Vol.2019-MPS-123 No.48
Vol.2019-BIO-58 No.48
2019/6/19

特徴出現頻度によるバッファオーバーフロー攻撃検知に関する考察

小林勇希[†1]　松田健[†2]　園田道夫[†3]　趙晋輝[†1]

概要：バッファオーバーフロー攻撃は，プログラムで用意された領域を超えるデータを送ることで本来想定していない領域を上書きし，情報を不正に操作する攻撃である．侵入検知システムにおいて既知の攻撃の特徴を利用するシグネチャマッチング方式等が研究されてきたが，シェルコードのパターンが無数に作成可能であることから全ての攻撃の検知は難しい．本研究では，攻撃コードのアセンブラ命令列の中に現れる特定命令とその順序に着目した特徴出現頻度解析によるバッファオーバーフロー攻撃の検知手法を，新たな入力データによって検討し，また誤認識問題を解決するために新規手法の提案を行った．

キーワード：バッファオーバーフロー，アセンブラ命令，出現頻度

Detection of Buffer Overflow Attacks Based on the Feature Appearance Frequency

KOBAYASHI YUUKI[†1]　MATSUDA TAKESHI[†2]
SONODA MICHIO[†3]　CHAO JINHUI[†1]

Abstract: Buffer overflow attacks are known as illegal manipulation of information by oversending data and overwriting data area malicously.Intrusion detection systems using signature match with features of known attacks have been proposed, however it is difficult to detect all patterns of attacks created by shell codes.In this research, we evaluate a detection method using analysis on feature frequency and patterns of specific assembler instruction with newly created attack data, and propose a new detection algorithm to reduce false alarms.

Keywords: buffer overflow, assembler instruction,appearance frequency

1. はじめに

バッファオーバーフロー攻撃は古くから知られているサイバー攻撃の 1 つであるが，2019 年 4 月の時点でも，UNIX 系 OS で利用されている wget というファイル取得関係のツールにおいてもその脆弱性が報告されている[1]．そのため，従来からこの攻撃に対する防御方法は様々な手法が検討されているが，新たな攻撃や未知の攻撃に対する有効な防御手法を講ずることは今なお必要性があると言える．

本研究では，攻撃コードのアセンブラ命令列の中に現れる特定命令とその順序に着目した特徴出現頻度解析によるバッファオーバーフロー攻撃の検知手法について，既存手法に対する追加実験と，同様に特徴出現頻度を利用した新規手法の提案を行う．

2. 研究背景

2.1 バッファオーバーフロー

ータの一時記憶領域に想定より長いデータを格納した際に，バッファーの領域を超えてデータが記録されてしまうセキュリティホールである．例として，Internet Explorer などのプログラムで発見されたゼロデイ脆弱性ではバッファオーバーフローを利用したものであった[2]．また，バッファオーバーフローを利用した攻撃をバッファオーバーフロー攻撃と呼び，種類としてスタック，ヒープ，静的領域などを攻撃の標的としたものが存在する．

3. 既存手法

3.1 シグネチャマッチング方式

広く知られているバッファオーバーフロー攻撃対策として，シグネチャマッチング方式が挙げられる．これは侵入検知システムで多く使用され，攻撃コードの特徴的な部分が記述されたシグネチャと呼ばれるデータベースと，ネットワーク上のパケットを比較して攻撃検知を行う方式である．しかし欠点があり，シグネチャに登録されていない攻撃は検出ができず，また攻撃に使用されるシェルコードのパターンが無数に作成できることから，すべての攻撃の検知は原理的に困難であるとされている．

3.2 アセンブラ命令ペアの出現頻度解析

南後らの研究[3]（以下，既存研究 1 とする）では，アセ

†1 中央大学
Chuo University
†2 長崎県立大学
University of Nagasaki
†3 情報通信研究機構
National Institute of Information and Communications Technology

Vol.2019-MPS-123 No.48
Vol.2019-BIO-58 No.48
2019/6/19

ンブラ命令列を，バッファオーバーフロー攻撃に使用されるコードによく見られる 6 命令とそれ以外の命令の 7 つに分類し，それらの各命令を中心に前後複数個の命令の出現順序に着目する近傍の概念を取り入れた頻度解析を行う．頻度解析の結果は攻撃データと正常データそれぞれの 49 次元の特徴出現頻度ベクトルに変換され，それらのベクトルと検知対象データのベクトルのコサイン尺度により差異を求める検知手法が提案されている．

4. 追加実験と提案手法

本研究では，既存研究 1 で用いられた特徴出現頻度を使用して，コサイン尺度による検知手法の追加実験と新手法の提案を行う．既存研究 1 の追加実験は，攻撃に含まれるアセンブラ命令のパターンに類似する正常データを新規に生成して検知を行い，誤検知の発生率を調査する．また，新手法では既存研究 1 の頻度解析の結果を用いて SVM の学習を行い，これに対しても新規に追加した正常データを検知させ誤検知率を調査する．

4.1 コサイン尺度を用いた判定手法の追加実験

まず，既存研究 1 のコサイン尺度を用いた判定手法において，攻撃に含まれるアセンブラ命令のパターンに類似する正常データを増やした追加実験を行い，誤検知の発生率を調査する．追加する正常データは CentOS 7 における/bin 下に存在する Linux コマンドを逆アセンブルした 966 個である．なお，頻度解析の学習に使用したデータは既存研究 1 で使用されたものと同一である．

4.2 SVM を用いた判定手法

次に，コサイン尺度の代わりに SVM を用いた検知実験を行う．学習データは既存研究 1 で使用されたものと同じ 49 次元の出現頻度ベクトルに，新たにラベルを追加した計 50 個のパラメーターを用いる．ラベルはそのデータが攻撃データなら-1，正常データなら 1 をとる値である．

実験では既存研究 1 で使用された 40 個のデータに 4.1 節で示した追加データを合わせた，計 1006 個のデータを使用する．また，実験は Microsoft Azure[4]を利用して行う．

5. 実装結果

まず，以下の表 1 に実験の種類をまとめた．なお，実験 0 は既存研究 1 の内容なので本研究では実験を行っていないが，比較のために結果を示す．

表 1 実験の手法と使用データ

実験番号	手法	使用データ
0	既存研究 1	既存研究 1
1	既存研究 1	新規追加データ
2	SVM	既存研究 1
3	SVM	新規追加データ

5.1 既存手法における追加正常データ検知実験

ここでは実験 0 と実験 1 について示す．実験 0 は既存研究 1 における検知実験であるが，攻撃データの誤検知は 0 だが，正常データ 20 個のうち 2 つのデータにおいて誤検知が発生している．実験 1 では追加した正常データの約 14%である 137 個において誤検知が発生した．

5.2 SVM による検知実験

次に，実験 2 と実験 3 について示す．実験 2 は既存研究 1 で使用された攻撃データ 20 個，正常データ 20 個に対する検知実験であるが，いずれも正しい検知結果を出力した．実験 3 は新規に追加した正常データの検知実験であるが，すべての正常データに対して正しい検知結果を出力した．

表 2 に実験 0 から 3 の結果をまとめて示す．

表 2 検知実験におけるデータ数と誤検知数

	攻撃データ		正常データ	
実験番号	データ数	誤検知数	データ数	誤検知数
0	20	0	20	2
1	0	0	966	137
2	20	0	20	0
3	0	0	966	0

6. 考察とまとめ

5.1 節より，実験 1 では既存研究 1 における新規データの誤検知率が 14%程度であった．また 5.2 節より，SVM を用いた判定手法では誤検知が 1 件も無かったことがわかる．コサイン尺度を用いた攻撃検知では意図的に誤検知を起こさせるデータを作成できたが，SVM による攻撃検知ではそれらのデータにおける誤検知は発生せず，既存手法 1 で用いた近傍の概念を取り入れたアセンブラ命令ペアの出現頻度は SVM を使用することでより効果的に働くことが確認できた．

今後の課題として，SVM 以外の手法についても特徴出現頻度を用いた検知実験を行い，より効果的な検知手法を探していくこと，そして誤検知を発生させやすい正常データの特徴を発見し，それらの特徴を持つデータについて誤検知を引き起こさない手法を提案することなどが挙げられる．

参考文献

[1] JVN，”JVN#25261088 GNU Wget におけるバッファオーバーフローの脆弱性”.https://jvn.jp/jp/JVN25261088，（参照 2019-05-13）.

[2]Jon Erickson 著，村上 雅章 訳，“Hacking 美しき策謀 脆弱性攻撃の理論と実際 第 2 版”，オライリー・ジャパン，2011.

[3] 南後 吉秀，松田 健，園田 道夫，趙 晋輝，“アセンブラ命令の出現状況に着目したバッファオーバーフロー攻撃の検知とその考察”，情報処理学会第 79 回全国大会，2017.

[4] Microsoft，“Microsoft Azure”. https://azure.microsoft.com/ja-jp/，（参照 2019-05-10）.

正誤表

下記の箇所に誤りがございました．お詫びして訂正いたします．

訂正箇所	誤	正
1 ページ 第 2 章 1 節 1 行目	ータの一時記憶領域に想定より長いデータを格納した際に，バッファーの領域を超えてデータが記録されてしまうセキュリティホールである．	バッファオーバーフローとは，バッファーと呼ばれるデータの一時記憶領域に想定より長いデータを格納した際に，バッファーの領域を超えてデータが記録されてしまうセキュリティホールである．