Vol.2012-MPS-87 No.30
2012/3/2

関連単語抽出アルゴリズムによる SQL インジェクション攻撃特徴抽出について

井上 繁之[†]　　松田 健[††]

本研究では，Web 検索における検索クエリ中のキーワードと関連性の強い語を抽出してクエリの拡張を行うための関連単語抽出アルゴリズムを SQL インジェクション攻撃検出アルゴリズムに適用する方法を提案する．その結果，攻撃特徴となる文字組合せを容易に選定することができ，さらに用意した攻撃サンプルに対しても高い攻撃検出率が得られることを実験により示した．

On the attack feature extraction of SQL Injection Attacks by the Related Word Extraction Algorithm

Shigeyuki Inoue[†]　and Takeshi Matsuda[††]

In this study, we proposed the method of SQL injection attacks detection applying the algorithm of the related word extraction. Then, we showed experimentally that attack feature characters are obtained easily and

1. はじめに

近年，データベース駆動型の Web アプリケーションを標的とする SQL インジェクション攻撃の増加が問題となっており，攻撃を自動検出する方法としてパターン認識や構文解析，ブラックリスト方式による検出法が開発されている．しかしながら，新種の攻撃に対応するために生じるリストの更新のために検出コストが増加してしまうため，攻撃に対処することが難しくなってきている．

この問題に対して文献 [1),2)]では，SQL インジェクション攻撃の文字列に含まれるセミコロンやシングルクォーテーションなどの記号を中心とする文字を攻撃特徴とする検出アルゴリズムとその数理モデルが提案されている．文献 [1),2)]では，攻撃検出に最適な文字を選定するために文字の組合せ方を変えて検出率が最大となる文字の組合せを試行錯誤で決めていた．

そこで本研究では，関連単語抽出アルゴリズム [3)]を用いて攻撃特徴文字の選定を行い，SQL インジェクション攻撃を検出する方法を提案した．

† サイバー大学
Cyber University
†† サイバー大学
Cyber University

2. 関連単語抽出アルゴリズムとその改良

2.1 関連単語抽出アルゴリズム

文献 [3)]では，Web 検索のクエリ拡張を行うための関連単語抽出アルゴリズムが提案されている．キーワード群 K に関するテキスト T における単語$w_l(l = 1,2,\cdots,\mathrm{o})$の評価は以下の手順で行われる．

(1) $k_i(i = 1,2,\cdots,\mathrm{m}), t_j(j = 1,2,\cdots,\mathrm{n})$をそれぞれテキスト T で出現するキーワード，センテンスとし，j_0をk_iを含むセンテンスt_{j0}の添数とする．それぞれのセンテンスt_jにおいて以下の計算を行う．

$$BV_{ki}(tj) = n - |j - j_0|,\ \mathrm{BV(tj)} = \sum_{i=1}^{m} BV_{ki}(tj) \ldots (a)$$

(2) センテンスt_jの出現位置による不公平を平滑化するため以下の計算を行う．

$$EBV(j) = \frac{1}{2n}\{n(n+2j-1) - 2j(j-1)\},\quad EBV(tj) = \frac{BV(tj)}{EBV(j)} \cdots \mathrm{(b)}$$

(3) センテンスt_jにある単語w_lの出現頻度tfについて以下の計算により，最終評価値$S(wl)$を求める．

$$EBV(tj) = EBV(w_l),\ V(w_l) = 1 + \frac{tf(w_l)}{n} log\, tf(w_l),\ S(w_l) = AveEBV(w_l) \times V(w_l) \cdots \mathrm{(c)}$$

2.2 SQL インジェクション攻撃特徴抽出への応用

関連単語抽出アルゴリズムではセンテンス間の距離を重要視するが，SQL インジェクション攻撃ではセンテンス間だけでなく単語間の位置関係も重要になるものと考え，センテンス間と単語間を重畳して評価するように式(a)を式(d)に，式(b)を式(e)に，式(c)を式(f)に変更した．

$$BV_{ki}(w_l) = \mathrm{o} - |\mathrm{l} - l_0|,\ \mathrm{BV}(w_l) = \sum_{i=1}^{m} BV_{ki}(w_l) \cdots \mathrm{(d)}$$

$$\mathrm{EBV(l)} = \frac{1}{2o}\{o(o+2l-1) - 2l(l-1)\},\ \mathrm{EBVW}(w_l) = \frac{BV(w_l)}{EBV(l)} \cdots \mathrm{(e)}$$

$$\mathrm{S}(w_l) = \mathrm{AveEBV}(w_l) \times \mathrm{V}(w_l) \times \mathrm{AveEBVW}(w_l) \cdots \mathrm{(f)}$$

さらに，文献 [3)]の関連単語抽出アルゴリズムではキーワードからの距離が 1 センテ

ンス離れるごとに評価値を 1 ずつ減じていたが，キーワードに近いセンテンスをより重要視するように式(a), (d)を以下の式に変更した．

$$BV_{ki}(\mathrm{tj}) = \mathrm{n} - |\mathrm{j} - j_0|^r \qquad (\mathrm{r} = 1.1,\ 1.2\cdots\cdot 2)$$
$$BV_{ki}(\mathrm{wl}) = \mathrm{o} - |\mathrm{l} - l_0|^p \qquad (\mathrm{p} = 1.1,\ 1.2\cdots\cdot 2)$$

3. 実験

3.1 実験条件

SQL インジェクション攻撃の攻撃特徴文字の抽出には，文献 [1),2)]で利用したサンプルの 20%にあたる 125 個のサンプルを用いた．なお，英字及び数値は複数の文字が連なった単語として，記号はそれ自体を単語として扱い，実験を行った．

3.2 実験結果

関連単語抽出アルゴリズムを実行するためには何らかのキーワードを選択する必要がある．”Datafile”や”半角スペース”などをキーワードとし，出現頻度の異なる文字で$S(w_l)$を計算したところ，いずれのキーワードにおいても表 1 に示す通り，上位 3 文字が変動することなく安定した結果を得ることが出来た．また，出現頻度が少ないキーワードで計算した場合においても得られた計算結果のうち最も評価値の高い文字をキーワードに加えて再度計算することで，ほぼ同等の結果を得ることが出来た．

3.3 考察

3.2 での実験により，文献 [1),2)]で試行錯誤の末に導き出された攻撃特徴文字を関連単語抽出アルゴリズムにより容易に抽出可能なことを確認することができた．また，出現頻度等に影響されることなく安定して上位 3 文字が抽出されることから，指定するキーワード及び攻撃文字列の内容に左右されることなく攻撃特徴文字を抽出することができる．更に，キーワードの選び方に依存するものの，コンマの出現頻度が右側丸括弧よりも多いのに対し，評価順位では右側丸括弧の方がコンマより高くなるという結果も得られた．

4. まとめ

本研究では，関連単語抽出アルゴリズム [3)]を用いて攻撃特徴文字の選定を行い，SQL インジェクション攻撃を検出する方法を提案した．この方法を他の SQL インジェクション攻撃のサンプルに適用すること，Web アプリケーション攻撃に適用することなどが今後の課題である．

表 1　キーワードによる評価順位の違い

	出現頻度	キーワード					文献[1),2)]
		Datafile	Parameter	Datafile 半角ｽﾍﾟｰｽ	Parameter 半角ｽﾍﾟｰｽ	半角ｽﾍﾟｰｽ	
半角ｽﾍﾟｰｽ	695	1	1	1	1	1	半角ｽﾍﾟｰｽ
ｼﾝｸﾞﾙｸｫｰﾃｰｼｮﾝ	200	2	2	2	2	2	ｼﾝｸﾞﾙｸｫｰﾃｰｼｮﾝ
ｾﾐｺﾛﾝ	129	3	3	3	3	3	ｾﾐｺﾛﾝ
ｺﾝﾏ	114	31	28	4	4	4	右側丸括弧
1	113	6	6	5	5	5	左側丸括弧
Select	101	5	5	6	6	6	－
右側丸括弧	93	9	7	7	7	7	－
左側丸括弧	85	21	20	8	8	8	－
Parameter	1	12	12	330	330	330	－
Datafile	1	8	10	332	332	332	－

参考文献

1) Takeshi Matsuda, Daiki Koizumi, Michio Sonoda, and Shigeichi Hirasawa：“On Predictive Errors of SQL Injection Attack Detection by the Feature of the Single Character”, Proceeding of 2011 IEEE International Conference on Systems, Man, and Cybernetics, pp.1722-1727, (2011).
2) Michio Sonoda, Takeshi Matsuda, Daiki Koizumi and Shigeichi Hirasawa：“On Automatic Detection of SQL Injection Attacks by the Feature Extraction of the Single Character”, Proceeding of 2011 International Conference on Security of Information and Networks, ACM, pp.81-86, (2011)
3) 大石 哲也，倉元 俊介，峯 恒憲，長谷川 隆三，藤田 博，越村 三幸:”. 関連単語抽出アルゴリズムを用いた Web 検索クエリの生成”, 電子情報通信学会論文誌. D, 情報・システム J92-D(3), pp.281-292, (2009)