文章编号:1674－2869(2015)07－0060－05

面向 SOAP 消息的 Web 服务注入攻击检测方法

吴长勤,王传安
安徽科技学院数学与信息学院,安徽 凤阳 233100

摘　要:面向 SOAP 消息的注入攻击对 Web 服务的发展前景和应用产生重要的影响,而目前尚缺乏有效的检测方法,为此我们提出了一种基于行为模式的注入攻击检测方法.该方法在 Web 服务访问日志行为建模阶段采用长度不同的特征子串来表示行为模式, 并通过序列模式的支持度来建立正常消息行为模式轮廓; 针对注入攻击复杂多变的特点, 进一步提出基于矩阵相似匹配结合相似判决阈值的注入攻击检测模型.实验结果表明,与多种其他经典检测方法相比,该检测方法在检测率和误报率等方面检测效果好,可有效地检测出注入攻击.

关键字:注入攻击;SOAP 消息;Web 服务;行为模式
中图分类号:TB35　　**文献标识码**:A　　**doi**:10. 3969/j. issn. 1674-2869. 2015. 07. 013

0　引　言

Web 服务是一个向其它应用过程或程序提供数据和服务的分布式模块化组件,它通过 XML、SOAP 等新技术实现各种服务之间的互操作[1]. Web 服务具有高度互操作、松散耦合和可重用等特点,已成为网格和云计算等计算技术的重要实现方式之一.

目前 Web 服务通信是以传输 SOAP 消息为基础的,由于 SOAP 协议在设计时并没有过多的考虑其安全性问题,从而导致 SOAP 消息在传输过程中极易受到诸如 SQL 注入攻击和 XML 注入攻击等其他形式的命令注入攻击,因此保证 SOAP 消息传输的安全性是当前研究的一个重要问题[2]. **文献[3]**描述了 Web 服务面临的安全威胁与状态,利用 OWL-S 与本体建立了服务安全攻击本体. Nils Gruschka 等人[4]指出消息的机密性、完整性和用户验证对于服务平台安全性的作用很小,从而提出采用 XML Schema 中的消息对接受到的 SOAP 消息进行确认,以此发现具有注入攻击的服务调用. **文献[5]**在研究中定义了一个消息安全本体来对 SOAP 消息进行本体验证,进而对攻击进行检测. 杨晓晖等人[6]提出基于信任度量的跨域访问控制模型,来实现 Web 服务访问安全.

笔者在充分考虑正常 Web 访问行为的特点后, 采用长度不同的多种特征子串来表示正常 SOAP 消息行为,并通过支持度来建立正常消息行为模式,最后利用提出的检测模型对注入攻击进行检测.

1　注入攻击检测方法

文中提出的注入攻击检测方法的实现过程可分为正常行为建模和和攻击检测两个阶段,具体检测模型如图 1 所示. 正常行为建模的主要任务是建立正常的 SOAP 消息的行为模式. 攻击检测阶段的任务是根据建立的正常行为模式,利用特定的检测模型来识别当前 SOAP 消息中的异常行为.

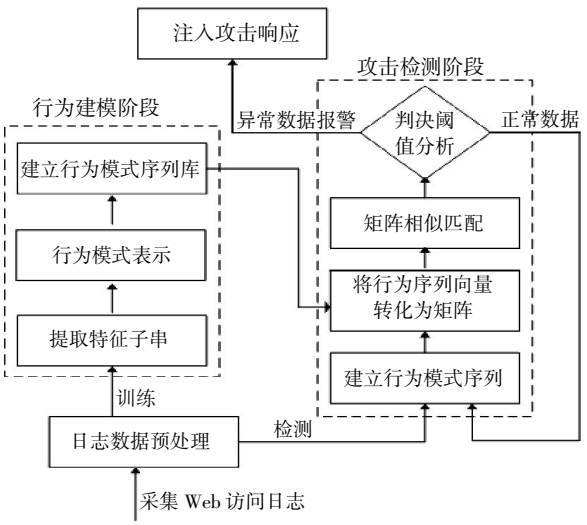

图 1　注入攻击检测模型

Fig.1　The detection model of injection attack

1.1　行为建模

Web 访问详细的记录了客户的 SOAP 信息行

收稿日期:2015-06-10
基金项目:安徽科技学院青年科研研究项目(ZRC2014426);安徽省自然科学研究项目(KJ2013Z048).
作者简介:吴长勤(1962-),男,安徽肥东人,副教授,研究方向:计算机应用技术.

为,通过观察 Web 访问日志,可发现消息中会大量重复出现一些字符串,将这些重复出现的字符串称为特征子串,其特征相对稳定,可代表 SOAP 消息的行为模式[7]. 行为建模阶段的主要工作如下.

(1) 获取训练数据: 将无注入攻击时产生的 Web 访问作为训练数据 $E=\{e_1,e_2,\cdots,e_n\}$,其中 e_i 表示按时间顺序排列的第 i 个 SOAP 消息事件,n 为 SOAP 消息事件序列长度.

定义 1:Web 访问日志记录的一次会话或一次执行过程称为 SOAP 消息事件,记为 $e=<l_a,t>$,其中 $l_a=<s_1,s_2,\cdots,s_m>$表示该 SOAP 消息事件中的 m 个字符,t 表示记录 SOAP 消息事件的时间.

(2)借鉴最长前缀提取算法[10]思想,采用最长前缀分析法提取训练数据 E 中所有长度大于 1 的重复出现的特征子串,组成特征子串集合 $T=\{t_1,t_2,\cdots,t_k\}$,t_j 表示第 j 个特征子串,k 为长度.

(3) 将训练数据 E 中的 SOAP 消息事件序列转换为按时间排序的 n 个长度分别为 $l(1)$, $l(2)$, $\cdots l(n)$的特征子串序列流 $C=\{C_1,C_2,\cdots,C_n\}$,其中 C_i 表示 eiSOAP 消息事件中包含的所有特征子串,记为 $C_i=\{t_j,t_{j+1},\cdots,t_m\}$.

(4) 计算集合 T 中每个特征子串在训练数据序列 E 中的支持度.

定义 2: 一个特征子串在序列 E 中的支持度等于该子串出现的次数除以所有特征子串总计出现次数,即:$\text{support}(t_i)=\text{number}(t_i)/\text{Number}(T)$,其中 $\text{Number}(T)$为集合 T 中所有特征子串出现的次数和.

(5)从 n 个特征子串流中,分别按照相应的支持度提取正常 Web 访问行为模式序列 $L(i)$.

考虑到不同的特征子串表示的行为模式意义可能不尽相同,其出现的频率也会有较大的区别,不能采用单一支持度,在文中设置了 k 个相应于不同特征子串的最小支持度 $\text{minsup}(t_1)$, $\text{minsup}(t_2)$, $\cdots,\text{minsup}(t_k)$. 对于 $1\leqslant i\leqslant k$,将 C_i 中等于或大于相应最小支持度 $\text{minsup}(t_i)$的特征子串提取出来,构成满足支持度要求的行为模式序列 $L(i)$.

(6)构建正常行为模式轮廓

将 S 个满足最小支持度要求的序列 $L(1)$, $L(2),\cdots L(S-1),L(S)$存储起来,用于描述正常行为模式轮廓,同时也作为下一步进行注入攻击检测的正常行为模式轮廓 $L=\{L(1),L(2),\cdots,L(K)\}$.

1.2　注入攻击检测

攻击检测阶段的任务是根据建立的正常行为模式, 利用特定的检测模型来识别当前 SOAP 消息事件中的异常行为. 攻击检测阶段的主要工作如下.

(1)获取测试数据:选取监测时间内 Web 访问日志作为测试数据,并进行预处理. 设预处理后得到的日志数据字符串序列为 $E'=\{la_1, la_2, \cdots, la_n\}$,其中 la_i 为按时间顺序排列的第 i 个字符串.

(2)按照行为建模阶段提出的方法,提取测试数据 E'中的 K 个行为模式序列,并以此构建当前日志行为模式轮廓 $L'=\{L'(1),L'(2),\cdots,L'(S)\}$.

(3)将两阶段构建的行为轮廓 L 和 L'分别转化成 $k\cdot n$ 的模板矩阵 $\boldsymbol{M}$ 和检测矩阵 $\boldsymbol{Q}$.

模板矩阵 $\boldsymbol{M}=(M_1,M_2\cdots M_k)^{\mathrm{T}}$, 其中 $\boldsymbol{M}_i$ 表示序列 $L(i)$转化的由 n 个 0 和 1 组成的行向量,即$\boldsymbol{M}_i=(a_1,a_2,\cdots,a_n)$,其物理意义表示:若子串集 T 中的某个特征子串在 $L(i)$中出现,将相应的矩阵位置元素置为 1,否则置为 0. 如 $L(i)$由集合 T 中的第 1,2,4 个特征子串组成,即 $L(i)=\{t_1,t_3,t_5\}$,则转化的行向量 $M_i=(1,1,0,1,0,0,\cdots,0)$. 同理,将序列库 L' 转化为行向量表示的检测矩阵 $\boldsymbol{Q}$.

(4)矩阵相似性度量:使用矩阵相似度来衡量检测矩阵 $\boldsymbol{Q}$ 与模板矩阵 $\boldsymbol{M}$ 之间的匹配程度. 下面引入矩阵间相似度的概念[8],设 $m\times n$ 矩阵全体为 $\boldsymbol{C}^{m\times n}$,假如 $\boldsymbol{A}\cdot\boldsymbol{B}\in\boldsymbol{C}^{m\times n}$,矩阵内积可定义为

$$<\boldsymbol{A}\cdot\boldsymbol{B}>=\text{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}) \tag{1}$$

式(1)中 tr(.)表示矩阵主对角线元素的和,而矩阵内积表示两矩阵相对应位置元素的乘积累加和,由矩阵内积可导出范数

$$\|\boldsymbol{A}\|=\sqrt{<\boldsymbol{A}\cdot\boldsymbol{B}>} \tag{2}$$

若 $\boldsymbol{A}$、$\boldsymbol{B}$ 为实数矩阵,则$|<\boldsymbol{A}\cdot\boldsymbol{B}>|\leqslant\|\boldsymbol{A}\|\cdot\|\boldsymbol{B}\|$,当且仅当 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 两矩阵完全相似时,$|<\boldsymbol{A}\cdot\boldsymbol{B}>|=\|\boldsymbol{A}\|\cdot\|\boldsymbol{B}\|$.

可知,$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 两矩阵的相似度为

$$\sin(\boldsymbol{A},\boldsymbol{B})=\cos\theta=\frac{|<\boldsymbol{A}\cdot\boldsymbol{B}>|}{\|\boldsymbol{A}\|\cdot\|\boldsymbol{B}\|} \tag{3}$$

式(3)中 θ 为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 两矩阵之间的夹角,当 $\theta=0°$ 时,表示 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相似性最好;当 $\theta=90°$ 时,$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 则完全不相似.

为计算检测矩阵 $\boldsymbol{Q}$ 与模板矩阵 $\boldsymbol{M}$ 之间的匹配程度,需将公式(3)改为

$$\text{sim}(\boldsymbol{M},\boldsymbol{Q})=\frac{|<\boldsymbol{A}\cdot\boldsymbol{B}>|}{\|\boldsymbol{A}\|\cdot\|\boldsymbol{B}\|} \tag{4}$$

当检测矩阵 $\boldsymbol{Q}$ 与模板矩阵 $\boldsymbol{M}$ 之间的匹配度 $\text{sim}(\boldsymbol{M},\boldsymbol{Q})$小于给定的判决阈值 w 时,则将矩阵 $\boldsymbol{Q}$ 标记为异常,并发出异常警报,最后将报警 SOAP 消息事件交由专家系统并结合人工进行验证,进而判定 Web 访问是否遭受注入攻击.

(5)判决阈值选择:选择合适的判决阈值是决定检测率和误报率的关键因素. 参照**文献**[9]中交叉验证的方法来确定文中的判决阈值 w, 将获得的 Web 访问日志按一定比例分成两部分,一部分用于获取行为模式, 而另一部分用于确定判决阈值和测试误报率,如此反复交叉的测试,并在测试过程中通过调整判决阈值来获得误报率与不同判决阈值间的对应关系,最终将期望误报率所对应的判决阈值 w 作为检测注入攻击的阈值.

2 实验与分析

本文的试验是在安徽科技学院校园网中进行的, 采集到的数据集既有校园网内部的 Web 访问数据,也有来自于或传输到教育网上的数据,并利用 C++语言完成文中相关的算法和实验. 具体实验环境见图 2 所示.

图 2 中的环境配置如表 1 所示.

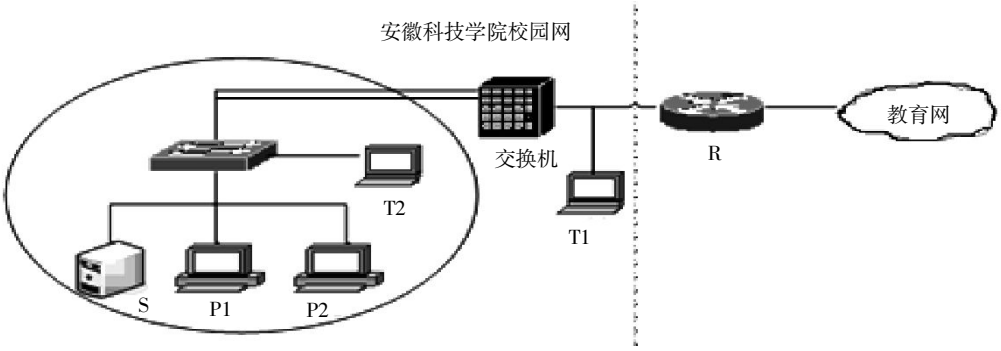

图 2 实验环境图

Fig.2 The platform of experimental environment

表 1 环境配置表

Tabel1 The configuration of experimental environment

名称	软件	作用
R		安徽科技学院出入口路由器
T1	Win7, Logparser	获取教育网流经江 R 的 Web 访问日志
T2	Win7	编写并向 S 发送正常 SOAP 消息
P1	Win7,注入	向 S 发送嵌有注入攻击
P2	攻击程序	命令的 SOAP 消息
S	Linux	注入攻击检测服务器

实验所采用的 Web 访问数据集如下:

训练数据集:将 T2 产生的 1 000 个 SOAP 消息记录作为训练数据集,用来建立正常的 SOAP 消息的行为模式.

测试验证数据集:将 T1 采集到的 Web 访问日志和 P1、P2 产生的 100 个 SOAP 消息记录作为测试数据集,用来验证文中提出的注入攻击检测效果.

根据训练数据集,运行特征子串提取算法,提取了 35 个用于表示行为模式的特征子串,进一步采用文中提出的行为模式序列提取方法,获得了21个行为模式序列库 $L=\{L(1),L(2),\cdots,L(21)\}$.

在建立正常行为模式轮廓后,对测试验证数据集采用文中提出的注入攻击检测方法进行检测. 为选择最佳的检测判决阈值,可通过已有的训练数据并通过判决阈值选择法来确定. 图 3 给出了实验过程中选择判决阈值 w 的曲线,图中的实线为未进行注入攻击对应的判决阈值曲线,而虚线为注入攻击后的判决阈值曲线. 从图 3 中可以看出,当 w=0.55 时,正常 SOAP 消息行为模式下的判决阈值曲线与注入攻击后对应的判决阈值曲线几乎没有的交叠部分,具有较好的区分度,因此文中将 0.55 作为最佳判决阈值.

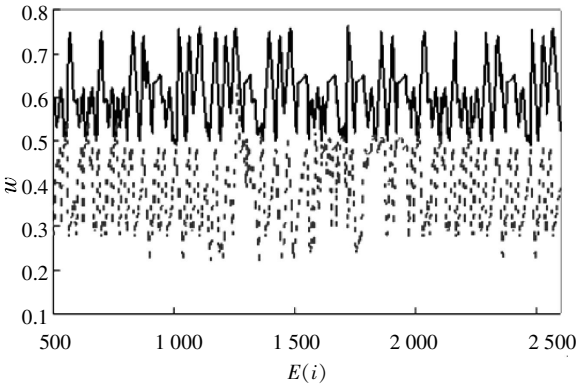

图 3 判决阈值曲线

Fig.3 The curve of decision threshold

注: —— 正常数据; - - - - 攻击数据

同时,还采用文中实验用的 Web 访问数据集分别对**文献**[10]中基于序列对比的检测方法、**文献**[11]中基于本体的检测方法以及**文献**[12]中基于文本聚类的攻击检测方法进行了对比实验,在检测率、误报率和实验时间方面进行了对比,实验对比结果见表 2 所示.

表 2　不同检测法的检测结果比较

Tabel 2　The comparison results of detection effect

	检测率/%	误报率/%	实验时间/s
本文检测法的实验结果	89.95	0.053	35
序列对比法的实验结果	79.93	0.33	28
本体检测法的实验结果	84.01	0.055	30
文本聚类法的实验结果	85.98	0.15	41

从表 2 中可以看出,文中提出的检测方法在检测率上明显优于其他 3 种检测方法,且误报率也明星低于序列对比和文本聚类的检测方法. 由于文中提出的检测方法需要先建立正常行为模式,然后再进行验证检测,所以在实验时间方面稍高于序列对比和本体检测法,但是其他两项检测指标具有明显的比较优势. 由实验对比结果可见,文中提出的注入攻击检测方法是有效的,具有很好的综合检测性能.

3　结　语

提出了一种新的面向 SOAP 消息的注入攻击攻击检测方法,该方法充分考虑了正常 Web访问行为的特点,采用长度不同的多种特征子串来表示正常消息行为,并通过支持度来建立正常消息行为模式,最后利用特定的检测模型来检测SOAP消息中是否含有注入攻击. 实验结果表明,文中提出的检测方法可有效的检测注入攻击,且具有很好的检测效果.

致　谢:

在此对文中实验和测试等提供支持和帮助的安徽科技学院计算中心表示感谢,并对在研究过程中提供的指导的各位老师帮助表示感谢!

参考文献:

[1] 刘玲霞,王东霞,黄敏桓. 一个 Web 服务可信体系结构[J].计算机科学,2014,12(41):30-32.
LIU ling-xia,Wang Dong-xia,Huang Min-heng. Trustworthy Architecture for Web Services[J]. Computer science,2014,12(41):30-32.(in Chinese)

[2] SABBARI M,ALIPOUR H S.A security model and its strategies for web services [J]. International Journal of computer Applications,2011,36(10):24-31.

[3] JENSEN M, GRUSCHKA N, HERKENHONER R, et al. Soa And Web Services: New Technologies, New Standards - New Attacks [C]//2007. ECOWS '07. Fifth European Conference on. Web Services, 2007: 35-44.

[4] NILS Gruschka, Norbert Luttenberger.Protecting Web Services from DoS attacks by SOAP Message Validation [J].IFIP International Federation for Information processing 2006, 9(201):171-182,

[5] DIEGO Z G,MARIA B F.Ontology -based Security Ploicies for supporting the management of Web Service Business Processes [C]//The IEEE international Conference on Semanitic computing.2008.

[6] 杨晓晖. 一种基于信任度量的 Web 服务跨域访问控制模型 [J]. 山东大学学报 (理学版),2014,9(49):115-122.
YANG Xiao-hui. A cross-domain access control model of Web service based on trust measurement[J].Journal of Shandong University(Natural Science),2014,9(49):115-122.(in Chinese)

[7] 王传安. SCADA 系统通信网中的高级持续性攻击检测方法[J].计算机科学与探索,2015,3(9):352-259.
Wang Chuan-an. Detection of Advanced Persistent Threats in SCADA Communication Network[J]. Journal of Frontiers of computer science and technology, 2015,9(27):352-359.(in Chinese)

[8] 翟东海.基于矩阵相似度的最佳样本块匹配算法及其在图像修复中的应用[J].计算机科学,2014,1(41):307-310.
ZHAI Dong-hai. Optimal Exemplar Matching Algorithm Based on Matrix Similarity and its Application in Image Inpainting[J]. Computer science,2014,1(41):307-310.(in Chinese)

[9] 田新广.基于 shell 命令和多重行为模式挖掘的用户伪装攻击检测[J].计算机学报,2010,33(4):697-705.
TIAN Xin-guang. Masquerade detection based on shell commands and multiple behavior patter mining [J]. Chinese Journal of computer,2010,33(4):697-705.(in Chinese)

[10] 孙义,胡雨霁,黄皓. 基于序列比对的 SQL 注入攻击检测方法[J].计算机应用研究,2010,9(27):3525-3528.
SUN YI,HU YU-JI,HUANG Hao. Method of defense SQL injection attacks based on sequence alignment [J]. Application Research of Computers,2010,9(27):3525:3528.(in Chinese)

[11] 陈军. 基于本体的 Web 服务攻击检测技术研究[J].计算机应用,2011,6(31):1515-1520.
CHEN Jun. Research of Web Services attack detection based on ontology [J]. Journal of Computer Applications,2011,6(31):1515:1520.(in Chinese)

[12] 杨晓峰. 基于文本聚类的网络攻击检测方法 [J].智能系统学院,2014,1(9):1-7.
YANG Xiaofeng. Text clustering based Web attack detection method[J]. CAAI Transactions on Intelligent Systems,2014,1(9):1-7.(in Chinese)

Detection of SOAP message-oriented injection attacks on Web services

WU Chang-qin, WANG Chuan-an

College of Mathematical and Information, Anhui Science and Technology University, Fengyang 233100, China

Abstract: SOAP message-oriented injection attacks have great impacts on prospects and applications of Web services; however, there is not an effective injection attacks detection method now. We proposed a method to detect injection attacks based on behavior patterns. In this method, the behavioral patterns of the legal behavior were characterized by characteristic substring sequences of different lengths, and the sequence supports were used to construct the normal behavior profiles in modeling stage of Web-access log message behavior. According to the complex and volatile features of injection attacks, a detection model based on matrix similarity matching and decision threshold was proposed. Compared with other classical detection models in experiments, the proposed method has better detection rate and false positive rate, showing good detection effects on injection attacks.

Keywords: injection attacks; SOAP messages; Web services; behavior profiles

本文编辑:陈小平