Vol.2020-IFAT-139 No.8
2020/7/31

マルウェア対策のための研究用データセット MWS Datasets ～コミュニティへの貢献とその課題～

寺田真敏[1,2]　秋山満昭[3]　松木隆宏[4]　畑田充弘[5]　篠田陽一[6]

概要：マルウェア対策研究を行う上での課題の一つに研究者・実務者間で利用できる「共通の素材」を作ることが挙げられる．ここでの素材とは，提案手法の評価に用いるマルウェアのサンプルや，感染活動に関わる通信データなどの研究用データセットのことである．本稿では，研究者コミュニティへの貢献とその課題の事例のひとつとして，2008年から開催しているマルウェア対策研究人材育成ワークショップ，共通の素材を流通させるためのマルウェア対策のための研究用データセット，そして，研究用データセットの利活用ならびに作成にあたって意識する必要がでてきたサイバーセキュリティ研究における倫理的な研究プロセスについて報告する．

キーワード：データセット，マルウェア，研究倫理，MWS

MWS Datasets for Anti-Malware Research ~ Contribution to the community and its challenges ~

Masato Terada[1,2]　Mitsuaki Akiyama[3]　Takahiro Matsuki[4]
Mitsuhiro Hatada[5]　Yoichi Shinoda[6]

Abstract: One of the challenges in conducting anti-malware research is to generate the available common materials for sharing researchers and practitioners. The materials mean datasets for the evaluation of proposed methodologies, such as malware samples and infection activity data. This paper shows the overview of Anti-Malware engineering WorkShop since 2008, Dataset for the anti-malware research and the ethical research process in cyber security research for promoting the utilization and creation of the dataset.

Keywords: Dataset, Malware, Research Ethics, MWS

1. はじめに

コロナ渦において，運送系企業を装ったフィッシング，マスク購入に関する悪質なショッピングサイトへの誘導だけではなく，FBI からは，新型ウイルスのワクチン開発など COVID-19 への対応に取り組んでいる医療，製薬および研究部門に対するサイバー攻撃の可能性を示唆する注意喚起が出されている．このようなサイバー攻撃において，不正なプログラムの総称であるマルウェアは，PC に潜伏し不正活動の継続，ランサムウエアとして身代金要求，スマートフォンを介した盗撮，盗聴など多様な用途に使用されている．

マルウェア対策研究を行う上で多くの課題があり，その一つに「共通の素材を作ること」が挙げられる．ここでの素材とは，提案手法の評価に用いるマルウェアのサンプルや，感染活動に関わる通信データなどの研究用データセットのことである．素材となるこのような研究用データセットは，これまで研究者らが独自に収集し，それぞれの解析手法や対策手法の妥当性を検証するために利用してきた．このため，同じテーマに取り組む研究者同士であっても，研究成果を単純に比較することが難しい．また，新たに研究を始めようとしても，昨今のマルウェアに起因する感染事件の発生にともない所属組織のセキュリティポリシーの制約強化から「研究用データセットを収集すること自体が難しくなっていること」も課題となってきている．

このような課題を抱えている状況において，更なる進化を続けるマルウェアに対峙していくために，2008 年，研究用データセット(MWS Datasets)を研究者に提供し，研究成果を共有・切磋琢磨する場として「マルウェア対策研究人材育成ワークショップ(MWS)」を開催するに至った．

本稿では，コミュニティへの貢献とその課題の事例のひとつとして，2008 年から開催しているマルウェア対策研究人材育成ワークショップと共通の素材として提供しているマルウェア対策のための研究用データセット MWS

1) 東京電機大学, Tokyo Denki University
2) 株式会社日立製作所, Hitachi Ltd.
3) NTT セキュアプラットフォーム研究所, NTT Secure Platform laboratories
4) 株式会社エヌ・エフ・ラボラトリーズ, N.F.Laboratories Inc.
5) 日本電信電話株式会社, Nippon Telegraph and Telephone Corporation
6) 北陸先端科学技術大学院大学,
Japan Advanced Institute of Science and Technology

Datasets について報告する．

2. マルウェア対策研究人材育成ワークショップ

2.1 開催の目的

インターネットのサイバー攻撃の脅威と実態など全般が見えにくくなってきている．その背景のひとつに，活動を見えにくくするためのマルウェア自身の機能や運用の高度化が挙げられる．このような状況下で，情報システムでのセキュリティ事故や事件の迅速な対処や未然防止のためには，先端的な研究者だけではなく，企業のネットワーク技術ならびにセキュリティ技術を開発する実務者もマルウェアに関する専門知識を備えていく必要がある．そこで，本ワークショップでは，研究用データセットの提供，研究成果の共有ならびに切磋琢磨する環境の提供を通して，マルウェアに関する専門知識を備えた研究者や実務者を育成していくことを目的としている．

(1) 研究用データセットの提供

トラヒック分析技術やマルウェア分析技術を研究および評価するための適切な素材を準備し，研究者(学生，ネットワーク技術ならびにセキュリティ技術を開発する実務者)に提供することで，次の二点を実現する．

(2) 研究成果の共有

共通の研究用データセットを用いて行った研究成果を本ワークショップで発表し，研究者間で共有することで，より具体的な成果の水平展開を図り，我が国のセキュリティ研究人材育成につなげる．

(3) 切磋琢磨する環境の提供

共通の研究用データセットに基づく研究内容を共有することで，具体的なスキルアップ目標や，先進的な研究テーマの発見など，研究者の評価育成の場を形成する．

2.2 研究用データセットの提供から研究成果の共有の実現

MWS は図 1 に示すとおり「研究用データセットの提供」，「分析ならびに対策技術の研究」，「研究成果の共有」というマルウェア対策研究のサイクルを継続的に回すことで，マルウェア対策研究活動を推進してきた．具体的な活動として，本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いる[1].

さらなる研究発展のため，研究用データセットの作成そのものが研究対象分野として立ち上がり，より活発に研究サイクルが回るよう後押しする活動を展開していきたいと考えている．

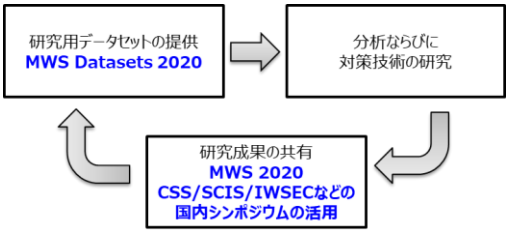

図 1：マルウェア対策研究のサイクル

2.2.1 研究用データセットの提供

マルウェア対策研究人材育成ワークショップに提供している研究用データセット(MWS Datasets)を表 1，図 3 に示す．研究用データセットは，実環境で発生している事象を収集した観測系データと，解析を目的とした環境下で収集した解析系データがある．

観測系データのひとつである BOS は，企業の情報システムを模擬するおとりシステムを用意し，組織内ネットワーク内に侵入した攻撃者の活動を観測したデータである．BOS2018 では，同じマルウェアを 2 つの異なる設定のおとりシステムで動かしたときに，来訪した攻撃者の挙動の類似点と相違点を観測データとして提供している(図 4，図 5)．図 4 はネットワークの共有資源を探査する活動，図 5 はネットワーク上の端末可達性の確認に関するものである．このような観測データから，攻撃者が実行するコマンド間隔に着目して，例えば攻撃者の挙動のシミュレーションにも活用することができる．

2.2.2 分析ならびに対策技術の研究

MWS から研究用データセットの提供を受けた組織は，毎年 50 組織前後であり，半数以上が学術系となっている(図 2)．学術系の新規は，学部生や院生がマルウェアに関する研究を進めるにあたり，素材を探している中で MWS にコンタクトをしてきたものである．

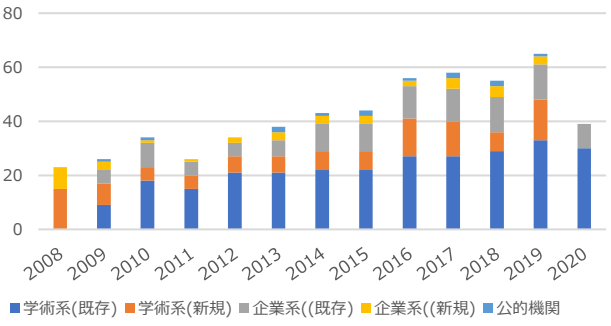

図 2：MWS Datasets の利用申請組織数の推移

表 1：MWS Datasets が提供する素材

区分	データセット名	概要	データ収集期間
観測系データ	CCC DATAset	ハニーポット(おとりシステム)で収集したマルウェア検体マルウェア検体(ハッシュ値)，攻撃通信データ，攻撃元データ	2008～2011
	D3M(Drive-by-Download Data by Marionette)	Web クライアントハニーポットで収集したマルウェア検体(ハッシュ値)，攻撃通信データ	2010～2015
	Augma Dataset	Web クライアントハニーポットで収集した攻撃通信データ	2020
	NICTER Dataset	ダークネットセンサーで収集した攻撃通信データ，リアルタイムダークネットセンサーの情報	2013～2020
	PRACTICE Dataset 2015	DRDoS(Distributed Reflection Denial of Service)攻撃の通信観測データ	2015
	BOS(Behavior Observable System)	攻撃者の行動を記録した研究用データセットで，マルウェア検体(ハッシュ値)，通信観測データ，プロセス観測データを含む	2014～2020
解析系データ	PRACTICE Dataset 2013	マルウェア動的解析による攻撃通信データ	2013
	FFRI Dataset	マルウェア自動解析によるマルウェアの挙動ログ	2013～2020
	Soliton Dataset	マルウェア解析環境で得られたマルウェアの挙動ログ	2018～2020
その他	NCD in MWS Cup 2014	MWS Cup 2014 会期中に収集したホワイトデータセット	2014
	MWS Cup 2015～2019 参加チームのスクリプト&スライド	MWS Cup 2015～2019 の参加チームによる発表スライドと課題を解くにあたって作成したスクリプト	2015～2019

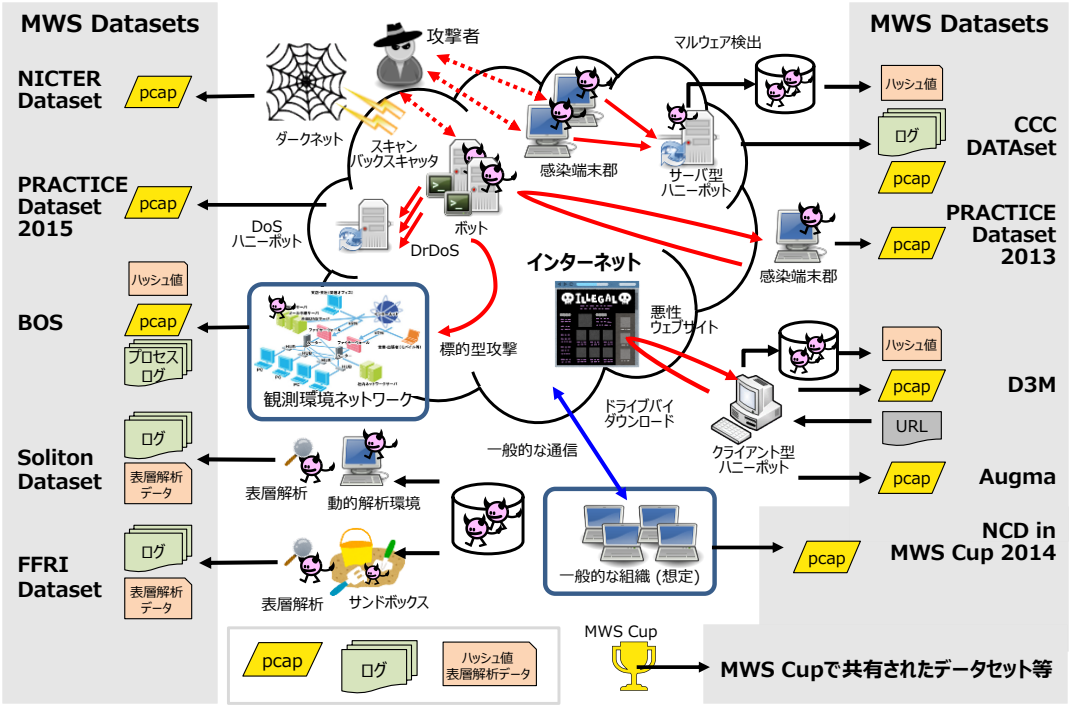

図 3：MWS Datasets を構成するデータ群

g14　総当たり的な探索

Date Time	Observable event
2018/01/19	
12:20	検体(.exe)を実行
12:26	*.*.102.145:443に接続
12:39	C:¥Windows¥SysWOW64¥cmd.exe
	ipconfig /all
12:45	net view
12:46	arp -a
	<中略>
2018/01/22	
09:15	*.*.102.145:443に接続
09:51	C:¥Windows¥SysWOW64¥cmd.exe
	ipconfig /all
	net view
	tracert www.yahoo.co.jp
10:18	ping -n 1 ActiveDirectory
10:50	net group /domain
10:53	net group "domain admins" /domain
10:56	net group "domain controllers" /domain
	net group "domain users" /domain
11:56	net view /domain
	net group /domain
	net user /domain
	net group "DnsUpdateProxy" /domain
	net groupコマンドによる探索繰り返し(10回以上)
	net user "1012000101" /domain
	net userコマンドによる探索繰り返し(100回以上)

g15　最小限の探索

Date Time	Observable event
2018/01/23	
18:42	検体(.exe)を実行
18:43	*.*.102.145:443に接続
	C:¥Windows¥SysWOW64¥cmd.exe
	ipconfig /all
18:45	net view
18:46	net group /domain
18:47	net group "Domain computers" /domain
19:05	net view /domain
2018/01/24	
12:21	*.*.102.145:443に接続
12:30	C:¥Windows¥SysWOW64¥cmd.exe
12:31	certutil -urlcache -split -f
	http://*.*.7.117:443/active.htm active.txt
12:32	*.*.7.117:443に接続
	<中略>
12:43	C:¥Temp¥asus.exe 10.139.8.1-10.139.8.255
	21,22,23,53,139,445,443,80,3389,3128,8080
	asus.exeコマンドによる探索繰り返し(10回以上)
	<中略>
13:02	powershell -exec bypass C:¥Temp¥profile.ps1
17:36	*.*.7.117:443に接続
	<省略>

図 4：来訪した攻撃者の異なる挙動 (BOS2018 データ)

g14　総当たり的な探索

Date Time	Observable event
2018/01/23	
10:11	*.*.102.145:443に接続
10:13	*.*.7.117:443に接続
10:26	C:¥IPtool¥asus.exe 10.16.117.2 ActiveDirectory:445 C:¥IPtool¥asus.exe 10.16.117.7:443 C:¥IPtool¥asus.exe 10.16.117.8:443 C:¥IPtool¥asus.exe 10.16.117.6:21 C:¥IPtool¥asus.exe 10.16.117.6:3389 C:¥IPtool¥asus.exe 10.16.117.10:445 C:¥IPtool¥asus.exe 10.16.117.11:3389 asus.exeコマンドによる探索繰り返し(150回以上) <中略>
17:04	*.*.102.145:443に接続 asus.exeコマンドによる探索繰り返し(80回以上)
17:35	C:¥Windows¥SysWOW64¥cmd.exe
17:39	C:¥IPtool¥asus.exe asus 10.32.1.1-10.32.1.255 21,22,23,53,139,445,443,80,3389,3128,8080
17:47	C:¥IPtool¥asus.exe asus 10.32.1.160 3128
17:48	C:¥IPtool¥asus.exe asus 192.168.12.1-192.168.12.255 21,22,23,53,139,445,443,80,3389,3128,8080

g15　総当たり的な探索

Date Time	Observable event
2018/01/24	
12:21	*.*.102.145:443に接続
12:30	C:¥Windows¥SysWOW64¥cmd.exe
12:31	certutil -urlcache -split -f http://*.*.7.117:443/active.htm active.txt
12:32	*.*.7.117:443に接続 <中略>
12:43	C:¥Temp¥asus.exe 10.139.8.1-10.139.8.255 21,22,23,53,139,445,443,80,3389,3128,8080 asus.exeコマンドによる探索繰り返し(10回以上)

図 5：来訪した攻撃者の類似した挙動 (BOS2018 データ)

表 2：MWS2008～MWS2019 における MWS Datasets を用いた論文の発表件数

	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19
CCC	22	27	16	15	9	3	3	3	3	4	1	2
MARS	-	-	1	1	0	0	-	-	-	-	-	-
D3M	-	-	4	3	3	9	14	9	2	7	3	1
IIJ MITF	-	-	-	1	-	-	-	-	-	-	-	-
FFRI	-	-	-	-	-	5	2	4	3	2	5	4
PRACTICE	-	-	-	-	-	3	1	0	1	0	0	0
NICTER	-	-	-	-	-	6	2	3	0		0	0
BOS	-	-	-	-	-	-	1	4	2	3	5	2
Soliton	-	-	-	-	-	-	-	-	-	-	1	0
NCD in MWS Cup 2014	-	-	-	-	-	-	-	0	0	3	0	1
(データセット概説)	0	1	1	1	0	1	0	0	0	0	0	0
合計	22	28	22	20	12	27	23	23	11	21	13	9
内，学生発表	8	15	10	9	9	10	10	14	8	12	9	6

注：一部の論文は複数の研究用データセットを利用している．
"ー"は研究用データセットの提供なし

2.2.3　研究成果の共有

2008 年からコンピュータセキュリティシンポジウム(CSS)と併催でマルウェア対策研究人材育成ワークショップ(MWS)を開催している．MWS では，研究用データセットの概説，解析データ解説も含めて，毎年 15～20 件近くの研究発表が行われ，約半数は学生による発表となっている(表 2)．MWS 以外への発表を含めて内容を概観すると，マルウェアの解析・検知・分類・防護に関する技術開発系の研究と，マルウェアの機能や活動を明らかにする動向調査系の研究がある(図 6)．

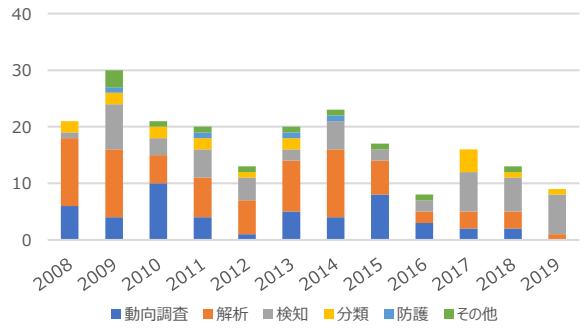

図 6：発表内容の分類(MWS 以外への発表も含む)

2.3 切磋琢磨する環境の提供の実現

2009 年から「切磋琢磨する環境」の場を用意するため，日頃の研究・運用で培った技術や開発ツールのレベルを競う MWS Cup を開始した．MWS Cup は，マルウェア対策に関する人材の育成と新たな技術やツールの発掘を目的とした実践的な競技会である(図 7)．MWS Cup の課題は事前課題と当日課題に分かれており，課題に対する正答率を競うテクニカルポイントと，課題取り組みを通して得らえたノウハウや作成したツールの有用性をアピールするプレゼンテーションポイントから構成している(表 3)．

表 3：MWS Cup 2019 の課題構成

区分	概要
テクニカルポイント	MWS 会期中に約 3 時間で取り組み，その場で解答を提出
プレゼンテーションポイント	MWS 会期前 2 ヶ月間で取り組み，MWS 会期中に，各参加チームが成果物のプレゼンテーションを実施

2017 年～2019 年の MWS Cup の参加者数と課題を表 4 に示す．なお，事前課題の採点は，下記指標に基づくプレゼンテーションの内容で評価し，当日課題の採点と合わせて最終的な順位を決定している．

- 新規性：従来になかったアイデア/データ/ツールであるか，新しい課題設定か等
- 実用性：第三者がすぐに活用できるアイデア/データ/ツールであるか等
- 有効性：幅広い第三者にとって有益なアイデア/データ/ツールであるか，現存する脅威やデータに効果的か等

表 4：MWS Cup の参加者数と課題

年	参加数	課題
2017	14 チーム 79 名	事前課題：自由課題 当日課題 1：DBD 攻撃解析 当日課題 2：マルウェア静的解析 当日課題 3：マルウェア動的解析
2018	16 チーム 86 名	事前課題：自由課題＋プレゼン 当日課題 1：DFIR(デジタルフォレンジックとインシデント対応) 当日課題 2：マルウェア静的解析 当日課題 3：マルウェア動的解析
2019	15 チーム 83 名	事前課題:MWS とって有意義なデータセットやツールを検討・作成せよ! 当日課題 1:悪性トラフィックデータ・クライアントログを分析せよ! 当日課題 2：マルウェア(RAT)を静的解析せよ! 当日課題 3:マルウェア表層解析ログを分析・分類せよ!

図 7：MWS Cup 2019 の当日課題の取組みの様子

3. サイバーセキュリティ研究における倫理的な研究プロセス

MWS の活動を進めていく過程で，研究用データセットの作成・利用において考慮すべき事項についても蓄積してきた．例えば，攻撃活動のデータを収集する際には，第三者に攻撃活動が波及しないようアクセス制御等の安全措置を講じるといったことが挙げられる．本章では，マルウェア対策研究に限らず，広くサイバーセキュリティ研究に視野を広げ，2016 年頃から取組み始めた課題であるサイバーセキュリティ研究における倫理的な研究プロセスについて述べる[2].

3.1 サイバーセキュリティ研究における研究倫理

ICT の進展にともない，誰も踏み入れたことがない，前例が十分にない倫理的領域を取り扱う機会が出てきた．特に，マルウェアを含めたサイバー攻撃の高度化および多様化のスピードは早く，研究者や実務者は最前線でサイバー攻撃の観測・分析・対策を実施する際に法制度や社会的責任の側面から準備をしておかなければ難しい判断を迫られるときがある．このため，ステークホルダ(利害関係者)の明確化，インパクトの見積もり，リスクの最小化努力，Responsible disclosure(研究成果の社会的な影響を考慮して，事前に必要な手続きを踏んだ後，研究成果を開示すること)を実施し，自身の研究を研究倫理的観点から実践して論じることの必要性が高まってきている．サイバーセキュリティ研究における研究倫理とは，自身の研究を研究倫理的観点から実践して論じることであり，そのために必要となる適切な手続きを倫理的な研究プロセスと呼んでいる(図 8).

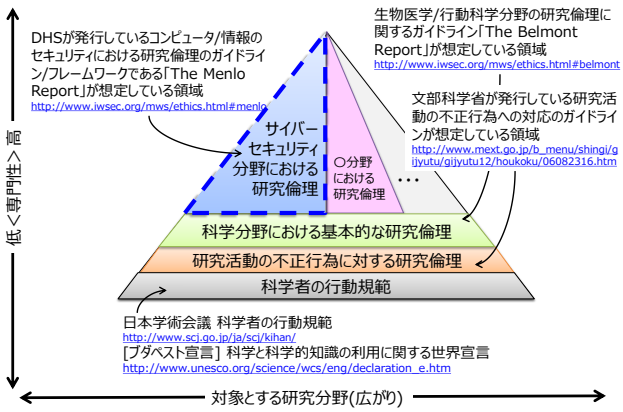

図 8：サイバーセキュリティ研究における
倫理的な研究プロセス

3.2 課題認識と対応方針

MWS 論文投稿における課題抽出と普及啓発の取り組みから(表 5)，サイバーセキュリティ研究における倫理的な研究プロセスの定着化にあたっては，各学会や研究会が自律的に倫理的な研究プロセスを実践できることが望ましいこと,そのためには投稿論文の書き方で解決できる問題と，研究そのもののアプローチの検証を通して解決できる問題とに取り組む必要があることがわかってきた．

- 投稿論文の書き方で解決できる問題
  論文投稿時に「チェックリスト」を用いて，研究者自身がセルフアセスメントできるようにする．
- 研究そのもののアプローチの検証を通して解決できる問題
  研究会やシンポジウムに研究倫理に関する「相談窓口」を設置し，技術や法律面から倫理的課題に対するアドバイスを行う．

相談窓口による対応は，2018 年のコンピュータセキュリティシンポジウム(CSS)から試行され，数件の投稿論文について倫理的課題の側面から相談窓口を通して研究者にアドバイスがなされている[3]．チェックリストによる対応は，2019 年にチェックリストが作成され，同年の CSS からセルフアセスメントを開始し，200 件を超える投稿論文においてチェックリストが活用されている[4]，

表 5：サイバーセキュリティ研究における
倫理的な研究プロセスに関する普及啓発

年	内容
2017	● SCIS 企画セッション(パネルディスカッション) 研究活動と Responsible disclosure ● MWS 企画セッション 研究倫理と Responsible Disclosure ● JSPS 第 192 委員会 公開シンポジウム(パネルディスカッション) サイバーセキュリティ研究における研究倫理
2018	● SCIS 企画セッション(パネルディスカッション) セキュリティに関する研究活動を進める上での倫理的課題 ● CSS 企画セッション(パネルディスカッション) サイバーセキュリティ研究のグレーゾーン
2019	● SCIS 企画セッション(パネルディスカッション) 製品セキュリティと研究活動 ～脆弱性発見者と対応者，それぞれの視点からの取組みを知ろう～ ● CSS 企画セッション(パネルディスカッション) 国内における脆弱性報告の調整活動 ～学術系が取り組むべき課題～

4. おわりに

切磋琢磨を通して，新たなサイバー攻撃に対応可能な研究人材の育成に寄与する MWS コミュニティは，マルウェア対策研究に必要となる研究用データセットを継続的に作成および提供し，その研究成果を共有するフレームワークを推進している．MWS はこの活動を 10 年以上に渡って続けることでコミュニティとしての持続的な発展を遂げた．

本稿では，学術コミュニティへの貢献とその課題の事例のひとつとして，2008 年から開催しているマルウェア対策研究人材育成ワークショップについて概要を述べた．本活動ならびに，研究用データセットが研究者間で共通言語としての役割を担うことや，本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，データセット作成そのものが研究対象分野として立ち上がり，研究活動をさらに発展させていくことが期待できる．

今後は，サイバーセキュリティ研究における倫理的な研究プロセスを含め，異なる分野間での活動を通して，研究成果を共有・切磋琢磨する場の発展を推進していく．

謝辞

本研究にあたって，有益な助言とデータセット作成の協力を頂いた研究者コミュニティ，ならびに総務省実証実験プロジェクトおよび CCC 運営連絡会の関係者各位に深く感謝いたします．

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",
https://www.iwsec.org/mws/
2)"サイバーセキュリティ研究における倫理的な研究プロセスについて"，http://www.iwsec.org/mws/ethics.html
3)"研究倫理相談窓口",
https://www.iwsec.org/css/2018/ethics.html
4)"サイバーセキュリティ研究における倫理的配慮のためのチェックリスト",
https://www.iwsec.org/css/2019/ethics_list.html