3E-1

SVMを利用したWAFの検知手法の提案

伊波靖† 安里梓† 高良富夫‡

沖縄工業高等専門学校メディア情報工学科† 琉球大学工学部情報工学科‡

1. はじめに

近年，Web アプリケーションの利用が増える一方で，XSS(Cross-Site Scripting)攻撃や SQL インジェクション攻撃も後を絶たず，情報漏えいなどの深刻な被害が報告されている．Web アプリケーションをそれらの攻撃から守る方法の一つに，WAF(Web Application Firewall)の使用があるが，WAF は入力値検査の問題を抱えている．Moosa は，この問題を解決するため，WAF に ANN（Artificial Neural Network）を利用し，その有効性を示したが，False Positive の割合を減少させることを課題とした(1)．一方，我々は，Windows 系の OS において SVM(Support Vector Machine)を利用した不正プログラムの検知手法を提案し，その検知能力の高さと，False Positive を低減させることに有効であることを示した(2)．

本研究は，WAF の入力値検査に SVM を利用し，False Positive を低減させながらも，未知の攻撃を検知出来る手法の提案を行い，実験によりその有効性を示す．

2. Web Application Firewall(WAF)

2.1 概要

WAF とは，Web アプリケーションを含む Web サイトと利用者の間で交わされる HTTP による通信を検査し，攻撃などの不正な通信を自動的に遮断するソフトウェア，もしくはハードウェアである．

本研究では，WAF の機能の中でも XSS や SQL インジェクションの脆弱性対策の基本となる入力値検査の問題について考える．

2.2 入力値検査の問題点

入力値検査におけるホワイトリストは，入力可能なパラメタ全てを設定することが難しく，また手間もかかる上，設定漏れの恐れもある．一方，ブラックリストは，既知の攻撃であれば問題はないが，未知の攻撃に対しては攻撃を見逃す検知漏れの恐れがある．

3. Support Vector Machine(SVM)

SVM は統計的学習理論に基づく新しい 2 クラスのパターン認識手法であり，ニューラルネットワークなどの従来法と比較して汎化能力が高い点と最適解が求まる点に特徴があり，学習に用いていないデータに対しても高い認識率を示す．SVM がこのような特徴を示すのは，その学習に認識誤りと汎化性能の両面から最適化が行われ，これが 2 次の凸計画問題として定式化されているため最適解を求める事ができるためである．

SVM は学習の最適解として求められた分離超平面による線形識別を行っているが，学習用データを線形分離することが不適切な場合，学習データを元のパターン空間からより高次のパターン空間に非線形写像を行い高次元空間で分離超平面を構築し線形識別を行う．

4. 提案手法

WAF の入力値検査に SVM を利用する手法を提案する．具体的には，SVM を予めホワイトリストとブラックリストのデータを使って学習させ，2 クラスのパターン識別器を構成しておく．これによって，入力値検査の問題が解決され，検知率の向上が期待出来る．図 1 に提案手法の枠組みを示す．

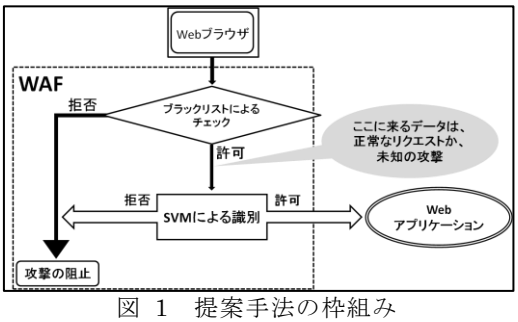

図 1 提案手法の枠組み

また SVM におけるデータの特徴表現は，ホワイトリストとブラックリストから得られた N-gram と N-gram の共起頻度によって行う．

5. 実験

本実験では，2 つの実験を行った．実験 1，2 共に使用した SVM は，TinySVM であり，線形カーネルを用いた．Solver Type は，C-SVM と One-Class-SVM の 2 つで実験を行う．

5.1 実験用データセット

ブラックリストのデータとして XSS，SQL インジェクションを各 100 個用意し，50 個を訓練用，50 個を評価用データとした．また，ホワイトリストのデータは，Web アプリケーションに入力されると考えられるデータを名前，住所，メッセージ，電話番号，パスワードの 5 つのカテゴリーに分類し，各カテゴリーとも 200 個用意，100 個を訓練用，100 個を評価用データとした．また，実験 2 で用いるため，名前，住所，メッセージについては，日本語と英語のデータ両方を揃えた．

A Proposal of SVM based detection method for WAF.
†Yasushi IHA, †Azusa ASATO
Dept. of Media Information Engineering, Okinawa National College of Technology
‡Tomio TAKARA
Dept. of Information Engineering, University of The Ryukyus

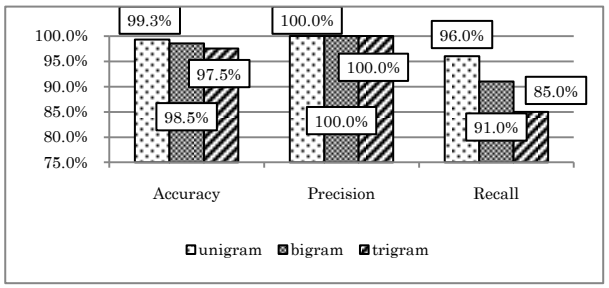

図 2　実験 1 の C-SVM の結果

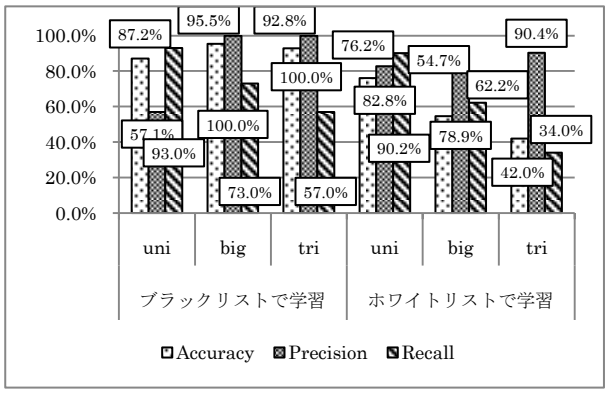

図 3　実験 1 の One-Class-SVM の結果

5.2　実験 1

実験 1 は，提案手法の有効性を示すために行った実験である．SVM に訓練用データを与え学習させた後，評価用データを用いて評価を行った．また，N-gram によって識別性能に変化が見られるかを調査するため，N の値を 1～3 まで変化させた．実験結果は，識別の Accuracy(適合率)の高さと F 値($F_{\beta=1}$)で評価した．図 2 に C-SVM の結果，図 3 に One-Class-SVM の結果を示す．また，図中の Recall は再現率，Precision は精度を示している．F 値は，unigram で 0.98，bigram で 0.95，trigram で 0.92 となり、良好な値となっている。

5.3　実験 2

実験 2 は，Moosa の研究において，英語のデータのみが扱われていたため，日本語の学習が SVM の識別結果に与える影響を調査するために行った．

英語のみで学習を行った SVM に英語，日本語の評価用データそれぞれを与えた結果と，日本語のみで学習を行った SVM に英語，日本語の評価用データそれぞれを与えた結果について調査する．なお，実験 1 の結果より，unigram を使用する．図 4 に C-SVM の結果，図 5 に One-Class-SVM の結果を示す．

6. 結論

本研究では，WAF の入力値検査に SVM を利用した検知手法を提案した．評価実験から，False Positive を低減させながらも，未知の攻撃を検知出来ることを示した．具体的には，実験 1 で行った C-SVM の unigram の実験において，適合率 99.3%，False Positive 0%と F 値 0.98 の結果より，SVM の有効性を確認した．False Negative の値は 4%であったが，

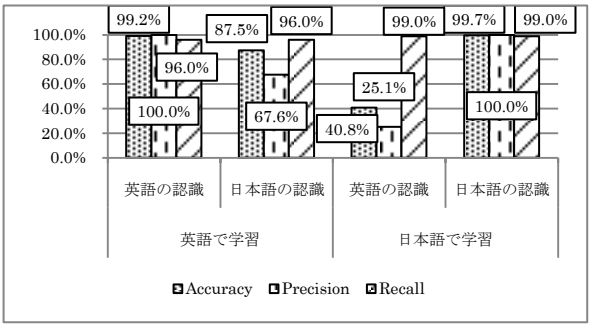

図 4　実験 2 の C-SVM の結果

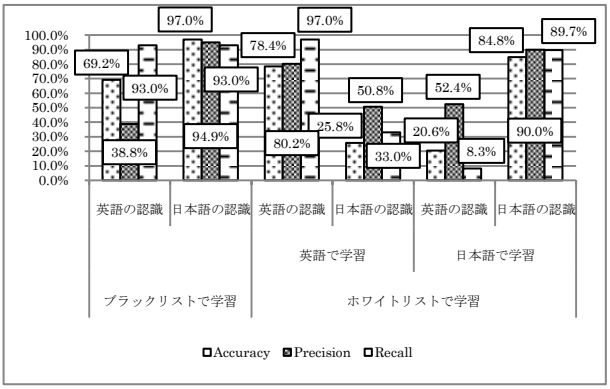

図 5　実験 2 の One-Class-SVM の結果

これは未知の不正リクエストの検知に対する値であること，またホワイトリストとブラックリストの訓練用データ数に差があり，SVM の学習に偏りがあったことから，値は妥当だと考える．

また，実験 2 では，C-SVM で日本語のデータを識別する際に，日本語の訓練用データを与えることで，適合率向上が期待出来ることを確認した．

今後の課題として，SVM のさらなる検知性能向上と提案手法の実装が挙げられる．まず検知性能向上については，訓練用データの数を増やすことと，特徴ベクトルの見直しが考えられる．また，実装については，オープンソースプロダクトで WAF を実現し，Apache のモジュールとして動作する mod_security へ組込むなどが考えられる．

参考文献

[1] Asaad Moosa,"Artificial Neural Network based Web Application Firewall for SQL Injection", World Academy of Science, Engineering and Technology, ISSUE 64, pp.12–21 (Apl. 2010)

[2] 伊波靖，高良富夫：危険なシステムコールに着目した Windows 向け異常検知手法，情報処理学会論文誌，Vol. 50, No. 9, pp.2173–2181 (Sep. 2009).