「潜在的要因を考慮した
SQL インジェクション攻撃検知システムの開発」
学位論文要旨

情報セキュリティ科学専攻
園田　道夫

インターネット技術の普及により，航空券の予約や書籍の購入，SNS によるコミュニケーションやネットバンキングなど，様々なサービスが Web 上で展開されるようになった．このような Web 上でサービスを展開するために開発されたアプリケーションのことを Web アプリケーションという．様々な Web アプリケーションが開発され，それらのサービスを利用するネットユーザーが増える一方で，Web アプリケーションを狙ったサイバー攻撃である，Web アプリケーション攻撃も技術の進化と共に巧妙化され，その被害も増え続けている．Web アプリケーションは，サービスのためのものやユーザーの個人情報を格納するためのデータベースを駆動させるものもあり，このような Web アプリケーションのデータベースの情報に不正にアクセスすることを目的とした SQL インジェクション攻撃の被害は特に深刻である．今後は，カメラやセンサーなどから取得されるデータを活用した IoT に関連する Web サービスの普及が見込まれるため，多くの情報を持つデータベースを安全に保護するための対策は急務であると言える．

一般的に，Web アプリケーションはユーザーの入力にしたがって動的に Web ページに表示するための内容を変更したり，ユーザーと取引をしたりする．その際に，Web アプリケーションがユーザーの入力を基にして SQL 文を組み立てるように開発されている場合，外部から任意の SQL 文を入力されることで Web アプリケーションのデータベースに不正にアクセスされる恐れがあり，このような攻撃手法のことを SQL インジェクション攻撃という．

SQL インジェクション攻撃の基本的な対策は，ユーザーの入力を SQL 文と解釈しないよう Web アプリケーションを開発することに加えて SQL の文法として特殊な意味を持つ記号を適切にエスケープ処理することである．しかしながら，開発上の様々な事由により脆弱性が作り出されることも考えられるため，SQL インジェクション攻撃から Web アプリケーションを防御する手法の開発

も重要である．SQL インジェクション攻撃に限らず，Web アプリケーション攻撃を防御するためのシステムのことを Web アプリケーションファイアウォール(WAF)と呼ぶ．WAF の基本的な動作原理は，攻撃のサンプルを抽象化した正規表現によるパターンマッチングであるため，防御の隙をつくバイパス攻撃が問題となる他，攻撃と構成が似ている正常な入力を誤検知する False Positive（第一種の過誤）が問題となっている．一方，攻撃を正常な入力と誤検知することを False Negative（第二種の過誤）というが，False Positive と False Negative の確率の大小はトレードオフの関係にあり，両者を同時に小さく保つことは一般的に難しいとされている．本研究の目的は，SQL インジェクション攻撃の文字列を数理的に分析して攻撃の特徴抽出を行うだけでなく，攻撃の特徴とよく似た正常な文字列を準備した上で正常の特徴抽出を行う手法を開発することで，False Positive と False Negative の両方を小さくし，未知の攻撃に対処できるだけでなく，バイパス攻撃の実現を困難にする WAF を開発することである．

初めに，SQL インジェクション攻撃の文字列を収集し，そのサンプル全体に含まれる要素を統計的に分析したところ，攻撃には区切り文字として意味をもつ特殊な記号が多く含まれていることを確認した．さらに，攻撃の特徴として使用する記号(以下，攻撃特徴記号という)を複数組み合わせて，文字列中にどの程度の割合の攻撃特徴記号が含まれるかを定量的に測るため，文字列中の攻撃特徴記号の含有率を収集したサンプルにおいて調べたところ，ほとんどの攻撃文字列において，半角スペース，シングルクォート，右側丸括弧と左側丸括弧の含有率が 0.1 程度であることを確認した．なお，収集したサンプルは SQL インジェクションの cheat sheet として公開されているものも含まれている．そこで，文字列中に含まれる攻撃特徴記号の含有率をzで表し，$0 < b < 1$を満足する実数のパラメータを用いて，その文字列が攻撃である確率をz^bとし，そうでない確率を$1 - z^b$とすると，攻撃特徴記号の含有率とその時の攻撃の可能性は

$$Pr(y|z,b) = (z^b)^y(1 - z^b)^{1-y} \quad \cdots ①$$

と確率モデルを用いて表現することができる．ただし，確率変数yは攻撃の場合は$y = 1$，正常の場合は$y = -1$と 2 値の値をとるものとして定義する．パラメータ$0 < b < 1$の値は，例えば，最尤推定法を利用する場合は，攻撃と正常のサンプルから得られる攻撃特徴記号のそれぞれの文字列における含有率から求めることができる．しかしながら，提案モデル①の尤度関数を最大化するパラメータは解析的に計算することができないため，本研究では，①の対数尤度関数のテー

ラー展開を用いて近似することで, 理論的に最尤推定量を導出した. 理論的に導出された最尤推定量は，数値計算の手法である Newton-Raphson 法で計算したものとほとんど一致することを確認した. これにより, 提案手法を実際の SQL インジェクション攻撃の検知システムに実装する際のアルゴリズムが単純化されることになる．さらに，攻撃特徴記号の含有率を用いた攻撃検知手法は，原理的には機械学習における線形分類器の一種であることを示し, Wang らによって開発された SCW(Exact Soft Confidence-Weight Learning)と呼ばれる線形分類器との比較実験を行なったところ，これまで実験を行なって経験的に設定していた攻撃特徴記号の含有率の攻撃検知のための閾値とほとんど同じ値の攻撃検知のための閾値が SCW のアルゴリズムからも導出されることを確認した.

すべての SQL インジェクション攻撃に上述で示した攻撃特徴記号が含まれるとは限らないが，よく知られている基本的な攻撃においては攻撃特徴記号を用いた SQL インジェクション攻撃の検知は有効であると言える．しかしながら，上述の攻撃特徴記号を含む正常な入力は数多く存在する. そこで本研究では, 潜在曲線モデルを応用した特徴抽出モデルを開発し，攻撃特徴記号同士の関連性を多項式で表現することで，攻撃と正常の両方の特徴を抽出する手法を提案した．なお，本研究では，半角スペース，シングルクォート，セミコロン，右側丸括弧，左側丸括弧の 5 つの記号を攻撃特徴記号として利用しているが，これはサンプルとして収集したデータの中でも，攻撃に頻出する記号の上位 5 つの記号を利用したものである．サンプルの収集の仕方によって選ばれる記号も異なる可能性もあるが, 基本的にこの 5 つの記号は SQL 文法として特殊な意味を持っているため，本研究ではこの 5 つの記号を攻撃特徴記号として使用することにした．このような 5 つの攻撃特徴記号の中でも，攻撃の文字列には半角スペースやシングルクォートの出現頻度は他の 3 つの記号と比較しても高く, かつ, 正常の文字列では頻度の低い一様な分布の状態であったため, 本研究では, 実験に使用したサンプルにおける攻撃と正常の両方の特徴を表現することができると考えられる 2 次の多項式を用いて特徴抽出をするための潜在曲線モデルを提案した. 攻撃特徴記号を増やしたりする場合は, その性質において特徴抽出に使用するモデルを変更することになるが，本研究では 2 次の多項式を用いた場合の潜在曲線モデルのパラメータ導出方法を示した. なお, パラメータの導出には最尤推定法を用いており，基本的には回帰分析におけるパラメータの導出方法と同じである. 以下, 簡単にモデルに潜在変数を導入する理由について述べる.

攻撃と正常の文字列を収集する際には，それぞれの文字列に攻撃・正常のラベルがつけられている状態であるとみなすことができるが，攻撃検知を行う場面では，その文字列が攻撃・正常のどちらのラベルかを正確に判断することは難しい．正確に攻撃・正常のラベルづけを行うには有識者の目視による確認が必要であるが，大量のデータの中からそのような作業を行うことは現実的でない．そのために保険的な対策である WAF を導入することは攻撃を未然に防ぐためにも重要なことであると言える．攻撃と正常を正確に分けることは困難であるが，どのような攻撃特徴記号がどの程度の含有率でどのような記号の組みが同時に観測されやすいかという情報を潜在変数に持たすことによってうまく攻撃検知できないかと考えたのが本研究の手法である．提案手法の有用性を示すために，オープンソースウェア(OSS)の WAF として有名である ModSecurity と攻撃検知の比較実験を行なった．その結果，False Negative については，提案システムは ModSecirity より僅かに大きくなったものの，False Positive については，提案システム は Modsecurity よりもはるかに小さくなった．なお，本研究で計算したパラメータ推定の結果に基づく攻撃検知システムを Apache のモジュールとして開発し，そのソースコードのコアとなる部分について以下の Web サイトで公開している．

http://matsudalab.office-server.co.jp/top/index.html

なお， ModSecurity のように OSS として公開されているシステムは見当たらないが，多くの研究において機械学習のアルゴリズムを用いた攻撃検知手法が提案されている．攻撃検知に機械学習の手法が期待される理由の一つに，未知の攻撃検知に対する汎化能力があることが挙げられる．本研究では，汎化能力が高いことで知られているサポートベクターマシン（SVM）を用いて攻撃検知実験を行なった． SVM による攻撃検知は，チューニング次第で提案手法や ModSecurity を用いたものと比較して特に良くも悪くもない結果が得られることを確認することができた．SQL インジェクション攻撃自体は ModSecurity が，正常文字列については提案手法が最も検知精度が高いという結果が得られた．今後の課題は攻撃特徴記号の数を増やしても False Positive と False Negative の両方を小さくできるように提案モデルを改良することである．