科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　１６　日現在

機関番号：　１１２０１
研究種目：　基盤研究(C)
研究期間：　2011～2013
課題番号：　２３５００００７４

研究課題名（和文）難読化されたウイルス攻撃を防御・検出する高速ベイジアンフィルタの研究

研究課題名（英文）An automatic unpacking method for computer virus effective in the virus　filter based on Bayesian theorem

研究代表者
厚井　裕司（Koi, Yuji）

岩手大学・工学部・非常勤講師

研究者番号：２０３３３７５０

交付決定額（研究期間全体）：（直接経費）　　4,000,000円、（間接経費）　　1,200,000円

研究成果の概要（和文）：近年のコンピュータウイルスは、圧縮や難読化が施されて解析が困難な状態になった実行可能圧縮とよばれる形式に変換されている。代表的な実行可能圧縮形式は、ASPack、UPX が挙げられる。そこで、本研究では圧縮形式に依存しないウイルスの自動解凍方式に関して考察し、学習型のアンチウイルスフィルタと組み合わせた実験を行った。すなわち、スパムメール向けの学習アルゴリズムであるGraham Bayes理論をウイルスに適用したもので、実行ファイルにおけるバイナリ情報の文字列の特徴から未知のマルウェアを抽出する。実験では、95%の検知率と0.02%の誤検出率をはるかに越える除去性能を達成した。

研究成果の概要（英文）：A rapid automatic virus detection algorithm using static code analysis is necessary.However,recent computer viruses are almost compressed into the executable compress format and are obfuscated.Thus,it is difficult to determine the characteristics of the binary code from the obfuscated computer viruses.
In this research,a method that unpacks compressed computer viruses automatically without restriction to compression type is proposed.The proposed method unpacks the common compression formats accurately 80% of the time,while unknown compression formats can also be unpacked.The proposed method is effective against unknown viruses by combining it with the existing known virus detection system like Bayesian Virus Filter.We could achieve to implement 95% detection rates and 0.02% false detection rates.

研究分野：　総合領域

科研費の分科・細目：　情報学・計算機システム・ネットワーク

キーワード：　ベイジアンフィルタ　ベイズの定理　難読化　暗号化　実行可能圧縮　未知ウイルス

１．研究開始当初の背景

政府のサイバークリーンセンターは、2007年末にウイルス対策ソフトで分析できない割合が17%であると発表した。この中には、構造を分析できないでウイルスらしいと判断している割合が相当高く含まれている。

このため、ウイルスのシグネチャを人手で抽出する現存技術の限界が叫ばれていた。本研究では、新しいベイジアンフィルタを実現して、このような状況を打開する

未知ウイルス(類似や亜種のウイルスを含む)の研究は、国内では皆無で海外に以下がある。

（１）IBM社がニューラルネットワーク技術を用いた自動免疫システムを研究開発
（２）コロンビア大がデータマイニング技術で大量の通信記録からウイルス攻撃をオフライン抽出

これらの中で定量的な成果を得ているものは無く、実用化は困難であると考えられる。

本研究では、圧縮・難読化された未知ウイルスに対して以下のような技術的視点から取り組む。

未知のウイルスを細分類すると、以前に相似性のあるウイルスが存在しない新種と以前に相似性のあるウイルスが存在する亜種の2種類に分類が可能である。90%以上の未知ウイルスはオリジナルのウイルスを一部改変して作成された亜種ウイルスであり、作成者が最初からウイルスを作るのに比べて作成に時間がかからない。これが、ボットウイルスの大量発生の原因である。亜種は多くの場合、送信されるメールの内容がほぼ同じものであったり、動作やPCに与える被害が同一のものであったりする。したがって、亜種と亜種、亜種とオリジナルのウイルスの間には相似性が存在する。また、全く異なったウイルス間にも共通した特徴が有る。したがって亜種ウイルスやウイルス間の共通した特徴を検出ができたら、95%以上の未知のウイルスを取り除ける。さらに、ウイルス対策ソフトを制作・販売している企業からは、実用化するためには誤検出率が0.02%以下になる事を求められている。

２．研究の目的

近年のコンピュータウイルスは実行可能圧縮と言う圧縮や難読化が施され、アセンブラコードの検査によるウイルス判定が難しくなっている。商用のアンチウイルスソフトは独自開発した解凍機能を有しているが、処理が複雑すぎて解凍できない場合が多発している。

そこで、私達はウイルスを実行させて、それ自身の解凍ルーチンでウイルスの中身をメモリに展開させた後に実行を止めて、その中身をベイジアンフィルタで検査する方法で80%の解凍・解析に成功した。しかしながら、この方法では毎回ウイルスを実行させるために、処理に時間が掛かる。

本研究では、上記ベイジアンフィルタに解凍前後の情報を記憶させることで、2回目以降の検査を実行可能圧縮のまま可能とする。これにより、実行可能圧縮型の検出率と処理速度の飛躍的な向上を実現する。

３．研究の方法

現在、ウイルスの判定が日々困難になっているが、原因としては以下がある。

・特定の人物や企業を狙うボットウイルスが増えて、検体の入手が難しくなっている。
・ウイルスのソースコードがオープンソースとして公開され、亜種ウイルスが急増している。
・最近の実行可能圧縮は、独自の解凍方法でアンチウイルスに解析不可な処理を採用している。

本研究では上記課題に対処すべく、未知ウイルス（類似ウイルスや亜種ウイルを含む）の特徴を検出する学習型の Bayesian Virus Filter に New Unpacking：ウイルスに気付かれずにその解凍ルーチンでウイルスの中身をメモリ上に展開する解凍機能と Signature Grouping：実行可能圧縮ウイルスの解凍前後の情報を対比することによって、解凍前のウイルスシグネチャを抽出する機能を組込んで、実行可能圧縮型ウイルスの検出率と処理速度の飛躍的な向上を目指す。以下に研究項目と研究内容および課題を纏める。

（１）検出率と処理速度を向上したベイジアンウイルスフィルタを実現
・研究内容
学習型 Bayesian Virus Filter に New Unpacking（ウイルスに気付かれずにウイルス自身で解凍させる機能）と Signature Grouping（解凍前後の情報を対比することによって、解凍前のウイルスシグネチャを抽出する機能）を組込む
・研究課題
- 類似ウイルスや亜種ウイルスを検出
- 95%の未知ウイルス検出率を達成
- 0.02%の誤検出率を達成

（２）ハニーポットシステム構築によるボットを含めた未知ウイルス収集
・研究内容
岩手大学総合情報処理センターと協力して、大規模のハニーポットを構築すると共に、それを利用して各種の未知ウイルスを

収集
・研究課題
➢　ボットウイルス収集機構の実現

（3）ベイジアンウイルスフィルタの改善と総合評価試験
・研究内容
実現したベイジアンウイルスフィルタを実際のネットワーク環境に設置し、性能試験・耐久試験を実施
・研究課題
95%の検出率並びに 0.02%の誤検出率を効率的に実現

既知ウイルスの特徴を学習して、類似点を持った未知ウイルス（類似ウイルスや亜種ウイルを含む）を検出するベイジアンウイルスフィルタに以下の 2 機能を組込む。

①New Unpacking：ウイルスに気付かれずにその解凍ルーチンでウイルスの中身をメモリ上に展開する解凍機能である。実行可能圧縮形式における解凍ルーチンがウイルスの中身に実行を移す時のアドレス空間における長距離ジャンプを検出して制御を止める。その後、メモリ上に展開されたウイルスの中身データを学習型のベイジアンウイルスフィルタに入力する

②Signature Grouping：上記ベイジアンウイルスフィルタに解凍前後の情報を記憶させることで、2 回目以降の検査を実行可能圧縮のまま可能とする。
これにより、実行可能圧縮型ウイルスの検出率と処理速度の飛躍的な向上を可能とする。

以下は、難読化されたウイルスの解凍方法となる。ウイルスがメモリ内で実行されると実行可能圧縮形式の解凍ルーチンすなわち Decompress code が実行されて、メモリ上に一番右のウイルスの中身を展開する。私達はデバッガーを使用してウイルスを実行させ、それ自身の解凍ルーチンでウイルスの中身をメモリに展開した後に実行を止めて、その中身をベイジアンフィルタで検査する方法で 80%の解凍・解析に成功した。しかしながら、この方法ではウイルスがデバッガーによる自身の解析に気が付いて実行を止める場合が多い。また、毎回ウイルスを実行させるために、処理に時間が掛かる。本研究では、対策として下記対応を図る。

・ウイルスに追跡を気付かれないような実行環境を実現する
・2 回目以降の検査を実行可能圧縮のまま可能とする
さらに、ハニーポットシステム構築による未知ウイルス収集では、検体が入手しにくいボットを含めて各種のウイルスを収集するには、大学程度の規模のネットワークが不可欠であり、複数の HP（ハニーポット）や侵入検知サーバさらにはログサーバから構成されるシステムを岩手大学構内に構築してボットを含む各種ウイルスを収集する。

４．研究成果
現在、ウイルスの判定が日々困難になっているが、原因としては以下がある。
・特定の人物や企業を狙うボットウイルスが増えて、検体の入手が難しくなっている。
・ウイルスのソースコードがオープンソースで公開され、亜種ウイルスが急増している。
・最近の実行可能圧縮は、独自の解凍方法でアンチウイルスに解析不可な処理を採用している。

今までの研究の実験手順で以下のような結果を得ている。
・実験データを自動解凍プログラムで解凍
・解凍したファイルから，Windows API である文字列を静的抽出し，その数を数えた。
全体として、53 種類の圧縮ウイルスに対して 42 種類が解凍でき、これは圧縮ウイルス全体の 79.2%に達している。また、11 種類の圧縮形式不明なウイルスは、8 種類が解凍できており、提案手法は圧縮形式を判断できないウイルスに対しても有効であるといえる。解凍できたウイルスに対して、解凍所用時間はすべて 10 秒以内になっている。この結果、リアルタイムでウイルス検出には実用レベルに達していると言える。
解凍結果を検証するため，解凍できたウイルスの中に Bagle シリーズのウイルスを用いて，Bayesian Virus Filter でウイルス検出率を調べた結果では、検出率が一致していることから、提案解凍手法で解凍できたウイルスはオリジナルウイルスと同じ特徴点を持っていると言える。また、今回実験を用いた亜種の数が多くなったことが分かり、これは解凍ツールがない圧縮方式に対して、提案解凍手法で解凍できたからである。

本研究では上記課題に対処すべく、未知ウイルス（類似ウイルスや亜種ウイルを含む）の特徴を検出する学習型の Bayesian Virus Filter で(1) New Unpacking：ウイルスに気付かれずにその解凍ルーチンでウイルスの中身をメモリ上に展開する解凍機能と(2) Signature Grouping：実行可能圧縮ウイルスの解凍前後の情報を対比することによって、解凍前のウイルスシグネチャを抽出する機能を組込んで、実行可能圧縮型ウイルスの検出率と処理速度の飛躍的な向上を可能にした。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　　0件）

〔学会発表〕（計　　0件）

６．研究組織

(1)研究代表者

厚井　裕司（Koi,Yuji）
岩手大学・工学部・非常勤講師
研究者番号：20333750

(2)研究分担者

中谷　直司（Nakaya,Naoshi）
岩手大学・工学部・准教授
研究者番号： 20322969