様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　　元 年　　６ 月 １２ 日現在

機関番号：３２６８９
研究種目：基盤研究(B)（一般）
研究期間：2016～2018
課題番号：１６Ｈ０２８３２

研究課題名（和文）セキュリティ解析の共通基盤となるマルウェア・インフォマティクスの確立

研究課題名（英文）Malware Informatics as a Power Base of Cyber Security Analysis

研究代表者
後藤　滋樹（Goto, Shigeki）

早稲田大学・理工学術院・教授

研究者番号：３０２８７９６６

交付決定額（研究期間全体）：（直接経費）　　13,270,000円

研究成果の概要（和文）：ネットワーク社会の最大の脅威となっているのがサイバー攻撃である．攻撃対策技術を確立することが社会的な要請であるが，従来の対策技術は個々の攻撃に対処するものであり，しかも人間の介在が必要であった．
本研究は，データ科学からのアプローチを採り，サイバー攻撃対策技術の研究成果を有効に蓄積するためのマルウェア・インフォマティックスの確立を目指した．
本研究では，実際に大量のデータを蓄積した．さらにデータの特徴を抽出・選択する際の素性(feature)エンジニアリングの技法を体系的に整理して比較した．データの特徴を使う対策技術において中心的な役割を担う機械学習のアルゴリズムを評価する方法を提案した．

研究成果の学術的意義や社会的意義
ネットワークの利用が拡大するにつれてサイバー攻撃の脅威が高まっている．攻撃対策技術の確立が望まれるが，攻撃の手法・対象が多岐にわたるために対症療法的な対策になっている．このような現状を認識して，本研究では成果が蓄積されて後に活用されることを重視している．
本研究で提案したのはマルウェア・インフォマティックスという総合的な枠組である．この中には大量のデータを整理して蓄積して多くの研究者に活用されたものがある．またデータの特徴を解析して攻撃対策の技術を確立する際に重要となる素性(feature)エンジニアリングがある．さらに機械学習のアルゴリズムの比較評価がある．この枠組が活用されることを望む．

研究成果の概要（英文）：In modern networked society, the most severe threat is Cyber Attack. There is a significant demand for establishing defense technology for cyber attacks. There have been many research projects on cyber attacks. However, they deal with a specific kind of attacks individually, and they include some manual operations in their methods. This project proposes the Malware Informatics which covers the large scale database of malware (malicious software). It also shows the feature engineering, which is useful and powerful in data analysis. It proposes a new method for evaluating machine learning algorithms which play central roles in our data science approach to cyber defense technology. We have published many papers and described detailed results on the Web page of our research project.

研究分野：情報セキュリティ

キーワード：セキュリティ　マルウェア　大規模データ解析　モバイル

様　式　C－１９、F－１９－１、Z－１９、CK－１９（共通）

１．研究開始当初の背景

(1) サイバー攻撃は高度に複雑・巧妙化している．攻撃の中核は悪意のあるソフトウェア，すなわちマルウェアである．マルウェアはコードの難読化，解析回避技術，攻撃手順の多段化等，多様な技術を駆使しているために，単純な方法で解析することは困難である．ウォール・ストリート・ジャーナル紙はマルウェア攻撃の実に 55%がアンチ・ウィルスソフトの検知をかいくぐることを紹介して，セキュリティ企業であるシマンテック社の上席副社長が「アンチウィルスソフトウェアは死んだ」と発言したことを報じている．このように従来のアプローチには限界がある．

(2) マルウェアの解析には静的解析と動的解析がある．静的解析はデバッガや逆アセンブラを利用して，マルウェアを実行することなく動作を詳細に解析する．静的解析は熟練した解析者が深い知識と経験に基づいて行うものであり，時間とコストがかかることが欠点である．動的解析は仮想環境等で マルウェアを実行し，その挙動を記録したデータを解析する．動的解析は自動化が可能であるためコスト面で静的解析よりも有利である．ただし，得られたデータの分析には依然として人手が必要である．

２．研究の目的

(1) 本研究の目的はサイバー攻撃の対策技術に資することである．サイバー攻撃ではマルウェアと呼ばれるウィルスやワーム等，悪意のあるソフトウェアが活動する．攻撃者が生成するマルウェアは膨大な数にのぼる．これを人手で解析するのは困難である．本研究では，大規模なデータを自動的に解析する手段としてデータ科学的アプローチを採用する．

(2) 従来からデータ科学的アプローチによるマルウェア解析技術が提案されているが，個々の技術が断片的であるために再利用が難しい．本研究の目的はデータ科学的アプローチによるマルウェア解析技術を再利用可能な共通基盤として体系化して「マルウェア・インフォマティクス」を確立することである．具体的には，①検体収集・蓄積，②素性エンジニアリング，③検体解析・性能評価の各要素技術を再利用可能な形で統合して体系化する．

３．研究の方法

本研究の特徴である科学的アプローチとは，機械学習，パターン認識，信号処理，時系列解析など，データを対象とする数理統計的技法を指す．本研究では特に機械学習に焦点を当てて，マルウェアの検知，検出，分類の課題に取り組む．具体的には以下の 3 つの WP (work package) の研究を遂行する．本研究では大規模データを用いた網羅的な実験を行う．

(1) WP1: 検体収集・蓄積

科学的に再現性のある解析を行うためには時間・空間の両軸を十分にカバーする大規模なデータを研究者間で共有する事が必要不可欠である．この課題を解決するために，本研究は時間軸（収集時刻）と空間軸(収集ポイント)の情報を明示した大規模なマルウェア検体のレポジトリを構築する．我々は従来から学会におけるマルウェア検体の情報共有の活動に参加している．本研究はレポジトリの構築で先鞭をつける．

(2) WP2: 素性(feature)エンジニアリング

マルウェア解析の静的解析や動的解析で得られるデータに対して機械学習アルゴリズムを有効に適用し，識別や分類のタスクを成功させるためには，適切な特徴を抽出・選択する素性エンジニアリング（feature engineering）が必要不可欠である．素性は経験的に決定されることが多く，決定の背後に潜む理由や直感が詳細に論文に記述されることは少ない．本研究では素性エンジニアリングの技法を体系的に整理して比較することにより，再利用可能な研究成果を蓄積する．

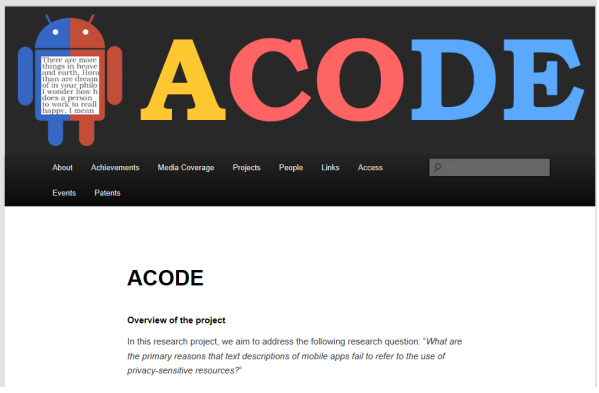
(3) WP3: 検体解析・性能評価

マルウェアを解析する目的に応じて適切な機械学習アルゴリズムを選択し，適切なメトリクス・方法を用いてその性能評価をすることが重要である．従来は研究毎に独自のメトリクスが用いられるケースが多く，異なる方式間の比較が困難であった．本研究ではマルウェア検知，マルウェア検出，マルウェア分類の 3 つの目的を実現するためにどのようなアルゴリズムが最適であるか，また得られた出力結果の性能を評価するためにはどのようなメトリクス・方法が最適であるかを明らかにする．

４．研究成果

(1) 本研究で得られたデータを研究者のコミュニティで共有するために，Web サイトを開設している．引用文献①

その中で紹介している研究成果の一つに ACODE がある．ACODE の意味は Analyzing COde and DEscription である．ここで対象としているのはモバイル機器のアプリケーション(アプリ)である．モバイル機器の利用者がアプリを入手するときには marketplace と呼ばれるサイトを利用する．

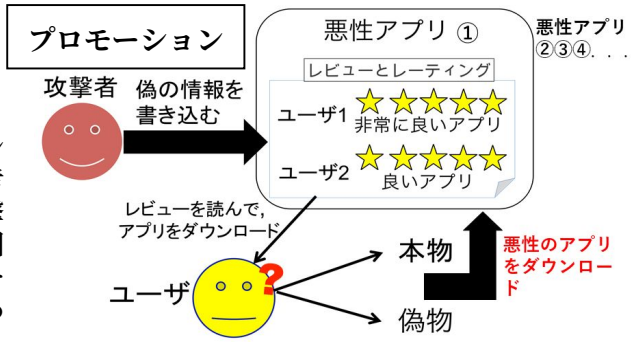


個々のアプリには文章による説明が記載されているが，それが正確とは限らない．特に問題が起きやすいのは個人情報を利用するアプリであるのに，その事実を明記していない場合である．スマートフォンユーザの意図しないプライバシー情報の漏洩を防ぐために，開発者はアプリが端末から情報を取得する旨を十分に説明すべきである．この問題を詳細に分析するために，本研究では ACODE という技術を提案した．ACODE はアプリのソフトウェアを静的に解析する．同時に説明文を自然言語として機械学習により解析する．ここでは人手による文章解析を必要としない．能率良く自動的に膨大なデータを処理することが可能である．本研究では，公式およびサード・パーティの marketplace にある 200,000 に及ぶアプリを解析した．説明文は多言語にわたる．この解析によって，個人情報を利用するアプリであるのに説明文には明記されていない理由を明らかにすることができた． 引用文献②，③

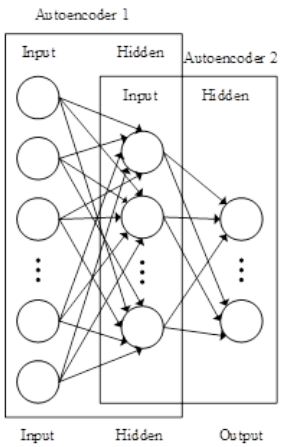
(2) Android アプリはリパッケージングによる改変が容易であるため，オリジナル版開発者に無許可で開発されたリパッケージングを施したアプリ(クローン)が広く流通している．クローンの多くが悪意のある目的で作成されている一方で，アプリ生成サービス等によって真正に作成されたものの，意図せずに元のアプリに類似したアプリ(関連アプリ)となる場合もある．本研究は公式およびサードパーティ・マーケットで収集した 130 万件を超える Android アプリを解析した．この目的はクローンと関連アプリの実態調査である．そのために類似アプリを抽出してクローンと関連アプリを分類する APPraiser と名付けたフレームワークを開発した．分析の結果，公式マーケットでは類似アプリの 76%が関連アプリであったのに対し，サードパーティ・マーケットでは類似アプリの 45%がクローンであると分かった．また公式マーケットからサードパーティにクローンされたアプリの 80%が悪性アプリであることが判明した．引用文献④

悪性コードがゲームに追加された状態

(3)モバイルアプリの配布プラットフォームのエコシステムにおいて，アプリストアは重要な役割を果たしている．ユーザが好みのアプリを選択する際に，アプリストアに掲載されている情報(レーティング，レビュー，ダウンロード数等)を参考にすることは一般的である．そこで，攻撃者は上記のような掲載情報を悪用し，意図的にアプリの評判をつりあげるプロモーショナル攻撃を仕掛けることが可能である．例えば，攻撃者が偽のレーティングとレビューを介して悪性アプリの評判を上げると，その悪性アプリを多くのユーザにダウンロードさせる攻撃を実行できる．本研究は，機械学習によりプロモーショナル攻撃を検知するシステムを開発した．提案システムはプロモーショナル攻撃を 90%以上の精度で検知できる．また大規模な評価ではシステムが特定した攻撃者を追跡することで、マルウェアを効果的かつ効率的に発見することが可能である．引用文献⑤

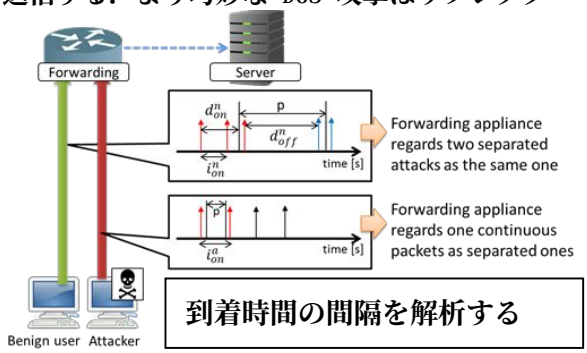


(4)接続経路を匿名化するプロトコルとして Tor (The Onion Routing)が利用されている．Tor の技術自体はサイバー攻撃とはならないが，Tor を使う目的が違法な取引を仲介する Web サイトへのアクセスとなる場合がある(dark Web)．本研究は Web サイトの指紋攻撃(fingerprinting attack)の新しい技術を提案して，Tor 利用者が匿名通信を用いてどの Web サイトへアクセスしたかを識別する方法を提案する．本研究では，Deep Learning の技術であるデノイジング・オートエンコーダや畳み込みニューラルネットワークを用いて特徴量を自動的に抽出する．本研究の成果は，Open World Test の評価において正しく識別した率(True Positive Rate)が 0.89，謝った識別をした率(False Positive Rate)が 0.006 である．この結果は既存の手法よりも高い精度で Website Fingerprinting Attack が行えることを意味する．
引用文献⑥



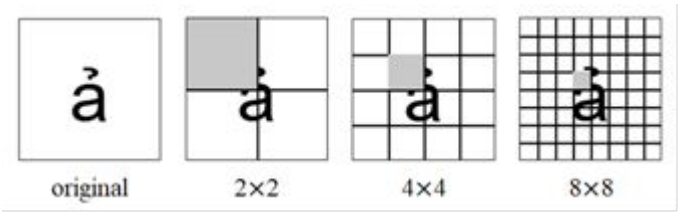
オートエンコーダを二段重ねる例

(5)インターネット上の攻撃の一つに DoS (Denial of Service，サービス妨害)がある．一般的な DoS 攻撃は大量のパケットを標的に向けて送信する．より巧妙な DoS 攻撃はリフレクター(reflector)と呼ばれるサーバを悪用する．その攻撃を DRDoS と呼ぶ．DRDoS では攻撃者が直接に送信するパケットの量を少なく抑えることができる．そのために DRDoS の検知が難しくなる．このような DRDoS 攻撃による被害が増加している．従来の方法で DRDoS 攻撃の特徴を用いて検出する場合にはパケットの中身を調べる必要がある．その方法ではネットワーク機器に対する負荷が増大する．また従来の特徴を示さない新規の DRDoS 攻撃に対応できない．本研究では，パケットの到着時間の間隔を分析する．具体的には統計的な外れ値検出と機械学習によるクラスタリングを用いて，DRDoS 攻撃に用いるプロトコルに依存せず，またトラヒックレートを抑えた攻撃にも有効な検知方法を提案した．引用文献⑦



到着時間の間隔を解析する

(6) 正規サイトのドメイン名に類似したドメイン名を利用してユーザを間違った Web サイトに誘導するフィッシング詐欺がある．例えば，小文字のエル(l)と数字のイチ(1)は紛らわしい．またオー(O)とゼロ(0)も類似している．この問題はドメイン名に多言語が導入されて，より深刻になった．ホモグラフ攻撃は正規サイトのドメイン名に含まれる文字を視覚的に類似する文字列で置換したドメイン名を生成して，ユーザを本来とは異なるサイトへ誘導する攻撃である．非 ASCII 文字を用いることが可能な国際化ドメイン名（IDN）では類似ドメイン名を多数生成することが可能であり，攻撃者がホモグラフ攻撃で生成した IDN（ホモグラフ IDN）がフィッシング攻撃で利用された事例が存在する．既存のホモグラフ IDN の検知手法には固定的な変換表を用いてドメイン名を検知する手法がある．この方法では変換表に登録されていない文字を検知することができない，変換表を人手で更新する必要があるという課題がある．本研究ではホモグラフ IDN が視覚的に正規サイトのドメイン名に類似しているという特徴に着目して，人間の視覚を模擬するために機械的な Optical Character Recognition (OCR)を利用してホモグラフ IDN を検知する手法を提案している．評価に当たっては実際に利用されていた 319 万件の IDN，および 1 万件の悪性な IDN を用いて従来手法と提案手法を比較した結果，従来手法では検知できなかったホモグラフ IDN を検知可能であることを確認した．引用文献⑧



OCR で視覚を模擬するためのマスク・パタン

＜引用文献＞

① https://nsl.cs.waseda.ac.jp/kibanb2016-2018/

② https://nsl.cs.waseda.ac.jp/projects/acode/

③ T. Watanabe, M. Akiyama, T. Sakai, H. Washizaki, and T. Mori, “Understanding the Inconsistency between Behaviors and Descriptions of Mobile Apps,” IEICE Transactions on Information and Systems, Vol.E101-D, No.11, pp. 2584--2599, November 2018.

④ Y. Ishii, T. Watanabe, M. Akiyama, and T. Mori, “APPraiser: A large scale analysis of Android clone apps,” IEICE Transactions on Information and Systems, Vol.E100-D, No.8, pp.1703--1713, Aug. 2017.

⑤ B. Sun, X. Luo, M. Akiyama, T. Watanabe and T. Mori, “PADetective: A Systematic Approach to Automate Detection of Promotional Attackers in Mobile App Store,” Journal of Information Processing (JIP), vol.26, pp.212--223, April 2018.

⑥ Kota Abe, and Shigeki Goto,“Fingerprinting Attack on Tor Anonymity using Deep Learning”, Proceedings of the Aaia-Pacific Advanced Network, Vol.42, pp.15--20, 2016.

⑦ Daiki Noguchi, Tatsuya Mori, Yota Egusa, Kazuya Suzuki, and Shigeki Goto, “Discriminating DRDoS Packets using Time Interval Analysis,”Proceedings of the Aaia-Pacific Advanced Network, Vol.44, pp.1--7, 2017.

⑧ Yuta Sawabe. Daiki Cjiba、Mitsuaki Akiyama, and Shigeki Goto, "Detection Method of Homograph Internationalized Domain Names with OCR", Journal of Information Processing (JIP), Vol.27, (印刷中), 2019.

５．主な発表論文等

〔雑誌論文〕（計９件）

① YOSHIDA Kanae, IMAI Hironori, SERIZAWA Nana, MORI Tatsuya, and KANAOKA Akira, Understanding the Origins of Weak Cryptographic Algorithms Used for Signing Android Apps, Journal of Information Processing (JIP), 査読有, Vol.27, (印刷中), 2019.
DOI: (印刷中のため未定)

② NAKAMORI Tomofumi, CHIBA Daiki, AKIYAMA Mitsuaki, GOTO Shigeki, Detecting Dynamic IP Addresses and Cloud Blocks Using the Sequential Characteristics of PTR Records, Journal of Information Processing (JIP), 査読有, Vol.27, (印刷中), 2019.
DOI: (印刷中のため未定)

③ SAWABE Yuta、CJIBA Daiki、AKIYAMA Mitsuaki, GOTO Shigeki, Detection Method of Homograph Internationalized Domain Names with OCR, Journal of Information Processing (JIP), 査読有, Vol.27, (印刷中), 2019.
DOI: (印刷中のため未定)

④ SUN Bo、FUJINO Akinori、MORI Tatsuya、BAN Tao、TAKAHASHI Takeshi、INOUE Daisuke, Automatically Generating Malware Analysis Reports Using Sandbox Logs, IEICE Transactions on Information and Systems, 査読有, Vol.E101.D, pp.2622--2632, 2018.
DOI: https://doi.org/10.1587/transinf.2017ICP0011

⑤ WATANABE Takuya、AKIYAMA Mitsuaki、SAKAI Tetsuya、WASHIZAKI Hironori、MORI Tatsuya, Understanding the Inconsistency between Behaviors and Descriptions of Mobile Apps, IEICE Transactions on Information and Systems, 査読有, Vol.E101.D, pp.2584--2599, 2018.
DOI: https://doi.org/10.1587/transinf.2017ICP0006

⑥ B. Sun, X. Luo, M. Akiyama, T. Watanabe and T. Mori, PADetective: A Systematic Approach to Automate Detection of Promotional Attackers in Mobile App Store, Journal of Information Processing (JIP), 査読有, Vol.26, pp.212--223, 2018.
DOI: https://doi.org/10.2197/ipsjjip.26.212

⑦ M. Hatada and T. Mori, Finding New Varieties of Malware with the Classification of Network Behavior, IEICE Transactions on Information and Systems, 査読有, Vol.E100-D, No. 8, pp. 1691-1702, Aug. 2017.
DOI: https://doi.org/10.1587/transinf.2016ICP0019

⑧ Y. Ishii, T. Watanabe, M. Akiyama, and T. Mori, APPraiser: A large scale analysis of Android clone apps, IEICE Transactions on Information and Systems, 査読有, Vol.E100-D, No.8, pp.1703-1713, Aug. 2017.
DOI: https://doi.org/10.1587/transinf.2016ICP0012

⑨ S. Mizuno, M. Hatada, T. Mori, and S. Goto, Detecting Malware-infected Devices Using the HTTP Header Patterns, IEICE Transactions on Information and Systems, 査読有, E101.D, Issue 5, pp.1370-1379, 2018.
DOI: https://doi.org/10.1587/transinf.2017EDP7294

〔学会発表〕（計３２件）以下には 20 件を記載

① T. Yasumatsu, T. Watanabe, F. Kanei, E. Shioji, M. Akiyama, and T. Mori, Understanding the Responsiveness of Mobile App Developers to Software Library Updates, the 9th ACM Conference on Data and Application Security and Privacy (CODASPY 2019), 2019.

② Tatsuya Mori, A measurement study of the Internationalized domain name (IDN) homograph attacks: present and future, Asia Pacific Advanced Network 47, 2019.

③ K. Yoshida, H. Imai, N. Serizawa, T. Mori, and A. Kanaoka, Understanding the Origins of Weak Cryptographic Algorithms Used for Signing Android Apps, the 10th IEEE International Workshop on Security Aspects in Processes and Services Engineering (SAPSE 2018), 2018.

④ 澤部 祐太, 千葉 大紀, 秋山 満昭, 後藤 滋樹, OCR を利用したホモグラフ IDN の検知法, コンピュータ・セキュリティ・シンポジウム 2018, 2018 年.

⑤ 中森 朋郁, 千葉 大紀, 秋山 満昭, 後藤 滋樹, PTR レコードの連続的設定を加味した動的 IP アドレスブロックおよびクラウド領域検出, コンピュータ・セキュリティ・シンポジウム 2018, 2018 年

⑥ 後藤 滋樹, インターネットの大域的な成功と局所的な反省, コンピュータ・セキュリティ・シンポジウム 2018, 2018 年.

⑦ Tomofumi Nakamori, Daiki Chiba, Mitsuaki Akiyama, and Shigeki Goto, Detecting Dynamic IP Addresses Using the Sequential Characteristics of PTR Records, Asia Pacific Advanced Network 46, Research Workshop, 2018.

⑧ Yuta Sawabe, Daiki Chiba, Mitsuaki Akiyama, and Shigeki Goto, Detecting Homograph IDNs using OCR, Asia Pacific Advanced Network 46, Research Workshop, 2018.

⑨ Shigeki Goto, APAN from Simple to Complex, Asia Pacific Advanced Network 46, 2018.

⑩ E. Pariwono, D. Chiba, M. Akiyama, and T. Mori, Don't throw me away: Threats Caused by the Abandoned Internet Resources Used by Android Apps, the 13th ACM ASIA Conference on Information, Computer and Communications Security (ASIACCS 2018), 2018.
⑪ Y. Ishii, T. Watanabe, F. Kanei, Y. Takata, E. Shioji, M. Akiyama, T. Yagi, B. Sun and T. Mori, Understanding the Security Management of Global Third-Party Android Marketplaces, the 2nd International Workshop on App Market Analytics (WAMA 2017), 2017.
⑫ M. Hatada and T. Mori, Detecting and Classifying Android PUAs by similarity of DNS queries, the 7th IEEE International COMPSAC Workshop on Network Technologies for Security, Administration and Protection (NETSAP 2017), 2017.
⑬ B. Sun, X. Luo, M. Akiyama, T. Watanabe and T. Mori, Characterizing Promotional Attacks in Mobile App Store, the 8th International Conference on Applications and Techniques in Information Security (ATIS 2017), 2017.
⑭ Daiki Noguchi, Tatsuya Mori, Yota Egusa, Kazuya Suzuki, Shigeki Goto, Discriminating DRDoS Packets using Time Interval Analysis, the 44th Meeting of the Asia-Pacific Advanced Network, 2017.
⑮ Tatsuya Mori, Large-scale network security measurement through the lens of darknet, The 44th APAN Meeting, Network Security Workshop, 2017.
⑯ Elkana Pariwono, Daiki Chiba, Mitsuaki Akiyama, and Tatsuya Mori, Measurement Study of the Hijackable Internet Resources inside Mobile Apps, 暗号と情報セキュリティシンポジウム (SCIS 2018), 2018 年.
⑰ 安松達彦, 金井文宏, 渡邉卓弥, 塩治榮太朗, 秋山満昭, 森達哉, モバイルアプリ開発者による脆弱性対応の実態調査, コンピュータセキュリティシンポジウム 2017, 2017 年.
⑱ 畑田充弘, 森達哉, DNS クエリ分析に基づく Android PUA の識別と亜種分類, コンピュータセキュリティシンポジウム 2017, 2017 年.
⑲ S. Mizuno, M. Hatada, T. Mori, and S. Goto, BotDetector: A robust and scalable approach toward detecting malware-infected devices, Proceedings of the IEEE International Conference on Communications (ICC 2017), May 2017.
⑳ Tatsuya Mori, How the Internet survey tools could affect network security monitoring, The 43rd APAN Meeting, Network Security Workshop, 2017.

〔その他〕
ホームページ等
https://nsl.cs.waseda.ac.jp/kibanb2016-2018/

６．研究組織

(1)研究分担者
研究分担者氏名：後藤　滋樹
ローマ字氏名：(GOTO, Shigeki)
所属研究機関名：早稲田大学
部局名：理工学術院
職名：教授
研究者番号（8 桁）：30287966

(2)研究協力者
研究協力者氏名：森　達哉
ローマ字氏名：(MORI, Tatsuya)