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　３ 年　 ４ 月　 ８ 日現在

機関番号：１７１０４
研究種目：基盤研究(C)（一般）
研究期間：2018～2020
課題番号：１８Ｋ１１２９６

研究課題名（和文）セキュアなキャンパスネットワークのための能動型マルウェア検出システムに関する研究

研究課題名（英文）An Active Malware Detection System for Secure Campus Networks

研究代表者
佐藤　彰洋（Satoh, Akihiro）

九州工業大学・情報基盤センター・助教

研究者番号：３０６０９３７６

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：大学におけるBYODの実現のため，キャンパスネットワークで観測される通信のみからマルウェアを検出することが求められる．しかしながら，ネットワークで観測されるマルウェアの通信は非常に限定的であり，且つマルウェアは自身の通信を隠蔽する仕組みを有する．本研究では，キャンパスネットワークにおいてマルウェアを検出するため，DNSに対する膨大な数の名前解決要求から存在期間が極端に短い悪性ドメインの判別を実現する．その特徴は，DNSの名前解決要求のみから感染の疑わしい端末を特定できる点，その名前解決の応答を書き換えコールバック先を強制的に変更することでマルウェアに関する能動的な情報収集ができる点にある．

研究成果の学術的意義や社会的意義
総務省は，オリンピックの東京開催を見据え，公衆無線LANの整備を推進している．加えて，高等教育の現場では，学生個人の端末を必携とするBYOD体制を検討する動きが盛んになってきている．このように，自身が所有する端末を外出先のネットワークに接続する利用形態は，今後増加するものと想像できる．その一方，マルウェアに感染した端末をネットワークに持ち込まれる可能性はより高まることになる．本研究の成果は，ネットワークに内在する感染端末を迅速に排除することを可能とする．故に，公衆無線LANやキャンパスネットワークなど，端末の持ち込みを前提としたネットワークにおいて，セキュリティの向上に大きく寄与する．

研究成果の概要（英文）：Some of the most serious security threats facing computer networks involve malware. To prevent malware-related damage, administrators must swiftly identify and remove the infected machines that may reside in their networks. However, many malware families have domain generation algorithms (DGAs) to avoid detection. In this research project, we develop a system to detect malware-infected machines from massive DNS queries. Here, we focus on queried domain names for the DNSs because name resolution is an unencrypted interaction that always occurs prior to malware communication. Our system mainly has two features to detect the infected machines by superficially analyzing DNS queries and to actively collect information about malware families by forcibly changing their callback destinations.

研究分野：ネットワークセキュリティ

キーワード：ネットワークセキュリティ　DGAマルウェア　ドメイン名　機械学習

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

モバイルデバイスの普及に伴い，私物情報端末の業務利用，すなわち BYOD(Bring Your Own Devices)を検討する動きが盛んになってきている．高等教育の現場でも，九州大学が先んじて学生個人の端末を必携とする BYOD 体制に移行した．これに続き，今後多くの大学で BYOD が検討・実施されるものと容易に想像できる．その一方，マルウェアに感染済みの端末をキャンパスネットワークに持ち込まれることが大きな課題となる．マルウェアに感染した端末は，C&C(Command-and-Control Server)を介して攻撃者の指令を受けることにより，ランサムウェアの配布，フィッシング詐欺，標的型攻撃への利用など，様々な犯罪活動を試みる．大学における BYOD では端末環境の多様性のため[1]，キャンパスネットワークで観測される通信のみからマルウェアを検出することが求められる．

これまでネットワーク内の端末の通信を監視するためには，ブラックリストやレピュテーションなど[2,3]，ドメインに関する付加的な情報を用いてきた．その検出を回避するために，高度なマルウェアでは DGA(Domain Generation Algorithm)が実装されている．DGA とは，C&C のドメインを頻繁に変更することで，マルウェアから C&C へのコールバック通信を隠蔽するための仕組みである[4]．具体的には，マルウェアは DGA により機械的にドメインを生成し，それらドメインに対して名前解決を試みる．その名前解決の結果，正しい応答が返ってきたものを C&C のドメインと見なす．

ネットワークで観測されるマルウェアの通信は非常に限定的であり，且つマルウェアは自身の通信を隠蔽する仕組みである DGA を有する．故に，マルウェアの検出ためには，自身のドメインを頻繁に変更する C&C，すなわち存在期間が極端に短い悪性ドメインに対する通信の判別が求められる．キャンパスネットワークにおいて大規模感染が発生した場合，大学の業務に多大な影響を及ぼすため，この問題の解決は急務である．このような厳しい制約の中，一般的な利用を阻害せずにマルウェアの脅威を取り除くことが本研究の試みである．

[1] F. L. Lévesque et al.,“Technological and Human Factors of Malware Attacks: A Computer Security Clinical Trial Approach,” ACM Transactions on Privacy and Security, Vol.21, No.4, pp.18:1–30, 2018.
[2] B. Rahbarinia et al.,“Efficient and Accurate Behavior-Based Tracking of Malware-Control Domains in Large ISP Networks,” ACM Transactions on Privacy and Security, Vol.19, No.2, pp.4:1–31, 2016.
[3] L. Bilge et al.,“Exposure: A Passive DNS Analysis Service to Detect and Report Malicious Domains,” ACM Transactions on Information and System Security, Vol.16, No.4, pp.14:1–28, 2014.
[4] D. Plohmann et al.,“A Comprehensive Measurement Study of Domain Generating Malware,” the USENIX Conference on Security Symposium, pp.263–278, 2016.

２．研究の目的

本研究の目的は，キャンパスネットワークにおいてマルウェアを検出するため，DNS に対する膨大な数の名前解決要求から存在期間が極端に短い悪性ドメインの判別することに集約される．DNS に着目した理由は，DGA により通信先を変更する仕組み上，マルウェアと C&C とのコールバック通信に先んじて，必ず名前解決が発生するためである．

本研究における独自性と創造性は，DNS の名前解決要求のみから感染の疑わしい端末を特定できる点，その名前解決の応答を書き換えコールバック先を強制的に変更することでマルウェアに関する能動的な情報収集ができる点にある．加えて，ネットワークへの導入が非常に容易な点も留意されたい．具体的には，DNS の名前解決のみに着目するため，本研究成果の導入は学内 DNS の名前解決を転送するのみである．

３．研究の方法

本研究の核となる，(3-1)C&C ドメイン検出技術，(3-2)ドメイン選択技術，(3-3)マルウェア解析技術の確立を目指す．以降，詳細を述べる．

(3-1) C&C ドメイン検出技術

DNS に対する名前解決要求から C&C のドメインを検出する技術を確立する．C&C は存在期間が極端に短い悪性ドメインであるため，ドメインに関する付加的な情報に依存したブラックリストやレピュテーションによる検出は困難である．そこで，良性・悪性の判別にドメイン文字列から得られる情報のみを利用する．その理由は，人為的に生成したドメインと機械的に生成したドメインには，その文字列に明白な違いが現れることが予想されるためである．

具体的なアプローチとして，辞書に基づいてドメイン文字列を意味のある単語群に分割すること，その単語群の占める割合からドメイン文字列のランダム性を評価することを検討する．加えて，日本語のアルファベット表記であるローマ字，中国語のアルファベット表記であるピンインなど，複数言語を考慮する．特筆すべきは，DGA に関する事前知識を全く必要とせず，ドメイン文字列のみで良性・悪性を判別できる点である．

(3-2) ドメイン選択技術

ネットワークにおける DNS の名前解決要求は膨大であるため，ドメインの文字列解析による C&C ドメイン検出技術では全てを処理しきれないことが予想される．それを補助するため，良性・悪性を判別すべきドメインを選定する技術を確立する．DGA ではコールバック通信先のドメインの変更に伴い，DNS の名前解決において幾つかの NXDOMAIN 応答が発生することが知られている．ここで，NXDOMAIN 応答はドメインが存在せず名前解決に失敗したことを意味する．この知見に基づき，NXDOMAIN 応答を返した名前解決のみに着目することで，良性・悪性を判別するドメインの数を大きく絞り込む．

NXDOMAIN 応答は，それ自身で原因が特定できるもの，それ以前の名前解決を参照することで原因が特定できるもの，原因の特定が困難なものに分類できる．具体的なアプローチは，類似性に基づいて原因が明確な NXDOMAIN 応答を除外する仕組みを検討する．次いで，原因の不明な NXDOMAIN 応答に対して，ドメインの良性・悪性を判定する．この結果を保持することで，これ以降に発生する同様の NXDOMAIN 応答を除外することが可能となる．

(3-3) マルウェア解析技術

C&C ドメイン検出技術は，ドメイン文字列のみに基づいて良性・悪性を判別するため，如何に運用に耐え得る精度を実現するかが課題となる．それを補助するため，感染が疑われる端末からマルウェアに関する情報を能動的に収集する技術を確立する．

具体的なアプローチは，C&C のドメインと判断された名前解決に対して，正規の内容に代わり情報収集システムのアドレスを返答する．これにより感染が疑われる端末が情報収集システムに送信する内容に基づいて感染の有無を判定する．留意すべきは，書き変えるのが本来 NXDOMAIN 応答を返す名前解決であるため，一般的な利用を全く阻害しない点である．

４．研究成果

先ず「C&C ドメイン検出技術」の有用性を確認するため，予め良性と悪性でラベル付されたドメインを用いて実証実験を行った．その結果，再現率 0.9960，適合率 0.9029 と非常に高い精度で感染端末を検出できることが明らかになった．これは従来手法と比較しても突出した値である．また，その成果を取り纏めたものは学術論文誌に採録されている[5]．しかしながら，計算時間の点で従来手法に劣っており，C&C ドメイン検出技術のみでは実用性の観点から不十分と言える．故に，「ドメイン選択技術」と「マルウェア解析技術」により，その大幅な改善を試みる必要性がある．

次に「ドメイン選択技術」と「マルウェア解析技術」の有用性を確認するため，九州工業大学のキャンパスネットワークにおいて実証実験を行った．その結果，良性と悪性の判別が必要なドメインの数を 1%未満まで低減できることが明らかになった．また，その成果を取り纏めたものは学術論文誌に採録されている[6]．これにより「C&C ドメイン検出技術」の抱える問題点の大幅な改善を実現した．

最後に 3 つの技術を統合した本システムの最終評価のため，九州工業大学のキャンパスネットワークにおいて実証実験を行った．その結果，1 ヶ月間に観測された 3,304,505 の NXDOMAIN 応答を返す DNS クエリから，6520 の感染が疑われる悪性クエリを検出できることが明らかになった．その 1 ヶ月分のデータを処理するのに必要とする計算時間は約 12000 秒であり，実用の範囲内に収まるであろうことを確認した．また，その成果を取り纏めたものは学術論文誌に採録されている[7]．これにより本研究成果の有用性を実証した．

[5] A. Satoh et al.,“Estimating the Randomness of Domain Names for DGA Bot Callbacks,” IEEE Communications Letters, Vol.22, No.7, pp.1378–1381, 2018.
[6] A. Satoh et al.,“Clustering Malicious DNS Queries for Blacklist-based Detection,” IEICE Transactions on Information and Systems, Vol.E102.D, No.7, pp.1404–1407, 2019.
[7] A. Satoh et al.,“A Superficial Analysis Approach for Identifying Malicious Domain Names Generated by DGA Malware,” IEEE Open Journal of the Communications Society, Vol.1, pp.1837–1849, 2020.

総務省は，2020 年オリンピック・パラリンピックの東京開催を見据え，公衆無線 LAN の整備を推進している．公衆無線 LAN の利用者数は増加傾向にあり，2018 年度末時点で約 5600 万人，2020 年度末時点で約 6400 万人と予想されている[8]．一方，高等教育の現場では，学生個人の端末を必携とする BYOD 体制を検討する動きが盛んになってきている．文部科学省の協力を得た大学 ICT 推進協議会の調査によると，BYOD を導入している大学は 30%を超えていることが報告されている[9]．このように，自身が所有する端末を外出先のネットワークに接続する利用形態は，今後増加するものと容易に想像できる．その一方，マルウェアに感染した端末をネットワークに持ち込まれる可能性はより高まることになる．

本研究の実現により，ネットワークに内在する感染端末を迅速に排除することが可能となる．その成果は，公衆無線 LAN やキャンパスネットワークなど，端末の持ち込みを前提としたネットワークにおいて，セキュリティを向上するための要素技術と成り得る．故に，本研究は，日本に

おける情報通信基盤の整備の方向性とも合致するものである．

[8] サイバーセキュリティタスクフォース 公衆無線 LAN セキュリティ分科会，“公衆無線 LAN セキュリティ分科会 報告書，” 2018．
[9] 大学 ICT 推進協議会 ICT 利活用調査部会，“BYOD を活用した教育改善に関する調査研究結果報告書，” 2018．

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　5件／うち国際共著　0件／うちオープンアクセス　2件）

１．著者名 Akihiro Satoh, Yutaka Fukuda, Toyohiro Hayashi, Gen Kitagata	４．巻 1
２．論文標題 A Superficial Analysis Approach for Identifying Malicious Domain Names Generated by DGA Malware	５．発行年 2020年
３．雑誌名 IEEE Open Journal of the Communications Society	６．最初と最後の頁 1837--1849
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 佐藤 彰洋, 林 豊洋, 和田 数字郎, 福田 豊	４．巻 62
２．論文標題 DGAマルウェアにより自動生成された悪性ドメインの判別	５．発行年 2021年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 --
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 佐藤 彰洋, 福田 豊, 井上 純一, 中村 豊	４．巻 62
２．論文標題 辞書に基づくDGAマルウェアに起因した悪性ドメインの判別	５．発行年 2021年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 829--837
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Akihiro Satoh, Yutaka Nakamura, Daiki Nobayashi, Kazuto Sasai, Gen Kitagata, Takeshi Ikenaga	４．巻 E102.D
２．論文標題 Clustering Malicious DNS Queries for Blacklist-based Detection	５．発行年 2019年
３．雑誌名 IEICE Transactions on Information and Systems	６．最初と最後の頁 1404--1407
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Akihiro Satoh, Yutaka Nakamura, Daiki Nobayashi, Takeshi Ikenaga	4．巻 22
2．論文標題 Estimating the Randomness of Domain Names for DGA Bot Callbacks	5．発行年 2018年
3．雑誌名 IEEE Communications Letters	6．最初と最後の頁 1378--1381
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計7件（うち招待講演　0件／うち国際学会　3件）

1．発表者名 Akihiro Satoh, Yutaka Nakamura, Gen Kitagata
2．発表標題 Identifying Malicious Domain Names Caused by Dictionary-based DGA Bots
3．学会等名 RIEC Annual Meeting on Cooperative Research Projects（国際学会）
4．発表年 2021年

1．発表者名 Akihiro Satoh, Yutaka Nakamura, Kazuto Sasai, Gen Kitagata
2．発表標題 A Malicious Domain Detection Approach for Callbacks of DGA Bots
3．学会等名 RIEC Annual Meeting on Cooperative Research Projects（国際学会）
4．発表年 2020年

1．発表者名 佐藤彰洋，福田豊，和田数字郎，中村豊
2．発表標題 辞書に基づくDGAボットにより生成された悪性ドメインの判別
3．学会等名 インターネットと運用技術シンポジウム
4．発表年 2019年

１．発表者名
Akihiro Satoh, Yutaka Nakamura, Kazuto Sasai, Gen Kitagata

２．発表標題
A Malicious DNS Query Clustering Approach for Blacklist-based Detection

３．学会等名
RIEC Annual Meeting on Cooperative Research Projects（国際学会）

４．発表年
2019年

１．発表者名
佐藤彰洋，中村豊，小倉光貴，野林大起，池永全志

２．発表標題
ブラックリストに基づく検出の効率化に向けた悪性DNSクエリ分類手法

３．学会等名
インターネットと運用技術シンポジウム

４．発表年
2018年

１．発表者名
佐藤彰洋，中村豊，小倉光貴，野林大起，池永全志

２．発表標題
ブラックリストにより検出された悪性DNSクエリの分類

３．学会等名
インターネットアーキテクチャ研究会

４．発表年
2018年

１．発表者名
佐藤彰洋，中村豊，笹井一人，北形元

２．発表標題
原因に基づく悪性DNSクエリの分類

３．学会等名
第25回先進的情報通信工学研究会

４．発表年
2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関