2] マルウェア対策研究人材育成ワークショップ 2009 (MWS2009) http://www.iwsec.org/mws/2009/
[3] マルウェア対策研究人材育成ワークショップ 2010 (MWS2010) http://www.iwsec.org/mws/2010/
[4] マルウェア対策研究人材育成ワークショップ 2011 (MWS2011) http://www.iwsec.org/mws/2011/
[5] マルウェア対策研究人材育成ワークショップ 2012 (MWS2012) http://www.iwsec.org/mws/2012/
[6] マルウェア対策研究人材育成ワークショップ 2013 (MWS2013) http://www.iwsec.org/mws/2013/
[7] マルウェア対策研究人材育成ワークショップ 2014 (NWS2014) http://www.iwsec.org/mws/2014/
[8] 畑田 充弘, 中津留 勇, 寺田 真敏, 篠田 陽一: マルウェア対策のための研究用データセットとワークショップを通じた研究成果の共有, CSS2009(MWS2009) (2009.10)
[9] 畑田 充弘, 中津留 勇, 秋山 満昭, 三輪 信介: マルウェア対策のための研究用データセット 〜MWS 2010 Datasets〜, CSS2010(MWS2010) (2010.10)
[10] 畑田 充弘, 中津留 勇, 秋山 満昭: マルウェア対策のための研究用データセット 〜MWS 2011 Datasets〜, CSS2011(MWS2011) (2011.10)
[11] 神薗 雅紀, 畑田 充弘, 寺田 真敏, 秋山 満昭, 笠間 貴弘, 村上 純一: マルウェア対策のための研究用データセット 〜MWS datasets 2013〜, CSS2003(MWS2013) (2013.10)
[12] CAIDA Data - Overview of Datasets, Monitors, and Reports, http://www.caida.org/data/overview/
[13] MAWILab, http://www.fukuda-lab.org/mawilab/
[14] PREDICT Dataset Catalog,

https://www.predict.org/Default.aspx?tabid=104
[15] MALICIA Project, http://malicia-project.com/dataset.html
[16] hpfriends, http://hpfeeds.honeycloud.net/#/home
[17] Contagio Malware Dump, http://contagiodump.blogspot.jp
[18] Android Malware Genome Project, http://www.malgenomeproject.org
[19] International Secure Systems Lab, http://www.iseclab.org
[20] DARPA Intrusion Detection Data Sets, http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/
[21] Mitsuaki Akiyama, Kazufumi Aoki, Yuhei Kawakoya, Makoto Iwamura, and Mitsutaka Itoh: Design and Implementation of High Interaction Client Honeypot for Drive-by-download Attacks, IEICE Transaction on Communication, Vol.E93-B No.5 pp.1131-1139 (2010.05)
[22] Mitsuaki Akiyama, Yuhei Kawakoya, and Takeo Hariu: Scalable and Performance-Efficient Client Honeypot on High Interaction System, The 12th IEEE/IPSJ International Symposium on Application and the Internet (SAINT2012)
[23] Kazufumi Aoki, Takeshi Yagi, Makoto Iwamura, and Mitsutaka Itoh: Controlling malware HTTP communication in dynamic analysis system using search engine, The 3rd International Workshop on Cyberspace Safety and Security (CSS2011)
[24] Cuckoo Sandbox: Automated Malware Analysis, http://www.cuckoosandbox.org/
[25] VirusTotal - Free Online Virus, Malware and URL Scanner, https://www.virustotal.com/ja/
[26] yara-project - A malware identification and classification tool, https://code.google.com/p/yara-project/
[27] Koji Nakao, Katsunari Yoshioka, Daisuke Inoue, Masashi Eto: A Novel Concept of Netw ork Incident Analysis based on Multi-layer Observations of Malware Activities, The 2nd Joint Workshop on Information Security (JWIS2007)
[28] Daisuke Inoue, Masashi Eto, Katsunari Yoshioka, Shunsuke Baba, Kazuya Suzuki, Junji Nakazato, Kazuhiro Ohtaka, and Koji Nakao: nicter: An Incid ent Analysis System Toward Binding Network Monitoring with Malware Analysis, WOMBAT Workshop on In formation Security Threats Data Collection and Sharing (WISTDCS 2008)
[29] Koji Nakao, Daisuke Inoue, Masahi Eto, and Katsunari Yoshioka: Practical Correlation Anal ysis between Scan and Malware Prof iles against Zero-Day Attacks based on Darknet Monitoring, IEICE Trans. Information and Systems, Vol. E9 2-D, No.5, pp. 787-798, 2009
[30] 竹久達也，井上大介，衛藤将史，吉岡克成，笠間貴弘，中里純二，中尾康二: サイバーセキュリティ情報遠隔分析基盤 NONSTOP, 電子情報通信学会 情報通信システムセキュリティ研究会 (ICSS), pp. 85-90, 2013 年 6 月