MWS2023ポストミーティング

D3M（Drive-by Download Data by Marionette）データセットの振り返り

2023年12月22日

秋山 満昭

D3M（Drive-by Download Data by Marionette）とは

- Web感染型マルウェアに関する研究用データセット
  - クライアント（ブラウザ）型ハニーポット[1][2]を用いて収集
  - ブラウザに対するエクスプロイト（ドライブバイダウンロード攻撃）およびインストールされるマルウェアに関連する情報が含まれる
  - 2010年～2015年の各年(6年間)で収集したデータを2010年～2023年の14年間に渡ってMWSで共有
- これまで95本の論文（国内研究会、国際会議、論文誌等）で活用
  - MWSデータセット全体では2023年までに357本の論文で活用された
  - そのうち約27%（95/357）はD3Mを活用
  - 攻撃傾向の変化にともない、2021年以降はデータセットは活用されていない

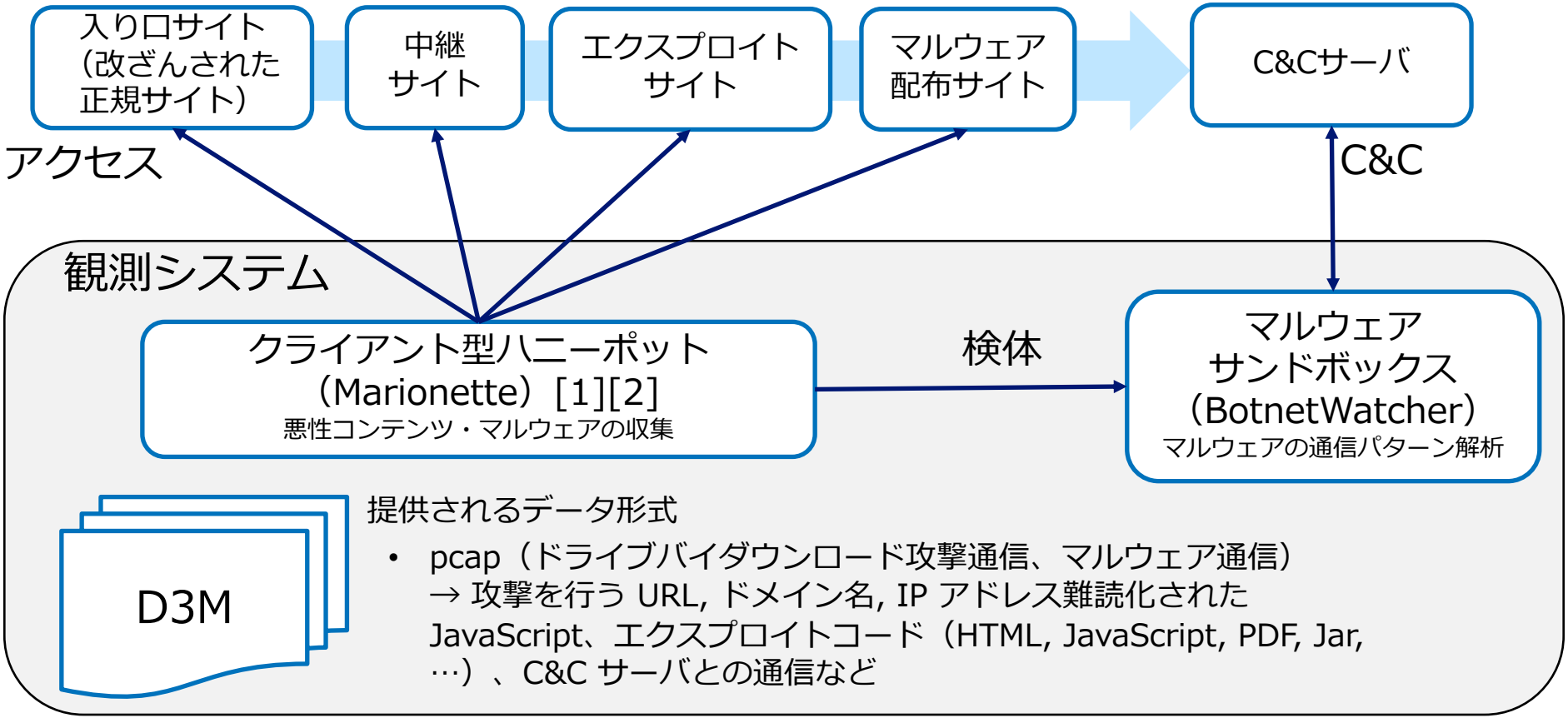
→ データセットの役割は十分に果たしたため、2023年度をもってD3Mの共有を終了します
（現在利用中の研究の発表は制約されません）

[1] Mitsuaki Akiyama, Kazufumi Aoki, Yuhei Kawakoya, Makoto Iwamura, and Mitsuataka Itoh, “Design and Implementation of High Interaction Client Honeypot for Drive-by-download Attacks,” IEICE Transactions on Communication, Vol.E93-B No.5 pp.1131–1139, May. 2010.
[2] Mitsuaki Akiyama, Yuhei Kawakoya, and Takeo Hariu, "Scalable and Performance-Efficient Client Honeypot on High Interaction System." IEEE/IPSJ International Symposium on Applications and the Internet (SAINT), 2012.

ドライブバイダウンロード攻撃

- ユーザがブラウザで特定のウェブサイトにアクセスした際に、そのユーザーの許可や意図なく、マルウェアを自動的にダウンロード・実行させる攻撃手法
  - 複数のウェブサイトをまたがる複雑な攻撃経路、ウェブコンテンツの難読化、ブラウザフィンガープリンティングによる標的の選定、クローキング、攻撃ツールキット（Exploit Kit）による攻撃サイトの自動構築、など攻撃側の技術が山盛り
- 2009年頃に全世界で大規模に発生（2010年代後半まで継続）
- アンダーグラウンドエコノミーの発達によって攻撃ツールキット（Exploit kit）の開発競争、攻撃者の分業化、攻撃サイト構築の自動化が進む

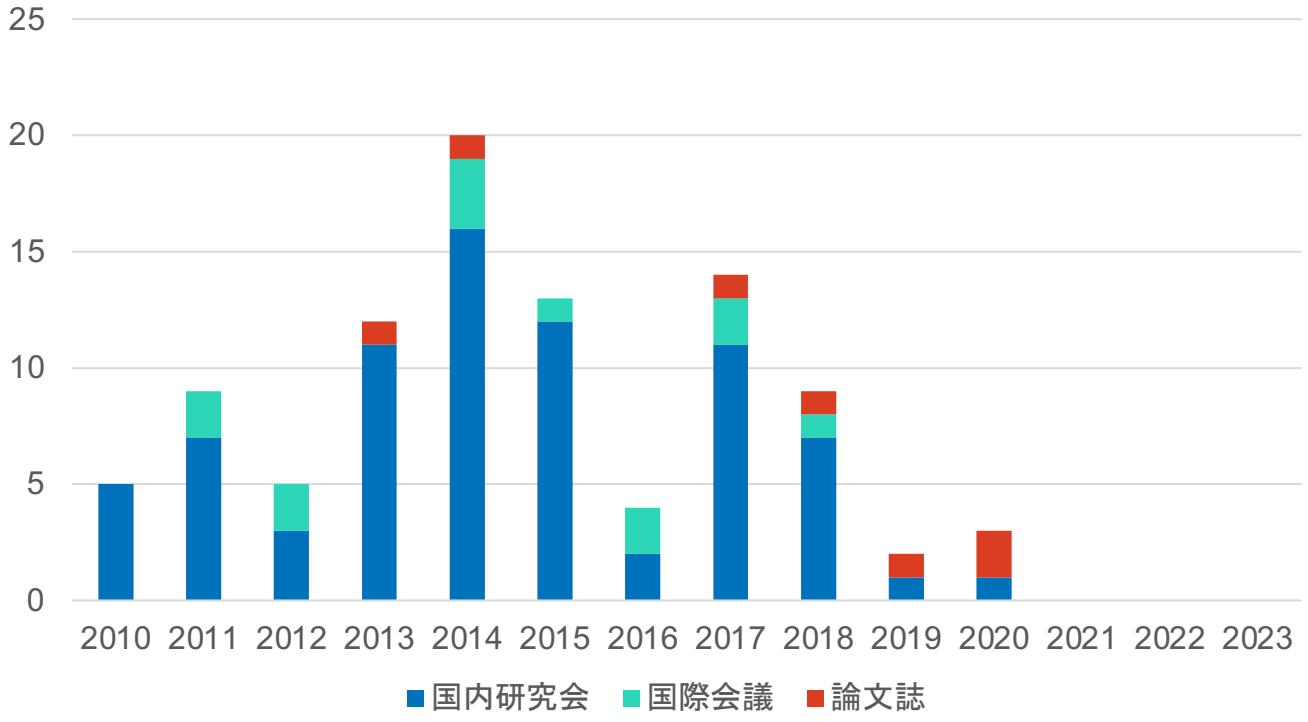
データセット収集方法



D3Mを用いた研究

https://www.iwsec.org/mws/achievements.html

合計：95本
　国内研究会：76本
　国際会議：12本
　論文誌：7本



研究への活用事例

- Webサイトのリダイレクト構造の解析と検知
- 難読化されたWebコンテンツの解析と検知
- エクスプロイトコードの解析と検知
- 攻撃ツールキット（Exploit kit）の解析と検知
- マルウェアの挙動解析と検知
- 悪性Webサイト検知/防御システムの設計・実装

D3Mを活用した論文一覧

2010

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～", MWS2010, [CCC,MARS,D3M].
福島祥郎, 堀良彰, 櫻井幸一: "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, [D3M].
寺田剛陽, 古川忠延, 東角芳樹, 鳥居悟: "検知を目指した不正リダイレクトの分析", MWS2010, [D3M].
桑原和也, 安藤槙悟, 藤原将志, 菊池浩明, 寺田真敏, 趙晋輝: "パスシーケンスに基づくDrive-by-Download 攻撃の分類", MWS2010, [D3M].
戸部和洋, 森達哉, 千葉大紀, 下田晃弘, 後藤滋樹: "実行ファイルに含まれる文字列の学習に基づくマルウェア検出方法", MWS2010, [D3M].

2011

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 ～", MWS2011, [CCC,D3M].
中村暢宏, 佐々木良一: "累積データを用いたボットネットのC&Cサーバ特定手法の評価", MWS2011, [D3M].
宮本大輔, ブラングレゴリー, 秋山満昭："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, [D3M].
ブラングレゴリー, 秋山満昭, 宮本大輔, 門林雄基："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, [D3M].
神薗雅紀, 西田雅太, 小島恵美, 星澤裕二："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, [D3M].
高田雄太, 森達哉, 後藤滋樹: "Web感染型マルウェアのリダイレクト解析", 情報処理学会全国大会講演論文集, Vol.2011, No.1, pp.497-499, (2011), [D3M].
Yuta Takata, Shigeki Goto, Tatsuya Mori: "Analysis of Redirection Caused by Web-based Malware", APAN 32nd Meeting, Network Research Workshop 2011, (2011), [D3M].
寺田真敏: "マルウェア対策研究人材育成ワークショップ~ 教育コミュニティへの貢献とその課題", 情報処理学会研究報告. コンピュータと教育研究会報告, Vol.2011, No.10, pp.1-6, (2011), [CCC,MARS,D3M].

D3Mを活用した論文一覧

2012

義則隆之, 伴拓也, 宮嵜仁志, 松井拓也, 佐藤両, 岡崎亮介, 篠田昭人, 廣友雅徳, 毛利公美, 神薗雅紀, 白石善明："通信可視化と動的解析の連携による攻撃解析支援", MWS2012, [D3M].
上西拓真, 神薗雅紀, 西田雅太, 森井昌克："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, [D3M].
寺田真敏, 岩本一樹, 遠藤基, 松坂志, 小林偉昭："サイバー攻撃対策のための観測記述データ表記に関する検討", MWS2012, [D3M].
Gregory Blanc, Daisuke Miyamoto, Mitsuaki Akiyama, and Youki Kadobayashi: "Characterizing Obfuscated JavaScript using Abstract Syntax Trees: Experimenting with Malicious Scripts", The 2012 IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA 2012), (2012), [D3M].
Daiki Chiba, Kazuhiro Tobe, Tatsuya Mori, Shigeki Goto: "Detecting Malicious Websites by Learning IP Address Features", SAINT2012: The 12th IEEE/IPSJ International Symposium on Applications and the Internet, pp.29-39, (2012), [D3M].

2013

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 ～", MWS2013, [CCC,MARS,D3M,FFRI,NICTER].
宮本久仁男: "偽装した名前解決レスポンスを用いた不正サイトアクセス防御システムの実装と評価", MWS2013, [D3M].
北野美紗, 大谷尚通, 宮本久仁男: "Drive-by-Download攻撃における通信の定性的特徴とその遷移を捉えた検知方式", MWS2013, [D3M].
笠間貴弘, 神薗雅紀, 井上大介："Exploit Kitの特徴を用いた悪性Webサイト検知手法の提案", MWS2013, [D3M].
大西祥生, 常松直樹, 永塚学, 野口雅矢: "改ざんサイトの傾向分析に基づいた検出手法", MWS2013, [D3M].
川島英之："データ管理基盤のマルウェア分析への適用", MWS2013, [D3M,PRACTICE].
川古谷裕平, 塩治榮太朗, 岩村誠, 針生剛男: "APIコール間のデータ依存関係を利用したマルウェア通信内容の特定", MWS2013, [D3M].
大月勇人, 瀧本栄二, 齋藤彰一, 毛利公一："Alkanetにおけるシステムコールの呼出し元動的リンクライブラリの特定手法", MWS2013, [CCC,D3M].
金岡晃, 加藤雅彦, 小出洋, 岡本栄司："組織内ネットワークとマルウェアのモデル化データを用いたマルウェア被害分析", MWS2013, [PRACTICE,D3M].
永井信弘, 千葉大紀, 後藤滋樹: "HTTP通信の時間軸解析によるWeb感染型マルウェア検知", 情報処理学会全国大会講演論文集, Vol.2013, No.1, pp.549-551, (2013), [D3M].
Daiki Chiba, Kazuhiro Tobe, Tatsuya Mori, Shigeki Goto: "Analyzing Spatial Structure of IP Addresses for Detecting Malicious Websites", Journal of Information Processing (JIP), Vol.21, No.3, pp.539-550, (2013), [D3M].
大月優輔, 市野将嗣, 木村聡一, 畑田充弘, 吉浦裕: "未知マルウェア検知のためのペイロード特徴量の有効性評価", 日本セキュリティ・マネジメント学会全国大会, pp.63-70, (2013), [D3M].

D3Mを活用した論文一覧

2014

松本浩明, 石井啓之, 薄羽大樹, 菊池浩明 : "Drive-by-download攻撃通信の可視化システム", MWS2014, [D3M].
高田雄太, 秋山満昭, 八木毅, 針生剛男: "プログラムスライシングを用いた環境依存コードの実行網羅性向上による潜在的URLの抽出", MWS2014, [D3M].
今野由也, 角田裕: "Drive-by-Download攻撃における悪性PDFの特徴に関する考察", MWS2014, [D3M].
松中隆志, 窪田歩, 星澤裕二: "Drive-by Download攻撃対策フレームワークにおけるWebアクセスログを用いたWebリンク構造の解析による悪性サイト検出手法の提案", MWS2014, [D3M].
進藤康孝, 佐藤彰洋, 中村豊, 飯田勝吉: "マルウェア感染ステップのファイルタイプ遷移に基づいたDrive-by Download攻撃検知手法", MWS2014, [D3M].
大谷尚通, 益子博貴, 重田真義 : "実環境におけるサイバー攻撃検知システムの有効性評価および検知範囲の拡大に向けた検討", MWS2014, [D3M].
田中恭之, 有川隼, 畑田充弘: "統計的手法を用いたマルウェア判定の実験結果", MWS2014, [D3M].
大月勇人, 瀧本栄二, 齋藤彰一, 毛利公一 : "ブランチトレース機能を用いたシステムコール呼出し元アドレス取得手法", MWS2014, [D3M].
黒米祐馬, 武田圭史 : "仮想マシンモニタにおけるテイント伝搬を用いた不正プログラムのステルス解析", MWS2014, [D3M].
千葉大紀, 八木毅, 秋山満昭, 青木一史, 針生剛男: "感染後通信検知のための通信プロファイリング技術の設計と評価", MWS2014, [D3M].
岩野透, 吉浦裕, 畑田充弘, 市野将嗣 : "LPCケプストラム分析を用いたマルウェア検知手法", MWS2014, [D3M].
渡辺喬之, 畑田充弘 : "動的解析に基づく評価用マルウェアの選定方式に関する検討", MWS2014, [D3M].
幾世知範, 青木一史, 八木毅, 針生剛男: "通信先と端末内の挙動との依存関係に基づくマルウェアダウンロードサイト特定手法", MWS2014, [D3M].
桑原和也, 櫻井陽大, 杉山ゆかり, 木村重規, 高橋則行: "次世代脅威に対する仮想化技術を用いた隔離システムの提案", MWS2014, [D3M].
Yusuke Otsuki, Masatsugu Ichino, Soichi Kimura, Mitsuhiro Hatada, Hiroshi Yoshiura: "Evaluating payload features for malware infection detection", Journal of Information Processing (JIP), Vol.22, No.2, pp.376-387, (2014), [D3M].
小崎頌太, 後藤滋樹: "HTTP通信の遷移に基づくWeb感染型マルウエア検出法", 電子情報通信学会総合大会講演論文集, Vol.2014, No.2, pp.180, (2014), [D3M].
Takashi MATSUNAKA, Ayumu KUBOTA, Takahiro KASAMA: "An Approach to Detect Drive-by Download by Observing the Web Page Transition Behaviors", The 9th Asia Joint Conference on Information Security (AsiaJCIS 2014), pp.19-25, (2014), [D3M].
高森健太郎, 岩本舞, 小島俊輔, 中嶋卓雄: "マハラノビス距離を用いた難読化マルウェアJavaScriptの検出", 情報処理学会研究報告. マルチメディア通信と分散処理研究会報告, Vol.2014, No.17, pp.1-7, (2014), [D3M].
Mitsuhiro Hatada, Masato Terada, and Tatsuya Mori: "Seven years in MWS: experiences of sharing datasets with anti-malware research community in Japan", 21st ACM Conference on Computer and Communications Security 2014 (ACM CCS 2014), (2014), [CCC,MARS,D3M,IIJ MITF,PRACTICE,FFRI,nicter darknet].
Kentaro Takamori, Mai Iwamoto, Shunsuke Oshima, Takuo Nakashima: "Detection of JavaScript of Malware with un-readability using Mahalanobis-distance", The 6th International Workshop on Network Traffic Control, Analysis and Applications (NTCAA 2014), pp.497-502, (2014), [D3M].

D3Mを活用した論文一覧

2015

佐藤祐磨, 中村嘉隆, 高橋修: "通信遷移とURLの属性情報を用いた悪性リダイレクト防止手法", MWS2015, [D3M].
益子博貴, 重田真義, 大谷尚通："Exploit Kitの変化への適応を目的としたサイバー攻撃検知システムの改良", MWS2015, [D3M].
蘇佳偉, 吉岡克成, 四方順司, 松本勉："情報理論的指標と異常検知に基づく難読化悪性JavaScript検知手法の提案", MWS2015, [D3M].
畑田充弘, 森達哉: "未知マルウェア検知に向けたマルウェア通信の実態調査", MWS2015, [D3M].
鮫島礼佳, 畑田充弘, 吉浦裕, 市野将嗣: "感染挙動の時系列情報のクラスタリングに基づくマルウェア検知手法", MWS2015, [CCC,D3M,PRACTICE].
黒米祐馬, 武田圭史："異なる解析環境を組み合わせたステルス性の高い動的バイナリ計装によるマルウェアの通信解析", MWS2015, [D3M].
中野進, 大月勇人, 明田修平, 瀧本栄二, 齋藤彰一, 毛利公一: "Windows 7 x64環境におけるマルウェア解析向けデータ取得法", MWS2015, [D3M].
大月勇人, 中野進, 明田修平, 瀧本栄二, 齋藤彰一, 毛利公一: "Windows 10 x64環境を対象とするシステムコールトレーサの実現手法", MWS2015, [D3M].
今野由也, 角田裕: "Exploit Kitで作成された悪性コンテンツの類似性調査", MWS2015, [D3M].
寺田成吾, 小林峻, 小出和弘, 羽藤逸文, 瀬戸口武研, 道根慶治, 山下康一："ネットワーク通信の相関性に基づくDrive-by Download攻撃検知手法", MWS2015, [D3M].
T. Adachi and K. Omote: "An Approach to Predict Drive-by-Download Attacks by Vulnerability Evaluation and Opcode," Proceedings of the 10th Asia Joint Conference on Information Security (AsiaJCIS 2015), IEEE, pp.145-151, (2015), [D3M].
安達貴志, 面和成: "脆弱性評価及びOpcodeを用いたDrive-by-Download攻撃予測手法の提案", The 32nd Symposium on Cryptography and Information Security (SCIS 2015), 2A1-3, (2015), [D3M].
岩本舞, 小島俊輔, 中嶋卓雄: "マハラノビス距離を用いた静的解析によるマルウェアの検出", 情報処理学会 研究報告マルチメディア通信と分散処理（DPS）, Vol.2015, No.49, pp.1-7, (2015), [D3M].

2016

青山佑平, 吉井章, 坂東翼, 尾崎幸也, 大倉佳歩, 小林孝史："Drive-by Download攻撃の解析支援アプリケーションの開発と評価", MWS2016, [D3M].
大坪雄平, 窪優司, 三村守, 田中英彦："攻撃種別に着目した悪性文書ファイルの特徴に関する考察", MWS2016, [D3M].
D. Adachi and K. Omote: “A Host-Based Detection Method of Remote Access Trojan in the Early Stage,” Proceedings of the 12th International Conference on Information Security Practice and Experience (ISPEC 2016), LNCS, Vol.10060, Springer-Verlag, pp.110-121, (2016), [D3M].
Mai Iwamoto, Shunsuke Oshima, Takuo Nakashima: "A Study of Malicious PDF Detection Technique," 2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), pp.197-203, (2016), [D3M].

D3Mを活用した論文一覧

2017

上川先之, 山内利宏 : "API操作ログ取得による難読化JavaScriptコード解析支援システム", MWS2017, [D3M].
三須剛史, 巻島和雄, 岡田晃市郎, 岩本一樹: "ソースコードの類似度に基づく悪性JavaScriptの分類に関する一検討", MWS2017, [D3M].
高田真資, 高橋健一, 川村尚生, 菅原一孔 : "Drive-by-Download攻撃検知の評価", MWS2017, [D3M].
尾崎幸也, 上山真也, 小西達也, 山崎雅斗, 坂東翼, 小林孝史: "悪性 URL の強調表示による Drive-by Download攻撃解析支援手法の提案", MWS2017, [D3M].
今健吾, 長瀬智行: "DBIを用いたPDFに含まれる悪性コード抽出手法の提案", MWS2017, [D3M].
金澤しほり, 中村嘉隆, 稲村浩, 高橋修: "悪性IPアドレスの分布特徴に基づく未知のWebサイトの判別手法", MWS2017, [NCD in MWS Cup 2014,CCC ,D3M].
三村守, 田中秀磨: "パラグラフベクトルへのプロキシサーバーログの丸投げ方式", MWS2017, [D3M,BOS,NCD in MWSCup 2014].
Mamoru Mimura and Hidema Tanaka (National Defense Academy): ”Long-term Performance of a Generic Intrusion Detection Method Using Doc2vec,” Proceedings of the 4th International Workshop on Information and Communication Security, pp.456-462 (2017), [NCD,D3M].
Mamoru Mimura and Hidema Tanaka (National Defense Academy): ”Heavy Log Reader: Learning the Context of Cyber Attacks Automatically with Paragraph Vector,” Proceedings of the 13th International Conference on Information Systems Security, LNCS, Vol.10717, pp.146-163 (2017), [NCD,D3M,BOS]
高田真資, 高橋健一, 川村尚生, 菅原一孔: "Drive-by-Download攻撃検知項目の評価", コンピュータセキュリティシンポジウム 2017 論文集, pp.811-816, (2017), [D3M].
高田真資, 高橋健一, 川村尚生, 菅原一孔: "Drive-by-Download攻撃検知手法における検知項目の調査", 第19回IEEE広島支部学生シンポジウム論文集, pp.281-284, (2017), [D3M].
岩本舞, 小島俊輔, 中嶋卓雄: "One-Class SVM を用いたマルウェア PDF 検出の一考察", 情報処理学会 第79回全国大会講演論文集, Vol.2017, No.1, pp.525-526, (2017), [D3M]
Mai Iwamoto, Shunsuke Oshima, Takuo Nakashima: "Malware Detection Method based on OC-SVM Focusing on Features of PDF Files," ICIC Express Letters, Vol.11, No.11, pp.1611-1618, (2017), [D3M]
岩本舞, 小島俊輔, 中嶋卓雄: "One-Class SVMを用いたマルウェアPDF検出手法の改良", 情報処理学会 研究報告コンピュータセキュリティ（CSEC）, Vol.2017, No.8, pp.1-8, (2017), [D3M]

D3Mを活用した論文一覧

2018

森本康太, 鄭俊俊, 齋藤彰一, 瀧本栄二, 毛利公一: "動的解析においてログが取得できないマルウェアの実態調査", MWS2018, (2018), [CCC,D3M,BOS].
高田真資, 高橋健一, 川村尚生, 菅原一孔: "Drive-by-Download攻撃時の通信データの関連性分析", MWS2018, (2018), [D3M].
三須剛史, 巻島和雄, 岩本一樹: "ソースコードの類似度に基づく悪性JavaScriptの検知", MWS2018, (2018), [D3M].
Mamoru Mimura and Hidema Tanaka (National Defense Academy): ”Reading Network Packets as a Natural Language for Intrusion Detection,” Proceedings of the 20th Annual International Conference on Information Security and Cryptology, LNCS, Vol.10779, pp.339-350 (2018), [NCD,D3M].
高田真資, 高橋健一, 川村尚生, 菅原一孔: "Drive-by-Download攻撃時の通信データの関連性分析", コンピュータセキュリティシンポジウム 2018 論文集, (2018), [D3M].
高田真資, 高橋健一, 川村尚生, 菅原一孔: "Drive-by-Download攻撃時の通信データの特徴の調査", 第20回IEEE広島支部学生シンポジウム論文集, (2018), [D3M].
岩本舞, 小島俊輔, 中嶋卓雄: "One-Class SVM を用いたマルウェアPDF検出手法", 情報処理学会 第80回全国大会講演論文集, Vol.2018, No.1, pp.451-452, (2018), [D3M].
石井将大, 猪俣漸(東京工業大学), オユンビレグ チングン, 京山剛大(NRI セキュアテクノロジ゛ーズ゛), 黒田渓介(東京工業大学), 佐藤敦(野村総合研究所), 菅谷光啓(NRI セキュアテクノロジ゛ーズ゛), 田中圭介, 千葉龍一郎, 西脇利知(東京工業大学), 橋本幸典, 廣野壮志(NRI セキュアテクノロジ゛ーズ゛), 松浦知史, 森健人(東京工業大学): "プロキシログのクラスタ間遷移に着目した異常検知手法の評価", 電子情報通信学会, 2018年 暗号と情報セキュリティシンポジウム (SCIS2018), (2018), [D3M]
Mamoru Mimura and Hidema Tanaka: ”Leaving All Proxy Server Logs to Paragraph Vector,” Journal of Information Processing, Vol.26, pp.804-812 (2018), [D3M,BOS,NCD].

D3Mを活用した論文一覧

2019

佐野涼太, 花田真樹, 早稲田篤志, 村上洋一, 布広永示, 折田彰, 関口竜也: "抽象構文木に基づくネスト構造に関する特徴を用いた悪性JavaScript検知手法", MWS2019, (2019), [D3M].
Mamoru Mimura (National Defense Academy): ”An Attempt to Read Network Traffic with Doc2vec,” Journal of Information Processing, Vol.27, pp.711-719, (2019), [D3M,NCD].

2020

Rozi Muhammad Fakhrur, Sangwook Kim, Seiichi Ozawa: "Deep Pyramid Convolutional Neural Networks for Detecting Obfuscated Malicious JavaScript Codes Using Bytecode Sequence Features", MWS2020, (2020), [D3M].
Mamoru Mimura: "Adjusting Lexical Features of Actual Proxy Logs for Intrusion Detection," Journal of Information Security and Applications, Vol.50, pp.102408, (2020), [D3M].
上川先之, 秋山満昭, 山内利宏: "Proxyオブジェクトを用いた解析妨害JavaScriptコード解析支援システムの実現", 情報処理学会論文誌, Vol.61, No.6, pp.1134-1145, (2020), [D3M].

D3Mデータセットを多数の研究でご活用いただき
ありがとうございました