6Z-3

効率的なマルウェア収集環境の構築

吉原大道　碓井利宣　武田圭史　村井純

慶應義塾大学環境情報学部

252-0882 神奈川県藤沢市遠藤 5322

{daido，alc，jun}@sfc.wide.ad.jp,keiji@sfc.keio.ac.jp

1. はじめに

2011 年に行われた McAfee Labs の調査[1]によると，2011 年の第 1 四半期におけるマルウェアの検出率は，例年の第 1 四半期のデータと比べて高い数値を示しており，今後マルウェアの数はさらに増加すると考えられる．本研究では，多様化し増加を続けるマルウェアに対応するため，受動型ハニーポットよりも効率的にマルウェアを収集することができる能動型ハニーポットの構築を行い，マルウェア研究者の支援を行う．本研究における能動型ハニーポットとは，自ら能動的に Web サイトにアクセスし，マルウェアを収集するハニーポットと定義する．

2. 事前調査

本研究における事前調査として，通常のハニーポットでの収集可能な検体数を確認するため，ハニーポットを構築しマルウェアを収集した．ハニーポットを構築したサーバには 1 つの IP アドレスを振り，Nepenthes を使用することで受動型マルウェア収集を行った．その結果，半年に 26 種類の検体を収集することができた．また，MalwareDomainList[2](以下 MDL)に掲載されている全ての URL にアクセスし，能動型マルウェア収集を行った．この方法では，一日に約 10 数検体の収集ができた．MDL に掲載されたサイトは数日後には既に削除または対策済みのものがほとんどであるため，迅速に収集を行う必要がある．

本研究ではこれらの調査結果を踏まえ，短期間でより多くのマルウェアを集めるために能動型マルウェア収集を行う．

3. 目的

第 2 節で示した通り低対話型ハニーポット，また MDL のみに頼る収集方法では得られる情報量が少ない．そのため，本研究ではマルウェアの解析を行う研究者を支援するために，MDL に頼らず効率的にマルウェアを収集できるハニーポットの環境構築を行う．

Structure of Efficient and Effective Malware Collecting System
Daido Yoshihara , Toshinori Usui , Keiji Takeda , Jun Murai, Faculty of Environment and Information Studies, Keio University 252-8520, Kanagawa, Japan

4. 関連研究

星澤裕二らの研究[3]では，効率的にマルウェアを収集するため，URL を示す文字列の特徴から優先的にアクセスする URL を定めている．このことから短時間で多くの URL にアクセスすることができることが示されている．氏らの研究では，asp，cgi といった拡張子の URL，ドメインに IP アドレスを含む URL，トップレベルドメインが特定の国コードである URL を MDS のものと判定し，優先的に巡回するとしている．本研究では，MDL に掲載される URL 情報の中に多く含まれる拡張子及びドメインを同時に含む URL，ドメインに IP アドレスを含む URL，実際にアクセスした際に得られる Web サイト内の情報及び通信の流れなどから MDS を判定し，効率的にマルウェアを収集する．

5. アプローチ

本研究では，Web クローリングを行った結果と事前に特定したマルウェアダウンロードサイト(以下 MDS)の特徴とを照らし合わせ，本研究にて定めた MDS の特徴に該当する項目が多い URL を MDS である確率が高いものとし，MDS である確率が高い URL のリストを作成する．そのリストをもとに，MDS へのアクセスを行うことで効率よく能動型のマルウェア収集を行う．

5.1 判定方法

MDL に掲載されている MDS の URL をドメイン名，ディレクトリ，ファイル名に分割する．また，各 URL に対応する IP アドレスの情報，MDL に掲載されている URL にアクセスし，そのレスポンスから得られる情報を複数組み合わせることで多次元的に分析し，MDS の特徴を推測する．

(a)MDL に掲載されている URL から抽出したドメインとファイルの拡張子の組み合わせにより MDS を判定する．掲載されている約 70000 個の URL の中で，100 個以上存在するドメイン

及び拡張子を抽出する．抽出された.cn や.ru といったトップレベルドメインや.info のような比較的値段の安いドメインと，SQL インジェクション攻撃が仕掛けられている可能性のある php，cgi などといった拡張子をそれぞれ組み合わせることにより MDS の URL を判定する．

(b)URL の中に IP アドレスが入っているものを MDS と判定する．MDS が新しく作成される場合に，特定されることを防ぐために，動的な IP アドレスを URL に使用していると推測している．そのため IP アドレスと，MDS に一番多く見られるファイルの拡張子とを組み合わせて URL を生成する．

(c)MDL に掲載されている URL にアクセスし，そのレスポンスから得られる情報をもとに MDS を判定する．Google の Niels Provos 氏らの研究[4]では，Web クローラにより収集した Web サイトの中から，別の悪意のあるページに対して iframe を用いてリンクされている Web サイトや難読化された Javascript を含んでいる Web サイトを「悪意のあるページ」として MapReduce を用いて判定している．本研究ではその方法に加え，MDS にアクセスした時の通信の流れの特徴確認や目視による Web サイト上のコンテンツの確認により，MDS を判定する．

5.2 収集環境

本研究で構築するハニーポットは高対話型のものであるため，VirtualBox4.1.4 上の OS にて構築している．使用 OS は WindowsXP(SP なし)，使用ブラウザは InternetExplorer6 である．

5.3 収集方法

仮想マシン上にて本研究における判定方法で MDS と判定した Web サイトに優先的にアクセスし，マルウェアを収集する．(図 1)

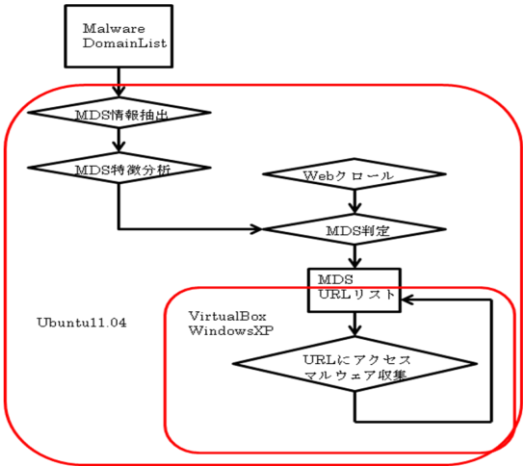

図 1

6. 評価方法

本研究では先述した MDL に掲載されているドメインにアクセスすることでマルウェアを収集し，その結果収集することができた検体数と，今回の判定方法を用いて収集することができたマルウェア検体数を比較することにより，本研究で構築したハニーポットのマルウェア収集効率の評価を行う．また，一定数以上の MDS，正常なサイトに対して本研究の判定方法を用いて確認した場合に，どの程度正確に判定できるのかを検証し，評価を行う．

7. 判定結果

4．関連研究で述べた様に，本研究における判定方法を用いて MDS の判定を行う場合，通常よりも効率よく判定を行うことが可能であることが示されている．そのため，星澤らの研究における MDS の判定方法よりも多くの方法を用いることで，より効率的な収集環境の構築が可能となる．

8. まとめ

本研究では，日々増加を続けるマルウェアについて分析を行う研究者を支援することを目的とし，MDS の URL が持つ特徴を分析することで効率的に MDS にアクセスし，短期間でより多くのマルウェアを収集するためのシステムを構築した．特定のドメインおよび拡張子を持つ URL，IP アドレスが存在する URL，アクセスした際に確認される MDS の特徴をもとに MDS を判定し，効率的に収集を行う．本システムによるマルウェア収集方法の結果と，事前調査および受動型ハニーポットである Nepenthes の使用，MDL に掲載される MDS へのアクセスにて行った収集方法における結果を比較することで検体数の評価とした．また，無作為に抽出した MDS と正常なサイトとを本研究の判定方法にて比較しその正確性を確認した．

参考文献

[1] McAfeeR Labs™，” McAfee 脅威レポート:2011 年第1 四半期”，2011

[2] MalwareDomainList，http://www.malwaredomainlist.com/

[3] 星澤裕二 他，“自律型クライアントハニーポットの提案”，電子情報通信学会，2009

[4] Niels Provos 他， The Ghost In The Browser Analysis of Web-based Malware，Google, Inc，2007