5K-03

攻撃データ以外の特徴によるOSコマンドインジェクション攻撃の可視化*

姜 雨†, 松田 健‡, 牛込 龍太郎§, 園田 道夫¶, 趙 晋輝‖
†§¶‖中央大学 ‡長崎県立大学

1 はじめに

Webアプリケーションに対する攻撃は増加傾向にあることが様々な機関の報告 [1], [2] からも読み取ることができる．Webアプリケーションに対する代表的な攻撃として，OSコマンドインジェクションやSQLインジェクションが有名である．OSコマンドインジェクション攻撃やSQLインジェクション攻撃は古くから知られている攻撃であり，多くの対策が研究されているが，上述の通り，被害は拡大し続けている．Webアプリケーションに対するこのようなインジェクション攻撃は，開発時に想定していない入力が原因となるため，本研究では，特定の入力データをターゲットにしたときに，そのターゲット以外の入力を正しく検知できるかどうか調査した．ここでターゲットにする入力は，URLや住所など，Webの入力フォームやブラウザのアドレス欄に入力する文字列とし，ターゲット以外の入力としてOSコマンドインジェクション攻撃のデータを用意した．

2 関連研究

Webアプリケーション攻撃の基本的な対策は，外部からの入力でSQLやOSのコマンドをサーバー側で実行させないようにすることである．しかしながら，何らかの理由でそのような対策がされていなかったり，出来なかったりすることが考えられるため，Webアプリケーションファイアウォール (WAF)[3] を導入することも重要な対策の1つであると言える．文献 [3] では文字列に含まれている記号の分布を調査することで攻撃を検知する仕組みを構築している．なお，Webアプリケーション攻撃の対策として特定の記号に対してエスケープ処理を施すことは基本的な対策と言えるが，アプリケーションの特性によってエスケープ処理が不都合になる場合もある．このような問題はWAFを導入する際にも起こり得ることであるため，本研究では，特定の入力データセットを入力ターゲットと想定した上で，その入力ターゲット以外の文字列を検知することができるかどうか調査した．

3 提案手法

3.1 データについて

本研究では，入力のターゲットを特定したときに，それ以外の入力を検知できるかどうか検討することを目的としている．そのために，通常，Webアプリケーションへ入力が想定されるURL，メールアドレス，郵便番号，住所，電話番号などの文字列を中心に，その他として，顔文字、プレーンテキスト、IPアドレス、アクティベーションコード，URLエンコードデータなどの入力をターゲットデータとする．一方，ターゲットとしないデータとしてはOSコマンドインジェクション攻撃のデータを用意した．使用したデータ数については，ターゲットデータについては，URLデータを1000個，その他を437個用意し，OSコマンドインジェクション攻撃は497個用意した．なお，本研究ではOSコマンドインジェクション攻撃のデータとして用意した497個のデータからランダムにいくつかのデータをサンプリングして，OSコマンドインジェクション攻撃を正しく検知できるかどうか調査することにした．

3.2 ターゲットデータの記号分布

ターゲットデータの文字列長はURLデータには長いものもあり，それ以外のものには短いものも含まれる．したがって，データに含まれる全ての文字をターゲットにするとデータが疎になることも考えられるため，ターゲットデータに含まれる特定の文字列に着目することにした．今回使用したデータには表1のような半角スペースを含む33個の記号が含まれていたため，本稿ではこれら33個の記号に着目して解析を進めていくことにする．なお，紙面の都合上33個の記号の中から5個の記号を選択して掲載した．

図1はターゲットデータの記号分布を表している．ターゲットデータのほとんどはURLデータであるため，スラッシュ，パーセント，ピリオド，アンダーバー，イコール，クエスチョンマークの頻度が高く，その他の記号としては，マイナス，カンマ，コロンも他の記号と比べると頻度が高いことが分かる．なお，ターゲッ

---

* Visualization of OS command injection attack by features other than attack data

† U Kyou , Chuo University
‡ Takeshi Matsuda , University of Nagasaki
§ Ryutaro Ushigome , Chuo University
¶ Michio Sonoda , Chuo University
‖ Jinhui Chao , Chuo University

記号	頻度	正規化頻度
:	1130	0.0621
/	4743	0.2608
sp	74	0.004
!	32	0.0017
” | 21 | 0.0011 |

表 1: ターゲットデータの記号分布の一部

トデータには URL 以外にもメールアドレスや電話番号などのデータも含まれており，どのようなデータセットに対しても図 1 と同様な記号の分布が得られる訳ではなく, また Web アプリケーションに通常入力されるデータは様々な種類があるため，データセットの選び方によって記号分布は様々な形状に変化する．

しかしながら，入力されるべき文字列がある程度特定できるターゲットデータの場合は，分布にある程度の誤差を許容しながら入力データを観測することで，ターゲットデータ以外の文字列の入力を観測できる可能性が考えられる．本研究では，図 1 のターゲットデータの記号分布を用いて，OS コマンドインジェクション攻撃のデータを検知できるかどうか調査した．

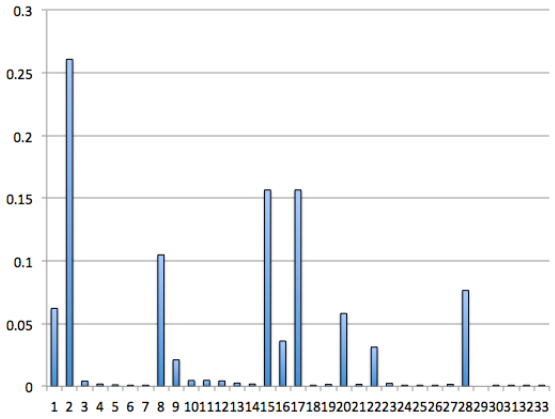

図 1: ターゲットデータの記号分布

3.3 攻撃検知に関する考察

図 1 に示したターゲットデータの分布を元にして，OS コマンドインジェクション攻撃の検知が可能であるかどうか検討する．本研究では，33 個の記号を用いて攻撃検知実験を行うため，文字列から 33 次元ベクトルを生成してターゲットデータの分布を構成している．実際には 497 個の攻撃データを準備しているが，紙面の都合のため 5 個の OS コマンドインジェクション攻撃のデータをランダムに抽出し，33 次元ベクトル化した

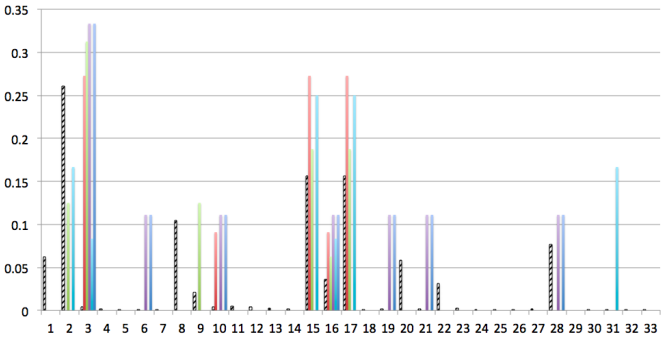

図 2: ターゲットデータと OS コマンドインジェクション攻撃データの比較

ものを図 2 に示す．図 2 は図 1 のターゲットデータの分布と攻撃データの分布を比較したものである．図 2 の棒グラフの斜線を引いているものがターゲットデータの分布であり，それ以外が攻撃データの分布である．図 2 より，本研究で想定したターゲットデータについては，シャープ (#)，シングルクォート (’)，不等号記号 (<, >)，パイプ (—) の使用頻度は 0 に近いため，これらの記号が観測されるかどうかを調査することで，OS コマンドインジェクション攻撃の検知が可能であると考えられる．なお，OS コマンドインジェクション攻撃についても，他のデータを検討することで異なる記号分布が得られる可能性があるため，より詳細な解析を行うことが今後の課題である．

参考文献

[1] IBM, “2017 年 上半期 Tokyo SOC 情報分析レポート,” https://www.ibm.com/blogs/tokyo-soc/wp-content/uploads/2017/09/tokyo_soc_report2017_h1.pdf (2017 年 12 月 29 日確認)

[2] IPA, “重要なセキュリティ情報一覧,” https://www.ipa.go.jp/security/announce/alert.html (2017 年 12 月 29 日確認)

[3] 園田 道夫, 松田 健, “攻撃特徴記号に基づく WAF 開発,” 情報処理学会論文誌, 56 巻, 9 号, pp.1826-1833 (2015)