基于 IIFA 算法的 SOA 安全模型①

蔡 亮，王 兵，李 辉

(北京化工大学 信息与科学技术学院，北京 100029)

摘 要: 针对 SOAP 消息传输过程中可能引发安全威胁，运用关联规则对 SOAP 消息进行挖掘，建立 IIFA-SOA 安全模型。基于系统实时性要求，采用索引生成频繁集算法(IIFA)进行关联规则的挖掘。实例表明，所建立的 IIFA-SOA 安全模型，可以进行安全威胁的预测与销毁，提高了系统的安全性。

关键词: SOA；SOAP；关联规则；WS-Security；IIFA

SOA Security Model Design Based on IIFA Algorithm

CAI Liang, WANG Bing, LI Hui

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract：For the SOAP message transmission process may lead to security threats. In this paper we address to mine SOAP messages using association rules, and establish IIFA-SOA security model. Based on the requirements of system real-time, we introduce Index Induce Frequent algorithm for frequent itemsets mining association rules. Experimental results show that the IIFA-SOA security model can make the prediction and elimination of security threats, accordingly improve system security.

Keywords：SOA; SOAP; association; WS-Security; IIFA

1 引言

基于 SOA 系统设计的一个主要挑战就是请求处理的安全性，它影响在 SOA 环境[1]中的每一个服务和应用程序。尽管 WS-Security[3]标准已经建立了组成这样一个框架的一个安全组扩展，也进一步扩展了一系列补充规范与标准，例如 WS-Policy[6]、WS-Authorization[7]、WS-Federation[8]。但是那些不知道应该应用那个标准或者是不知道这个标准是否能保护 Web 服务的 SOA 开发者可能会拒绝使用这些 Web 服务安全标准。

本文描述了一个嵌入服务内确保 SOA[16]服务安全的智能化安全模型[13]。主要研究构建在安全服务层和消息安全层的安全模型，消息安全层由数字签名、数据加密、安全令牌[14]三部分组成。为保证消息层信息不包含攻击，安全服务层协助提供安全策略。安全模型和身份验证协同工作保证用户身份合法性和服务提供者免受任何可能的网络攻击。同时安全模型根据历史数据利用 IIFA 算法产生一组关联规则。最后，安全服务利用数据挖掘模型预测并阻止 SOAP[10]消息攻击，并根据规则提出一个有效的消息安全级别。

2 相关概念

2.1 频繁项集

设 $I=\{i_1,i_2,\ldots,i_m\}$是项的集合。设任务相关的数据 D 是数据库事务的集合，其中每个事务 T 是项的集合，使得 $T \subseteq I$。每一个事务有一个标识符，称作 TID。设 A 是一个项集，事务 T 包 A 当且仅当 $A \subseteq T$。关联规则是形如 $A \Rightarrow B$ 的蕴涵式，其中 $A \subset B$, $B \subset I$ 并且 $A \cap B = \varnothing$。规则 $A \Rightarrow B$ 在事务集 D 中成立，具有支持度 s，其中 s 是 D 中事务包含 $A \cup B$ 的百分比，其概率为 P($A \cup B$)，即是 support($A \Rightarrow B$)=P($A \cup B$)项的集合称为项集(itemset)。包含 k 个项的

① 收稿时间:2010-04-20;收到修改稿时间:2010-05-22

项集称为 k 项集。项集满足小最小持度 min_support，如果项集的出现频率≥min_support 与 D 中事务总数的乘积。如果项集满足最小支持度，则称它为频繁项集(frequent_itemset)。

2.2 IIFA 算法

Apriori 算法采用迭代的方式，产生候选集，再扫描数据库来计算候选集的频度(支持度数)，消除非频繁项集，当数据库宽度较大时，会产生较庞大的候选集。

因此，本文引入 IIFA 算法。IIFA 算法:为减少候选集的生成，引入索引数组概念：Index[]的元素由一个二元组组成(F1，G(C2-F2))。其中，Fi 为频繁 i-项集，Ci 为候选 i-项集，C2-F2 即为非频繁 2-项集，而 G(C2-F2)则是与频繁 1-项集构成的非频繁 2-项集的项集。例：若频繁 1-项集为{m},{n},{i},{j},{k}；非频繁 2-项集为{n,i}{n,j}{I,j},{j,k}。

3 安全模型设计

3.1 安全模型简介

本文的目的是使用数据挖掘技术提高 SOA 中的 Web 服务安全及相关安全策略。具体来说，数据挖掘并不是一种解决 SOA 安全性的新方法，本文利用数据挖掘衡量由 WS-Security 及其相关策略定义的安全功能的准确性，即可产生一种新的安全策略，具体过程是分析收到的 SOAP 消息，验证消息属性及其相关策略并将结果保存在 SMDB(Security Model database)中，然后运用关联规则算法进行数据挖掘建立安全模型。模型通过检验攻击与安全属性及策略之间的关联产生新模型弥补先前策略存在的不足。例如，当系统有一个验证请求时，身份验证令牌本身存在的漏洞可能引发一个安全攻击。通过将安全攻击与特定的安全策略进行匹配，挖掘模型可以产生一条规则，管理员根据规则提供的安全威胁描述，部署一个更强的令牌或者混合令牌进行安全策略的调整，提高系统的安全级别。

综上所述，数据挖掘方法用于安全调整有以下几个优点：

(1) 通过分析 SOAP 消息的长度及其解析时间可以预测潜在攻击。因此，任何可能威胁系统的攻击将被删除。

(2) 数据挖掘根据 SOAP 消息请求及这些消息可能引起的攻击把用户分成安全、怀疑、禁止三个不同的信息安全级别。安全用户是指请求服务用户没有发出过任何有害信息并且也没用攻击过系统，这些服务请求者被认为是可信任的。怀疑用户:请求服务用户发送的消息中很可能包含危险的内容，这些用户需要服务器进一步确认。禁止用户:请求服务用户发送过攻击服务器的信息，这些用户将被拒绝访问。

(3) 在新的安全策略发布之前测试策略本身潜在的漏洞。安全管理员通过挖掘安全服务的结构来调整策略使之到达安全要求。例如，管理员采取调整令牌类型、加密算法等措施，后面章节中将给出详细说明。

3.2 安全模型

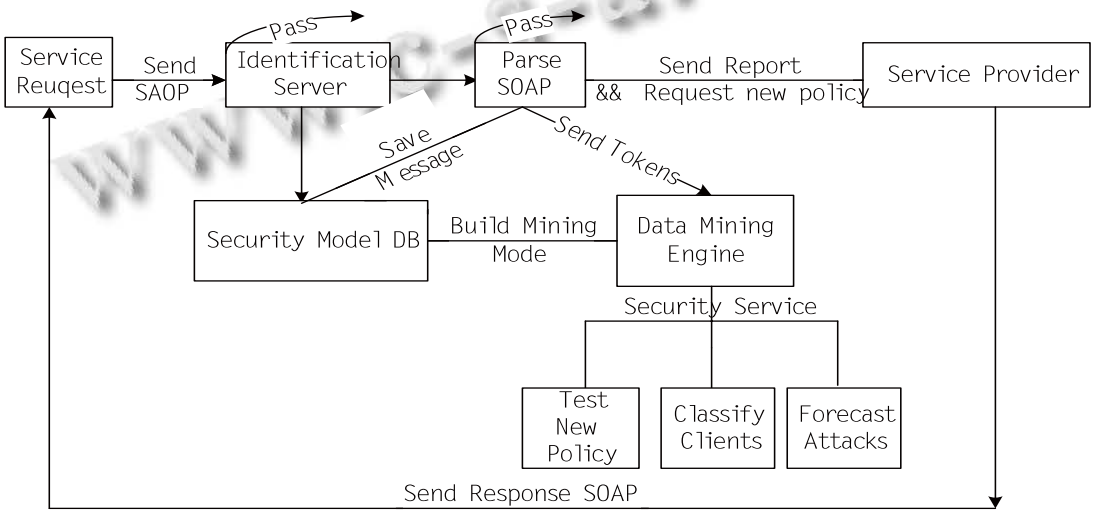

图 1 SOA 系统安全模型

SOA 安全模型如图 1 所示，可分为以下几个步骤：

(1) 服务请求者向服务器发送 SOAP 消息请求[15]。

(2) 服务器验证服务申请者身份是否合法。

(3) 智能安全服务与身份验证服务同时对收到的 SOAP 消息请求进行处理，主要实现以下功能：

① 解析 SOAP 消息

② 将 SOAP 消息的安全特征存储到数据库

③ 运用所建造的挖掘模型预测 SOAP 消息可能引发的攻击

④ 根据已有数据对请求者进行分类

(4) 向服务器产生报表，将 SOAP 消息响应反馈给消息请求者

(5) 安全管理员在服务器端测试策略的有效性

上述安全服务的主要功能包括：请求者身份验证、SOAP 消息请求解析、安全特征的存储、潜在攻击的预测及销毁、基于数据挖掘模型动态安全策略的构建。

4 实验与分析

4.1 实验环境

将所建安全模型运用于基于 J2EE 开发的某多 Web 服务系统中。Web 服务[4]提供大量不同长度和不同解析时间的 SOAP 消息，系统中已发布的安全策略包括签名、加密签名两种。仿真过程中，假设随机选取包含 Web 攻击的 SOAP 消息。这些 Web 攻击是在 SMDB 中定义和存储的，他们具有不同的令牌类型、数字签名、加密算法。

4.2 IIFA 挖掘模型与实验

依前所述，本文在 ODM[5]中采用 IIFA 算法构建初始安全模型。为确保 SOA 系统安全最简单的策略是采用虚拟专用网(CPN)来传输服务请求，这种策略只能满足简单、粗粒度的安全性要求。另一种策略是基于应用服务器，立足于服务实现平台中的 SOA 安全功能，服务平台可以根据实际最终用户来保留安全环境，这种策略便于实现高级的授权策略。但如果用户有多个平台，所需的配置及集成工作可能会导致在工时上所花的成本相当于或超过购买及配置 SOA 专门产品所需的成本。由于 CPN 策略不能处理复杂的应用，基于应用服务器的策略不适用于多个平台。Apriori[19]算法采用迭代的方式，产生候选集，再扫描数据库来计算候选集的频度(支持度数)，消除非频繁项集，当数据库宽度较大时，会产生较庞大的候选集，且因多次扫描数据库将导致大量的 I/O 开销,FP-growth 算法数据库投影到一颗频繁树 FP-tree 上，然后通过寻找 FP-tree 中节点路径来穷尽挖掘频繁项集，FP-growth 算法不产生大量候选集，将数据库的主要信息以高比率压缩的 FP-tree 的形式存放在内存中，减少多次扫描数据库的 I/O 开销，但数据库庞大时，FP-tree 深度加深由此将占据较大内存空间。本文提出采用索引生成频繁集算法(IIFA)，引入索引并在此基础上生成候选集，减少候选集的数量，无需占据大量内存空间，大大提高了频繁项集的挖掘效率。

IIFA 算法:为减少候选集的生成，引入索引数组概念：index[]的元素由一个二元组组成(F_1，$G(C_2-F_2)$)。其中，F_i 为频繁 i-项集，C_i 为候选 i-项集，C_2-F_2 即为非频繁 2-项集，而 $G(C_2-F_2)$则是与频繁 1-项集构成的非频繁 2-项集的项集。

算法思想：该算法是在 $F_{k-1}\times F_1$ 算法的基础上，利用 $G(C_2-F_2)$进行过滤，有效避免了产生较多冗余候选集并减少了产生重复候选集的可能性。

产生频繁项集算法如下：

设 $I=\{i_1,i_2,\ldots,i_m\}$是项的集合，C_k 为候选 k-项集的集合，而 F_k 为频繁 k-项集的集合。

基于 Index[]的频繁项集产生。

(1) $k=1$

(2) $F_k=\{i|i\in I\wedge\sigma(i)\geq N\times\min\_support\}$ {发现所有的频繁 1-项集}

(3) $k=k+1$

(4) $C_k=F_{k-1}\oplus F_{k-1}$ {产生 2-项候选集}

(5) $F_k=\{f_k\in C_2\wedge\sigma(f_k\geq N\times\min\_support\}$

(6) generate-index(F1,C_2-F_2)

(7) while(k<n||(C_k,$\sigma(C_k)\geq N\times\min\_support$) (n 为频繁项集限定层数)

(8) $k=k+1$

(9) $C_k=F_{k-1}\oplus G_j(\forall i_p,i_q,i_r\in C_k(p<q<r,i_m\in\{i_p,i_q,i_r\}$与$i_1\in F_1(1<p)$ 且$((i_m,G_j)\in Index[]$

(10) $F_k=\{f_k\in C_2\wedge\sigma(f_k\times\min\_supprot\}$

(11) *end* while

(12) *return* F_k

本文将采用 IIFA 算法得出三个不同的数据挖掘模型；第一个初始模型处理安全令牌及其相关的攻击。

第二、三个模型用于处理数字签名和加密算法及其引起的攻击。三个模型极其相似，下面将对其中一个进行详细介绍，挖掘模型主要包括 3 个表：SOAP 消息分析表、令牌攻击解决表、令牌信息表。SOAP 消息分析表存储 SOAP 消息长度、SOAP 消息解析时间、请求开始时间戳、请求结束时间戳等信息。令牌攻击解决表存储 SOAP 消息及消息引起的攻击等信息。令牌信息表存储系统中使用的安全令牌等信息。为了预测可能危害服务的攻击，利用收集的数据测试根据安全令牌产生的关联规则，下面列出了挖掘模型产生三条关联规则：

Username = Existing ,Parsing Time = 512 - 1210ms ,SOAP Length = 8910 - 11911bytes

-> Username = Message Rewriting

Kerberos = Existing SOAP length = 5191 - 7109 bytes Password = right

-> Kerberos = XML Injection Attacks

X.509 = Existing SOAP Length = 13421 – 16001 bytes

-> X.509 = DoS Attacks

每条规则都包含概率和重要性等因素。第一条规则显示，服务器利用用户名/密码作为安全令牌验证服务请求者身份，服务器处理进程衡量 SOAP 消息解析时间在 512 - 1210ms 之间，SOAP 消息长度在 8910 - 11911bytes 之间，这样的消息可能引起消息重写攻击。第二条规则是 SOAP 消息内部使用 Kerberos 证书，消息长度在 5191 - 7109 bytes 之间，这样的请求信息可能引起 XML 注入攻击，第三条规则是 SOAP 消息内部使用 X.509 安全令牌，消息长度在 13421 – 16001 bytes 之间，这样的请求信息可能引起 DoS 攻击。

挖掘模型由概率与重要性两个重要的因素组成，概率表示请求消息引起攻击的可能性，重要性表示产生规则的 SOAP 消息的重要指标。例如，第一条规则可能性为 0.5，重要性为 1.85，安全管理员设定允许或是拒绝请求 SOAP 消息的可能性阀值。如果对一个服务攻击的预测的可能性大于之前定义的安全阀值，安全服务模型将发出一个警告，此次请求可能引起攻击。

预测进程通过 Web 服务代码动态的运行，执行 SQL 语句根据发布的安全令牌和预先定义的阀值预测可能的攻击。

```
slelect Ast.ReqSOAPNum,Ast.ParsingTime,
Ast.ReqSOAPLength,Predict(ReqAttackinfo.CertificateN
ame,ReqAttackinfo.AttackName,ReqAttackinfo.Support,
ReqAttackinfo.Probability as AttackPredition
from SOAPMessageAnalysis Ast
join (select ReqSOAPNum,ParsingTime,
ReqSOAPLength from SOAPMessageAnalysis
where ReqSOAPNum = '20090909') as st
on Ast.ParsingTime = st.ParsingTime
and Ast.ReqSOAPLength = st.ReqSOAPLength;
```

图 3 中显示了编号为 20090909、解析时间为 6189ms、消息长度为 9001 字节的 SOAP 消息请求，它可能引起三个依赖于用户名/密码、Kerberos 证书、X.509 安全令牌的攻击，用户名/密码被怀疑可能一起消息重写攻击、Kerberos 容易遭受 XML 重写攻击、X.509 安全令牌容易遭受 DoS 攻击。但是，攻击的预测是依靠预先定义的概率与支持率，如图，新的 SOAP 消息引起消息重写攻击的概率为 0.12391，引起 XML 注入攻击的概率为 0.12981，引起 DoS 攻击的概率是 0.18911，都小于预先定义的阀值 0.4，因此，服务器将允许消息通过。

ReqSOAPNum	ParsingTime	ReqSOAPLength	AttackPrediction
20090909	6819	9001	

AttackPredition

CertificateName	AttackName	Support	Probability
UserName	Message Rewriting	12	0.12391
Kerberos	XML Injection Attacks	10	0.12981
X.509	DoS Attacks	11	0.18911

图 3 实验查询结果

随着 SMDB 中记录的增长和服务器管理员的经验越来越丰富，FP-树频繁集算法将产生更加精确的关联规则。服务器管理员能够更加精确的分析请求 SOAP 消息，能够提出更加有效的安全方案。安全模型预测的精确性如图 2 所示。图表包括两行，其中实线表示理想精度，虚线表示规则模型精度。服务器需要 90%的数据到达 75%的精度。因此，发布的利用 FP-树频繁集算法产生的挖掘模型能够预测 75%的服务攻击。

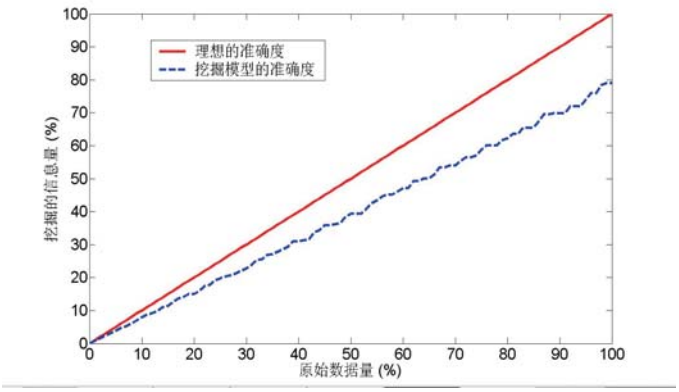

图 2 关联规则预测准确度

4.3 验证安全策略

服务提供者在策略单独实施之前利用与预测攻击相同的方法对其进行验证，验证可以脱机进行，但是必须在 Web 服务运行期间收集请求信息。因此服务提供者可以脱机进行多次测试以到达配置最理想的安全策略。因此，服务提供者可以借鉴从过去安全模型的经验，这可以帮助他们为 SOA 安全建立更可靠的安全策略。

5 总结

本文提出了一种新的基于 IIFA-SOA 模型的安全服务，模型利用 SOA 环境接收到的请求 SOAP 消息运用 FP-树频繁集数据挖掘算法预测 Web 服务攻击，建议模型可以根据安全令牌、加密算法等安全特性指定服务攻击的类型。当安全服务接收到 SOAP 请求消息时，依靠根据 SOAP 消息长度、解析时间、加密算法等特性产生的数据挖掘模型判断接收消息还是拒绝消息，安全服务接收到的 SOAP 请求消息越多，则预测结果将越准确。

此外，数据挖掘模型能够用于验证 WS-SecurityPolicy 管理的新的安全策略。本文提出的安全模型通过反复配置验证安全令牌、数字签名等安全属性使安全服务受到攻击的概率达到最低。

本文提出的安全模型的准确性、可执行性、优化特性及功能性需要在商业应用环境中进行验证，使安全模型达到更好预测效果需要获得大量有效的训练数据。

参考文献

1 Igor Sedukhin. End-to-End Security for Web Services and Services Oriented Architectures, White Paper, March 2003.
2 WS-SecurityPolicyV1.0,http://www.oasisopen.org/committees/download.php/15979/oasis-wssx-wssecuritypolicy-1.0.pdf.
3 Gunnar Peterson. Service Oriented Security Architecture. Information Security Bulletin, Nov. 2005,10:325－330.
4 ThomasErl. Service-Oriented Architecture–Concepts and Design. Pearson Education, Inc, 2005.
5 Oracle SOA Suite Developer’s Guide. http://www.oracle.com/lang/business _intelligence /data-mining.html.
6 WS-Policyv1.2, http://www.w3.org/Submission/WS-Policy/
7 WS-Authorization, http://www.ws-standards.com/ws- security.asp.
8 WS-Federation,ftp://www6.software.ibm.com/software/per/lib/develorary/wsfed. pdf.
9 Han JW, Micheline Kamber.数据挖掘概念与技术,机械工业出版社, 2007.146－181.
10 Web Service Security SOAP Messages with Attachments Profile..http://www.oasisopen.org/committees/download.php/10902/wss-swaprofile-1.0-&e=7152.
11 Chamberlin D, Florescu D, Robie J. Xuery 1.0: An XML Query Language.W3C Working Draft, 2001:35－40.
12 Mohammad Ashiqur Rahaman, Maarten Rits, Andreas Schaad. An Inline Approach for Secure SOAP Requests and Early Validation. OWASP Europe Conference, 2006.
13 Huang Y, Kumaran S, Chung J. A service management framework for service-oriented enterprises. Proc. of the IEEE International Conference on E-Commerce Techonlogy. Beijing, China, 2004. 181－186.
14 Douglas R. Stinson, Journal of Combinatorial Design, 密码学原理实践.电子工业出版社, 2008.1.3.
15 Hany F, Yamany EL, Miriam AM. Capretz Use of Data Mining to Enhance Security for SOA 2008 IEEE. [doi: 10.1109/ICCIT. 2008.173]
16 Agrawal R, Srikant R. Fast algorithms for mining association rules. Proc. of VLDB94, 1994(9):487－499.
17 朱彦霞,张雪萍,王家耀.改进的频繁项集挖掘算法.计算机工程与应用, 2009,45(4):143.
18 李洪奇,武装.基于 SOA 的企业应用集成.微型计算机, 2010.1－3.