マルウェア解析の効率化手法の検討

山口 和晃　　堀合 啓一　　田中 英彦
情報セキュリティ大学院大学　情報セキュリティ研究科
{mgs081101,dgs063101,tanaka}@iisec.ac.jp

1　はじめに

近年のマルウェアは、プロの手により複雑化しており、数そのものも爆発的に増えていることから、人手による解析のみでは解析者の負担が多く、解析が困難になっている。そのため、自動的なマルウェアの解析が必要であるが、最近は仮想マシンで実行されているか、ネットワーク接続の有無、日時などの動作環境要因から処理を分岐することで、振る舞いを変化させる耐解析機能を備えたマルウェアが増えている。このことから、自動的解析の精度の低下や解析効率の低下が起きている。そのため、本研究では、マルウェアの実行環境や実行方法の工夫により、自動的解析の精度の向上や解析効率の向上に対する有効性の検討を行う。

2　マルウェアの動的挙動自動解析システム

動的挙動解析の全体構成を図１[1]に示す。ネットワーク環境はLinuxのカーネル・パケット・フィルタに使用されているiptablesの機能を利用して構築している。模擬サーバの一部は、Truman[7]の機能を利用した。

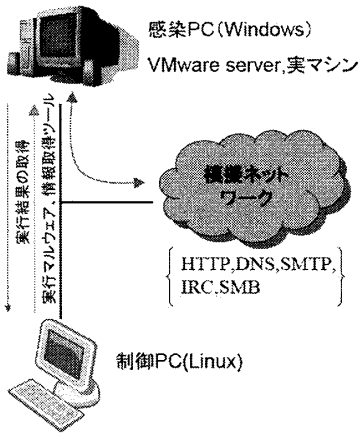

図１ 動的挙動解析の全体構成[1]

Efficiency Improvement of Malware Analysis

Kazuaki YAMAGUCHI
Keiichi HORIAI
Hidehiko TANAKA
Institute of information security

この模擬ネットワークには、マルウェアを実行する感染PCが接続され、模擬DNS、IRC、SMTP、SMB、HTTPの各サーバ群は、実際には制御PCの中に実装している。また、Windows内の挙動については、主にAPI CALLの情報を解析するのではなく、マルウェアを実行する前の状態と実行後の状態を記録したログを比較し、それらの差分をマルウェアの挙動を示す情報として抽出している。また、マルウェアを実行するWindows OSの種類としてWindows XP各サービスパックの場合の比較、マルウェアを実行する感染PC環境を仮想マシンにするか、実マシンでするかによる比較、API CALL監視の有無による比較ができる環境を構築した。

2.1　マルウェアの実行制御

次にマルウェアの実行制御について述べる。マルウェアの挙動解析は、実行制御用のホスト（以下ホスト）OSと、マルウェアを実行する感染PC (仮想マシンのVMware Serverまたは、実マシン)上で作動しマルウェアを実行する。感染PCのOSが起動すると続いて感染PC内のローダが起動し、感染PC内の情報取得に必要なソフトウェアをホストから受信し、マルウェア実行前のログを取得する。続いて解析対象のマルウェアをホストから受信してこのマルウェアを実行し、指定した時間経過後にマルウェア実行後のログを取得する。また、マルウェアの実行時間は２００秒とした[3]。

3　実マシン環境と仮想マシン環境の比較

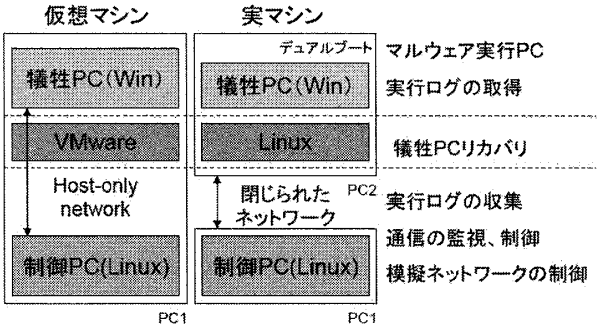

図２ 実マシン環境と仮想マシン環境の違い[1]

マルウェアの挙動を自動的に解析するには、マルウェ

アの実行環境、自動解析処理の全体制御や挙動の記録と分析、感染後のPC の復旧などの機能が必要となる。

実マシン環境と仮想マシン環境の違いを整理した結果を、図２に示す。同図に示すように、ほとんどの処理は共通化が可能であり、違いのある部位は、感染後に復旧する機能を実現する部分だけとなっている。

4　マルウェアの実行で取得する情報

挙動解析の結果から、レジストリ改竄、システムファイルの変更、プロセスリスト、Hostファイルの改竄、ルートキットの有無、通信ログ、などの項目を文字列の情報として抽出し、これを解析データとして蓄積する 。

5　実行環境によるマルウェアの挙動の変化

動作環境に依存せずに実行時刻などから確率的に挙動パターンを変更するマルウェアについて考察する。マルウェアの挙動を解析することにおいて、１種類の実行パターンしか持たないマルウェアに対して何度も実行していては無駄が多い。だが、複数の実行パターンを持つマルウェアにおいてはできるだけ全てのパターンを網羅する必要がある。実行回数の効率化を考えると、ある回数において、あるパターン数が確認されていたとして、その次の解析においてまた新たな挙動パターンを得られる確率を考え、解析を終了するかどうかを判断すればよいと考えられる。この確率は式（１）の大数の法則から導き出せる。

$$\Pr\left(\left|\bar{x}-\mu\right|\geq\varepsilon\right)\leq\frac{\sigma^2}{n\varepsilon^2} \qquad (1)$$

ここで、注意が必要なことは、この式が成り立つには確率が独立であり、一様分布している必要がある。この式から、最低１０回以上の解析を行わなければ、９０％以上の確率で全ての挙動パターンを網羅したとは言えないことがわかる。

6　解析精度の評価方法

解析精度の評価は、取得した解析データを文字列に置き換えることにより、ハミング距離によってマルウェアの類似度の算出をすることでおこなった。

7　実験

本大学院のハニーポットで取得したマルウェアを使用して解析を行った。仮想環境と実環境の解析システムを使用し、マルウェアの実行回数の比較、Windows OS の種類の比較、API CALL 監視の有無による比較をおこなった。

得られた解析データと亜種のマルウェアの結果との比較検討により、自動的解析の精度の向上に対する有効性の検討をおこなった。

8　まとめと今後の課題

今回の実験によって、マルウェアの挙動を正確に把握するためには、マルウェアの実行条件を複数変更する必要性あることがわかった。また、Windows OS の種類の変更や、実行時刻の違い、マルウェアのパッキング状況、API CALL 監視の有無の各有効性の検討により、動的解析の解析効率の向上と精度の向上が可能であることがわかった。また、仮想マシンと実マシンとの併用によって、仮想マシンの解析結果から解析に時間がかかる実マシンの解析の回数を低減することで解析効率の向上が可能であることがわかった。

今後の課題としては、マルウェアの挙動パターンの変更のメカニズムの調査から更なる効率化の検討が必要である。また、解析精度においては、今回の解析装置は、仮想環境の模擬ネットワークを使用したが、安全な通信のみを実ネットワークに接続させることにより、より向上が可能である[5]と考えられることから、検討の必要がある。

9　参考文献

[1] 堀合 啓一、今泉 隆文、田中 英彦、"定点観測によるボットネットの観測と Malware の動的挙動解析システムの提案" 、情報処理学会論文誌 Vol.49 No.4 1-12、2008/4

[2] 堀合 啓一、今泉 隆文、田中 英彦、" ハミング距離によるマルウェア亜種の自動分類"、第４１回 CSEC、2008/5

[3] 笠間 貴弘、吉岡 克成、井上 大介、衛藤将史、中尾康二、松本 勉、"マルウェア動的解析におけるマルウェア実行時間に関する検討"、SCIS 2009、2009/1

[4] 星澤 裕二、岡田 晃市郎、山村 元昭、椎木 孝斉、マルウェアの動作条件の抽出、情報処理学会研究報告、2007

[5] 吉岡 克成、松本 勉、" 自動マルチパス解析によるマルウェア動的解析の提案"、SCIS2009、2009

[6] Tony Lee、Jigar J.Mody、" Behavioral Classification"、In Proceedings of EICAR 2006、2006/4

[7] TJoe Stewar 、Truman、" The Reusable Unkonwn Malware Analysis Net"